



# **GeoMatch Algorithm Technical Documentation**

Partnership with the Dutch Central Agency for the Reception of Asylum Seekers (COA)

Prepared for: Deloitte Risk Advisory BV (Deloitte)

Prepared by: GeoMatch research team at the Immigration Policy Lab (IPL)

Last updated: Feb 16, 2024

|                                                                                                      |           |
|------------------------------------------------------------------------------------------------------|-----------|
| <b>1. Overview of the GeoMatch Algorithm</b>                                                         | <b>5</b>  |
| <b>2. Stage 1: Modeling</b>                                                                          | <b>5</b>  |
| a. General goal of the modeling stage                                                                | 5         |
| b. Technical details                                                                                 | 6         |
| i. Choice of Machine Learning tool                                                                   | 6         |
| ii. Validation techniques used                                                                       | 8         |
| iii. Model and parameters used in the backtests ran for COA                                          | 8         |
| iv. Need for statistical overlap                                                                     | 9         |
| <b>3. Stage 2: Mapping</b>                                                                           | <b>10</b> |
| <b>4. Stage 3: Matching</b>                                                                          | <b>10</b> |
| a. General goal of the matching stage                                                                | 10        |
| b. Technical details                                                                                 | 11        |
| Geographic balance in backtest                                                                       | 13        |
| c. Ongoing research and options for GeoMatch implementation                                          | 14        |
| i. Fairness options                                                                                  | 14        |
| ii. Allocation/load balancing options                                                                | 15        |
| iii. Incorporation of clients' preferences                                                           | 16        |
| iv. Incorporation of multiple outcomes                                                               | 17        |
| <b>5. Explaining GeoMatch</b>                                                                        | <b>18</b> |
| a. Explaining GeoMatch: general explanation                                                          | 18        |
| b. Quantifying the relative strength of each LMR: the Employment Impact Score                        | 19        |
| c. Explaining the GeoMatch's individual predictions: Variable Importance Plots                       | 20        |
| <b>6. Dutch Context: description of the data sources, target population, predictors and outcomes</b> | <b>20</b> |
| a. Target Population                                                                                 | 21        |
| b. Impact on out-of-scope population                                                                 | 21        |
| c. Data Sources and Data Quality                                                                     | 22        |
| d. Choice of outcome                                                                                 | 23        |
| e. Predictors                                                                                        | 24        |
| f. Model retraining                                                                                  | 24        |
| g. Estimation of gains: backtest and pilot                                                           | 25        |
| i. Employment gains                                                                                  | 25        |
| ii. Financial gains                                                                                  | 26        |
| iii. Beyond backtest: RCT and pilot                                                                  | 26        |
| h. Downstream effects on other outcomes                                                              | 26        |
| i. Evaluating the pilot: prospective study                                                           | 27        |
| j. Engaging users: continuous feedback loop                                                          | 28        |
| <b>7. References</b>                                                                                 | <b>28</b> |
| <b>8. Appendices</b>                                                                                 | <b>29</b> |
| 1. Appendix 1: Model Performance                                                                     | 29        |

|                                                                                                                                                                                                      |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| a. Model Performance on multiple outcomes                                                                                                                                                            | 29 |
| Performance metrics, GBM model                                                                                                                                                                       | 29 |
| Calibration Plot - Probability of finding work in year 1                                                                                                                                             | 30 |
| Calibration Plot - Percent of time pursuing education in year 1                                                                                                                                      | 31 |
| Calibration Plot - Percent of time with work in first 2 years                                                                                                                                        | 31 |
| Calibration plot - Wages in first two years                                                                                                                                                          | 32 |
| b. Model performance: comparison with alternative methods                                                                                                                                            | 32 |
| Performance Metrics, alternative methods                                                                                                                                                             | 32 |
| Logit model calibration plot                                                                                                                                                                         | 33 |
| OLS calibration plot                                                                                                                                                                                 | 33 |
| 2. Appendix 2: Data Quality Checks                                                                                                                                                                   | 34 |
| a. Type of each variables used in backtest                                                                                                                                                           | 34 |
| b. Missing data on key variables                                                                                                                                                                     | 35 |
| c. Missing data on outcome variables                                                                                                                                                                 | 36 |
| Missing outcome variable proportions                                                                                                                                                                 | 36 |
| d. Descriptives on number and type of cases across years                                                                                                                                             | 37 |
| Filtering variables for GeoMatch backtest and pilot                                                                                                                                                  | 37 |
| Descriptive summaries, type and number of cases across years                                                                                                                                         | 38 |
| All tables below will filter out minors and individuals with family reunifiers and hard criteria. The additional filters applied for each specific table will be specified in the table description. | 38 |
| e. Descriptives, outcomes of SM interview over the years                                                                                                                                             | 38 |
| Descriptive summaries, PS regional advice                                                                                                                                                            | 39 |
| Descriptive summaries, PS regional advice - AA vs VA                                                                                                                                                 | 39 |
| Descriptive summaries, PS advice and final linking                                                                                                                                                   | 39 |
| Descriptive summaries, PS advice and final linking                                                                                                                                                   | 39 |
| Compliance with PS recommendation, by month of linking                                                                                                                                               | 40 |
| Percentage of PS recommendations in UR of first AZC                                                                                                                                                  | 41 |
| f. Descriptives, PLEs' journeys at AZCs                                                                                                                                                              | 41 |
| Descriptive summaries, events while at AZC (all cases with housing interview)                                                                                                                        | 41 |
| Descriptive summaries, events while at AZC (cases with at least 1 PS rec)                                                                                                                            | 41 |
| Most common pairwise combinations of AZCs for PLEs going through at least 2 AZCs                                                                                                                     | 43 |
| First AZC and PS recommendation                                                                                                                                                                      | 44 |
| Second AZC (or beyond) and PS recommendation                                                                                                                                                         | 44 |
| Full journey summary, cases with housing interview                                                                                                                                                   | 45 |
| g. Descriptives, dates checks                                                                                                                                                                        | 46 |
| Summary distribution linkings/day and gap between SM and linking                                                                                                                                     | 47 |
| Summary distribution linkings/day and gap between SM and linking, AA                                                                                                                                 | 47 |
| Summary distribution linkings/day and gap between SM and linking, VA                                                                                                                                 | 48 |
| h. Covariate imbalances in historical data                                                                                                                                                           | 48 |

|                                                                                                           |    |
|-----------------------------------------------------------------------------------------------------------|----|
| i. Proportion of hard criteria to non hard criteria individuals linked to each LMR every month            | 49 |
| Linkings per bimonth, 2015/2020                                                                           | 50 |
| Linkings per month, 2016/2017 (top 20 LMRs)                                                               | 51 |
| Cumulative linkings, 2020                                                                                 | 53 |
| j. Length of stay at AZC and employment                                                                   | 53 |
| Distribution summaries, length of stay at any AZC (days)                                                  | 54 |
| Employment based on total time spent in AZC, AA vs VA                                                     | 55 |
| k. Local linkings                                                                                         | 55 |
| UR level                                                                                                  | 56 |
| Percentage of local linkings, local = same UR                                                             | 56 |
| Percentage of local linkings, among cases having received at least 1 PS advice                            | 56 |
| Percentage of local linkings, among cases having received a blanco advice from PS                         | 57 |
| Percentage of local linkings, among cases having received no SM interview                                 | 57 |
| LMR level                                                                                                 | 58 |
| Percentage of local linkings, local = same LMR                                                            | 58 |
| Percentage of local linkings, among cases having received at least 1 PS advice                            | 59 |
| Percentage of local linkings, among cases having received a blanco advice from PS                         | 59 |
| Percentage of local linkings, among cases having received no SM interview                                 | 59 |
| Muni level                                                                                                | 60 |
| Percentage of local linkings, local = same Muni                                                           | 60 |
| Diversity of LMRs                                                                                         | 61 |
| Diversity of LMRs linkings at each AZC, AA and VA (34% of the combinations remain)                        | 61 |
| Diversity of LMRs linkings at each AZC, AA (10% of the combinations remain)                               | 61 |
| Diversity of LMRs linkings at each AZC, VA (30% of the combinations remain)                               | 62 |
| Employment based on whether the case was linked to local LMR or not (ie same LMR as the AZC they were in) | 63 |
| l. Nationality distribution                                                                               | 63 |
| Most common nationalities since 2014                                                                      | 64 |
| m. Secondary migration                                                                                    | 64 |
| 3. Appendix 3: Variable Importance Plots                                                                  | 65 |
| Achterhoek                                                                                                | 66 |
| Amersfoort                                                                                                | 67 |
| Drechtsteden                                                                                              | 68 |
| Drenthe                                                                                                   | 69 |
| Flevoland                                                                                                 | 70 |
| Food Valley                                                                                               | 71 |
| Friesland                                                                                                 | 72 |
| Gooi- en Vechtstreek                                                                                      | 72 |

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| Gorinchem                                                          | 74  |
| Groningen                                                          | 75  |
| Groot Amsterdam                                                    | 76  |
| Haaglanden                                                         | 77  |
| Helmond-De Peel                                                    | 78  |
| Holland Rijnland                                                   | 79  |
| Midden-Brabant                                                     | 80  |
| Midden-Gelderland                                                  | 81  |
| Midden-Holland                                                     | 82  |
| Midden-Limburg                                                     | 83  |
| Midden-Utrecht                                                     | 84  |
| Noord-Holland (Noord)                                              | 85  |
| Noord-Limburg                                                      | 86  |
| Noordoost-Brabant                                                  | 87  |
| Regio Zwolle                                                       | 88  |
| Rijk van Nijmegen                                                  | 89  |
| Rijnmond                                                           | 90  |
| Rivierenland                                                       | 91  |
| Stedendriehoek en Noordwest Veluwe                                 | 92  |
| Twente                                                             | 93  |
| West-Brabant                                                       | 94  |
| Zaanstreek – Waterland                                             | 95  |
| Zeeland                                                            | 96  |
| Zuid-Holland Centraal                                              | 97  |
| Zuid-Kennemerland en IJmond                                        | 98  |
| Zuid-Limburg                                                       | 99  |
| Zuidoost-Brabant                                                   | 100 |
| 4. Appendix 4: Model Counts for Backtest                           | 101 |
| Count of individuals included in the backtest training data by LMR | 101 |
| Count of individuals included in the backtest test dataset         | 102 |

# 1. Overview of the GeoMatch Algorithm

The GeoMatch algorithm (referred to as “the algorithm” throughout this document) relies on three stages: modeling, mapping, and matching. This document summarizes the key elements of each of the three stages. It also provides updates on ongoing research aimed at improving the algorithm and offering more options to partners. Finally, we explain the specificities and decisions that were made while conducting backtests and designing a pilot in the specific context of our partnership with the Dutch Central Agency for the Reception of Asylum Seekers (COA).

The technical foundations of the GeoMatch algorithm were described in a 2018 *Science* paper (Bansak et al. 2018<sup>1</sup>). For additional documentation, please refer to the [Supplementary material](#) and the [replication code](#) for the 2018 *Science* Paper on this algorithm within this file.

## 2. Stage 1: Modeling

### a. General goal of the modeling stage

The modeling stage involves a supervised machine learning process that predicts the expected success for any quantifiable metric - for example, employment after a year - of new refugee arrivals across all possible resettlement locations. We designate historical resettlement data for model training, in which the unit of observation is a single refugee and which contains information on the refugees’ background characteristics (e.g., country of origin, language skills, gender, age, etc.), time of arrival, assigned location, and measured employment success.

This training data is then used to build a bundle of supervised learning models that predict the relevant integration metric chosen by the partner (for example, employment success) as a function of their background characteristics. A separate model is fit for subgroups of refugees assigned to each location, thus yielding different models for each location and allowing for the discovery of refugee/location synergies. These fitted models are then applied to new, out-of-sample refugee arrival

---

<sup>1</sup> Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373), 325-329.

data to predict the expected employment success of each new arrival at each possible resettlement location.

## b. Technical details<sup>2</sup>

In the modeling stage of the algorithm, the model training data is used to build a bundle of statistical models that predict refugees' probability of the integration outcome (or value of the outcome, if continuous outcomes were to be used), and those models are then applied to the out-of-sample prediction data to generate their predicted probabilities.

The goal of the algorithm is to discover synergies between refugee characteristics and resettlement locations such that refugees can be matched to the resettlement locations that best fit their own personal profiles. Thus, it is not sufficient to simply achieve a model that is highly predictive of the integration outcome in general; instead, the modeling stage must also prioritize interactions between location recommendation and refugee characteristics if it is to provide useful information to generate optimized recommendations. For this reason, we train different models for each location, which allows us to identify different synergies in different locations.

In the Dutch case, locations are Labor Market Regions (LMRs). There are 35 LMRs in the Netherlands. Specifically, for each resettlement location, the model training data is first subsetting to those refugees who were recommended to that location, and a machine learning model is then fitted that uses those refugees' characteristics to predict their integration outcome. That fitted model is then applied to the out-of-sample data to predict the probability of the integration outcome for the new refugee arrivals, should they be sent to the location in question. This process is performed separately and independently for each individual location, which yields for each individual refugee in the prediction data a vector of predicted probabilities, one for each of the 35 locations. Collectively for all refugees in the out-of-sample data, the final result is a matrix of predicted probabilities with rows representing individual refugees and columns representing resettlement locations.

### i. Choice of Machine Learning tool

Modeling and estimation of the predicted probabilities should ideally be undertaken using a flexible supervised machine learning technique that automatically performs feature selection and can identify complex non-linearities and interactions in the

---

<sup>2</sup> For further details, please refer to the Supplemental Materials for the *Improving refugee integration through data-driven algorithmic assignment* 2018 Science Paper, included in the files sent with this document (page iii of the Supplemental Material)

feature space, with cross-validation employed in the model training in order to achieve flexibility without over-fitting. Possible machine learning techniques for modeling and estimating include boosted trees, random forests, elastic-net logistic regression, and kernel-based regularized least squares.

Two primary criteria are used to inform which technique(s) are considered and then actually employed in our implementation of the modeling stage. The first, and most important, is predictive performance. For testing purposes, we can use historical data for both the model training data and the test data, thereby allowing us to empirically assess out-of-sample model fit for each machine learning technique, as described below.

The second criterion is the ability to handle many predictors (both continuous and categorical) and perform implicit variable selection. The refugee data contain many features that can be considered as possible predictors, yet the degree of predictiveness offered by each feature will be unknown in advance and may vary across resettlement locations and across time in ways that are difficult to anticipate, even through a theory-driven approach. This issue is exacerbated by the fact that feature non-linearities and interactions of unknown order may also provide significant predictiveness. Furthermore, there are also certain key interactions that must be considered, namely interactions between case type and all other variables, since the prediction data will only pertain to free cases (in the Netherlands, cases with no hard criteria) while the model training data will include all case types. Thus, the machine learning technique must be capable of efficiently exploring a high-dimensional feature space and automatically identifying predictive features and feature interactions while ignoring or remaining robust to irrelevant features. This is particularly important for resettlement locations with sparser historical data in order to guard against overfitting to too many, and possibly irrelevant, features.

In general, we only include variables that have been shown quantitatively or qualitatively to potentially be relevant to the integration outcome of interest. The variable importance plots provided in the Appendix 3 of this document are one of the sources used to select the relevant predictors. Note that because we train a different model for each Labor Market Region, it is expected that different predictors will matter in different models. We still have to include the full range of predictors in all the models, since it is impossible to know which predictor will matter in which model in advance, and the predictors that matter in each LMR can change overtime. Finally, the variable selection may be further restricted by data availability and legal constraints, such as GDPR rules around the use of sensitive variables (such as religion and ethnicity). Besides this initial feature selection, we rely on the algorithm to select the variables that matter most in each location and each period. Any additional manual feature selection would inhibit the algorithm's ability to learn changes that occur in synergies across time and/or locations.

## ii. Validation techniques used

As is general practice in supervised machine learning, a number of validation techniques and metrics should be used in both model building and model assessment. In particular, for any technique involving tuning parameters, cross-validation can be used to select the parameter values. In addition, as mentioned above, backtests that use separate subsets of historical data for the model training and test data can be performed, which allows for the assessment of out-of-sample model fit for each machine learning technique. A number of model fit metrics can be used, including mean squared error and percent reduction in error given a particular classification threshold.

The modeling stage also includes the option of calibrating the predicted probabilities. Given binary integration outcome variables of interest, classification machine learning models should be used. While model fit is often most easily assessed via some metric of classification accuracy, the ultimate outputs of interest of the fitted models for our algorithm given a binary outcome variable are not classification predictions, but rather class probabilities (i.e. predicted probabilities of a positive integration outcome). As has been shown by previous studies in the literature on supervised learning, classification models can sometimes output unreliable or biased predicted class probabilities even when they offer respectable classification accuracy. While some studies have provided evidence on which classification techniques, by virtue of their mathematical and algorithmic features, are likely to yield unreliable predicted probabilities and in what ways any bias is likely to manifest, the empirics and theory establishing general results are limited in this regard, and the problem is likely to be context- and data-dependent. Nonetheless, calibration plots can be used to assess the possibility of unreliable predicted probabilities outputted from any supervised learning model, and various calibration methods have been introduced to correct for any observed bias. Two common calibration methods, which are considered for use in our algorithm, are Platt scaling and isotonic regression. Appendix 1 to this technical document shows the relevant performance metrics used for model evaluation in the Dutch context, as well as comparison performance metrics for non-ML alternative methods (OLS, logit regression).

## iii. Model and parameters used in the backtests ran for COA

We assessed various machine learning techniques that can efficiently handle mixed data and remain robust in the presence of many and possibly irrelevant predictors, including gradient boosted trees, random forests, elastic-net logistic regression, and kernel-based regularized least squares. We found gradient boosted trees to exhibit

the best performance on the basis of out-of-sample classification accuracy (percent reduction in error) and predicted probability reliability (calibration plots). Specifically, we used stochastic gradient boosted trees (bag fraction of 0.5) with a binomial deviance loss function, which we implemented in R using the GBM package.

Gradient boosted trees have a number of practical characteristics that make them well-suited for use in our algorithm, namely their robustness to irrelevant predictors, automatic variable selection, and ability to discover complex interactions without requiring researcher specification. In addition to the well-known predictive performance of boosted trees, they also have attractive theoretical characteristics. In particular, it has been shown that given i.i.d. samples of a binary outcome and a feature vector that predicts the outcome, boosted tree ensembles achieve consistency and converge to the minimal error rate (Bayes risk) in the large sample limit given reasonable regularity conditions and appropriate regularization techniques, such as early stopping.

We used cross-validation within the training data to select tuning parameter values, including the interaction depth, learning rate, and number of boosting iterations (the early stopping point) in our implementation of gradient boosted trees. Parameters were tuned independently for each location-specific model, which means we find and use the ‘best’ parameters for each of our 35 models. In addition, on the basis of our model fit metrics, we also found model performance to be best without the use of probability calibration methods. The following parameters were used in the backtests: the vector of tree depths to be considered in cross-validation was set to `c(3,4,5,6,7,8)`, the maximum number of boosting iterations to be considered in cross-validation was set to 600, the shrinkage rate was set to 0.01, and the minimum number of observations in the trees terminal nodes was set to 5.

#### iv. Need for statistical overlap

Regardless of the demographic composition of the Labor Market Regions in the training data, the algorithm can draw inferences from small groups, and automatically weights characteristics accordingly without the need to manually add weights. However, the algorithm still requires statistical overlap in order to generate predictions for each combination of predictors. If certain groups are entirely absent from certain Labor Market Regions, the algorithm would be unable to learn anything about the synergies between the missing group and the LMR. We have not encountered such an issue yet in the data and backtest, but the monitoring plan includes monitoring the level of statistical overlap in all regions during the pilot. Moreover, technical innovations IPL is working on, such as load balancing measures, could further decrease this risk. When models are trained, we review the data for the

latest year of PLEs that have entered the COA system to ensure that we do not see sudden compositional changes that we have not accounted for in our models. We have seen annual fluctuations in the share of some demographic characteristics (e.g., increase in the share of Syrian asylum seekers), but we have not seen a sudden shift that would violate the requirement for statistical overlap.

### **3. Stage 2: Mapping<sup>3</sup>**

As described above, the final output of the modeling stage is a matrix of refugee-location predicted probabilities, with rows representing the individual refugees in the prediction data and columns representing the resettlement locations under consideration. However, the information in this form is not immediately usable for the matching stage of the algorithm because refugees are typically not recommended to locations on an individual basis but rather on a case-level basis, where cases are most often family units.

Thus, a case-level metric must be constructed. In other words, refugee-location predicted probabilities must be mapped to a case-location metric, where the metric is some mapping function of the predicted probabilities. Specifically, for each case-location pair, the mapping function is applied to the refugee-location predicted probabilities for all refugees belonging to that case, yielding a single value for that case-location pair. This results in a new matrix with the same number of columns (locations) as previously but now as many rows as cases rather than refugees.

Various mapping functions can be used, such as the average or maximum probabilities among case members. Our preferred case-level metric, and the one used in our backtest for COA, is the predicted probability that at least one refugee in the case would find employment at the location in question.

### **4. Stage 3: Matching**

#### **a. General goal of the matching stage**

---

<sup>3</sup> For further details, please refer to the Supplemental Materials for the *Improving refugee integration through data-driven algorithmic assignment* 2018 Science Paper, included with this technical document (page vii of the Supplemental Material).

The matching stage involves recommending each case to a specific location to fulfill a chosen optimality criterion subject to constraints. Our algorithm is flexible and can accommodate multiple criteria and constraints. The optimality criterion we usually use in our applications is to maximize the average of the case-level metric (i.e., the global average of the probability that at least one refugee in each family gains employment). We also impose constraints that represent real-world recommendation restrictions, such as how many cases can be assigned to different locations.

In the Dutch case, the two constraints we impose on the algorithm are (i) status holders must be assigned across labor market regions according to predetermined capacity and proportionality guidance. This includes the status holder housing quotas across municipalities (i.e. “taakstelling”), and (ii) permit holders in the extended asylum (VA) procedure have to be assigned only to LMRs that are in the same allocation region (UR) as their AZC<sup>4</sup>.

## b. Technical details<sup>5</sup>

The final stage of the algorithm aims at providing recommendations to specific locations for each case as they arrive. This is an *online assignment problem* because cases must be assigned as they arrive, without knowledge of future arrivals. The method that we use to solve this online problem is based on the solution to the *static assignment problem*. In the static assignment problem, we know all of the future arrivals within the relevant cohort we are trying to optimize over (a year worth of arrivals, for example, or any policy relevant period such as a taakstelling period in the Netherlands case). When we know all of the cases that will arrive, we can solve for the optimal assignment using mathematical optimization software. The optimal assignment is the assignment of cases to locations that maximizes total (or average) employment, or any chosen metric of interest.

The optimization method we use also provides us with the necessary information to determine a full ranking of locations for every case. Each time the static assignment problem is solved, it outputs a *dual variable* for each region. Dual variables are a concept used in the linear optimization field. They measure the estimated marginal cost (in terms of total employment) of removing one slot from each region. In other words, the dual variable estimates the cost for future cases of placing a case in one LMR, since there is now one less slot available for future cases at this LMR. With this information, along with the estimated employment probabilities for the current case,

---

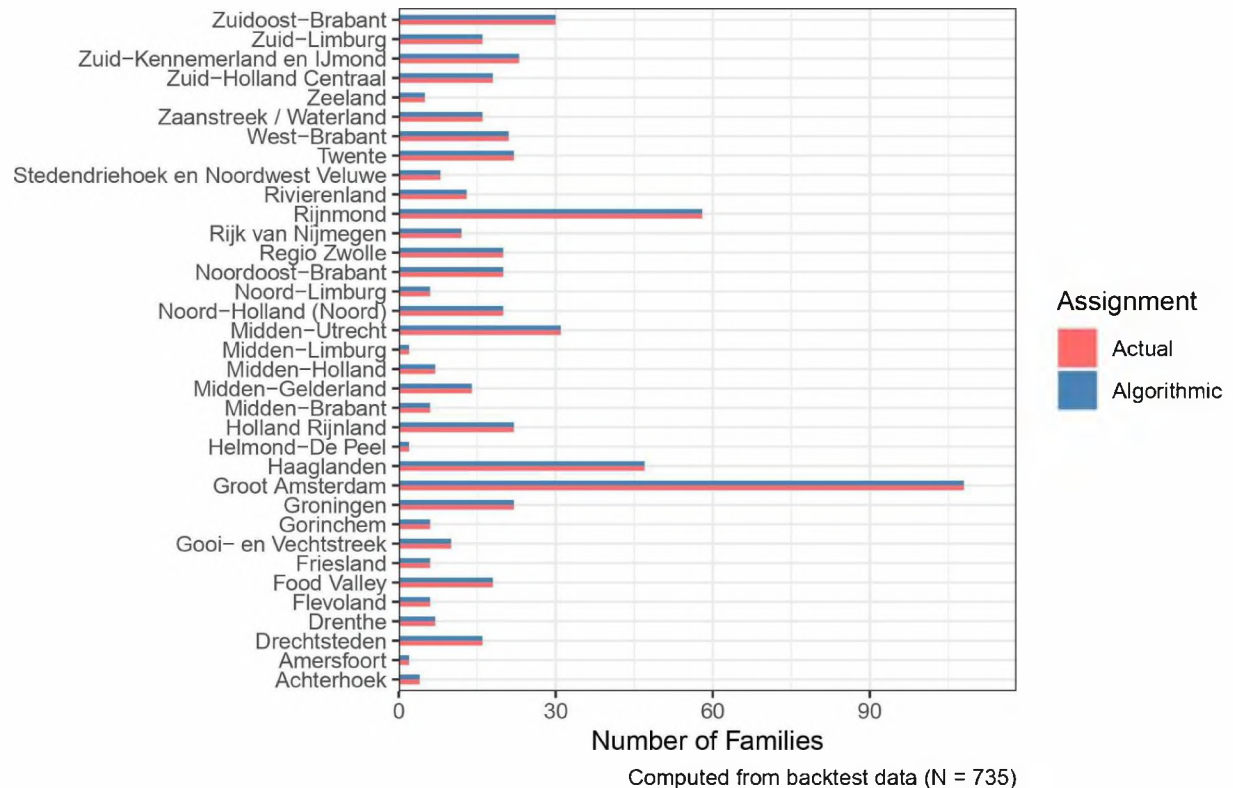
<sup>4</sup> This reflects current COA policy. Should the policy change and allow VA cases to be placed in any UR, the GeoMatch algorithm would be immediately updated to reflect the new policy

<sup>5</sup> For further details, please refer to the Supplemental Materials for the *Improving refugee integration through data-driven algorithmic assignment* 2018 Science Paper, included with this technical document (page ix of the Supplemental Material)

we can compute the location score for each region as a function of the difference between the estimated current case's employment probability at the location and a dual variable that estimates the value of a slot at each region. The higher the current case's employment probability at the location, and the lower the region's dual variable, the higher the location will be ranked for this case.

Prior to actually performing the assignment according to the chosen optimality criterion, however, constraints must be built into the process (see above). The constraints primarily involve limiting the number of assignments to each location and prohibiting certain case-location pairs (in the Dutch context for example, for extended asylum process cases, we prohibit locations outside of the relocation region of the AZC). From the implementation standpoint, such constraints can be easily built into the process by pre-specifying the number of available assignment slots for each location, and by setting the case-location metric to an arbitrary low value for the prohibited case-location pairs. The optimal static matching finds the assignment that maximizes the average employment score while respecting the constraints.

Through these constraints, the algorithm can ensure that any geographic balance goal (taakstelling in the Netherlands case) is fully reached. In the backtest, for example, we assumed that each Labor Market Region should receive the exact same number of cases under the new algorithmic assignment than they received in reality (in the actual assignment). The plot below shows that each LMR received the same number of cases under both assignments (actual and GeoMatch) in the backtest. This is a feature of the algorithm. The gains in employment were therefore achieved by linking different cases to each LMR, not by changing how many cases each LMR received. In a pilot, the algorithm will fully respect the target numbers for the number of cases each LMR will receive, as provided by COA through their taakstelling targets. The respect of these targets is a built-in feature of the algorithm.



## ***Geographic balance in backtest***

In practice, refugees and asylum seekers must often be assigned to a locality virtually immediately upon being processed by resettlement authorities, without knowing who will arrive in the future. As a result, each arriving refugee or asylum seeker case is typically assigned in an online, dynamic, fashion, and these assignments cannot be reversed. The dynamic aspect of this problem introduces a key trade-off between immediate and future rewards: assigning a current case to a location results in an immediate reward (namely, the employment score of the current case at that location), but also uses up a slot at that location for the unknown future arrivals.

Therefore, for every case that arrives, we simulate future arrivals. In each simulation, we choose cases in the historical data uniformly at random and act as if they were part of the current cohort. For each simulation, we can solve the static assignment problem described above. Each solution of the static assignment problem maximizes the total (or average) employment level for the current case in addition to the simulated future cases. For each case that arrives, we solve the static assignment problem with simulated future arrivals between 5 and 10 times. For each of these simulations, we look at the location that the current case is assigned to. We then recommend that the current case be assigned to the location that they were most

often assigned to in these simulations. The locations ranked 2 and 3 are determined using the dual variables described above.

The intuition for this method is the following: if the current case is assigned to location A in all (or most) of the simulations, then location A is the best location for the current case with high probability. Since we do not know who will arrive in the future, we cannot be 100% sure that the assigned location will be the best, but using these simulations allows us to find the location that has the highest probability of being the best.

The proposed method can be extended to settings with batching (i.e., the setting where multiple cases are assigned together). Batching only makes the problem easier, and any online algorithm can be extended to settings with batching by ignoring the batches and assigning members of a batch one by one. However, certain methods also allow us to take advantage of the additional information obtained under batching to improve performance. In the Dutch case, arrivals are assigned one-by-one and therefore we do not use batching in this setting.

These methods are described in more detail in Bansak & Paulson 2022<sup>6</sup>.

## c. Ongoing research and options for GeoMatch implementation

### i. Fairness options

We can think of fairness from two different perspectives: the permit holders and the welcoming municipalities. From the perspective of groups of permit holders, there are multiple possible definitions of fairness. Perhaps the most salient definition of fairness is that every group of people (based on specified covariates such as education or country of origin) are better off after using GeoMatch than before. For example, we want to make sure that both low educated and high educated permit holders are benefited by using GeoMatch. In the backtests we have run so far in several contexts and countries, we have not found any group to be worse off from using GeoMatch than compared to the status quo. Moreover, the gains are relatively evenly balanced across groups. Another way to define fairness is through a benchmark based on a random allocation. Namely, we could compute the average outcome for each group if all cases were assigned uniformly at random to regions. In some sense, a random allocation is *a priori* fair, and thus any allocation that results in

---

<sup>6</sup> Bansak, K. Paulson, E. Outcome-Driven Dynamic Refugee Assignment with Allocation Balancing. Working paper. [Preprint](#).

all groups being better off than this benchmark is also fair. A third and stronger benchmark is the following: consider randomly allocating slots to each group (so if a group comprises 30% of all cases, they would also receive 30% of the slots at each location), and then allow each group to optimize their allocation within those slots. This is a strictly stronger fairness benchmark than the random allocation benchmark. For the Netherlands implementation, the monitoring plan provides details on the fairness metrics GeoMatch and COA agreed to track.

The predicted gains by subgroups are essential elements of the monitoring plan we build and consider when deploying pilots, and we would be able to detect if gains for some subgroups in the pilot start to drop below the benchmark set with our partners. If partners are interested in taking a more active approach to ensuring fairness in outcomes among especially relevant subgroups, we are developing new methods aimed at guaranteeing that these notions of fairness hold with high probability. These methods can be adapted to any chosen benchmark. This is achieved by modifying the static matching problem (described above) to include fairness constraints according to the chosen definition of fairness.

Another notion of fairness is from the perspective of the municipalities. Similarly to the fairness definition above, our goal could be to ensure that every location is benefited by using GeoMatch compared to the status quo, or to ensure that the gains for each group are above a random allocation benchmark. We monitor this fairness criteria by tracking the predicted (and actual) employment probabilities (or other outcome of interest) in every location. We are also actively developing new built-in methods for ensuring that every location has positive gains relative to the chosen benchmark, again by modifying the static assignment problem to include these fairness constraints. The monitoring plan includes more details on how we track fairness for the municipalities.

## ii. Allocation/load balancing options

*[This section describes a potential extension of the GeoMatch algorithm that is not part of the currently planned implementation in the Netherlands].*

The algorithm described above can still result in a periodically imbalanced allocation to localities over time, which can lead to undesirable workload inefficiencies for resettlement resources and agents. Many reasons can account for this periodic imbalance over time. Consider the non-stationarity of arrivals. If a batch of ‘similar’ cases arrives, imbalance will likely occur. Because refugee inflows are, in part, due to international events, there can be clustering of arrivals with specific background characteristics. Thus, it can be common for batches of ‘similar’ cases to arrive together, contributing to an imbalanced allocation. However, because each locality has a given amount of resources (e.g. number of resettlement officers, or temporary

housing) that cannot be transferred across localities, maintaining a steady workload is crucial and is a first-order concern of many host countries. These observations can motivate the use of an algorithm that actively balances the allocation over time to each location.

In order to address these potential inefficiencies and tensions on the resettlement systems, we have developed a new assignment algorithm that dynamically matches refugees and asylum seekers to localities within host countries in order to improve employment outcomes while maintaining a smooth balance on the load placed over time on the receiving localities (see Bansak & Paulson 2022<sup>3</sup>). Building on the algorithm described above and borrowing ideas from queuing theory, the allocation balancing algorithm incorporates imbalance minimization into the assignment process. This allows refugees to be dynamically assigned to localities in a way that improves their expected employment scores while also maintaining a balanced allocation across the localities over time.

We show in Bansak & Paulson 2022<sup>3</sup> that the allocation balancing algorithm is able to improve balance over time with very little loss in efficiency (i.e., employment gains). We introduce a single parameter which allows us to control the trade-off between allocation balancing and outcome maximization. Therefore, the policymaker can choose or tune this parameter to obtain the desired level of employment and allocation balance. In situations in which the payoff/cost of outcomes, wait time, and idle time can all be measured in or converted to a common metric (such as payoff/cost in dollars), policymakers might want to set the parameter to the specific value that leads to maximization of that common metric. Another approach would be to use empirical experiments to tune the parameter to a value that establishes an acceptable degree of balance across locations over time.

Finally, the allocation balancing algorithm also offers several ancillary benefits. In particular, the balancing algorithm provides an immediate solution to the issue of an unknown total number of arrivals. For example, when the final number of arriving refugees or asylum seekers ends up being lower than initial plans/projections, the final allocation could end up being far from the desired year-end allocation proportions, as quotas will be filled for certain localities but not others. In smoothing the allocation across localities over time, the balancing algorithm averts this problem. As an additional benefit, the balancing algorithm also helps to improve the resilience of the underlying learning system through greater exploration.

### iii. Incorporation of clients' preferences

*[This section describes a potential extension of the GeoMatch algorithm that is not part of the currently planned implementation in the Netherlands].*

In Acharya, Bansak and Hainmueller 2021<sup>7</sup>, we introduce a method to allow our partners to incorporate refugee preferences into the algorithm matching and recommendation. In this setting, the partner can set a threshold  $g$  for the integration metric of their choice: for example, the partner can choose a minimum employment rate that will constrain the algorithm in the locations that can be selected. In parallel, clients would list preferred locations (they do not need to list all locations, but can list as many as they want – all non-listed locations are treated as equally undesirable by the algorithm). The algorithm will then recommend the highest location on the client's preference list that satisfies two key constraints. First, there needs to be spots left in the location. Second, the average expected outcome score (for example, the average employment rate) can not fall below the threshold set by the partner.

This mechanism requires the partner to select a value for the outcome threshold,  $g$ , and this choice implies a tradeoff between an outcome-based and preference-based matching. From the partner's perspective, it is desirable to achieve the highest possible value of  $g$  to ensure that the clients' outcomes are optimized. However, setting a higher value of  $g$  comes at the cost of assigning clients to locations that are, in expectation, lower in their preference rankings. That is, while the mechanism simultaneously considers both outcomes and preferences, there is a trade-off between the two, where the balance of that trade-off changes as  $g$  increases. Moreover, not all values of  $g$  can produce a feasible matching. We can provide partners with  $g_{max}$ , the highest possible average outcome score that can be generated by a feasible matching.

#### iv. Incorporation of multiple outcomes

*[This section describes a potential extension of the GeoMatch algorithm that is not part of the currently planned implementation in the Netherlands].*

In general, when choosing the specific integration outcome variable to feed into the algorithm for optimization, any data-driven algorithmic decision-making procedure must balance several imperatives. The first is the goal of helping the clients (i.e. status holders) achieve long-term and lasting success. The second is the need to take shorter-term information into account, such that the algorithmic methods are resilient and can adapt to changes over time with higher velocity. Finally, the third is the need for an outcome with sufficient variation across status holders and locations for algorithmic recommendations to be viable from a statistical and mathematical perspective.

---

<sup>7</sup> Acharya, A., Bansak, K., & Hainmueller, J. (2022). Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. *Political Analysis*, 30(1), 89-112.

In cases where partners are interested in optimizing over multiple outcomes, several options exist. First, a composite outcome can be built by taking a weighted average of the outcomes the partner is interested in optimizing over. The algorithm would then use the single weighted average as the integration outcome of interest. Second, we could adapt the method described in Acharya, Bansak and Hainmueller 2021<sup>3</sup> (see above) to optimize on one outcome, but subject to meeting specific constraints or threshold on another outcome. This would allow our partners to set a minimum acceptable threshold on one outcome (for example, a minimum acceptable retention rate in the locations recommended by the algorithm) and let the algorithm optimize over another outcome (for example, expected employment gains at the locations recommended by the algorithm) under this new set of constraints.

## 5. Explaining GeoMatch

As described above, the full GeoMatch algorithm is composed of three stages: modeling, mapping and matching. This multi-stage approach mirrors our users' goal, which is not simply to predict permit holders' chances of employment, but to find the optimal location for them in a dynamic, constrained environment. This complex, multi-stage approach also means that we cannot entirely rely on conventional Machine Learning explanatory metrics, since at best they would only explain one part of the GeoMatch algorithm (the modeling and individual predictions), and not the overall results and recommendations. We are committed to provide as much explanation as we can about the GeoMatch recommendations to users, but cannot rely on existing tools to do so. The paragraphs below detail our general approach to explaining GeoMatch, which will be applied in the Dutch pilot.

### a. Explaining GeoMatch: general explanation

We always provide users with a general overview of what the algorithm does (such as finding synergies between individual characteristics, locations, and likelihood of employment by learning from historical data) and doesn't do (such as finding the best location for minors or based on non-employment considerations). This general overview can include an explanation of the mapping stage. Besides this general overview, two more specific explainability frameworks can be provided: insights about how the individual predictive probabilities of employment were obtained in the modeling stage (i.e., information about the variables that have a high predictive power and are most relevant in training the model for each model we trained), and insights about how the final GeoMatch recommendations were produced as a result of the matching stage.

## b. Quantifying the relative strength of each LMR: the Employment Impact Score

To meet our users' needs of receiving simple, clear, and useful information, we focus on providing them with the numbers most directly relevant to their goals and decision-making:

- the predicted individual probabilities of employment for the current case produced in the modeling stage,
- as well as the predicted total number of future jobs estimated after the matching stage.
- in contexts where load balancing considerations are incorporated into the GeoMatch algorithm (see methodological innovation above), a score detailing how far above or below its target each location is, the target being when the location would receive a case in a perfectly load balanced world. This load balancing score is customized to each partners' contexts and priorities.

This information directly informs which locations are recommended by the algorithm, and is most relevant to the users' decision-making task. Moreover, when and if users are presented with multiple recommendations, this information can empower them to weigh the trade-off between the individual case and the cohort's overall employment, or between employment and load balancing, while using their own insights about the current case. The exact information we provide in each context is decided in collaboration with our partners. In the Dutch pilot for example, IPL and COA decided to provide information about the current and future trade-offs to the PS, but not the AOs. As of April 2023, IPL and COA are still in discussions about whether load balancing will be included in the Dutch pilot, and about which information will be provided to users about load balancing. More details on load balancing implementation will be provided to Deloitte as part of the second part of the second audit.

We aggregate these three numbers into a normalized score out of 10, which we call the Employment Impact Score (EIS). For each LMR and case, this score provides a snapshot of the employment gains expected for both the case and the future arrivals in the taakstelling period, if the case is linked to this LMR. If load balancing is included in the algorithm, the EIS also provides a snapshot of how much each location currently 'needs' a case to satisfy the goal of staggered placements throughout the year. Because the EIS aggregates multiple components, including the individual prediction of employment for this case at this LMR, the expected total number of jobs for future arrivals under several simulated future arrival cohorts, and a custom load balancing score, no metric exists to compute a confidence level around the EIS. If we wanted to build a measure of uncertainty around the EIS, we would need to use a bootstrap statistical approach. However, such an approach

would require to bootstrap both the individual predictions, as well as the simulations of future arrivals in the cohort, for each case at each LMR. This would considerably lengthen the running time necessary to produce recommendations for a given case. Such long running times to produce recommendations are not desirable for, nor realistic to impose on, users in an implementation.

### c. Explaining the GeoMatch’s individual predictions: Variable Importance Plots

We provide variable importance plots for each Labor Market Region in the Appendix 3 of this technical document. However, IPL and COA have agreed that these plots will not be provided to users, for multiple reasons. First, as explained above, the GeoMatch recommendation incorporates more than the predictions of employment for a case at each LMR. Focusing too much on the individual predictions would present an incomplete picture of what COA is trying to achieve with GeoMatch. Second, by definition, variable importance plots can only provide a partial explanation of much more complex models and synergies. While useful, such explanations can lead to incomplete interpretations, with potential harmful consequences in COA’s context, in which users will be directly interacting with permit holders.

Moreover, although our predictive models are valid for estimating the causal effect of a location on the outcome of interest, they do not provide causal estimates of the impact of case-level covariates on the outcomes. Importance plots, and other explanatory frameworks for predictions, therefore do not provide information as to the direct causal effect that one covariate has on the outcome. This is not a problem, since the goal of our predictive models is to predict the outcome for a case at a given location, not explain what makes the case more likely to find employment at that location. If the relationship between a variable and outcome is driven by an unmeasured, confounding variable, the predictions are still valid as long as the new data for which predictions are generated is drawn from the same distribution as the training data. In this case, it is still true that the predicted variable and outcome “move” together. Therefore, although these complex machine learning models are valuable for our prediction task at hand, explanatory frameworks should be used with caution and with an understanding of the difference between correlation and causation.

## **6. Dutch Context: description of the data sources, target population, predictors and outcomes**

## a. Target Population

The target population for the deployment of GeoMatch in the Netherlands consists of status holders who are granted residence permits and are assigned to a municipality through the regular housing procedure. Further, in building our predictive models and providing recommendations, we only consider data and outcomes for adults (i.e. 18 years or older).

We exclude some subsets of status holders who fall outside the scope of the objectives or for whom data are unreliable. First, we exclude status holders who fell under the 2019 Children's Pardon. Second, we exclude resettlers/relocants/asylum seekers covered by the EU-Turkey deal due to some ambiguity in the data recorded on their registration. Third, we exclude status holders who have "hard criteria" that pre-determine their locations. Beyond family reunification constraints, these hard criteria may be motivated by other determinants such as employment contracts and medical issues. Status holders with hard criteria would not receive algorithmic recommendations since their locations are already predetermined. Finally, the target population for algorithmic recommendations further excludes family reunifiers, since such status holders are automatically reunified with their family in the Netherlands. Note, however, that data for family reunifiers is included in the training data in order to supply the predictive models with more information to detect synergies between background characteristics and locations.

We only include permit holders for whom we can collect employment information from the CBS Asielcohort dataset. We do not necessarily expect these two groups to be similar, and qualitative insights will be required to understand the reasons behind missing employment data. However, after we filter out permit holders that are out of our scope (those who died or left the Netherlands), we are able to build an employment outcome variable for most permit holders in our dataset (see Appendix 2c).

## b. Impact on out-of-scope population

As described above, we restricted the scope of the backtest analysis and will restrict the scope of the pilot to permit holders who we can infer are eligible to work. We make this determination based on available variables, such as the age of the permit holders (we filter minors out) or the determination by the program supervisor that a case has "hard criteria" that should guide their recommendation to a specific location. Hard criteria include medical conditions for example. Because our outcome of interest is employment, or employment-related variables such as wages, we mostly look at the impact of GeoMatch on these employment, or employment-related,

outcomes. Therefore, we do not consider the impact of the algorithm on populations that are not eligible for work. Specific attention to this out-of-scope population and how to handle them within the new process will be given in the instructions for employees and infosheets for status holders that COA will provide before the start of the pilot.

It is theoretically possible that some permit holders in our backtest and pilot scope could still not be eligible to work. However, because we do not have the relevant variables to identify these potential permit holders in the data, we are also not able to estimate the impact of GeoMatch on them. If we receive more information about how a measured variable impacts a permit holder's work eligibility, we would immediately adjust the scope of our study rather than wait to measure the impact on these new subgroups of permit holders. We do not envision a scenario where we would have the data to analyze the impact but not adjust the scope.

### c. Data Sources and Data Quality

Two data sources are leveraged to implement and test the use of the GeoMatch tool in the Dutch context. The first is COA's administrative data on status holders ('IBIS data'). This is a comprehensive data set covering all status holders in the Netherlands containing their background characteristics, procedural information, and location data from 2014 onwards. This is the core data set upon which the GeoMatch recommendation procedure in the Netherlands can be based.

Second, an additional data source is also required for studying outcomes, which are not covered by COA's administrative data. We rely on the Asielcohort microdata compiled by Statistics Netherlands (CBS) every year to find outcomes data for all permit holders in our dataset. The Asielcohort microdata is composed of merged data from a number of administrative databases, several of which track various measures of status holders in the Netherlands after arrival. These include economic and educational integration that can be evaluated as mainstream and downstream outcomes in the algorithm.

As recipients of the data (not collectors, producers, or owners of the data), we undertake routine data quality checks to establish our confidence in its integrity. This includes steps such as inspecting the types, ranges and distributions of all relevant variables, to ensure variables did not contain nonsensical values that would be suggestive of miscoding/unreliability, as well as checking the contents of the variables in reference with the codebooks provided by COA. We also look at the number of missing values for the variables used in our analyses. We also conduct checks on the variables we use as outcomes, including making sure the average and range of the wage variable were at reasonable levels, checking the correlations between employment after one, two or three years as well as between employment and wage

levels. We also look at the current distribution of cases across Labor Market Regions and at the current outcome averages in each location and checked for any patterns that would not align with our qualitative understanding of current patterns. Finally, we look at evolutions across time of the types of cases in IBIS to check if they align with COA's and IPL's qualitative understanding of trends in arrivals and changes in COA processes (overall arrivals, proportions of family reunifiers or hard criteria cases, proportion of cases having a housing interview, proportion of minors and males across the years).

Because IPL must abide by CBS' data use agreements, we cannot export all of these analyses out of the Remote Environment. For example, we cannot export minimum and maximum values from the Remote Environment. We check minimums and maximums for all variables as part of the scope check, but cannot export the results. We provide the analyses we were able to export in Appendix 2 of this document.

During the pilot, we will build in data quality checks in the algorithm code and return a warning or an error if the data does not match the expected format or specification.

## d. Choice of outcome

As explained above, in choosing the specific outcome variable to feed into the algorithm we seek to balance several imperatives. The first is the goal of helping the clients (i.e. permit holders) achieve long-term and lasting success. The second is the need to take shorter-term information into account, such that the algorithmic methods are resilient and can adapt to changes over time with higher velocity. Finally, the third is the need for an outcome with sufficient variation across status holders and locations for algorithmic recommendations to be viable from a statistical and mathematical perspective.

For these reasons, we chose the medium-term outcome of employment attainment within a full year after moving into housing in the municipality (i.e. the date when the status holder departs the AZC and moves into their assigned housing in the municipality) as our target outcome for optimization. Specifically, we consider the binary outcome variable that indicates whether a status holder has found any employment in any of the 12 months within their first year after moving into housing in the municipality. It is essential that these outcomes are measured after moving into housing in the municipality because only then do status holders arrive in their destination LMR (the target geography for the algorithm's recommendations and the structure around which the predictive models are trained).

GeoMatch seeks to optimize employment in the Netherlands given initial linking to a Labor Market Region - regardless of whether the permit holder stays and finds employment in the initial LMR, or in another (as long as it is still in the Netherlands). This is because the initial linking is the only exogenous linking, i.e. is decided by COA

and where GeoMatch recommendations would be considered. The permit holder can self-select into their second or third LMR, and this self-selection could bias our models if we used the second or third LMR as the relevant LMR. Moreover, if we assume that the initial linking can influence the likelihood that a permit holder finds employment or not in a different LMR, we would further bias our models by using these endogenous LMRs in our modeling. Therefore, training models using the LMR where the permit holder found employment, as opposed to the LMR of the initial linking, would carry a significant risk of biasing the models.

## e. Predictors

The background characteristics that are used as predictors in the modeling procedure determine the basis upon which the algorithm can identify person-location synergies to leverage. Hence, it is important to include as many predictors that have a plausible link to the target outcome as possible. In addition, since the models are designed to be applied in a real-world implementation to status holders before they have been assigned, only pre-assignment variables can be employed as predictors.

Based on data availability, the characteristics we include as predictors are age, gender, marital status, prior education, number of family members, country of origin, native language, and prior work experience and industry. In addition, we also include several key variables related to the assignment procedure as predictors, including the month of assignment, year of assignment, whether or not someone is a family reunifier, and whether they had their interview (and hence received labor market regional advice) at a POL, AZC, or other type of location.

## f. Model retraining

Models will be retrained once a year at a minimum. COA can provide updated data on employment variables once a year, when CBS publishes the Asielcohort Dataset. IPL and COA are exploring the possibility of computing employment variables more frequently (twice a year), if an employment variable can be built at this frequency. This would allow us to retrain models twice a year, in addition to being able to monitor results more frequently.

Moreover, should the scope of the pilot change, models would also be retrained accordingly.

Each time the models are retrained, quality checks mentioned above in this document as well as detailed in the monitoring plan will be conducted. Similarly,

model performance metrics will be checked (see Appendix for details on which model performance metrics are used) to monitor that the predictive accuracy of the model stays at acceptable levels.

## g. Estimation of gains: backtest and pilot

### i. Employment gains

The backtest described in this technical document is further detailed in the report, “GeoMatch Netherlands: Retrospective Impact Evaluation and Feasibility Study for the Central Agency for the Reception of Asylum Seekers (COA).” The data used in that backtest combined data from an IBIS export in 2021 with the Asielcohort 202010. The backtest was conducted by training a model on in-scope permit holders that were assigned to a municipality before June 2018. The out-of-sample test cohort was made up of permit holders assigned to municipalities between June 2019 and August 2019 (since the last available employment outcomes in the dataset were from August 2020). The report provides the full details on the population scope included in the backtest. In the backtest, the assignment date is the date that an asylum seeker moved out of COA housing and arrived at the municipality where they have been linked.

In order to assess the potential impact of using data-driven location recommendations on status holders’ integration outcomes in the Netherlands, we performed a retrospective impact evaluation (backtest), where we compared the outcomes that status holders achieved under their actual (status quo) assignment procedure to what we estimate they could have achieved had they instead been recommended to labor market regions by the GeoMatch tool. The main outcome for the backtest focused on gains in employment in the 12 months after assignment. Employment was defined as any employment of any length in the 12 months following assignment to a municipality. The backtest report includes analyses to indicate how this outcome is correlated with other outcomes of interest, including earnings and proportion of time worked. The algorithm considers the likelihood that any adult member of a case obtains employment, but the gains are presented at the individual level rather than the case level.

The cohort used in the backtest report is the cohort of permit holders who arrived at their municipality between June and August 2019. This cohort’s one year employment overlapped with the start of the Covid-19 pandemic, since we looked at employment between June 2019 and August 2020. During the pilot, IPL will use the most up-to-date employment data available to train the models used in the pilot, and these models will be regularly updated (see model retraining section above). Each

time IPL retrains the models used in the pilot, performance metrics will be monitored to ensure that the predictive quality of the models is acceptable.

## ii. Financial gains

The backtest report includes an estimation of the financial benefits of using the GeoMatch recommendations (section 5.3 of the backtest report). This computation relies on assumptions listed in the report and built in concertation with COA, and is based on the two-year employment outcomes in terms of time worked over that time period. Assuming a 17% increase above the baseline in the proportion of time worked over the first two years, we estimated the two-year financial benefits for the government for each annual cohort both in terms of social assistance benefits cost savings and additional income tax revenue generated from additional months worked from using GeoMatch recommendations to be roughly 5.3 million euros<sup>8</sup>. We emphasize that this is an estimate based on several assumptions, and that it is likely to represent a lower bound, both due to conservative assumptions made throughout the analysis and with full VA restrictions applied to the GeoMatch recommendations. To go further in the estimation of the financial gains, we can provide a confidence interval around the 5.3 million euros estimate, using the confidence interval we computed for the proportion of time worked over the first two years [15.8;17.8]. The confidence interval for the two-year financial benefits for the government for each annual cohort would be [4.9 million ; 5.7 million].

## iii. Beyond backtest: RCT and pilot

COA and IPL are planning to rigorously evaluate the effects of GeoMatch when it is launched in the Netherlands. We will rely on the scientific gold standard for impact evaluation, a randomized control trial (RCT), to estimate the real effect of GeoMatch in the Netherlands. The monitoring plan attached to this technical document also includes key details on how the RCT will be monitored.

## h. Downstream effects on other outcomes

Although the goal of GeoMatch is to maximize employment, we expect the benefits of algorithmic recommendations to go beyond the domain of short term economic

---

<sup>8</sup> This estimate assumes that all predictors used in the backtest can be used in the pilot. If GeoMatch cannot use religion and ethnicity as predictors, the estimate for financial gains is roughly 4.5 million euros.

integration. Past research documents a close and intertwined relationship between economic and other dimensions of integration success, such as linguistic and social integration (Harder et al. 2018). Thus, we believe there is a solid basis to expect positive “downstream” effects of using GeoMatch on related integration outcomes.

In the backtest, we provided COA with an estimation of the impact of GeoMatch on other available outcomes, including longer term employment outcomes (percent of time with work in first two years), wage outcomes (wages in first two years) and educational outcomes (percent of time pursuing education in first year). We evaluated the expected impact of targeting one-year employment on these other outcomes. Specifically, we continued to use the standard algorithmic procedure that optimized on first-year employment, but estimated the resulting gains with respect to the other outcomes. All downstream outcomes are estimated with the GBM package under the same configurations (covariates, model training approach and validation) we used to estimate gains on the main outcome. Appendix 1 below provides some relevant performance metrics (including calibration and accuracy) for the outcome variables used in the backtest report.

Besides the outcomes we were able to measure, we expect that status holders who find work sooner through optimized geographic recommendations are also quicker to learn Dutch, expand their social network, form social ties with Dutch citizens through work-related contacts, and face a lower propensity to engage in secondary migration. Existing research has also found evidence that employment participation can confer broader psychosocial benefits within refugee populations beyond the value produced by cash or financial gain itself (Hussam et al. 2021). However, because of data limitation issues, we are not able to measure these other outcomes in the Dutch context.

In concert with COA, we used the two-year employment outcome as the basis for an estimation of the financial benefits for the Dutch government of using GeoMatch. The estimated gains in the proportion of time worked over the first two years is the key input into a financial calculus whose components and assumptions are detailed in pages 22 and 23 of the backtest report.

## i. Evaluating the pilot: prospective study

In order to rigorously measure the impact of using GeoMatch in the Dutch context on permit holders’ employment, COA and IPL will deploy a prospective study (Randomized Control Trial) along with the pilot. For details on the design of this prospective study, readers can consult the Research Design presentation attached to this memo. For details on how the prospective study will be monitored throughout

the implementation of the pilot, readers can consult the Monitoring Plan attached to this memo.

## j. Engaging users: continuous feedback loop

Throughout the pilot, we will gather feedback from users (program supervisors and allocation officers) at frequent intervals (at least once per month for the first 6 months; then every three to six months). We will use both quantitative (surveys) and qualitative (focus groups) methods to gather feedback, and will focus on several key aspects of using the algorithm:

- How users understand the tool
- How useful the tool is to users
- The main reasons for compliance and non compliance with the tool
- Desired changes to the tool

## 7. References

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373), 325-329.

[Paper](#)

[Supplemental material](#)

[Replication Code](#)

Bansak, K. Paulson, E. Outcome-Driven Dynamic Refugee Assignment with Allocation Balancing. Working paper. [Preprint](#).

Acharya, A., Bansak, K., & Hainmueller, J. (2022). Combining Outcome-Based and Preference-Based Matching: A Constrained Priority Mechanism. *Political Analysis*, 30(1), 89-112.

[Paper](#)

Bansak, K., Fei, J., Adams-Cohen, N., Ferwerda, J., Hangartner, D., Hainmueller, J. (2021) GeoMatch Netherlands: Retrospective Impact Evaluation and Feasibility Study for the Central Agency for the Reception of Asylum Seekers (COA). Internal COA Report.

Harder, Niklas, Lucila Figueroa, Rachel M Gillum, Dominik Hangartner, David D Laitin and Jens Hainmueller. 2018. "Multidimensional measure of immigrant integration." *Proceedings of the National Academy of Sciences* 115(45):11483–11488.

Hussam, Reshmaan N, Erin M Kelley, Gregory V Lane and Fatima T Zahra. 2021. The Psychosocial Value of Employment. Technical report National Bureau of Economic Research.

## 8. Appendices

### 1. Appendix 1: Model Performance

#### a. Model Performance on multiple outcomes

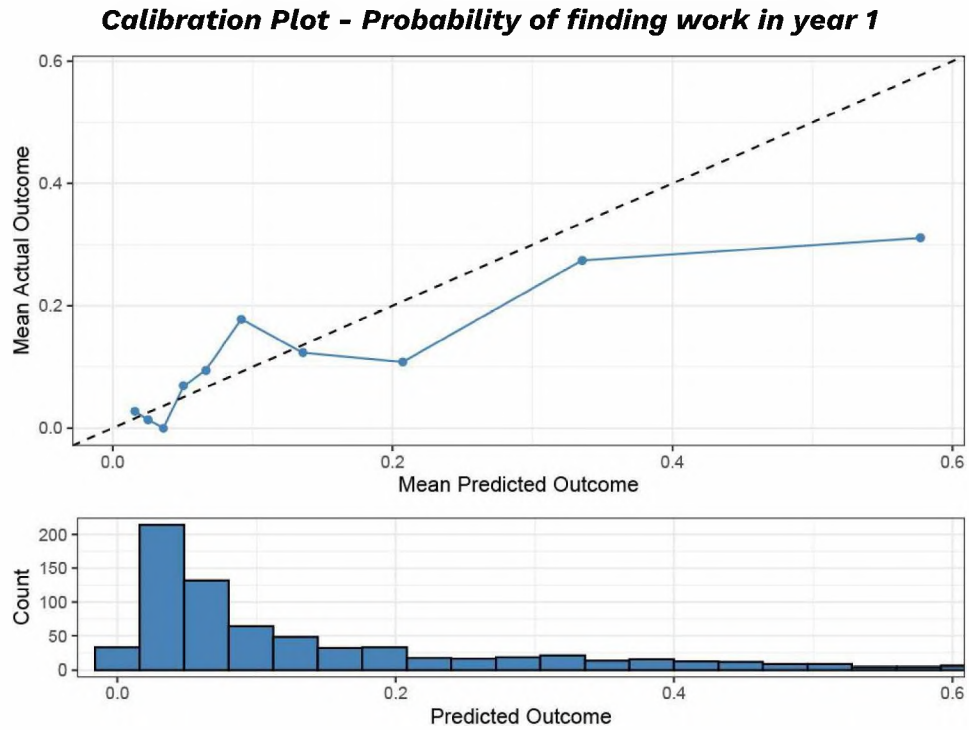
The table below provides summary metrics of model performance for several outcome variables used in the backtest report. We provide RMSE, R Square and Correlation metrics for the continuous variables, and accuracy and RMSE for the binary variable. Because GBM provides probabilities predictions even for binary variables, before measuring accuracy we classified any probability above 0.5 as a 1 and any probability equal or below 0.5 as a 0.

**Performance metrics, GBM model**

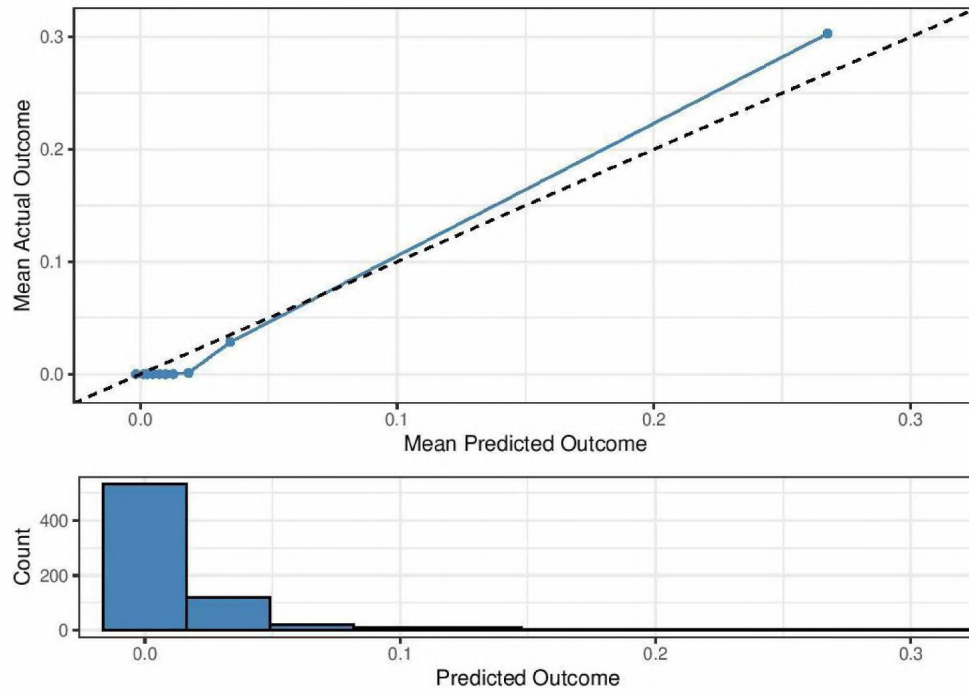
| Outcome                                             | Accuracy | RMSE                                               | R Square | Correlation |
|-----------------------------------------------------|----------|----------------------------------------------------|----------|-------------|
| <i>Probability of finding work in year 1</i>        | 0.91     | 0.27<br>(0.25 among AA cases, 0.28 among VA cases) | NA       | NA          |
| <i>Percent of time with work in first 2 years</i>   | NA       | 0.19                                               | 0.48     | 0.69        |
| <i>Percent of time pursuing education in year 1</i> | NA       | 0.1                                                | 0.58     | 0.76        |
| <i>Wages in first two years</i>                     | NA       | 7352                                               | 0.41     | 0.64        |

Another key performance metric we use in our evaluations is the calibration. We provide the calibration plots for the outcome variables used in the backtest report below. The calibration plots evaluate the predictive performance of a model by comparing the mean predicted outcomes against the mean actual outcomes across different segments of the data. The 10 bullets in each calibration plot represent the decile groups of the predicted values, where each bullet's position reflects the mean predicted outcome (x axis) versus the mean actual outcome (y axis) for that decile. These bullets visually demonstrate how well the model's predictions align with the

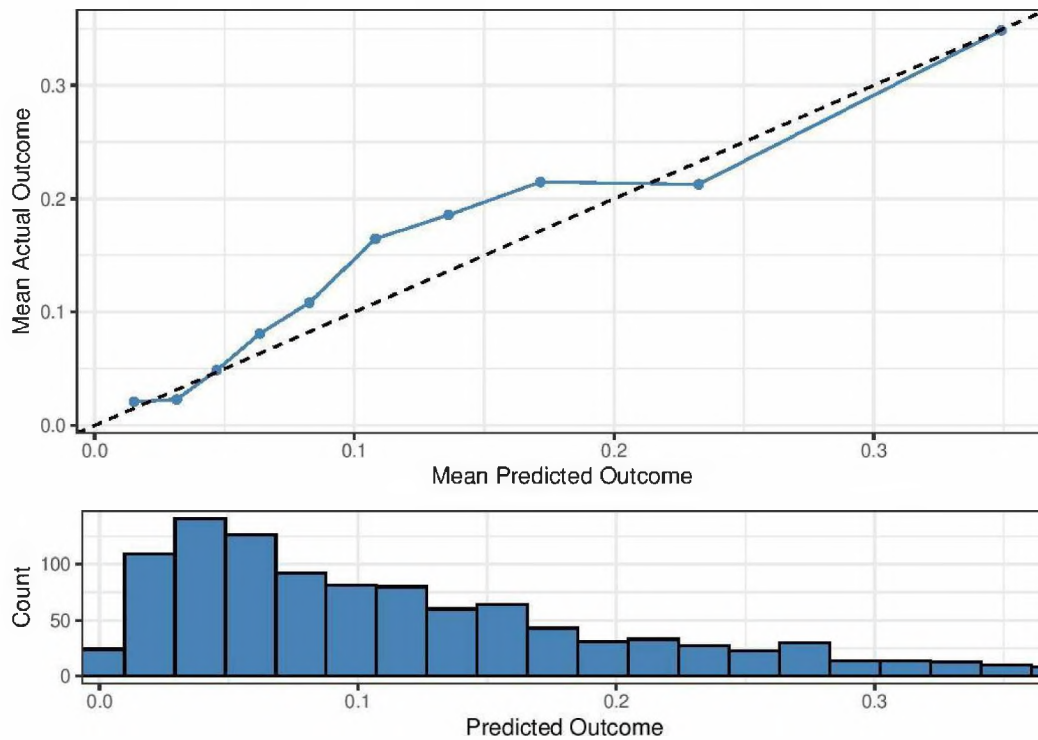
actual outcomes across the spectrum of predicted values, indicating the model's calibration accuracy within each quantile segment. A perfectly calibrated model would follow the identify (diagonal) line.

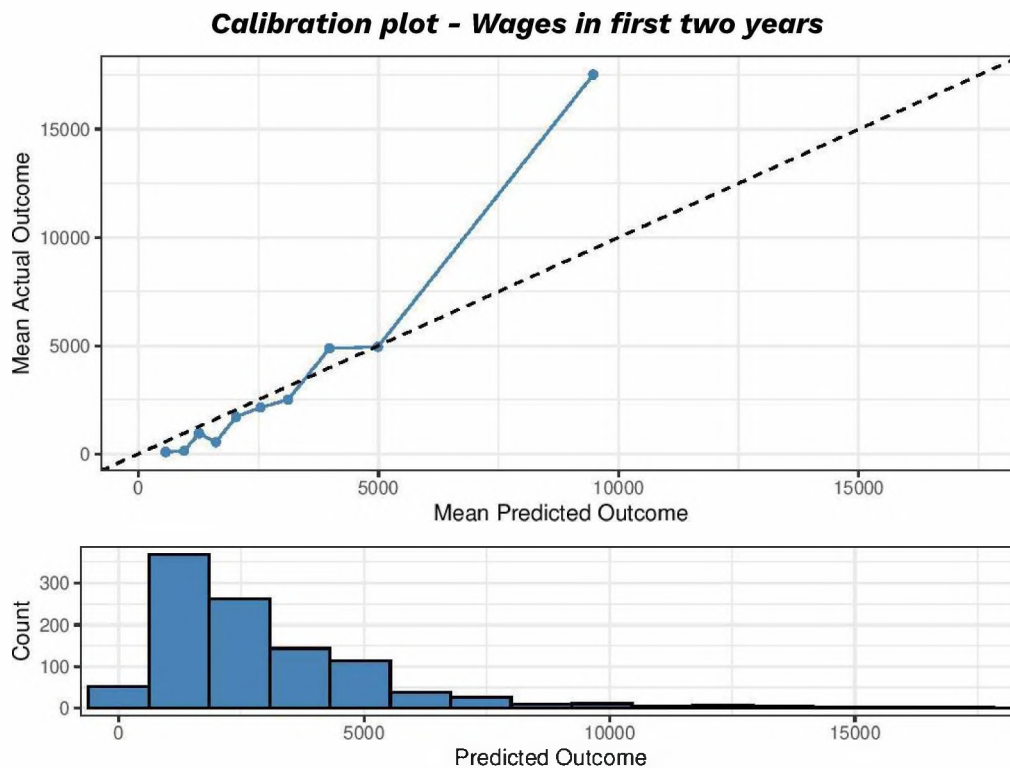


**Calibration Plot - Percent of time pursuing education in year 1**



**Calibration Plot - Percent of time with work in first 2 years**





## b. Model performance: comparison with alternative methods

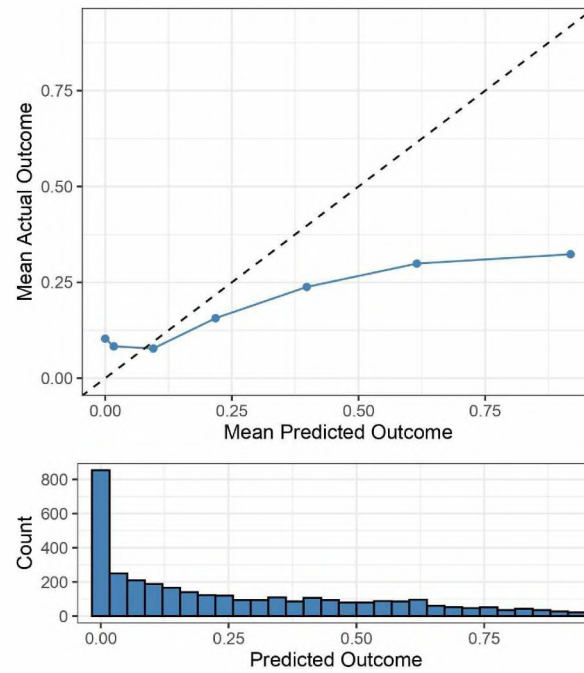
For comparison purposes, we provide performance metrics on our primary outcome variable (probability of finding employment within the first 12 months in a municipality) for a few key relevant alternative statistical methods. We also provide calibration plots for these alternative methods. These comparison metrics show that the method we use in the Dutch pilot, GBM, outperforms more traditional predictive models.

### **Performance Metrics, alternative methods**

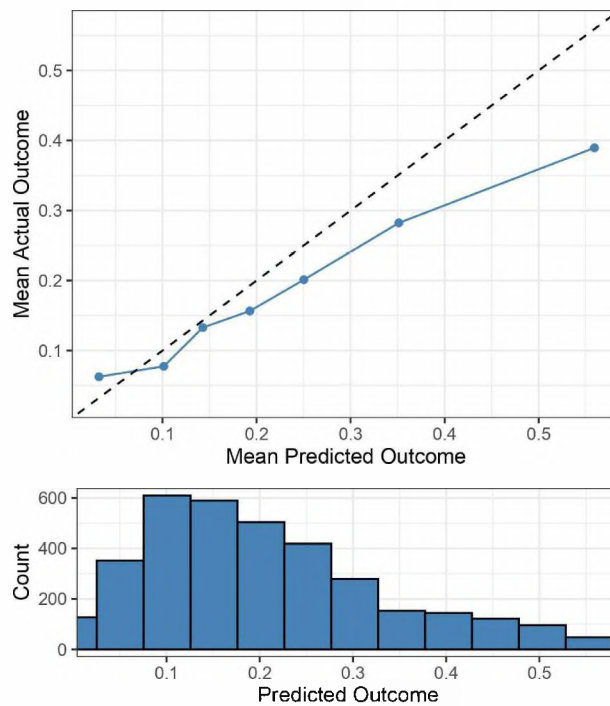
| Method       | Accuracy <sup>9</sup> | RMSE |
|--------------|-----------------------|------|
| <b>OLS</b>   | 0.81                  | 0.37 |
| <b>LOGIT</b> | 0.73                  | 0.44 |
| <b>GBM</b>   | 0.91                  | 0.27 |

<sup>9</sup> For accuracy, we assume that all predictions with a probability over 0.5 predict employment == 1

**Logit model calibration plot**



**OLS calibration plot**



## 2. Appendix 2: Data Quality Checks

### a. Type of each variables used in backtest

We check the type of each variable we use in the algorithm, and make sure that numeric/factor/character variables are coded as such. See `backtest_types.csv` file attached for the full list.

## b. Missing data on key variables

| Variable Name                     | Info            | Vergunninghouder_op_enig_moment | N      |
|-----------------------------------|-----------------|---------------------------------|--------|
| Verblijfplaats_1                  | contains_info   | no                              | 37568  |
| Verblijfplaats_1                  | contains_info   | yes                             | 126494 |
| Verblijfplaats_1                  | no_info         | no                              | 11223  |
| Verblijfplaats_1                  | no_info         | yes                             | 231    |
| Code_Geslacht                     | contains_info   | no                              | 1866   |
| Code_Geslacht                     | contains_info   | yes                             | 126725 |
| Code_Geslacht                     | no_info         | no                              | 46925  |
| Code_Geslacht                     | no_info         | yes                             | 0      |
| Gekoppeld_aan_gemeente            | contains_info   | no                              | 86     |
| Gekoppeld_aan_gemeente            | contains_info   | yes                             | 116278 |
| Gekoppeld_aan_gemeente            | no_info         | no                              | 48705  |
| Gekoppeld_aan_gemeente            | no_info         | yes                             | 10447  |
| Datum_koppeling_aan_gemeente      | contains_info   | no                              | 86     |
| Datum_koppeling_aan_gemeente      | contains_info   | yes                             | 117224 |
| Datum_koppeling_aan_gemeente      | no_info         | no                              | 48705  |
| Datum_koppeling_aan_gemeente      | no_info         | yes                             | 9501   |
| merged_indicator                  | Both            | no                              | 2      |
| merged_indicator                  | Both            | yes                             | 101388 |
| merged_indicator                  | IBIS_Only       | no                              | 48789  |
| merged_indicator                  | IBIS_Only       | yes                             | 25337  |
| Code_land_werkervaring_opgedaan_1 | contains_info   | adult                           | 30992  |
| Code_land_werkervaring_opgedaan_1 | contains_info   | minor                           | 2      |
| Code_land_werkervaring_opgedaan_1 | no_info         | adult                           | 38151  |
| Code_land_werkervaring_opgedaan_1 | no_info         | minor                           | 32243  |
| Code_land_van_educatie_1          | contains_info   | adult                           | 36937  |
| Code_land_van_educatie_1          | contains_info   | minor                           | 7      |
| Code_land_van_educatie_1          | no_info         | adult                           | 32206  |
| Code_land_van_educatie_1          | no_info         | minor                           | 32238  |
| aantal_ple_leden_DIFF_num_cid     | 0               | yes                             | 96952  |
| aantal_ple_leden_DIFF_num_cid     | 1               | yes                             | 3445   |
| aantal_ple_leden_DIFF_num_cid     | 2               | yes                             | 508    |
| aantal_ple_leden_DIFF_num_cid     | 3               | yes                             | 254    |
| aantal_ple_leden_DIFF_num_cid     | 4               | yes                             | 131    |
| aantal_ple_leden_DIFF_num_cid     | 5               | yes                             | 64     |
| aantal_ple_leden_DIFF_num_cid     | 6               | yes                             | 17     |
| aantal_ple_leden_DIFF_num_cid     | 7               | yes                             | 10     |
| aantal_ple_leden_DIFF_num_cid     | 8               | yes                             | 3      |
| aantal_ple_leden_DIFF_num_cid     | 9               | yes                             | 4      |
| Omschrijving_nationaliteit        | Nederlandse     | adult                           | 6542   |
| Omschrijving_nationaliteit        | Nederlandse     | minor                           | 5132   |
| Omschrijving_nationaliteit        | not_Nederlandse | adult                           | 62601  |
| Omschrijving_nationaliteit        | not_Nederlandse | minor                           | 27113  |

### c. Missing data on outcome variables

For each of the outcome variables used by GeoMatch in the backtest, we show the percentage of backtest adults with outcome information. More information on the variables and construction of these outcomes is provided in the report, “GeoMatch Netherlands: Retrospective Impact Evaluation and Feasibility Study for the Central Agency for the Reception of Asylum Seekers.”

#### ***Missing outcome variable proportions***

| <b>Outcome</b>                                                 | <b>Outcome Variable</b> | <b>Prop with info</b> |
|----------------------------------------------------------------|-------------------------|-----------------------|
| Any work in the first 12 months                                | anywork_first12         | 1                     |
| Any work in the first 24 months                                | anywork_first24         | 0.91                  |
| Any work in the first 36 months                                | anywork_first36         | 0.78                  |
| Any education in the first 12 months                           | anyeduc_first12         | 1                     |
| Any education in the first 24 months                           | anyeduc_first24         | 0.91                  |
| Any education in the first 36 months                           | anyeduc_first36         | 0.78                  |
| Proportion of months with any work in the first 12 months      | propwork_first12        | 1                     |
| Proportion of months with any work in the first 24 months      | propwork_first24        | 0.91                  |
| Proportion of months with any work in the first 36 months      | propwork_first36        | 0.78                  |
| Proportion of months with any education in the first 12 months | propeduc_first12        | 1                     |
| Proportion of months with any education in the first 24 months | propeduc_first24        | 0.91                  |
| Proportion of months with any education in the first 36 months | propeduc_first36        | 0.78                  |

|                                 |  |  |
|---------------------------------|--|--|
| any work in the first 36 months |  |  |
|---------------------------------|--|--|

#### d. Descriptives on number and type of cases across years

The first table below shows how many individuals we find in the IBIS data, based on the following filters we applied for the backtest.

##### ***Filtering variables for GeoMatch backtest and pilot***

| Filtering                                                                | Variable filtered                                       | Unique Vnummer Crypt |
|--------------------------------------------------------------------------|---------------------------------------------------------|----------------------|
| Raw                                                                      |                                                         | 270,832              |
| People holding permits                                                   | IndicatieVergunninghouder 1 == yes                      | 158,251              |
| People linked through regular housing procedure                          | SoortPlaatsing 1 == 'Regulier via omgekeerd bemiddelen' | 141,001              |
| People who are not resettlers, relocants or associates of EU-Turkey deal | ClassificatiesHervestigerVervallen != RUV, REL or HVT   | 133,781              |
| People who are not part of Children's pardon 2019                        | Opvangvorm has Gezinslocatie in it?not exactly sure     | 133,011              |
| People who have a muni assignment                                        | Gekkoppeldaangemeente != ""                             | 130,676              |
| People who have gone to their muni                                       | IngangsdatumVerblijfplaats 1 != "?"                     | 123,046              |
| Remove duplicates                                                        |                                                         | 123,045              |

In the table below, we start with the individuals who could be included in a pilot ('*N power*'). From the table above, this means we apply the first 4 filters (have permit, regular housing procedure, no resettlers and no Children's Pardon), but include people who don't have a municipality assignment and people who have not moved to their municipality yet, since this filter was relevant for the backtest but would not be for the pilot. The '*N power*' column in the table below sums up to the 133,011 in the table above (minus 6 duplicates that we also delete, so the column sums up to 133,005).

'*N housing*' refers to the number of individuals receiving a housing interview (if 1 member of a PLE has received a housing interview, we consider that all members of the PLE have had a housing interview), '*N SM*' refers to the number of individuals

having received a Screening and Matching interview, ‘*N family reunifiers*’ refers to the number of individuals coded as family reunifiers, ‘*N hard criteria*’ refers to the number of individuals with hard criteria, ‘*N minors*’ refers to the number of individuals under 18 when they moved to their municipality, and ‘*N male*’ the individuals coded as male.

‘*N pilot*’ refers to the number of adults, with a housing interview, with no family reunifier and no hard criteria. This is the number of adults we would have been able to enroll in the pilot.

Finally, ‘*% total N in pilot*’ refers to the percentage of people having received a permit each year that end up pilot eligible (i.e., in ‘*N pilot*’).

***Descriptive summaries, type and number of cases across years***

| <b><i>Koppeling Year</i></b> | <b><i>N power</i></b> | <b><i>N housing</i></b> | <b><i>N SM</i></b> | <b><i>N Family reunifier</i></b> | <b><i>N hard criteria</i></b> | <b><i>N minors</i></b> | <b><i>N male</i></b> | <b><i>N pilot</i></b> | <b><i>% total in pilot</i></b> |
|------------------------------|-----------------------|-------------------------|--------------------|----------------------------------|-------------------------------|------------------------|----------------------|-----------------------|--------------------------------|
| 2014                         | 10483                 | 9159                    | 0                  | 3878                             | 0                             | 3739                   | 6550                 | 5482                  | 0.505                          |
| 2015                         | 27551                 | 21924                   | 1                  | 13191                            | 1                             | 10409                  | 15726                | 12089                 | 0.425                          |
| 2016                         | 31874                 | 26104                   | 659                | 12964                            | 64                            | 11353                  | 18903                | 15208                 | 0.434                          |
| 2017                         | 21048                 | 13828                   | 1692               | 15767                            | 402                           | 9962                   | 9840                 | 3735                  | 0.144                          |
| 2018                         | 11749                 | 9285                    | 3254               | 8248                             | 690                           | 5218                   | 5846                 | 2540                  | 0.182                          |
| 2019                         | 9720                  | 7404                    | 5310               | 5311                             | 1185                          | 3923                   | 5257                 | 3066                  | 0.259                          |
| 2020                         | 13232                 | 11004                   | 8354               | 5478                             | 1558                          | 3525                   | 7832                 | 3963                  | 0.281                          |
| 2021                         | 5004                  | 3849                    | 3098               | 2027                             | 544                           | 474                    | 2940                 | 229                   | 0.042                          |
| NA                           | 2344                  | 652                     | 65                 | 1218                             | 25                            | 1476                   | 1374                 | 238                   | 0.019                          |

All tables below will filter out minors and individuals with family reunifiers and hard criteria. The additional filters applied for each specific table will be specified in the table description.

#### e. Descriptives, outcomes of SM interview over the years

To produce the table below, we filtered the data to cases who received a SM interview. In the first table, we show the proportion of PLEs whose first PS recommendation is a blanco, blank, or actual recommendation (‘% of first rec LMR’),

as well as the proportion of PLEs who received more than 1 actual recommendation from the PS.

***Descriptive summaries, PS regional advice***

| Koppeling Year | % of first rec Blanco | % of first rec Blank | % of first rec LMR | % more than 1 rec |
|----------------|-----------------------|----------------------|--------------------|-------------------|
| 2016           | 37%                   | 0%                   | 63%                | 21%               |
| 2017           | 26%                   | 11%                  | 63%                | 7%                |
| 2018           | 19%                   | 21%                  | 59%                | 12%               |
| 2019           | 9%                    | 30%                  | 60%                | 17%               |
| 2020           | 15%                   | 36%                  | 48%                | 9%                |
| 2021           | 13%                   | 35%                  | 52%                | 8%                |

***Descriptive summaries, PS regional advice - AA vs VA***

|          | % of first rec Blanco | % of first rec Blank | % of first rec LMR | % more than 1 rec |
|----------|-----------------------|----------------------|--------------------|-------------------|
| AA cases | 25%                   | 6%                   | 69%                | 16%               |
| VA cases | 10%                   | 43%                  | 47%                | 9%                |

In the tables below, we only keep cases that received at least 1 regional advice from the PS (i.e., no blank or blanco advice). We look at the percentage of PLEs that were linked to a municipality in one LMR recommended by the PS.

***Descriptive summaries, PS advice and final linking***

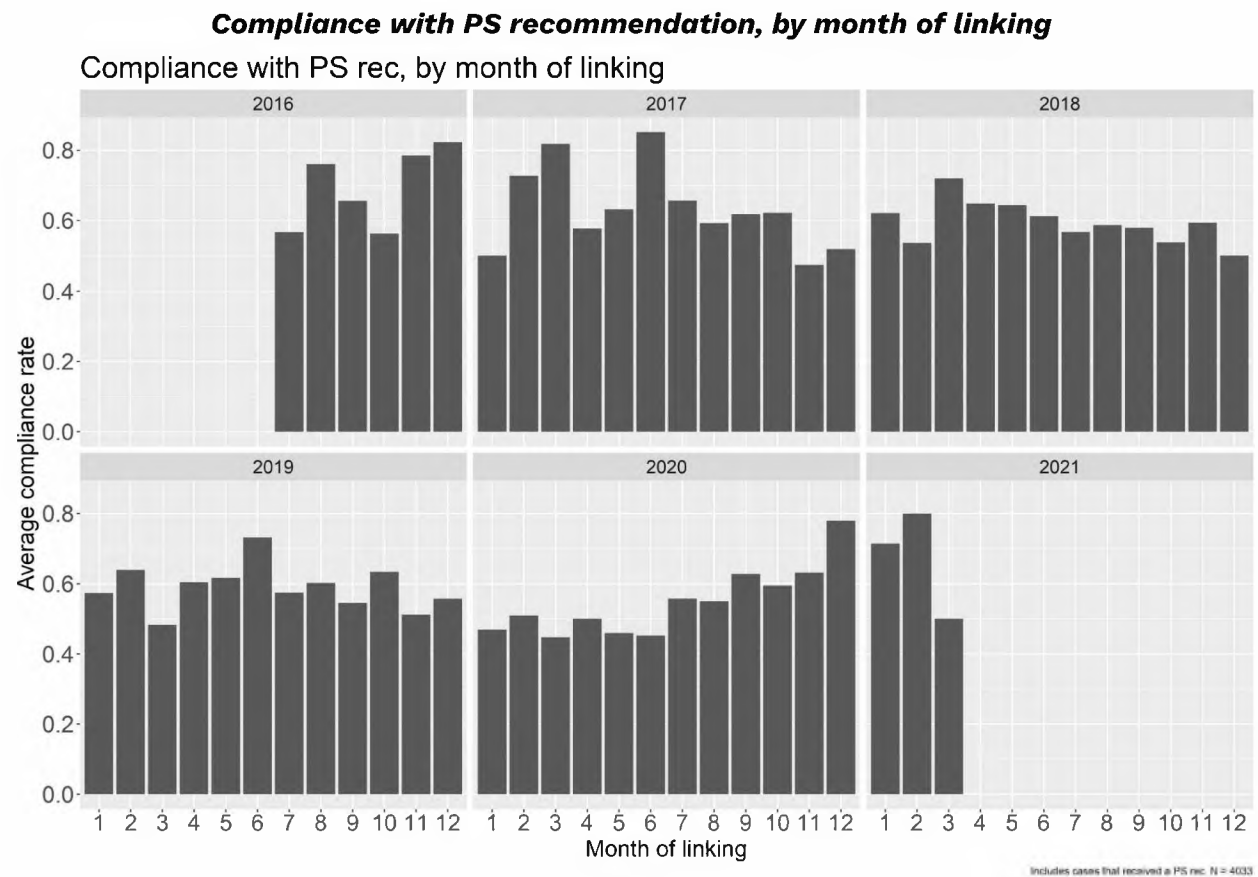
|          | % of cases linked to an LMR recommended by PS |
|----------|-----------------------------------------------|
| Overall  | 59%                                           |
| AA cases | 58%                                           |
| VA cases | 59%                                           |

***Descriptive summaries, PS advice and final linking***

| Koppeling Year | % of cases linked to an LMR recommended by PS |
|----------------|-----------------------------------------------|
| 2016           | 67%                                           |

|                     |     |
|---------------------|-----|
| 2017                | 59% |
| 2018                | 59% |
| 2019                | 59% |
| 2020                | 56% |
| 2021<br>(Jan-March) | 72% |

We also look at the percentage of cases linked to a PS recommended LMR based on the month of linking for the case. No clear pattern emerges that would indicate varying degree of compliance throughout the year from AOs.



Finally, we look at the average number of PS recommendations for a case (among cases that receive at least 1 recommendation), as well as the percentage of these recommendations that are for an LMR in the UR of the permit holder's AZC. For VA cases, the PS already knows the AZC the case is staying at (and is presumably aware

of the VA restriction policy). For AA cases, the PS does not know which AZC the case will be placed in when they make their recommendations.

**Percentage of PS recommendations in UR of first AZC**

|          | Average N of recommendations | % of PS recs in UR of first AZC |
|----------|------------------------------|---------------------------------|
| AA cases | 1.46                         | 68%                             |
| VA cases | 1.31                         | 45%                             |

#### f. Descriptives, PLEs' journeys at AZCs

In the table below, we look at the percentage of PLEs that receive their housing interview and their linking (koppeling) while staying at an AZC. This table only includes cases who have received a housing interview. We also look at the percentage of cases who stay in multiple AZCs. Note that by 'stay at multiple AZCs', we mean that a given case has multiple entries in IBIS coded as being in an AZC, and the name of the locations differ in any way. These numbers align with our understanding of the linking process: almost all AA cases receive their housing interview at a POL, not an AZC, but are linked to a municipality after they moved to the AZC. VA cases, on the other end, receive their housing interview and their linking while staying at an AZC. AA and VA cases appear to be as likely to go through at least 2 AZCs in their journeys.

**Descriptive summaries, events while at AZC (all cases with housing interview)**

|          | Received housing interview at AZC | Linked while at AZC | Stay at multiple AZCs |
|----------|-----------------------------------|---------------------|-----------------------|
| AA cases | 0.4%                              | 94%                 | 28%                   |
| VA cases | 99.9%                             | 97%                 | 28%                   |

We now look at the same numbers, but filtered only to cases that have received at least 1 recommendation from the PS.

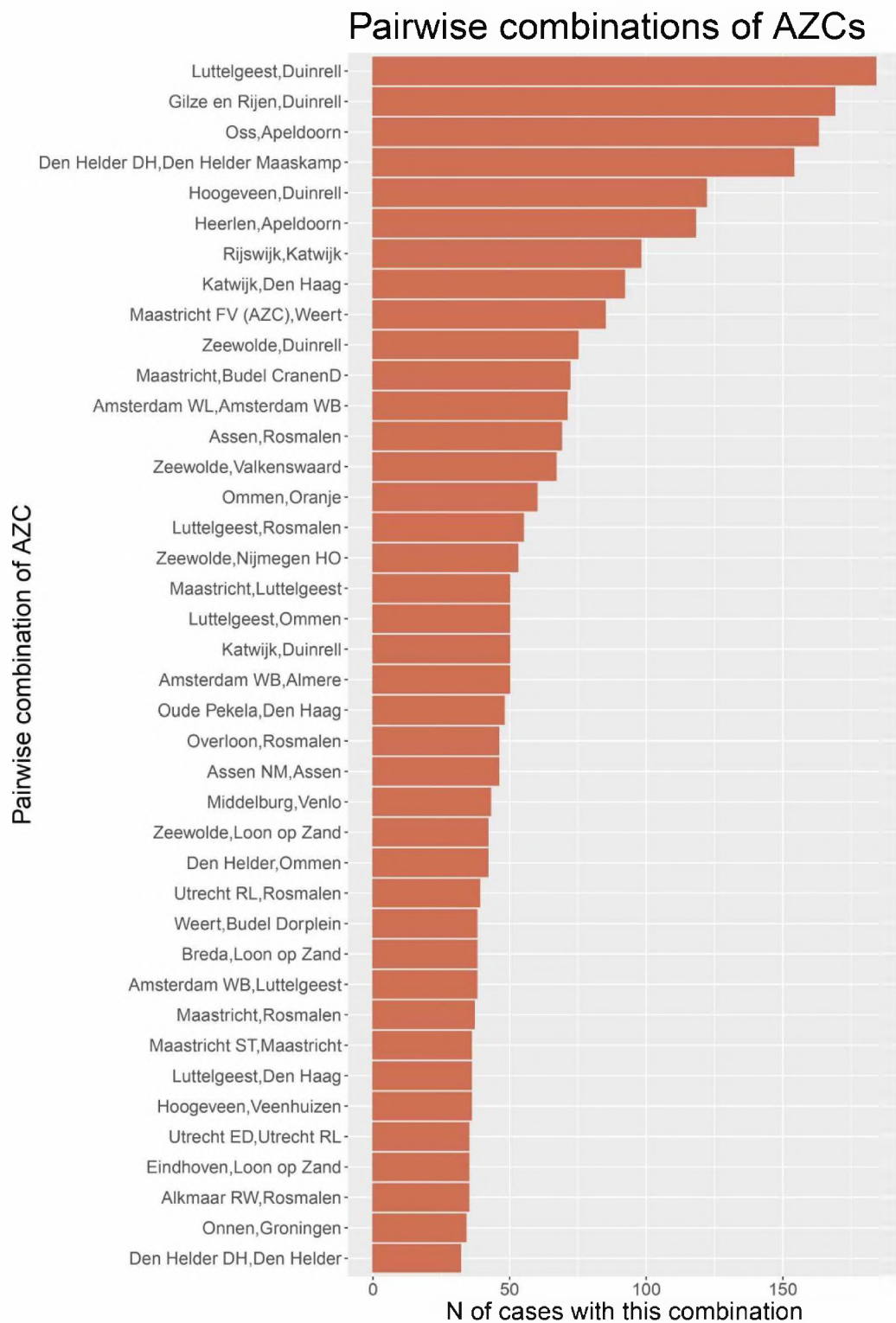
**Descriptive summaries, events while at AZC (cases with at least 1 PS rec)**

|          | Receive housing interview at AZC | Linked while at AZC | Stay at multiple AZCs |
|----------|----------------------------------|---------------------|-----------------------|
| AA cases | 0%                               | 96%                 | 22.7%                 |

|                 | Receive housing<br>interview at AZC | Linked while at<br>AZC | Stay at multiple<br>AZCs |
|-----------------|-------------------------------------|------------------------|--------------------------|
| <b>VA cases</b> | 99.8%                               | 99.4%                  | 26%                      |

Among cases that go through at least 2 AZCs, we look at what the most common pairwise combinations are (we only plot the 40 most common combinations).

**Most common pairwise combinations of AZCs for PLEs going through at least 2 AZCs**



ases that went through multiple AZCs. Only 40 most frequent combinations shown. N cases = 11,200

We also look at the match between any of the PS recommendations and the first AZC the case was sent to by the placement unit. For AAs, this shows how often the placement unit ‘follows’ one of the PS advice. For VAs, this shows how often the PS tries to make a local recommendation. The table below only includes cases that received at least 1 PS advice, and we look at the percentage of cases whose 1st AZC was in one of the LMR recommended by the PS. Overall, 32% of the cases that had received at least 1 PS regional advice were first sent to an AZC within an LMR recommended by the PS. The percentages are much higher for AAs than they are for VAs, which seem to indicate that the placement unit is sometimes willing to place the case in an AZC according to the PS recommendation.

***First AZC and PS recommendation***

| Year | AA            | VA            |
|------|---------------|---------------|
| 2016 | 44% (N = 212) | NA (N = 1)    |
| 2017 | 42% (N = 354) | 19% (N = 68)  |
| 2018 | 48% (N = 524) | 20% (N = 250) |
| 2019 | 47% (N = 682) | 27% (N = 554) |
| 2020 | 30% (N = 233) | 14% (N = 982) |
| 2021 | NA (N = 6)    | 17% (N = 75)  |

We also look at cases that have received a PS recommendation, have gone through at least 2 AZCs, and whose 1st AZC was not in a PS-recommended LMR. Among these cases, we look at all the AZCs the case went through after the first one (excluding the first one), and look at the percentage of cases for whom at least one of the AZCs besides the 1st one is in an LMR recommended by the PS. There are many reasons why a case would be transferred to a different AZC. Being transferred to an AZC in a PS-recommended LMR could indicate that at least one of the reasons might be the AO’s willingness to follow the PS advice.

***Second AZC (or beyond) and PS recommendation***

| Year | AA           | VA           |
|------|--------------|--------------|
| 2016 | 60% (N = 73) | NA (N = 1)   |
| 2017 | 52% (N = 61) | 36% (N = 22) |
| 2018 | 54% (N = 70) | 38% (N = 86) |

| Year | AA           | VA            |
|------|--------------|---------------|
| 2019 | 50% (N = 84) | 33% (N = 138) |
| 2020 | 44% (N = 43) | 22% (N = 198) |
| 2021 | NA (N = 3)   | NA (N = 7)    |

Finally, we look at different summary scenarios and journeys. The cells in grey in the table below marks different types of cases, based on their track (AA vs VA), whether they received non-blanco PS recommendation, whether their first AZC is in a PS recommended UR or LMR. Within each of these types of cases, we then look at the percentage who were linked to a PS recommended LMR, the percentage linked within their first UR/LMR/muni, and the percentage transferred to a different AZC. We also look at how many cases fall within each of the types and what percentage of overall cases each type represents.

***Full journey summary, cases with housing interview***

| Case Type | PS rec       | First AZC is in PS recommended UR | First AZC is in PS recommended LMR | % linked to a PS recommended LMR | %linked within their first UR | % linked in their first LMR | % linked in their first muni |
|-----------|--------------|-----------------------------------|------------------------------------|----------------------------------|-------------------------------|-----------------------------|------------------------------|
| AA        | Exists       | 1                                 | 1                                  | 75                               | 93                            | 69                          | 32                           |
| AA        | Exists       | 1                                 | 0                                  | 48                               | 89                            | 15                          | 6                            |
| AA        | Exists       | 0                                 | 0                                  | 38                               | 27                            | 9                           | 2                            |
| AA        | Blanco/Blank | NA                                | NA                                 | NA                               | 62                            | 26                          | 8                            |
| AA        | No SM itw    | NA                                | NA                                 | NA                               | 64                            | 28                          | 12                           |
| VA        | Exists       | 1                                 | 1                                  | 81                               | 94                            | 76                          | 27                           |
| VA        | Exists       | 1                                 | 0                                  | 68                               | 93                            | 6                           | 1                            |
| VA        | Exists       | 0                                 | 0                                  | 47                               | 27                            | 7                           | 0.9                          |
| VA        | Blanco/Blank | NA                                | NA                                 | NA                               | 64                            | 23                          | 7                            |

| Case Type | PS rec    | First AZC is in PS recommended UR | First AZC is in PS recommended LMR | % linked to a PS recommended LMR | %linked within their first UR | % linked in their first LMR | % linked in their first muni |
|-----------|-----------|-----------------------------------|------------------------------------|----------------------------------|-------------------------------|-----------------------------|------------------------------|
| VA        | No SM itw | NA                                | NA                                 | NA                               | 51                            | 23                          | 9                            |

| Case Type | PS rec          | First AZC is in PS recommended UR | First AZC is in PS recommended LMR | % transferred to different AZC | % of overall cases (since 2018) | N cases (since 2018) |
|-----------|-----------------|-----------------------------------|------------------------------------|--------------------------------|---------------------------------|----------------------|
| AA        | Exists          | 1                                 | 1                                  | 12                             | 7.6                             | 634                  |
| AA        | Exists          | 1                                 | 0                                  | 15                             | 4.6                             | 389                  |
| AA        | Exists          | 0                                 | 0                                  | 33                             | 5                               | 422                  |
| AA        | Blanco/Blank    | NA                                | NA                                 | 18                             | 6.6                             | 556                  |
| AA        | No SM itw       | NA                                | NA                                 | 21                             | 2.4                             | 198                  |
| VA        | Exists          | 1                                 | 1                                  | 12                             | 4.2                             | 351                  |
| VA        | Exists          | 1                                 | 0                                  | 22                             | 6.5                             | 541                  |
| VA        | Exists          | 0                                 | 0                                  | 32                             | 11.6                            | 969                  |
| VA        | Blanco/Blank    | NA                                | NA                                 | 25                             | 25.6                            | 2137                 |
| VA        | No SM interview | NA                                | NA                                 | 32                             | 22.3                            | 1871                 |

### g. Descriptives, dates checks

The table below shows distribution for two summary statistics (still among cases with no hard criteria or family reunifiers):

- The number of days between the housing interview (housing interview, Datum huisvestingsgesprek 1) and the linking by the AO (Datum koppeling aan gemeente), among cases who receive a housing interview
- The number of linkings (koppeling aan gemeente) per day, across all AOs, for cases with no family reunifier or hard criteria

We show distribution summaries for all data since 2014, as well as for data starting in 2019.

**Summary distribution linkings/day and gap between SM and linking**

|              | N days between housing and linking (since 2014) | N days between housing and linking (since 2019) | N linkings/days (since 2014) | N linkings/day (since 2019) |
|--------------|-------------------------------------------------|-------------------------------------------------|------------------------------|-----------------------------|
| Min          | -5843                                           | -211                                            | 1                            | 1                           |
| 1st quartile | 4                                               | 2                                               | 5                            | 4                           |
| Median       | 8                                               | 5                                               | 11                           | 8                           |
| Mean         | 38.79                                           | 14.65                                           | 18.22                        | 9.4                         |
| 3rd quartile | 25                                              | 10                                              | 22                           | 13                          |
| Max          | 1723                                            | 555                                             | 185                          | 48                          |

**Summary distribution linkings/day and gap between SM and linking, AA**

|              | N days between housing and linking (since 2014) | N days between housing and linking (since 2019) | N linkings/days (since 2014) | N linkings/day (since 2019) |
|--------------|-------------------------------------------------|-------------------------------------------------|------------------------------|-----------------------------|
| Min          | -5843                                           | 2                                               | 1                            | 1                           |
| 1st quartile | 5                                               | 4                                               | 1                            | 1                           |
| Median       | 8                                               | 6.5                                             | 3                            | 2                           |
| Mean         | 28.17                                           | 19.56                                           | 4.58                         | 3.4                         |
| 3rd quartile | 17                                              | 12                                              | 5                            | 5                           |
| Max          | 1723                                            | 545                                             | 52                           | 20                          |

**Summary distribution linkings/day and gap between SM and linking, VA**

|              | N days between housing and linking (since 2014) | N days between housing and linking (since 2019) | N linkings/days (since 2014) | N linkings/day (since 2019) |
|--------------|-------------------------------------------------|-------------------------------------------------|------------------------------|-----------------------------|
| Min          | -406                                            | -211                                            | 1                            | 1                           |
| 1st quartile | 4                                               | 2                                               | 4                            | 3                           |
| Median       | 8                                               | 5                                               | 8                            | 6                           |
| Mean         | 40.39                                           | 13.48                                           | 15.03                        | 7.3                         |
| 3rd quartile | 27                                              | 9                                               | 18                           | 10                          |
| Max          | 1352                                            | 555                                             | 169                          | 43                          |

#### h. Covariate imbalances in historical data

For the purposes of the monitoring plan, we will track the distributional balance for the following covariates and groups, which have been agreed upon by COA and IPL:

- Age: individuals 40 years old and older
- Gender: female
- Country of origin: Syrians
- Case size (cases with 5 or more individuals)
- Education level: individuals with low education levels

Because imbalances in terms of covariates distributions across LMRs also appear in the current process, we want to be able to detect “extra” imbalance potentially generated by GeoMatch. Imbalances at each LMR for each year during 2014 and 2021 were computed in the following way:

- Percentage of all linkages who were female during a given year:  $X\%$
- Percentage of all linkages at a given LMR,  $i$ , who were female during a given year:  $Y_i\%$
- We define the imbalance at LMR  $i$  during a given year as:  $\text{abs}(Y_i - X)$ . This is the percentage point difference between LMR  $i$ 's percentage of linkages that were females in a given year and the overall percentage of linkages that were females that were linked that year.

We then look at the maximum value for imbalances at any LMR that received at least 30 individuals or 10 cases during the year. We also export the standard deviations of the full distributions of imbalances on each covariate. We compute the threshold for

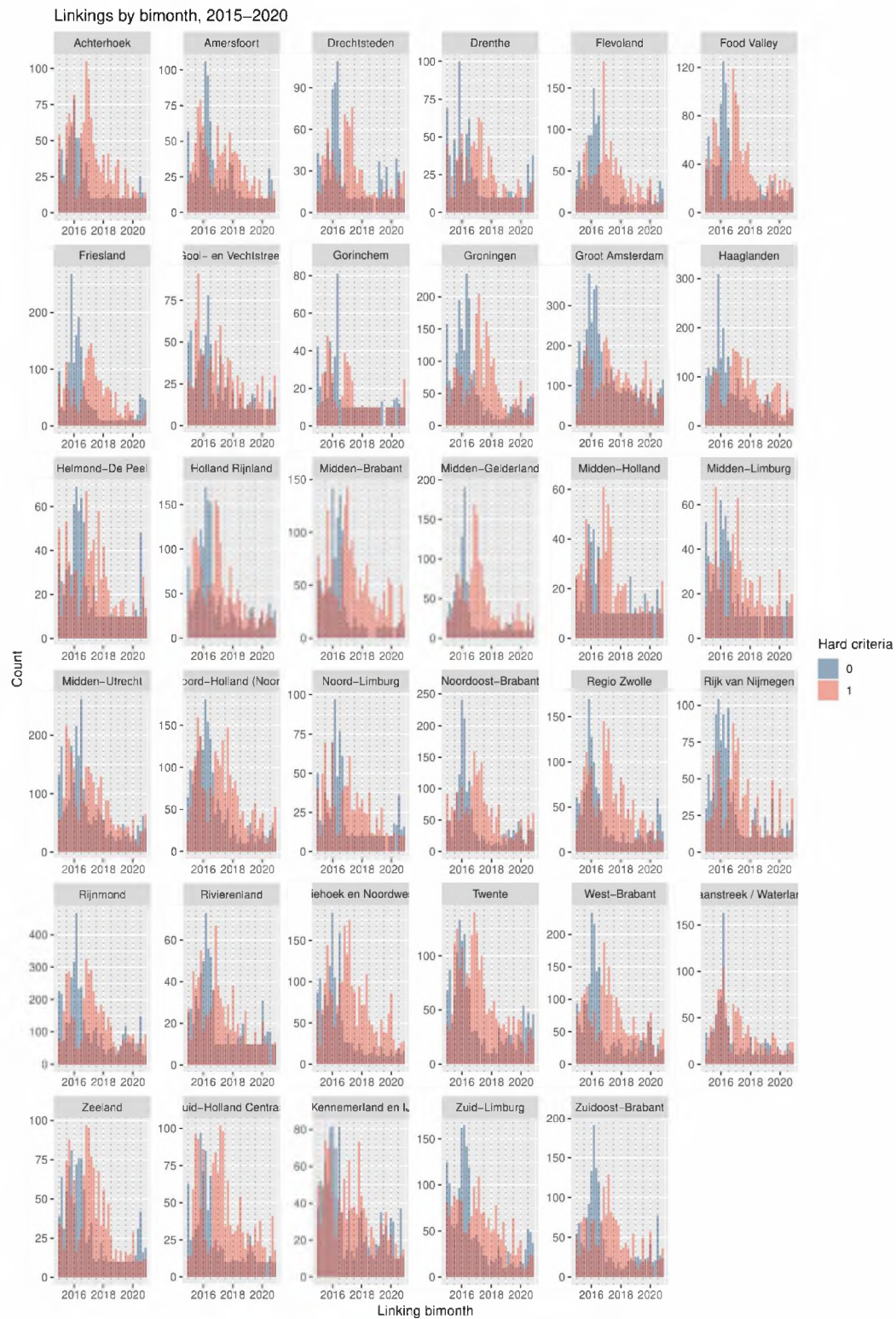
‘extra’ imbalances as any imbalance in the pilot that would be more than x standard deviations from the maximum that already happened naturally in the historical data. The value of x is determined by the confidence level we apply to the analysis: 90%, 95% or 99%. We opted for the most commonly used 95% confidence level (see monitoring plan for more details).

| <b>Covariate</b>  | <b>Max(Imbalance), in percentage points in historical data</b> | <b>Standard Deviation</b> | <b>Threshold for ‘extra’ imbalance if measured in pilot (percentage points) 95% confidence level</b> | <b>Threshold for ‘extra’ imbalance if measured in pilot (percentage points) 90% confidence level</b> | <b>Threshold for ‘extra’ imbalance if measured in pilot (percentage points) 99% confidence level</b> |
|-------------------|----------------------------------------------------------------|---------------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Female            | 21.4                                                           | 4                         | 29.2                                                                                                 | 27.94                                                                                                | 31.66                                                                                                |
| Low Education     | 15.8                                                           | 6.3                       | 28.1                                                                                                 | 26.07                                                                                                | 31.91                                                                                                |
| Syrians           | 27.8                                                           | 6.3                       | 40.09                                                                                                | 38.11                                                                                                | 43.96                                                                                                |
| Over 40 years old | 21.3                                                           | 3.1                       | 27.46                                                                                                | 26.47                                                                                                | 29.37                                                                                                |
| Cases of size 5+  | 11.2                                                           | 1.5                       | 14.04                                                                                                | 13.58                                                                                                | 14.95                                                                                                |

i. Proportion of hard criteria to non hard criteria individuals linked to each LMR every month

When setting up the capacity constraints and load balancing for GeoMatch, it is important to understand current patterns of linkings under the status quo COA procedure. In the following plots, we look at how many hard criteria/family reunifiers (red) and non hard criteria/family reunifiers individuals are linked to each municipality every month. Because of CBS export limitations, we cannot export or plot any count below 10. This prevents us from looking at more granular counts at the week level, for example. In the first plot below, we look at all 35 LMRs over the 5 years during which we have full data (2015-2020) and plot counts of linkings every 2 months. All counts below 10 were rounded up to 10

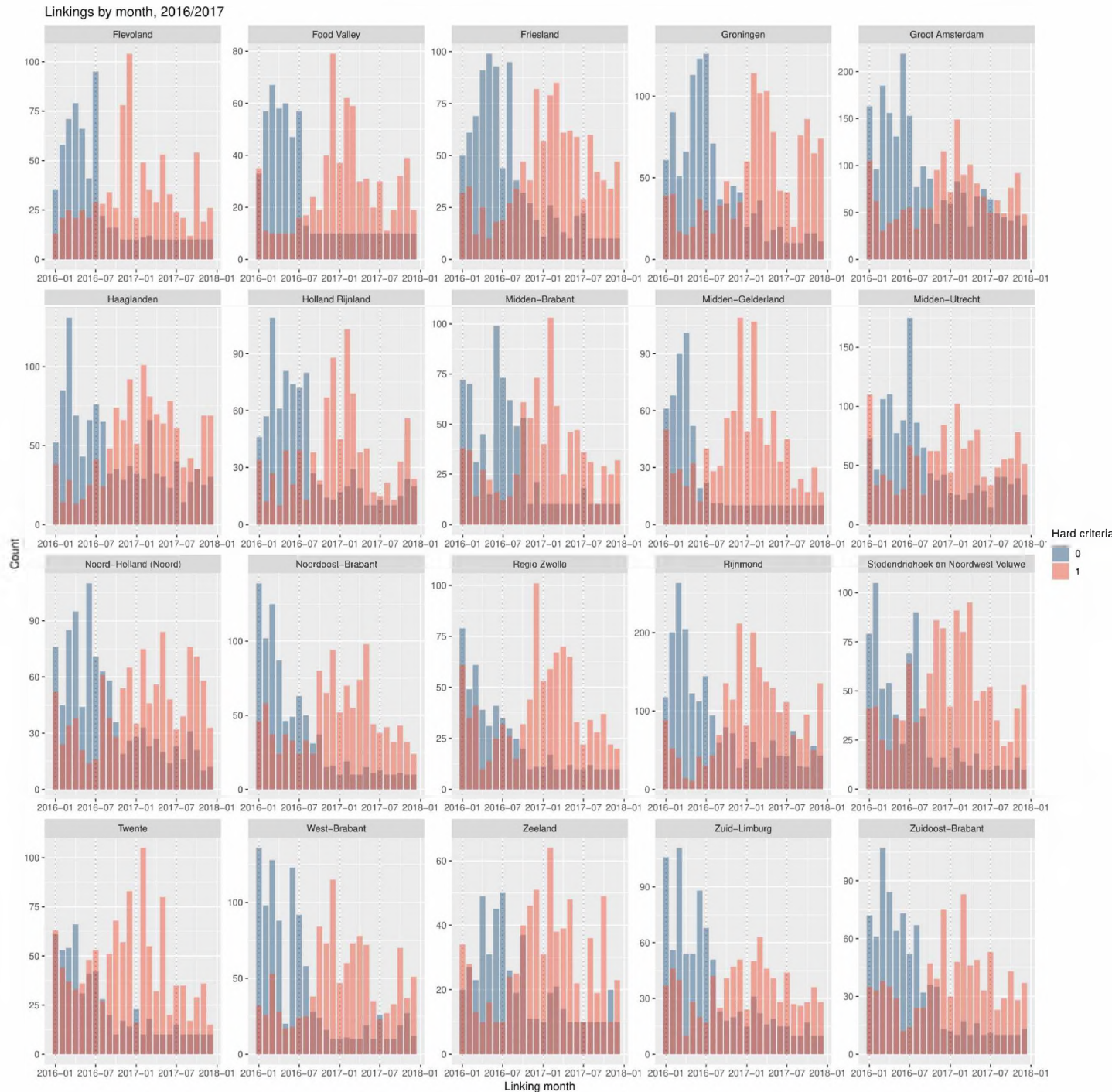
## Linkings per bimonth, 2015/2020



Black vertical lines are new taasktelling periods in January and July. N = 123,045.  
394/2498 of the subgroup counts were under 10 and were rounded up to 10 to follow CBS export rules

In the second plot, we zoom in on two years (2016/2017). Because of CBS export limitations, we could only export the 2 years with the highest numbers of linkings, and only for the 20 largest LMRs. All counts below 10 were rounded up to 10.

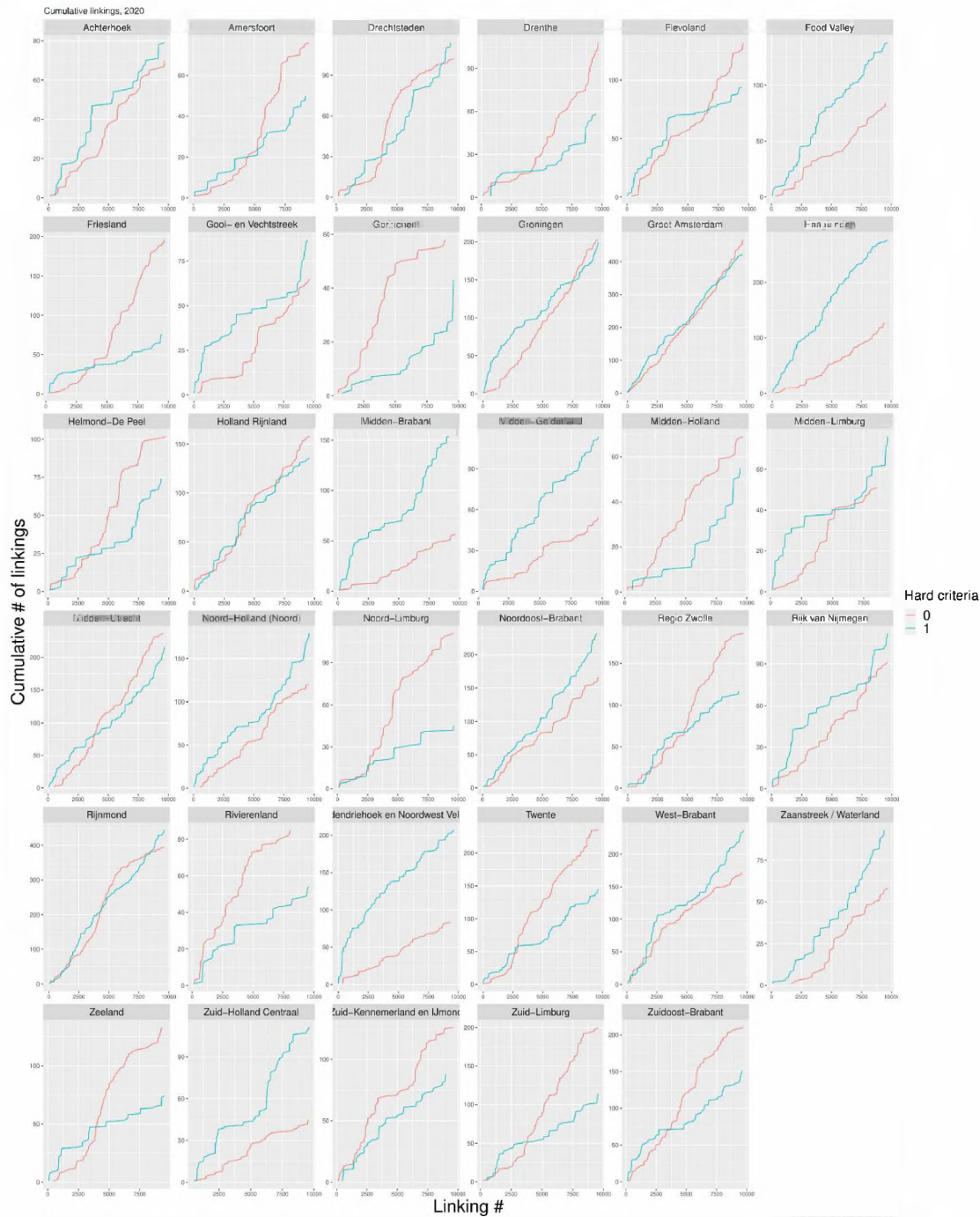
## Linkings per month, 2016/2017 (top 20 LMRs)



Black vertical lines are new taaskstelling periods in January and July. N = 52636. Only top 20 LMRs plotted. 131/960 of the subgroup counts were under 10 and were rounded up to 10 to follow CBS export rules

In the third plot, we look at the cumulative number of linkings at each LMR in 2020. The x axis plots each linking at any LMR. The Y axis plots the cumulative number of linkings at each LMR. In a perfectly load balanced world, the lines would be diagonal. This plot shows deviation from perfect load balancing in the current status quo linking process.

## Cumulative linkings, 2020



### j. Length of stay at AZC and employment

One justification for the VA restriction policy within COA is that once a VA case has spent a lot of time in an AZC before being linked, they are more likely to integrate if they are linked to a municipality near their current AZC. The assumption is that the permit holders have had time to develop local networks and habits that eventually

facilitate their integration. We cannot fully test this assumption, as this is a causal question - i.e, it is possible that permit holders who stay longer or are linked locally are different from permit holders who stay less long or are linked elsewhere, preventing us from drawing any conclusion from a descriptive finding. With this important caveat in mind, we still explored descriptive patterns of employment.

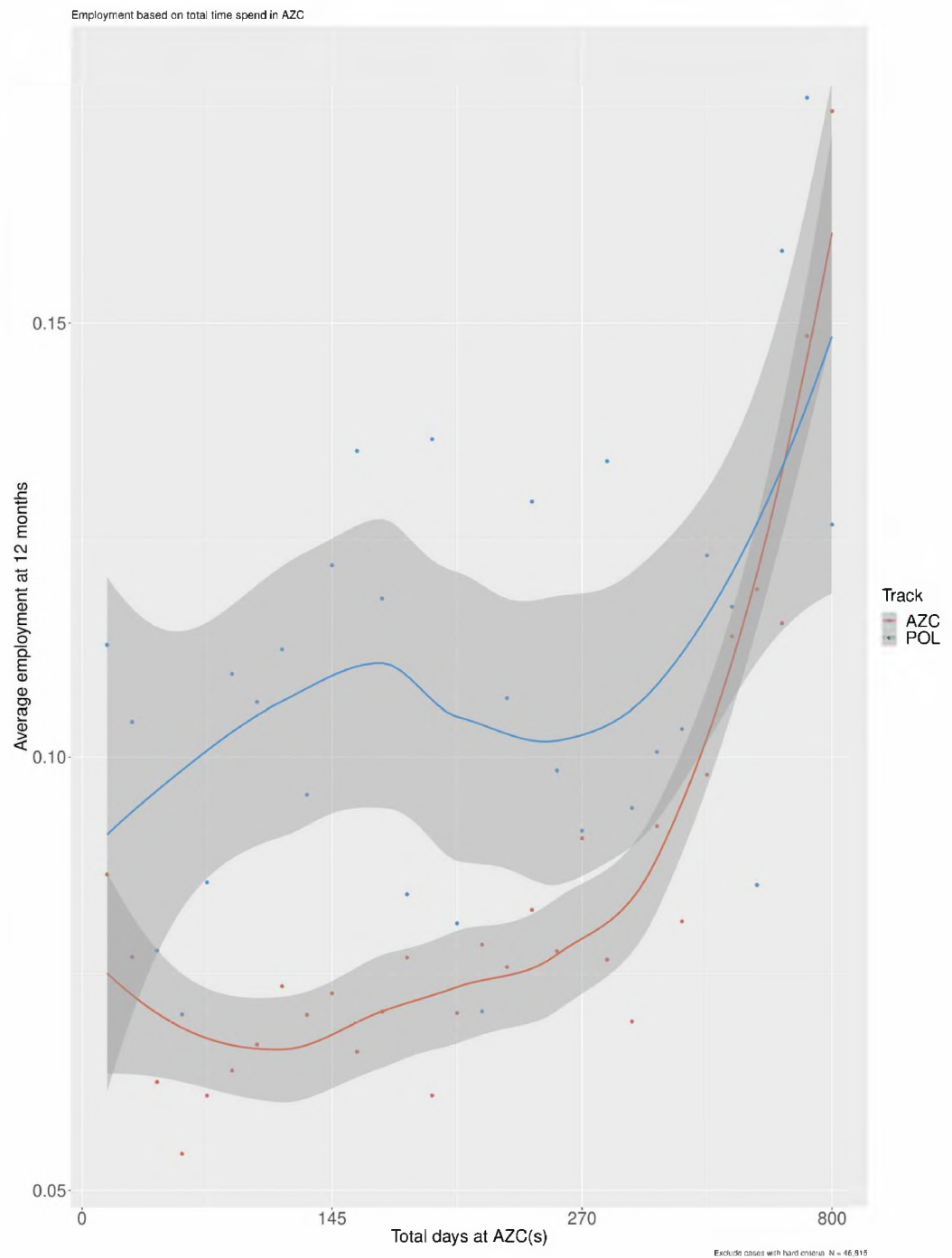
First, we look at the distribution of length of stay in days at an AZC for AA and VA cases. This includes all adults with no family reunifier or hard criteria and, for adults who went through multiple AZCs, includes the sum of days spent at any AZC.

***Distribution summaries, length of stay at any AZC (days)***

|        | AA    | VA    |
|--------|-------|-------|
| Min    | 0     | 0     |
| Q1     | 91    | 147   |
| Median | 153   | 226   |
| Mean   | 192.6 | 273.3 |
| Q3     | 252   | 337   |
| Max    | 1694  | 2930  |

In the plot below, we plot the average employment rate based on the length of stay at the AZC, for both VA and AA cases. This plot suggests that the longer an individual stays in an AZC, the more likely they are to find employment one year after having left the AZC. This pattern is especially true for VA cases. Again, we cannot causally attribute these higher employment rates to having stayed at the AZC for a longer period: it is also possible that individuals who stayed longer were different from individuals who left the AZC quickly.

**Employment based on total time spent in AZC, AA vs VA**



k. Local linkings

All the data in this section is based on a geographic mapping of AZCs to LMRs: we found the localization of each of the 120 AZCs in the IBIS data, and matched them to

the LMR they are geographically in. We know that most AZCs ‘serve’ the UR and LMR they are geographically in, but that is not always the case. Every 3 months, COA adjusts the mapping of AZC to UR, in order to respond to differing needs of cases from different URs. Therefore, it can happen that an AZC is geographically in UR1, but ‘serves’ a different UR. Those situations remain exceptions, and since we do not have access to the mappings for each time period in our historical data, all the analyses below assume a geographic mapping.

## **UR level**

The table below shows the percentage of cases linked locally (i.e., linked to a municipality that is geographically in the same UR as the AZC the case was staying in on the date the case was linked/gekoppeld).

**Percentage of local linkings, local = same UR**

| Year    | AA  | VA  |
|---------|-----|-----|
| 2014    | 71% | 62% |
| 2015    | 66% | 64% |
| 2016    | 65% | 69% |
| 2017    | 64% | 60% |
| 2018    | 72% | 59% |
| 2019    | 77% | 66% |
| 2020    | 70% | 70% |
| 2021    | 64% | 75% |
| Overall | 69% | 65% |

The table below shows the same thing, but only among cases that received at least 1 regional advice from the PS.

**Percentage of local linkings, among cases having received at least 1 PS advice**

| Year | AA            | VA         |
|------|---------------|------------|
| 2016 | 88% (N = 212) | NA (N = 1) |

| Year | AA            | VA            |
|------|---------------|---------------|
| 2017 | 78% (N = 354) | 50% (N = 68)  |
| 2018 | 80% (N = 524) | 60% (N = 250) |
| 2019 | 70% (N = 682) | 66% (N = 554) |
| 2020 | 72% (N = 233) | 67% (N = 982) |
| 2021 | NA (N = 6)    | 70% (N = 75)  |

The table below shows the same thing, but only among cases that received a blanco or blank advice (but still received a SM interview).

***Percentage of local linkings, among cases having received a blanco advice from PS***

| Year | AA            | VA              |
|------|---------------|-----------------|
| 2016 | 76% (N = 123) | NA (N = 3)      |
| 2017 | 52% (N = 207) | 44% (N = 44)    |
| 2018 | 57% (N = 267) | 63% (N = 254)   |
| 2019 | 72% (N = 184) | 71% (N = 625)   |
| 2020 | 77% (N = 104) | 73% (N = 1,185) |
| 2021 | NA (N = 1)    | 86% (N = 73)    |

The table below shows the same thing, but only among cases that did not receive a SM interview).

***Percentage of local linkings, among cases having received no SM interview***

| Year | AA            | VA             |
|------|---------------|----------------|
| 2014 | 70% (N = 317) | 62% (N = 4503) |
| 2015 | 66% (N = 309) | 64% (N = 9753) |

| Year | AA             | VA             |
|------|----------------|----------------|
| 2016 | 64% (N = 2360) | 69% (N = 9424) |
| 2017 | 59% (N = 390)  | 60% (N = 1957) |
| 2018 | 72% (N = 72)   | 58% (N = 778)  |
| 2019 | 73% (N = 15)   | 60% (N = 472)  |
| 2020 | 61% (N = 107)  | 65% (N = 588)  |
| 2021 | NA (N = 4)     | 61% (N = 33)   |

### **LMR level**

The table below shows the percentage of cases linked locally (i.e., linked to a municipality that is geographically in the same LMR as the last AZC the case was staying in).

#### **Percentage of local linkings, local = same LMR**

| Year    | AA             | VA             |
|---------|----------------|----------------|
| 2014    | 20% (N = 317)  | 20% (N = 4503) |
| 2015    | 25% (N = 309)  | 21% (N = 9753) |
| 2016    | 32% (N = 2695) | 29% (N = 9428) |
| 2017    | 35% (N = 951)  | 32% (N = 2069) |
| 2018    | 36% (N = 863)  | 28% (N = 1282) |
| 2019    | 42% (N = 881)  | 33% (N = 1651) |
| 2020    | 35% (N = 444)  | 26% (N = 2755) |
| 2021    | 45% (N = 11)   | 25% (N = 181)  |
| Overall | 33%            | 26%            |

The table below shows the same thing, but only among cases that received at least 1 regional advice from the PS.

**Percentage of local linkings, among cases having received at least 1 PS advice**

| Year | AA            | VA            |
|------|---------------|---------------|
| 2016 | 52% (N = 212) | NA (N = 1)    |
| 2017 | 42% (N = 354) | 19% (N = 68)  |
| 2018 | 41% (N = 524) | 30% (N = 250) |
| 2019 | 45% (N = 682) | 33% (N = 554) |
| 2020 | 39% (N = 233) | 22% (N = 982) |
| 2021 | NA (N = 6)    | 19% (N = 75)  |

The table below shows the same thing, but only among cases that received a blanco or blank advice (but still received a SM interview).

**Percentage of local linkings, among cases having received a blanco advice from PS**

| Year | AA            | VA             |
|------|---------------|----------------|
| 2016 | 31% (N = 123) | NA (N = 3)     |
| 2017 | 26% (N = 207) | 25% (N = 44)   |
| 2018 | 27% (N = 267) | 25% (N = 254)  |
| 2019 | 31% (N = 184) | 31% (N = 625)  |
| 2020 | 37% (N = 104) | 28% (N = 1185) |
| 2021 | NA (N = 1)    | 27% (N = 73)   |

The table below shows the same thing, but only among cases that did not receive a SM interview).

**Percentage of local linkings, among cases having received no SM interview**

| Year | AA            | VA             |
|------|---------------|----------------|
| 2014 | 19% (N = 317) | 20% (N = 4503) |

| Year | AA             | VA             |
|------|----------------|----------------|
| 2015 | 25% (N = 309)  | 21% (N = 9753) |
| 2016 | 30% (N = 2360) | 29% (N = 9424) |
| 2017 | 34% (N = 390)  | 32% (N = 1957) |
| 2018 | 39% (N = 72)   | 29% (N = 778)  |
| 2019 | 47% (N = 15)   | 34% (N = 472)  |
| 2020 | 24% (N = 107)  | 30% (N = 588)  |
| 2021 | NA (N = 4)     | 36% (N = 33)   |

### ***Muni level***

The table below shows the percentage of cases linked locally (ie linked to a municipality that is geographically in the same municipality as the last AZC the case was staying in).

#### ***Percentage of local linkings, local = same Muni***

| Year    | AA    | VA    |
|---------|-------|-------|
| 2014    | 1.6%  | 3.1%  |
| 2015    | 3.9%  | 5.2%  |
| 2016    | 10.4% | 10.2% |
| 2017    | 11.6% | 11.3% |
| 2018    | 15.8% | 10.4% |
| 2019    | 17.3% | 13%   |
| 2020    | 13.7% | 7.9%  |
| 2021    | 9%    | 11.6% |
| Overall | 11.8% | 7.7%  |

## Diversity of LMRs

Another way to look at the willingness of AOs to link cases non-locally is to look at the diversity of LMRs a given AZC links cases to. We look at all the linkings an AZC makes over the course of 2 months. Among all combinations of 120 AZCs and 46 bimonths, we only keep combinations with at least 5 linkings (we keep 34% of the combinations after applying this filter). Within these combinations, we look at 2 metrics (see tables below for distribution summaries of these metrics):

- The number of unique LMR linked to during each bimonth
- The proportion of linkings made at this AZC during that bimonth that each LMR receives.

For example, if AZC 1 during the Jan-March 2017 period linked 20 cases to LMR 1, 10 cases to LMR 2, 1 case to LMR 3 and 1 to LMR 4, the number of unique LMR linked to for this AZC x bimonth combination is 4, and the proportion of linkings we save would be 62.5 (20/32), 16.1 (10/32), 3 (1/32) and 3 (1/32).

### ***Diversity of LMRs linkings at each AZC, AA and VA (34% of the combinations remain)***

|        | % of this AZC's bimonth linkings going to each LMR | N Unique LMRs |
|--------|----------------------------------------------------|---------------|
| Min    | 0.36%                                              | 1             |
| Q1     | 3.7%                                               | 5             |
| Median | 7.7%                                               | 7             |
| Mean   | 12%                                                | 8.3           |
| Q3     | 14.2%                                              | 10            |
| Max    | 100%                                               | 32            |

### ***Diversity of LMRs linkings at each AZC, AA (10% of the combinations remain)***

|     | % of this AZC's bimonth linkings going to each LMR | N Unique LMRs |
|-----|----------------------------------------------------|---------------|
| Min | 1.23%                                              | 1             |
| Q1  | 8.33%                                              | 4             |

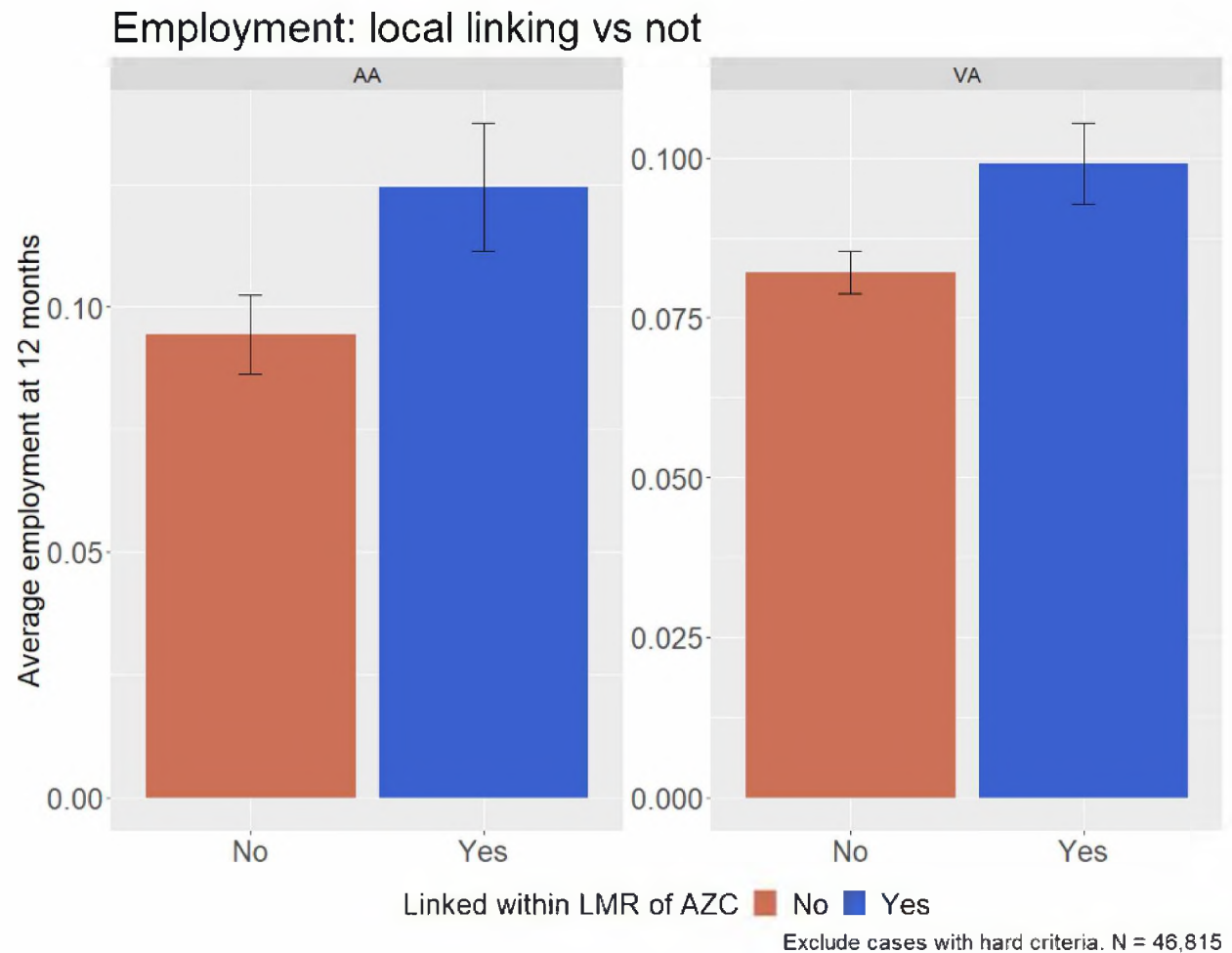
|        | % of this AZC's bimonth linkings going to each LMR | N Unique LMRs |
|--------|----------------------------------------------------|---------------|
| Median | 13.33%                                             | 5             |
| Mean   | 19%                                                | 5.3           |
| Q3     | 22.22%                                             | 6             |
| Max    | 100%                                               | 20            |

***Diversity of LMRs linkings at each AZC, VA (30% of the combinations remain)***

|        | % of this AZC's bimonth linkings going to each LMR | N Unique LMRs |
|--------|----------------------------------------------------|---------------|
| Min    | 0.38%                                              | 1             |
| Q1     | 4.2%                                               | 5             |
| Median | 8.8%                                               | 7             |
| Mean   | 13.2%                                              | 7.6           |
| Q3     | 16.7%                                              | 9             |
| Max    | 100%                                               | 23            |

In the plot below, we look at employment levels among cases placed locally and among cases not placed locally (defining local as the municipality and AZC being in the same LMR). This plot does not show the causal effect of being placed locally, however, since it's possible that permit holders placed locally are different from permit holders not placed locally.

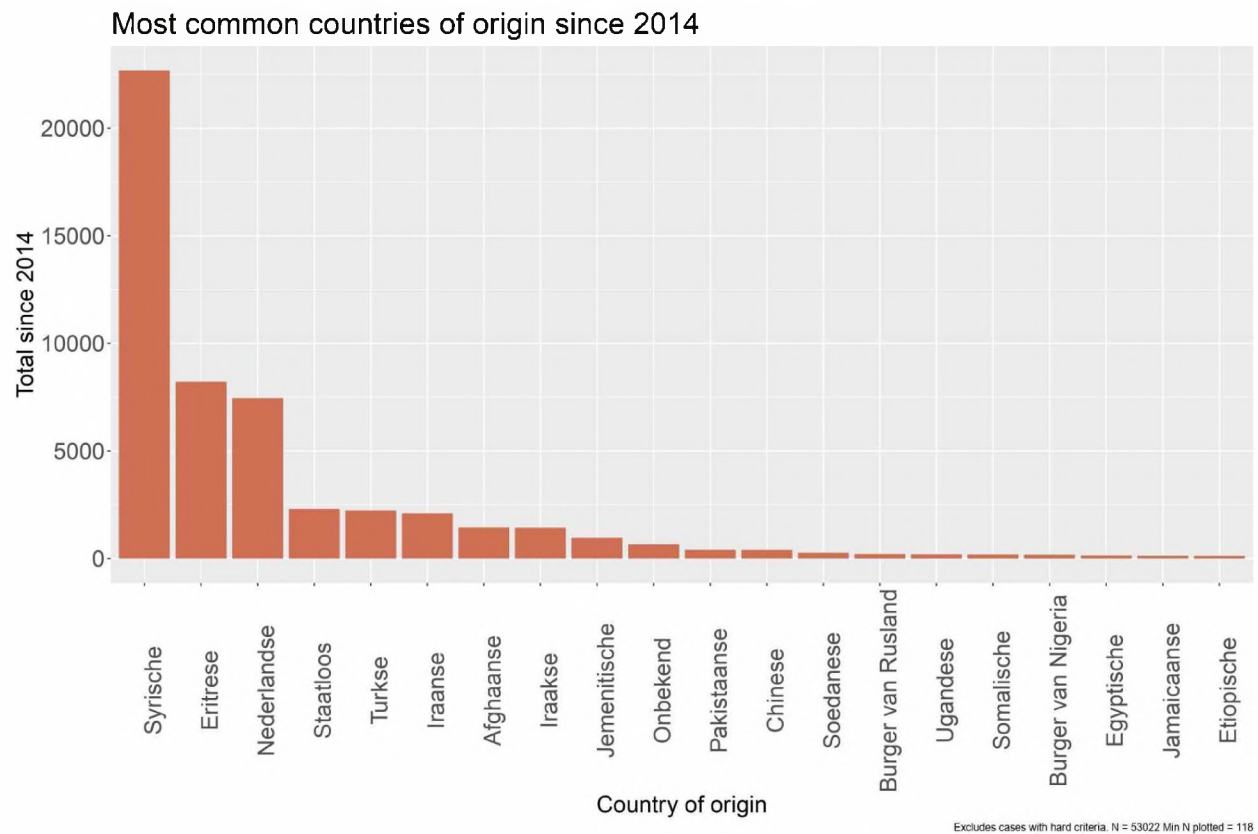
**Employment based on whether the case was linked to local LMR or not (ie same LMR as the AZC they were in)**



## I. Nationality distribution

The plot below shows the most frequent nationalities for cases with no hard criteria since 2014. If a permit holder arrived in 2014 and since then received the Dutch nationality, IBIS would be updated and show Dutch as the nationality for this case. For model training purposes, we will use country of origin rather than nationality.

**Most common nationalities since 2014**



**m. Secondary migration**

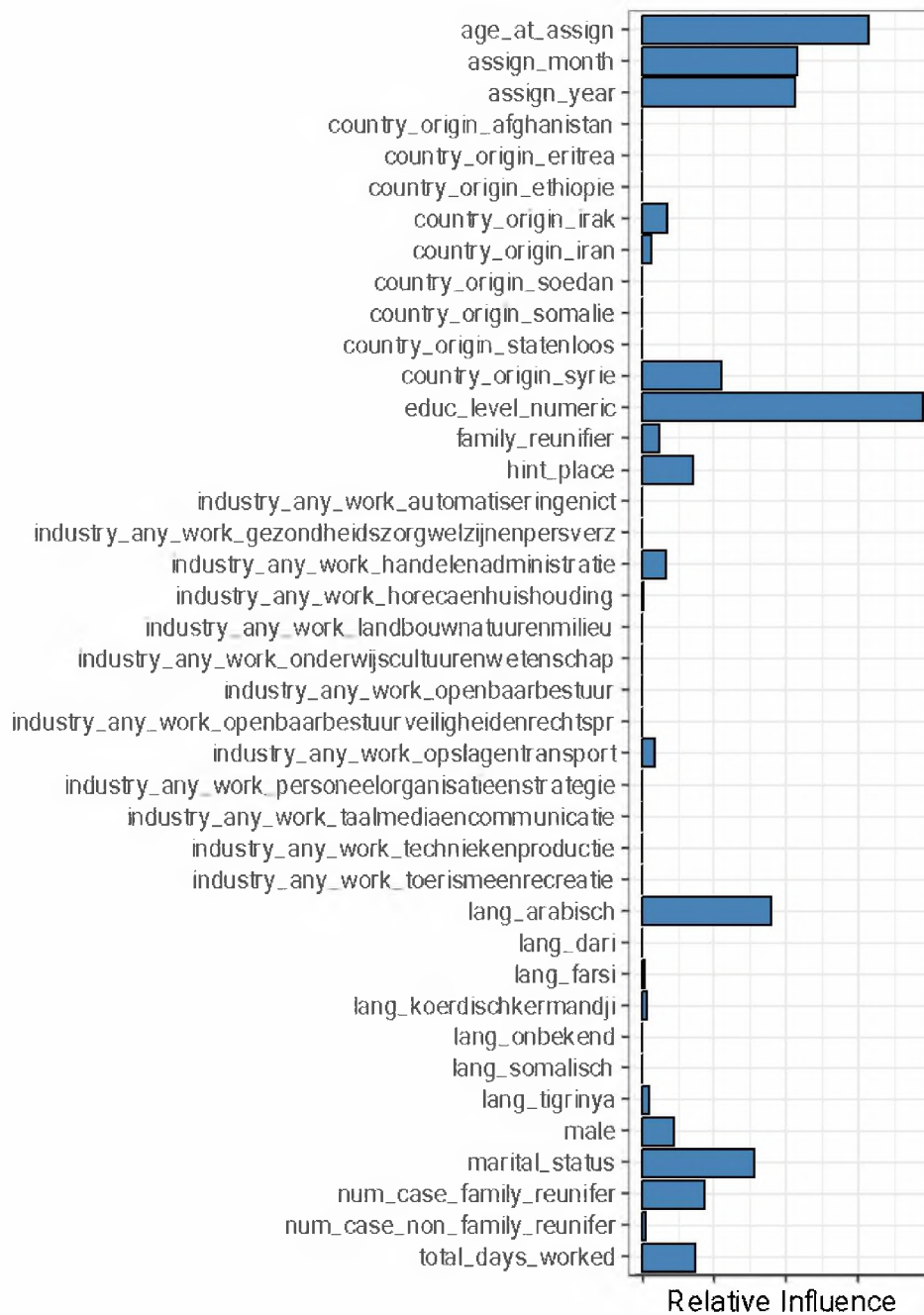
The table below shows the share of adult permit holders without hard criteria that remained in the municipality or LMR that they were assigned to for the 12 months after assignment. This was calculated using the variable *gemeente2021* from the Asielcohort dataset and the variable *Gekoppeldaangemeente* from the IBIS data. Less than five percent of permit holders outmigrate to another LMR.

| <b>Year of Assignment</b> | <b>% Remaining in Municipality of Assignment</b> | <b>% Remaining in LMR of Assignment</b> |
|---------------------------|--------------------------------------------------|-----------------------------------------|
| <b>2014</b>               | 92%                                              | 96%                                     |
| <b>2015</b>               | 93%                                              | 96%                                     |
| <b>2016</b>               | 94%                                              | 97%                                     |
| <b>2017</b>               | 93%                                              | 97%                                     |
| <b>2018</b>               | 94%                                              | 97%                                     |
| <b>2019</b>               | 96%                                              | 97%                                     |
| <b>2020</b>               | 96%                                              | 97%                                     |
| <b>2021</b>               | 95%                                              | 97%                                     |

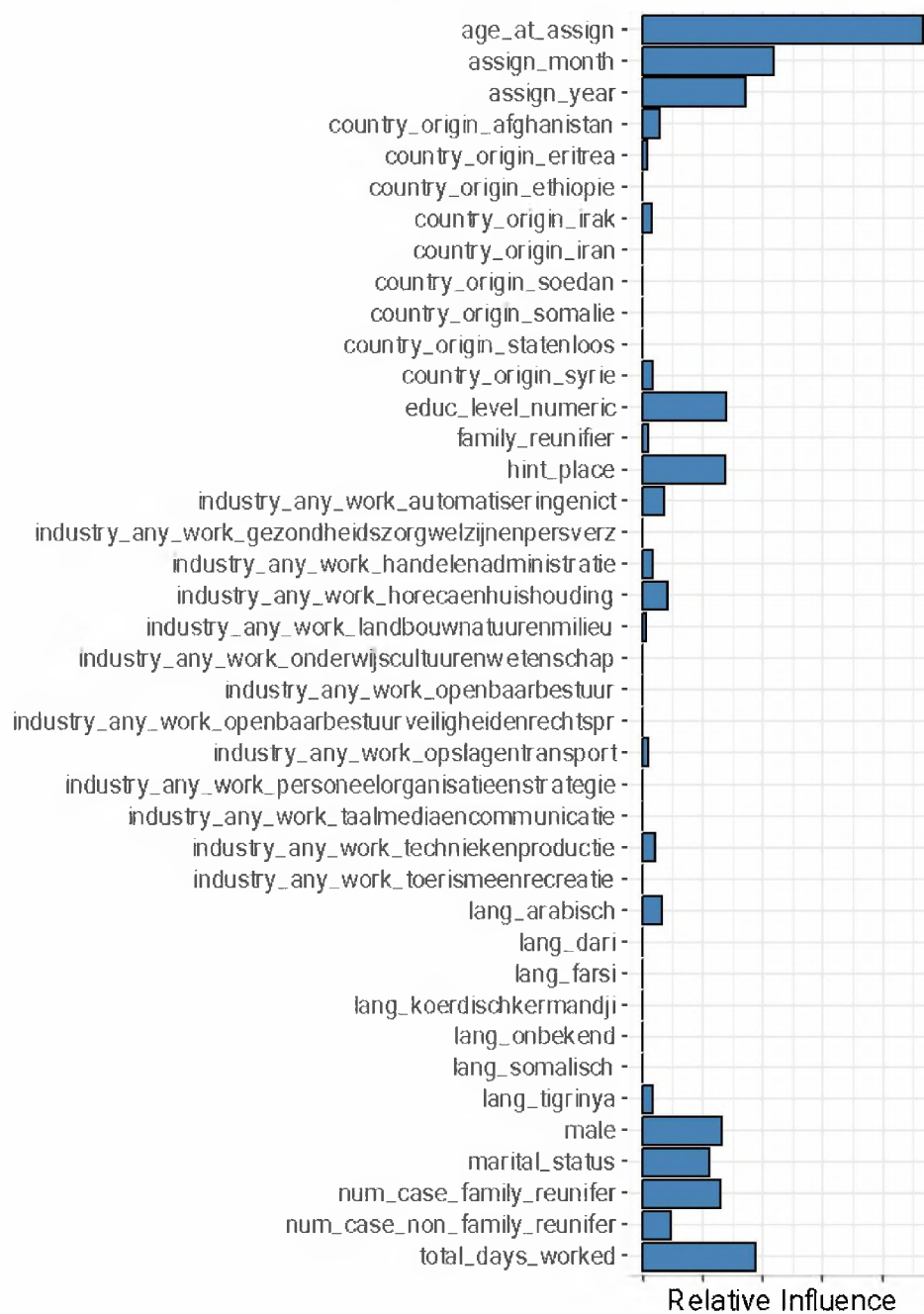
### 3. Appendix 3: Variable Importance Plots

These variable importance plots were prepared based on the set of predictors used in the backtest. It means they include two predictor variables (religion and ethnicity) that will not be used in the pilot because of GDPR restrictions.

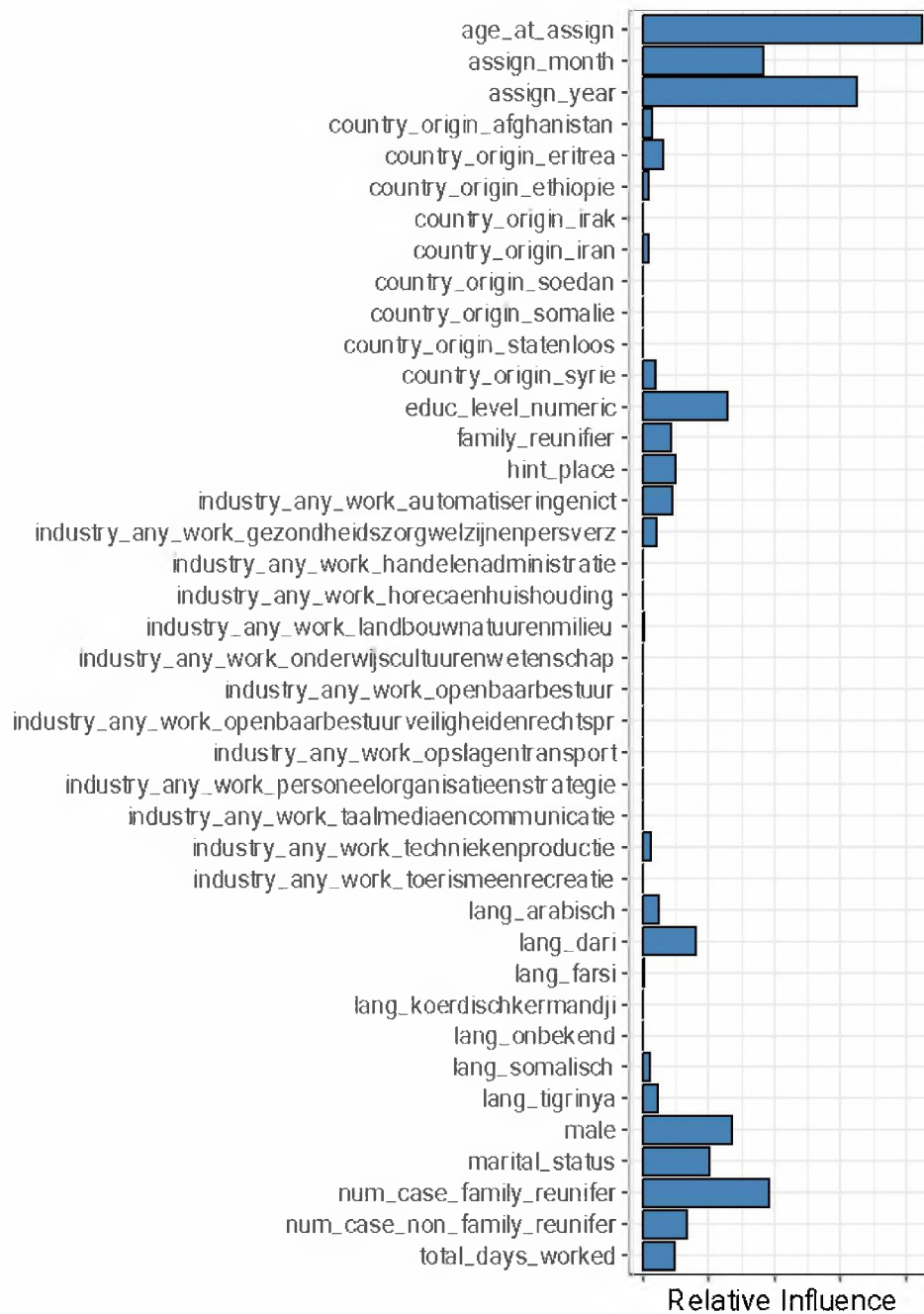
### Achterhoek



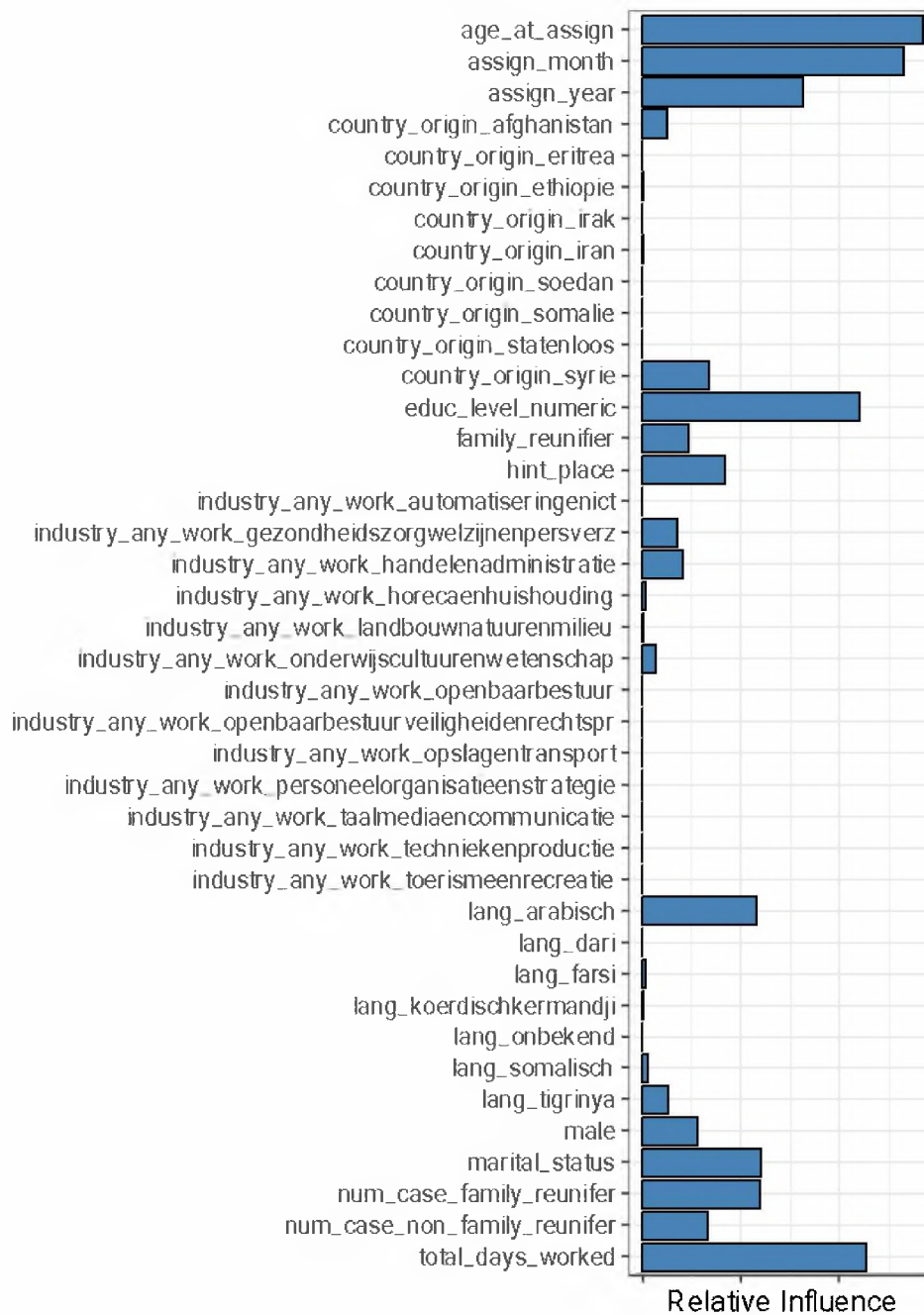
**Amersfoort**



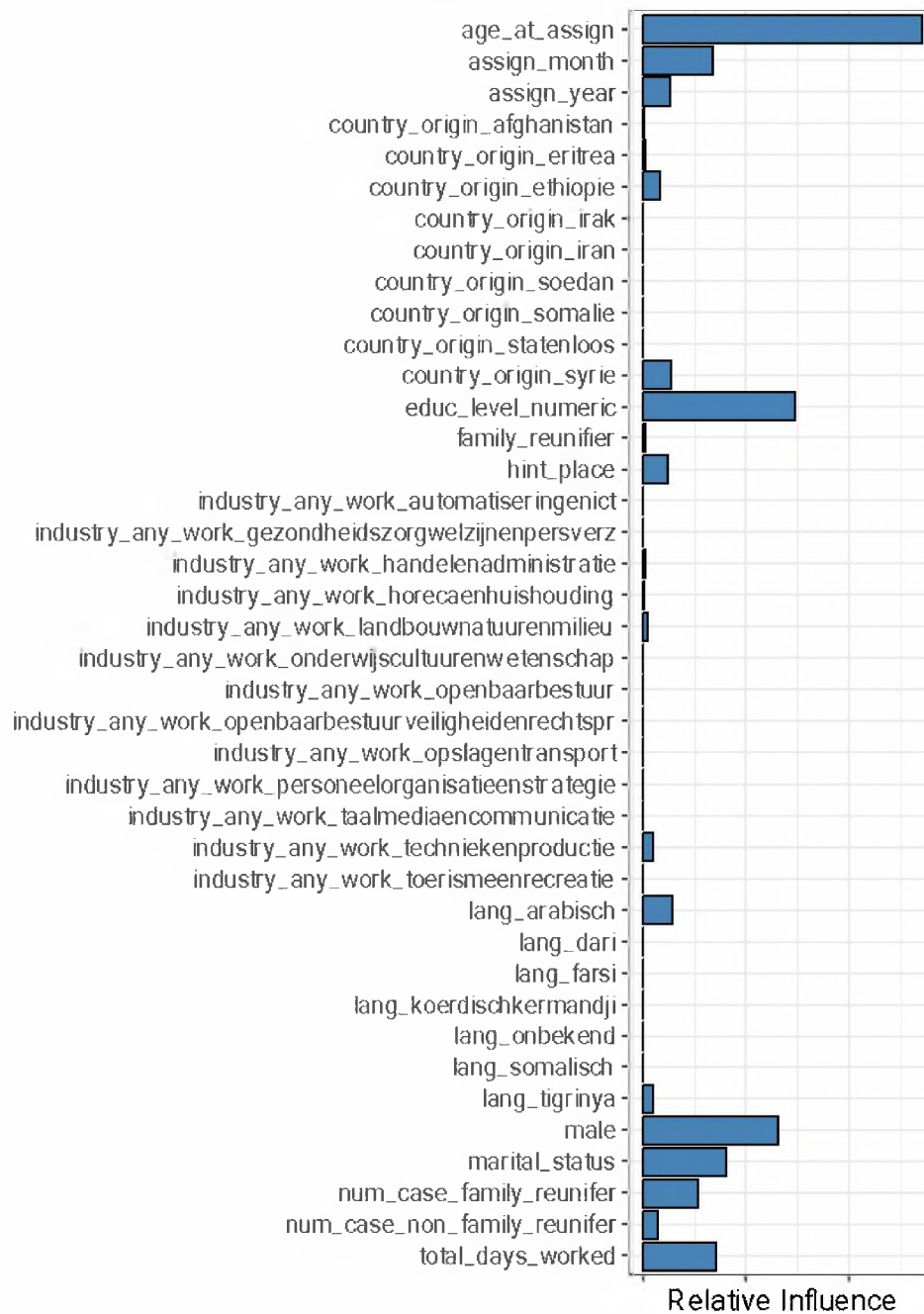
## ***Drechtsteden***



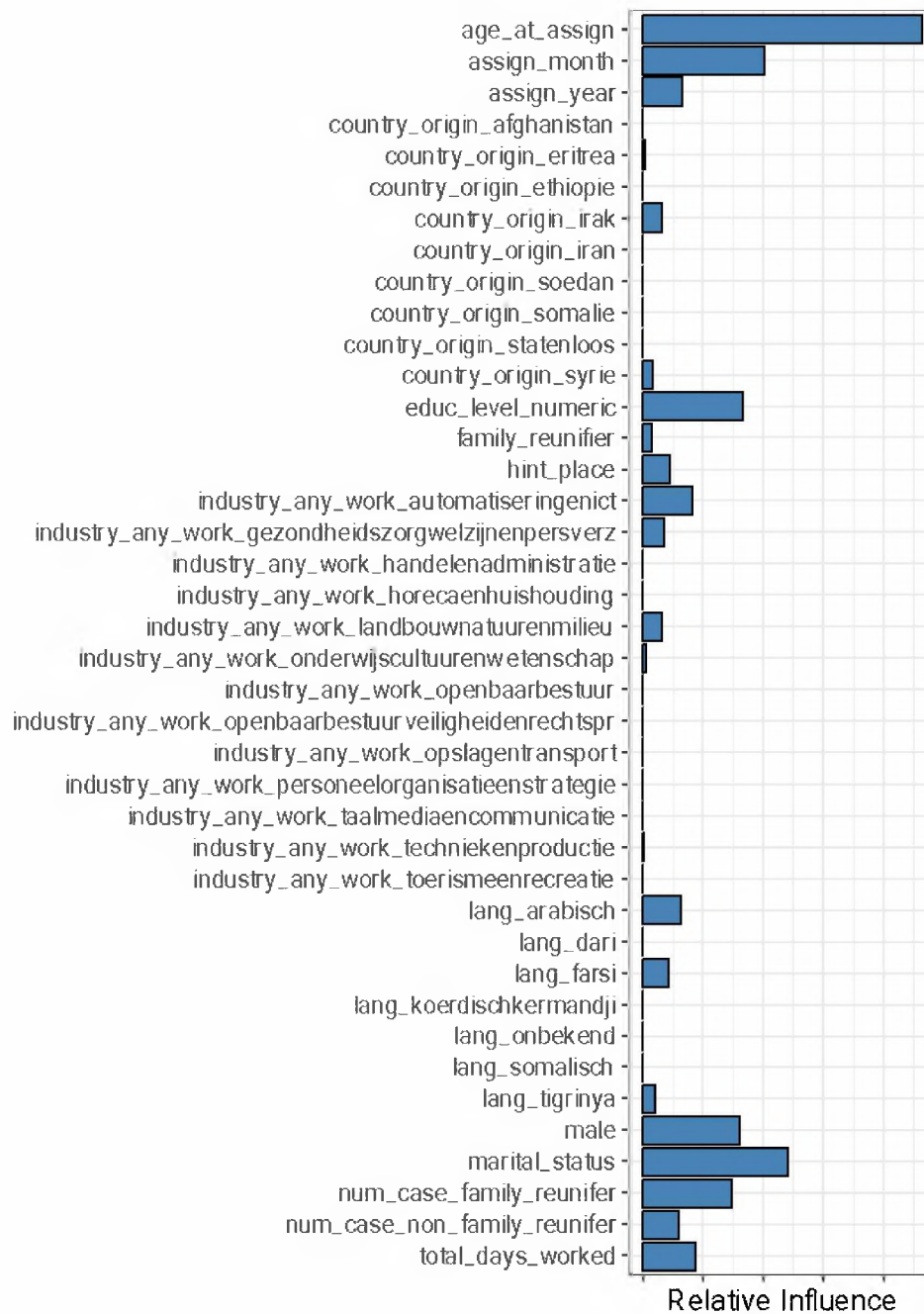
**Drenthe**



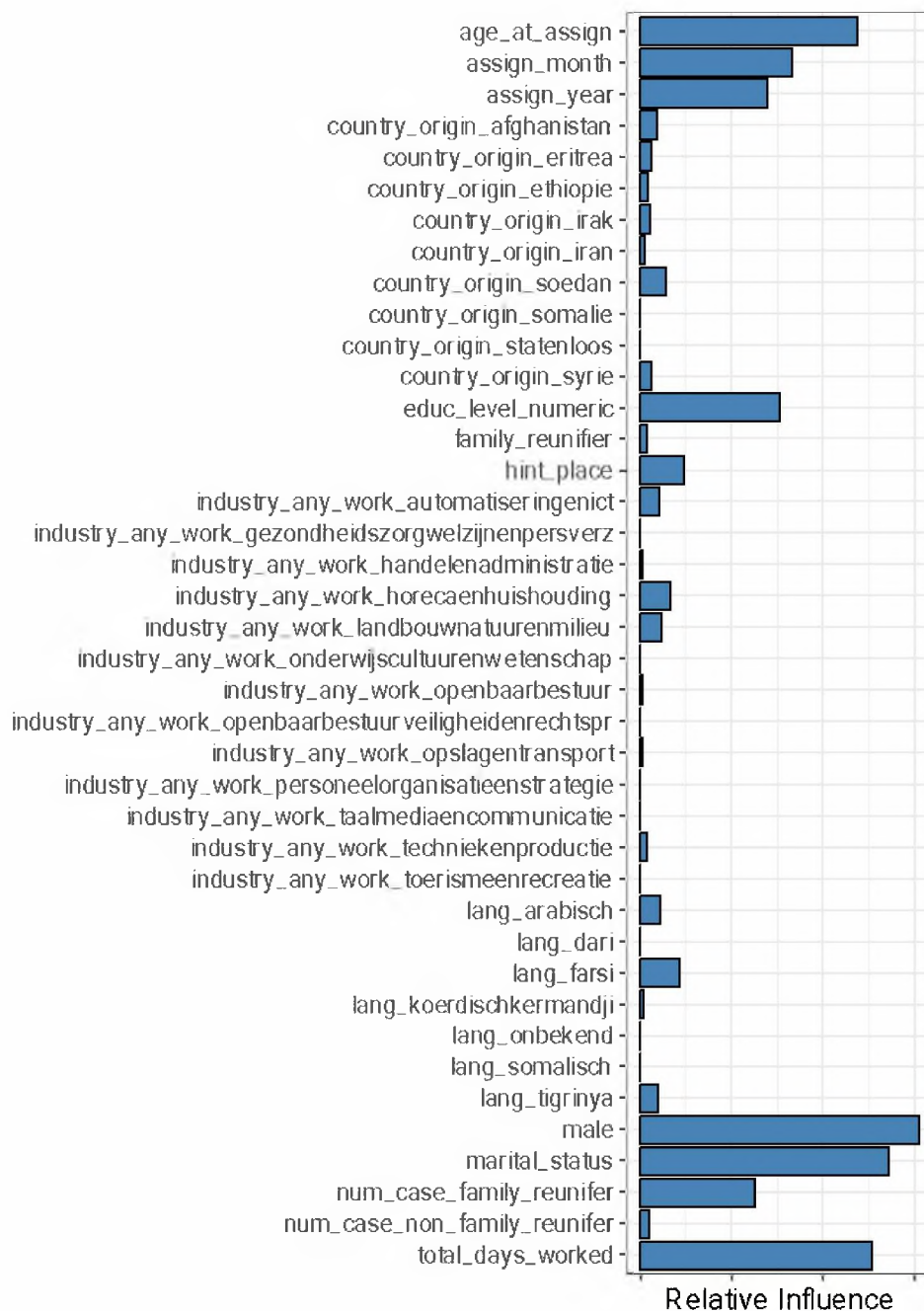
**Flevoland**



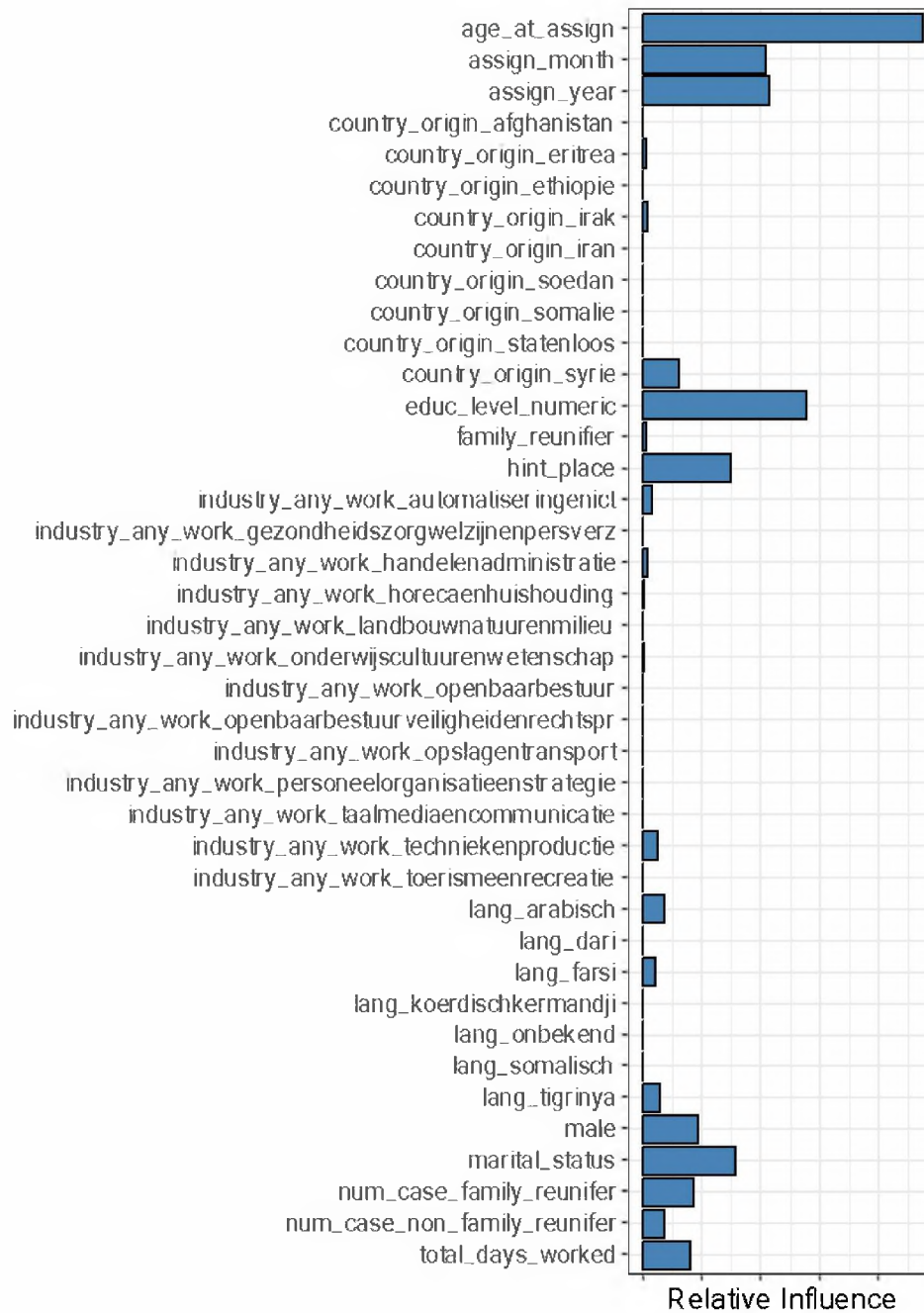
**Food Valley**



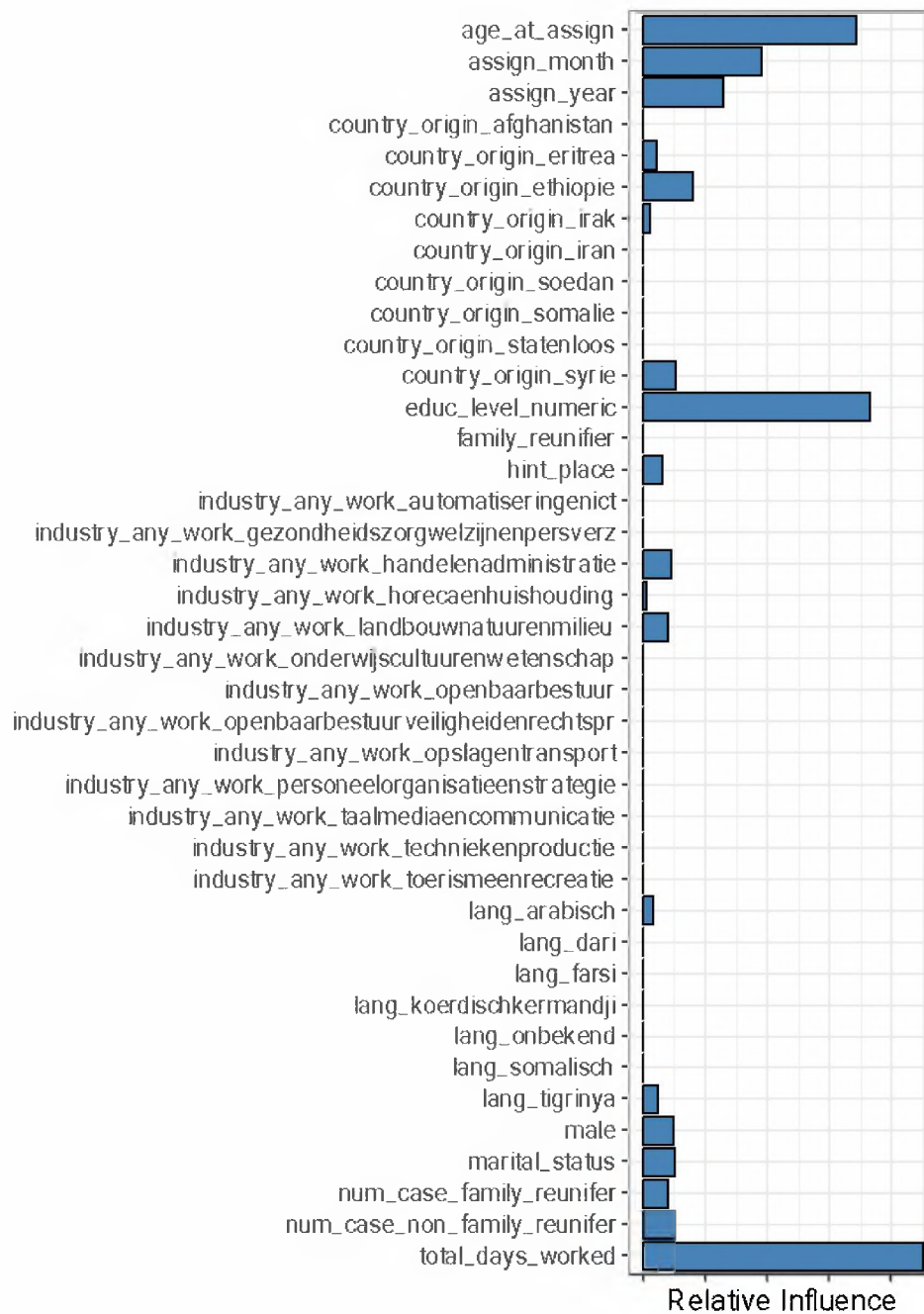
**Friesland**



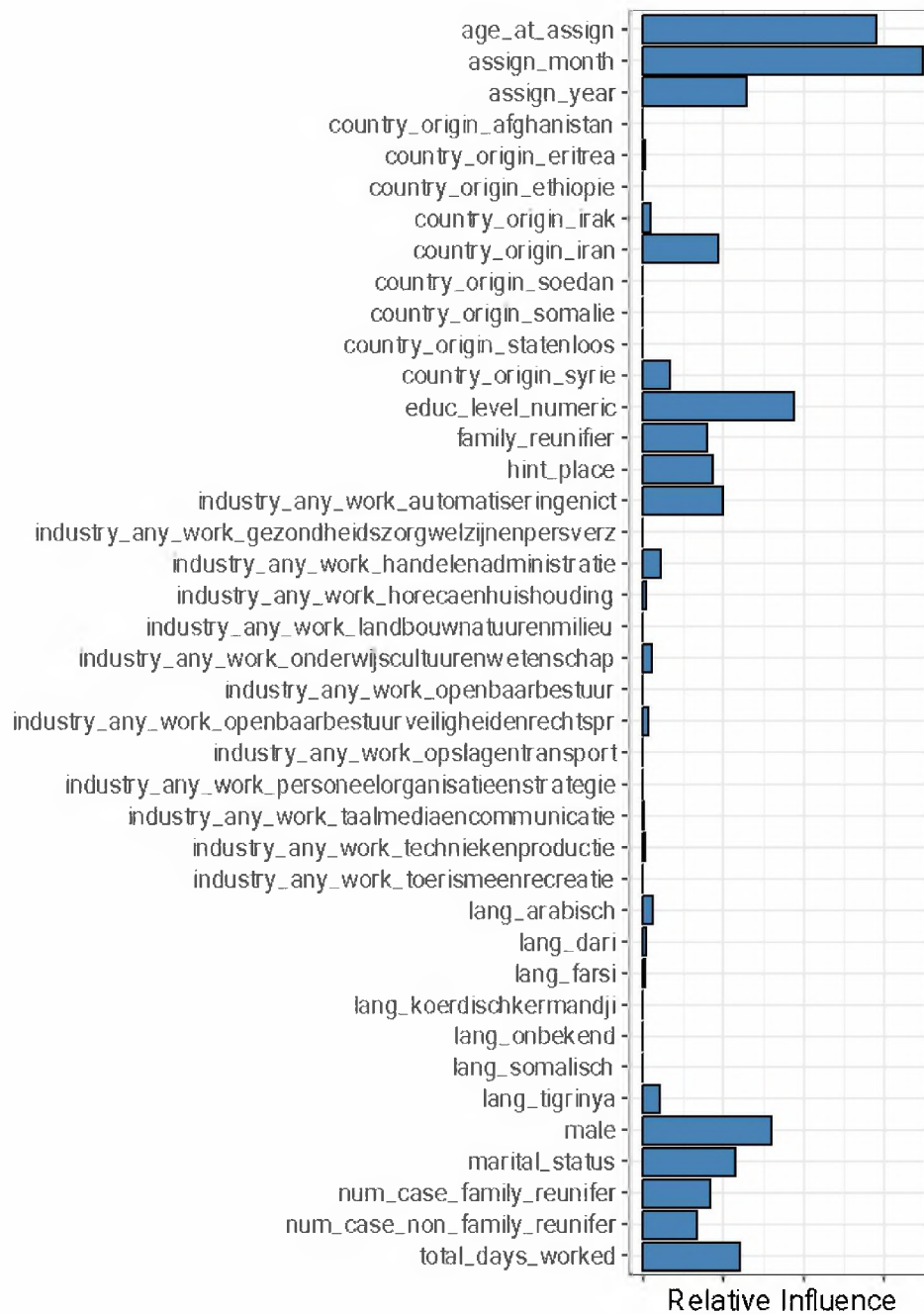
**Gooi- en Vechtstreek**



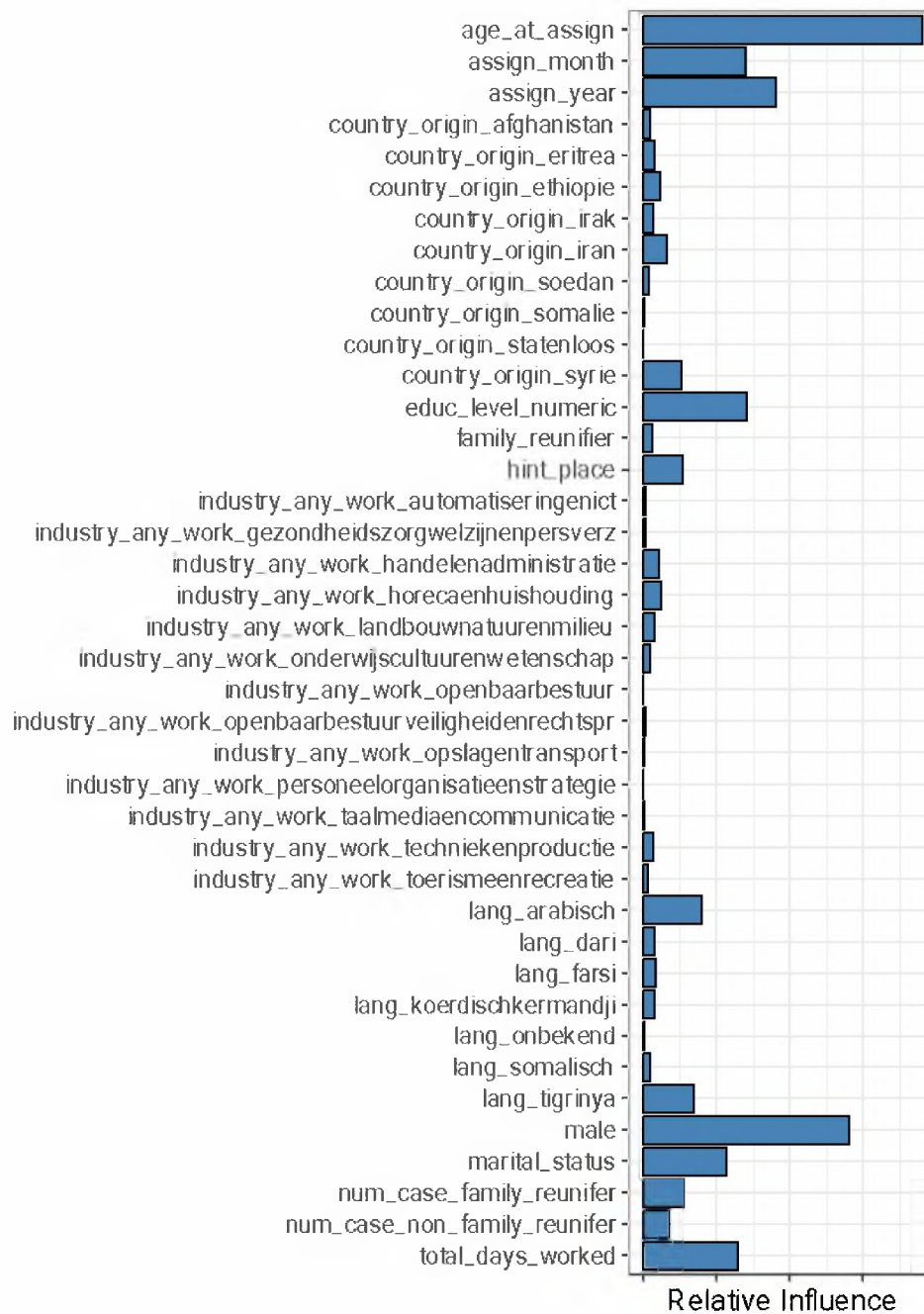
**Gorinchem**



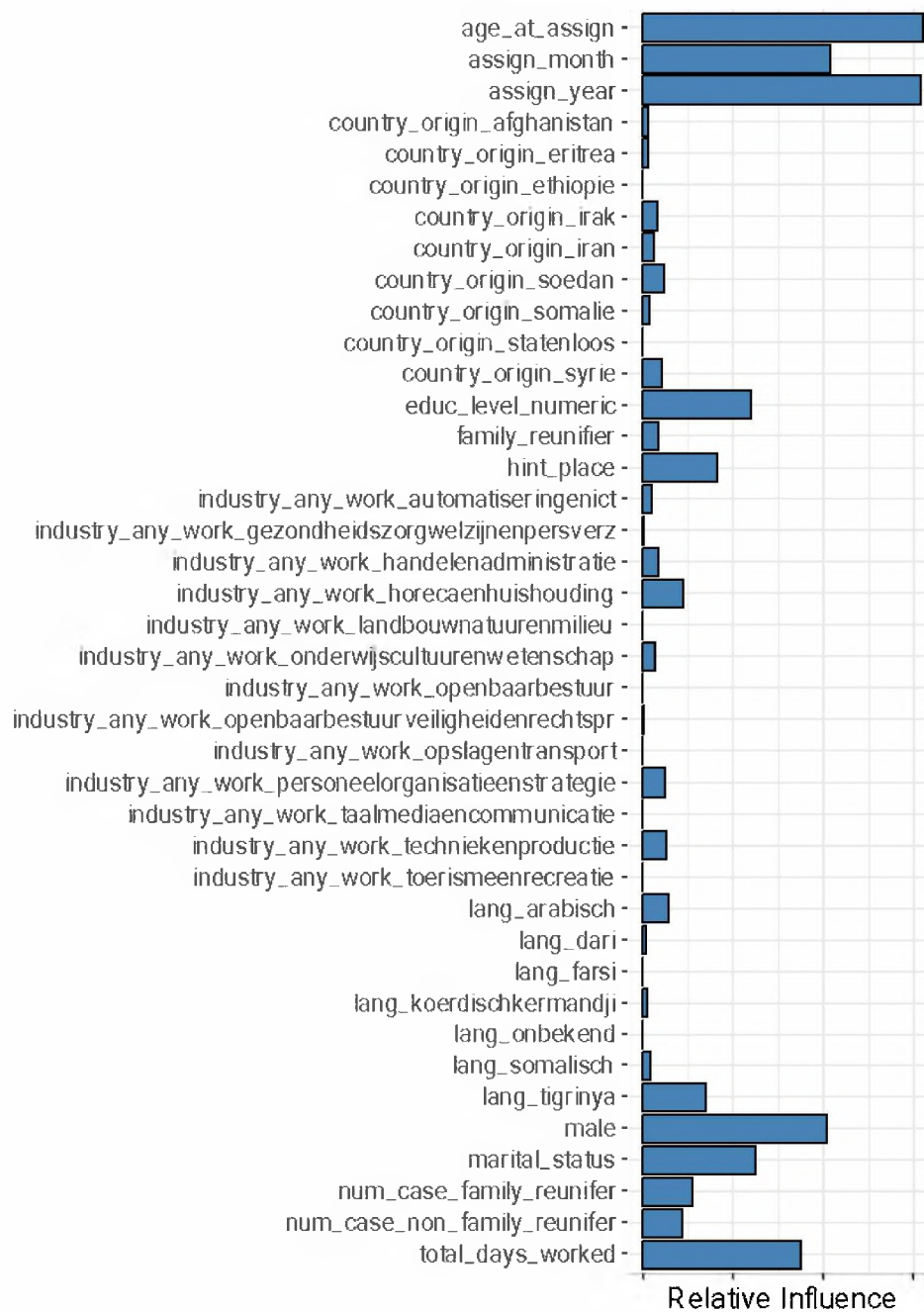
**Groningen**



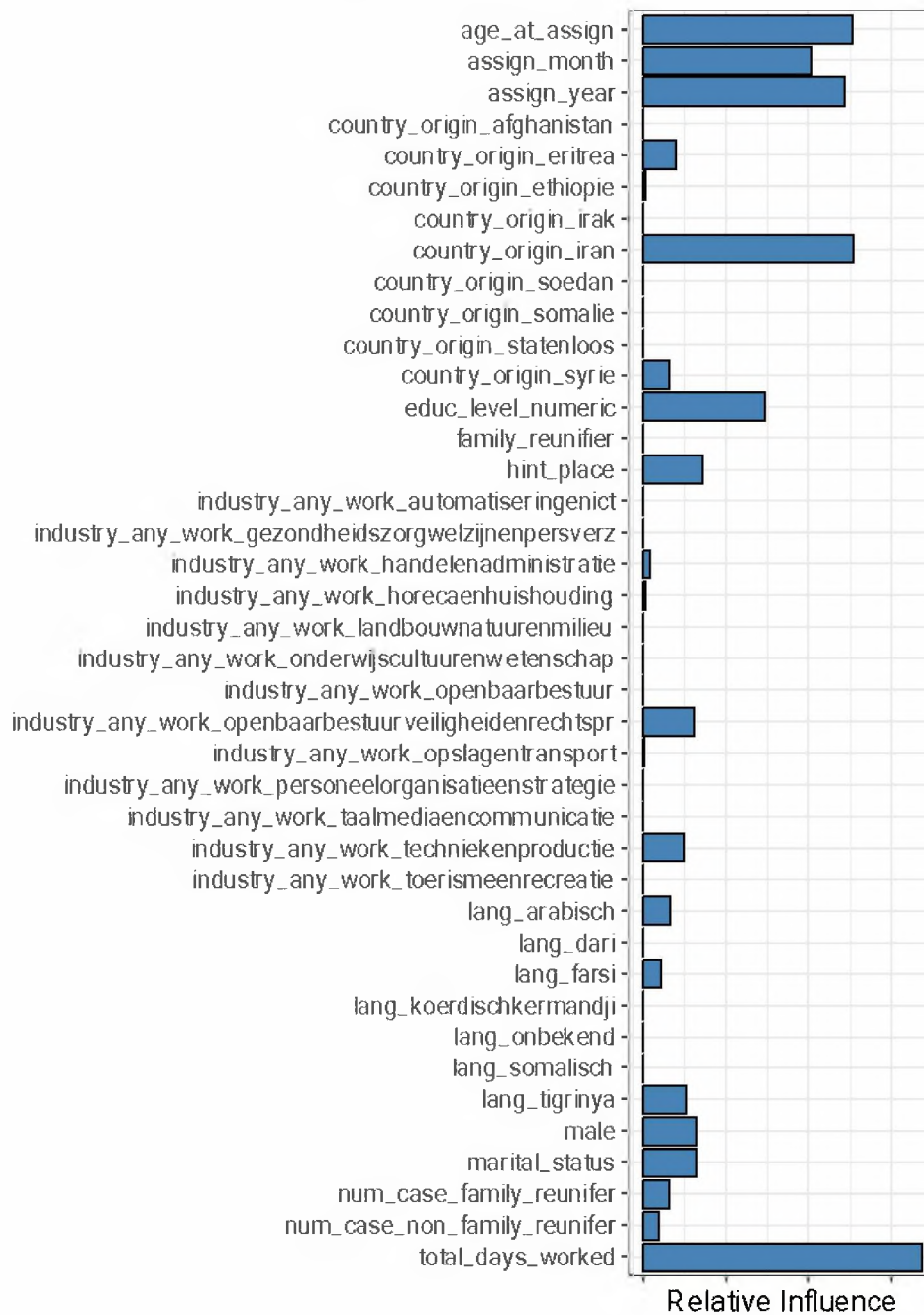
**Groot Amsterdam**



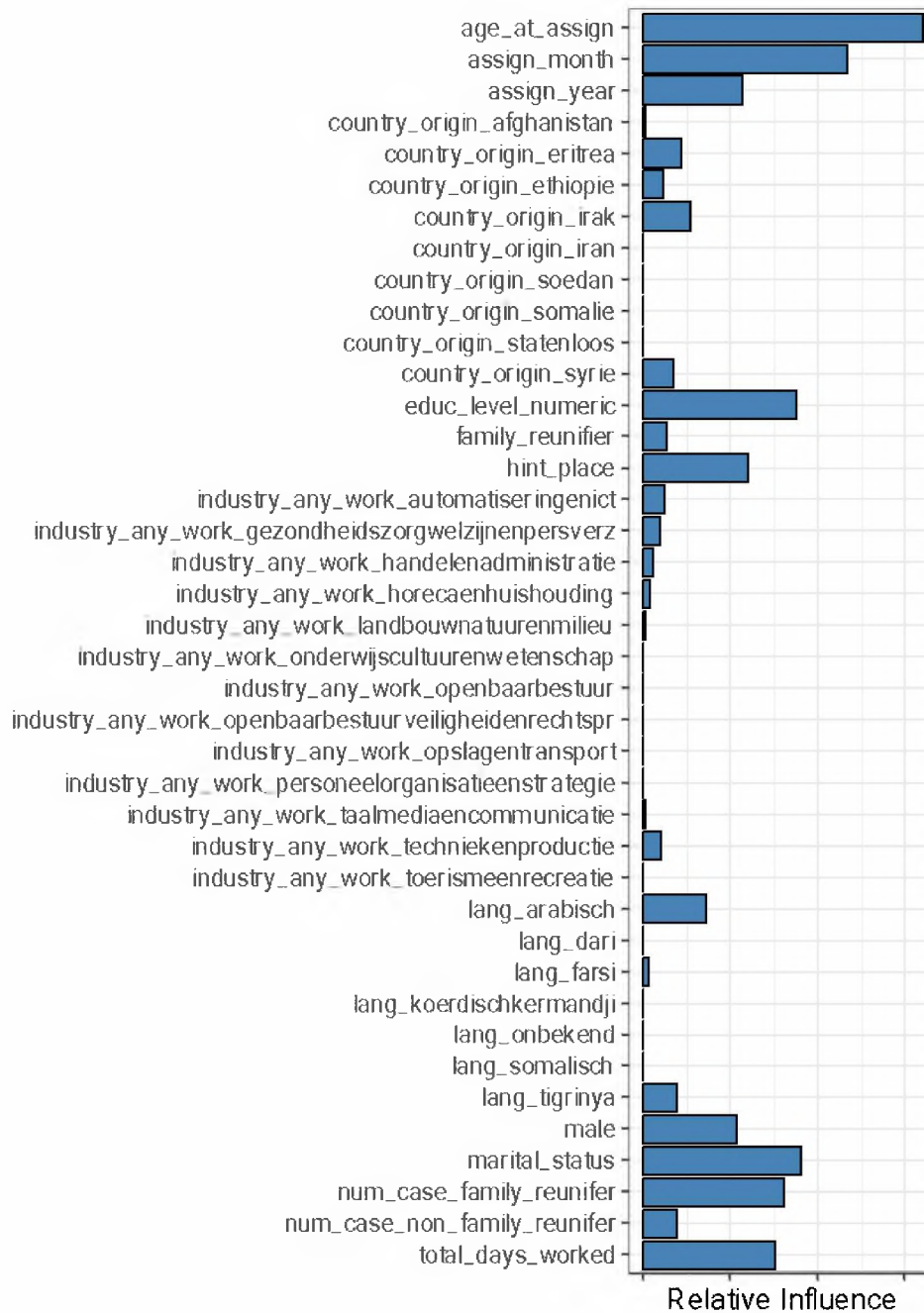
**Haaglanden**



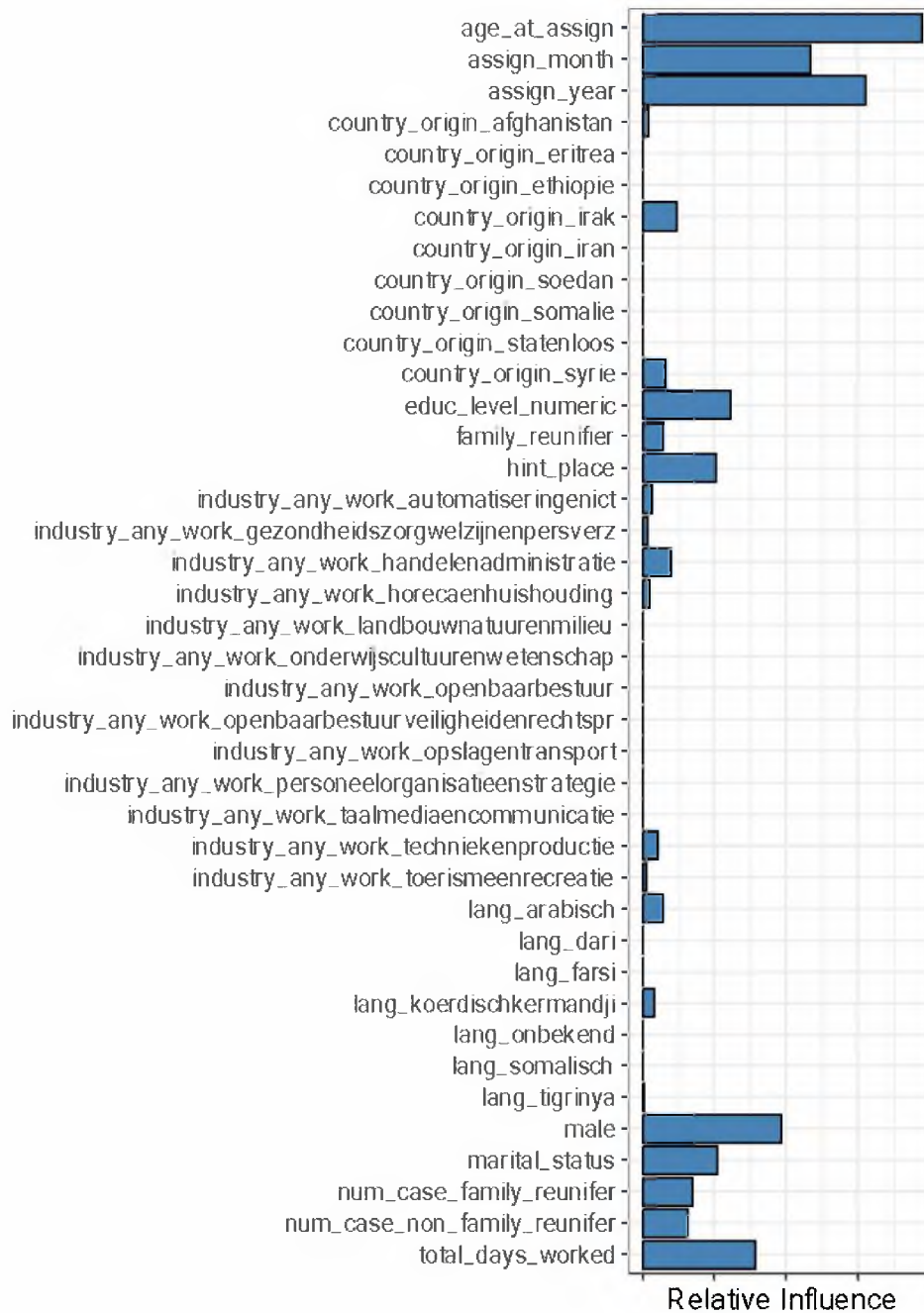
**Helmond-De Peel**



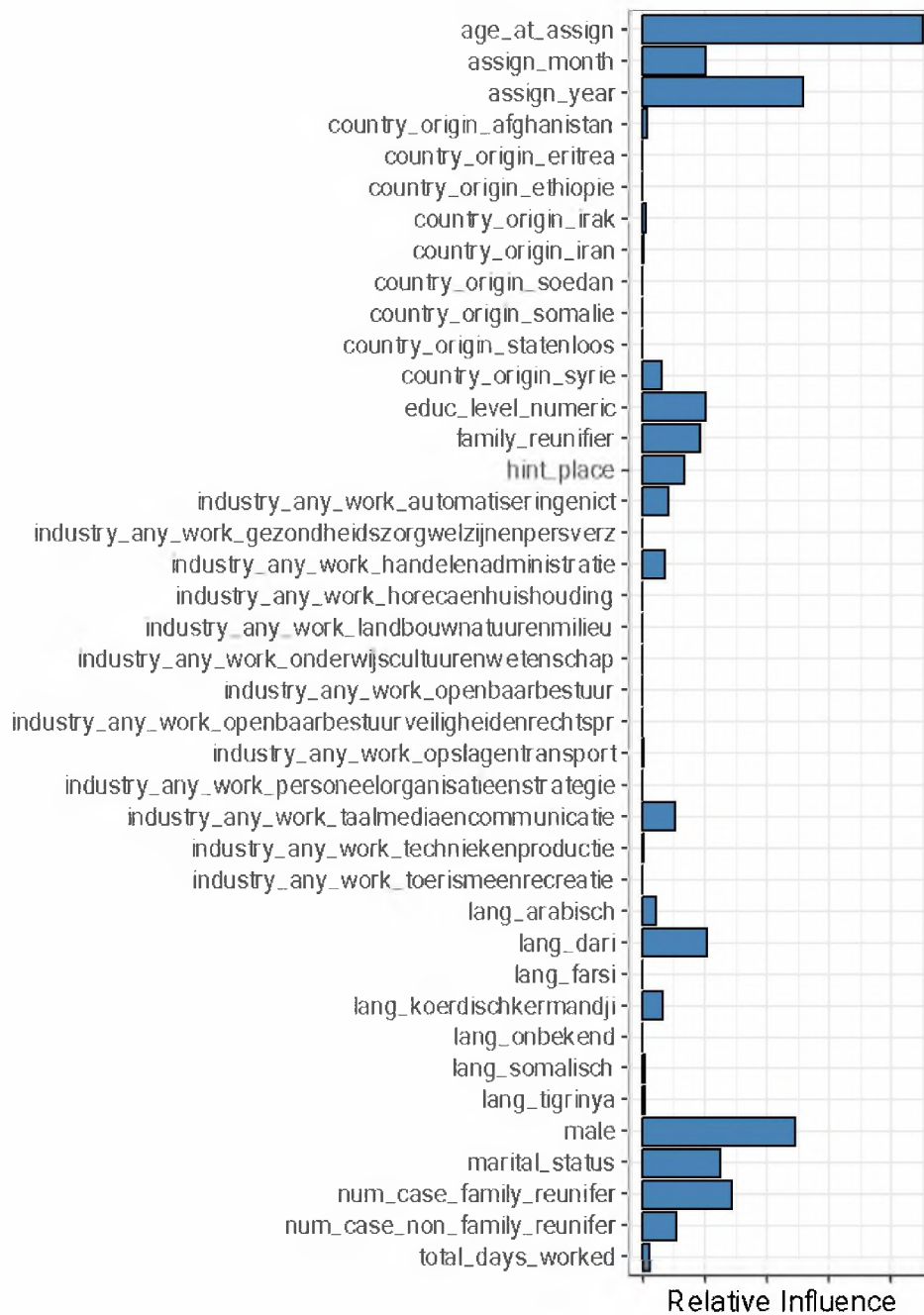
**Holland Rijnland**



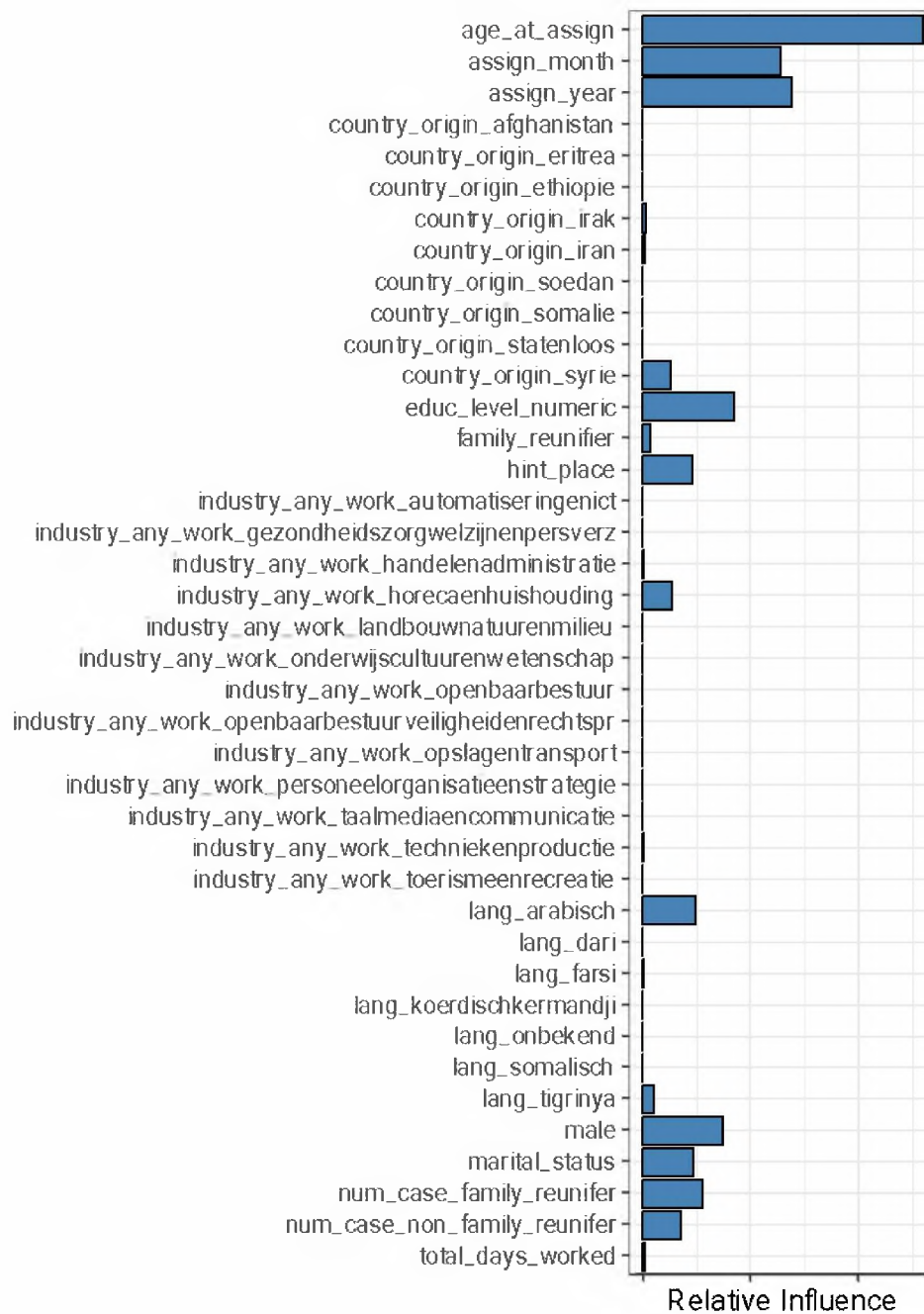
**Midden-Brabant**



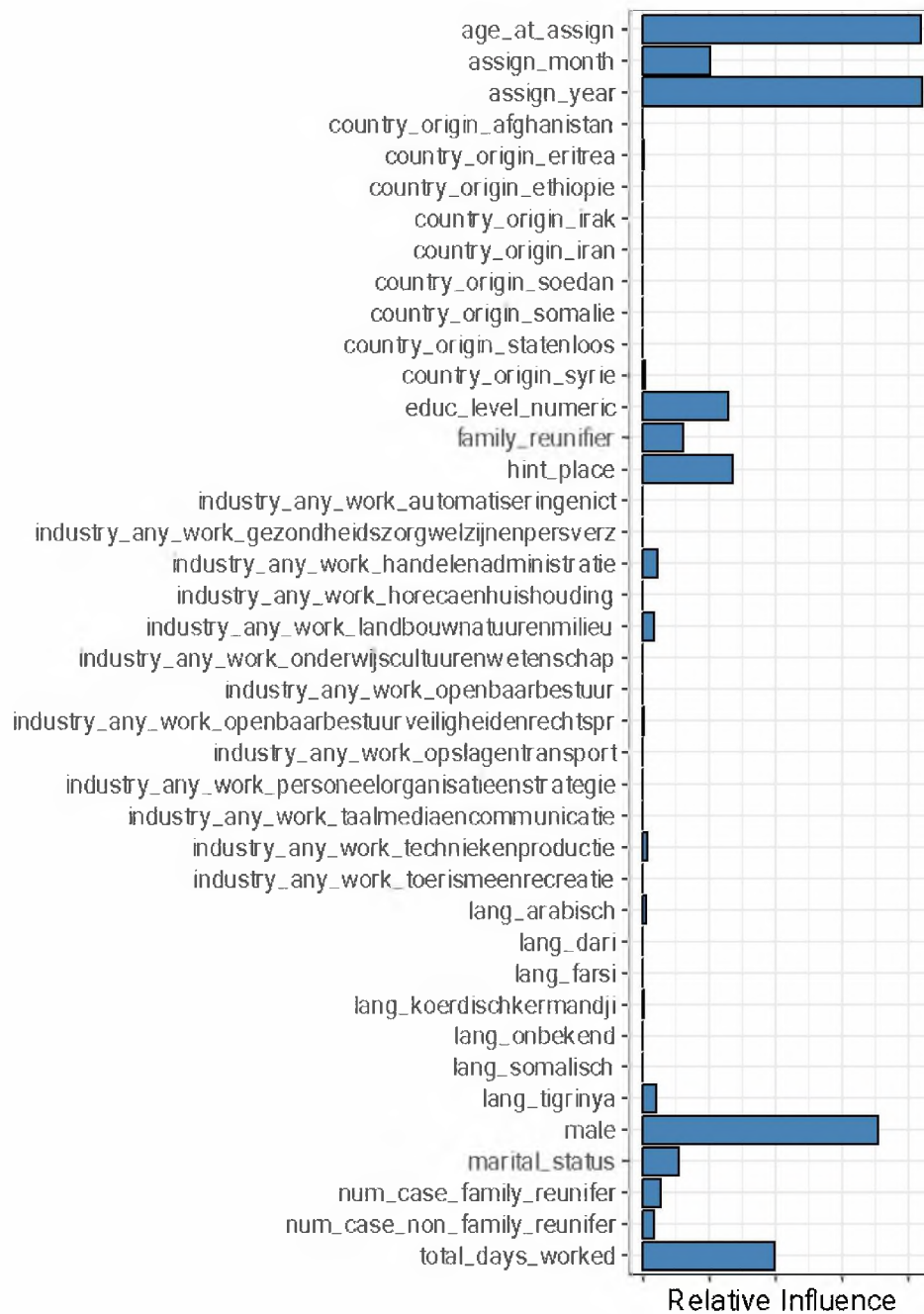
**Midden-Gelderland**



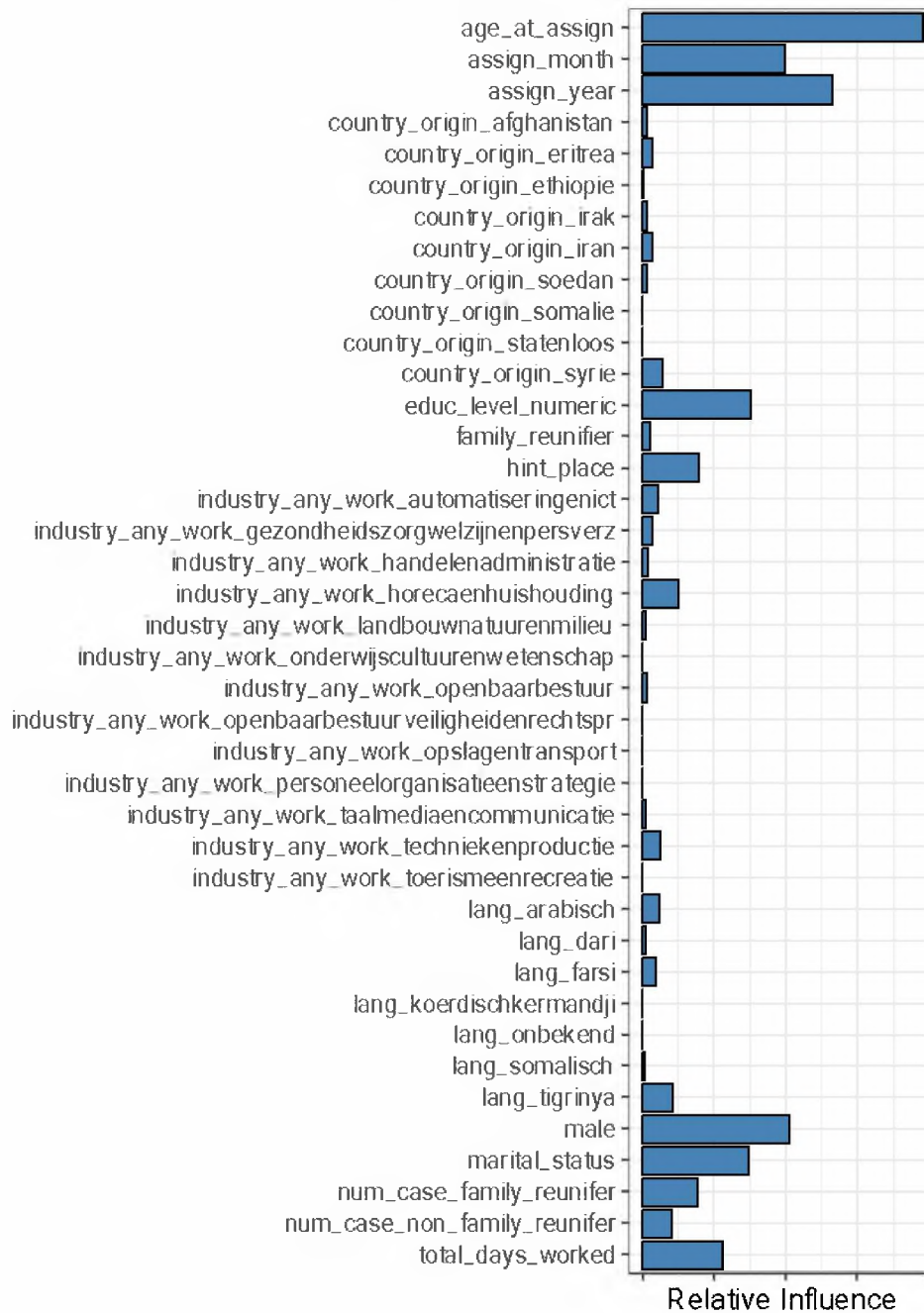
**Midden-Holland**



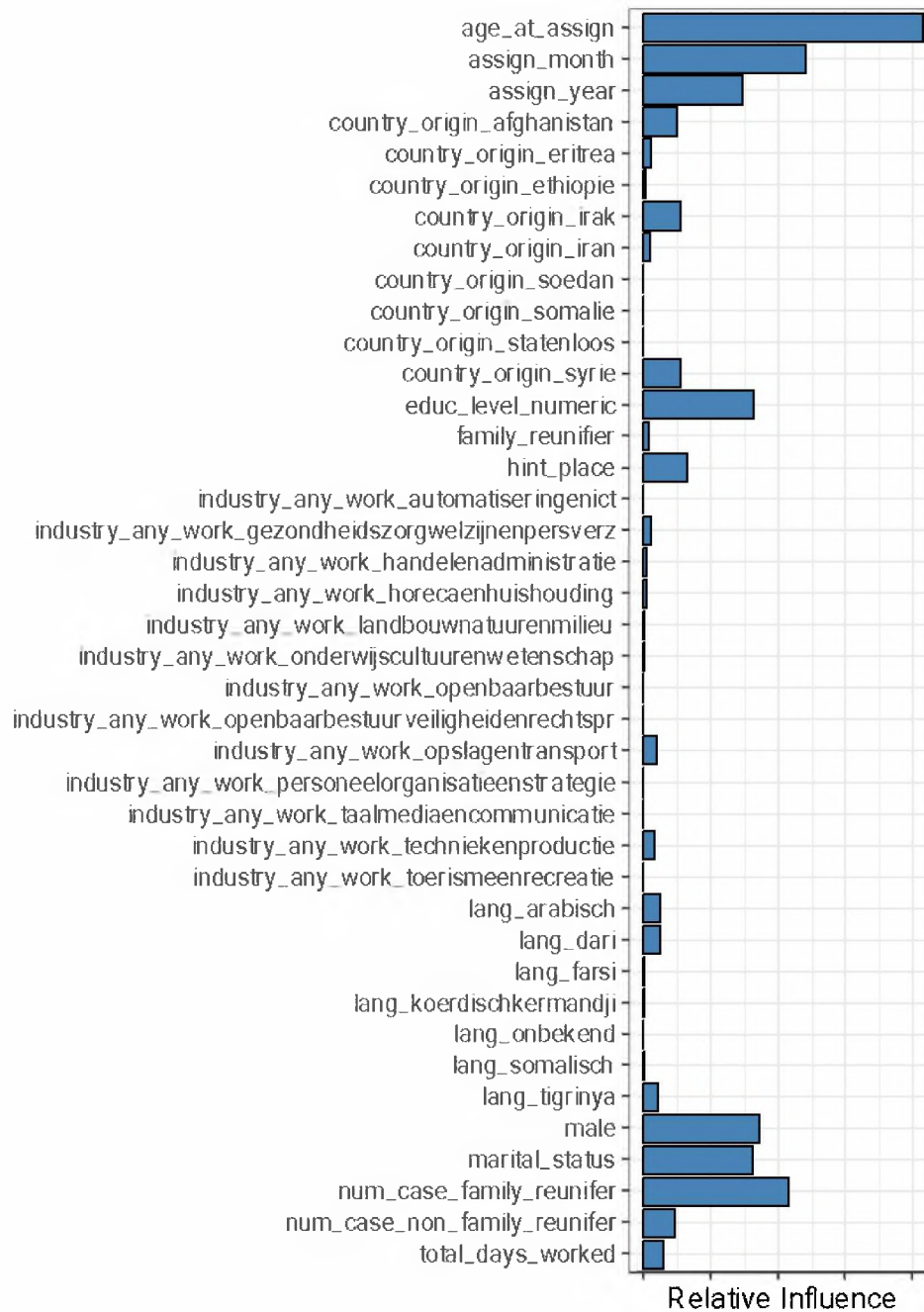
**Midden-Limburg**



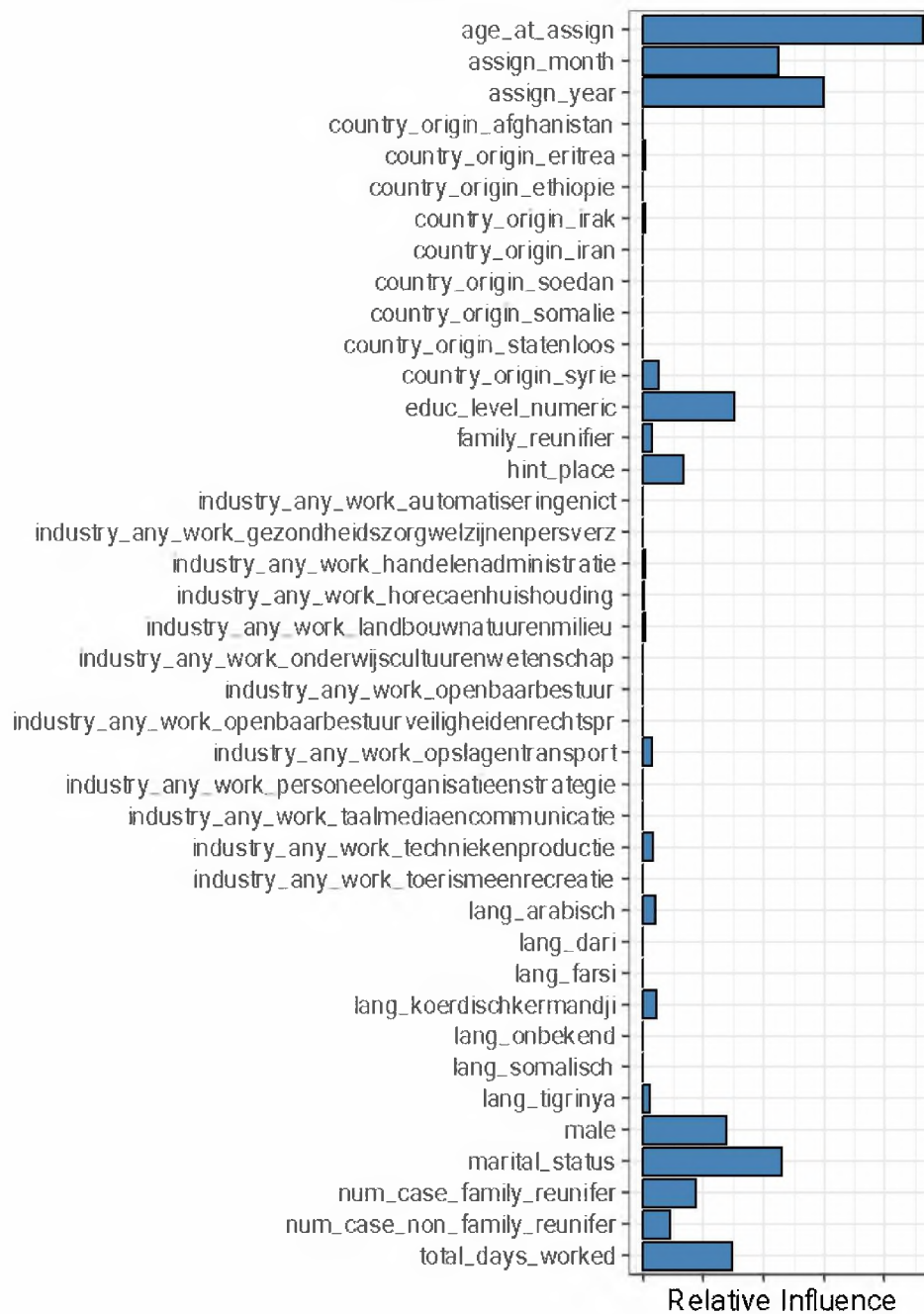
**Midden-Utrecht**



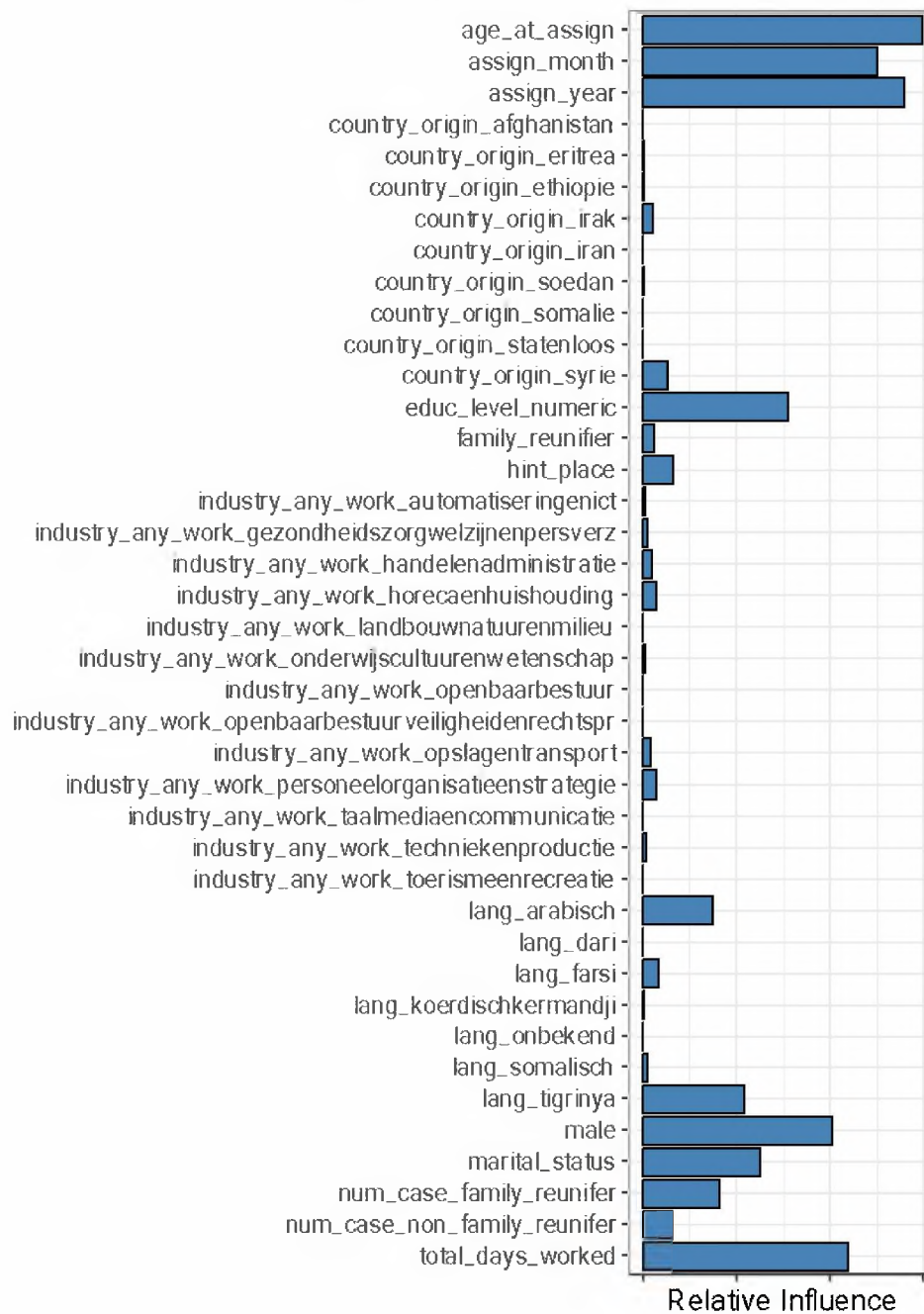
**Noord-Holland (Noord)**



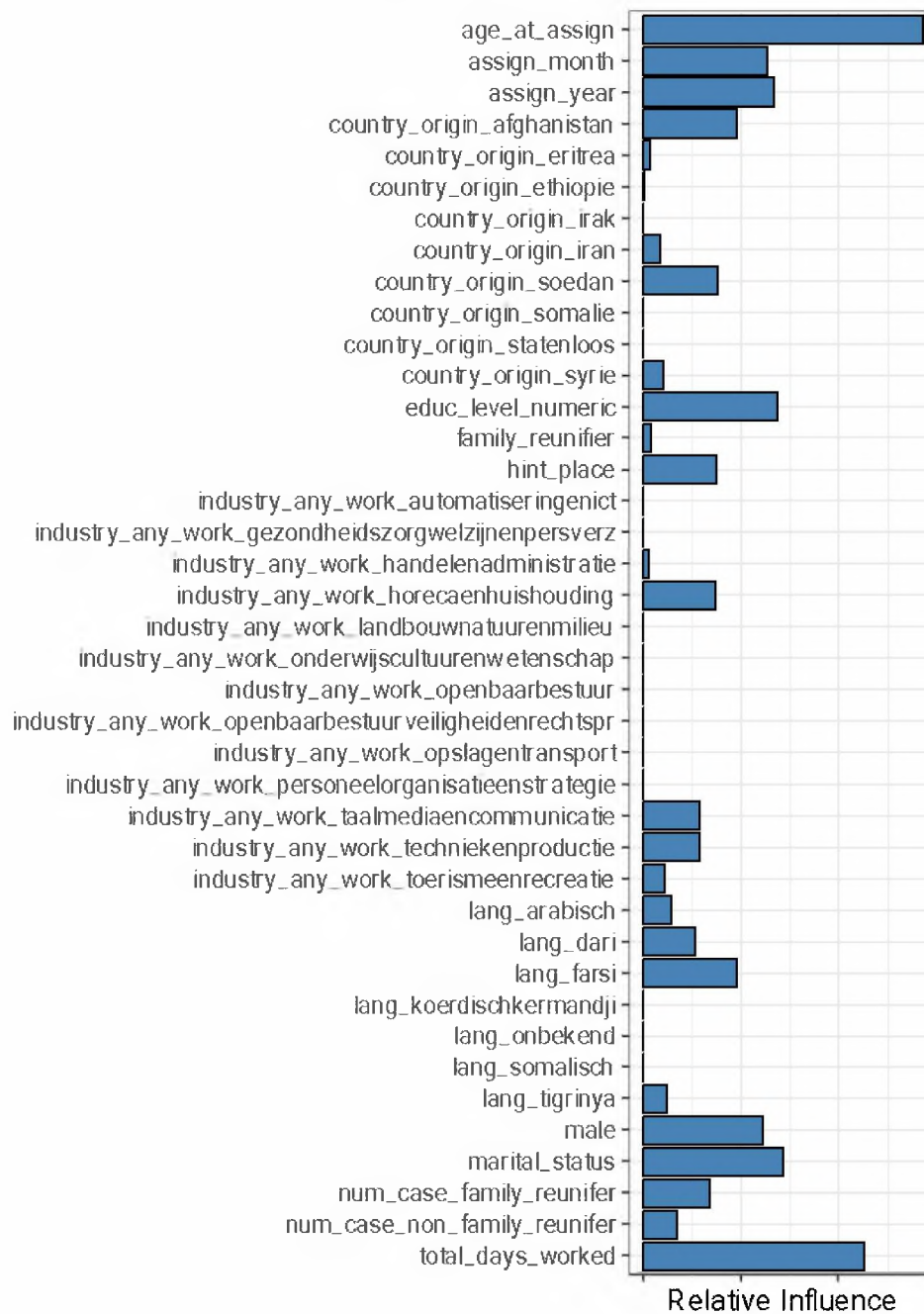
**Noord-Limburg**



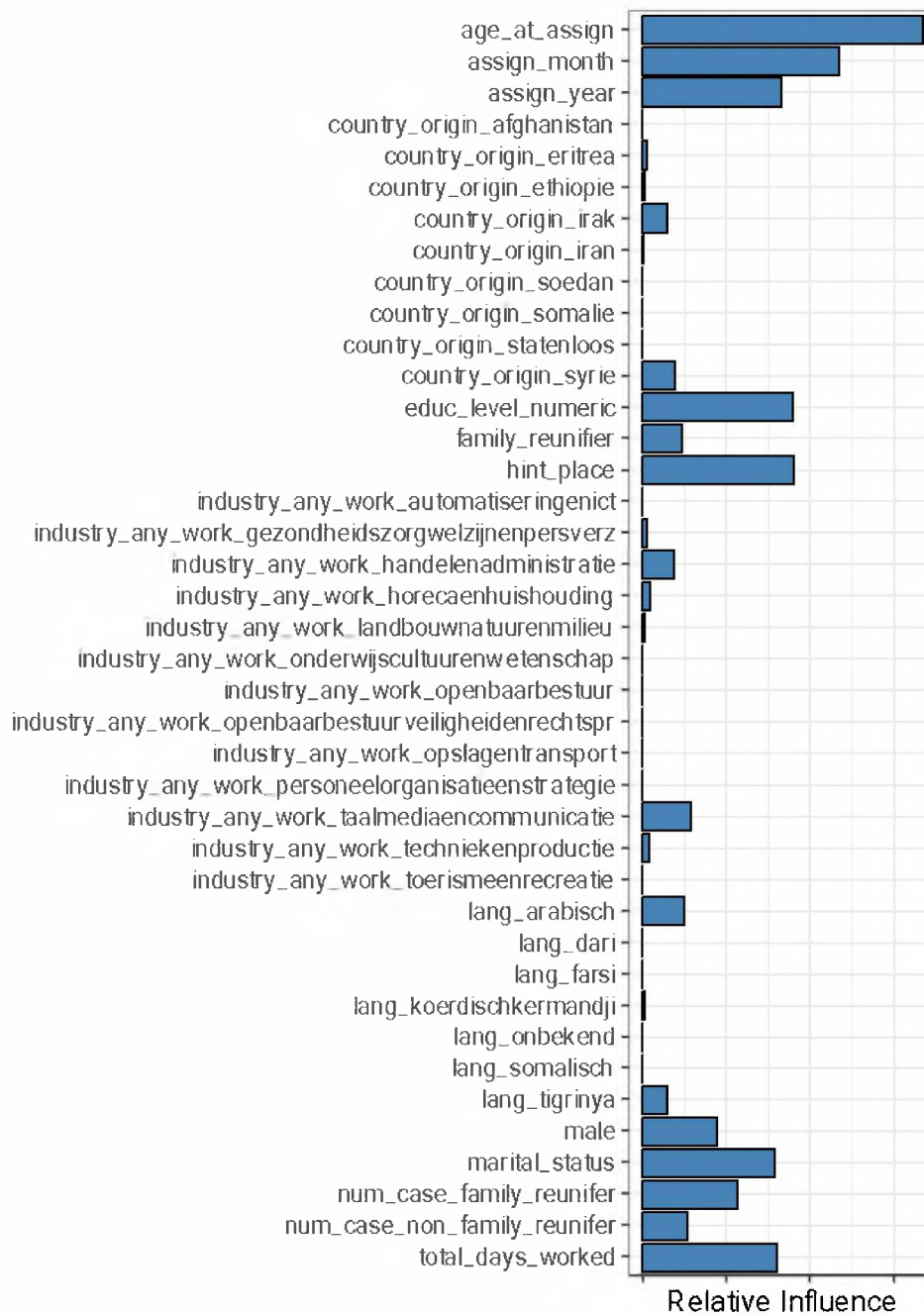
**Noordoost-Brabant**



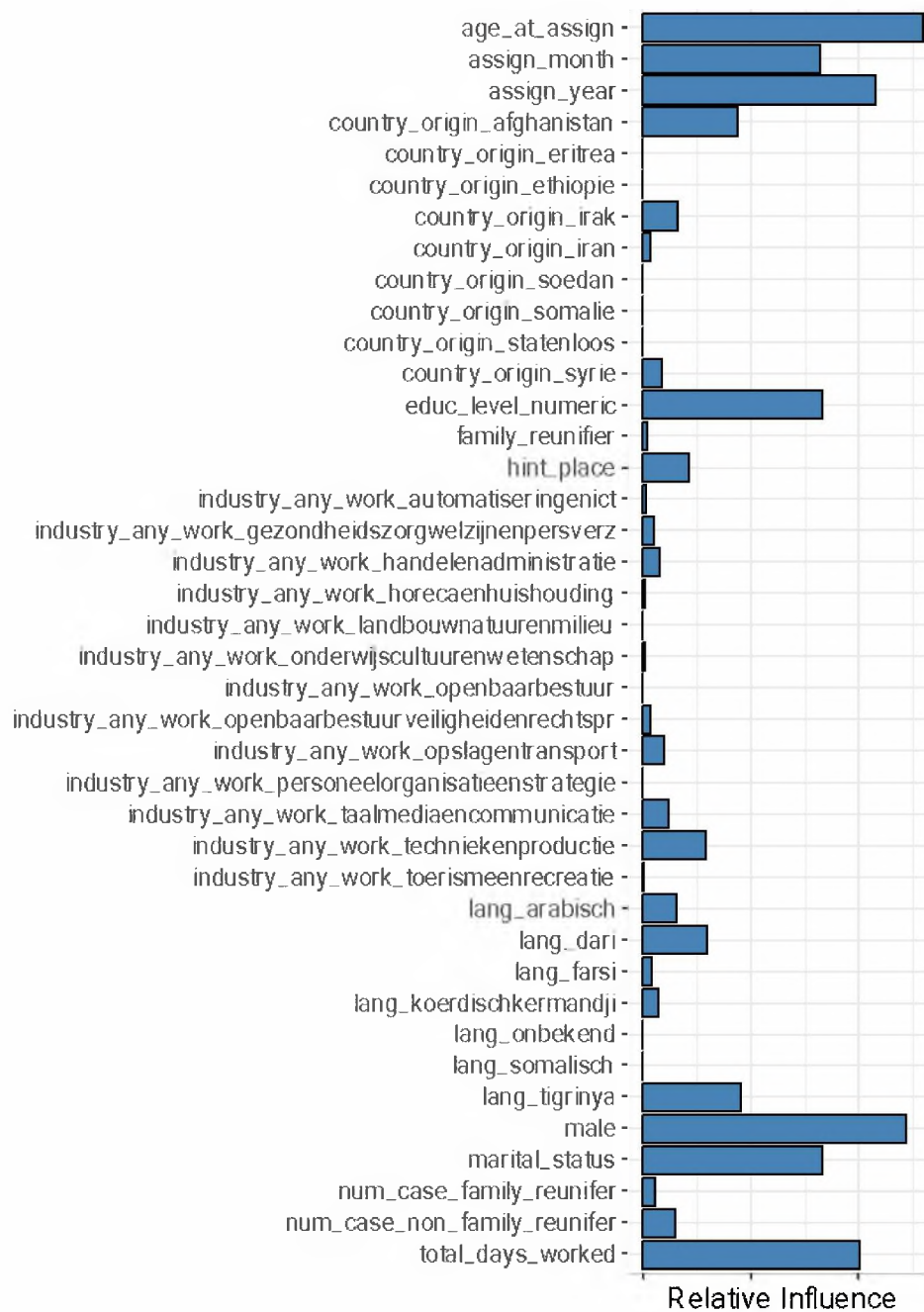
**Regio Zwolle**



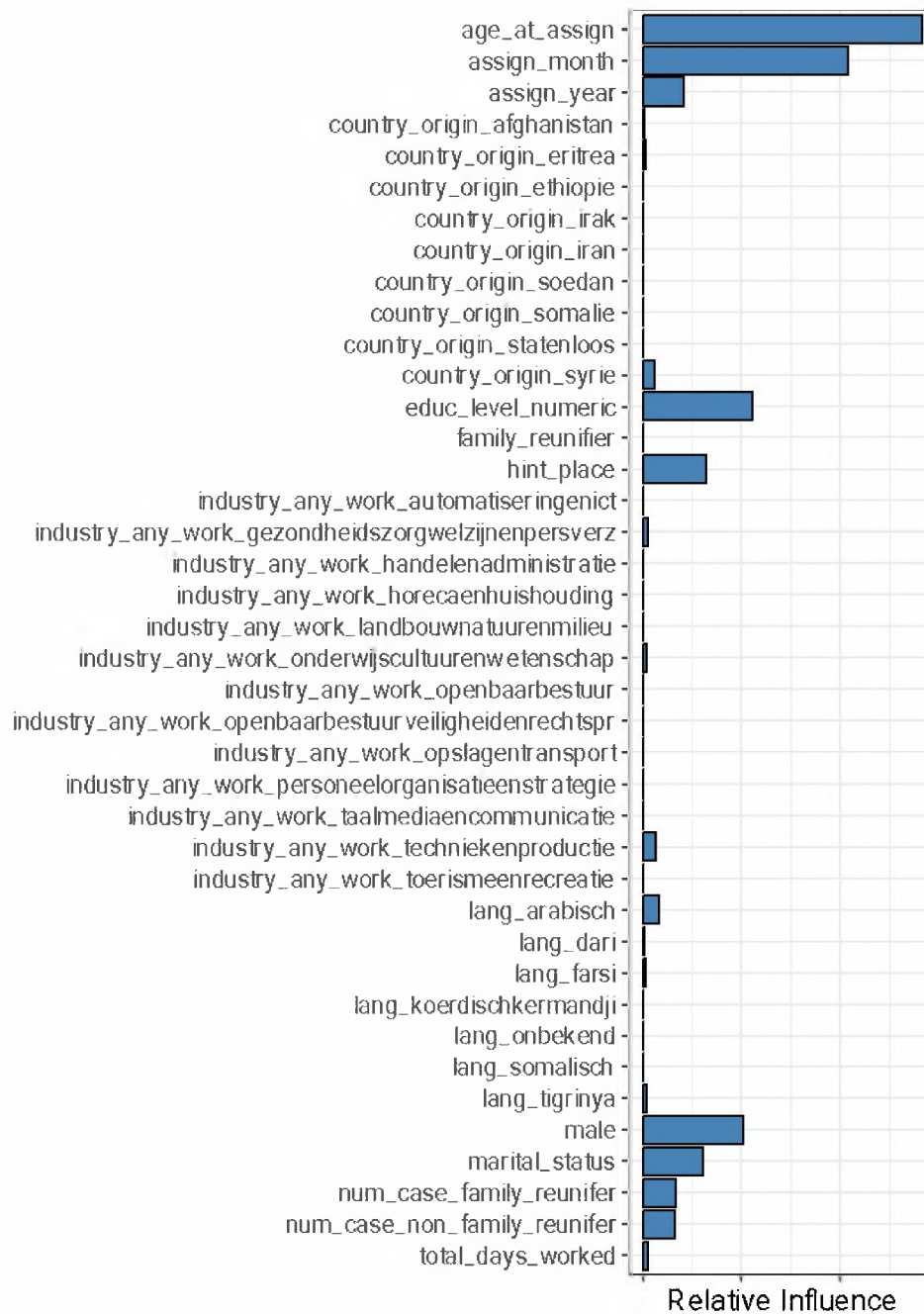
**Rijk van Nijmegen**



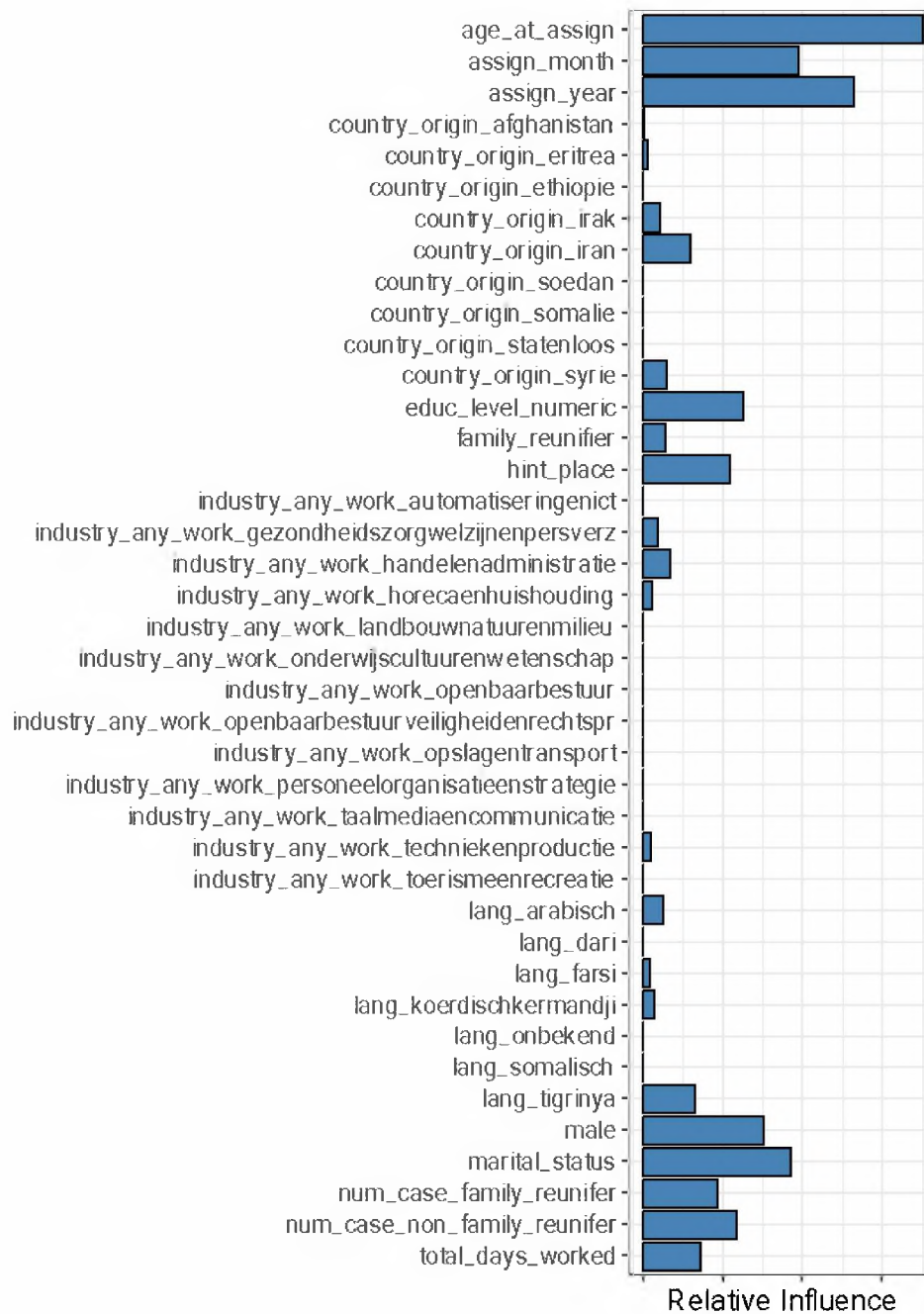
**Rijnmond**



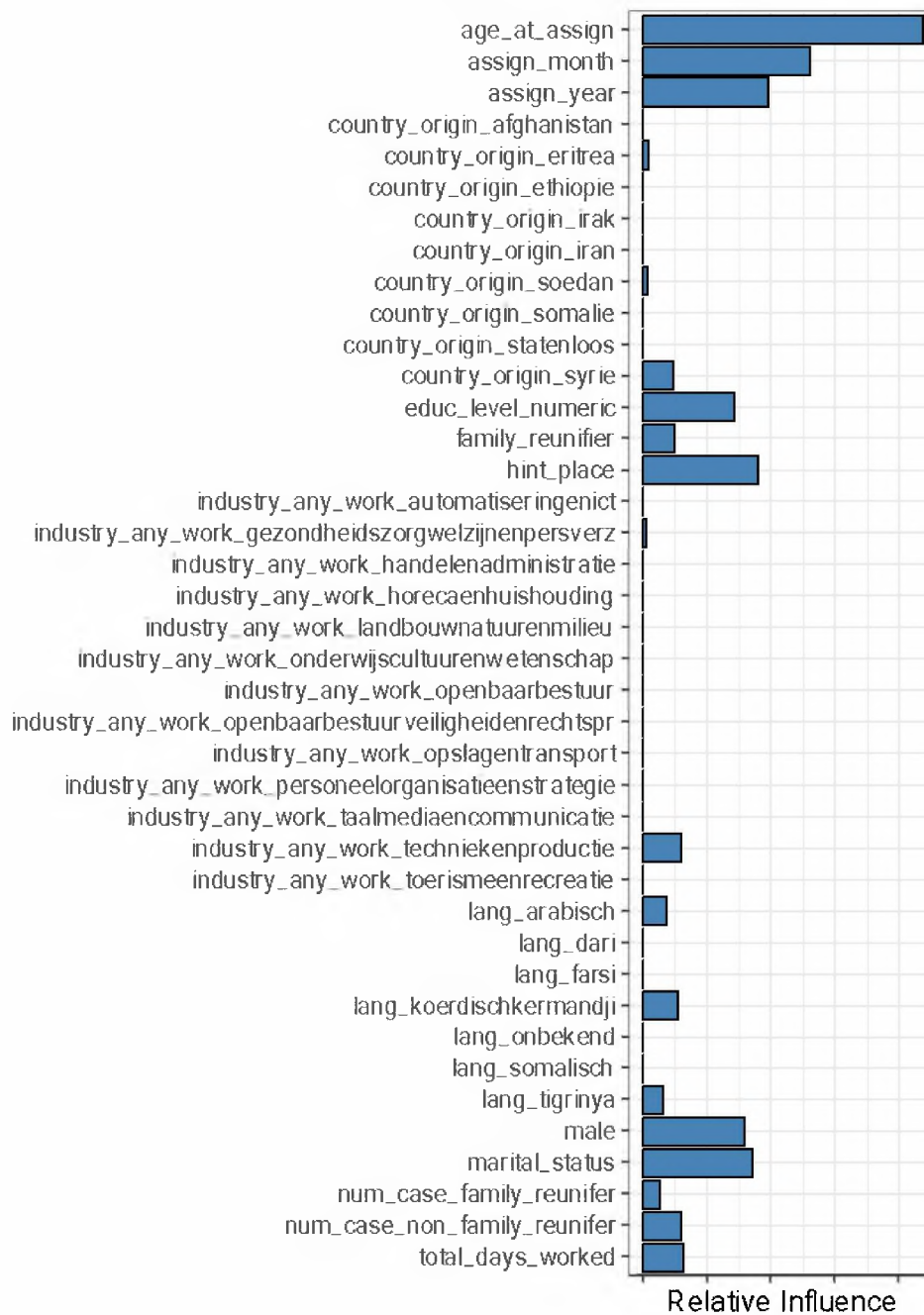
**Rivierenland**



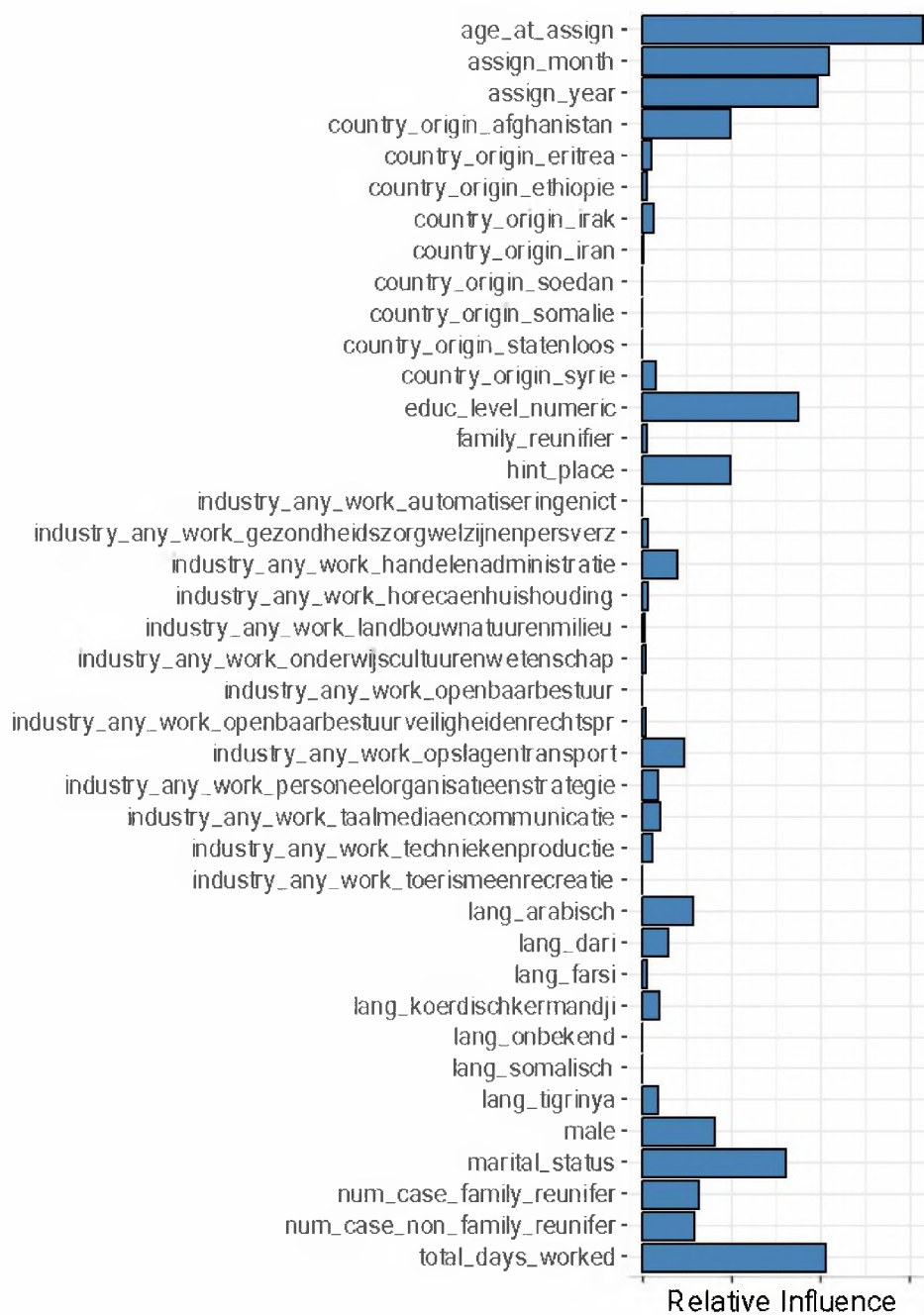
**Stedendriehoek en Noordwest Veluwe**



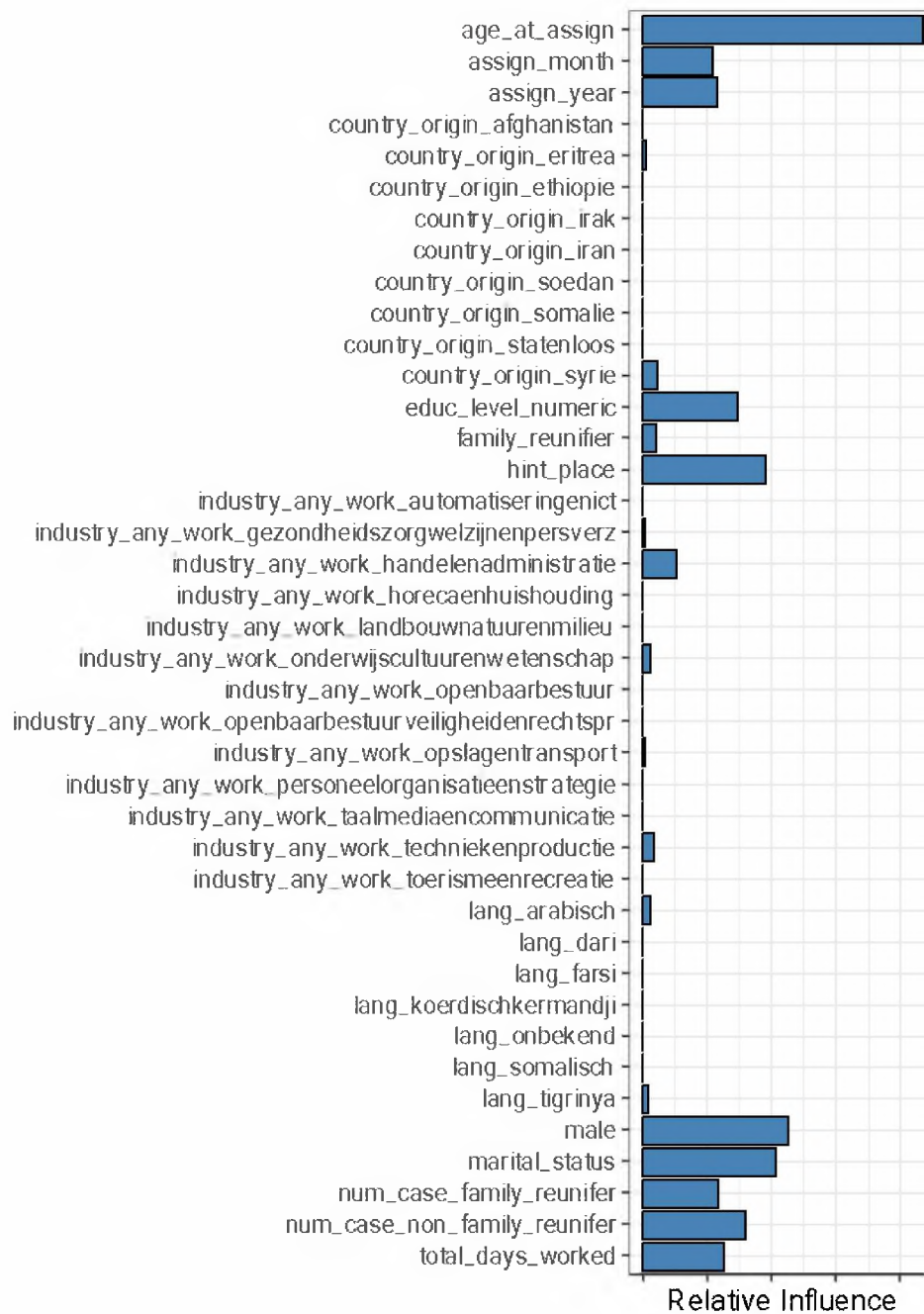
**Twente**



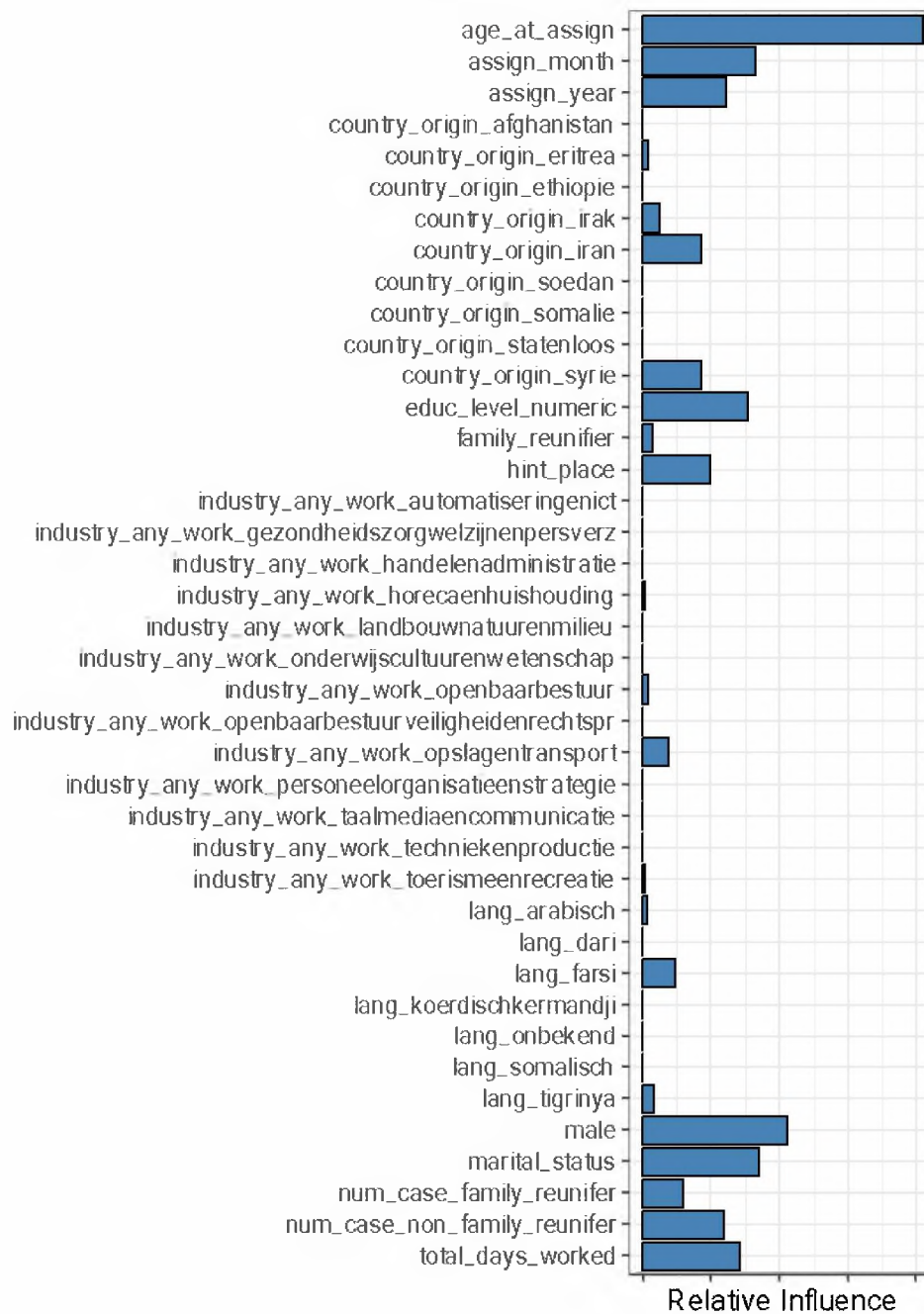
**West-Brabant**



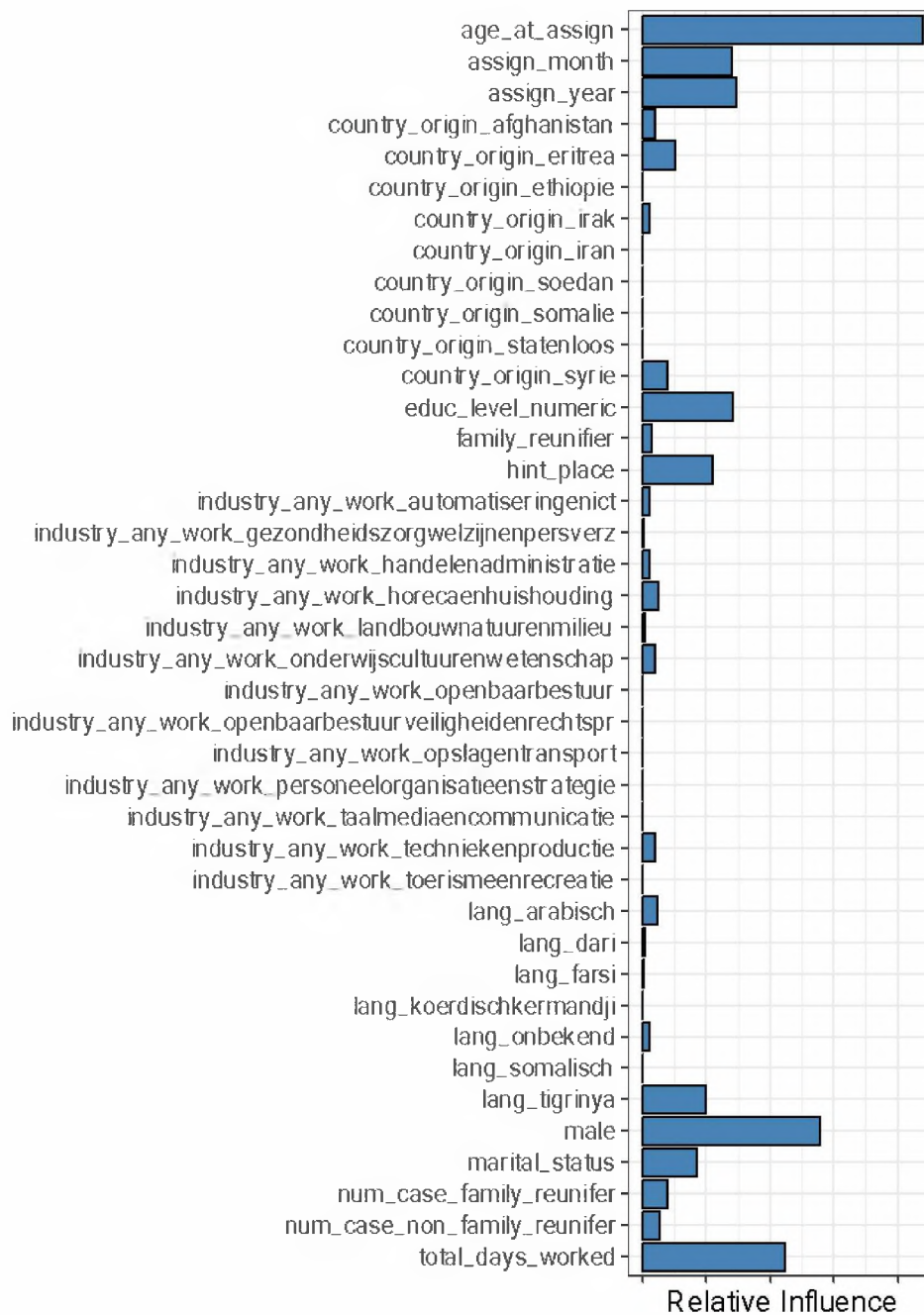
**Zaanstreek – Waterland**



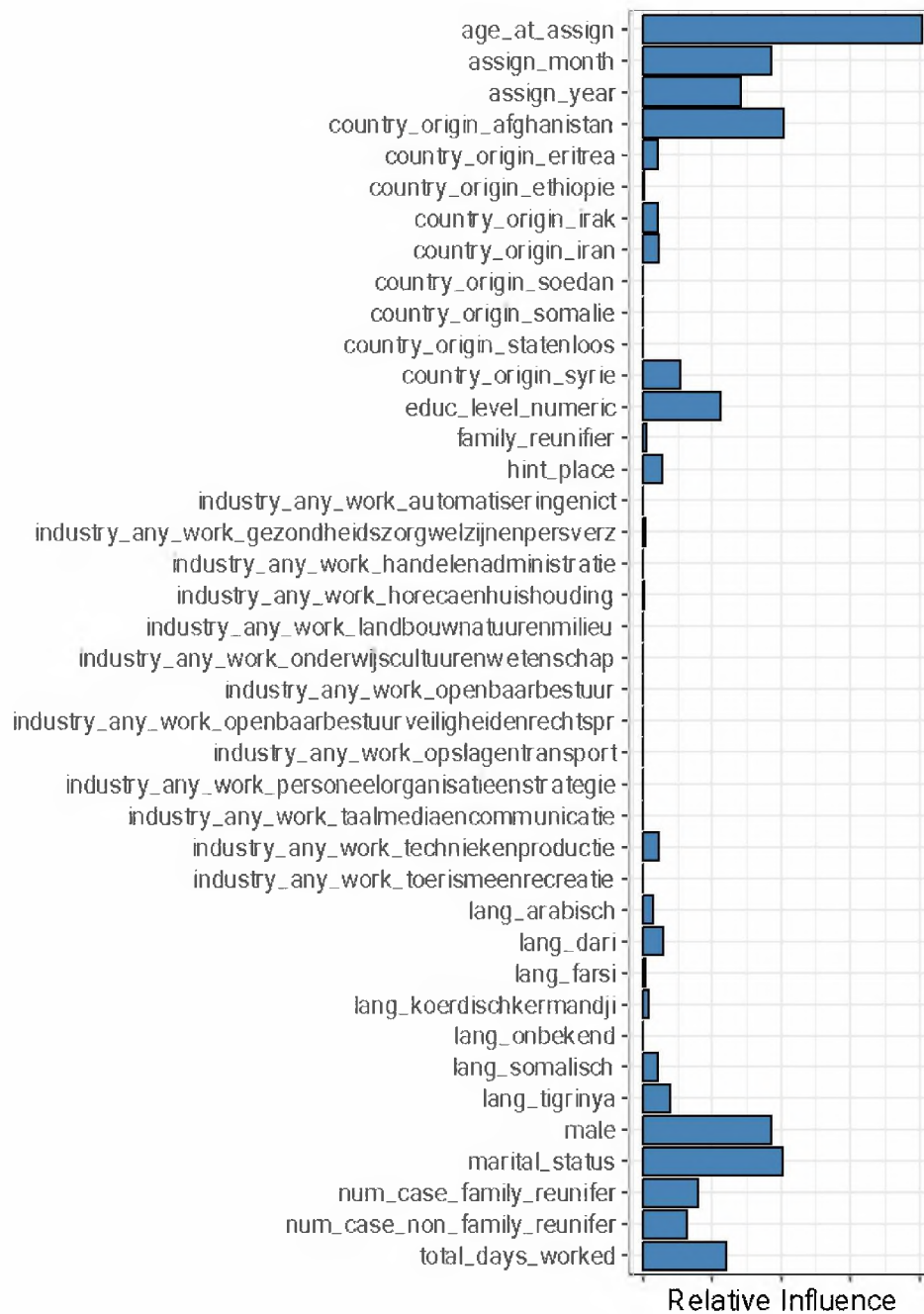
**Zeeland**



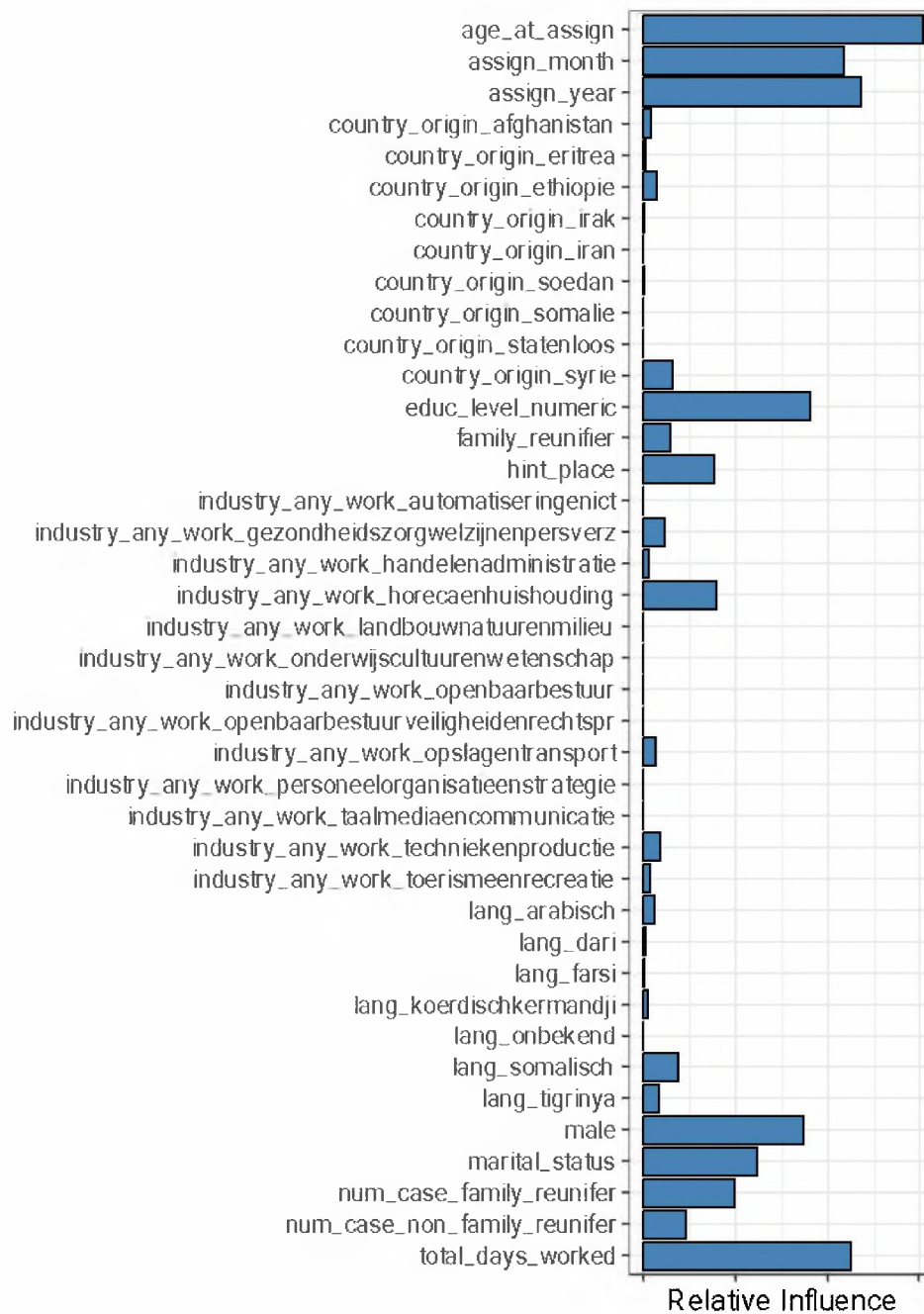
**Zuid-Holland Centraal**



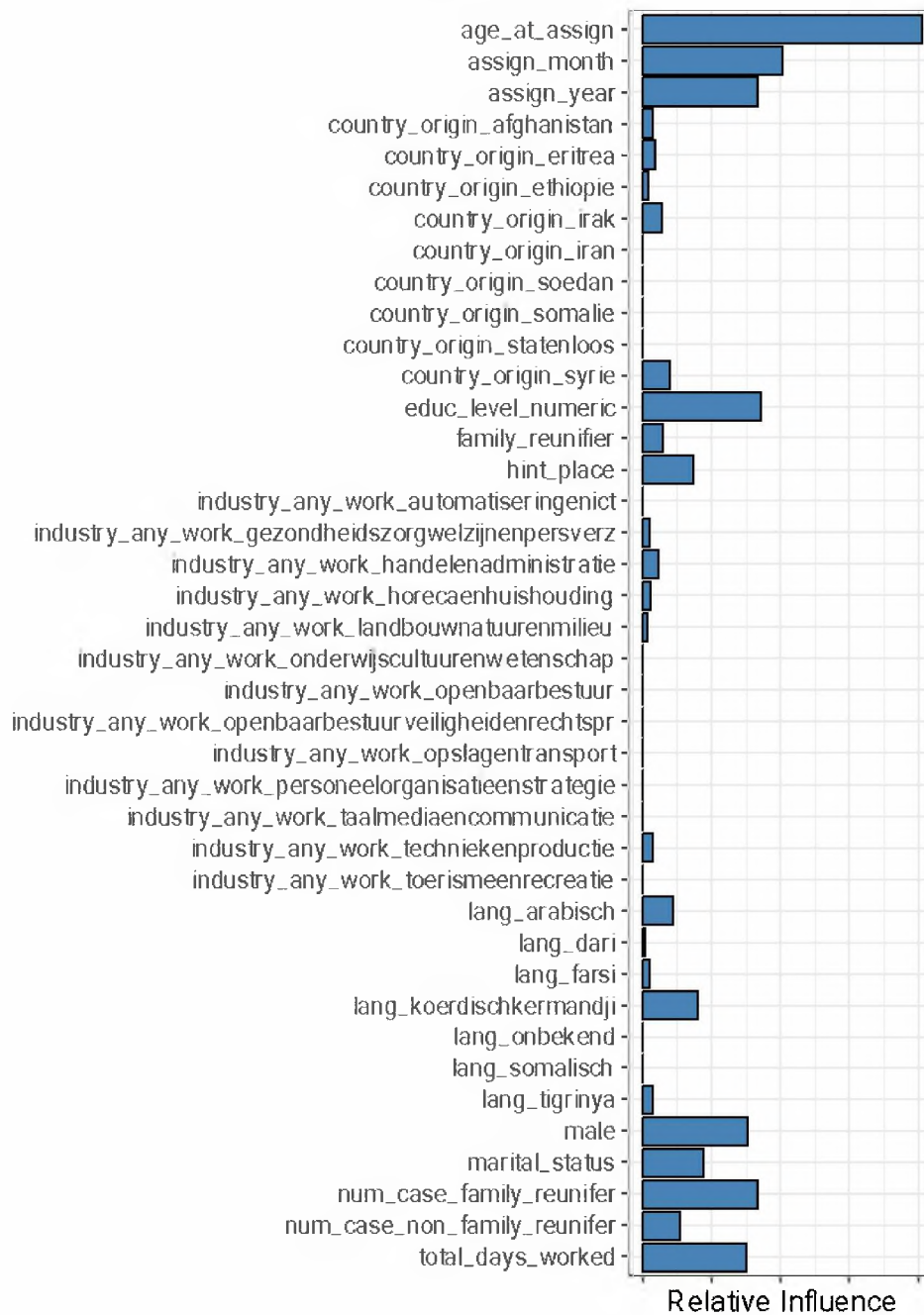
**Zuid-Kennemerland en IJmond**



**Zuid-Limburg**



**Zuidoost-Brabant**



## 4. Appendix 4: Model Counts for Backtest

The backtest was conducted using a training dataset and test dataset. The training dataset included 55,090 total individuals. For the backtest, a model was trained per LMR (35 models in total). The test data included 735 individuals who were assigned to municipalities between June 2019 and August 2019. The tables below show the count of individuals who comprise the training and test datasets from each LMR.

***Count of individuals included in the backtest training data by LMR***

| <b>Labor Market Region</b> | <b>Number included in backtest training data</b> |
|----------------------------|--------------------------------------------------|
| Achterhoek                 | 989                                              |
| Amersfoort                 | 988                                              |
| Drechtsteden               | 897                                              |
| Drenthe                    | 761                                              |
| Flevoland                  | 1229                                             |
| Food Valley                | 1108                                             |
| Friesland                  | 1980                                             |
| Gooi- en Vechtstreek       | 874                                              |
| Gorinchem                  | 442                                              |
| Groningen                  | 2315                                             |
| Groot Amsterdam            | 4220                                             |
| Haaglanden                 | 2955                                             |
| Helmond-De Peel            | 883                                              |
| Holland Rijnland           | 1777                                             |

|                                    |      |
|------------------------------------|------|
| Midden-Brabant                     | 1462 |
| Midden-Gelderland                  | 1362 |
| Midden-Holland                     | 575  |
| Midden-Limburg                     | 726  |
| Midden-Utrecht                     | 2829 |
| Noord-Holland (Noord)              | 2218 |
| Noord-Limburg                      | 844  |
| Noordoost-Brabant                  | 2029 |
| Regio Zwolle                       | 1572 |
| Rijk van Nijmegen                  | 1187 |
| Rijnmond                           | 4433 |
| Rivierenland                       | 701  |
| Stedendriehoek en Noordwest Veluwe | 2007 |
| Twente                             | 1611 |
| West-Brabant                       | 2174 |
| Zaanstreek / Waterland             | 928  |
| Zeeland                            | 1121 |
| Zuid-Holland Centraal              | 1114 |
| Zuid-Kennemerland en IJmond        | 1132 |
| Zuid-Limburg                       | 1893 |
| Zuidoost-Brabant                   | 1754 |

***Count of individuals included in the backtest test dataset***

| <b>Labor Market Region</b>                                                                                                                        | <b>Number included in backtest test data</b> |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------|
| Achterhoek, Amersfoort, Drenthe, Flevoland, Friesland, Gorinchem, Helmond-De Peel, Midden-Brabant, Midden-Holland, Midden-Limburg, Noord-Limburg, | 80                                           |

|                                                           |     |
|-----------------------------------------------------------|-----|
| Stedendriehoek en Noordwest Veluwe, Zeeland <sup>10</sup> |     |
| Drechtsteden                                              | 18  |
| Food Valley                                               | 21  |
| Gooi- en Vechtstreek                                      | 12  |
| Groningen                                                 | 23  |
| Groot Amsterdam                                           | 114 |
| Haaglanden                                                | 53  |
| Holland Rijnland                                          | 25  |
| Midden-Gelderland                                         | 17  |
| Midden-Utrecht                                            | 35  |
| Noord-Holland (Noord)                                     | 24  |
| Noordoost-Brabant                                         | 23  |
| Regio Zwolle                                              | 24  |
| Rijk van Nijmegen                                         | 13  |
| Rijnmond                                                  | 67  |
| Rivierenland                                              | 15  |
| Twente                                                    | 24  |
| West-Brabant                                              | 23  |
| Zaanstreek / Waterland                                    | 22  |
| Zuid-Holland Centraal                                     | 26  |
| Zuid-Kennemerland en IJmond                               | 27  |
| Zuid-Limburg                                              | 18  |
| Zuidoost-Brabant                                          | 31  |

<sup>10</sup> These LMRs were combined in this table to meet the requirements from exporting tables from the CBS environment because there were 10 or fewer people from each LMR in the test dataset.



ONDERZOEK

# DPIA/AIIA

Risicoanalyse voor het project  
Kansrijke Koppeling met behulp van  
AI

72068 - Vertrouwelijk - 22 maart 2024

5.1.2.e

## Versiebeheer

| Datum versie  | Toelichting                           |
|---------------|---------------------------------------|
| 22 april 2024 | Oplevering versie 1.0 door Berenschot |

---

## Managementsamenvatting

In het licht van de voortdurende inspanningen om de integratie en participatie van statushouders in Nederland te bevorderen, initieert het Centraal Orgaan opvang asielzoekers (COA) een project genaamd 'Kansrijke Koppeling met behulp van AI'. Dit project is gewijd aan het ontwikkelen en implementeren van een innovatief algoritme dat tot doel heeft statushouders effectief te koppelen aan passende arbeidsmarktregio's.

Omdat de voorgenomen AI-toepassing een verhoogd privacyrisico inhoudt, is het uitvoeren van een Data Protection Impact Assessment (DPIA) verplicht. Het COA voert dit DPIA uit ter voorbereiding op de pilotfase van het project Kansrijke Koppeling met behulp van AI, waarin de AI-toepassing van IPL mogelijk wordt geïmplementeerd in Nederland. Het DPIA wordt verrijkt met een risico-inschattinginstrument voor algoritmen van het Electronic Commerce Platform (ECP): Artificial Intelligence Impact Assessment (AIIA). De AIIA is een breder instrument dat zich richt op alle mogelijke ethische en juridische vraagstukken die geassocieerd kunnen worden met de toepassing van AI. De AIIA kijkt niet alleen naar risico's, maar biedt het ook een kader voor het maken van ethische keuzes voor de inzet van AI.

### Risico's van de AI-toepassing

We beschrijven in dit DPIA enkel de belangrijkste risico's op hoofdlijnen en lichten de belangrijkste maatregelen toe. Op hoofdlijnen signaleren we de volgende risico's, welke we in deel C van dit rapport nader toelichten:

1. De doelstellingen van het project worden niet uniform beschreven. Hierdoor ontstaat het risico dat de doelstellingen op basis waarvan de algoritmen worden afgesteld afwijken van de doelstellingen die zijn beoogd door het management.
2. In de projectdocumentatie wordt verschillende keren het beeld geschetst dat de data die wordt gebruikt binnen de AI-toepassing objectief is. Een verkeerd begrip van data kan leiden tot bias.
3. Het is onbekend in hoeverre de voorgenomen acties rondom privacy zijn opgevolgd. De eerdere PIA en AIIA bevatten een hele lijst aan (voorgenomen) acties. Het is ook nog niet duidelijk of de informele dimensie van privacy beschermd wordt. Mogelijk is privacy nog steeds een risico.

### (Persoons-)gegevens

4. Er worden diverse demografische gegevens van vergunninghouders gebruikt. Een deel van deze variabelen zegt onmiskenbaar iets over de etniciteit van de betrokkenen. Dit leidt tot hoog risico op discriminatie. Het risico van sample selection bias ligt ook op de loer. Dit kan ook versterkt worden door de keuzes die medewerkers maken op basis van mogelijke eigen vooringenomenheid.
5. Bewaartermijnen voor het verwerken van persoonsgegevens die zijn verwerkt ten behoeve van training en monitoring zijn niet vastgesteld en geïmplementeerd.
6. Er zijn verschillende afspraken gemaakt met IPL rondom de verwerking van persoonsgegevens. Deze afspraken hebben we niet kunnen valideren. Hierdoor ontstaat het vermoeden dat doorgifte van data buiten de EU toch zal plaatsvinden, tenzij hiervoor aanvullende afspraken worden gemaakt.

### AI-toepassing

7. De uitlegbaarheid van beslissingen in individuele gevallen kan lastig zijn. Ook voor medewerkers is het niet altijd duidelijk waarom tot een bepaald besluit is gekomen. Dit brengt uitdagingen mee voor medewerkers die op basis van dat advies een besluit moeten nemen.
8. Wanneer de human-in-the-loop in de praktijk weinig tot niet afwijkt van de besluitvorming zal er waarschijnlijk sprake zijn van geautomatiseerde besluitvorming. Het risico bestaat dat als

hier sprake van is, dat persoonsgegevens van vergunninghouders onrechtmatig worden verwerkt.

9. Waarborg dat informatiebeveiligingsmaatregelen nader uitgewerkt worden en specifiek gemaakt worden voor deze desbetreffende verwerkingsactiviteiten.

### **Betrokkenen**

10. Er is beperkt inzicht in de belangen van de vergunninghouders doordat zij niet vanaf de start zijn betrokken bij dit project. Het uiterste risico is dat deze groep onbedoeld gedehumaniseerd wordt doordat zij zelf geen autonomie hebben in dit project.
11. De rechten van vergunninghouders zijn potentieel onvoldoende gewaarborgd. We hebben bijvoorbeeld tegenstrijdige informatie over de mogelijkheid van vergunninghouders om bezwaar te maken tegen het gebruik van de AI-toepassing voor hun plaatsing.

### **Maatregelen**

#### **Pilotfase project Kansrijke Koppeling**

1. Datawijsheid. Waarborg voldoende bewustzijn van de risico's en de ethische overwegingen rondom het gebruik van data bij alle stakeholders die betrokken zijn bij het project.
2. Inclusiviteit. Zorg ervoor dat er voldoende stakeholders betrokken worden die de belangen van de vergunninghouders vertegenwoordigen. Als stakeholders niet betrokken kunnen worden bij dit proces, beargumenteer dan waarom zij niet betrokken dienen te worden.
3. (Juridische) ondersteuning. Zorg ervoor dat er voldoende ondersteuning aanwezig is in de vorm van privacy officers, experts, functionarissen gegevensbescherming en ethici om gedurende het project op verschillende vlakken te ondersteunen.
4. (Semi-)geautomatiseerde besluitvorming. Overweeg om de AI-toepassing toch te kwalificeren als een vorm van geautomatiseerde besluitvorming en implementeer de daarbij behorende maatregelen.

#### **Ingebruikname van algoritme na pilotfase**

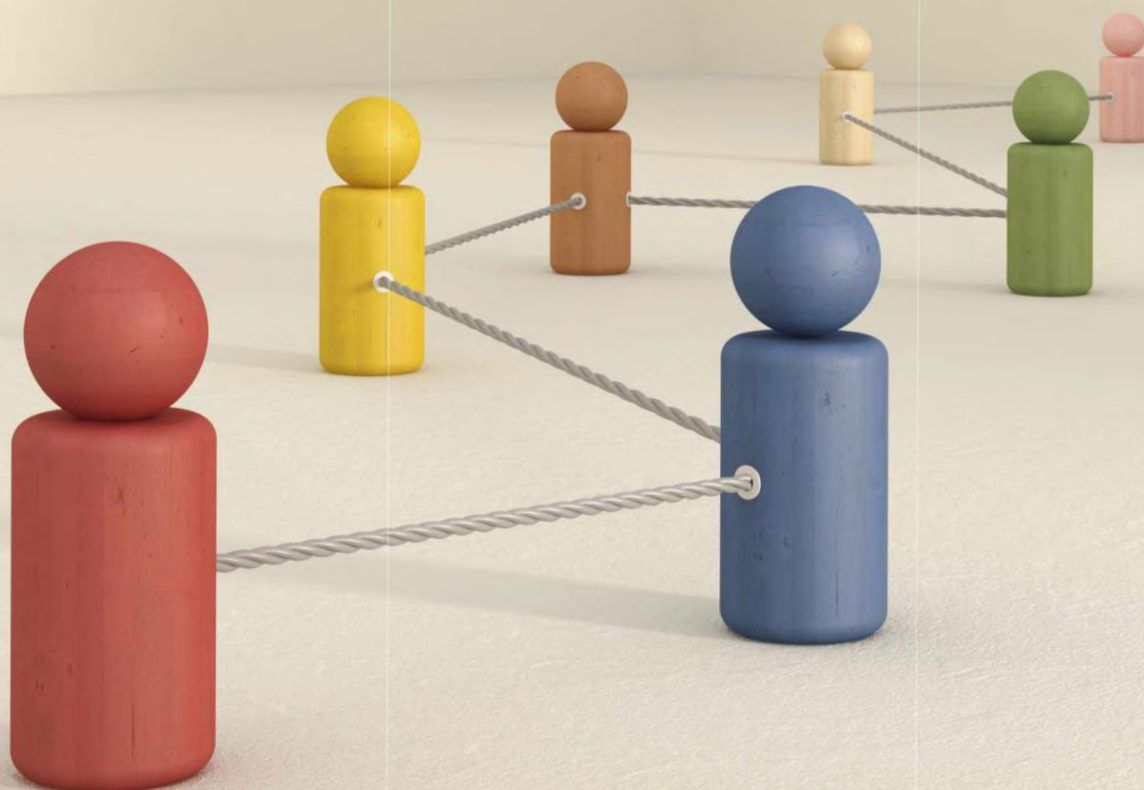
5. Juridische controle. Voer een laatste juridische check, controleer of de eerder vastgestelde kaders nog voldoende worden nageleefd. Heb hierbij ook in het bijzonder aandacht voor de ontwikkelingen rondom de AI-act.
6. Periodiek evalueren. We adviseren om periodiek, lees ten minste elke twee jaar, het algoritme te evalueren en te controleren of de assessments nog van toepassing zijn. Regel hier een standaard procedure voor in, voor zover dit nog niet geregeld is binnen het COA.
7. Bijdrage aan het plaatsingsproces. Het is nog onduidelijk hoe goed de pilot van het algoritme uitgevoerd kan worden, gezien de taakstelling vrij dominant is in het plaatsingsproces. Evalueer na de pilot, afhankelijk van of de rust bij het COA is teruggekeerd, hoe goed het algoritme bijdraagt aan de werkzaamheden van de programmabegeleider en regievoerder.

#### **Wezenlijke veranderingen in het algoritme gedurende levenscyclus**

8. Check op wettelijke kaders. Controleer of het algoritme met deze wijzigingen nog steeds binnen het wettelijke kader past. Evalueer of het algoritme niet langer gebruikt mag worden of dat het met aanpassingen wel mogelijk is.
9. Controleer de ethische impact. Controleer de impact van deze wijzigingen op de (organisatie)waarden. Kijk bijvoorbeeld of de wijziging impact heeft op het mensgerichte karakter van de dienstverlening. Bepaal op basis van deze controle of er aanvullende maatregelen nodig zijn.

# Inhoudsopgave

|                                                                |                                            |
|----------------------------------------------------------------|--------------------------------------------|
| <b>Hoofdstuk 1</b>                                             | <b>6</b>                                   |
| <b>Inleiding</b>                                               | <b>6</b>                                   |
| 1.1 Inleiding                                                  | 7                                          |
| 1.2 Achtergrond DPIA/AIIA                                      | 7                                          |
| 1.3 Aanpak en leeswijzer                                       | 8                                          |
| <b>Hoofdstuk 2</b>                                             | <b>9</b>                                   |
| <b>Deel A: Algemene kenmerken gegevensverwerkingen</b>         | <b>9</b>                                   |
| 2.1 Voorstel                                                   | 10                                         |
| 2.2 Persoonsgegevens                                           | 12                                         |
| 2.3 Gegevensverwerkingen                                       | 15                                         |
| 2.4 Technieken en methoden van gegevensverwerking              | 19                                         |
| 2.5 Verwerkingsdoeleinden                                      | 21                                         |
| 2.6 Betrokken partijen                                         | 22                                         |
| 2.7 Belangen bij de gegevensverwerking                         | 23                                         |
| 2.8 Verwerkingslocaties                                        | 27                                         |
| 2.9 Juridisch en beleidsmatig kader                            | 28                                         |
| 2.10 Bewaartermijnen                                           | 29                                         |
| <b>Hoofdstuk 3</b>                                             | <b>30</b>                                  |
| <b>Deel B: Beoordeling rechtmatigheid gegevensverwerkingen</b> | <b>30</b>                                  |
| 3.1 Rechtsgrond                                                | 31                                         |
| 3.2 Bijzondere persoonsgegevens                                | 34                                         |
| 3.3 Doelbinding                                                | 35                                         |
| 3.4 Noodzaak en evenredigheid                                  | 36                                         |
| 3.5 Rechten van de betrokkene                                  | 36                                         |
| <b>Hoofdstuk 4</b>                                             | <b>40</b>                                  |
| 4.1 Risico's voor de betrokkenen                               | 41                                         |
| <b>Hoofdstuk 5</b>                                             | <b>44</b>                                  |
| 5.1 Maatregelen                                                | 45                                         |
| <b>Bijlage 1. Variabelen</b>                                   | <b>47</b>                                  |
| <b>Bijlage 2. Gedragscode Artificiële Intelligentie</b>        | <b>63</b>                                  |
| <b>Bijlage 3. Invulschema grondrechten IAMA</b>                | <b>Fout! Bladwijzer niet gedefinieerd.</b> |
| <b>Bijlage 4. Geraadpleegde documenten</b>                     | <b>64</b>                                  |



## HOOFDSTUK 1

# Inleiding

Dit hoofdstuk behandelt het doel van het DPIA/AIIA, de gebruikte modellen en de structuur van het document.

## 1.1 Inleiding

Het Centraal Orgaan opvang asielzoekers (COA) heeft als taak om de materiële en immateriële opvang van asielzoekers te verzorgen, alsmede het plaatsen van asielzoekers op gemeentelijke opvangplaatsen en het verrichten van werkzaamheden met betrekking tot de bemiddeling bij de uitstroom van verblijfsgerechtigden.<sup>1</sup> Het gaat hier om een wettelijke taak. In het verlengde van deze wettelijke taak, is het in belang van het COA om de koppeling van vergunninghouders aan gemeenten te verbeteren. Immigration Policy Lab (IPL) heeft een Artificiële Intelligentie-toepassing (AI-toepassing) ontwikkeld die overheden kan helpen bij het plaatsen van vergunninghouders in regio's waar zij het meeste kans hebben op succesvolle integratie. IPL heeft hiervoor wetenschappelijk onderzoek uitgevoerd dat laat zien dat er bij gebruik van de AI-toepassing van IPL significante potentiële voordelen zijn te behalen voor arbeidsmarktparticipatie van vergunninghouders. De AI-toepassing van IPL wordt al gebruikt in Zwitserland. In Zweden vindt onderzoek plaats naar de mogelijke voordelen van de AI-toepassing. In Canada, de Verenigde Staten en Nederland heeft vergelijkbaar onderzoek plaatsgevonden. Het beoogde doel van de inzet van de AI-toepassing is het verbeteren van de kwaliteit van de koppeling tussen vergunninghouder en gemeente.

## 1.2 Achtergrond DPIA/AIIA

Omdat de voorgenomen AI-toepassing een verhoogd privacyrisico inhoudt, is het uitvoeren van een Data Protection Impact Assessment (DPIA) verplicht op grond van artikel 35 van de Algemene verordening gegevensbescherming (AVG). Het COA voert dit DPIA uit ter voorbereiding op de pilotfase van het project Kansrijke Koppeling met behulp van AI, waarin de AI-toepassing van IPL mogelijk wordt geïmplementeerd in Nederland. Het DPIA wordt verrijkt met een risico-inschattinginstrument voor algoritmen van het Electronic Commerce Platform (ECP): Artificial Intelligence Impact Assessment (AIIA). De AIIA is een breder instrument dat zich richt op alle mogelijke ethische en juridische vraagstukken die geassocieerd kunnen worden met de toepassing van AI. De AIIA kijkt niet alleen naar risico's, maar biedt het ook een kader voor het maken van ethische keuzes voor de inzet van AI.

### Data Protection Impact Assessment

In het DPIA wordt in beeld gebracht of kan worden voldaan aan de beginselen van de AVG, of de AI-toepassing juridisch haalbaar is, wat de privacyrisico's van de voorgenomen verwerking(en) van persoonsgegevens zijn en welke maatregelen passend zijn om risico's die volgen uit gegevensverwerkingen te beheersen. Ook dient dit DPIA om de procesmatige totstandkoming en implementatie vast te leggen en zorgvuldig handelen te onderbouwen, zowel bij de ontwikkeling als bij de inzet van de AI-toepassing. Mede aan de hand van dit DPIA kan worden besloten over de wenselijkheid om het project voort te zetten door te starten met een pilotfase. We hanteren het Model DPIA Rijksdienst, dit model gebruikt 17 punten om een DPIA uit te voeren.

Dit DPIA wordt ter advies voorgelegd aan de functionaris gegevensbescherming (FG) van het COA. Expliciet wordt de FG gevraagd te adviseren over de rechtmatigheid van de voorgenomen verwerkingen van persoonsgegevens, in het bijzonder rekening houdend met proportionaliteit en subsidiariteit, en de wijze waarop geïdentificeerde risico's worden gemitigeerd. Dit advies zal van belang zijn bij definitieve besluitvorming door het COA.

---

<sup>1</sup> Zoals bedoeld in artikel 1, eerste lid, onderdeel g, van de Huisvestingswet 2014 naar door burgemeester en wethouders beschikbaar gestelde huisvesting (zie artikel 3 van de Wet COA).

## Artificial Intelligence Impact Assessment

Inzet van nieuwe technologie kan enerzijds maatschappelijke en economische voordelen hebben, maar kan anderzijds risico's voor de veiligheid en fundamentele rechten van vergunninghouders met zich meebrengen. Het uitvoeren van het AIIA is nuttig en noodzakelijk, omdat meerdere screeningvragen van stap 1 van het AIIA met 'ja' beantwoord kunnen worden. De vragen hebben betrekking op de maatschappelijke en politieke context, de kenmerken van de technologie en de processen waarvan de technologie deel uitmaakt. Door dit AIIA uit te voeren streeft het COA ernaar om enerzijds ruimte te bieden voor innovatie en het benutten van de kansen die AI biedt en anderzijds de risico's voor de veiligheid en fundamentele rechten van vergunninghouders in beeld te brengen en te beperken.

De screeningvragen wijzen op een aantal mogelijk risicovolle aspecten van de toepassing van het algoritme van IPL die maken dat ethische beraadslaging wordt aangeraden.

### 1.3 Aanpak en leeswijzer

Tussen het uitvoeren van een DPIA en AIIA bestaat de nodige overlap. Beide onderzoeken vergen gelijksoortige informatie van dezelfde stakeholders. Dat betekent dat we voor zowel het DPIA als het AIIA dezelfde aanpak volgden. In de uitvoering hebben we beide assessments dan ook parallel aan elkaar uitgevoerd. Omdat het COA eerder al een DPIA heeft opgesteld en ook al een (vroeg) concept-AIIA heeft vastgesteld, gebruiken we deze documenten als basis voor het nieuwe DPIA en AIIA. Onze aanpak bestaat uit de volgende stappen:



- **Stap 1:** Beschrijven context (hoofdstuk 2)
- **Stap 2:** Beoordeling rechtmatigheid (hoofdstuk 3)
- **Stap 3:** Beschrijving en beoordeling risico's (hoofdstuk 4)
- **Stap 4:** Definieren maatregelen (hoofdstuk 5)



## HOOFDSTUK 2

# Deel A: Algemene kenmerken gegevensverwerkingen

Een overzicht van de relevante feiten van de gegevensverwerkingen.

## 2.1 Voorstel

| Rijksmodel DPIA                                                                                                                                                                                                                                                        | AIIA                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Beschrijf het voorstel waar het DPIA op toeziet op hoofdlijnen.</li> <li>2. Benoem hoe het voorstel tot stand is gekomen.</li> <li>3. Benoem wat de beweegredenen zijn achter de totstandkoming van het voorstel.</li> </ol> | 2.1 Beschrijf de toepassing en het doel van de toepassing. |

In het licht van de voortdurende inspanningen om de integratie en participatie van statushouders in Nederland te bevorderen, initieert het Centraal Orgaan opvang asielzoekers (COA) een project genaamd 'Kansrijke Koppeling met behulp van AI'. Dit project is gewijd aan het ontwikkelen en implementeren van een innovatief algoritme dat tot doel heeft statushouders effectief te koppelen aan passende arbeidsmarktregio's. Het betreft het komen tot een (significant) beter regioadvies voor vergunninghouders, te meten via een verhoogde arbeidsmarktparticipatie van vergunninghouders.

### Het project Kansrijke Koppeling met behulp van AI

'Kansrijke Koppeling met behulp van AI' is ontworpen om de overgang van statushouders naar de Nederlandse samenleving te vergemakkelijken door hen te koppelen aan regio's waar hun vaardigheden en ervaring het meest waardevol en nodig zijn. Dit wordt bereikt door een algoritme dat rekening houdt met een reeks variabelen, waaronder de individuele vaardigheden, kwalificaties en voorkeuren van de statushouders, evenals de specifieke behoeften en kansen van verschillende arbeidsmarktregio's. Het algoritme formuleert een niet-bindend regioadvies dat helpt ter ondersteuning bij het koppelen van een vergunninghouder aan de gemeente waar deze zo kansrijk mogelijk is. De implementatie van dit algoritme is gebaseerd op de overtuiging dat een doelgerichte plaatsing van statushouders niet alleen hun individuele kansen op succesvolle integratie en economische zelfstandigheid zal vergroten, maar ook zal bijdragen aan de diversiteit en vitaliteit van de Nederlandse arbeidsmarkt.

De AI-toepassing is tijdens een ontwikkelingsfase getraind. Dat proces is qua verwerkingen vergelijkbaar met fase 1 van het project 'Kansrijke koppeling met behulp van AI'. Het COA (MIT) deelt hiertoe een bestand (met persoonsgegevens) vanuit IBIS met het CBS. Door een specialistische afdeling van het CBS worden de persoonsgegevens gepseudonimiseerd en 'verRIND'.<sup>2</sup> Door binnen de RA-omgeving (Remote Accesomgeving) van CBS onderzoek te doen aan de hand van deze historische data, bouwt IPL de AI-toepassing op een wijze die passend is voor de Nederlandse situatie binnen de RA-omgeving van CBS. Daarna heeft IPL de AI-toepassing, die passend is voor de Nederlandse situatie beschikbaar gesteld aan het COA. De AI-toepassing wordt dan gehost binnen de omgeving van het COA (on-premise). Het COA gebruikt de AI-toepassing om ambtenaren te ondersteunen door hen te adviseren bij de plaatsing van vergunninghouders. Vervolgens zal er een terugkoppeling zijn over onder meer het wel/niet opvolgen van het regioadvies (de voorspelling). Aan de hand hiervan kan IPL de AI-toepassing in de toekomst optimaliseren.

<sup>2</sup> Dat houdt in dat de data door CBS automatisch wordt versleuteld met een uniek RIN-nummer, op basis van specifieke koppeldata zoals het vreemdelingennummer.

## Businesscase

De businesscase voor 'Kansrijke Koppeling met behulp van AI' is gebaseerd op de volgende overwegingen:

**Effectiviteit:** door statushouders te koppelen aan regio's waar hun vaardigheden het meest nodig zijn, wordt de kans op succesvolle integratie en participatie vergroot.

**Economische impact:** het project zal bijdragen aan het verminderen van de werkloosheid onder statushouders en hun economische bijdrage aan de samenleving vergroten.

**Sociale cohesie:** door de integratie en participatie van statushouders te bevorderen, wordt de sociale cohesie en diversiteit binnen de Nederlandse samenleving versterkt.

Het algoritme biedt het COA datagedreven inzichten voor betrouwbare besluitvorming.

- Het algoritme stimuleert bij het COA een cultuur van continu leren en verbeteren.
- Data-analyse helpt het COA bij het identificeren van trends en het ontwikkelen van proactieve strategieën.
- Het gebruik van data verhoogt de transparantie en versterkt het vertrouwen in het COA bij stakeholders.
- Het algoritme stelt het COA in staat om ondersteuning aan statushouders te personaliseren, wat duurzame integratie bevordert.
- De implementatie van het algoritme markeert een belangrijke stap in de digitale transformatie van het COA.
- Het algoritme wordt als positief en bruikbaar gezien als er een significante stijging van 4%-punt toename arbeidsparticipatie onder statushouders is in het eerste jaar na uitstroom naar een gemeente. Ten opzichte van 12% in de huidige situatie, is dat een relatieve verbetering van 33% op arbeidsmarktparticipatie is ten opzichte van de testgroep die geen match beschikbaar krijgt.<sup>3</sup>

## Doelstellingen en verwachtingen

Het primaire doel van dit project is het ontwikkelen en implementeren van het algoritme om de koppeling van statushouders aan arbeidsmarktregio's te optimaliseren. We streven ernaar om door middel van dit initiatief de toekomstige levenskwaliteit van statushouders te verbeteren, de integratieprocessen te versnellen en bij te dragen aan de arbeidsparticipatie in Nederland.

Statushouders hebben op dit moment een minder zekere positie op de arbeidsmarkt en hun arbeidsparticipatie is relatief laag.<sup>4</sup> Het betreft hier vergunninghouders die met ondersteuning van een algoritme worden gekoppeld aan een zo kansrijk mogelijke arbeidsmarktregio op basis van hun kenmerken.<sup>5</sup> De verhoging in arbeidsmarktparticipatie wordt gerealiseerd (en gemonitord) wanneer de vergunninghouders al zijn uitgestroomd uit de COA-opvang.

Dit project kan bijdragen aan de realisatie van diverse tactische (t) en operationele (o) doelstellingen van de COA Meerjarenstrategie 2020-2025.

- Kansrijke koppeling nieuwe statushouders (o).
- Het creëren van gezamenlijke oplossingen voor nationale en lokale maatschappelijke vraagstukken (t).
- Resultaten en meer draagvlak door goede samenwerking met betrokken partners (o).

---

<sup>3</sup> COA – Project Initiation Document.

<sup>4</sup> <https://www.cbs.nl/nl-nl/nieuws/2021/15/arbeidsparticipatie-statushouders-stagneert>.

<sup>5</sup> <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-Investeren-in-de-arbeidsmarktintegratie-van-statushouders.pdf>.

- Begrip, draagvlak en reputatie door open proactieve communicatie (o) en door integratie van de statushouder in de Nederlandse maatschappij.<sup>6</sup>

## 2.2 Persoonsgegevens

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                       | AIIA                                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Beschrijf alle persoonsgegevens die worden verwerkt.</li> <li>2. Classificeer alle persoonsgegevens.</li> <li>3. Geef per categorie persoonsgegevens aan welke persoonsgegevens worden verzameld.</li> <li>4. Geef aan wat de bron is van deze persoonsgegevens.</li> </ol> | <ol style="list-style-type: none"> <li>1. Beschrijf welke data worden gebruikt in het kader van de toepassing: de aard van de data, de kwaliteit van de data en de bron van de data.</li> </ol> |

### Betrokkenen

In deze paragraaf beschrijven we de gebruikte persoonsgegevens aan de hand van een aantal categorieën. De toepassing waar dit DPIA op toeziet, richt zich primair op de verwerking van persoonsgegevens van vergunninghouders/asielzoekers. Deze groep betrokkenen kenmerkt zich als een kwetsbare groep binnen de maatschappij. Deze groep bevindt zich vaak een lange tijd in onzekerheid over een toekomst in Nederland en over onder welke omstandigheden deze toekomst plaatsvindt. Daarbij heeft deze groep slechts beperkte invloed op verschillende aspecten als wonen, werk, en inkomen.

De scope van dit DPIA richt zich op de vergunninghouder van wie de gegevens binnen de AI-toepassing worden verwerkt. De volledige beschrijving van alle variabelen treft u aan in **Bijlage 1**. Onder iedere categorie persoonsgegevens kunnen meerdere variabelen vallen.

### Persoonsgegevens

In onderstaande tabel hebben we de verschillende variabelen gecategoriseerd en geclassificeerd.

| Categorie                        | Variabelen                                                                                                                                                                                                                                                                                                                                                                     | Classificatie                    |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| Identificatienummers             | Vnummer_crypt, BSN_crypt                                                                                                                                                                                                                                                                                                                                                       | Wettelijke identificatiegegevens |
| Demografische gegevens           | Omschrijving_Geslacht, Geboortedatum, Omschrijving_Burg_Staat, Omschrijving_moedertaal, Omschrijving_land_van_herkomst onderwijs_ja, Geschat_niveau, Diploma_behaald, Omschrijving_land_vaneducatie, Startopleiding, Eindopleiding, Diploma_bezit, Omschrijving_land_werkervaringopgedaan, Branche_Omschrijving, Aanstelling, Begindatum_werkervaring, Einddatum_werkervaring, | Gewoon/gevoelig                  |
| Asielpprocedure vergunninghouder | VDLibisnr, ingangsdatum_Verblijfplaats, Nummer_plaatsingseenheid_PLE_crypt, classificatie_indicatie_nareiziger, Groepsrol_MVV_indicatie_nareiziger, Datum_huisvestings_gesprek, Opvangvorm, Gekoppeld Aan Gemeente, Datum koppeling aan gemeente,                                                                                                                              | Gewoon                           |

<sup>6</sup> In de Meerjarenstrategie van het COA is 'Kansrijke koppeling nieuwe vergunninghouders' opgenomen als operationele doelstelling. Ook is het project Kansrijke koppeling met behulp van AI vermeld als verbeteropgave in het jaarplan 2022.

|            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |        |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
|            | Indicatie_Vergunninghouder, Soort_Plaatsing, Dienst, Classificatie_Hervestiger_Vervallen, Hardecriteriavantoeppassing                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |        |
| Advies IPL | geomatch_regioadvies_1, geomatch_regioadvies_eis_1, geomatch_regioadvies_current_1, geomatch_regioadvies_future_1, geomatch_regioadvies_ps_endorsement_1, geomatch_regioadvies_2, geomatch_regioadvies_eis_2, geomatch_regioadvies_current_2, geomatch_regioadvies_future_2, geomatch_regioadvies_ps_endorsement_2, geomatch_regioadvies_3, geomatch_regioadvies_eis_3, geomatch_regioadvies_current_3, geomatch_regioadvies_future_3, geomatch_regioadvies_ps_endorsement_3, ps_geomatch_feedback, ple_geomatch_treatment, Omschrijvingarbeidsregio, Omschrijvingarbeidsregio_eis, Verblijfplaats (ple_sm_location_id) | Gewoon |

## Databronnen

De AI-toepassing wordt getraind binnen de RA-omgeving van het CBS. Waar het tot personen herleidbare gegevens betreft, zal er sprake zijn gepseudonimiseerde data. De te gebruiken data zijn afkomstig uit de volgende databronnen:

- **Microdatabestand CBS.** Het COA heeft inzichtelijk uit welke bronnen het Microdatabestand CBS is opgebouwd en van welke organisaties de data afkomstig zijn. Te denken valt aan onderwijsgegevens en inkomensgegevens. Voor het asiel cohortonderzoek deelt het COA al langere tijd gegevens met het CBS. Deze gegevens zijn eveneens opgenomen in het Microdatabestand van het CBS.
- **Comprehensive IBIS Variables Spreadsheet - IPL Data Request.** In deze lijst staan extra gegevensvelden uit IBIS die een relatie hebben met het proces Kansrijke Koppeling. Het betreft variabelen die het COA – naast de IBIS-variabelen voor het asiel cohortonderzoek – mag delen met het CBS. Omdat het hier variabelen betreft die betrekking hebben op de kenmerken van vergunninghouders, zoals persoonsgegevens en zachte criteria, houdt dit mogelijk een privacyrisico in. De gegevens zijn primair verwerkt in het kader van de uitvoering van het primaire proces door het COA. Het betreft gegevens die door het COA zelf zijn verzameld, bijvoorbeeld tijdens het screeningsgesprek en gegevens die in het kader van het vervullen van wettelijke taken beschikbaar zijn gekomen voor het COA via de Basisvoorziening Vreemdelingen (hierna: 'BVV').

| Bron                   | Variabelen                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Asielcohort</b>     | EconPositie_Werkende, SLNLBPH, onderwijs_ja, PEILMAAND, Vnummer_crypt, BSN_crypt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>IBIS</b>            | Vnummer_crypt, BSN_crypt, VDLibisnr, Omschrijving_Geslacht, Omschrijving_Burg_Staat, Omschrijving_land_van_herkomst, Omschrijving_moedertaal, Ingangsdatum_Verblijfplaats, Geboortedatum, Geschat_niveau, Diploma_behaald, Omschrijving_land_vaneducatie, Startopleiding, Eindopleiding, Diploma_bezit, Omschrijving_land_werkervaringopgedaan, Branche_Omschrijving, Aanstelling, Begindatum_werkervaring, Einddatum_werkervaring, Nummer_plaatsingseenheid_PLE_crypt, classificatie_indicatie_nareiziger, Groepsrol_MVV_indicatie_nareiziger, Datum_huisvestings_gesprek, Opvangvorm, Gekoppeld Aan Gemeente, Datum koppeling aan gemeente, Indicatie_Vergunninghouder, Soort_Plaatsing, Dienst, Classificatie_Hervestiger_Vervallen, Hardecriteriavantoepassing, geomatch_regioadvies_1, geomatch_regioadvies_eis_1, geomatch_regioadvies_current_1, geomatch_regioadvies_future_1, geomatch_regioadvies_ps_endorsement_1, geomatch_regioadvies_2, geomatch_regioadvies_eis_2, geomatch_regioadvies_current_2, geomatch_regioadvies_future_2, geomatch_regioadvies_ps_endorsement_2, geomatch_regioadvies_3, geomatch_regioadvies_eis_3, geomatch_regioadvies_current_3, geomatch_regioadvies_future_3, geomatch_regioadvies_ps_endorsement_3, ps_geomatch_feedback, ple_geomatch_treatment, Omschrijvingarbeidsregio, Omschrijvingarbeidsregio_eis, Verblijfplaats (ple_sm_location_id), ple_ps_ids, Einddatum_Verblijfplaats, Date of arrival in municipality, location_in_geomatch_pilot, geomatch_ple_percentage, LMR for Municipality, UR for LMR, PLE_aantal_vergunninghouders, Referentiekenmerk, Datum_Bekend_Referentiekenmerk, ple_sm_ur (Verblijfplaats UR) |
| <b>Geomatch Server</b> | GeoMatch Monitoring Export                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>SPOLIS</b>          | PEILMAAND, SDATUMAANVANG, SDATUMEINDE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>TVS</b>             | Recent Linkings, TVS Forecast, Regioadvies overgenomen, Reden koppeling, Reden niet overnemen regioadvies, Toelichting niet overnemen regioadvies, Opmerking, gemeente_percent_full, ple_ao_ids                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

## Datakwaliteit en dataminimalisatie

AI-toepassingen hebben naar hun aard grote behoefte aan data. Tijdens de gehele levenscyclus van de AI-toepassing moet worden vermeden dat persoonsgegevens onrechtmatig worden verzameld, verwerkt of gegenereerd. Zo kan het beginsel van dataminimalisatie worden gewaarborgd.

Om te voorkomen dat er te veel persoonsgegevens worden gebruikt en daarmee het beginsel van dataminimalisatie wordt overschreden, zal bij het ontwerpen en gedurende de toepassing van de (mogelijk te introduceren) AI-toepassing moeten worden geborgd dat zo min mogelijk persoonsgegevens worden gebruikt. Er dient actief te worden gestuurd op dataminimalisatie. Dat vergt expliciete afwegingen met betrekking tot proportionaliteit. Het COA waarborgt proportionele omgang met persoonsgegevens bijvoorbeeld door de data eerst te onderzoeken zodat kan worden bepaald welke geschikt zijn voor gebruik in fase 2. Aan de hand daarvan kan door het COA worden besloten wat wenselijk is om te gebruiken. Als tijdens de pilotfase blijkt dat bepaalde inputgegevens overbodig zijn, kan de verwerking daarvan ook worden gestaakt.

De kwaliteit van de gegevens die worden gebruikt om de AI-toepassing te trainen is essentieel om een kwalitatieve uitkomst te verzekeren. Daarbij is het van groot belang dat de gegevens accuraat, voldoende breed en representatief zijn, maar ook dat ze geen sociaal geconstrueerde vertekeningen of onnauwkeurigheden (bias) bevatten. De persoonsgegevens die worden verwerkt moeten juist zijn, om te voorkomen dat de AI-toepassing leidt tot onvoorzienbare uitkomsten en ongewenste effecten voor de betrokkene. Bij AI-toepassingen wordt extra waarde gehecht aan het beginsel van juistheid omdat dat eventuele vertekeningen (bias) voorkomt. In aanbeveling 11 van het rapport van Deloitte wordt daarom ook aangegeven dat het wenselijk is om de analyses die zijn uitgevoerd om de datakwaliteit te controleren, inclusief de resultaten en conclusies die voortkomen uit de analyses, te documenteren.<sup>7</sup> Verder is het aanbevolen om de datakwaliteit te monitoren wanneer het algoritme opnieuw getraind gaat worden. Beide acties zijn opgenomen in de actielijst pilotfase. Verder onderneemt IPL ook acties om de datakwaliteit routinematig te testen, zo checken zij bijvoorbeeld de verschillende variabelen op afwijkingen en foutieve labels.<sup>8</sup>

Omdat het COA intern en - waar relevant - extern verantwoording moet kunnen afleggen over deze beheersmaatregelen, wordt in zijn algemeenheid een 'algoritmische audit' aanbevolen. Hiermee wordt door een onafhankelijke derde partij het gebruik van de algoritmen en data getoetst. Als onderdeel van de algoritmische audit heeft het COA aan het CBS gevraagd de (input-)data van de AI-toepassing te toetsen. Ook kijkt Deloitte naar de gebruikte data (en de analyse van het CBS) en het gebruikte model. Naar verwachting zullen de analyses van het CBS en Deloitte ook resulteren in nieuw te nemen mitigerende maatregelen. Daarmee wordt ook de kwaliteit en integriteit van de data gedocumenteerd.

## 2.3 Gegevensverwerkingen

| Rijksmodel DPIA                                                                                                                                                                           | AIIA                |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| <ol style="list-style-type: none"> <li>1. Beschrijf alle gegevensverwerkingen.</li> <li>2. Geef aan welke categorieën persoonsgegevens worden verwerkt per gegevensverwerking.</li> </ol> | Niet van toepassing |

In het eerdere Privacy Impact Assessment (PIA) werden de volgende drie voorgenomen gegevensverwerkingen onderscheiden:<sup>9</sup>

1. trainen
2. adviseren
3. monitoren.

In deze paragraaf worden de voorgenomen gegevensverwerkingen binnen deze stappen grafisch weergegeven en beschreven. Wanneer dat passend is, zullen deze stappen terugkomen in andere paragrafen van dit DPIA.

### 1. Trainen

Het trainen van de AI-toepassing gebeurt aan de hand van historische data van het COA en het CBS door IPL. Ook voert IPL nieuw wetenschappelijk onderzoek uit om in beeld te brengen wat de nieuwe werkwijze concreet zal opleveren.

<sup>7</sup> Algoritme Analyse Kansrijke Koppeling, Deloitte, 19 juli 2022

<sup>8</sup> GeoMatch Algorithm Technical Documentation, IPL, 16 februari 2024

<sup>9</sup> PIA Project Kansrijke koppeling v0.8, 26 juni 2022

Daartoe verstrekt het COA (MIT) een bestand (met persoonsgegevens) in elektronische vorm aan het CBS (vanuit IBIS, via het veilige CBS-uploadportaal). CBS pseudonimiseert de van het COA verkregen gegevens volgens de standaard pseudonimiseringstechnieken van het CBS ('verRINnen'). Na koppeling met de gegevens van het CBS, zet het CBS de gegevens klaar voor onderzoek in zijn remote access (Microdata-)omgeving. Deze beveiligde RA-onderzoeksomgeving van het CBS wordt opengesteld voor IPL. IPL verwerkt de gepseudonimiseerde persoonsgegevens ten behoeve van wetenschappelijk onderzoek.

De dataflow van 'stap 1 – trainen' ziet er als volgt uit:

Bij het trainen van de AI-toepassing worden de volgende gegevens verwerkt:



## Gegevens

EconPositie\_Werkende, SLNLBPH, onderwijs\_ja, PEILMAAND, PEILMAAND, SDATUMAANVANG, SDATUMEINDE, Vnummer\_crypt, BSN\_crypt, VDLibisnr, Omschrijving\_Geslacht, Omschrijving\_Burg\_Staat, Omschrijving\_land\_van\_herkomst, Omschrijving\_moedertaal, Ingangsdatum\_Verblijfplaats, Geboortedatum, Geschat\_niveau, Diploma\_behaald, Omschrijving\_land\_vaneducatie, Startopleiding, Eindopleiding, Diploma\_bezit, Omschrijving\_land\_werkervaringopgedaan, Branche\_Omschrijving, Aanstelling, Begindatum\_werkervaring, Einddatum\_werkervaring, Nummer\_plaatsingseenheid\_PLE\_crypt, classificatie\_indicatie\_nareiziger, Groepsrol\_MVV\_indicatie\_nareiziger, Datum\_huisvestings\_gesprek, Opvangvorm, Gekoppeld Aan Gemeente, Datum koppeling aan gemeente, Indicatie\_Vergunninghouder, Soort\_Plaatsing, Dienst, Classificatie\_Hervestiger\_Vervallen, Hardecriteriaavantoepassing, Verblijfplaats (ple\_sm\_location\_id), Einddatum\_Verblijfplaats, LMR for Municipality, UR for LMR, Regioadvies overgenomen, PLE\_aantal\_vergunninghouders, Referentiekenmerk, Datum\_Bekend\_Referentiekenmerk

## 2. Adviseren

Het voorstel is dat de AI-toepassing on-premise binnen de COA ICT-omgeving wordt geïnstalleerd. Als de geïnstalleerde AI-toepassing wordt gebruikt om te adviseren, ziet het proces – en de daarmee samenhangende verwerking(en) van persoonsgegevens – er als volgt uit:

De stappen 2a.3; 2a.6 en 2b. zijn (deels) nieuw ten opzichte van de huidige situatie.

De AI-toepassing heeft invloed op de stappen 2a.6 tot en met a.10.

2a.0 Als tijdens het huisvestingsgesprek blijkt dat er harde plaatsingscriteria van toepassing zijn, wordt er gekoppeld aan een gemeente volgens dit harde criterium en wordt de AI-toepassing niet gebruikt.

2a.1 Screeningsgesprek waarbij de programmabegeleider zachte plaatsingscriteria ophaalt bij vergunninghouder.

2a.2 De programmabegeleider registreert zachte plaatsingscriteria in IBIS.

2a.3 Persoonsgegevens vergunninghouder (kenmerken) worden verzonden naar de AI-toepassing.

2a.4 De AI-toepassing verwerkt de gegevens en geeft een regioadvies.

2a.5 IBIS geeft het regioadvies door aan de betreffende programmabegeleider.

2a.6 De programmabegeleider voert het volgende in IBIS in:

Regioadvies AI-toepassing overgenomen ja/nee (checkbox)

Indien nee: regioadvies programmabegeleider

Reden waarom regioadvies AI-toepassing niet is overgenomen (nog nader te bepalen of het multiple choice of een vrij tekstveld moet zijn, of allebei).

2a.7 Er bestaat een geautomatiseerde koppeling tussen IBIS en TVS.

Ten opzichte van de huidige situatie, zullen daarin de volgende velden nieuw worden meegenomen:

Regioadvies AI-toepassing

Regioadvies overgenomen ja/nee (checkbox)

Indien nee: regioadvies programmabegeleider

Reden waarom regioadvies AI-toepassing niet is overgenomen

2a.8 Op basis van de informatie over het regioadvies in TVS, koppelt de regievoerder de vergunninghouder vervolgens aan een gemeente.

Ten opzichte van de huidige situatie, is het regioadvies mede tot stand gekomen met behulp van de AI-toepassing.

2a.9 De regievoerder legt in TVS vast aan welke gemeente de vergunninghouder wordt gekoppeld, of het regioadvies wordt opgevolgd en – indien van toepassing – wat de reden is voor het niet opvolgen daarvan.

Ten opzichte van de huidige situatie, is de koppeling (mogelijk) mede tot stand gekomen met behulp van AI-ondersteuning.

2a.10 Er bestaat een geautomatiseerde koppeling tussen TVS en IBIS.

Ten opzichte van de huidige situatie, is de koppeling (mogelijk) mede tot stand gekomen met behulp van AI-ondersteuning.

2b. Feedbackloop waarin IBIS geautomatiseerd onder andere de volgende informatie doorgeeft aan de AI-toepassing ten behoeve van het – waar nodig – leren van en het waar nodig aanpassen (optimaliseren) aan de hand van feedback over gegeven regioadviezen (machine learning).

- Regioadvies overgenomen door programmabegeleider + indien van toepassing de reden voor het niet overnemen daarvan.
- Regioadvies overgenomen door regievoerder + indien van toepassing de reden voor het niet overnemen daarvan.

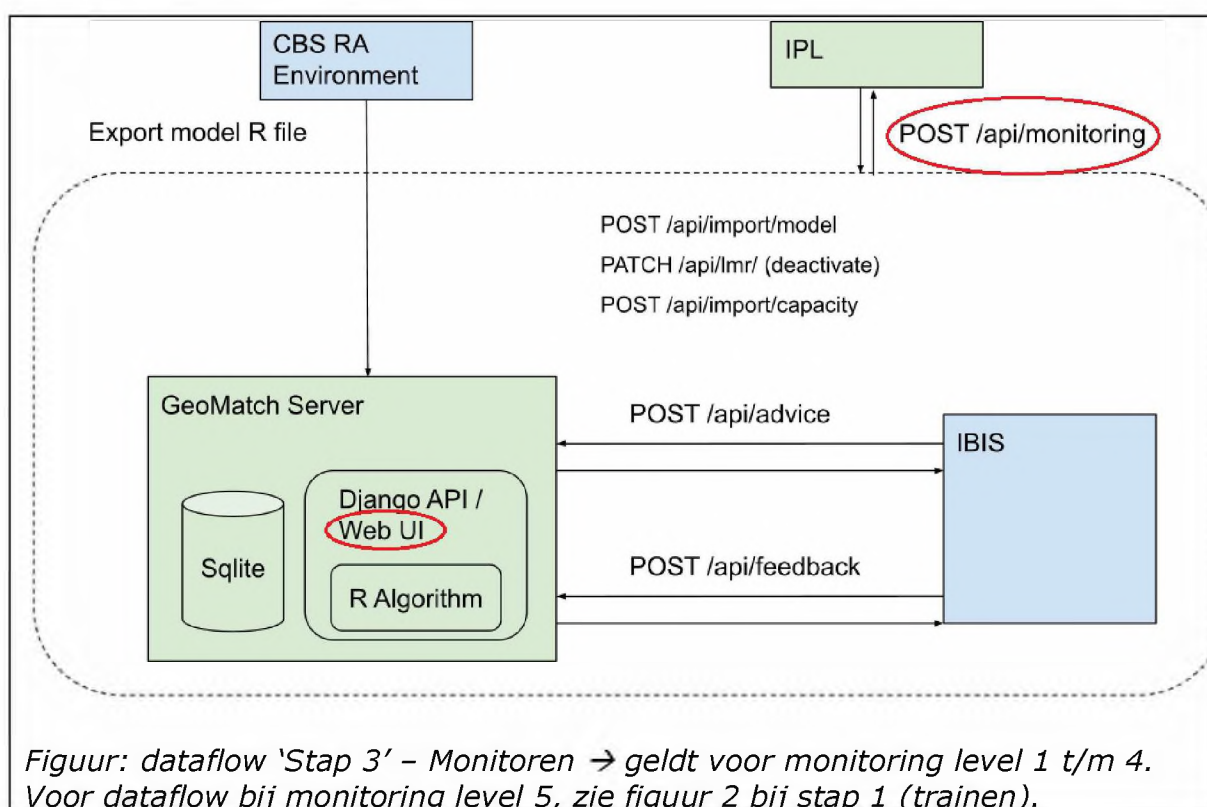
- Gekoppelde gemeenten (nodig om nog te vullen taakstelling aan te passen voor volgende adviezen) en datum koppeling.

### 3. Monitoren

De leverancier (IPL) van de AI-toepassing moet monitoren of de regioadviezen goed worden overgenomen en of de parameters in de AI-toepassing moeten worden aangepast. Daarom zullen, als de AI-toepassing daadwerkelijk wordt gebruikt, de uitkomsten worden gemonitord en – waar nodig – verder worden aangepast en verbeterd (optimalisatie). In het monitoringplan worden conform aanbeveling 7 van het Deloitte-rapport, de monitoringsactiviteiten, toegestane marges, governance en een stopknop beschreven.

IPL heeft een monitoringplan aangeleverd waarin ze aangeven hoe de monitoring zal plaatsvinden. Het monitoringplan is te vinden als bijlage 1 van dit PIA. IPL onderscheidt 5 monitoringlevels, waarbij gegevens tijdens stap 3 op twee verschillende manieren van het COA naar IPL gaan. Voor monitoringlevel 1 tot en met 4 zal het delen van de gegevens kunnen via de Web User Interface (UI), zoals weergegeven in onderstaande afbeelding. Voor monitoringlevel 5 zullen gegevens (minimaal één keer per jaar en maximaal vier keer per jaar) gedeeld worden op de manier zoals ook beschreven wordt in figuur 1 bij stap 1 (trainen).

Schematisch ziet dat er als volgt uit:



Uitgangspunt: idealiter vindt monitoring plaats aan de hand van anonieme, statistische en/of geaggregeerde gegevens.

Hierover is aanvullende informatie vereist van IPL. Indien de kwaliteit van de monitoring niet toereikend is, gebruikmakend van geaggregeerde gegevens, is het uitgangspunt om alleen

gepseudonimiseerde gegevens te verstrekken. Momenteel kunnen we geen conclusie trekken over of dit uitgangspunt wordt nageleefd, het monitoringsplan bevat niet voldoende details.<sup>10</sup>

### Gegevens monitoren

EconPositie\_Werkende, SLNLBPH, onderwijs\_ja, PEILMAAND, PEILMAAND, SDATUMAANVANG, SDATUMEINDE, Vnummer\_crypt, BSN\_crypt, VDLibisnr, Omschrijving\_Geslacht, Omschrijving\_Burg\_Staat, Omschrijving\_land\_van\_herkomst, Omschrijving\_moedertaal, Ingangsdatum\_Verblijfplaats, Geboortedatum, Geschat\_niveau, Diploma\_behaald, Omschrijving\_land\_vaneducatie, Startopleiding, Eindopleiding, Diploma\_bezit, Omschrijving\_land\_werkervaringopgedaan, Branche\_Omschrijving, Aanstelling, Begindatum\_werkervaring, Einddatum\_werkervaring, Nummer\_plaatsingseenheid\_PLE\_crypt, classificatie\_indicatie\_nareiziger, Groepsrol\_MVV\_indicatie\_nareiziger, Datum\_huisvestings\_gesprek, Opvangvorm, Gekoppeld Aan Gemeente, Dbeslotenatum koppeling aan gemeente, Indicatie\_Vergunninghouder, Soort\_Plaatsing, Dienst, Classificatie\_Hervestiger\_Vervallen, Hardecriteriaavtoepassing, geomatch\_regioadvies\_1, geomatch\_regioadvies\_eis\_1, geomatch\_regioadvies\_current\_1, geomatch\_regioadvies\_future\_1, geomatch\_regioadvies\_ps\_endorsement\_1, geomatch\_regioadvies\_2, geomatch\_regioadvies\_eis\_2, geomatch\_regioadvies\_current\_2, geomatch\_regioadvies\_future\_2, geomatch\_regioadvies\_ps\_endorsement\_2, geomatch\_regioadvies\_3, geomatch\_regioadvies\_eis\_3, geomatch\_regioadvies\_current\_3, geomatch\_regioadvies\_future\_3, geomatch\_regioadvies\_ps\_endorsement\_3, ps\_geomatch\_feedback, ple\_geomatch\_treatment, Omschrijvingarbeidsregio, Omschrijvingarbeidsregio\_eis, Verblijfplaats (ple\_sm\_location\_id), ple\_ps\_ids, Einddatum\_Verblijfplaats, Date of arrival in municipality, location\_in\_geomatch\_pilot, geomatch\_ple\_percentage, Recent Linkings, TVS Forecast, LMR for Municipality, UR for LMR, Regioadvies overgenomen, Reden koppeling, Reden niet overnemen regioadvies, GeoMatch Monitoring Export, Toelichting niet overnemen regioadvies, PLE\_aantal\_vergunninghouders, Opmerking, gemeente\_percent\_full, Referentiekenmerk, Datum\_Bekend\_Referentiekenmerk, ple\_sm\_ur (Verblijfplaats UR), ple\_ao\_ids

## 2.4 Technieken en methoden van gegevensverwerking

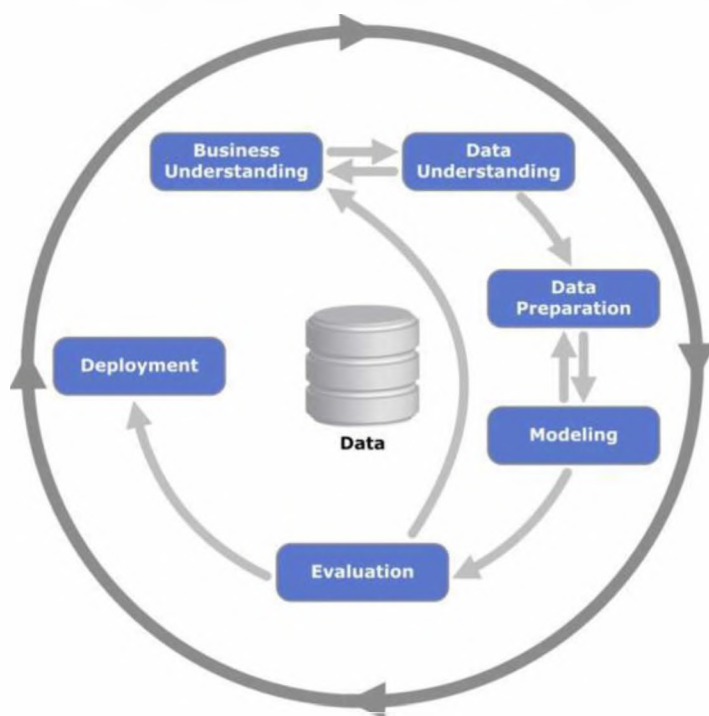
| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                  | AIIA                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Beschrijf op welke wijze en met welke (technische) middelen en methoden de persoonsgegevens worden verwerkt.</li> <li>2. Benoem of sprake is van (semi-) geautomatiseerde besluitvorming, profilering, een cloudoplossing of big dataverwerkingen en, zo ja, beschrijf waaruit dat bestaat.</li> </ol> | <ol style="list-style-type: none"> <li>1. Beschrijf welke AI-technologie wordt ingezet om het doel te bereiken.</li> </ol> |

Uit het project initiation-document worden drie verschillende technieken genoemd die worden toegepast in het kader van de AI-toepassing:

- **Data-analyse:** geavanceerde data-analysetechnieken zullen worden gebruikt om de data te verzamelen en te analyseren, waaronder de vaardigheden en voorkeuren van statushouders, en de behoeften en kansen van arbeidsmarktregio's.

<sup>10</sup> GeoMatch Netherlands : Pilot Monitoring Plan, IPL, februari 2024

- **Machine learning:** machine learning-algoritmen zullen worden geïmplementeerd om patronen en trends in de data te identificeren en om voorspellende modellen te ontwikkelen die de koppeling van statushouders aan arbeidsmarktregio's optimaliseren.
- **Continue evaluatie:** het project zal een cyclus van continue evaluatie en verbetering volgen, waarbij het algoritme regelmatig wordt getest, geëvalueerd en verfijnd om de nauwkeurigheid, relevantie en impact te maximaliseren. We hanteren hierbij de CRISP-DM-methodiek.<sup>11</sup>



De technische onderbouwing van het algoritme en de inzet daarvan worden nader onderzocht in de Algorithmic Audit, uitgevoerd door Deloitte. De beschrijving van het gebruikte algoritme is te vinden in de technische documentatie van IPL.<sup>12</sup> Het DPIA gaat hier verder niet op in. Wel beschrijven we hieronder nog in hoeverre er sprake is van geautomatiseerde besluitvorming.

### Geautomatiseerde besluitvorming en profilering

De werking van het algoritme heeft tot gevolg dat er automatisch sprake is van profilering van vergunninghouders. Op basis van de verschillende kenmerken worden ze immers gekoppeld aan een bepaalde arbeidsmarktregio.

Het algoritme is bedoeld om aanvullend te zijn aan de werkzaamheden van de programmabegeleider en regievoerder; het algoritme is daarmee niet bedoeld als leidend in de besluitvorming van de gebruikers. In het geval van het COA lijkt het principe van 'human-in-command' het meest relevant. Dat verwijst naar de capaciteit van medewerkers om de algehele activiteit van de AI-toepassing te overzien (inclusief de bredere economische, maatschappelijke, juridische en ethische gevolgen van de AI) en het vermogen om te kiezen wanneer en hoe het systeem in iedere specifieke situatie wordt gebruikt. Daarbij kan het bijvoorbeeld gaan om de keuze om de AI-toepassing in een bepaalde situatie niet te gebruiken, om een bepaalde mate van menselijk oordeel vast te stellen tijdens het gebruik van het systeem of om te garanderen dat een door het systeem genomen beslissing kan worden herroepen.

<sup>11</sup> COA, Project Initiation Document, 23 december 2023, p. 16.

<sup>12</sup> IPL, GeoMatch Algorithm Technical Documentation, 16 februari 2024.

Organisaties mogen mensen niet onderwerpen aan geautomatiseerde besluiten die rechtsgevolgen voor hen hebben of hen 'anderszins in aanmerkelijke mate treffen'.<sup>13</sup> Of er sprake is van geautomatiseerde besluitvorming, dient te worden bepaald op basis van een aantal criteria die door het Europees Hof zijn vastgesteld:

1. Er moet een besluit worden genomen.
2. Dit besluit moet uitsluitend gebaseerd zijn op geautomatiseerde verwerking, inclusief profilering.
3. Het moet juridische gevolgen hebben voor het individu of andere, een gelijkwaardig of vergelijkbaar effect hebben in zijn impact op het individu.

In een recente zaak voor het Europees Hof van Justitie is invulling gegeven aan deze criteria.<sup>14</sup> Omdat uit deze zaak een brede opvatting van het begrip geautomatiseerde besluitvorming volgt, bestaat er een mogelijkheid dat ook de implementatie van dit project onder deze kwalificatie valt. Dit zal onder meer afhangen van de wijze waarop de pilot daadwerkelijk wordt vormgegeven. Een hoge scoringskans op het volgen van adviezen wordt door het hof gezien als een teken dat er sprake is van de geautomatiseerde besluitvorming. Hierover kan, voorafgaand aan de start van de pilot, nog geen uitsluitsel worden gegeven. In het eerder uitgevoerde DPIA is hiervoor weliswaar een uitgebreide toelichting geboden waarom hier naar mening van de schrijvers geen sprake van is. Bovendien zijn een aantal maatregelen voorgesteld om te waarborgen dat COA-medewerkers zelf een afgewogen keuze kunnen maken.<sup>15</sup> Een concrete uitwerking van deze maatregelen hebben we niet kunnen vaststellen op het moment van schrijven. Hierdoor bestaat het risico dat er toch sprake zal zijn van geautomatiseerde besluitvorming.

**NB.** Indien er sprake is van geautomatiseerde besluitvorming dan heeft dit verstrekken gevolgen voor het COA. Bij de risico's zullen we hier nader op ingaan.

## 2.5 Verwerkingsdoeleinden

| Rijksmodel DPIA                                           | AIIA                |
|-----------------------------------------------------------|---------------------|
| 1. Beschrijf de doeleinden van alle gegevensverwerkingen. | Niet van toepassing |

Het COA biedt asielzoekers leefbare en veilige opvang en begeleidt hen naar een toekomst in Nederland of daarbuiten, in samenwerking met partners in de samenleving. In de Wet COA zijn de taken van het COA vastgelegd. Een van deze taken betreft het verrichten van werkzaamheden met betrekking tot de bemiddeling bij de uitstroom van verblijfsgerechtigden als bedoeld in artikel 1, eerste lid, onderdeel g, van de Huisvestingswet 2014 naar door burgemeester en wethouders beschikbaar gestelde huisvesting (artikel 3, eerste lid, onderdeel d, Wet COA).

In het eerder uitgevoerde DPIA zien we de volgende doeleinden gespecificeerd worden:

- **Verwerkingsdoeleinde stap 1 – trainen:** het verwerkingsdoel van de voorgenomen gegevensverwerkingen betreft het ontwikkelen van de AI-toepassing ten behoeve van de dienstverlening van het COA.
- **Verwerkingsdoeleinde stap 2 – adviseren:** het verwerkingsdoel van de voorgenomen gegevensverwerking(en), en, meer specifiek, de inzet van de AI-toepassing binnen het proces 'Uitstroom vergunninghouder' betreft het uitvoeren van de wettelijke taak van het COA om werkzaamheden te verrichten met betrekking tot de bemiddeling bij de uitstroom van

<sup>13</sup> Artikel 22 AVG.

<sup>14</sup> ECLI:EU:C:2023:957 – SCHUFA Case.

<sup>15</sup> COA – Privacy Impact Assessment, p. 40.

verblijfsgerechtigden als bedoeld in artikel 1, eerste lid, onderdeel g, van de Huisvestingswet 2014 naar door burgemeester en wethouders beschikbaar gestelde huisvesting (artikel 3, eerste lid, onderdeel d, Wet COA) en in die zin te komen tot een regioadvies voor vergunninghouders.

- **Verwerkingsdoeleinde stap 3 – monitoren:** het verwerkingsdoel van de voorgenomen gegevensverwerkingen betreft het verder ontwikkelen van de AI-toepassing en het verbeteren van de dienstverlening van het COA.

## 2.6 Betrokken partijen

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                                                                                     | AIIA                                                                                                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>1. Benoem alle partijen die betrokken zijn en deel deze in per gegevensverwerking. Deel deze partijen in onder de rollen: verwerkingsverantwoordelijke, gezamenlijke verwerkingsverantwoordelijke, verwerker, sub-verwerker, verstrekker, betrokkene(n) en ontvanger.</p> <p>2. Wanneer bekend, benoem ook welke functionarissen/afdelingen binnen deze partijen toegang krijgen tot welke persoonsgegevens.</p> | <p>2.4 Beschrijf welke actoren een rol spelen in de toepassing</p> <p>4.1 Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?</p> |

Vanuit de eerder uitgevoerde PIA hebben we de volgende betrokken partijen geïdentificeerd en zij vervullen conform de AVG de volgende rollen:

| Wie                     | Omschrijving                                                                                                                                                                                                                                                                                                                                                                                   | Rol                                                  | Toegang                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Vergunninghouder</b> | De vergunninghouder is de betrokkene, de natuurlijke persoon waarvan (historische) persoonsgegevens worden verstrekt ten behoeve van wetenschappelijk onderzoek (stap 1), die is 'onderworpen' aan de AI-toepassing (stap 2) en waarvan mogelijk persoonsgegevens worden gedeeld ten behoeve van de monitoring (stap 3).                                                                       | Betrokkene                                           | N.v.t.                                                                                                                                                                                                                                                                                                                                                              |
| <b>COA</b>              | Namens het COA worden gegevens vanuit IBIS verstrekt aan het CBS door het Management Informatie Team.                                                                                                                                                                                                                                                                                          | Verwerkingsverantwoordelijke, Verstrekker            | ICT krijgt toegang tot alle persoonsgegevens die nodig zijn voor het algoritme.                                                                                                                                                                                                                                                                                     |
| <b>CBS</b>              | De taak van het CBS volgt uit artikel 3 van de CBS-wet. Die taak omvat het verrichten van statistisch onderzoek ten behoeve van praktijk, beleid en wetenschap, en het openbaar maken van de op grond van zodanig onderzoek samengestelde statistieken. De bevoegdheden voor het uitvoeren van die taak zijn - voor zover relevant - nader uitgewerkt in de artikelen 33 en 41 van de CBS-wet. | Verwerkingsverantwoordelijke, ontvanger, verstrekker | CBS heeft toegang tot de gegevens in de RA-omgeving en de trainingsdataset die vanuit IBIS wordt verstrekt. Gegevens voor monitoringlevel 5 worden aan het CBS verstrekt op dezelfde manier zoals dat bij de trainingsdataset gebeurt. Het CBS koppelt de gegevens aan hun gegevens (integratieuitkomsten), pseudonimiseert ze en stelt ze ter beschikking aan IPL. |

| Wie              | Omschrijving                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | Rol                                     | Toegang                                                                                                    |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------------------------------------------------------------------------|
| <b>Solvinity</b> | Solvinity is de vaste leverancier van clouddiensten voor het Ministerie van Justitie en Veiligheid. Wanneer het COA systemen nodig heeft, maakt het gebruik van Solvinity als cloudleverancier.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Verwerker                               | De gegevens worden on-premise gehost via Solvinity. Deze zijn goed beveiligd.                              |
| <b>IPL</b>       | IPL ontwikkelt de AI-toepassing. Het is gebruikelijk dat deze trainingsdata selecteert en voorbereidt. De ontwikkelaar adviseert over de toepassing van het algoritme en eventuele aanpassingen aan de bedrijfsvoering die daarbij horen. Naar aanleiding van de PIA-workshop in juli 2019 gaf het CBS aan dat het is toegestaan om IPL toegang te geven tot de gepseudonimiseerde dataset. Het CBS en IPL hebben met elkaar een overeenkomst gesloten over het gebruik van de gepseudonimiseerde dataset ten behoeve van het wetenschappelijke onderzoek door IPL. Overleg tussen de betrokken partijen is noodzakelijk om te bepalen onder welke voorwaarden IPL toegang mag verkrijgen tot (niet-herleidbare) gegevens, bestemd voor monitoring. Ten aanzien van het wetenschappelijke onderzoek is IPL zelfstandig verwerkingsverantwoordelijk, ook als gebruik wordt gemaakt van de RA-omgeving van CBS. | Verwerkingsverantwoordelijke, ontvanger | IPL ontvangt de gepseudonimiseerde trainingsdataset en de maandelijkse gepseudonimiseerde monitoringsdata. |

## 2.7 Belangen bij de gegevensverwerking

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                    | AIIA                                                                                                                                                                                                                                                                                                                                                                                |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Beschrijf alle belangen die de betrokken partijen hebben bij de gegevensverwerkingen.</li> <li>2. Vraag betrokkenen of hun vertegenwoordigers ook naar hun mening over de verwerking, indien relevant.</li> <li>3. Licht deze mening toe onder het belang van de betrokkenen.</li> </ol> | <ol style="list-style-type: none"> <li>2.4 Beschrijf welke actoren een rol spelen in de toepassing</li> <li>3.1 Wat zijn de baten voor de organisatie?</li> <li>3.2 Wat zijn de baten voor het individu?</li> <li>3.3 Wat zijn de baten voor de maatschappij als geheel?</li> <li>4.1 Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?</li> </ol> |

In de eerdere PIA heeft het COA de onderstaande (groepen) belanghebbenden geïdentificeerd.

### Belangen van vergunninghouders

Asielzoekers met een verblijfsvergunning gaan deel uitmaken van de Nederlandse samenleving. Vergunninghouders gelden als kwetsbare doelgroep. Daarvan is het COA zich vanzelfsprekend van nature bewust. De voorgenomen AI-toepassing kan bijdragen aan het verhogen van de slagingskansen van vergunninghouders in Nederland. De AI-toepassing en de lessen die ervan kunnen worden geleerd, zullen ervoor zorgen dat voor vergunninghouders sneller en beter inzichtelijk wordt in welke regio zij het meest kansrijk worden geacht. De mogelijk te implementeren nieuwe werkwijze kan leiden tot de volgende baten:

- Vergrote kans op een baan en daarmee de kans op een zelfredzame toekomst voor vergunninghouders.<sup>16</sup> Het hebben van een baan werkt bespoedigend voor andere integratiedoelen om de zelfredzaamheid te vergroten, zoals het spreken van de taal, het hebben van een netwerk, het aangaan van contacten met Nederlanders en het leidt ook tot vermindering van doorstroom naar andere gemeenten. De verwachting is dat vergunninghouders sneller, of langduriger aan het werk kunnen, of met een hoger inkomen, omdat ze in een regio terechtkomen die past bij de specifieke achtergrond van de vergunninghouder.
- Een andere uitkomst kan zijn dat vergunninghouders een hoger onderwijsniveau halen, met betere kansen op de arbeidsmarkt later. Dit leidt ook tot meer geluk.
- Het is in het belang van de betrokkenen dat zij beter worden geplaatst en slagen in hun nieuwe thuisland, om hen te beschermen tegen oneigenlijk gebruik van gegevens en te zorgen dat er geen ongewenst(e) onderscheid/discriminatie ontstaat of (indien dat onderscheid er in de huidige situatie al is) wordt vergroot.

Verder hebben vergunninghouders belang bij:

- een menswaardige behandeling
- transparante communicatie over de werkwijze van het algoritme, de gegevens die gebruikt worden en het doel van het algoritme
- een rechtvaardig proces en gelijke behandeling
- autonomie
- veiligheid in het kader van de privacy, emotionele veiligheid, een veilige leefomgeving, maar ook in het kader van mogelijke bescherming tegen het regime van het thuisland.

## Belangen van het COA

Het COA<sup>17</sup> heeft op basis van de businesscase voor dit project de volgende belangen:

- Het verhogen van de effectiviteit van de plaatsingen van vergunninghouders en het verbeteren van de dienstverlening naar vergunninghouders op basis van datagedreven inzichten, bijvoorbeeld door meer transparantie te bieden over het plaatsingsproces.
- Het verminderen van de werkloosheid onder statushouders, het bevorderen van duurzame integratie in de samenleving en de economische bijdrage van statushouders aan de samenleving vergroten.
- De sociale cohesie binnen Nederland verhogen.
- Het gebruikmaken van het GeoMatch-algoritme binnen de beveiligde ICT-omgeving van het COA, waardoor het risico op informatiebeveiligings- en privacy-incidenten aanzienlijk kleiner wordt. Daarbij hoort ook het belang om te voldoen aan de Baseline Informatiebeveiliging Overheid en het toepassen van een betrouwbaar algoritme.
- Het opbouwen van kennis en vaardigheden rondom datagedreven werken en het ontwikkelen van een leer- en verbetercultuur, wat mogelijk ook leidt tot andere, nieuwe innovatietrajecten. Het is in het belang van het COA dat de medewerkers de AI-toepassing goed begrijpen en de werking kunnen uitleggen. Bovendien helpen de datagedreven inzichten om de mensen om wie dit algoritme draait, beter te leren kennen en begrijpen.

---

<sup>16</sup> Het COA heeft nog geen duidelijke definitie voor wat een 'goede' integratie inhoudt, en in hoeverre het COA hierin een rol kan spelen.

<sup>17</sup> Directieteam, bestuur, stuurgroep en Projectteam Project Kansrijke Koppeling met behulp van. AI, eindgebruikers AI-toepassing.

- Het ontwikkelen en toepassen van een positief en bruikbaar algoritme. Het algoritme wordt als positief en bruikbaar gezien als er een significante stijging van 4%-punt toename arbeidsparticipatie onder statushouders is in het eerste jaar na uitstroom naar een gemeente. Ten opzichte van 12% in de huidige situatie, is dat een relatieve verbetering van 33% op arbeidsmarktparticipatie is ten opzichte van de testgroep die geen match beschikbaar krijgt.<sup>18</sup>

## Belangen van IPL

'Als het gaat om immigratiebeleid, vertrouwen mensen eerder op anekdotes en ideologie dan op bewijs. Ons doel is om hen te helpen weloverwogen beslissingen te nemen om het leven en de gemeenschappen van immigranten te verbeteren,' zo stelt Duncan Lawrence, Executive Director van IPL. IPL gebruikt grote datasets, creatieve onderzoeksmethoden en geavanceerde, analytische hulpmiddelen om urgente problemen rondom immigratiebeleid te verhelpen. Door beleidsbepalers te begeleiden en onderzoek te doen, biedt IPL oplossingen om levens van immigranten verbeteren.<sup>19</sup>

## Belangen van het CBS

Het belang van het CBS strekt zich uit tot het ontvangen en verwerken van datasets voor wetenschappelijk onderzoek. Uit CBS PIA, onderdeel 6: 'Het CBS zelf heeft geen belang bij de uitkomsten van het onderzoek anders dan dat onderzoek kwalitatief op de juiste wijze geschied. De voorgenomen verwerking is altijd noodzakelijk om de taken van het CBS te kunnen uitvoeren.'<sup>20</sup>

## Belangen van de samenleving

De AI-toepassing (en het onderliggende onderzoek)<sup>21</sup> zal inzicht opleveren op basis waarvan bestaand beleid kan worden aangepast of nieuw beleid kan worden ontworpen. Kansrijke koppeling kan positieve gevolgen hebben voor de samenleving als geheel. Bij succes zal het bijvoorbeeld kunnen zorgen voor een positiever beeld op de asielketen en het Nederlandse immigratie- en integratieproces. Een succesvollere kansrijke koppeling leidt tot meer vergunninghouders aan het werk, wat betekent dat vergunninghouders beter integreren (door actieve deelname aan de samenleving, verbeterde taalontwikkeling en opbouw netwerk) en dat de belastinginkomsten stijgen en de bijstandskosten dalen.

IPL heeft de significante potentiële voordelen in kaart gebracht van het gebruik van het Geomatch-algoritme voor de Nederlandse situatie. De baten voor de maatschappij als geheel zijn als volgt:

- Het aandeel personen (doelpopulatie volwassen statushouders) dat in staat was werk te vinden in het eerste jaar na aankomst in een gemeente zou toenemen met 4,5 procent. Dit komt neer op een stijging van 37% boven het huidige percentage (de basislijn) van 12,0%. Als het huisvestingsproces niet gehinderd zou worden door de beperkingen van de va-procedure (als een vergunninghouder vanuit een azc buiten zijn uitplaatsingsregio geplaatst zou kunnen worden) zou naar verwachting een stijging van 66% kunnen worden behaald (8,0% boven de basislijn).
- Als het advies van de GeoMatch-tool wordt overgenomen is de kans groter dat een statushouder ook op de langere termijn positieve effecten ervaart. De impactevaluatie suggereert dat statushouders in het eerste jaar 10% en het tweede jaar 13% meer loon hebben verdiend, ten opzichte van de huidige resultaten.

<sup>18</sup> Dit percentage is afkomstig uit het vooronderzoek van IPL van mei 2022 en dit is de basis voor de business case: GeoMatch Netherlands Updated Backtests, IPL, 5 mei 2022

<sup>19</sup> Zie mission IPL via <https://immigrationlab.org/about-us/>.

<sup>20</sup> CBS, Standaard PIA, 24 maart 2021, te raadplegen via [deze link](#).

<sup>21</sup> Rapport op aanvraag beschikbaar: GeoMatch Netherlands, Retrospective Impact Evaluation and Feasibility Study for the Central Agency for the Reception of Asylum Seekers (COA) Kirk Bansak, Jennifer Fei, Nicholas Adams-Cohen, Jeremy Ferwerda, Dominik Hangartner and Jens Hainmueller.

- Een rekensom met voorzichtige aannames laat zien dat het gebruik van het algoritme € 5,3 miljoen per jaar kan opleveren (vanaf het tweede jaar). Over tien jaar tijd is de verwachte maatschappelijke opbrengst/besparing dan € 47,9 miljoen. Dit bestaat uit besparing op uitkeringen en opbrengsten uit inkomstenbelasting. Er is daarnaast een directe winst voor vergunninghouders met een baan (totale loon, gerekend over de gehele groep: € 3,5 miljoen per jaar).
- Een minder conservatieve rekensom laat zien dat het algoritme een nog veel hoger bedrag kan opleveren. Daarnaast is niet meegenomen dat vergunninghouders die in jaar twee een baan hebben waarschijnlijk ook in jaar drie, vier, vijf, etc. een baan hebben. De opbrengsten hiervan zijn niet meegenomen in de bepaling van het bedrag van € 47,9 miljoen.
- Thema's als AI en algoritmes staan enorm in de belangstelling bij zowel burgers als de rijksoverheid en de verwachtingen erover zijn hooggespannen. AI is volgens de Nederlandse regering een sleuteltechnologie. Vrijwel elk departement ontwikkelt toepassingen of zet deze al in.

## Belangen van overige relevante partijen

Het COA is voornemens om dit algoritme te registreren in het landelijk algoritmeregister. Op deze wijze kan iedereen lezen over de toepassing en werking van het algoritme.

| Partij                                                           | Belang                                                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Gemeenten (VNG)                                                  | De betrokkenen worden geplaatst in een gemeente binnen een uitplaatsingsregio. Hoewel de gemeenten geen primaire belanghebbenden zijn, is redelijkerwijs te verwachten dat de AI-toepassing wel gevolgen heeft voor hen. Gemeenten moeten vergunninghouders passende woonruimte aanbieden, zoals vastgesteld in de taakstelling. Gemeenten hebben vanzelfsprekend baat bij een succesvollere kansrijke koppeling. |
| Ministerie van Justitie en Veiligheid en de 'Vreemdelingenketen' | Het Ministerie van Justitie en Veiligheid (JenV) draagt verantwoordelijkheid voor de vreemdelingenketen. Het COA is een zelfstandig bestuursorgaan (zbo). Het bestuur is verantwoordelijk voor de dagelijkse leiding van het COA. JenV is opdrachtgever én toezichthouder. De politieke verantwoordelijkheid ligt bij de staatssecretaris van JenV.                                                               |
| Ministerie van Sociale Zaken en Werkgelegenheid                  | Het Ministerie van Sociale Zaken en Werkgelegenheid (SZW) is onder andere verantwoordelijk voor arbeidsmarktbeleid en uitkeringen. Ook was SZW betrokken bij de totstandkoming van screening en matching, en ondersteunen ze (financieel) de werkzaamheden op het gebied van integratie van vergunninghouders binnen het COA, waaronder de kansrijke koppeling.                                                   |
| Ministerie van Binnenlandse Zaken                                | Het Ministerie van Binnenlandse Zaken (BZK) is verantwoordelijk voor de vaststelling van de taakstelling huisvesting vergunninghouders en de verdere ontwikkeling van digitalisering bij de overheid. Het is denkbaar dat dit project van het COA voor BZK leidt tot nieuwe, verbeterde inzichten voor slimme digitale oplossingen bij overheden.                                                                 |

| Partij                                             | Belang                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Uitvoeringsinstituut Werknemersverzekeringen (UWV) | Het UWV is verantwoordelijk voor het uitvoeren van de werknemersverzekeringen WW, WAO, WIA, de Ziektewet en verder de Wajong, de Toeslagenwet en de WAZ. De gegevens over werkgelegenheid – die het CBS koppelt aan de gegevens van het COA – zijn afkomstig van het UWV. Met UWV is contact gezocht om te onderzoeken hoe de manier van monitoring (tijdens de pilot) verder vormgegeven kan worden.                                                      |
| Autoriteit Persoonsgegevens (AP)                   | De Autoriteit Persoonsgegevens (AP) is de Nederlandse gegevensbeschermingsautoriteit en het zelfstandig bestuursorgaan dat in Nederland bij wet als toezichthouder is aangesteld voor het toezicht op het verwerken van persoonsgegevens. Zoals ook beschreven in het meerjarig visiedocument Focus AP 2020-2023: 'Dataprotectie in een digitale samenleving' zijn AI en algoritmes een van de drie focusgebieden in het toezicht van de AP. <sup>22</sup> |
| Overige belangenorganisaties (niet uitputtend)     |                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| VluchtelingenWerk Nederland                        | De vereniging VluchtelingenWerk Nederland is een niet-gouvernementele organisatie die de belangen van vluchtelingen en asielzoekers in Nederland behartigt.                                                                                                                                                                                                                                                                                                |
| Privacy First                                      | Privacy First is een onafhankelijke stichting met als doel het behoud en de bevordering van het recht op privacy.                                                                                                                                                                                                                                                                                                                                          |
| Bits of Freedom                                    | Stichting Bits of Freedom is een onafhankelijke Nederlandse stichting die opkomt voor digitale burgerrechten. Bits of Freedom verdedigt het grondrecht op privacy en het grondrecht op communicatievrijheid van de Nederlandse burger.                                                                                                                                                                                                                     |
| Amnesty International                              | Amnesty International is een vereniging die de naleving van de mensenrechten beoogt zoals die zijn vastgelegd in de Universele Verklaring van de Rechten van de Mens en andere internationale mensenrechtendocumenten. Op 26 oktober 2021 publiceerde Amnesty International het rapport 'Xenofobe machines, discriminatie door ongereguleerd gebruik van algoritmen in het Nederlandse toeslagenschandaal'. <sup>23</sup>                                  |

## 2.8 Verwerkingslocaties

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                           | AIIA                                                                                                                                                                          |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Benoem in welke landen de gegevensverwerkingen plaatsvinden.</li> <li>2. Beschrijf het doorgiftemechanisme dat van toepassing is wanneer verwerkingslocaties zich buiten de Europese Economische Ruimte bevinden</li> <li>3. Noem welke aanvullende maatregelen van toepassing zijn.</li> </ol> | <p>4.3 Welke waarden en belangen spelen een rol in de context van mijn toepassing van AI?</p> <p>5.2 Welke maatregelen zijn genomen om de veiligheid van de AI te borgen?</p> |

<sup>22</sup> Autoriteit Persoonsgegevens, *Focus AP 2020-2023: 'Dataprotectie in een digitale samenleving'*, 11 november 2019, te raadplegen via: <https://www.autoriteitpersoonsgegevens.nl/nl/over-de-autoriteit-persoonsgegevens/focus-ap-2020-2023>.

<sup>23</sup> Amnesty International, *Xenofobe machines, discriminatie door ongereguleerd gebruik van algoritmen in het Nederlandse toeslagenschandaal*, 26 oktober 2021, te raadplegen via: [https://www.amnesty.nl/content/uploads/2021/10/NL\\_Xenophobe-Machines\\_Amnesty-International\\_EUR-35\\_4686\\_2021.pdf?x42580](https://www.amnesty.nl/content/uploads/2021/10/NL_Xenophobe-Machines_Amnesty-International_EUR-35_4686_2021.pdf?x42580).

De verwerkingsactiviteiten binnen dit project Kansrijke Koppeling met behulp van AI bevinden in beginsel binnen de Europese Economische Ruimte (EER) plaats. Meer specifiek zullen de verwerkingsactiviteiten worden uitgevoerd in Nederland binnen de systemen van het COA en CBS. Het trainen van de AI-oplossing vindt onder meer plaats door medewerkers van IPL. Hoewel het IPL een samenwerking betreft tussen de universiteit van Zurich in Zwitserland en Stanford University in de Verenigde Staten, zullen verwerkingsactiviteiten enkel plaatsvinden in Zwitserland.

Vanwege de verwerking van persoonsgegevens in Zwitserland is er sprake van doorgifte van persoonsgegevens buiten de EU en is een beroep op een doorgiftemechanisme noodzakelijk. Voor Zwitserland kan hiervoor een beroep worden gedaan op een adequaatheidsbesluit.

Omdat we deze werkwijze niet volledig kunnen valideren hebben we hierover een risico opgenomen in Deel C van het DPIA.

## 2.9 Juridisch en beleidsmatig kader

| Rijksmodel DPIA                                                                                                                                                | AIIA                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| 1. <i>Benoem alle wet- en regelgeving en beleid met mogelijke gevolgen voor de gegevensverwerkingen. De AVG en de Richtlijn hoeven niet genoemd te worden.</i> | <i>Bijlage 2 grondrechten IAMA</i> |

Het Centraal Orgaan opvang asielzoekers (Opdrachtgever), hierna: 'COA' heeft als taak om de materiële en immateriële opvang van asielzoekers te verzorgen, alsmede het plaatsen van asielzoekers op gemeentelijke opvangplaatsen en het verrichten van werkzaamheden met betrekking tot de bemiddeling bij de uitstroom van verblijfsgerechtigden, zoals bedoeld in artikel 1, eerste lid, onderdeel g, van de Huisvestingswet 2014 naar door burgemeester en wethouders beschikbaar gestelde huisvesting (zie artikel 3 van de Wet COA). Het gaat hier om een wettelijke taak. In het verlengde van deze wettelijke taak, is het in belang van het COA om de koppeling van vergunninghouders aan gemeenten te verbeteren.

De taak van het CBS volgt uit artikel 3 van de CBS-wet. Die taak omvat het verrichten van statistisch onderzoek ten behoeve van praktijk, beleid en wetenschap, en het openbaar maken van de op grond van zodanig onderzoek samengestelde statistieken. De bevoegdheden voor het uitvoeren van die taak zijn - voor zover relevant - nader uitgewerkt in de artikelen 33 en 41 van de CBS-wet. Op grond van artikel 33 van de CBS-wet is het CBS bevoegd om registraties die worden bijgehouden in verband met de uitvoering van een wettelijke taak te verwerven ten behoeve onderzoeksdoeleinden. Artikel 41 van de CBS-wet bepaalt dat ten behoeve van statistisch en wetenschappelijk onderzoek, op verzoek, gegevens kunnen worden verstrekt, dan wel daartoe toegang kan worden verleend, mits passende maatregelen worden genomen om herkenning van afzonderlijke personen, huishoudens, ondernemingen of instellingen te voorkomen.

Daarnaast zijn ook andere algemene wettelijke kaders relevant. Met name de volgende wetten en verdragen zijn in dit verband relevant:

- Europees Verdrag voor de Rechten van de Mens (EVRM).
- Algemene wet gelijke behandeling (Awgb).
- Voor overheidsorganisaties gelden regels voor het maken van onderscheid. Het maken van onderscheid is enkel toegestaan als het onderscheid objectief kan worden gerechtvaardigd en proportioneel is.<sup>24</sup>

<sup>24</sup> Reactie op het verslag van het Schriftelijk Overleg over de brief over de uitvoering van de motie-Marijnissen (TK 26643 nr 751 / TK 35510 nr 21), 13 juli 2021, p. 6, te raadplegen via: <https://www.rijksoverheid.nl/documenten/kamerstukken/2021/07/13/beantwoording-kamervragen-over-voorstel-opruimen-vervuilde-data-binnen-overheidsinstellingen>.

## 2.10 Bewaartermijnen

In de eerdere PIA is ten aanzien van bewaartermijnen enkel opgenomen dat voor het onderdeel adviseren de bewaartermijnen uit het [Basis selectiedocument van het COA](#) worden gehanteerd. Echter, ook voor verwerkingen in het kader van trainen en monitoren zullen bewaartermijnen moeten worden vastgesteld en worden afgedwongen. De uitwerking hiervan hebben we niet kunnen valideren. De bewaartermijnen voor gegevens in de bewonersdossiers is vastgesteld op twintig jaar (handeling 18 van het Basis selectiedocument).

Bij het vaststellen van deze bewaartermijnen zal het COA rekening dienen te houden met het beginsel van gegevensbeperking. Gegevens zullen na het verstrijken van de bewaartermijnen vernietigd moeten worden.

Ten aanzien van het verwijderen zullen hierover ook afspraken gemaakt dienen te worden met IPL voor zover zij als verwerker of verwerkingsverantwoordelijke kwalificeren. Deze afspraken zijn weliswaar in theorie opgenomen in de overeenkomst, maar we hebben niet kunnen valideren of, en op welke wijze, dit in de praktijk plaatsvindt. Hierdoor ontstaan enkele risico's die we in hoofdstuk 5 nader zullen toelichten.



### HOOFDSTUK 3

# Deel B: Beoordeling rechtmatigheid gegevensverwerkingen

Beoordeel de rechtsgrond, noodzaak en doelbinding van de gegevensverwerkingen en rechten van de betrokkene.

### 3.1 Rechtsgrond

| Rijksmodel DPIA                                                                   | AIIA                                                                             |
|-----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| 1. <i>Bepaal op welke rechtsgronden de gegevensverwerkingen worden gebaseerd.</i> | 4.2 <i>Zijn deze waarden en belangen geconcretiseerd in wet- en regelgeving?</i> |

**NB.** De rechtsgronden in dit hoofdstuk zijn overgenomen uit de eerder uitgevoerde PIA. De reden hiervoor is dat er in beginsel geen redenen zijn om aan te nemen dat deze rechtsgronden veranderd zijn.

In deze paragraaf is het relevant om de voorgenomen gegevensverwerkingen te verdelen in de volgende stappen:

1. trainen
2. adviseren
3. monitoren.

#### 1. Trainen

Aangenomen wordt dat AI-ontwikkelaars, -onderzoekers en zelfs -gebruikers zich in bepaalde omstandigheden bezighouden met wat onder de AVG geldt als 'wetenschappelijk onderzoek'. Het trainen van een AI-toepassing tijdens de ontwikkelingsfase kan bijvoorbeeld vallen onder de technologische ontwikkeling van het systeem. Ook fundamenteel AI-onderzoek, ongeacht of dit privaat of publiek wordt gefinancierd, kan onder dit uitzonderingsregime vallen. Het trainen van een AI-toepassing kan worden beschouwd als technologische ontwikkeling.<sup>25</sup> Derhalve zal ten behoeve van 'stap 1' gebruik worden gemaakt van de ruimte die Overweging 50 bij de AVG biedt.

Het COA beschikt over een set met gegevens over vergunninghouders. Deze gegevensset is vastgelegd in IBIS, het hoofdsysteem van het COA. Als wettelijke grondslag geldt dat het COA-persoonsgegevens van asielzoekers in het kader van zijn taken mag verwerken op grond van de vervulling van een taak van algemeen belang of van een taak in het kader van de uitoefening van openbaar gezag.

De voorgenomen verstrekking is ten behoeve van wetenschappelijke en/of statistische doeleinden. Wetenschappelijk en/of statistisch onderzoek is verwant aan de wettelijke taak van het COA die is neergelegd in artikel 3 van de Wet COA, samengevat: de materiële en immateriële opvang van asielzoekers.

De persoonsgegevens die het COA zal gebruiken voor de ontwikkeling van de AI-toepassing zijn al op een eerder moment verzameld ten behoeve van het vervullen van zijn wettelijke taak. Uit het beginsel van doelbinding, zoals opgenomen in artikel 5, eerste lid, onderdeel b, van de AVG volgt dat die persoonsgegevens niet verder mogen worden verwerkt op een met het oorspronkelijke verwerkingsdoel onverenigbare wijze.

<sup>25</sup> T. Gils, E. Wauters, B. Bénichou, J. De Bruyne en P. Valcke, *Artificiële Intelligentie en gegevensbescherming: een verkennende gids*, Kenniscentrum Data & Maatschappij, Brussel, België, mei 2020, par. 4.5A, p. 67, te raadplegen via: [https://data-en-maatschappij.ai/uploads/publications/20200602\\_Rapport-AI-GDPR\\_aug2020.pdf](https://data-en-maatschappij.ai/uploads/publications/20200602_Rapport-AI-GDPR_aug2020.pdf).

De eerste vraag die rijst, is dus of de gegevens verder mogen worden verwerkt voor een ander doeleinde, namelijk ten behoeve van wetenschappelijke en/of statistische doeleinden. Dat vereist een beoordeling van de verenigbaarheid met het originele verwerkingsdoel. Echter, in dit geval zal er sprake zijn van een rechtsvermoeden van doelverenigbaarheid. Op grond van artikel 5, eerste lid, onderdeel b, van de AVG wordt de verwerking van persoonsgegevens voor wetenschappelijk onderzoek immers verenigbaar geacht met het oorspronkelijke verwerkingsdoel. Het beginsel van doelbinding staat daardoor niet in de weg aan het secundaire gebruik van de gegevens over vergunninghouders: deze mogen worden verwerkt om wetenschappelijk onderzoek uit te voeren. Dit betreft een tamelijk breed uitgangspunt dat ook van toepassing is bij wetenschappelijk onderzoek op ander terrein dan het oorspronkelijke verzameldoel.

De tweede vraag betreft de grondslag voor het delen van gegevens met CBS ten behoeve van wetenschappelijk onderzoek. Uit Overweging 50 bij de AVG volgt dat als er sprake is van een verdere verwerking van persoonsgegevens die verenigbaar is met het oorspronkelijke verwerkingsdoel, er geen afzonderlijke grondslag is vereist voor de secundaire verwerking van persoonsgegevens. Dat zal aan de orde zijn bij het secundaire gebruik (de verstrekking) van de gegevensset ten behoeve van wetenschappelijk onderzoek.

Artikel 89, eerste lid, van de AVG vereist dat er passende technische en organisatorische maatregelen, zoals onder meer bedoeld in artikel 5, eerste lid, onderdeel f, van de AVG, moeten worden getroffen voor de verwerking van persoonsgegevens ten behoeve van onderzoeksdoeleinden, onder meer om het beginsel van dataminimalisatie van artikel 5, eerste lid, onderdeel c, van de AVG te garanderen. De onderzoekers van IPL zullen gepseudonimiseerde persoonsgegevens gebruiken. Door een specialistische afdeling van CBS worden de persoonsdatabestanden gepseudonimiseerd en 'verRINd'.

Ten aanzien van het beheer van de RA-omgeving door CBS, bestaande uit de verwerving, versleuteling, en verstrekking van (micro)data, kwalificeert het CBS als zelfstandig verwerkingsverantwoordelijk, en heeft daarmee zelfstandig zeggenschap over verwerving, versleuteling, en verstrekking van de betreffende (micro)data. Ten aanzien van het wetenschappelijke onderzoek is IPL zelfstandig verwerkingsverantwoordelijk, ook als gebruik wordt gemaakt van de RA-omgeving van het CBS.

## 2. Adviseren

Artikel 5, eerste lid, aanhef en onderdeel a, van de AVG bepaalt dat persoonsgegevens moeten worden verwerkt op een wijze die ten aanzien van de betrokkene rechtmatig, behoorlijk en transparant is.

Het COA is, aldus de algemene missie, de organisatie in de vreemdelingenketen die asielzoekers leefbare en veilige opvang biedt en hen begeleidt naar een toekomst in Nederland of daarbuiten. Op grond van artikel 3, eerste lid, onderdeel d, Wet COA is het COA belast met 'werkzaamheden met betrekking tot de bemiddeling bij de uitstroom van verblijfsgerechtigden als bedoeld in artikel 1, eerste lid, onderdeel g, van de Huisvestingswet 2014 naar door burgemeester en wethouders beschikbaar gestelde huisvesting.' Dat maakt dat de gegevensverwerkingen door het COA kan worden gebaseerd op de rechtsgrond 'de vervulling van een taak in het kader van de uitoefening van het openbaar gezag die aan de verwerkingsverantwoordelijke is opgedragen middels Unierecht of het lidstatelijke recht', zoals bepaald in artikel 6, eerste lid, onderdeel e, van de AVG. Het betreft hier een verwerking van persoonsgegevens ten behoeve van het primaire proces van het COA.

Artikel 6, derde lid, van de AVG bepaalt dat de rechtsgrond voor de in het eerste lid, onderdeel e, bedoelde verwerking moet worden vastgesteld bij Unierecht of lidstatelijk recht dat op de verwerkingsverantwoordelijke van toepassing is. Het doel van de verwerking wordt in die rechtsgrond vastgesteld of is met betrekking tot de in het eerste lid, onderdeel e, bedoelde verwerking noodzakelijk voor de vervulling van een taak van algemeen belang of voor de uitoefening van het openbaar gezag dat aan de verwerkingsverantwoordelijke is opgedragen.

Het is niet vereist dat deze taak is vastgelegd in een wet in formele zin. De regeling waarin de taak is vastgelegd moet voldoende duidelijk en nauwkeurig zijn, zodat de toepassing ervan voorspelbaar is. Het is daarnaast ook niet vereist dat voor elke afzonderlijke gegevensverwerking specifieke wetgeving aanwezig is. Dit volgt uit de Overwegingen 41 en 45 bij de AVG. In Bijlage 1 bij de Kamerbrief over uitkomsten externe toetsing verstrekking persoonsgegevens van COA aan politie (29 januari 2021) is opgenomen dat 'het COA op basis van een taak van algemeen belang (artikel 6, eerste lid, onderdeel e, van de AVG) een gerechtvaardigde grondslag heeft voor de verwerking van persoonsgegevens'.<sup>26</sup>

De grondslag van artikel 6, eerste lid, onderdeel e, van de AVG vormt geen vrijbrief om alle denkbare persoonsgegevens ongelimiteerd te verwerken. Op grond van de AVG is het verplicht te kunnen aantonen waarom het gebruik van persoonsgegevens noodzakelijk en evenredig is. Dat doet het COA in de paragrafen 2.2 en 3.4 van dit DPIA. Deze onderbouwing is ook van belang vanwege het transparantiebeginsel.

### 3. Monitoren

Het uitgangspunt is dat er slechts geaggregeerde gegevens worden verstrekt aan IPL. Indien daarmee de kwaliteit van de monitoring niet toereikend is, is het uitgangspunt om alleen gepseudonimiseerde gegevens te verstrekken.

Als er persoonsgegevens worden verwerkt ten behoeve van 'stap 3 – monitoring', is het aannemelijk dat deze gegevensverwerking wordt gebaseerd op artikel 6, eerste lid, onderdeel e, van de AVG of op de verenigbaarheidstoets van artikel 6, vierde lid, van de AVG. De betrokkene dient er op voorhand van te worden geïnformeerd dat persoonsgegevens zullen worden verwerkt ten behoeve van dit secundaire verwerkingsdoel.

---

<sup>26</sup> Bijlage 1 bij de Kamerbrief over uitkomsten externe toetsing verstrekking persoonsgegevens van COA aan politie, 29 januari 2021, te raadplegen via: <https://www.rijksoverheid.nl/documenten/rapporten/2021/01/29/tk-bijlage-1-uitkomsten-externe-toetsing-verstrekking-persoonsgegevens-van-coa-aan-politie>.

## 3.2 Bijzondere persoonsgegevens

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                                                                  | AIIA |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| <ol style="list-style-type: none"> <li>1. <i>Als bijzondere of strafrechtelijke persoonsgegevens worden verwerkt, beoordeel of een van de wettelijke uitzonderingen op het verwerkingsverbod van toepassing is of, in geval van nieuwe regelgeving, er zou moeten komen.</i></li> <li>2. <i>Bij verwerking van een wettelijk identificatienummer, beoordeel of dit is toegestaan.</i></li> </ol> |      |

### Bijzondere persoonsgegevens (artikel 9 AVG)

Op grond van artikel 9, eerste lid, van de AVG geldt een verwerkingsverbod voor persoonsgegevens waaruit ras of etnische afkomst blijken. De termen ras of etnische afkomst worden niet gedefinieerd in de AVG. Volgens overweging 51 van de AVG gaat het bij de voornoemde termen in ieder geval om persoonsgegevens die door hun aard bijzonder gevoelig zijn wat betreft de grondrechten en fundamentele vrijheden. Deze gegevens verdienen specifieke bescherming aangezien de context van de verwerking ervan significante risico's met zich kan meebrengen.

Het ras of de etnische afkomst blijkt uit een persoonsgegeven wanneer een rechtstreeks verband bestaat tussen het persoonsgegeven en het gevoelige kenmerk. Ook gegevens die indirect dergelijke informatie onthullen, worden aangemerkt als bijzondere categorieën persoonsgegevens. Gegevens die hooguit een indicatie geven dat het om gevoelige persoonsgegevens zou kunnen gaan, zijn niet voldoende om te spreken over een rechtstreeks verband op grond waaraan het verwerkingsverbod geldt.

Om te beoordelen of er sprake is van bijzondere persoonsgegevens achten we het van belang om de gegevens in samenhang te bekijken en niet afzonderlijk. Door gegevens te verwerken over het land van afkomst, taal, (land van ) opleiding en achtergrond worden er veel culturele eigenschappen verwerkt die iets over de etniciteit van het individu vertellen. Daarmee lijkt het onvermijdelijk om te concluderen dat er (in samenhang) bijzondere persoonsgegevens worden verwerkt van de vergunninghouders.

Concluderend is er sprake van de verwerking van bijzondere persoonsgegevens over etniciteit in de zin van artikel 9, eerste lid van de AVG. Voor de verwerking van deze categorieën persoonsgegevens geldt een verbod tenzij er sprake is van een uitzonderingsgrond zoals vermeld in artikel 9, tweede lid van de AVG. Hiervoor hebben wij nog geen onderbouwing kunnen vaststellen.

### Wettelijk identificatienummer

Voor de overheid is het gebruik van het Burgerservicenummer (BSN) geregeld in artikel 10 van de Wet algemene bepalingen Burgerservicenummer (Wabb). Het COA kan bij het verwerken van persoonsgegevens in het kader van de uitvoering van zijn publieke taak gebruikmaken van het BSN, zonder dat daarvoor nadere regelgeving is vereist.

Omdat het vreemdelingennummer slechts wordt gebruikt door ketenpartners in hun onderlinge informatie-uitwisseling, is het op grond van de Wabb niet verplicht om het gebruik van dit nummer van een wettelijke basis te voorzien. (Zie MvT bij de Wijziging van de Vw in verband met de uitbreiding van het gebruik van biometrische kenmerken in de vreemdelingenketen in verband met het verbeteren van de identiteitsvaststelling van de vreemdeling, Kamerstuk 33192, nr. 3.)

### 3.3 Doelbinding

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                                       | AIIA                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|
| 1. <i>Als de persoonsgegevens voor een ander doeleinde worden verwerkt dan het doeleinde waarvoor de persoonsgegevens oorspronkelijk zijn verzameld, beoordeel of deze (nieuwe) verdere verwerking toelaatbaar is op grond van Unie- of lidstaatrechtelijk recht, dan wel verenigbaar is met het doel waarvoor de persoonsgegevens oorspronkelijk zijn verzameld.</i> | Niet van toepassing |

**NB.** De informatie in deze paragraaf blijft ongewijzigd ten aanzien van een eerdere versie van het DPIA en is daardoor letterlijk overgenomen.

Ook voor doelbinding wordt onderscheid gemaakt tussen de verschillende voorgenomen verwerkingen:

1. trainen
2. adviseren
3. monitoren.

#### 1. Trainen

In paragraaf 3.1 van dit document is uitgewerkt dat de persoonsgegevens die het COA zal gebruiken voor de ontwikkeling van de AI-toepassing al op een eerder moment zijn verzameld voor een ander doel.

In dit geval zal er sprake zijn van een rechtsvermoeden van doelverenigbaarheid. Op grond van artikel 5, eerste lid, onderdeel b, van de AVG wordt de verwerking van persoonsgegevens voor wetenschappelijk onderzoek immers verenigbaar geacht met oorspronkelijke verwerkingsdoelen. Het beginsel van doelbinding staat daardoor niet in de weg aan het secundaire gebruik van de gegevens over vergunninghouders: deze mogen worden verwerkt om wetenschappelijk onderzoek uit te voeren. Dit betreft een tamelijk breed uitgangspunt dat ook van toepassing is bij wetenschappelijk onderzoek op ander terrein dan het oorspronkelijke verzameldoel. Uit Overweging 50 bij de AVG volgt dat als er sprake is van een verdere verwerking van persoonsgegevens die verenigbaar is met het oorspronkelijke verwerkingsdoel, er geen afzonderlijke grondslag is vereist voor de secundaire verwerking van persoonsgegevens. Dit maakt dat wordt voldaan aan het beginsel van doelbinding.

#### 2. Adviseren

Tijdens de gebruiksfase worden de persoonsgegevens duidelijk verzameld voor het verwerkingsdoel en verwerkt ten behoeve van dat verwerkingsdoel.

#### 3. Monitoren

Het uitgangspunt is dat er slechts geaggregeerde gegevens worden verstrekt aan IPL. Indien daarmee de kwaliteit van de monitoring niet toereikend is, is het uitgangspunt om alleen gepseudonimiseerde gegevens te verstrekken.

Zoals uiteengezet in paragraaf 3.1 van dit document, is het aannemelijk dat – als er persoonsgegevens worden verwerkt ten behoeve van 'stap 3 – monitoring' – het aannemelijk is dat deze gegevensverwerking wordt gebaseerd op artikel 6, eerste lid, onderdeel e, van de AVG of op de verenigbaarheidstoets van artikel 6, vierde lid, van de AVG. De betrokkene dient op voorhand te worden geïnformeerd dat persoonsgegevens zullen worden verwerkt ten behoeve van dit secundaire verwerkingsdoel.

Als artikel 6, eerste lid, onderdeel e, van de AVG geldt als grondslag, zullen de persoonsgegevens duidelijk worden verwerkt voor dit verwerkingsdoel en worden verwerkt ten behoeve van dat verwerkingsdoel. Als artikel 6, vierde lid, van de AVG geldt als grondslag, is aannemelijk dat onderbouwd zal moeten worden dat er sprake is van verenigbare doelen. Het COA zal daartoe dan de verenigbaarheidstoets uitvoeren zoals bepaald in dat lid.

### 3.4 Noodzaak en evenredigheid

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | AIIA                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><i>Beoordeel of de gegevensverwerkingen noodzakelijk zijn voor het verwezenlijken van de verwerkingsdoeleinden. Ga hierbij in ieder geval in op:</i></p> <p><i>a. Proportionaliteit: staat de inbreuk op de persoonlijke levenssfeer en de bescherming van de persoonsgegevens van de betrokkenen in evenredige verhouding tot de verwerkingsdoeleinden?</i></p> <p><i>b. Subsidiariteit: kunnen de verwerkingsdoeleinden in redelijkheid niet op een andere, voor de betrokkene minder nadelige wijze, worden verwezenlijkt?</i></p> | <p><i>6.1. Is de toepassing proportioneel?</i></p> <p><i>6.2. Is met minder ingrijpende middelen hetzelfde doel te verwezenlijken?</i></p> <p><i>6.3. Is de keuze voor de toepassing positive sum of zero sum?</i></p> <p><i>6.4. Wat zijn restrisiko's en waarom zijn deze acceptabel?</i></p> <p><i>6.5. Wordt rekening gehouden met verder gebruik (downstream responsibility)?</i></p> |

**NB.** In overleg met de opdrachtgever is afgesproken dat de noodzaak voor het verwerken van persoonsgegevens wordt beoordeeld door Deloitte in de Algorithmic Audit en dat dit buiten de scope van dit DPIA valt.

### 3.5 Rechten van de betrokkene

| Rijksmodel DPIA                                                                                                                                                                                                                                | AIIA                                                                                                                                                        |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><i>Beschrijf de procedure waarmee invulling wordt gegeven aan de rechten van de betrokkenen.</i></p> <p><i>Als de rechten van de betrokkene worden beperkt, beschrijf op grond van welke wettelijke uitzondering dat is toegestaan.</i></p> | <p><i>3.2 Wat zijn de baten voor het individu?</i></p> <p><i>4.1 Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?</i></p> |

De AVG geeft betrokkenen rechten waardoor ze controle over hun gegevens kunnen uitoefenen. Deze rechten kunnen van toepassing zijn op persoonsgegevens die worden (1) opgenomen in de trainingsgegevens, (2) worden gebruikt om te adviseren en (3) te monitoren.

(Oud-)bewoners kunnen zich jegens COA op hun AVG-rechten beroepen. Het COA heeft een procedure voor de uitvoering van de rechten van betrokkenen. In het rechten en plichtengesprek wordt een verkorte versie van de privacyverklaring beschikbaar gesteld, zoveel mogelijk in de eigen taal. Hier wordt ook in vermeld dat de betrokkene het recht heeft om een verzoek in te dienen. Opgemerkt wordt dat per geval dient te worden beoordeeld of een van de beperkingen op de rechten van betrokkenen, zoals opgenomen in artikel 23 AVG, artikel 41 Uitvoeringswet AVG en de relevante bepalingen in Hoofdstuk 2 van de AVG, van toepassing is. Het COA dient de betrokkenen actief op de hoogte te stellen van hun rechten.<sup>27</sup>

<sup>27</sup> B. van der Sloot e.a., *Handreiking Non-discriminatie by design*, opgesteld in opdracht van het Ministerie van Binnenlandse Zaken, 10 juni 2021, p. 58, te raadplegen via: <https://www.rijksoverheid.nl/documenten/rapporten/2021/06/10/handreiking-non-discriminatie-by-design>.

## Recht op informatie (artikel 12/13 AVG)

Een nadelig gevolg van de inzet van een AI-toepassing is dat het kan leiden tot beperkte uitlegbaarheid en controleerbaarheid van verwerkingen en daarop gebaseerde (menselijke) besluiten. Op grond van de AVG moet het COA in staat zijn om AI-beslissingen uit te leggen en de verwachte gevolgen voor de betrokkene (de logica van het systeem). Het COA tracht al vanaf de ontwerpfase rekening te houden met de mogelijkheid om controle op de AI-toepassing uit te voeren, bijvoorbeeld door middel van een algoritmische audit. Het COA is voornemens de richtlijnen voor het toepassen van algoritmen door overheden te nemen als uitgangspunt en aan de hand daarvan – in samenwerking met de afdeling communicatie – te komen tot een algemeen en specifiek informatiepakket.<sup>28</sup> Voorts is het COA voornemens medewerkers die zullen werken met de AI-toepassing te trainen en zal het maatschappelijke verantwoordelijkheid afleggen, bijvoorbeeld via een algoritmeregister op [coa.nl](https://coa.nl) of een Rijksbreed algoritmeregister. In de documenten die het COA tot stand brengt wordt de menselijke tussenkomst door medewerkers van het COA geëxpliciteerd. Tevens informeert het COA statushouders over de AI-toepassing door middel van een brief die in meerdere talen wordt aangeboden. Op deze wijze tracht het COA transparant te zijn over de inzet van de AI-toepassing. Verder heeft het COA:

- Doorlopend contact met IPL teneinde de voor het ontwerp en de ontwikkeling van het algoritmische systeem gebruikte methoden te documenteren. Keuzes over welke inputdata er zijn verzameld en geselecteerd en de manier waarop dat is gebeurd zijn helder.
- Het proces waarbinnen de AI-toepassing zal adviseren helder beschreven.
- Onderzoek gedaan naar de inzet van dit specifieke systeem op dit specifieke gebied.
- Het bedrijfsmodel van dit systeem onderzocht (hoe het waarde creëert voor de organisatie).
- In samenspraak met IPL besloten dat de AI-toepassing wordt ontworpen met interpreteerbaarheid in het achterhoofd. Er is aandacht voor welk model het eenvoudigst en het meest interpreteerbaar is voor de betreffende toepassing, de trainings- en testgegevens en de mogelijkheden om de interpreteerbaarheid van het model te beoordelen.

## Recht van inzage, vergetelheid en bezwaar (artikel 15/17/21 AVG)

Betrokkenen hebben op grond van de AVG het recht om controle uit te oefenen over de verwerking van hun persoonsgegevens. Bijvoorbeeld door middel van het inzagerecht, het recht op vergetelheid, en het recht om bezwaar te maken tegen de gegevensverwerking. Het is daarom raadzaam de organisatie en bedrijfsprocessen in te richten zodat het COA verzoeken van betrokkenen kan afhandelen. Daartoe is het COA verplicht. Er wordt, conform aanbeveling 9 van het Deloitte-rapport, een aanspreekpunt ingericht waar vergunninghouders, algoritmegebruikers en derden terecht kunnen met vragen, zorgen en algemene feedback.<sup>29</sup> Ook wordt er een eindverantwoordelijke aangewezen binnen het COA die verantwoordelijk is voor afstemming met IPL over het verwerken van de feedback. Beide punten zijn opgenomen op de actielijst pilotfase.

---

<sup>28</sup> Ministerie van Justitie en Veiligheid, *Richtlijnen voor het toepassen van algoritmen door overheden en publieksvoorlichting over data-analyses*, 19 mei 2021, te raadplegen via: <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/richtlijnen/2021/09/24/richtlijnen-voor-het-toepassen-van-algoritmen-door-overheden-en-publieksvoorlichting-over-data-analyses/Richtlijnen+Algoritmen+08032021.pdf>.

<sup>29</sup> Rapportage analyse Algoritme Kansrijke Koppeling, Deloitte, 19 juli 2022.

Recht van bezwaar (artikel 21) Omdat er sprake is van een verwerking voor wetenschappelijke doeleinden die is gebaseerd op taken van algemeen belang, kan de uitzondering die artikel 21, zesde lid, van de AVG van toepassing zijn die het recht van bezwaar uitsluit. Omdat het recht van bezwaar niet van toepassing is op de datasets die worden verwerkt, kan het recht op verwijdering – in casu afhankelijk van geslaagd bezwaar – ook geen toepassing vinden (artikel 17, eerste lid, onderdeel c, van de AVG). Omdat het COA op basis van een taak van algemeen belang, zoals bedoeld in artikel 6, eerste lid, onderdeel e, van de AVG een gerechtvaardigde grondslag heeft voor de verwerking van persoonsgegevens, hebben betrokkenen volgens artikel 21 AVG te allen tijde het recht om – vanwege met zijn specifieke situatie verband houdende redenen – bezwaar te maken tegen de verwerking van die persoonsgegevens van deze betrokkene betreft. Een geslaagd bezwaar zal ertoe leiden dat het COA de verwerking van persoonsgegevens staakt, tenzij hij dwingende gerechtvaardigde gronden voor de verwerking aanvoert die zwaarder wegen dan de belangen, rechten en vrijheden van de betrokkene.

Informatie over het recht op bezwaar zal uiterlijk op het moment van het eerste contact met de betrokkene onder de aandacht van de betrokkene gebracht en duidelijk en gescheiden van enige andere informatie worden weergegeven (artikel 21, vierde lid, van de AVG).

### Recht op rectificatie en aanvulling, beperking van de verwerking (artikel 16/18 AVG)

Ter waarborging van de doelstellingen van algemeen belang met betrekking tot wetenschappelijk onderzoek lijken tevens, ingevolge artikel 44 van de Uitvoeringswet AVG, de artikelen 16 en 18 (recht op inzage, rectificatie en beperking van de verwerking) niet van toepassing.

### Recht om niet onderworpen te worden aan geautomatiseerde individuele besluitvorming en profilering (artikel 22 AVG)

Het algemene verbod om niet te onderworpen worden op een uitsluitend op geautomatiseerde verwerking gebaseerd besluit van artikel 22 van de AVG kan niet worden omzeild door menselijke tussenkomst te garanderen. Als een COA-medewerker routinematig automatisch gegenereerde regioadviezen toepast op personen zonder het resultaat daadwerkelijk te beïnvloeden, is dat nog steeds een uitsluitend op geautomatiseerde verwerking gebaseerd besluit. Het behoeft daarom doorlopend aandacht, training en kunde om te waarborgen dat regioadviezen worden gebruikt als advies. Het COA is voornemens menselijke tussenkomst bij de voorgenomen AI-toepassing doorlopend te garanderen. Mogelijke waarborgen die het COA treft om menselijke tussenkomst te garanderen omvatten:

- Doorlopende gesprekken met stakeholders en gebruikers waarbij onder meer de vraag aan de orde komt hoe de betrokkene te allen tijde centraal kan staan en hoe de uitlegbaarheid van beslissingen kan worden bevorderd.
- Gebruikers krijgen voorafgaand aan de start van de pilot voorlichting en instructies over het gebruik van de AI-tool zoals dat de tool ter ondersteuning dient en in wezen niet leidt tot grote afwijkingen ten opzichte van het bestaande proces, maar met name tot objectivering.
- Gebruikers dienen bewust het advies van de AI-toepassing niet routinematig op te volgen. Het routinematig opvolgen van de AI-toepassing is simpelweg onwenselijk omdat een aantal kenmerken niet worden meegenomen door de AI-toepassing.
- Conform aanbeveling 10 van het Deloitte-rapport ontvangt de algoritmegebruiker verplichte kennis en hulpmiddelen over de reikwijdte van het advies van het algoritme, zodat de gebruiker zich bewust is van het risico op automation bias.

- Er komt een 'rode stopknop' voor de toepassing zodat het systeem kan worden stilgelegd als er fouten gemaakt (dreigen te) worden. De specificaties van deze knop worden beschreven in het monitoringplan zoals beschreven in aanbeveling 7 van het Deloitte-rapport.
- De feedbackloop zal realtime zijn, zodat de AI-toepassing direct kan worden aangepast ten behoeve van opvolgende adviezen.
- De vergunninghouder wordt geïnformeerd over de inzet van de AI-toepassing en zal de mogelijkheid krijgen om bezwaar aan te tekenen tegen het gebruik van de tool. Dan wordt het bestaande proces gevolgd.
- De menselijke tussenkomst wordt voorts gegarandeerd doordat de programmabegeleider en de regievoerder beide kunnen afwijken van het advies van de AI-toepassing. Het is in die zin een ondersteunend advies. Met name de regievoerder is in de positie om een toets te doen (vierogenprincipe). Daarbij kan deze tevens rekening houden met factoren die vaak tijdelijk zijn, trendafhankelijk zijn of samenhang hebben met prioriteiten die volgen uit de taakstelling van individuele gemeenten.



#### HOOFDSTUK 4

## **Deel C: Beschrijving en beoordeling risico's voor de betrokkenen**

Beschrijf en beoordeel de risico's van de gegevensverwerkingen voor de rechten en vrijheden van de betrokkenen. Houd hierbij rekening met de aard, omvang, context en doelen van de gegevensverwerkingen.

## 4.1 Risico's (voor de betrokkenen)

| Rijksmodel DPIA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | AIIA                                                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><i>Beschrijf en beoordeel alle mogelijke risico's van de gegevensverwerkingen voor de rechten en vrijheden van de betrokkenen, zoals het recht op privacy en het verbod op discriminatie. Ga in ieder geval in op:</i></p> <ul style="list-style-type: none"> <li>• <i>Welke negatieve gevolgen de gegevensverwerkingen kunnen hebben voor de rechten en vrijheden van de betrokkenen, zoals het verbod op discriminatie;</i></li> <li>• <i>De oorsprong van deze gevolgen;</i></li> <li>• <i>De waarschijnlijkheid (kans) dat deze gevolgen zullen intreden; en</i></li> <li>• <i>De ernst (impact) van deze gevolgen voor de betrokkenen wanneer deze intreden.</i></li> </ul> | <p><i>4.1 Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?</i></p> <p><i>6.4 Wat zijn restrisico's en waarom zijn deze acceptabel?</i></p> <p><i>Bijlage 3 Risico's en maatregelen Algemene Rekenkamer</i></p> |

We beschrijven in dit DPIA enkel de belangrijkste risico's op hoofdlijnen en lichten enkel de belangrijkste maatregelen toe, in afwijking van de eerder uitgevoerde PIA. Op deze wijze beogen we het COA inzicht te bieden in de belangrijkste risico's. Het is vervolgens aan het COA om te oordelen welke risico's acceptabel zijn, en welke risico's gemitigeerd dienen te worden.

### Project

1. We zien in verschillende documenten dat doelstellingen soms anders en dus niet uniform geformuleerd worden. Hierdoor ontstaat het risico dat de doelstellingen op basis waarvan de algoritmen worden afgesteld afwijken van de doelstellingen die zijn beoogd door het management.
2. In de projectdocumentatie wordt verschillende keren het beeld geschetst dat de data die wordt gebruikt binnen de AI-toepassing objectief is. Echter, het is belangrijk om op te merken dat data niet neutraal zijn en daarmee nooit objectief kunnen zijn, daar kan dus een belangrijk risico ontstaan. Data (en ook algoritmen) zijn altijd door mensen gemaakt of gecreëerd met een specifiek doel in gedachten. Het is belangrijk om daarvan bewust te zijn, ook met het doel om vooringenomenheid tegen te gaan.
3. Het is onbekend in hoeverre de voorgenomen acties rondom privacy zijn opgevolgd. De eerdere PIA en AIIA bevatten een hele lijst aan (voorgenomen) acties. Zo hebben wij de brief voor de vergunninghouders en de privacyverklaring nog niet kunnen inzien. Het is ook nog niet duidelijk of de informele dimensie van privacy beschermd wordt. Daarnaast is er nog geen uitleg beschikbaar gesteld via een Rijksbreed register waar in begrijpelijke taal uitleg wordt gegeven over onder andere de werking van het algoritme, welke gegevens worden gebruikt en wat het doel van de toepassing is.

## (Persoons-)gegevens

4. Om te bepalen welke locatie het meest geschikt is voor vergunninghouders worden diverse demografische gegevens van vergunninghouders gebruikt. Een deel van deze variabelen zegt onmiskenbaar iets over de etniciteit van de betrokkenen, denk bijvoorbeeld aan variabelen rondom nationaliteit en taal. Daarnaast kunnen ook andere gegevens zoals leeftijd, gender of burgerlijke staat een rol spelen bij het advies op welke locatie iemand wordt ondergebracht. Dit leidt tot hoog risico op discriminatie.<sup>30</sup> Dit kan mogelijk versterkt worden doordat er enkel data worden gebruikt van de tien landen waar de meeste statushouders vandaan komen. Het risico van *sample selection bias* ligt op de loer. Dat risico kan zich bijvoorbeeld ook uiten doordat de AI-toepassing groepen die minder kansrijk zijn in Nederland altijd koppelt aan een gemeente waar zij een kleine slagingskans hebben. Er wordt immers historische data gebruikt. Tevens kan dit ook versterkt worden door de keuzes die medewerkers maken op basis van mogelijke eigen vooringenomenheid.
5. Bewaartermijnen voor het verwerken van persoonsgegevens die zijn verwerkt ten behoeve van training en monitoring zijn niet vastgesteld en geïmplementeerd.
6. Uit de gesprekken komt naar voren dat er verschillende afspraken zijn gemaakt met IPL rondom de verwerking van persoonsgegevens, bijvoorbeeld dat persoonsgegevens enkel binnen de EU verwerkt zullen worden. Deze afspraken hebben we niet kunnen valideren. De door ons ontvangen verwerkersovereenkomsten verwijzen bijvoorbeeld naar Stanford University als contractuele partij binnen de overeenkomst. Hierdoor ontstaat bij ons een vermoeden dat doorgifte van data buiten de EU toch zal plaatsvinden, tenzij hiervoor aanvullende afspraken worden gemaakt. Dit heeft gevolgen voor de veiligheid van de betrokkene alsmede de vertrouwelijkheid van de organisatie.

## AI-toepassing

7. De uitlegbaarheid van beslissingen in individuele gevallen kan lastig zijn. Ook voor medewerkers is het niet altijd duidelijk waarom tot een bepaald besluit is gekomen. Dit brengt uitdagingen mee voor medewerkers die op basis van dat advies een besluit moeten nemen. Wanneer besluit je bijvoorbeeld om ervan af te wijken?
8. Wanneer de human-in-the-loop in de praktijk weinig tot niet afwijkt van de besluitvorming zal er waarschijnlijk sprake zijn van geautomatiseerde besluitvorming. Hiervoor dient een andere wettelijke basis te worden gezocht. Het risico bestaat dat als hier sprake van is, dat persoonsgegevens van vergunninghouders onrechtmatig worden verwerkt.
9. Waarborg dat informatiebeveiligingsmaatregelen nader uitgewerkt worden en specifiek gemaakt worden voor deze desbetreffende verwerkingsactiviteiten.
10. De overeenkomsten die we hebben ontvangen en gezien zijn vaak nog niet ondertekend of definitief. Ook ontbreekt een verwerkersovereenkomst van bijvoorbeeld Solvinity. Hierdoor kunnen we niet vaststellen of aan alle juridische verplichtingen is voldaan.

## Betrokkenen

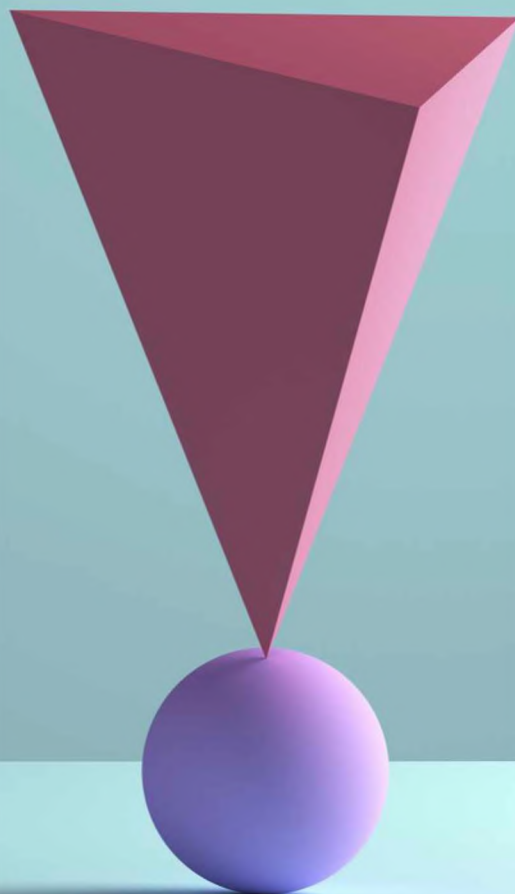
11. Er is beperkt inzicht in de belangen van de vergunninghouders doordat zij niet vanaf de start zijn betrokken bij dit project. Voor zover bij ons bekend, is er tot op heden geen ruimte geweest voor participatie vanuit deze groep, die toch de hoofdrol spelen in deze AI-toepassing.

---

<sup>30</sup> Hierbij maken we een onderscheid in directe en indirecte discriminatie. Directe discriminatie doelt op 'openlijk' onderscheid op basis van bijvoorbeeld nationaliteit. Directe discriminatie is verboden, er zijn enkele wettelijke uitzonderingen mogelijk. Indirecte discriminatie doelt op een verborgen of 'verkapt' onderscheid, formeel gezien wordt er geen onderscheid gemaakt, bijvoorbeeld op nationaliteit. Informeel kan door een combinatie van factoren toch indirect onderscheid gemaakt worden, bijvoorbeeld door een combinatie aan taal, land van herkomst en geslacht.

Daardoor is het onduidelijk wat zij zelf van dit proces vinden en kan de AI-toepassing niet aangepast worden op de input van deze groep. Het uiterste risico is dat deze groep onbedoeld gedehumaniseerd wordt doordat zij zelf geen autonomie hebben in dit project.

12. De rechten van vergunninghouders zijn potentieel onvoldoende gewaarborgd. We hebben bijvoorbeeld tegenstrijdige informatie over de mogelijkheid van vergunninghouders om bezwaar te maken tegen het gebruik van de AI-toepassing voor hun plaatsing. Zo is het onduidelijk of er een opt-out is ingebouwd in de techniek. Ook als een regievoerder of programmabegeleider het advies van de AI-toepassing niet opvolgt, worden de gegevens van de vergunninghouder verwerkt. Tevens is de informatie over de AI-toepassing niet in alle talen beschikbaar, als de vergunninghouder niet is geïnformeerd, dan kan hij/zij ook geen bezwaar maken.



## HOOFDSTUK 5

# Deel D: Geadviseerde maatregelen

Beschrijf de voorgenomen maatregelen om de hiervoor beschreven risico's van de gegevensverwerkingen voor de vrijheden en rechten van de betrokkenen aan te pakken.

## 5.1 Maatregelen

| Rijksmodel DPIA                                                                                                                                                                                                                                             | AIIA                                                                                                                                                                                                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><i>Beoordeel welke technische, organisatorische en juridische maatregelen in redelijkheid kunnen worden getroffen om de hiervoor beschreven risico's te voorkomen of te verminderen.</i></p> <p><i>Beschrijf welke maatregel welk risico aanpakt</i></p> | <p><i>5.1 Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de AI te borgen?</i></p> <p><i>5.2 Welke maatregelen zijn genomen om de veiligheid van de AI te borgen?</i></p> <p><i>5.3 Welke maatregelen zijn genomen om de transparantie van het handelen van de AI te borgen?</i></p> |

De maatregelen zijn opgesteld op basis van de fasering en levenscyclus van het algoritme, de urgente problematiek en de meest haalbare, concrete acties. We houden hierbij rekening met het feit dat risico's onmogelijk helemaal gereduceerd kunnen worden, maar stellen maatregelen vast die passen binnen de inspanning die het COA kan leveren.<sup>31</sup>

### Pilotfase project Kansrijke Koppeling

De maatregelen die tijdens deze fase genomen worden, dienen om de meest urgente risico's te mitigeren.<sup>32</sup>

1. **Datawijsheid.** Waarborg voldoende bewustzijn van de risico's en de ethische overwegingen rondom het gebruik van data bij alle stakeholders die betrokken zijn bij het project. Dat betekent onder meer ook dat stakeholders voldoende bewust moeten worden dat data in zichzelf over een bepaalde bias beschikken (data zijn nooit neutraal). Ook moet duidelijk gemaakt worden dat het feit dat iets juridisch mag, dit niet altijd ethisch wenselijk is. Laat bijvoorbeeld separaat ethische beraadslaging plaatsvinden met alle (of een selectie van) relevante stakeholders (bestuur, communicatie, projectleider, eindgebruikers, data-analisten, etc.) om samen te reflecteren over de waarden die geraakt worden met de inzet van deze AI-toepassing. Mogelijk botsen organisatiewaarden, dat wordt in zo'n proces goed blootgelegd. Dit proces helpt ook om alle onderbuikgevoelens die nog leven op tafel te krijgen en finale passende maatregelen vast te stellen. Denk hierbij aan het uitvoeren van een DEDA, een instrument dat gebruikt kan worden om ethische beraadslaging vast te leggen.
2. **Inclusiviteit.** Zorg ervoor dat er voldoende stakeholders betrokken worden die de belangen van de vergunninghouders vertegenwoordigen. Zo kan ook beter worden getoetst welke zorgen vergunninghouders hebben, in hoeverre de aannames op basis waarvan de AI-toepassing adviseert ook daadwerkelijk aansluiten bij de beleving van betrokkenen en welke verbeteringen mogelijk zijn. Als stakeholders niet betrokken kunnen worden bij dit proces, beargumenteer dan waarom zij niet betrokken dienen te worden.
3. **(Juridische) ondersteuning.** Zorg ervoor dat er voldoende ondersteuning aanwezig is in de vorm van privacy officers, experts, functionarissen gegevensbescherming en ethici om gedurende het project op verschillende vlakken te ondersteunen. Bijvoorbeeld in het inrichten van privacy gerelateerde processen, afspraken met stakeholders zoals IPL, Solvinity en het CBS en het doorlopend controleren en uitvoeren van het DPIA. Op verschillende onderdelen hebben we niet kunnen valideren of er voldoende expertise aanwezig is. Een privacy officer zou ook de

<sup>31</sup> Deze maatregelen zijn gebaseerd op de situatie van nu (8 april 2024), mogelijk worden er nog indringende besluiten gemaakt die impact hebben op deze maatregelen. Zoals de keuze om etniciteit en/of religie als indicator op te nemen in het algoritme.

<sup>32</sup> Deze maatregelen dienen genomen te worden in aanvulling op de aanbevelingen uit het Deloitte rapport.

eerder vastgestelde acties vanuit de PIA kunnen toetsen. Tevens kan de privacy officer adviseren over het aspect van dataminimalisatie.

4. **(Semi-)geautomatiseerde besluitvorming.** Overweeg om de AI-toepassing toch te kwalificeren als een vorm van geautomatiseerde besluitvorming en implementeer de daarbij behorende maatregelen. Hiervoor zullen voor het COA twee oplossingsrichtingen zijn. Ofwel er wordt gekozen om te werken met uitdrukkelijke toestemming, ofwel er dient een duidelijkere wettelijke grondslag te worden vastgelegd.

## Ingebruikname van algoritme na pilotfase

Het algoritme zal een onderdeel worden van de standaard werkwijze, wanneer de pilot als succesvol wordt bestempeld. Ook in deze fase is het belangrijk om nog een aantal maatregelen te treffen.

5. **Juridische controle.** Voer een laatste juridische check, controleer of de eerder vastgestelde kaders nog voldoende worden nageleefd. Heb hierbij ook in het bijzonder aandacht voor de ontwikkelingen rondom de AI-act.
6. **Periodiek evalueren.** We adviseren om periodiek, lees ten minste elke twee jaar, het algoritme te evalueren en te controleren of de assessments nog van toepassing zijn. Regel hier een standaard procedure voor in, voor zover dit nog niet geregeld is binnen het COA. Leg bijvoorbeeld vast om elke twee jaar het DPIA en AIIA te herhalen. De oude input van de assessments kan benut worden wanneer deze nog steeds up-to-date is. Heb bij de evaluatie vooral aandacht voor de genomen maatregelen, toets of deze nog voldoende accuraat en relevant zijn. Evalueer ook de technische werking van het algoritme. Past dit nog steeds bij de situatie en binnen de juridische kaders? Controleer ook periodiek de geschiktheid van de data, met inachtneming van de inspanning die het COA kan leveren hierop. Evalueer hoe menselijke tussenkomst effect heeft gehad op het algoritmische besluitvormingsproces.
7. **Bijdrage aan het plaatsingsproces.** Het COA heeft momenteel ontzettend veel werk aan het op orde krijgen van de huisvesting en de ruimtelijke puzzel rondom de taakstelling per gemeente. Het is nog onduidelijk hoe goed de pilot van het algoritme uitgevoerd kan worden, gezien de taakstelling vrij dominant is in het plaatsingsproces. Evalueer na de pilot, afhankelijk van of de rust bij het COA is teruggekeerd, hoe goed het algoritme bijdraagt aan de werkzaamheden van de programmabegeleider en regievoerder.

## Wezenlijke veranderingen in het algoritme gedurende levenscyclus

Het algoritme dient ook herzien te worden wanneer er wezenlijke veranderingen worden doorgevoerd, buiten de periodieke evaluatie. Denk hierbij bijvoorbeeld aan situaties waarbij er een andere dataset gebruikt gaat worden, bij een verandering in het algoritme of de methodiek, wanneer de resultaten anders gebruikt worden in het werkproces, wanneer resultaten beginnen af te wijken of wanneer er een verandering is in de wijze waarop besluiten worden gemaakt. In dit geval:

8. **Check op wettelijke kaders.** Controleer of het algoritme met deze wijzigingen nog steeds binnen het wettelijke kader past. Evalueer of het algoritme niet langer gebruikt mag worden of dat het met aanpassingen wel mogelijk is.
9. **Controleer de ethische impact.** Controleer de impact van deze wijzigingen op de (organisatie)waarden. Kijk bijvoorbeeld of de wijziging impact heeft op het mensgerichte karakter van de dienstverlening. Bepaal op basis van deze controle of er aanvullende maatregelen nodig zijn.



## BIJLAGE

# Bijlage 1. Variabelen

In deze bijlage volgt de volledige beschrijving van de variabelen die gebruikt worden voor de AI-toepassing alsmede de relevante monitoringslevels.

Tabel 1: te verwerken persoonsgegevens

| #  | Variable                       | Source               | Description                                             | Anonymization | Categorie persoonsgegevens                      |
|----|--------------------------------|----------------------|---------------------------------------------------------|---------------|-------------------------------------------------|
| 1  | EconPositie_Werkende           | Asielcohort          | Employment indicator                                    |               | Gewoon persoonsgegevens                         |
| 2  | SLNLBPH                        | Asielcohort          | Wages                                                   |               | Gewoon persoonsgegevens                         |
| 3  | Onderwijs_ja                   | Asielcohort          | Education indicator                                     |               | Gewoon persoonsgegevens                         |
| 4  | PEILMAAND                      | Asielcohort          | Year and month of measurement                           |               | Gewoon persoonsgegevens                         |
| 5  | PEILMAAND                      | SPOLIS               | Whether person was employed (excluding self-employment) |               | Gewoon persoonsgegevens                         |
| 6  | SDATUMAANVANG                  | SPOLIS               | Starting date of the period in Spolis data              |               |                                                 |
| 7  | SDATUMEINDE                    | SPOLIS               | End date of the period in Spolis data                   |               |                                                 |
| 8  | Vnummer_crypt                  | IBIS and Asielcohort | Individual identification number                        |               | Gewoon persoonsgegevens (identificerend nummer) |
| 9  | BSN_crypt                      | IBIS and Asielcohort | Individual identification number                        |               | Gewoon persoonsgegevens (identificerend nummer) |
| 10 | VDLibisnr                      | IBIS                 | IBIS individual identification number                   |               |                                                 |
| 11 | Omschrijving_Geslacht          | IBIS                 | Gender                                                  |               | Gewoon persoonsgegevens                         |
| 12 | Omschrijving_Burg_Staat        | IBIS                 | Marital Status                                          |               | Gewoon persoonsgegevens                         |
| 13 | Omschrijving_land_van_herkomst | IBIS                 | Country of Origin                                       |               | Gewoon persoonsgegevens                         |
| 14 | Omschrijving_moedertaal        | IBIS                 | Language                                                |               | Gewoon persoonsgegevens                         |

| #  | Variable                               | Source | Description                | Anonymization                                                       | Categorie persoonsgegevens |
|----|----------------------------------------|--------|----------------------------|---------------------------------------------------------------------|----------------------------|
| 15 | Ingangsdatum_Verblijf plaats           | IBIS   | Location start date        | 1st of the month (only applies to training and sending data to CBS) | Gewoon persoonsgegevens    |
| 16 | Geboortedatum                          | IBIS   | Date of Birth              | Could use age instead                                               | Gewoon persoonsgegevens    |
| 17 | Geschat_niveau                         | IBIS   | Estimated education        |                                                                     | Gewoon persoonsgegevens    |
| 18 | Diploma_behaald                        | IBIS   | Graduated                  |                                                                     | Gewoon persoonsgegevens    |
| 19 | Omschrijving_land_van educatie         | IBIS   | Country of education       |                                                                     | Gewoon persoonsgegevens    |
| 20 | Startopleiding                         | IBIS   | Date Start Education       |                                                                     | Gewoon persoonsgegevens    |
| 21 | Eindopleiding                          | IBIS   | Date End education         |                                                                     | Gewoon persoonsgegevens    |
| 22 | Diploma_bezit                          | IBIS   | Diploma in possession      |                                                                     | Gewoon persoonsgegevens    |
| 23 | Omschrijving_land_werkervaringopgedaan | IBIS   | Country of work experience |                                                                     | Gewoon persoonsgegevens    |
| 24 | Branche_Omschrijving                   | IBIS   | Sector of industry         |                                                                     | Gewoon persoonsgegevens    |
| 25 | Aanstelling                            | IBIS   | Type of Job                |                                                                     | Gewoon persoonsgegevens    |
| 26 | Begindatum_werkervaring                | IBIS   | Work start date            | 1st of the month (only applies to training and sending data to CBS) | Gewoon persoonsgegevens    |

| #  | Variable                            | Source | Description                                          | Anonymization                                                        | Categorie persoonsgegevens |
|----|-------------------------------------|--------|------------------------------------------------------|----------------------------------------------------------------------|----------------------------|
| 27 | Einddatum_werkervaring              | IBIS   | Work end date                                        | 1st of the month (only applies to training and sending data to CBS)  | Gewoon persoonsgegevens    |
| 28 | Nummer_plaatsingseenheid_PLE_crypt  | IBIS   | Case number                                          |                                                                      | Gewoon persoonsgegevens    |
| 29 | classificatie_indicatie_nareiziger  | IBIS   | Family reunifier indicator                           |                                                                      | Gewoon persoonsgegevens    |
| 30 | Groepsrol_MVV_indicatie_nareiziger  | IBIS   | Group MVV family reunifier indicator                 |                                                                      | Gewoon persoonsgegevens    |
| 31 | Datum_huisvestings_gesprek          | IBIS   | Housing interview date                               | 1st of the month? (Only applies to training and sending data to CBS) | Gewoon persoonsgegevens    |
| 32 | Opvangvorm                          | IBIS   | Reception form                                       |                                                                      | Gewoon persoonsgegevens    |
| 33 | Gekoppeld Aan Gemeente              | IBIS   | Municipality that the person is eventually linked to |                                                                      | Gewoon persoonsgegevens    |
| 34 | Datum koppeling aan gemeente        | IBIS   | Date linked to municipality                          |                                                                      | Gewoon persoonsgegevens    |
| 35 | Indicatie_Vergunninghouder          | IBIS   | Permit holder                                        |                                                                      |                            |
| 36 | Soort_Plaatsing                     | IBIS   | Placement Type                                       |                                                                      | Gewoon persoonsgegevens    |
| 37 | Dienst                              | IBIS   | Service                                              |                                                                      | Gewoon persoonsgegevens    |
| 38 | Classificatie_Hervestiger_Vervallen | IBIS   | Resettlement classification                          |                                                                      | Gewoon persoonsgegevens    |
| 39 | Hardecriteriaavontoesing            | IBIS   | Hard criteria present                                |                                                                      | Gewoon persoonsgegevens    |

| #  | Variable                              | Source | Description                                                              | Anonymization | Categorie persoonsgegevens |
|----|---------------------------------------|--------|--------------------------------------------------------------------------|---------------|----------------------------|
| 40 | geomatch_regioadvies_1                | IBIS   | The first labor market region recommended by algorithm (rank == 1)       |               |                            |
| 41 | geomatch_regioadvies_eis_1            | IBIS   | The EIS score from 0-10 for the 1st GeoMatch advice                      |               |                            |
| 42 | geomatch_regioadvies_current_1        | IBIS   | The current PLE score from 0-10 for the 1st GeoMatch advice              |               |                            |
| 43 | geomatch_regioadvies_future_1         | IBIS   | The future PLEs score from 0-10 for the 1st GeoMatch advice              |               |                            |
| 44 | geomatch_regioadvies_ps_endorsement_1 | IBIS   | "True" if the PS checked the endorsement box for the 1st GeoMatch advice |               |                            |
| 45 | geomatch_regioadvies_2                | IBIS   | The second labor market region recommended by algorithm (rank == 2)      |               |                            |
| 46 | geomatch_regioadvies_eis_2            | IBIS   | The EIS score from 0-10 for the 1st GeoMatch advice                      |               |                            |
| 47 | geomatch_regioadvies_current_2        | IBIS   | The current PLE score from 0-10 for the 1st GeoMatch advice              |               |                            |
| 48 | geomatch_regioadvies_future_2         | IBIS   | The future PLEs score from 0-10 for the 1st GeoMatch advice              |               |                            |
| 49 | geomatch_regioadvies_ps_endorsement_2 | IBIS   | "True" if the PS checked the endorsement box for the 1st GeoMatch advice |               |                            |
| 50 | geomatch_regioadvies_3                | IBIS   | The third labor market region recommended by algorithm (rank == 3)       |               |                            |
| 51 | geomatch_regioadvies_eis_3            | IBIS   | The EIS score from 0-10 for the 1st GeoMatch advice                      |               |                            |

| #  | Variable                              | Source | Description                                                                                                                                                           | Anonymization | Categorie persoonsgegevens |
|----|---------------------------------------|--------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------|
| 52 | geomatch_regioadvies_current_3        | IBIS   | The current PLE score from 0-10 for the 1st GeoMatch advice                                                                                                           |               |                            |
| 53 | geomatch_regioadvies_future_3         | IBIS   | The future PLEs score from 0-10 for the 1st GeoMatch advice                                                                                                           |               |                            |
| 54 | geomatch_regioadvies_ps_endorsement_3 | IBIS   | "True" if the PS checked the endorsement box for the 1st GeoMatch advice                                                                                              |               |                            |
| 55 | ps_geomatch_feedback                  | IBIS   | Free text. Feedback that the PS can optionally give for overall GeoMatch advice                                                                                       |               |                            |
| 56 | ple_geomatch_treatment                | IBIS   | Assignment to treat or control                                                                                                                                        |               |                            |
| 57 | Omschrijvingarbeidsregio              | IBIS   | Regional advice                                                                                                                                                       |               |                            |
| 58 | Omschrijvingarbeidsregio_eis          | IBIS   | For each regional advice from the PS, the EIS score from 0-10 (can be retrieved from GeoMatch API)                                                                    |               |                            |
| 59 | Verblijfplaats (ple_sm_location_id)   | IBIS   | The ID of the location (POL or AZC) where the SM interview was performed for the PLE                                                                                  |               | Gewoon persoonsgegevens    |
| 60 | ple_ps_ids                            | IBIS   | The anonymized ID of each PS who conducted an interview and filled out the information in IBIS for both S+M and housing interviews (i.e. the PS interviewing the PLE) |               |                            |
| 61 | Einddatum_Verblijfplaats              | IBIS   | Location end date                                                                                                                                                     |               | Gewoon persoonsgegevens    |
| 62 | Date of arrival in municipality       | IBIS   | The date that the PLE arrived in the municipality (and is no longer tracked in IBIS)                                                                                  |               |                            |

| #  | Variable                   | Source | Description                                                                                                                                                                                                                                                        | Anonymization | Categorie persoonsgegevens |
|----|----------------------------|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------|
| 63 | location_in_geomatch_pilot | IBIS   | True if the AZC or POL is participating in the GeoMatch pilot                                                                                                                                                                                                      |               |                            |
| 64 | geomatch_ple_percent age   | IBIS   | This is the % of the overall target number that should be used for GeoMatch treatment and control cases for each municipality.                                                                                                                                     |               |                            |
| 65 | Recent Linkings            | TVS    | This is all of the linkings or relinkings that have occurred since the previous TVS import. This is sent to GeoMatch after each import from TVS to IBIS. See the OpenAPI documentation for a full accounting of these variables.                                   |               |                            |
| 66 | TVS Forecast               | TVS    | For each municipality, this is all of the information from the TVS Forecast page (i.e. Fase2, Taakstelling, etc). This is sent to GeoMatch after each import from TVS to IBIS. See the OpenAPI documentation or TVS page for a full accounting of these variables. |               |                            |
| 67 | LMR for Municipality       | IBIS   | The Labour Market Region for each municipality                                                                                                                                                                                                                     |               |                            |
| 68 | UR for LMR                 | IBIS   | The Uitplaatsingsregio for the LMR                                                                                                                                                                                                                                 |               |                            |
| 69 | Regioadvies overgenomen    | TVS    | Whether the AO accepted the PS or GeoMatch regional advice.                                                                                                                                                                                                        |               |                            |
| 70 | Reden koppeling            | TVS    | The reason the Allocation Officer chose the linked municipality                                                                                                                                                                                                    |               |                            |

| #  | Variable                               | Source          | Description                                                                                                                                                                                                                                                                                                                                                                  | Anonymization | Categorie persoonsgegevens |
|----|----------------------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|----------------------------|
| 71 | Reden niet overnemen regioadvies       | TVS             | The reason the Allocation Officer chose to override the regional advice                                                                                                                                                                                                                                                                                                      |               |                            |
| 72 | GeoMatch Monitoring Export             | GeoMatch server | This includes a snapshot of the GeoMatch server state for each day of the past month. It should be sent to IPL along with the IBIS monitoring variables each month. It includes the taakstelling state of each municipality for the day, as well as the most recent machine learning model in use. This is useful for performing comparisons to control PLEs for monitoring. |               |                            |
| 73 | Toelichting niet overnemen regioadvies | TVS             | Free text. The reason the AO chose not to adopt the regional advice.                                                                                                                                                                                                                                                                                                         |               |                            |
| 74 | PLE_aantal_vergunninghouders           | IBIS            | Number of permit holders in this PLE                                                                                                                                                                                                                                                                                                                                         |               |                            |
| 75 | Opmerking                              | TVS             | Free text. Comments from the AO on why they chose the municipality for linking.                                                                                                                                                                                                                                                                                              |               |                            |
| 76 | gemeente_percent_full                  | TVS             | A list of each municipality and how full it currently is based on taakstelling.                                                                                                                                                                                                                                                                                              |               |                            |
| 77 | Referentiekennmerk                     | IBIS            | Used for GeoMatch to determine if the PLE is in AA or VA procedure                                                                                                                                                                                                                                                                                                           |               |                            |
| 78 | Datum_Bekend_Refere ntiekennmerk       | IBIS            | [If possible] Date of referentiekennmerk                                                                                                                                                                                                                                                                                                                                     |               |                            |
| 79 | ple_sm_ur (Verblijfplaats UR)          | IBIS            | The Uitplaatsingsregio where the permit holder is residing at the time of the interview.                                                                                                                                                                                                                                                                                     |               |                            |

| #  | Variable   | Source | Description                                                                                                    | Anonymization | Categorie persoonsgegevens |
|----|------------|--------|----------------------------------------------------------------------------------------------------------------|---------------|----------------------------|
| 80 | ple_ao_ids | TVS    | The anonymized ID of each AO who was assigned each PLE in TVS, even if that AO did not end up linking the PLE. |               |                            |

Tabel 2: Justification keys (relevantie en noodzakelijkheid)

| Justification Code Key |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                      | Needed to link the variables to the variables of the RA-environment of Statistics Netherlands                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 2                      | Needed as an input into the statistical models used to predict integration outcomes. We expect predicted outcomes to differ across locations as a function of this variable, and thus inclusion of this variable will allow us to detect and leverage important synergies and make better match recommendations.                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 3                      | Needed as an input into the statistical models used to predict integration outcomes, and to ensure fairness across different groups and protected characteristics. We expect predicted outcomes to differ across locations as a function of this variable, and thus inclusion of this variable will allow us to detect and leverage important synergies and make better match recommendations regardless of which group (segmented by this variable) one belongs to. As a side note, it is worth mentioning that the literature has shown that, even for people of the same nationality, there can be important network effects based on different religions, and thus religion and ethnicity are important to include in addition to nationality. |
| 4                      | Needed to understand current assignment mechanism, build in real-world constraints, and focus only on the population of interest, in order for our statistical modeling to accurately represent the process and hence provide accurate predictions.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 5                      | Needed because the machine learning models use this variable to predict optimal locations and generate recommendations (regional advice).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 6                      | Needed for monitoring the algorithm during pilot implementation. This tool designed to be a system that continues to learn over time as things changes. It is impossible to know which variables might become more or less important over time, and the only way to figure that out and then make the best possible decisions on that basis is by having access to these variables.                                                                                                                                                                                                                                                                                                                                                                |
| 7                      | Needed to implement or improve the research design (for example by improving power, spillovers, or length of pilot)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

Tabel 3: Gebruik van persoonsgegevens ten behoeve van training

| # | Variable             | Justification |
|---|----------------------|---------------|
| 1 | EconPositie_Werkende | 2             |
| 2 | SLNLBPH              | 2             |
| 3 | onderwijs_ja         | 2             |

| #  | Variable                               | Justification |
|----|----------------------------------------|---------------|
| 4  | PEILMAAND                              | 2             |
| 5  | PEILMAAND                              |               |
| 6  | SDATUMAANVANG                          |               |
| 7  | SDATUMEINDE                            |               |
| 8  | Vnummer_crypt                          | 1             |
| 9  | BSN_crypt                              | 1             |
| 10 | VDLibisnr                              | 1             |
| 11 | Omschrijving_Geslacht                  | 2,3           |
| 12 | Omschrijving_Burg_Staat                | 2             |
| 13 | Omschrijving_land_van_herkomst         | 2,3           |
| 14 | Omschrijving_moedertaal                | 2,3           |
| 15 | Ingangsdatum_Verblijfplaats            | 4             |
| 16 | Geboortedatum                          | 2             |
| 17 | Geschat_niveau                         | 2             |
| 18 | Diploma_behaald                        | 2             |
| 19 | Omschrijving_land_vaneducatie          | 2             |
| 20 | Startopleiding                         | 2             |
| 21 | Eindopleiding                          | 2             |
| 22 | Diploma_bezit                          | 2             |
| 23 | Omschrijving_land_werkervaringopgedaan | 2             |
| 24 | Branche_Omschrijving                   | 2             |
| 25 | Aanstelling                            | 2             |
| 26 | Begindatum_werkervaring                | 2             |
| 27 | Einddatum_werkervaring                 | 2             |
| 28 | Nummer_plaatsingseenheid_PLE_crypt     | 1             |
| 29 | classificatie_indicatie_nareiziger     | 2,4           |

| #  | Variable                            | Justification |
|----|-------------------------------------|---------------|
| 30 | Groepsrol_MVV_indicatie_nareiziger  | 2,4           |
| 31 | Datum_huisvestings_gesprek          | 2,4           |
| 32 | Opvangvorm                          | 4             |
| 33 | Gekoppeld Aan Gemeente              | 2,4           |
| 34 | Datum koppeling aan gemeente        | 2,4           |
| 35 | Indicatie_Vergunninghouder          | 4             |
| 36 | Soort_Plaatsing                     | 4             |
| 37 | Dienst                              | 4             |
| 38 | Classificatie_Hervestiger_Vervallen | 4             |
| 39 | Hardecriteriauntoepassing           | 4             |
| 59 | Verblijfplaats (ple_sm_location_id) |               |
| 61 | Einddatum_Verblijfplaats            |               |
| 67 | LMR for Municipality                |               |
| 68 | UR for LMR                          |               |
| 69 | Regioadvies overgenomen             |               |
| 74 | PLE_aantal_vergunninghouders        |               |
| 77 | Referentiekenmerk                   |               |
| 78 | Datum_Bekend_Referentiekenmerk      |               |

Tabel 4: Gebruik van persoonsgegevens ten behoeve van het adviseren van een regio

| #  | Variable                       | Justification |
|----|--------------------------------|---------------|
| 10 | VDLibisnr                      | 1             |
| 11 | Omschrijving_Geslacht          | 5             |
| 12 | Omschrijving_Burg_Staat        | 5             |
| 13 | Omschrijving_land_van_herkomst | 5             |
| 14 | Omschrijving_moedertaal        | 5             |
| 15 | Ingangsdatum_Verblijfplaats    | 5             |

| #  | Variable                               | Justification |
|----|----------------------------------------|---------------|
| 16 | Geboortedatum                          | 5             |
| 17 | Geschat_niveau                         | 5             |
| 18 | Diploma_behaald                        | 5             |
| 19 | Omschrijving_land_vaneducatie          | 5             |
| 20 | Startopleiding                         | 5             |
| 21 | Eindopleiding                          | 5             |
| 22 | Diploma_bezit                          | 5             |
| 23 | Omschrijving_land_werkervaringopgedaan | 5             |
| 24 | Branche_Omschrijving                   | 5             |
| 25 | Aanstelling                            | 5             |
| 26 | Begindatum_werkervaring                | 5             |
| 27 | Einddatum_werkervaring                 | 5             |
| 28 | Nummer_plaatsingseenheid_PLE_crypt     | 1             |
| 29 | classificatie_indicatie_nareiziger     | 4,5           |
| 30 | Groepsrol_MVV_indicatie_nareiziger     | 4,5           |
| 31 | Datum_huisvestings_gesprek             | 4,5           |
| 32 | Opvangvorm                             | 4             |
| 35 | Indicatie_Vergunninghouder             | 4             |
| 36 | Soort_Plaatsing                        | 4             |
| 37 | Dienst                                 | 4             |
| 38 | Classificatie_Hervestiger_Vervallen    | 4             |
| 56 | ple_geomatch_treatment                 | 4             |
| 57 | Omschrijvingarbeidsregio               |               |
| 59 | Verblijfplaats (ple_sm_location_id)    |               |
| 61 | Einddatum_Verblijfplaats               |               |
| 65 | Recent Linkings                        | 6             |

| #  | Variable                       | Justification |
|----|--------------------------------|---------------|
| 66 | TVS Forecast                   | 6             |
| 67 | LMR for Municipality           | 6             |
| 68 | UR for LMR                     | 6             |
| 74 | PLE_aantal_vergunninghouders   |               |
| 77 | Referentiekenmerk              |               |
| 78 | Datum_Bekend_Referentiekenmerk |               |
| 79 | ple_sm_ur (Verblijfplaats UR)  |               |

Tabel 5: Gebruik van persoonsgegevens ten behoeve van monitoring<sup>33</sup>

| #  | Variable                       | Monitoring Level | Justification |
|----|--------------------------------|------------------|---------------|
| 1  | EconPositie_Werkende           | 5                | 6             |
| 2  | SLNLBPH                        | 5                | 6             |
| 3  | onderwijs_ja                   | 5                | 6             |
| 4  | PEILMAAND                      | 5                | 6             |
| 5  | PEILMAAND                      | 5                |               |
| 6  | SDATUMAANVANG                  | 5                |               |
| 7  | SDATUMEINDE                    | 5                |               |
| 8  | Vnummer_crypt                  | 5                | 1             |
| 9  | BSN_crypt                      | 5                | 1             |
| 10 | VDLibisnr                      | 1,2,3,4,5        | 1             |
| 11 | Omschrijving_Geslacht          | 3,4,5            | 6             |
| 12 | Omschrijving_Burg_Staat        | 3,4,5            | 6             |
| 13 | Omschrijving_land_van_herkomst | 3,4,5            | 6             |
| 14 | Omschrijving_moedertaal        | 3,4,5            | 6             |
| 15 | Ingangsdatum_Verblijfplaats    | 1,2,3,4,5        | 6             |
| 16 | Geboortedatum                  | 3,4,5            | 6             |

<sup>33</sup> Zie ook: GeoMatch Netherlands Pilot Monitoring Plan, IPL, februari 2024

| #  | Variable                               | Monitoring Level | Justification |
|----|----------------------------------------|------------------|---------------|
| 17 | Geschat_niveau                         | 3,4,5            | 6             |
| 18 | Diploma_behaald                        | 3,4,5            | 6             |
| 19 | Omschrijving_land_vaneducatie          | 3,4,5            | 6             |
| 20 | Startopleiding                         | 3,4,5            | 6             |
| 21 | Eindopleiding                          | 3,4,5            | 6             |
| 22 | Diploma_bezit                          | 3,4,5            | 6             |
| 23 | Omschrijving_land_werkervaringopgedaan | 3,4,5            | 6             |
| 24 | Branche_Omschrijving                   | 3,4,5            | 6             |
| 25 | Aanstelling                            | 3,4,5            | 6             |
| 26 | Begindatum_werkervaring                | 3,4,5            | 6             |
| 27 | Einddatum_werkervaring                 | 3,4,5            | 6             |
| 28 | Numer_plaatsingseenheid_PLE_crypt      | 1,2,3,4,5        | 1             |
| 29 | classificatie_indicatie_nareiziger     | 1,2,3,4,5        | 6             |
| 30 | Groepsrol_MVV_indicatie_nareiziger     | 1,2,3,4,5        | 6             |
| 31 | Datum_huisvestings_gesprek             | 1,2,3,4,5        | 6             |
| 32 | Opvangvorm                             | 1,2,3,5          | 6             |
| 33 | Gekoppeld Aan Gemeente                 | 1,2,3,4,5        | 6             |
| 34 | Datum koppeling aan gemeente           | 1,2,3,4,5        | 6             |
| 35 | Indicatie_Vergunninghouder             | 1                | 6             |
| 36 | Soort_Plaatsing                        | 1                | 6             |
| 37 | Dienst                                 | 3,4,5            | 6             |
| 38 | Classificatie_Hervestiger_Vervallen    | 1,2,3,4,5        | 6             |
| 39 | Hardecriteriavantoepassing             | 2,4,5            | 6             |
| 40 | geomatch_regioadvies_1                 | 1,2,3,4,5        | 6             |
| 41 | geomatch_regioadvies_eis_1             | 1,7              | 7             |
| 42 | geomatch_regioadvies_current_1         | 1,7              | 7             |
| 43 | geomatch_regioadvies_future_1          | 1,7              | 7             |

| #  | Variable                              | Monitoring Level | Justification |
|----|---------------------------------------|------------------|---------------|
| 44 | geomatch_regioadvies_ps_endorsement_1 | 1,7              | 6             |
| 45 | geomatch_regioadvies_2                | 1,2,3,4,5        | 6             |
| 46 | geomatch_regioadvies_eis_2            | 1,7              | 7             |
| 47 | geomatch_regioadvies_current_2        | 1,7              | 7             |
| 48 | geomatch_regioadvies_future_2         | 1,7              | 7             |
| 49 | geomatch_regioadvies_ps_endorsement_2 | 1,7              | 6             |
| 50 | geomatch_regioadvies_3                | 1,2,3,4,5        | 6             |
| 51 | geomatch_regioadvies_eis_3            | 1,7              | 7             |
| 52 | geomatch_regioadvies_current_3        | 1,7              | 7             |
| 53 | geomatch_regioadvies_future_3         | 1,7              | 7             |
| 54 | geomatch_regioadvies_ps_endorsement_3 | 1,7              | 6             |
| 55 | ps_geomatch_feedback                  | 1,7              | 6             |
| 56 | ple_geomatch_treatment                | 1,2,3,4,5        | 6             |
| 57 | Omschrijvingarbeidsregio              | 1,4,5            | 6             |
| 58 | Omschrijvingarbeidsregio_eis          | 1,7              | 6             |
| 59 | Verblijfplaats (ple_sm_location_id)   | 1,2,3            | 6             |
| 60 | ple_ps_ids                            | 1,5              | 6             |
| 61 | Einddatum_Verblijfplaats              | 1,2,3,4,5        | 6             |
| 62 | Date of arrival in municipality       | 1,2,3,4,5        | 6             |
| 63 | location_in_geomatch_pilot            | 1,2,3,4,5        | 7             |
| 64 | geomatch_ple_percentage               |                  | 7             |
| 65 | Recent Linkings                       | 2,6              | 7             |
| 66 | TVS Forecast                          | 2,6              | 7             |
| 67 | LMR for Municipality                  | 2,3,4,5,6        | 6             |
| 68 | UR for LMR                            | 2,3,4,5,6        | 6             |
| 69 | Regioadvies overgenomen               | 1,2,3,4,5        | 7             |
| 70 | Reden koppeling                       | 1,2,3,4,5        | 7             |

| #  | Variable                               | Monitoring Level | Justification |
|----|----------------------------------------|------------------|---------------|
| 71 | Reden niet overnemen regioadvies       | 1,2,3,4,5        | 7             |
| 72 | GeoMatch Monitoring Export             | 1,2,3,4,5        | 7             |
| 73 | Toelichting niet overnemen regioadvies | 1,2,3,4,5        | 7             |
| 74 | PLE_aantal_vergunninghouders           | 1,2,3,4,5        |               |
| 75 | Opmerking                              | 1,2,3,4,5        | 7             |
| 76 | gemeente_percent_full                  | 1,2,3,4,5        |               |
| 77 | Referentiekenmerk                      | 1,2,3,4,5        | 7             |
| 78 | Datum_Bekend_Referentiekenmerk         | 1,2,3,4,5        | 7             |
| 79 | ple_sm_ur (Verblijfplaats UR)          | 1,2,3            | 6             |
| 80 | ple_ao_ids                             | 1,5              | 6             |

# Bijlage 2. Gedragscode Artificiële Intelligentie

De Gedragscode Artificiële Intelligentie maakt onlosmakelijk deel uit van het AIIA. Deze set gedragsregels vormt het fundament onder het AIIA.

## ***Deel 1: Ethische principes***

Toepassing van AI moet voldoen aan de volgende algemene ethische principes, gebaseerd op de European Group on Ethics in Science and New Technologies.

1. Wij maken geen inbreuk op de menselijke waardigheid
2. Wij respecteren de menselijke autonomie
3. Wij onderzoeken en ontwikkelen AI in overeenstemming met de mensenrechten en universele waarden
4. Wij dragen bij aan rechtvaardigheid, gelijke kansen en solidariteit
5. Wij respecteren de uitkomsten van democratische besluitvorming
6. Wij passen AI toe conform de principes van de rechtsstaat
7. Wij waarborgen de veiligheid en integriteit van gebruikers
8. Wij voldoen aan wet- en regelgeving omtrent databescherming en privacy
9. Wij voorkomen schadelijke invloed op het milieu

## ***Deel 2: Praktijkregels***

De praktijkregels zijn concrete handvatten om AI verantwoord toe te passen in de praktijk. Deze set van regels is gebaseerd op en een update van de "Handreiking voor gedragsregels autonome systemen" van ECP.

10. Wij maken de gebruiker waar noodzakelijk identificeerbaar
11. Wij geven inzicht in de werking en handelingsgeschiedenis van AI-systemen
12. Wij dragen zorg voor de integriteit van AI-systemen, opgeslagen informatie en overdracht daarvan
13. Wij dragen zorg voor vertrouwelijkheid van informatie
14. Wij dragen zorg voor continuïteit
15. Wij dragen zorg voor traceerbaarheid, toetsbaarheid en voorspelbaarheid van AI-handelingen
16. Wij maken geen inbreuk op intellectuele eigendommen
17. Wij respecteren de privacy van mensen, en de wet- en regelgeving op dat gebied
18. Wij maken verantwoordelijkheden in de keten helder
19. Wij laten de informatieverwerking door AI-systemen auditen

# Bijlage 3. Geraadpleegde documenten

- 20230719\_Selectielijst COA - definitief voor publicatie\_0
- 220626 PIA Project Kansrijke koppeling v0.8
- 220706 AIIA v0.5
- 220719 Deloitte rapport - Algoritme Analyse Kansrijke koppeling (22\_1275\_KD\_Ir) (2)
- 230723 Algoritmische audit Deloitte
- 231207 PID Kansrijke koppeling mbv AI - powerpoint
- 231212 PID Kansrijke koppeling mbv AI
- COA-IPL\_PilotVariables-20231128
- GeoMatch Variable Importance Plots
- GeoMatchNetherlands\_COA\_rd\_launchv2
- GeoMatchNetherlands\_ExternalPilotMonitoringPlan\_round2\_COAreview
- GeoMatchNetherlands\_IEReport\_070221
- GeoMatchNetherlands\_MonitoringPlan\_2024
- GeoMatchNetherlands\_RCT\_flow\_restart\_20230817
- GeoMatchNetherlands\_scenarios\_all
- GeoMatchNetherlands\_UpdatedBacktests\_051022\_gdpr\_financial
- Informatiebeveiligingsbeleid COA 20190719 v1.0
- NLD\_GeoMatch\_TechnicalDocumentation\_20240216
- PID Kansrijke koppeling mbv AI PowerPoint
- Regiomatch Werkinstructie Pleinpagina Pilot - huisvestings en screeningsgesprek
- Service Level Agreement COA - Stanford University 0.4 (feedback ES) (002)
- Service Level Agreement COA - Stanford University 1.0 definitief
- Verwerkersovereenkomst COA-IPL
- Werkinstructie v0.9



## WIJ ZIJN BERENSCHOT, GRONDLEGGER VAN VOORUITGANG

Nederland is continu in ontwikkeling. Maatschappelijk, economisch en organisatorisch verandert er veel. Al ruim 85 jaar volgen wij als adviesbureau deze ontwikkelingen op de voet en werken we aan een vooruitstrevende samenleving. De behoefte om iets fundamenteels te betekenen voor mens en maatschappij zit in onze genen. Met onze adviezen en oplossingen hebben we dan ook actief meegebouwd aan het Nederland van vandaag. Altijd op zoek naar duurzame vooruitgang.

Alles wat we doen, is onderzocht, onderbouwd en vanuit meerdere invalshoeken bekeken. Zo komen we tot gefundeerde adviezen en slimme oplossingen. Die zijn op het eerste gezicht misschien niet altijd de meest voor de hand liggende. Juist deze eigenzinnigheid maakt ons uniek. Daarbij zijn we niet van symptoombestrijding. En gaan pas naar huis als het is opgelost.

### Berenschot B.V.

Van Deventerlaan 31-51, 3528 AG UTRECHT  
Postbus 8039, 3503 RA UTRECHT  
030 2 916 916  
[www.berenschot.nl](http://www.berenschot.nl)



Deloitte Risk Advisory B.V.  
Gustav Mahlerlaan 2970  
1081 LA Amsterdam  
Postbus 58110  
1040 HC Amsterdam  
Nederland

Tel: 512e  
www.deloitte.nl

Van:

5.1.2.e

Datum:

29 april 2024

Naar:

Centraal Orgaan opvang asielzoekers  
T.a.v. 5.1.2.e

Onze referentie:

KD/wk/24-0725

Onderwerp:

Rapportage analyse Algoritme Kansrijke Koppeling

Geachte 5.1.2.e

Het Centraal Orgaan opvang asielzoekers ("COA") heeft Deloitte Risk Advisory B.V. ("Deloitte") gevraagd om een analyse uit te voeren van het algoritme 'kansrijke koppeling' ("Initiële Algoritme"). Het COA wil dit algoritme in gebruik nemen om de koppeling van vergunningshouders aan gemeentes te verbeteren. Deloitte heeft ook in het verleden analyses uitgevoerd op dit Initiële Algoritme:

- **Analyse 1:** Deloitte heeft in de zomer van 2022 werkzaamheden uitgevoerd als onderdeel van de onderzoeksfase van het Initiële Algoritme. Die werkzaamheden hebben geresulteerd in een rapport met kenmerk KD/lr/22-1275 met bevindingen, bestaande uit observaties en aanbevelingen. Op basis van deze observaties en aanbevelingen heeft COA het Initiële Algoritme aangepast ("Algoritme") in voorbereiding op de pilotfase.
- **Analyse 2:** In 2023 en 2024 heeft Deloitte het Algoritme in voorbereiding op de pilotfase geanalyseerd. De werkzaamheden die Deloitte heeft uitgevoerd op het Algoritme in Analyse 2 staan beschreven in de getekende overeenkomst met kenmerk KS/js/23-0546 en bestaan uit twee stappen:
  - **Stap 1:** Tijdens de eerste stap van Analyse 2 ("Stap 1") heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de volgende aanbevelingen uit Analyse 1: 1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, CBS-1, CBS-3, CBS-4 en CBS-5. De werkzaamheden van Stap 1 zijn afgerond in juli 2023 en hebben geresulteerd in een tussentijdse rapportage met kenmerk KD/avl/23-0957 ("Tussentijdse Rapportage").
  - **Stap 2:** Tijdens de tweede stap van Analyse 2 ("Stap 2") heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de aanbevelingen die niet zijn geanalyseerd in Stap 1, en de opvolging van aanbevelingen die niet zijn gesloten in de Tussentijdse Rapportage na afronding van Stap 1.

Dit rapport presenteert de gecombineerde resultaten van Stap 1 en Stap 2 ("Rapport").



29 april 2024  
KD/wk/24-0725

Mocht u naar aanleiding van dit rapport vragen of opmerkingen hebben, aarzel dan niet contact op te nemen met 5.1.2 e (tel. 5.1.2 e).

Met vriendelijke groet,

Namens Deloitte Risk Advisory B.V.





29 april 2024  
KD/wk/24-0725

**Rapportage analyse**  
**Algoritme Kansrijke Koppeling**

|                      |                                              |
|----------------------|----------------------------------------------|
| Afdeling             | : Directie vakontwikkeling en -ondersteuning |
| Eigenaar             | : 5.1.2.e                                    |
| Gedelegeerd eigenaar | : 5.1.2.e                                    |
| Auteur(s)            | : Deloitte Risk Advisory B.V.                |
| Datum                | : 29 april 2024                              |
| Status               | : Finaal                                     |
| Type analyse         | : Tweede analyse                             |



29 april 2024  
KD/wk/24-0725

## 1. Managementsamenvatting

Het COA heeft Deloitte gevraagd om een analyse uit te voeren van het algoritme 'kansrijke koppeling' ("Initiële Algoritme"). Het COA wil dit algoritme in gebruik nemen om de koppeling van vergunningshouders aan gemeentes te verbeteren. Deloitte heeft ook in het verleden analyses uitgevoerd op dit Initiële Algoritme:

- **Analyse 1:** Deloitte heeft in de zomer van 2022 werkzaamheden uitgevoerd als onderdeel van de onderzoeksfase van het Initiële Algoritme. Die werkzaamheden hebben geresulteerd in een rapport met bevindingen, bestaande uit observaties en aanbevelingen. Op basis van deze observaties en aanbevelingen heeft COA het Initiële Algoritme aangepast in voorbereiding op de pilotfase.
- **Analyse 2:** In 2023 en 2024 heeft Deloitte het Algoritme in voorbereiding op de pilotfase geanalyseerd. De werkzaamheden die Deloitte heeft uitgevoerd op het Algoritme in Analyse 2 staan beschreven in de getekende overeenkomst met kenmerk KS/js/23-0546 en bestaan uit twee stappen:
  - **Stap 1:** Tijdens de eerste stap van Analyse 2 heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de volgende aanbevelingen uit Analyse 1: 1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, CBS-1, CBS-3, CBS-4 en CBS-5. De werkzaamheden van Stap 1 zijn afgerond in juli 2023 en hebben geresulteerd in een tussentijdse rapportage met kenmerk KD/avl/23-0957.
  - **Stap 2:** Tijdens de tweede stap van Analyse 2 heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de aanbevelingen die niet zijn geanalyseerd in Stap 1, en de opvolging van aanbevelingen die niet zijn gesloten in de Tussentijdse Rapportage na afronding van Stap 1.

Dit rapport presenteert de gecombineerde resultaten van Stap 1 en Stap 2, waarbij de nadruk ligt op de bevindingen en aanbevelingen uit het Deloitte rapport van de initiële analyse opgevolgd zijn door het COA. Dit rapport zal door het COA gebruikt worden om te bepalen of het Algoritme op een verantwoorde en ethische wijze de plaatsing van vergunninghouders kan ondersteunen, en met name of de Pilot op een verantwoorde en ethische wijze kan starten.

Het Centraal Bureau voor de Statistiek ("CBS") heeft observaties gerapporteerd tijdens de initiële analyse [1]. Op het verzoek van het COA heeft Deloitte de observaties van het CBS opnieuw geanalyseerd tijdens Analyse 2. De resultaten van deze tweede analyse zijn opgenomen in dit rapport.

De analyses die zijn uitgevoerd door Deloitte hebben geleid tot het sluiten van 13 bevindingen. COA heeft niet voldoende opvolging gegeven aan 12 bevindingen waardoor deze niet gesloten konden worden. Deloitte heeft deze bevindingen herschreven zodat de open bevindingen de huidige status reflecteren.



29 april 2024  
KD/wk/24-0725

- De bevindingen zijn ingedeeld in vier categorieën documentatie, governance, model en CBS:
- 6 bevindingen in de categorie documentatie hebben betrekking op ontbrekende of onvolledige informatie in de algoritmedocumentatie.
  - 3 bevindingen in de categorie governance hebben betrekking op de verdeling van rollen en verantwoordelijkheden en de besluitvorming daarover, in de gehele algoritmelevenscyclus van het Algoritme die onvoldoende zijn uitgewerkt.
  - 1 bevinding in de categorie model is gerelateerd aan inputselectie wat niet is uitgevoerd tijdens de ontwikkeling van het Algoritme.
  - 2 bevindingen in de categorie CBS hebben betrekking op de bevindingen die CSB heeft gemaakt tijdens de initiële analyse.

De meeste openstaande bevindingen hebben betrekking op twee terugkerende thema's:

- 1) Het Algoritme is getraind op data tot en met juni 2018, en vrijwel alle resultaten in de algoritmedocumentatie zijn gebaseerd op dit Algoritme. Sinds 2018 zijn er veranderingen opgetreden in de samenleving, zoals veranderingen in instroom in de asielketen en veranderingen in thuiswerken door bijvoorbeeld COVID-19. COA is voornemens in 2024 een Pilot te starten. Het risico bestaat dat het besluit of het Algoritme op een verantwoorde manier ingezet kan worden, gebaseerd is op verouderde resultaten.
- 2) Het IPL geeft onvoldoende handvatten aan het COA om het COA te begeleiden in het maken van belangrijke besluiten rondom het Algoritme. Dit bevat zowel het identificeren van wat belangrijke keuzes zijn, als het uitwerken van deze keuzes rondom het Algoritme (bijvoorbeeld op het gebied van methodologie datakwaliteit, model performance en monitoring).

Tabel 1: Overzicht van geeft een overzicht weer van de openstaande bevindingen die Deloitte heeft gemaakt.

**Tabel 1: Overzicht van openstaande Deloitte bevindingen**

| Nr. | Categorie  | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Aanbeveling                                                                                                                                                                                                                                                  |
|-----|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3   | Governance | <p>In 2022 heeft een sessie plaatsgevonden met verschillende belanghebbenden. Hierin is vastgesteld welke ethische waarden de belanghebbenden het belangrijkste vonden in de context van het Algoritme. COA heeft deze ethische waarden als uitgangspunt genomen voor de verdere ontwikkeling van het Algoritme.</p> <p>Echter, er is daarna geen directe input op de door COA gemaakte keuzes gevraagd of terugkoppeling gegeven aan belanghebbenden over hoe de voor hun belangrijke ethische waardes zijn meegenomen tijdens de ontwikkeling van het Algoritme.</p> | <p>Betrek de eerder betrokken belanghebbenden en toets of de belangrijkste waarden van deze belanghebbenden voldoende terugkomen in het Algoritme, en of ze vertrouwen hebben dat het Algoritme ethisch en verantwoord kan worden toegepast in de Pilot.</p> |



29 april 2024  
KD/wk/24-0725

|          |              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>4</b> | Documentatie | COA heeft de rollen en verantwoordelijkheden van belanghebbenden geïdentificeerd. Echter, deze rollen en verantwoordelijkheden komen onvoldoende naar voren in de algoritmedocumentatie.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Verrijk de algoritmedocumentatie met een verduidelijking over de rol van belanghebbenden, namelijk dat de externe belanghebbenden SZW en VNG enkel geïnformeerd zullen worden, maar geen consulterende of besluitvormende rol zullen krijgen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b>5</b> | Governance   | <p>Het COA is afhankelijk van externe partijen voor het gebruik en beheersing van het Algoritme. De opgestelde SLA tussen Stanford University en COA is niet ondertekend en de duur van de overeenkomst is niet vastgelegd, waardoor onduidelijkheid blijft bestaan over de verdere samenwerkingsafspraken tussen COA en het IPL.</p> <p>Verder observeert Deloitte dat COA voor de gemaakte keuzes, conclusies over de kwaliteit van de data en het Algoritme, en interpretatie van de monitoringresultaten, vaak in eerste instantie naar het IPL kijkt.</p> <p>COA zou echter vanuit haar maatschappelijke verantwoordelijkheid moeten kunnen onderbouwen waarom keuzes zijn gemaakt en of die bij haar visie en waardes past. Ook zou COA zelf conclusies moeten trekken op basis van de door IPL aangeleverde performance tests, datakwaliteitschecks en monitoringresultaten, om op basis daarvan te bepalen of op een verantwoorde manier gestart kan worden met de Pilot.</p> | <p>Formaliseer de samenwerkingsafspraken tussen COA en IPL door de tijdsduur van de overeenkomst op te nemen in de opgestelde SLA en deze door beide partijen te ondertekenen.</p> <p>Maak duidelijke afspraken tussen COA en IPL over de benodigde handvatten die IPL zou moeten bieden om COA te begeleiden in het maken van belangrijke besluiten rondom het Algoritme. IPL zou moeten ondersteunen in de vertaling van de tabellen met datakwaliteit en model performance resultaten naar wat dit daadwerkelijk voor het gebruik van het Algoritme betekent, zodat COA een weloverwogen besluit kan nemen over de start van de Pilot.</p> <p>Daarnaast zou IPL als kennishouder uit zichzelf contact moeten zoeken met COA over of gemaakte (ontwerp)keuzes bij COA's visie en waarde passen.</p> |
| <b>6</b> | Governance   | Het is onduidelijk waarom bepaalde (ontwerp)keuzes zijn gemaakt (zoals de onderbouwing van de gekozen "Mapping" functies en optimalisatie criteria), en dat conclusies op basis van de uitgevoerde performance tests en datakwaliteitschecks ontbreken. Hierdoor blijft het onduidelijk of de data en het Algoritme van voldoende kwaliteit zijn.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | IPL dient de algoritmedocumentatie te verrijken met de onderbouwing waarom bepaalde (ontwerp)keuzes zijn gemaakt, zoals de gekozen "Mapping" functies en optimalisatie criteria. Vervolgens dient COA te besluiten of deze onderbouwing past bij de sociaal-maatschappelijke context van COA en het Algoritme.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |



29 april 2024  
KD/wk/24-0725

|           |              |                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |              | Dit brengt het risico met zich mee dat de stuurgroep van COA niet de juiste handvatten heeft om een weloverwogen besluit te kunnen nemen over de start van de Pilot met het Algoritme.                                                                                                                                                                                  | IPL moet COA de juiste handvatten bieden zodat COA aan de hand van de gepresenteerde datakwaliteitsmetriecken en performance metriecken een conclusie kan trekken of de datakwaliteit en model performance voldoende zijn om te kunnen starten met de Pilot en waarom dit het geval is. COA dient vervolgens te concluderen of de datakwaliteit en model performance voldoende zijn om te kunnen starten met de Pilot.                                                                                                         |
| <b>11</b> | Documentatie | De algoritmedocumentatie bevat een beschrijving van de uitgevoerde datakwaliteitschecks en de resultaten ervan. Echter, er zijn geen onderbouwde conclusies gedocumenteerd over de uitgevoerde datakwaliteitschecks. Hierdoor is het onduidelijk of de data van voldoende kwaliteit is om gebruikt te kunnen worden voor de Pilot.                                      | IPL moet COA de juiste handvatten bieden om aan de hand van de gepresenteerde datakwaliteitsmetriecken een conclusie te kunnen trekken of de datakwaliteit voldoende is om te kunnen starten met de Pilot en waarom dit het geval is. COA dient dan te concluderen of de datakwaliteit voldoende is om te kunnen starten met de Pilot.                                                                                                                                                                                         |
| <b>12</b> | Documentatie | Er is geen recente documentatie beschikbaar over beide eerlijkheidsperspectieven. Daarnaast is de definitie van eerlijkheid niet getoetst met een diverse groep betrokkenen.                                                                                                                                                                                            | IPL dient de eerlijkheidsperspectieven opnieuw te analyseren op basis van de laatste versie van het Algoritme dat gebruikt zal worden in de Pilot. Op basis van de resultaten van deze analyse, dient COA te concluderen of het Algoritme voldoet aan de door COA gestelde definitie van eerlijkheid.<br><br>Daarnaast dient COA de eerlijkheidsperspectieven met een diverse groep betrokkenen te toetsen zodat COA draagvlak kan creëren voor deze perspectieven. COA kan ook beargumenteren waarom zij dit niet nodig acht. |
| <b>13</b> | Model        | Het is niet gedocumenteerd waarom de geselecteerde kenmerken relevant zijn voor het Algoritme. Het risico ontstaat dat het Algoritme hierdoor kenmerken gebruikt die niet relevant zijn voor de toekomst.<br><br>Ook kan het gebruiken van onnodig veel data in strijd zijn met het dataminimalisatie principe uit de AVG wanneer de data bestaat uit persoonsgegevens. | IPL dient per attribuut een heldere beschrijving en duidelijke onderbouwing op te nemen in de algoritmedocumentatie waarom de geselecteerde kenmerken relevant zijn voor het Algoritme.<br><br>COA dient IPL hierbij te ondersteunen door input te leveren waarom de geselecteerde kenmerken relevant zijn in de sociaal maatschappelijke context van het Algoritme en COA.                                                                                                                                                    |



29 april 2024  
KD/wk/24-0725

|           |              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |              | Verder kan het Algoritme leren van verkeerde relaties in de data (spurious correlations) die niet aanwezig zijn in de data uit de praktijk.                                                                                                                                                                                                                                                                                                                                                                                                                               | Verder dient COA het eindbesluit te nemen welke kenmerken in het Algoritme gebruikt worden.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>14</b> | Documentatie | COA heeft de verbanden van het Algoritme niet onderbouwd in de algoritmedocumentatie. Daarnaast zijn de variable importance plots verouderd. Het is niet gedocumenteerd of de Post Covid Algoritmes en het Pilot Algoritme vergelijkbare relaties vinden op basis van recentere, post COVID data. Het risico ontstaat dat het Pilot Algoritme wordt gebruikt in de Pilot zonder dat de werking daarvan wordt begrepen of dat er goedkeuring wordt gegeven voor een verouderd Algoritme.                                                                                   | COA en IPL dienen samen de algoritmedocumentatie te verrijken met een onderbouwing over hoe de gemodelleerde verbanden in het Algoritme te verklaren zijn. COA kan de gemodelleerde verbanden verklaren door het gesprek aan te gaan met experts (bijv. algoritmegebruikers), deze betrokkenen kunnen bevestigen of de gemodelleerde verbanden niet onlogisch of onwenselijk zijn vanuit de praktijk. Hierbij kunnen bijvoorbeeld de variable importance plots als uitgangspunt worden genomen, en kan met experts worden getoetst of het belang van de verschillende attributen verklaarbaar is. |
| <b>18</b> | Documentatie | De financiële voordelen van het Algoritme worden gecommuniceerd naar interne betrokkenen met mogelijk externe communicatie als gevolg terwijl de accuraatheid van dit getal onduidelijk en verouderd is.                                                                                                                                                                                                                                                                                                                                                                  | Wees transparant over de accuraatheid, en daarmee de betrouwbaarheid, van de impact op loon analyse en de financiële voordelen van het Algoritme.<br><br>Deze aanbeveling kan ook opgevolgd worden door de financiële voordelen van het Algoritme niet te communiceren, en deze dus ook niet op te nemen in het algoritmeregister en geen rol te laten spelen in besluitvorming rond het gebruik van persoonsgegevens (zie bevinding 20).                                                                                                                                                         |
| <b>20</b> | Documentatie | COA heeft onderbouwd dat het gebruik wil maken van de persoonsgegevens land van herkomst en moedertaal in het Algoritme, en dat IPL zal gaan monitoren dat er geen etnische groepen worden benadeeld. Echter, een juridische onderbouwing ontbreekt nog voor het mogen gebruiken van deze persoonsgegevens.<br><br>COA heeft aangegeven dat deze juridische onderbouwing niet alleen voor dit Algoritme ontbreekt, maar een breder probleem is waar ze zich bewust van is en aan werkt. Ook is nog niet uitgewerkt hoe de monitoring op etniciteit zal gaan plaatsvinden. | Finaliseer de juridische onderbouwing voor het gebruik van persoonsgegevens, waaronder de persoonsgegevens land van herkomst, moedertaal en etniciteit (voor bias monitoringdoeleinden), en leg dit vast in de DPIA.<br><br>Geef meer details over de door IPL uitgevoerde analyse. Werk ook verder uit hoe de monitoring op etniciteitsgroepen zal plaatsvinden.                                                                                                                                                                                                                                 |



29 april 2024  
KD/wk/24-0725

Als laatste ontbreken details over de door IPL uitgevoerde analyse waarmee wordt aangetoond dat er geen statistisch verband is tussen land van herkomst, moedertaal en etniciteit. Meer details over de uitgevoerde analyse zijn gewenst om de uitkomst te kunnen plaatsen.

Tabel 2 geeft een overzicht weer van de bevindingen die het CBS heeft gemaakt tijdens de Initiële Analyse. De bevindingen zijn opnieuw geanalyseerd door Deloitte. Voor verdere uitleg van de CBS bevindingen wordt verwezen naar de analyse van het CBS [1].

Tabel 2: Overzicht van openstaande CBS bevindingen

| Nr.       | Observatie                                                                                                                                                                                                                                                                                                                                                                                                         | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3-<br>CBS | In de algoritmedocumentatie komt niet naar voren wat de prestaties zijn van het Pilot Algoritme. Ook is het onduidelijk in hoeverre de datakwaliteit en prestaties van de Post COVID Algoritmes robuust zijn en vergelijkbaar zijn met de resultaten van het Algoritme zoals vastgelegd in de algoritmedocumentatie.                                                                                               | IPL dient te onderbouwen waarom het Algoritme, dat deels getraind is op een pre-COVID periode, ook kan worden toegepast op een post-COVID periode waarin de arbeidsmarkt erg is veranderd, thuiswerken is bijvoorbeeld veel gebruikelijker. Dit kan bijvoorbeeld door de prestaties van het Algoritme te toetsen op een out-of-time post-COVID periode.                                                                                                                                                                                                                                                   |
|           | Daarnaast ontbreken heldere procedures voor het hertrainen van het Algoritme in de algoritmedocumentatie, en ontbreekt de vastlegging van relevante documentatie over een hertrainde versie van het Algoritme.                                                                                                                                                                                                     | Indien de prestaties van het Algoritme voor die out-of-time post-COVID periode niet veel afwijken van de prestaties zoals gedocumenteerd in de algoritmedocumentatie, dan kun je daaruit concluderen of het effect van de COVID periode mogelijk beperkt is.<br><br>Daarnaast dient COA met IPL vast te stellen wat er minimaal gedocumenteerd zal worden bij hertrainen van het Algoritme. Hierbij zou tenminste moeten blijken wat er geanalyseerd is rond datakwaliteit, model performance en robuustheid, zodat op basis hiervan COA kan concluderen of het hertrainde Algoritme voldoende presteert. |
| 4-<br>CBS | De WIS laat de gebruiker drie scores zien. Een huidig casussen niveau, toekomstig casussen niveau en een gecombineerde score. Het is onduidelijk hoe deze informatie de Algoritmegebruiker ondersteunt.<br><br>De regio met de hoogste score zal altijd een WIS van 10 krijgen. Hierdoor lijkt het Algoritme altijd een geschikte arbeidsmarktregio gevonden te worden ook al is de geschatte baankans hier klein. | IPL dient samen met COA te concluderen of de WIS de juiste informatie levert aan de Algoritmegebruiker. Pas waar nodig de WIS berekening aan, of beargumenteer waarom de huidige WIS berekening wordt toegepast.                                                                                                                                                                                                                                                                                                                                                                                          |



29 april 2024  
KD/wk/24-0725

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>De WIS levert hiermee niet de informatie aan de Algoritmegebruiker om de robuustheid van het advies van het Algoritme in te schatten.</p> <p>Daarnaast is het onvoldoende gedocumenteerd waarom de keuze is gemaakt voor een toekomstige casussen niveau in de WIS, waarmee het belang van toekomstige casussen mogelijk boven de Vergunninghouder in kwestie stelt, en of de benodigde data voldoende representatief is voor de berekening van een toekomstige casussen niveau.</p> | <p>Verder dient het IPL in overleg met COA de algoritmedocumentatie te verrijken met een beschrijving waarom de keuze is gemaakt voor een fictief toekomstige casussen niveau, en een analyse die beschrijft dat de benodigde data voldoende representatief is voor de berekening van dit niveau. Waar nodig kan IPL/COA de WIS berekening wederom aanpassen, of beargumenteren waarom de huidige WIS berekening wordt toegepast.</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Tabel 3 geeft een overzicht weer van bevindingen die zijn gesloten, maar waar Deloitte nog een aanbeveling heeft.

**Tabel 3: Overzicht van overige Deloitte aanbevelingen (zonder bevinding)**

| Nr. | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7   | Stel aan de hand van de eerste monitoringresultaten vast te of de verdelingen stabiel genoeg zijn om te werken met statistisch significante verschillen, of dat drempelwaarden beter op een andere manier kunnen worden vastgesteld om veel verklaarbare overschrijdingen van drempelwaarden te voorkomen. Daarnaast is de aanbeveling om aan de hand van de eerste monitoringresultaten vast te stellen welke opvolgingstermijnen gekoppeld zouden moeten worden aan de verschillende monitoring levels. |
| 8   | Geef voldoende aandacht aan het effect van het Algoritme in de Pilot om de accuraatheid van het Algoritme in de praktijk objectief te toetsen.                                                                                                                                                                                                                                                                                                                                                            |
| 9   | Indien wordt besloten daadwerkelijk te starten met de Pilot, rond dan de publicatie van het algoritmeregister af en plaats een aankondiging van de pilot op COA.nl.                                                                                                                                                                                                                                                                                                                                       |



29 april 2024  
KD/wk/24-0725

## 2. Inhoudsopgave

|                                              |           |
|----------------------------------------------|-----------|
| <b>1. Managementsamenvatting</b>             | <b>4</b>  |
| <b>2. Inhoudsopgave</b>                      | <b>11</b> |
| <b>3. Introductie en achtergrond</b>         | <b>12</b> |
| 3.1. Introductie van het Algoritme           | 12        |
| 3.2. Betrokkenen Algoritme                   | 13        |
| 3.3. Ontvangen documenten                    | 14        |
| 3.4. Scope van de werkzaamheden van Deloitte | 14        |
| <b>4. Analyse activiteiten</b>               | <b>16</b> |
| 4.1. Governance en proces                    | 16        |
| 4.2. Doel, gebruik en inbedding              | 24        |
| 4.3. Dataverzameling en datakwaliteit        | 26        |
| 4.4. Bias en datagebruik                     | 31        |
| 4.5. Algoritme input en componenten          | 36        |
| 4.6. Algoritme theorie                       | 39        |
| 4.7. Algoritme uitvoer                       | 41        |
| 4.8. AVG                                     | 43        |
| <b>5. Disclaimers</b>                        | <b>48</b> |
| <b>6. Referenties</b>                        | <b>49</b> |



29 april 2024  
KD/wk/24-0725

### 3. Introductie en achtergrond

Het Centraal Orgaan opvang asielzoekers ("COA") heeft Deloitte Risk Advisory B.V. ("Deloitte") gevraagd om een analyse uit te voeren van het algoritme 'kansrijke koppeling' ("Initiële Algoritme"). Het COA wil dit algoritme in gebruik nemen om de koppeling van vergunningshouders aan gemeentes te verbeteren. Deloitte heeft ook in het verleden analyses uitgevoerd op dit Initiële Algoritme:

- **Analyse 1:** Deloitte heeft in de zomer van 2022 werkzaamheden uitgevoerd als onderdeel van de onderzoeksfase van het Initiële Algoritme. Die werkzaamheden hebben geresulteerd in een rapport met bevindingen, bestaande uit observaties en aanbevelingen. Op basis van deze observaties en aanbevelingen heeft COA het Initiële Algoritme aangepast ("Algoritme") in voorbereiding op de pilotfase ("Pilot").
- **Analyse 2:** In 2023 en 2024 heeft Deloitte het Algoritme in voorbereiding op de pilotfase geanalyseerd. De werkzaamheden die Deloitte heeft uitgevoerd op het Algoritme in Analyse 2 staan beschreven in de getekende overeenkomst met kenmerk KS/js/23-0546 en bestaan uit twee stappen:
  - **Stap 1:** Tijdens de eerste stap van Analyse 2 ("Stap 1") heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de volgende aanbevelingen uit Analyse 1: 1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, CBS-1, CBS-3, CBS-4 en CBS-5. De werkzaamheden van Stap 1 zijn afgerond in juli 2023 en hebben geresulteerd in een Tussentijdse Rapportage met kenmerk KD/avl/23-0957.
  - **Stap 2:** Tijdens de tweede stap van Analyse 2 ("Stap 2") heeft de nadruk van de werkzaamheden van Deloitte gelegen op de opvolging van de aanbevelingen die niet zijn geanalyseerd in Stap 1, en de opvolging van aanbevelingen die niet zijn gesloten in de Tussentijdse Rapportage na afronding van Stap 1.

Dit rapport presenteert de gecombineerde resultaten van Stap 1 en Stap 2 ("Rapport").

#### 3.1. Introductie van het Algoritme

Het COA beschrijft dat het doel van het Algoritme is om te komen tot een beter regioadvies voor vergunninghouders vergeleken met het huidige proces, te meten via een verhoogde arbeidsmarktparticipatie van vergunningshouders. Het Algoritme maakt gebruik van machine learning technologie en persoonsgegevens om een advies te geven aan de werknemers van het COA. Het advies bestaat uit een aanbeveling voor welke arbeidsmarktregio('s) de vergunningshouder een verhoogde arbeidsmarktparticipatie heeft. Als persoonsgegevens worden bijvoorbeeld opleiding, leeftijd en burgerlijke staat gebruikt. Naar aanleiding van bevinding 20 van Deloitte in de Tussentijdse Rapportage is COA voornemens geen bijzondere persoonsgegevens zoals etniciteit en religie meer te gebruiken; dit is nog niet geformaliseerd en zal in april 2024 ter advies worden voorgelegd aan de directie van COA (zie Sectie 4.8.2). Een volledige lijst van de gebruikte gegevens is beschikbaar in de algoritmedocumentatie van het COA. De algoritmedocumentatie bestaat uit meerdere documenten die zijn beschreven in sectie 3.3 van dit rapport.

Het Algoritme bevindt zich nu in de "pre-pilot" fase. Tijdens de pre-pilot start de technische en organisatorische implementatie van het Algoritme. Ter afronding van de pre-pilot, is COA voornemens dit Rapport te gebruiken om aan de COA directie voor te leggen al dan niet met de pilot van het Algoritme te starten. Na goedkeuring zal, tijdens de pilot, het Algoritme voor het eerst gebruikt en gemonitord worden. Na de pilot neemt het COA een definitief besluit om het Algoritme te borgen in het reguliere proces.

Het Algoritme is door het COA in samenwerking met het Immigration Policy Lab ("IPL") ontwikkeld. Het IPL is een onderzoeksinstituut gelieerd aan Stanford University en de ETH Zürich. Het Algoritme van het IPL wordt al gebruikt in Zwitserland. In Zweden, Canada en de Verenigde Staten vindt onderzoek naar de mogelijke voordelen van het Algoritme van het IPL plaats, vergelijkbaar met het onderzoek dat door het COA wordt verricht.



29 april 2024  
KD/wk/24-0725

Het Algoritme is ontwikkeld met data die beschikbaar is gesteld door het COA en het CBS. De data is voor het IPL beschikbaar via een data omgeving van het CBS waar alleen het IPL toegang tot heeft.

3.2. Betrokkenen Algoritme

Tabel 3 geeft de betrokkenen bij het Algoritme weer.

Tabel 3: Overzicht van betrokkenen

| Rol                                                 | Beschrijving                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Algoritme eigenaar                                  | De eigenaar van het Algoritme is de werknemer van het COA die eindverantwoordelijk is voor de beslissingen in een specifieke fase. Het COA heeft de ontwikkeling van het Algoritme opgedeeld in de onderzoeksfase, pre-pilot en borging van het reguliere proces.                                                                                                                                                                                                                                                                                                    |
| Algoritmegebruiker                                  | De gebruiker is een werknemer van het COA die het Algoritme gebruikt. Het Algoritme geeft een regioadvies, vervolgens nemen de gebruikers de uiteindelijke beslissing omtrent de koppeling van een vergunninghouder aan een gemeente.                                                                                                                                                                                                                                                                                                                                |
| Centraal Bureau voor de Statistiek ("CBS")          | Het CBS levert in samenwerking met het COA data aan het IPL.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| COA                                                 | Het Algoritme valt onder de verantwoordelijkheid van het COA. In de onderzoeksfase is de Directie Vakontwikkeling en -ondersteuning verantwoordelijk.                                                                                                                                                                                                                                                                                                                                                                                                                |
| Immigration Policy Lab ("IPL")                      | Het IPL heeft het Algoritme ontwikkeld en zorgt voor de monitoring en (door-)ontwikkeling van het Algoritme.                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Landelijke Regietafel Migratie & Integratie ("LRT") | De Landelijke Regietafel wordt als gremium ingezet om externe belanghebbenden te spreken (zoals Vereniging van Nederlandse Gemeenten, en Ministeries Justitie en Veiligheid, Sociale Zaken en Werkgelegenheid en Binnenlandse Zaken en Koninkrijksrelaties) om de brede maatschappelijke gevolgen van het Algoritme voorafgaand aan het gebruik te onderzoeken en te documenteren.<br><br>COA heeft in Stap 2 besloten om alleen de externe belanghebbenden Ministerie van Sociale Zaken en Werkgelegenheid en de Vereniging van Nederlands Gemeentes te informeren. |
| Vergunningshouders                                  | De vergunningshouders zijn asielzoekers die een verblijfsvergunning hebben ontvangen. De vergunningshouders zijn de data subjecten. De uitkomsten van het Algoritme zullen betrekking op hen hebben.                                                                                                                                                                                                                                                                                                                                                                 |
| Stuurgroep                                          | De Stuurgroep binnen COA is het goedkeuringsorgaan dat beslissingen neemt over het Algoritme. Het gaat hierbij bijvoorbeeld om goedkeuring te starten met de pilot of advies over het wel of niet inzetten van bijzondere persoonsgegevens.                                                                                                                                                                                                                                                                                                                          |



29 april 2024  
KD/wk/24-0725

### 3.3. Ontvangen documenten

Deloitte voert de analyse uit op basis van documenten die zijn aangeleverd door het COA als onderdeel van de algoritmedocumentatie tijdens zowel Stap 1 en Stap 2 van Analyse 2. De hoofddocumenten van Stap 1 zijn:

- NLD\_GeoMatch\_TechnicalDocumentation\_20230418.pdf (hierna: "Technische Documentatie 1").
- Annex\_GeoMatchNetherlands\_PilotMonitoringPlan\_20230416.pdf (hierna: "Monitoringsplan 1").
- 230419 Audit 2.1 Deloitte project Kansrijke Koppeling m.b.v. AI.xlsx (hierna: "Bevindingen Status Excel 1").
- Annex\_GeoMatchNetherlands\_UpdatedBacktests\_060222 (hierna: "Backtest Rapport").
- Annex DARE chart.PNG (hierna: "DARE chart").
- Annex Stakeholdersession.docx (hierna: "Stakeholdersession 1").

De documenten van Stap 2 zijn:

- Annex Stakeholdersession.docx (hierna: "Stakeholdersession 2").
- COA IPL Audit 2.b (hierna: "Bevindingen Status Excel 2").
- Service Level Agreement COA - Stanford University 1.0 definitief.pdf (hierna: "SLA").
- NLD\_GeoMatch\_TechnicalDocumentation\_20240216.pdf (hierna: "Technische Documentatie 2").
- GeoMatchNetherlands\_MonitoringPlan\_2024-02-26.pdf (hierna: "Monitoringsplan 2").
- PID Kansrijke koppeling m.b.v. AI PowerPoint.pdf (hierna: "Project Initiation Document").
- 240312 Communicatieplan Kansrijke koppeling met AI v1.1.pdf (hierna: "Communicatieplan").
- Werkinstructie v0.9.docx (hierna: "Werkinstructie").
- Printscreen IBIS.png (hierna: "Printscreen IBIS").
- geomatch\_nld\_variables\_20240305.xlsx (hierna: "Feature Lijst").
- 240320-invullijst-algoritmeregister-Kansrijke-Koppeling-v0.2-concept (hierna: "Concept Invullijst Algoritmeregister").

De documenten Bevindingen Status Excel 1 en Bevindingen Status Excel 2 zijn geen onderdeel van de algoritmedocumentatie van het COA, hierin is enkel opgenomen hoe COA opvolging heeft gegeven aan de gemaakte observaties en aanbevelingen.

Deloitte heeft niet gecontroleerd of de aangeleverde set van informatie compleet is.

### 3.4. Scope van de werkzaamheden van Deloitte

Tijdens Analyse 1 is een werkplan is opgesteld door Deloitte in samenwerking met het COA met daarin de analyseactiviteiten. Deze analyse activiteiten zijn beschreven in Appendix 1 van het rapport met kenmerk KD/lr/22-1275 waarin de werkzaamheden van Deloitte van Analyse 1 zijn vastgelegd. Deze analyseactiviteiten hebben destijds geleid tot 20 bevindingen vanuit Deloitte en 6 bevindingen vanuit het CBS.

Voor Analyse 2 is deze lijst met bevindingen als uitgangspunt genomen, en is aan de hand van de aangepaste algoritmedocumentatie per bevinding beoordeeld wat de huidige status hiervan is (open, gesloten, conditioneel gesloten). Daarnaast kan de beoordeling van de aangepaste algoritmedocumentatie leidt tot nieuwe bevindingen. De status van de bestaande bevindingen en eventuele nieuwe bevindingen zijn opgenomen in dit Rapport.

Tijdens de analyse heeft Deloitte gebruikt gemaakt van een vraag- en antwoordenlijst voor het COA en het IPL ("Q&A") [2].



29 april 2024  
KD/wk/24-0725

Deloitte heeft tijdens het uitvoeren van de analyses geen toegang gehad tot de data en code van het Algoritme. Hierdoor was het voor Deloitte niet mogelijk om, als onderdeel van de dimensie 'Algoritme uitvoer' te analyseren hoe de stabiliteit van het Algoritme wordt afgezet tegen overfitting en generalisatie normen. Deloitte heeft daardoor ook niet de reproduceerbaarheid en stabiliteit van het Algoritme kunnen analyseren.

In overleg met COA zijn de analyseactiviteiten uit de dimensies "Security en back-ups" en "Implementatie, change controls en audit logs" buiten de reikwijdte van de opdracht geplaatst, evenals de verwerkersovereenkomst.



29 april 2024  
KD/wk/24-0725

#### 4. Analyse activiteiten

De paragrafen in deze sectie beschrijven voor elk van de volgende acht dimensies de eventuele bevindingen die daaruit zijn voortgekomen tijdens Analyse 1 van Deloitte.<sup>1</sup> Per bevinding wordt aangegeven hoe hier opvolging aan gegeven is door COA, en of daarmee de bevinding gesloten kan worden of (gedeeltelijk) open blijft staan.

- 4.1 Governance en proces
- 4.2 Doel, gebruik en inbedding
- 4.3 Dataverzameling en datakwaliteit
- 4.4 Bias en datagebruik
- 4.5 Algoritme input en componenten
- 4.6 Algoritme theorie
- 4.7 Algoritme uitvoer
- 4.8 AVG

Elke bevinding bestaat uit een observatie en een aanbeveling.

##### 4.1. Governance en proces

Als onderdeel van de analyse van de dimensie 'governance en proces' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van de governance van het Algoritme (rollen, verantwoordelijkheden, deskundigheid, betrokkenheid en algoritmelevenscyclus) rondom het gebruik van algoritmes, aan de hand van het 'three lines of defense model' voor algoritmerisico.

##### 4.1.1. Bevinding 1

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De vaststelling van rollen en verantwoordelijkheden in de gehele algoritmelevenscyclus is beperkt uitgewerkt in de algoritmedocumentatie voor interne en externe belanghebbenden die betrokken zijn bij de ontwikkeling van het Algoritme.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met de rollen en verantwoordelijkheden van interne en externe actoren die (eind)verantwoordelijk is (zijn) voor bepaalde werkzaamheden.

---

<sup>1</sup>Een gedetailleerd overzicht van de analyse activiteiten uit Analyse 1 die hebben geleid tot deze bevindingen is te vinden in het rapport met kenmerk KD/Ir/22-1275.



29 april 2024  
KD/wk/24-0725

### Stap 1 van Analyse 2

In de algoritmedocumentatie van Stap 1 heeft COA het document DARE chart toegevoegd aan de algoritmedocumentatie. Dit bestand omschrijft de rollen, COA stakeholders die betrokken zullen zijn voorafgaand en/of tijdens de pilot, en de stakeholder impact op het project. Tijdens de uitvoering van Stap 1 heeft COA hier stakeholders aan toegevoegd, namelijk CBS als data provider en IPL als ontwikkelaar en beheerder van het Algoritme.

Met deze beschrijving van de rollen en verantwoordelijkheden is deze bevinding gesloten.

#### 4.1.2. Bevinding 2

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De externe belanghebbenden worden pas tijdens de pre-pilot via de Landelijke Regietafel Migratie & Integratie ("LRT") betrokken. Hiermee ontstaat het risico dat fundamentele veranderingen later in het project worden doorgevoerd en/of hun belangen niet volledig worden gewaarborgd, wat tot een risico op vooringenomenheid van het Algoritme kan leiden.
- **Aanbeveling:** Betrek alle belanghebbenden zo vroeg mogelijk in de pre-pilot bij de besluitvorming door te vragen naar de waarden, uitgangspunten en belangen die voor hen belangrijk zijn bij de toepassing van het Algoritme.

### Stap 1 van Analyse 2

In de algoritmedocumentatie van Stap 1 heeft COA het document Stakeholdersession 1 opgenomen. Dit document bevat de notulen van een twee uur durende stakeholdersessie op 7 december 2022, waarin externe stakeholders zijn geïnformeerd over het project en gevraagd om hun ideeën, waardes en principes in de context van het Algoritme. Verder heeft COA aangegeven voornemens te zijn voorafgaand aan de pilot nog een stakeholder sessie te organiseren waarin meer details worden gedeeld.

De externe stakeholder Vereniging van Nederlandse Gemeenten (hierna: "VNG") was niet aanwezig bij bovengenoemde stakeholdersessie. COA heeft aangegeven dat dezelfde sessie met VNG is gehouden op 19 december 2022, en dat hieruit geen additionele punten naar voren zijn gekomen.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding gesloten kan worden indien COA in de algoritmedocumentatie opneemt dat bovenstaande sessie met VNG heeft plaatsgevonden en wat de uitkomsten daarvan waren.

### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 heeft COA in het document Stakeholdersession 2 de volgende toevoeging gedaan:

*"At a later time (December 19, 2022) there was an additional meeting with 5.1.2 e from the Vereniging Nederlandse Gemeenten. There is no additional information to add to the minutes based on this meeting."*

Met het toevoegen van de vermelding van de sessie met VNG is de aanbeveling van Deloitte opgevolgd, en daarmee is de bevinding gesloten.



29 april 2024  
KD/wk/24-0725

#### 4.1.3. Bevinding 3

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Vergunningshouders, belangenbehartigers van vergunningshouders en een diverse groep burgers, met verschillende perspectieven op de maatschappij, worden niet geconsulteerd voorafgaand aan de pilot van het Algoritme.
- **Aanbeveling:** Betrek vergunningshouders, belangenbehartigers van vergunninghouders en een diverse groep burgers bij de ontwikkeling van het Algoritme om te bevestigen dat er een maatschappelijk draagvlak voor het Algoritme is en het Algoritme ethisch en verantwoord kan worden toegepast.

#### Stap 1 van Analyse 2

Op verzoek van COA is deze bevinding niet geanalyseerd in Stap 1.

#### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 heeft COA het document Stakeholdersession 2 opgenomen. In dit document staat beschreven welke ethische waarde per aanwezige betrokkene (o.a. vanuit afgevaardigden vanuit het ministerie van Sociale Zaken en Werkgelegenheid, ministerie Binnenlandse Zaken en Koninkrijksrelaties, Vluchtelingenwerk Nederland, Ministerie Justitie en Veiligheid, Vereniging Nederlandse Gemeentes, IPL en COA) als het belangrijkste wordt gezien. In de Q&A heeft COA vervolgens toegelicht hoe de ethische waarden rechtvaardigheid, zelfbeschikking, transparantie en de menselijke factor, zijn meegenomen in het ontwikkelproces.

Deloitte observeert dat er echter geen directe input of betrokkenheid is vanuit Vergunningshouders, belangenbehartigers van vergunninghouders en een diverse groep burgers op de door COA gemaakte keuzes, zoals de keuzes over hoe invulling gegeven is aan de ethische waarden rechtvaardigheid, zelfbeschikking, transparantie en menselijke factor. Het risico op beperkt maatschappelijk draagvlak blijft hiermee bestaan, waarmee de bevinding open blijft staan.



29 april 2024  
KD/wk/24-0725

De aanbeveling blijft om Vergunningshouders, belangenbehartigers van vergunninghouders en een diverse groep burgers bij de verdere ontwikkeling van het Algoritme en de evaluatie van de Pilot te betrekken. Hierdoor kan COA bevestigen dat er een maatschappelijk draagvlak voor het Algoritme is en het Algoritme ethisch en verantwoord kan worden toegepast.

| Nr. | Categorie  | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Aanbeveling                                                                                                                                                                                                                                           |
|-----|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3   | Governance | In 2022 heeft een sessie plaatsgevonden met verschillende belanghebbenden. Hierin is vastgesteld welke ethische waarden de belanghebbenden het belangrijkste vonden in de context van het Algoritme. COA heeft deze ethische waarden als uitgangspunt genomen voor de verdere ontwikkeling van het Algoritme. Echter, er is daarna geen directe input op de door COA gemaakte keuzes gevraagd of terugkoppeling gegeven aan belanghebbenden over hoe de voor hun belangrijke ethische waardes zijn meegenomen tijdens de ontwikkeling van het Algoritme. | Betrek de eerder betrokken belanghebbenden en toets of de belangrijkste waarden van deze belanghebbenden voldoende terugkomen in het Algoritme, en of ze vertrouwen hebben dat het Algoritme ethisch en verantwoord kan worden toegepast in de Pilot. |

#### 4.1.4. Bevinding 4

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De besluitvorming en samenwerkingsvormen met externe belanghebbenden, zoals de LRT, zijn beperkt uitgewerkt in de algoritmedocumentatie of nog niet uitgevoerd, zoals het overleg met de LRT. Onverenigbare standpunten tijdens de samenwerkingsvormen kunnen leiden tot vertraging en (tijdelijke) stillegging van het project.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met een verduidelijking over de rol van belanghebbenden, en het onderscheid dat eventueel wordt gemaakt tussen de consulterende en/of besluitvormende rol van de belanghebbenden.

#### Stap 1 van Analyse 2

Op verzoek van COA is deze bevinding niet geanalyseerd in Stap 1.

#### Stap 2 van Analyse 2

COA heeft aangegeven in de Bevindingen Status Excel 2 dat de externe belanghebbenden SZW en VNG enkel geïnformeerd zullen worden, maar geen consulterende of besluitvormende rol zullen krijgen. Dit beperkt het risico op onverenigbare standpunten.

Deloitte observeert dat deze informatie niet is meegenomen is in de algoritmedocumentatie. Het risico ontstaat dat er in de toekomst onduidelijkheden ontstaan over de rollen en verantwoordelijkheden van externe belanghebbenden.



29 april 2024  
KD/wk/24-0725

Deze bevinding kan worden gesloten indien de rollen en verantwoordelijkheden van de externe belanghebbenden worden opgenomen in de algoritmedocumentatie zodat de rol van deze partijen voor iedereen duidelijk is.

| Nr. | Categorie    | Observatie                                                                                                                                                                               | Aanbeveling                                                                                                                                                                                                                                   |
|-----|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4   | Documentatie | COA heeft de rollen en verantwoordelijkheden van belanghebbenden geïdentificeerd. Echter, deze rollen en verantwoordelijkheden komen onvoldoende naar voren in de algoritmedocumentatie. | Verrijk de algoritmedocumentatie met een verduidelijking over de rol van belanghebbenden, namelijk dat de externe belanghebbenden SZW en VNG enkel geïnformeerd zullen worden, maar geen consulterende of besluitvormende rol zullen krijgen. |

#### 4.1.5. Bevinding 5

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Het COA is afhankelijk van externe partijen voor het gebruik en beheersing van het Algoritme. Er zijn geen uitgewerkte overeenkomsten wat de diensten en verantwoordelijkheden van het IPL en het CBS definieert, en het COA beschikt zelf nog niet over de nodige expertise.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met de rollen en verantwoordelijkheden van interne en externe actoren die (eind)verantwoordelijk is (zijn) voor bepaalde werkzaamheden.

#### Stap 1 van Analyse 2

COA heeft aangegeven een service level agreement tijdens Stap 2 te delen om opvolging te geven aan deze bevinding.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de service level agreement.

#### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 heeft COA de SLA aangeleverd die is opgesteld door COA en welke de afspraken beschrijft tussen COA en Stanford University. De SLA beschrijft onder andere de verantwoordelijkheden van beide partijen:

- **COA:** beschikbaar stellen van twee servers met Docker platform, en maandelijks leveren van data t.b.v. monitoring.
- **Stanford University:** trainen van de AI toepassing aan de hand van historische data van COA en CBS, leveren van gebruiksrecht op het Algoritme, zonder financiële verplichtingen, en actief monitoren van de AI toepassing in lijn met het opgestelde monitoringsplan.

Daarnaast bevat de SLA o.a. een lijst met contactpersonen, een twee jaarlijkse evaluatie van de SLA die door COA zal worden geïnitieerd, een beschrijving dat beide partijen zullen voldoen aan Nederlandse en Europese wetgeving, een beknopte beschrijving van de eerstelijns- (service desk), tweedelijns- (incident management) en derdelijns (change management) support van beide partijen, de meldplicht datalekken voor beide partijen, en een beknopte beschrijving van overlegstructuren, rapportage frequentie en escalatie mogelijkheid.

De SLA neemt echter niet alle onduidelijkheden weg over de verdere samenwerking:

- De SLA is in het Nederlands geschreven, waarmee deze niet leesbaar is voor IPL.
- De SLA beschrijft geen looptijd van de overeenkomst, waarmee onduidelijk is voor welke periode de samenwerkingsovereenkomst zal gelden, of onder welke voorwaarden deze samenwerking kan worden stopgezet.



29 april 2024  
KD/wk/24-0725

- Daarnaast beschrijft de SLA de rollen en verantwoordelijkheden van het CBS niet, waardoor ook deze afhankelijkheid onduidelijk blijft. In de SLA aangegeven dat het IPL een overeenkomst heeft gesloten met het CBS voor het verrijken van de gegevens om het Algoritme in de Nederlandse context te kunnen laten werken. Echter, in hoeverre deze contractvorm tussen IPL en CBS ook zekerheid biedt aan COA en daarmee continuïteit van het Algoritme kan garanderen is onduidelijk, omdat de details van de overeenkomst niet worden beschreven in de SLA.
- De SLA beschrijft niet welke Nederlandse- en Europese wetgeving van toepassing is, waarmee onduidelijk is aan welke wetgeving voldoen dient te worden en wie hiervoor verantwoordelijk is.
- Als laatste is de SLA eenzijdig opgesteld door COA en nog niet ter goedkeuring aan Stanford University voorgelegd en/of ondertekend, waarmee de afspraken vanuit de SLA momenteel nog niet geformaliseerd zijn en het risico blijft bestaan dat beide partijen hier niet hetzelfde mee om zullen gaan.

Deloitte observeert dat de SLA nog niet is ondertekend door Stanford University, en dat de duur van de overeenkomst niet is vastgelegd, waardoor onduidelijkheid blijft bestaan over de verdere samenwerkingsafspraken tussen COA en het IPL, waarmee de bevinding open blijft. Verder observeert Deloitte dat COA voor de gemaakte keuzes, conclusies over de kwaliteit van de data en het Algoritme, en interpretatie van de monitoringresultaten, vaak in eerste instantie het IPL vraagt dit aan te leveren. COA zou echter vanuit haar maatschappelijke verantwoordelijkheid moeten kunnen onderbouwen waarom keuzes zijn gemaakt en of die bij haar visie en waarden past. Ook zou COA zelf conclusies moeten trekken op basis van de door IPL aangeleverde performance tests, datakwaliteitschecks en monitoringresultaten, om op basis daarvan te bepalen of op een verantwoorde manier gestart kan worden met de Pilot.

Deloitte beveelt aan om aan de stuurgroep van COA te wijzen op haar verantwoordelijkheid, en haar de vraag voor te leggen of ze (zowel nu voor de start van de Pilot als naar de toekomst toe bij het monitoren van het Algoritme) voldoende handvatten en kennis hebben om zelfstandig conclusies te trekken over gemaakte keuzes, en om te beoordelen of de datakwaliteit en model performance voldoende aansluiten op hun verwachtingen.

| Nr. | Categorie  | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5   | Governance | <p>Het COA is afhankelijk van externe partijen voor het gebruik en beheersing van het Algoritme. De opgestelde SLA tussen Stanford University en COA is niet ondertekend en de duur van de overeenkomst is niet vastgelegd, waardoor onduidelijkheid blijft bestaan over de verdere samenwerkingsafspraken tussen COA en het IPL.</p> <p>Verder observeert Deloitte dat COA voor de gemaakte keuzes, conclusies over de kwaliteit van de data en het Algoritme, en interpretatie van de monitoringresultaten, vaak in eerste instantie naar het IPL kijkt.</p> | <p>Formaliseer de samenwerkingsafspraken tussen COA en IPL door de tijdsduur van de overeenkomst op te nemen in de opgestelde SLA en deze door beide partijen te ondertekenen.</p> <p>Maak duidelijke afspraken tussen COA en IPL over de benodigde handvatten die IPL zou moeten bieden om COA te begeleiden in het maken van belangrijke besluiten rondom het Algoritme. IPL zou moeten ondersteunen in de vertaling van de tabellen met datakwaliteit en model performance resultaten naar wat dit daadwerkelijk voor het gebruik van het Algoritme betekent, zodat COA een weloverwogen besluit kan nemen over de start van de Pilot.</p> |



29 april 2024  
KD/wk/24-0725

COA zou echter vanuit haar maatschappelijke verantwoordelijkheid moeten kunnen onderbouwen waarom keuzes zijn gemaakt en of die bij haar visie en waardes past.

Daarnaast zou IPL als kennishouder uit zichzelf contact moeten zoeken met COA over of gemaakte (ontwerp)keuzes bij COA's visie en waarde passen.

Ook zou COA zelf conclusies moeten trekken op basis van de door IPL aangeleverde performance tests, datakwaliteitschecks en monitoringresultaten, om op basis daarvan te bepalen of op een verantwoorde manier gestart kan worden met de Pilot.

#### 4.1.6. Bevinding 6

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Er is gebrek aan beschikbaarheid van analyses en gemaakte (ontwerp)keuzes binnen het Algoritme.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met de uitgevoerde analyses, waar mogelijk de uitkomsten van de analyses, de gemaakte ontwerpkeuzes en de mechanismen om gedefinieerde werkbegrippen te waarborgen.

#### Stap 1 van Analyse 2

In de Technische Documentatie 1 van Stap 1 heeft COA in sectie 6.c beschreven welke datakwaliteitschecks en analyses zijn uitgevoerd (zie bevinding 11), waarbij de resultaten van deze checks zijn opgenomen in Appendix 2. Verder heeft COA meerdere analyses toegevoegd in de Technische Documentatie 1 rondom uitlegbaarheid (zie bevinding 13 en 14). Daarnaast beschrijft COA in sectie 2, 3 en 4 dat de methodologie ontwerpkeuzes bevat, zoals "Mapping" functies en optimalisatie criteria (zie bevinding 16). De uitkomsten en conclusies van checks en analyses worden niet beschreven. Daarnaast worden ontwerpkeuzes aangegeven, maar de uitkomst en argumentatie van deze ontwerpkeuzes niet beschreven.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de technische documentatie.

#### Stap 2 van Analyse 2

In de Technische Documentatie 2 van Stap 2 heeft COA meer informatie toegevoegd over de uitgevoerde analyses:

Appendix 1 en 2 bevatten respectievelijk de resultaten van de uitgevoerde performance tests (o.a. accuracy, RMSE, etc.) en datakwaliteitschecks (o.a. missende waarden, aantal unieke waarden, een samenvatting van de dataset, etc.). COA heeft per tabel met resultaten een korte beschrijving toegevoegd over de uitgevoerde tests en hoe de resultaten tot stand zijn gekomen. Conclusies ontbreken echter over de uitgevoerde tests, zoals of het aantal missende waarden per variabele in Appendix 2.b naar verwachting is, of dit verklaarbaar is of dat dit mogelijk een probleem vormt.

- Daarnaast is de Technische Documentatie 2 niet geüpdatet met een onderbouwing van de gekozen "Mapping" functies (zie hoofdstuk 3 van de Technische Documentatie 2) en optimalisatie criteria (zie hoofdstuk 4.a van de Technische Documentatie 2).



29 april 2024  
KD/wk/24-0725

Deloitte observeert dat het onduidelijk is waarom bepaalde (ontwerp)keuzes zijn gemaakt (zoals de onderbouwing van de gekozen "Mapping" functies en optimalisatie criteria), en dat conclusies op basis van de uitgevoerde performance tests en datakwaliteitschecks ontbreken. Hierdoor blijft het onduidelijk of de data en het Algoritme van voldoende kwaliteit zijn. Dit brengt het risico met zich mee dat de stuurgroep van COA niet de juiste handvatten heeft om een weloverwogen besluit te kunnen nemen over de start van de Pilot met het Algoritme.

Deloitte beveelt aan de algoritmedocumentatie te verrijken met de onderbouwing waarom bepaalde (ontwerp)keuzes zijn gemaakt en waarom niet voor alternatieve (ontwerp)keuzes is gekozen. Daarnaast beveelt Deloitte aan de conclusies op basis van de uitgevoerde performance tests en datakwaliteitstest inclusief onderbouwing op te nemen in de algoritmedocumentatie.

| Nr. | Categorie  | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6   | Governance | Het is onduidelijk waarom bepaalde (ontwerp)keuzes zijn gemaakt (zoals de onderbouwing van de gekozen "Mapping" functies en optimalisatie criteria), en dat conclusies op basis van de uitgevoerde performance tests en datakwaliteitschecks ontbreken. Hierdoor blijft het onduidelijk of de data en het Algoritme van voldoende kwaliteit zijn. Dit brengt het risico met zich mee dat de stuurgroep van COA niet de juiste handvatten heeft om een weloverwogen besluit te kunnen nemen over de start van de Pilot met het Algoritme. | <p>IPL dient de algoritmedocumentatie te verrijken met de onderbouwing waarom bepaalde (ontwerp)keuzes zijn gemaakt, zoals de gekozen "Mapping" functies en optimalisatie criteria. Vervolgens dient COA te besluiten of deze onderbouwing past bij de sociaal-maatschappelijke context van COA en het Algoritme.</p> <p>IPL moet COA de juiste handvatten bieden zodat COA aan de hand van de gepresenteerde datakwaliteitsmetrieken en performance metrieken een conclusie kan trekken of de datakwaliteit en model performance voldoende zijn om te kunnen starten met de Pilot en waarom dit het geval is. COA dient vervolgens te concluderen of de datakwaliteit en model performance voldoende zijn om te kunnen starten met de Pilot.</p> |

#### 4.1.7. Bevinding 7

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Er is geen monitoringsplan opgesteld, waarin afspraken over het monitoren en beheer gedurende de algoritmelevenscyclus van het Algoritme worden gedefinieerd; evenals de afspraken met het IPL, het CBS en eventueel UWV. Dit brengt als risico dat de monitoringsactiviteiten niet of te weinig uitgevoerd zullen worden.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met een monitoringsplan, waarin bijvoorbeeld activiteiten, toegestane marges, governance en een stopknop gedefinieerd zijn inclusief de frequentie van de activiteiten. Deze documentatie en afspraken dienen op orde te zijn voorafgaand aan het gebruik van het Algoritme in de pilot. In de pilot wil het COA de werking van het Algoritme onderzoeken door het Algoritme te gebruiken onder een representatief deel van de vergunninghouders.



29 april 2024  
KD/wk/24-0725

### Stap 1 van Analyse 2

Het Monitoringsplan 1 beschrijft verschillende 'monitoring levels' ofwel monitoringsactiviteiten. Per monitoring level wordt het doel, benodigde data en de monitoring frequentie beschreven. Daarnaast worden details van de uit te voeren analyses gegeven, met bijbehorende drempelwaarden en vervolgacties. Er is geen invulling gegeven aan de hoogte van de drempelwaarden voor monitoringslevel 1, 4 en 5 (nu beschreven als X% en Y%) en binnen welke termijn de vervolgacties dienen te worden opgevolgd.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding kan worden gesloten indien COA de drempelwaarden vaststelt en opneemt in het Monitoringsplan, en opneemt binnen welke termijn er opvolging gegeven dient te worden aan vervolgacties. Deloitte heeft aangegeven dat de bevinding open blijft en in Stap 2 opnieuw geanalyseerd zal worden op basis van het monitoringplan.

### Stap 2 van Analyse 2

COA heeft het Monitoringplan 2 verrijkt met drempelwaarden in lijn met de aanbeveling van Deloitte. Bij monitoring level 1 wordt bijvoorbeeld een drempelwaarde van 70% gehanteerd, wat betekent dat verwacht wordt dat minimaal 70% van de adviezen van het Algoritme worden overgenomen. Voor monitoring level 4 en 5 wordt gekeken naar statistisch significante verschillen voor de voorspellingen (level 4) en daadwerkelijke uitkomsten (level 5) tussen de behandelingsgroep en controlegroep. Daarmee is deze bevinding gesloten.

Wel beveelt Deloitte aan om aan de hand van de eerste monitoringresultaten vast te stellen of de verdelingen stabiel genoeg zijn om te werken met statistisch significante verschillen, of dat drempelwaarden beter op een andere manier kunnen worden vastgesteld om veel verklaarbare overschrijdingen van drempelwaarden te voorkomen. Daarnaast beveelt Deloitte aan om aan de hand van de eerste monitoringresultaten vast te stellen welke opvolgingstermijnen gekoppeld zouden moeten worden aan de verschillende monitoring levels.

## 4.2. Doel, gebruik en inbedding

Als onderdeel van de analyse van de dimensie 'doel, gebruik en inbedding' is in Analyse 1 het volgende uitgevoerd:

- Analyseren of het COA de waarden van de relevante belanghebbenden (zoals in sectie 2.4 van de AIIA geïdentificeerd, inclusief het COA zelf) op een juiste, eenduidige en adequate wijze heeft vastgelegd en of is vastgelegd hoe deze worden meegenomen in het gebruik van het Algoritme.

### 4.2.1. Bevinding 8

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De doelstellingen worden onvoldoende concreet uitgewerkt in lange termijn prestatie-indicatoren. Verder zijn er geen acceptabele toleranties en (fout)marges gedefinieerd op zowel korte- als lange termijn doelstellingen.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met meetbare (kwantitatieve) kwaliteitseisen van het Algoritme, voorafgaand aan de pre-pilot.

### Stap 1 van Analyse 2

In de algoritmedocumentatie van Stap 1 heeft COA het Monitoringsplan 1 opgenomen zoals beschreven onder bevinding 7. COA heeft tijdens Stap 1 aangegeven drempelwaarden vast te stellen en op te nemen in de finale versie van het Monitoringsplan. Het Monitoringsplan wordt verder geanalyseerd onder bevinding 7. Er is niet beschreven in de algoritmedocumentatie hoe het Algoritme aan het eind van de pilot geëvalueerd wordt en wanneer COA de pilot als geslaagd ziet.



29 april 2024  
KD/wk/24-0725

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van het Monitoringsplan en/of de algoritmedocumentatie.

### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 is het Project Initiation Document opgenomen. In dit document worden drie doelstellingen beschreven, op het vlak van gebruik, effect en cultuur. De doelstellingen beschrijven wanneer de Pilot een succes is. Twee van de doelstellingen zijn kwantitatief meetbaar en betreffen het gebruik en het effect van het Algoritme. Daarnaast beschrijft de cultuur doelstelling dat de pilot een succes is voor COA wanneer werknemers positief zijn over datagedreven werken, beslissingsondersteunende instrumenten en de nieuwe manier van werken. In de Q&A heeft COA bevestigd dat deze doelstellingen leidend zijn. De doelstellingen zijn vastgesteld door de stuurgroep in December 2023.

Met het toevoegen van de drie doelstellingen op het vlak van gebruik, effect en cultuur is de aanbeveling van Deloitte opgevolgd, en is de bevinding gesloten voor de Pilot. Het advies voor de Pilot evaluatie is wel om voldoende aandacht te geven aan het effect van het Algoritme om de accuraatheid van het Algoritme in de praktijk objectief te toetsen.

#### 4.2.2. Bevinding 9

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Er zijn geen processen en/of mechanismen ingesteld om algemene feedback op te halen ten aanzien van het Algoritme. Verder komt het doel van het gebruikersoverleg niet overeen met het doel van het COA om interne en externe feedback te ontvangen.
- **Aanbeveling:** Stel een aanspreekpunt aan waar derden en Algoritmegebruikers terecht kunnen met zowel hun vragen/zorgen, als algemene feedback. Wijs daarnaast een eindverantwoordelijke aan binnen het COA die zorgt voor afstemming met het IPL over (mogelijke) aanpassingen aan de hand van ontvangen feedback.

### Stap 1 van Analyse 2

Tijdens Stap 1 beschrijft COA in de Bevindingen Status Excel 1 de rol van centrale coördinator voor de verschillende focus groepen, en dat de project manager eindverantwoordelijk is voor het verwerken van feedback en het afstemmen met IPL over (mogelijke) aanpassingen naar aanleiding van ontvangen feedback. Ook heeft COA aangegeven wie deze rollen zullen vervullen. Verder geeft COA aan dat in een algoritmeregister zal worden opgenomen wat het aanspreekpunt is waar derden en algoritmegebruikers terecht kunnen.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding gesloten kan worden indien COA de beschrijvingen opneemt in de algoritmedocumentatie en/of het algoritmeregister.

### Stap 2 van Analyse 2

In sectie 6.j van de Technische Documentatie 2 beschrijft COA hoe het feedback van Algoritmegebruikers middels enquêtes en focusgroepen. In deze feedback wil COA zich focussen wat het begrip van de Algoritmegebruikers is over het Algoritme, hoe bruikbaar het Algoritme voor de Algoritmegebruikers is, de redenen voor het (niet) overnemen van de Algoritme uitkomsten en gewenste aanpassingen aan het Algoritme. Tijdens Stap 1 heeft COA beschreven dat de projectleider de coördinator is voor het verzamelen van deze feedback.

In het Communicatieplan beschrijft COA de communicatie strategie. Daarin beschrijft COA dat persoonlijke communicatie vanuit COA naar Vergunningshouders het belangrijkste communicatiemiddel is. Daarnaast zijn er aanvullende communicatiemiddelen zoals COA Plein, COA.nl en MyCoa. In de Q&A beschrijft COA dat op COA Plein wordt beschreven wie het aanspreekpunt is voor Vergunningshouders.



29 april 2024  
KD/wk/24-0725

Daarnaast zal in de eerste maand van de pilot het algoritmeregister worden gevuld inclusief contactgegevens van COA voor derden. Deloitte heeft de informatie die zal worden opgenomen in het algoritmeregister niet geanalyseerd.

COA heeft opvolging gegeven aan de aanbeveling van Deloitte door informatie toe te voegen aan de algoritmedocumentatie over het verzamelen van feedback, en de aanspreekpunten voor Algoritmegebruikers, Vergunningshouders en derden. De bevinding is daarmee gesloten. Wel is de aanbeveling van Deloitte om, indien wordt besloten daadwerkelijk te starten met de Pilot, de publicatie van het algoritmeregister af te ronden en een aankondiging van de pilot te plaatsen op COA.nl.

#### 4.2.3. Bevinding 10

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Het Algoritme geeft enkel adviezen aan de Algoritmegebruiker. Er is een risico dat de Algoritmegebruiker niet bewust is van de beperkingen van het Algoritme bij het geven van adviezen en er daarbij teveel vertrouwen ontstaat in het Algoritme.
- **Aanbeveling:** Verrijk de verplichte kennis en hulpmiddelen voor de Algoritmegebruiker met een onderdeel over de reikwijdte van het advies van het Algoritme, zodat de Algoritmegebruiker zich bewust is van het risico op automation bias.

#### Stap 1 van Analyse 2

Op verzoek van COA is deze bevinding niet geanalyseerd in Stap 1.

#### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 heeft COA de Werkinstructie opgenomen. De Werkinstructie beschrijft voor de Algoritmegebruikers wat de scope van het Algoritme is en hoe het Algoritme toegepast dient te worden. Daarnaast beschrijft COA in de Werkinstructie dat het Algoritme een beslissingsondersteunend Algoritme is dat helpt bij het formuleren van een regioadvies op basis van kansen op de arbeidsmarkt. COA beschrijft hier ook dat de Algoritmegebruiker een breder beeld heeft dan alleen de arbeidsmarkt en het daarom niet erg is als van het Algoritme advies wordt afgeweken. Daarnaast heeft COA een screenshot aangeleverd van een concept versie van de user interface (Printscreen IBIS). Naast een advies over geschikte arbeidsregio's vanuit het Algoritme komt hierin ook terug of sprake is van harde criteria, of er voorkeursregio's zijn, en welke zachte criteria van toepassing zijn. De Printscreen IBIS bevestigt dat de Algoritmegebruiker een breder beeld heeft dan alleen de arbeidsmarkt en daarmee dus zelf een besluit kan nemen dat passend is bij de statushouder.

COA heeft opvolging gegeven aan de aanbeveling van Deloitte door in de Werkinstructie te beschrijven wat de reikwijdte van het advies is, i.e. kansen op de arbeidsmarkt, en dat van het advies afgeweken mag worden. Dit wordt ook gereflecteerd in de user interface voor de gebruiker. De bevinding kan daarmee gesloten worden.

#### 4.3. Dataverzameling en datakwaliteit

Als onderdeel van de analyse van de dimensie 'dataverzameling en datakwaliteit' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van zowel het data verzamelingsproces als de kwaliteit en representativiteit van de verzamelde data met behulp van beschrijvende statistieken.



29 april 2024  
KD/wk/24-0725

#### 4.3.1. Bevinding 11

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Er is onvoldoende documentatie van de analyses die uitgevoerd zijn door het IPL en/of het COA om de kwaliteit en integriteit van de data te waarborgen. Verder zijn er geen data beschrijvende statistieken beschreven in de modeldocumentatie.
- **Aanbeveling:** Documenteer de analyses, die zijn uitgevoerd, inclusief de resultaten en conclusies, om datakwaliteit te analyseren. Verder is het aanbevolen om datakwaliteit te monitoren wanneer het Algoritme opnieuw getraind gaat worden.

##### Stap 1 van Analyse 2

In sectie 6.c van de Technische Documentatie 1 van Stap 1 zijn beschreven welke datakwaliteitschecks zijn uitgevoerd. De conclusies van deze checks zijn niet opgenomen in de algoritmedocumentatie. COA heeft aangegeven additionele details op te nemen in de algoritmedocumentatie

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de algoritmedocumentatie.

##### Stap 2 van Analyse 2

In sectie 6.c van de Technische Documentatie 2 van Stap 2 heeft COA additionele informatie beschreven over de uitgevoerde datakwaliteitschecks. De resultaten van de datakwaliteitschecks worden beschreven in appendix 2 van de Technische Documentatie 2 (o.a. missende waarden, aantal unieke waardes, een samenvatting van de dataset, etc.). COA heeft per tabel met resultaten een korte beschrijving toegevoegd over de uitgevoerde tests en hoe de resultaten tot stand zijn gekomen. COA beschrijft ook dat het niet mogelijk is om alle resultaten uit de data omgeving te exporteren door afspraken met het CBS over data gebruik. COA geeft verder aan dat tijdens de Pilot wordt geanalyseerd of de data in het juiste format staat en juist gespecificeerd is. Sectie 6.f van de Technische Documentatie 2 beschrijft dat de datakwaliteitschecks worden herhaald wanneer het Algoritme opnieuw getraind wordt.

Deloitte observeert dat er geen conclusies zijn gedocumenteerd over de uitgevoerde datakwaliteitschecks, zoals of het aantal missende waarden per variabele in Appendix 2.b naar verwachting is, of dit verklaarbaar is of dat dit mogelijk een probleem vormt. Een voorbeeld is dat in 2014 en 2015 slechts één record in de dataset zit met een harde criteria, terwijl dit in de jaren erna veel hoger ligt (bijv. 1,558 records in 2015). Het is onduidelijk wat het (verwachte) effect hiervan is op de kwaliteit van de voorspellingen van het Algoritme. Hierdoor blijft het onduidelijk of de data van voldoende kwaliteit is om gebruikt te kunnen worden (zie ook bevinding 6). Dit brengt het risico met zich mee dat de data van onvoldoende kwaliteit is om te gebruiken voor de ontwikkeling van het Algoritme.

Deloitte beveelt aan om heldere conclusies te documenteren over de datakwaliteit en dat de datakwaliteit voldoende is om gebruikt te kunnen worden. Daarnaast beveelt Deloitte aan om dit te herhalen wanneer het Algoritme opnieuw getraind wordt.

| Nr. | Categorie    | Observatie                                                                                                                                                                                                           | Aanbeveling                                                                                                                                                                                                                            |
|-----|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 11  | Documentatie | De algoritmedocumentatie bevat een beschrijving van de uitgevoerde datakwaliteitschecks en de resultaten ervan. Echter, er zijn geen onderbouwde conclusies gedocumenteerd over de uitgevoerde datakwaliteitschecks. | IPL moet COA de juiste handvatten bieden om aan de hand van de gepresenteerde datakwaliteitsmetriekeken een conclusie te kunnen trekken of de datakwaliteit voldoende is om te kunnen starten met de Pilot en waarom dit het geval is. |



29 april 2024  
KD/wk/24-0725

Hierdoor is het onduidelijk of de data van voldoende kwaliteit is om gebruikt te kunnen worden voor de Pilot.

COA dient dan te concluderen of de datakwaliteit voldoende is om te kunnen starten met de Pilot.

#### 4.3.2. CBS-1

De onderstaande bevinding komt uit de analyse activiteiten onder de dimensie dataverzameling en datakwaliteit van het CBS. Voor de uitgevoerde analyse van het CBS wordt verwezen naar het document CBS Analyse [1].

- **Observatie:** Er zijn vergunningshouders waarvan het label (de doelvariabele wel/geen baan) ontbreekt en die dus ontbreken in de trainingsdata.
- **Aanbeveling:** Vergelijk de samenstelling van de doelpopulaties met en zonder doelvariabele.

##### Stap 1 van Analyse 2

In de algoritmedocumentatie van Stap 1 heeft COA de Technische Documentatie 1 toegevoegd. In dit document beschrijft COA in sectie 6.a hoe de doelpopulatie wordt vastgesteld. COA beschrijft dat de doelvariabele voor de volledige populatie beschikbaar is wanneer Vergunningshouders die niet in scope zijn uit de data worden gehaald. Dit is ook op te maken uit de tabel in Appendix 2.c van de Technische Documentatie.

COA heeft opvolging gegeven aan de aanbeveling van CBS door informatie toe te voegen over de doelpopulatie. De bevinding is daarmee gesloten.

#### 4.3.3. CBS-2

De onderstaande observatie komt uit de analyse activiteiten onder de dimensie dataverzameling en datakwaliteit van het CBS. Voor de uitgevoerde analyse van het CBS wordt verwezen naar het document CBS Analyse [1].

- **Observatie:** Het gebruikte Algoritme (gradient boosting) is gebaseerd op beslisbomen. Dit soort algoritmen zijn robuust tegen ontbrekende waarden op kenmerken. Eenheden met een ontbrekende waarde op een kenmerk waarop gesplitst wordt, worden toegewezen aan het blad dat het beste past bij het doel. Er is echter een set kenmerken die voor alle personen in de trainingsdata ontbreken. Deze kenmerken zijn verkregen uit gesprekken met vergunningshouders, maar zijn niet opgenomen in de dataset, bijvoorbeeld het netwerk en de ambities van de vergunningshouders. COA-medewerkers nemen deze mee in de afweging over de best passende arbeidsmarktregio als aanvulling op de andere kenmerken die wel in de dataset zijn meegenomen, zoals opleiding of werkgeschiedenis. Ten opzichte van de COA-medewerker, mist de inputdata voor het Algoritme dus informatie om de baankansen te voorspellen.
- **Aanbeveling:** Draag zorg dat alle werknemers die met het Algoritme werken zich bewust zijn dat de voorspelde kansen zijn gemaakt op basis van alleen de kenmerken die in de trainingsdata zijn opgenomen, en dat ze hun toewijzing baseren op de combinatie van voorspelde modelkansen en de overige kenmerken (bijvoorbeeld netwerk en ambitie).

##### Stap 1 van Analyse 2

Op verzoek van COA is deze bevinding niet geanalyseerd in Stap 1.

##### Stap 2 van Analyse 2

COA beschrijft in de Bevindingen Status Excel 2 dat in de training en het bijbehorende instructiemateriaal is opgenomen dat het advies van het Algoritme een belangrijk onderdeel is van verschillende elementen die gewogen worden in een regio-advies. Naast dit element zijn er ook zachte criteria die meegenomen dienen te worden door de Algoritmegebruiker worden in het advies, zoals netwerk.



29 april 2024  
KD/wk/24-0725

De Werkinstructie beschrijft op pagina 7 dat het Algoritme een beslissingsondersteunend Algoritme is en dat de Algoritmegebruiker een breder beeld heeft en daarom ook andere zachte criteria dient mee te nemen (zoals ambitie, ervaring en opleidingsachtergrond). Daarnaast beschrijft de Werkinstructie dat het niet het erg is als van het Algoritme advies wordt afgeweken. Zoals beschreven in Bevinding 10 wordt dit bevestigd door de user interface, waarin ook deze zachte criteria worden getoond aan de Algoritmegebruiker.

COA heeft hiermee opvolging gegeven aan de aanbeveling van het CBS. De bevinding is daarmee gesloten.

#### 4.3.4. CBS-3

De onderstaande bevinding komt uit de analyse activiteiten onder de dimensie dataverzameling en datakwaliteit van het CBS. Voor de uitgevoerde analyse van het CBS wordt verwezen naar het document CBS Analyse [1].

- **Observatie:** Augustus 2020 is geen representatief meetmoment voor het meten van baankansen onder vergunningshouders (of alle inwoners van Nederland), vanwege de uitbraak van het coronavirus in Nederland en de bijbehorende maatregelen in maart 2020 en later. Vergunningshouders werken relatief vaak op tijdelijke contracten in sectoren als de horeca. Het is aannemelijk dat deze groep relatief hard geraakt is door de maatregelen en op dat moment noodgedwongen minder is gaan werken of werkloos is geworden.
- **Aanbeveling:** Verkrijg inzicht in de verschillen tussen augustus 2020 en de periode vóór coronamaatregelen (in hoeverre werken vergunningshouders minder uren, of in andere sectoren). Om aan de hand van deze resultaten te bepalen hoe representatief de resultaten zijn voor een typisch jaar. Indien mogelijk: herhaal de analyse op een periode zonder coronamaatregelen.

#### Stap 1 van Analyse 2

COA beschrijft in de Bevindingen Status Excel 1 dat er momenteel nog geen post COVID data beschikbaar is, en dat het daarmee nog niet haalbaar is een analyse op deze periode uit te voeren. Er is geen analyse uitgevoerd op de verschillen tussen de pre COVID periode en de COVID periode, omdat COA van mening is dat de pre COVID periode niet representatief is voor de post COVID periode. Het COA is voornemens voorafgaand aan de start van de pilot het Algoritme opnieuw te trainen op basis van de meest recente data.

In de Tussentijdse Rapportage heeft Deloitte aangegeven deze bevinding opnieuw te analyseren tijdens Stap 2 op basis van de algoritmedocumentatie.

#### Stap 2 van Analyse 2

Tijdens Stap 2 heeft COA aangegeven dat het Algoritme opnieuw is getraind op een post COVID dataset ("Post Covid Algoritmes"). COA heeft na vragen in de Q&A een tabel aangeleverd met enkele performance metrics over twee mogelijke Post COVID Algoritmes. Daarnaast heeft COA aangegeven dat het Algoritme nogmaals opnieuw getraind gaat worden voordat een pilot van start gaat ("Pilot Algoritme"). De algoritmedocumentatie, inclusief uitkomsten van datakwaliteitschecks, performance tests, variabel importance plots, resultaten van de backtests, zijn echter volledig gebaseerd op het Algoritme en niet aangepast naar aanleiding van de hertrainde Post Covid Algoritmes.

Deloitte observeert dat de aangeleverde tabel niet bijdraagt aan de vraag of het Algoritme dat getraind is op o.a. pre COVID en COVID data ook representatief is voor een post COVID periode waarin thuiswerken bijvoorbeeld veel gangbaarder is. Het blijft hiermee onduidelijk of het Algoritme dat getraind is op o.a. pre COVID en COVID data ook kan worden toegepast op een post COVID periode.

Verder ontbreken heldere procedures voor het hertrainen van het Algoritme en de vastlegging hiervan. Sectie 6.f van de Technische Documentatie 2 beschrijft dat bij hertrainen bepaalde kwaliteitschecks zoals beschreven in de technische documentatie en het monitoringplan worden uitgevoerd.



29 april 2024  
KD/wk/24-0725

Echter, doordat de uitkomsten en conclusies van deze checks niet zijn vastgelegd is onduidelijk of deze analyses zijn uitgevoerd voor de Post Covid Algoritmes, zoals checks op datakwaliteit, een analyse van de robuustheid van de voorspellingen (e.g., zijn de voorspellingen voor dezelfde statushouder vergelijkbaar tussen het Algoritme en Post Covid Algoritmes), en de onderbouwing van de data periode die is gebruikt voor hertrainen (e.g., het Algoritme bevatte 4.5 jaar aan data met ruim 123 duizend observaties, de Post COVID Algoritmes slechts 5 maanden met minder dan 3 duizend observaties).

Ook komt in de algoritmedocumentatie niet naar voren of de Post COVID Algoritmes voldoende presteren om succesvol te zijn in de Pilot. Ook is het onduidelijk in hoeverre de datakwaliteit en prestaties van de Post COVID Algoritmes robuust zijn en vergelijkbaar zijn met de resultaten zoals vastgelegd in de algoritmedocumentatie. Dit brengt het risico met zich mee dat de Post Covid Algoritmes niet voldoende presteren om succesvol te zijn in de Pilot. Daarnaast ontbreken heldere procedures voor het hertrainen van het Algoritme in de algoritmedocumentatie, en ontbreekt de vastlegging van de relevante documentatie over een hertrainde versie van het Algoritme.

Deloitte beveelt aan om de prestaties (o.a. datakwaliteit, robuustheid) van het Pilot Algoritme te documenteren en te onderbouwen of de prestaties van dit Pilot Algoritme voldoende zijn om te gebruiken in de Pilot, zodat de stuurgroep van COA dit mee kan nemen in het besluit om al dan niet te starten met de Pilot.

| Nr.          | Categorie    | Observatie                                                                                                                                                                                                                                                                                                           | Aanbeveling                                                                                                                                                                                                                                                                                                                                             |
|--------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>CBS-3</b> | Documentatie | In de algoritmedocumentatie komt niet naar voren wat de prestaties zijn van het Pilot Algoritme. Ook is het onduidelijk in hoeverre de datakwaliteit en prestaties van de Post COVID Algoritmes robuust zijn en vergelijkbaar zijn met de resultaten van het Algoritme zoals vastgelegd in de algoritmedocumentatie. | IPL dient te onderbouwen waarom het Algoritme, dat deels getraind is op een pre-COVID periode, ook kan worden toegepast op een post-COVID periode waarin de arbeidsmarkt erg is veranderd, thuiswerken is bijvoorbeeld veel gebruikelijker. Dit kan bijvoorbeeld door de prestaties van het Algoritme te toetsen op een out-of-time post-COVID periode. |
|              |              | Daarnaast ontbreken heldere procedures voor het hertrainen van het Algoritme in de algoritmedocumentatie, en ontbreekt de vastlegging van relevante documentatie over een hertrainde versie van het Algoritme.                                                                                                       | Indien de prestaties van het Algoritme voor die out-of-time post-COVID periode niet veel afwijken van de prestaties zoals gedocumenteerd in de algoritmedocumentatie, dan kun je daaruit concluderen of het effect van de COVID periode mogelijk beperkt is.                                                                                            |
|              |              |                                                                                                                                                                                                                                                                                                                      | Daarnaast dient COA met IPL vast te stellen wat er minimaal gedocumenteerd zal worden bij hertrainen van het Algoritme. Hierbij zou tenminste moeten blijken wat er geanalyseerd is rond datakwaliteit, model performance en robuustheid, zodat op basis hiervan COA kan concluderen of het hertrainde Algoritme voldoende presteert.                   |



29 april 2024  
KD/wk/24-0725

#### 4.4. Bias en datagebruik

Als onderdeel van de analyse van de dimensie 'bias en datagebruik' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van de eerlijkeheidsstatistieken, onwenselijke bias en profilerings eigenschappen van het Algoritme en de gebruikte data, afgezet tegen marktconforme (gangbare) acceptatiecriteria voor deze statistieken.

##### 4.4.1. Bevinding 12

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De definitie van eerlijkheid is niet volledig voor wat betreft de context van het Algoritme. Het is niet gedefinieerd welke groepen er met elkaar vergeleken worden, welke marges er zijn en hoe eerlijkheid gemeten wordt.
- **Aanbeveling:** Werk de definitie van eerlijkheid uniform en volledig uit inclusief procedures om eerlijkheid te meten op de test set en tijdens het monitoren van het Algoritme. Toets de definitie van eerlijkheid met een diverse groep (minimaal) bestaande uit het COA, de LRT, de vergunningshouders en/of belangenbehartigers maar ook een onafhankelijke afvaardiging vanuit de maatschappij in Nederland.

#### Stap 1 van Analyse 1

Sectie 4.c.ii van de Technische Documentatie 1 beschrijft de definitie van eerlijkheid vanuit (1) het perspectief van de Vergunningshouders, en (2) het perspectief van de arbeidsmarktregio's. De analyse en conclusie vanuit het perspectief van de arbeidsmarktregio's is niet opgenomen in de Technische Documentatie 1. COA heeft aangegeven dit voor Stap 2 op te zullen nemen.

In het Monitoringsplan 1 is opgenomen onder 'monitoring level 3', dat jaarlijks wordt uitgevoerd, of de verdeling van specifieke subgroepen (o.a., leeftijd  $\geq 40$ , vrouwen, herkomstland Syrië) over arbeidsmarktregio's statistisch niet meer afwijkt dan de historisch geobserveerde maximum afwijking. COA geeft in het Monitoringsplan aan dat de lijst van subgroepen nog niet finaal is en voor Stap 2 wordt gefinaliseerd.

COA geeft aan voor Stap 2 de opgestelde definitie van eerlijkheid en monitoring hiervan te toetsen met stakeholders.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en in Stap 2 opnieuw geanalyseerd zal worden op basis van de technische documentatie, het monitoringsplan en het stakeholder proposal.

#### Stap 2 van Analyse 2

De definitie van eerlijkheid vanuit het perspectief van de arbeidsmarktregio's is dat elke arbeidsmarktregio beter af is met gebruik van het Algoritme vergeleken met de status quo. In de Q&A geeft COA aan dat het niet mogelijk is om de resultaten of conclusies te delen over eerlijkheid vanuit het perspectief van de arbeidsmarktregio's, omdat er niet genoeg observaties zijn in elke arbeidsmarktregio waardoor de resultaten niet geëxporteerd mogen worden vanuit de CBS omgeving. Het is daardoor voor Deloitte niet mogelijk om eerlijkheid vanuit dit perspectief te analyseren. COA beschrijft in de Q&A dat de effecten over de arbeidsmarktregio's gemonitord zullen worden in de Pilot.

De analyses voor het eerlijkheidsperspectief vanuit Vergunningshouders zijn gebaseerd op het Algoritme dat is ontwikkeld met pre COVID data, net als andere analyses zoals beschreven in sectie 4.3.4. Het is daarmee niet duidelijk of het Pilot Algoritme voldoet aan de eerlijkheidsperspectieven vanuit Vergunningshouders.



29 april 2024  
KD/wk/24-0725

COA heeft in het Monitoringsplan 2 de finale lijst van subgroepen onder ‘monitoring level 3’ opgenomen, waarbij gekeken zal worden naar leeftijd ( $\geq 40$ ), geslacht, land van herkomst (Syrië), casus grootte ( $\geq 5$ ) en onderwijsniveau (laag).

Deloitte observeert dat er onvoldoende is gedocumenteerd over de eerlijkheid van het Algoritme. Het is onduidelijk of het Algoritme voldoet aan de definitie van eerlijkheid vanuit perspectieven van zowel de Vergunningshouders als de arbeidsmarktregio’s. Daarnaast is het onduidelijk of de definitie van eerlijkheid is getoetst met een diverse groep. Het risico ontstaat dat het Algoritme niet voldoet aan de door COA opgestelde definitie van eerlijkheid, en dat draagvlak ontbreekt voor de opgestelde definitie van eerlijkheid.

De aanbeveling is om het Pilot Algoritme te analyseren op beide eerlijkheidsperspectieven en op basis van deze analyse te concluderen of het Algoritme voldoet aan de eerlijkheidsperspectieven, zodat de stuurgroep van COA dit mee kan nemen in het besluit om al dan niet te starten met de Pilot. Verder blijft de aanbeveling om de definitie van eerlijkheid te toetsen met een diverse groep betrokkenen om draagvlak te creëren of te beargumenteren waarom COA dit niet nodig acht.

| Nr. | Categorie    | Observatie                                                                                                                                                                   | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12  | Documentatie | Er is geen recente documentatie beschikbaar over beide eerlijkheidsperspectieven. Daarnaast is de definitie van eerlijkheid niet getoetst met een diverse groep betrokkenen. | <p>IPL dient de eerlijkheidsperspectieven opnieuw te analyseren op basis van de laatste versie van het Algoritme dat gebruikt zal worden in de Pilot.</p> <p>Op basis van de resultaten van deze analyse, dient COA te concluderen of het Algoritme voldoet aan de door COA gestelde definitie van eerlijkheid.</p> <p>Daarnaast dient COA de eerlijkheidsperspectieven met een diverse groep betrokkenen te toetsen zodat COA draagvlak kan creëren voor deze perspectieven. COA kan ook beargumenteren waarom zij dit niet nodig acht.</p> |

4.4.2. CBS-4

De onderstaande observatie komt uit de analyse activiteiten onder de dimensie ‘bias en datagebruik’ van het CBS. Voor de uitgevoerde analyse van het CBS wordt verwezen naar het document CBS Analyse [1].

- Observatie:** De samenstelling van de trainingsdata per arbeidsregio kan afwijken van de samenstelling van de gehele populatie vergunningshouders (door de selectieve allocatie) en van de samenstelling van toekomstige. Elk regionaal model zal zwaarder leunen op de relatief veel voorkomende groepen in die regio (een regio heeft mogelijk vooral vergunningshouders met een bepaalde nationaliteit of werksector). Dit leidt wellicht tot vertekening van de voorspellingen van relatief zeldzame groepen in deze regio. Daarnaast kan het zo zijn dat bepaalde groepen helemaal niet voorkomen in bepaalde regio’s. Het Algoritme kan dan geen inschatting maken van het succes van mensen met kenmerk X in regio A. Dit maakt de vergelijkbaarheid tussen regio’s, en de afweging van het Algoritme over in welke regio’s de vergunningshouders de hoogste baankans heeft, moeilijk.



29 april 2024  
KD/wk/24-0725

- **Aanbeveling:** Dit representativiteitsrisico is inherent aan de regionale opzet van het Algoritme. Dit risico kan gemitigeerd worden door het creëren van bewustzijn bij medewerkers, of er rekening mee te houden in de modelopzet. Om de disbalans in de regionale trainingsdata te balanceren zou het Algoritme bijvoorbeeld getraind kunnen worden met grotere gewichten voor relatief zeldzame groepen en lagere gewichten voor relatief veel voorkomende groepen. Een procentuele vergelijking van welke subgroepen in welke regio's wonen geeft al een eerste beeld van de mate van balans/disbalans. Ook worden zo subgroepen in beeld gebracht die in bepaalde regio's niet of amper voorkomen en waarvoor de voorspellingen van het Algoritme minder betrouwbaar zijn. Om inzicht te krijgen in de mate van afwijking tussen regio's, en het risico dat de gemiddelde baankans voor die regio vertekend is, kan gebruikt worden gemaakt van de variatiecoëfficiënt (CV). Bij de COA-medewerkers kan bewustzijn gecreëerd worden door de wijze waarop zij de uitkomst van het Algoritme te zien krijgen. Als zij, naast de aanbevolen regio's, ook de geschatte baankans en een betrouwbaarheidsscore per regio te zien krijgen, geeft hen dit waardevolle informatie over de robuustheid van de aanbeveling.

### Stap 1 van Analyse 2

In Stap 1 beschrijft COA in sectie 6.b van de Technische Documentatie 1 het representativiteitsrisico zoals beschreven in de observatie. COA beschrijft dat wanneer bepaalde groepen ontbreken in arbeidsmarktregio's, het Algoritme dan geen inschatting kan maken voor deze groepen in de specifieke regio. Verder beschrijft COA dit niet te zijn tegengekomen in de data.

COA heeft aangegeven dit verder te monitoring als onderdeel van "monitoring level 3". Echter wordt deze monitoringsactiviteit niet omschreven in monitoring level 3 in het Monitoringsplan 1. COA heeft aangegeven de monitoringsactiviteit in het monitoringsplan te beschrijven in Stap 2.

Daarnaast beschrijft de aanbeveling dat er bewustzijn gecreëerd kan worden bij de COA medewerkers door de wijze waarop zij de output van het Algoritme te zien krijgen. In Stap 2 zal de werkinstructie en voorlichting naar COA medewerkers worden aangeleverd door COA.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van het monitoringsplan en de werkinstructie.

### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 is het Monitoringsplan 2 opgenomen. In het Monitoringsplan 2 is een beschrijving opgenomen hoe de verdeling van verschillende subgroepen gemonitord gaat worden tijdens de pilot. Als onderdeel van "monitoring level 3" is opgenomen dat de verdeling van de subgroepen door het Algoritme vergeleken gaat worden met de historische verdeling.

Verder heeft COA een de Werkgelegenheid Impact Score (WIS) geïntroduceerd in de Werkinstructie. Op basis van de WIS wordt een adviesrang opgeteld voor de drie meest kansrijke regio's. De regio met de hoogste score zal altijd een WIS van 10 krijgen. Hierdoor lijkt het Algoritme altijd een geschikte arbeidsmarktregio gevonden te worden ook al is de geschatte baankans hier klein. De WIS is daarmee niet een representatieve weergave van de geschatte baankans en geeft geen inzicht in de robuustheid van de schatting. Het is waardevol voor de Algoritmegebruiker om te weten indien de geschatte baankans voor elke arbeidsmarktregio laag is, echter is deze informatie nu niet beschikbaar voor de Algoritmegebruiker. Met deze informatie zou de Algoritmegebruiker kunnen beslissen om de zachte criteria meer mee te wegen in het eindadvies.

Bij het vaststellen van de WIS wordt een correctie gemaakt voor mogelijk banenverlies doordat toekomstige casussen niet naar de arbeidsmarktregio kunnen waar de huidige casus geplaatst wordt. Deze correctie doet COA met een "toekomstige casussen niveau".



29 april 2024  
KD/wk/24-0725

COA simuleert toekomstige casussen op basis van historische patronen. Het is onvoldoende gedocumenteerd of de historische data en simulatie voldoende representatief zijn voor dit doeleinde. Verder betekent het gebruik van het “toekomstige casussen niveau” dat een Vergunningshouder een locatie advies kan krijgen wat voor deze persoon een niet optimaal advies kan zijn, maar wat positief is voor de Vergunningshouders van de toekomst. Alternatief zou COA kunnen kiezen deze correctie niet toe te passen en een ‘first come, first served’ principe te hanteren per arbeidsmarktregio. De correctie voor toekomstige casussen brengt ook als risico met zich mee dat de Algoritmegebruiker minder begrip heeft van de getoonde WIS en daardoor mogelijk deze minder goed kan meewegen in de finale plaatsing. Het komt uit de algoritmedocumentatie onvoldoende naar voren waarom COA deze (ethische) keuze heeft gemaakt.

Deloitte observeert dat de WIS niet de informatie levert aan de Algoritmegebruiker om de robuustheid van het advies van het Algoritme in te schatten. Het risico ontstaat dat de Algoritmegebruikers teveel leunen op het advies van het Algoritme terwijl de voorspellingen weergeven dat de geschatte baankans laag is. Daarnaast maakt de WIS gebruik van een “toekomstig casussen niveau”. Hiermee stelt COA het belang van toekomstige casussen mogelijk boven de Vergunningshouder in kwestie. Het is onduidelijk of COA deze keuze heeft getoetst, of er voldoende draagvlak voor de keuze is, en of de data voldoende representatief is voor de berekening van een toekomstige casussen niveau.

De aanbeveling is om bewustzijn bij de Algoritmegebruikers te creëren door aanvullende informatie aan de gebruiker te laten zien. Naast de aanbevolen regio’s zijn ook de geschatte baankans en een betrouwbaarheidsscore per regio waardevolle informatie. Daarnaast is het aanbevolen om de keuze en methode voor toekomstige casussen te toetsen met betrokkenen, zowel intern als extern.

| Nr.   | Categorie | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CBS-4 | Model     | <p>De WIS laat de gebruiker drie scores zien. Een huidig casussen niveau, toekomstig casussen niveau en een gecombineerde score. Het is onduidelijk hoe deze informatie de Algoritmegebruiker ondersteunt.</p> <p>De regio met de hoogste score zal altijd een WIS van 10 krijgen. Hierdoor lijkt het Algoritme altijd een geschikte arbeidsmarktregio gevonden te worden ook al is de geschatte baankans hier klein. De WIS levert hiermee niet de informatie aan de Algoritmegebruiker om de robuustheid van het advies van het Algoritme in te schatten.</p> | <p>IPL dient samen met COA te concluderen of de WIS de juiste informatie levert aan de Algoritmegebruiker. Pas waar nodig de WIS berekening aan, of beargumenteer waarom de huidige WIS berekening wordt toegepast.</p> <p>Verder dient het IPL in overleg met COA de algoritmedocumentatie te verrijken met een beschrijving waarom de keuze is gemaakt voor een fictief toekomstige casussen niveau, en een analyse die beschrijft dat de benodigde data voldoende representatief is voor de berekening van dit niveau. Waar nodig kan IPL/COA de WIS berekening wederom aanpassen, of beargumenteren waarom de huidige WIS berekening wordt toegepast.</p> |

29 april 2024  
KD/wk/24-0725

---

Daarnaast is het onvoldoende gedocumenteerd waarom de keuze is gemaakt voor een toekomstige casussen niveau in de WIS, waarmee het belang van toekomstige casussen mogelijk boven de Vergunninghouder in kwestie stelt, en of de benodigde data voldoende representatief is voor de berekening van een toekomstige casussen niveau.

---

#### 4.4.3. CBS-5

De onderstaande observatie komt uit de analyse activiteiten onder de dimensie 'bias en datagebruik' van het CBS. Voor de uitgevoerde analyse van het CBS wordt verwezen naar het document CBS Analyse [1].

- **Observatie:** Het is mogelijk dat vergunningshouders verhuizen in het eerste jaar na toewijzing aan de arbeidsmarktregio. Dit gegeven wordt in de inputdata van het Algoritme niet meegenomen. Het zou zo kunnen zijn dat vergunningshouders verhuizen en in de nieuwe arbeidsmarktregio een baan vinden. In de inputdata wordt dit geregistreerd alsof een baan in de toegewezen arbeidsmarktregio wordt gevonden. Deze keuze is gemaakt om bias te verminderen: het Algoritme wordt niet getraind op factoren die je nog niet kan weten op het moment van toewijzing aan een arbeidsmarktregio. Tegelijkertijd zou deze keuze juist een mogelijke bron van bias kunnen zijn als er arbeidsmarktregio's zijn waar een groot deel van de vergunningshouders snel weer verhuist. De voorspellingen van het Algoritme voor deze arbeidsmarktregio kloppen dan niet. Wij raden aan om een extra analyse uit te voeren om te bepalen in hoeverre dit risico van toepassing is.
- **Aanbeveling:** Algoritme trainen op vergunningshouders die nog steeds woonachtig zijn in de arbeidsmarktregio waaraan ze zijn toegewezen en kijken of dit de resultaten verandert, of controleer hoe vaak het voorkomt dat personen naar een andere arbeidsmarktregio verhuizen.

#### Stap 1 van Analyse 2

In Stap 1 beschrijft COA in sectie 6.d van de Technische Documentatie 1 waarom ze van mening is dat statushouders die verhuizen en een nieuwe baan in een andere arbeidsmarktregio vinden ook een positieve doelvariabele zouden moeten krijgen.

Indien kan worden aangetoond dat de grote van deze verhuizende populatie beperkt is in de verschillende arbeidsmarktregio's, en daarmee de impact van deze verhuizende populatie ook beperkt zal zijn, kan deze bevinding gesloten worden. COA heeft aangegeven om voor Stap 2 te onderzoeken of het mogelijk is om te analyseren of de impact van de verhuizende populatie beperkt is.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de algoritmedocumentatie.

#### Stap 2 van Analyse 2

In Stap twee heeft COA in Appendix 2.m van de Technische Documentatie 2 een tabel toegevoegd waarin wordt beschreven hoeveel personen in de toegewezen arbeidsmarktregio blijven wonen in het eerste jaar. Het percentage van het aantal personen dat verhuist is 4% in de jaren 2014 en 2015, en 3% in de jaren daarna. In de Q&A geeft COA aan dat dit percentage voor COA acceptabel is.

COA heeft hiermee opvolging gegeven aan de aanbeveling van het CBS door te laten zien dat het percentage personen dat verhuist voor COA acceptabel is. De bevinding is daarom gesloten.



29 april 2024  
KD/wk/24-0725

#### 4.5. Algoritme input en componenten

Als onderdeel van de analyse van de dimensie 'algoritme input en componenten' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van de input en componenten op hun uitlegbaarheid, relevantie en impact op de uitkomst op basis van gangbare marktpraktijk en algoritmetheorie.

##### 4.5.1. Bevinding 13

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Het Algoritme gebruikt alle attributen uit IBIS die potentieel relevant zijn voor het te behalen doel. Er wordt geen input selectie (feature selection) toegepast om te analyseren of attributen relevant zijn.  
Het Algoritme kan hierdoor onnodig complex zijn en onnodig veel data gebruiken. Het gebruiken van onnodig veel data is in strijd met het dataminimalisatie-principe uit de AVG wanneer de data bestaat uit persoonsgegevens.
- **Aanbeveling:** Pas input selectie toe om te onderbouwen waarom attributen nodig zijn in het Algoritme.

#### Stap 1 van Analyse 2

Tijdens Stap 1 beschrijft COA in de Bevindingen Status Excel 1 dat het input selectie heeft toegepast door middel van een kwantitatieve en kwalitatieve analyse. COA geeft aan dat voor de kwalitatieve analyse interviews met COA medewerkers en proces experts binnen Nederland zijn gehouden. Voor de kwantitatieve analyse is gekeken naar variable importance plots (opgenomen in de Technische Documentatie 1), en naar resultaten van het algoritme van IPL in andere landen, en academische literatuur.

In de algoritmedocumentatie van Stap 1 is het niet opgenomen hoe de input van experts is meegenomen en of dit geleid heeft tot aanpassingen van de gebruikte attributen. Daarnaast zijn de variable importance plots in de Technische Documentatie 1 verouderd, en is er geen onderbouwde conclusie opgenomen in de Technische Documentatie 1 die COA hieraan verbindt. COA heeft verder aangegeven dat er vanuit (interne en externe) stakeholder sessies geen directe feedback is geweest op de gebruikte attributen. Het is onduidelijk of expliciet is getoetst met stakeholders dat het wenselijk is dat persoonskenmerken zoals geslacht, leeftijd en land van herkomst gebruikt worden voor de plaatsing van statushouders. In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de algoritmedocumentatie.

#### Stap 2 van Analyse 2

In de algoritmedocumentatie van Stap 2 heeft COA de Feature Lijst toegevoegd. In de Q&A beschrijft COA dat de Feature Lijst nog niet finaal is omdat het nog niet is besloten of de kenmerken land van herkomst en moedertaal in het Algoritme gebruikt gaan worden. Het gebruik van deze twee kenmerken wordt verder geanalyseerd in Sectie 4.8.2.

De Feature Lijst beschrijft de kenmerken die in het Algoritme gebruikt zullen worden. COA beschrijft in deze Feature Lijst niet waarom deze kenmerken gebruikt worden. In sectie 6.e van de Technische Documentatie 2 beschrijft COA dat het de kenmerken leeftijd, geslacht, burgerlijke staat, opleiding, aantal familie leden, land van herkomst, moedertaal, en werkervaring in een industrie selecteert op basis van data beschikbaarheid. Verder worden er verschillende kenmerken over de toewijzingsprocedure, zoals maand en jaar van toewijzing, geselecteerd. Er wordt niet onderbouwd in de algoritmedocumentatie waarom deze kenmerken relevant zijn voor het Algoritme en of het COA deze kenmerken wel wil gebruiken voor de doeleinden van het Algoritme.



29 april 2024  
KD/wk/24-0725

Tijdens Stap 1 heeft COA aangegeven kwantitatieve en kwalitatieve analyses te hebben uitgevoerd als onderdeel van de input selectie. Deze analyses zijn echter niet opgenomen in de algoritmedocumentatie van zowel Stap 1 als Stap 2. Deze analyses zijn daarmee niet zichtbaar voor Deloitte waardoor het onduidelijk blijft wat COA heeft gedaan om de relevantie van de geselecteerde kenmerken te analyseren. Daarnaast zijn deze analyses gebaseerd op een oude versie van het Algoritme, zie Sectie 4.3.4. Het is daardoor onduidelijk of de geselecteerde kenmerken ook relevant zijn op basis van recentere data en of COA verwacht dat de gevonden verbanden ook relevant zijn voor de toekomst.

Deloitte observeert dat het niet is gedocumenteerd waarom de geselecteerde kenmerken relevant zijn in het Algoritme. Het is enkel beschreven dat kenmerken worden geselecteerd omdat deze beschikbaar zijn. Het risico ontstaat dat het Algoritme hierdoor kenmerken gebruikt die niet relevant zijn voor de toekomst. Ook kan het gebruiken van onnodig veel data in strijd zijn met het dataminimalisatie principe uit de AVG wanneer de data bestaat uit persoonsgegevens. Verder kan het Algoritme leren van verkeerde relaties in de data (spurious correlations) die niet aanwezig zijn in de data uit de praktijk.

De aanbeveling is om in de algoritmedocumentatie te beschrijven welke attributen worden gebruikt in het Algoritme, hoe deze keuze is gemaakt door COA en waarom de attributen relevant zijn voor het Algoritme.

| <b>Nr.</b> | <b>Categorie</b> | <b>Observatie</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <b>Aanbeveling</b>                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>13</b>  | Model            | Het is niet gedocumenteerd waarom de geselecteerde kenmerken relevant zijn voor het Algoritme. Het risico ontstaat dat het Algoritme hierdoor kenmerken gebruikt die niet relevant zijn voor de toekomst. Ook kan het gebruiken van onnodig veel data in strijd zijn met het dataminimalisatie principe uit de AVG wanneer de data bestaat uit persoonsgegevens. Verder kan het Algoritme leren van verkeerde relaties in de data (spurious correlations) die niet aanwezig zijn in de data uit de praktijk. | IPL dient per attribuut een heldere beschrijving en duidelijke onderbouwing op te nemen in de algoritmedocumentatie waarom de geselecteerde kenmerken relevant zijn voor het Algoritme. COA dient IPL hierbij te ondersteunen door input te leveren waarom de geselecteerde kenmerken relevant zijn in de sociaal maatschappelijke context van het Algoritme en COA. Verder dient COA het eindbesluit te nemen welke kenmerken in het Algoritme gebruikt worden. |



29 april 2024  
KD/wk/24-0725

#### 4.5.2. Bevinding 14

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** De relatie van de attributen ten opzichte van de uitkomst is alleen geanalyseerd op het belang van de attributen. Echter, de directe relatie tot de uitkomst (bijvoorbeeld met het SHAP raamwerk, Partial Dependence Plots en/of Individual Expectations Plots) is niet gedocumenteerd. Hierdoor is het Algoritme niet technisch uitlegbaar en kan het Algoritme leren van verkeerde relaties (spurious correlations) in de data.
- **Aanbeveling:** Documenteer de werking van het Algoritme aan de hand van globale uitlegbaarheidstechnieken.

##### Stap 1 van Analyse 2

In algoritmedocumentatie van Stap 1 is de Technische Documentatie 1 toegevoegd. In de Technische Documentatie 1 zijn de variable importance plots in de Technische Documentatie verouderd, en er is geen onderbouwde conclusie opgenomen bij deze variable importance plots.

Verder blijkt uit de variable importance plots dat leeftijd een sterk verklarende variabele is in veel arbeidsmarktregio's. Er is niet geanalyseerd in hoeverre dit erin resulteert dat het Algoritme bepaalde leeftijdsgroepen altijd in bepaalde arbeidsmarktregio's plaatst en of dit wenselijk is.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de algoritmedocumentatie.

##### Stap 2 van Analyse 2

In de Bevindingen Status Excel 2 beschrijft COA dat nieuwe variable importance plots zijn toegevoegd aan de Technische Documentatie 2. Deze variable importance plots zijn gebaseerd op het verouderde Algoritme dat getraind is op data uit de tijdsperiode Januari 2014 tot Juni 2018. Verder is er geen onderbouwde conclusie opgenomen bij de variable importance plots die de werking van het Algoritme beschrijven. Het is daarmee ook niet beschreven of de relaties en de werking van het Algoritme wenselijk en verklaarbaar zijn voor COA, bijvoorbeeld dat leeftijd een sterk verklarende variabele is in veel arbeidsmarktregio's en dat dit er mogelijk toe kan leiden dat bepaalde leeftijdsgroepen altijd in bepaalde arbeidsmarktregio's worden geplaatst.

Deloitte observeert dat COA de verbanden van het Algoritme niet heeft onderbouwd in de algoritmedocumentatie aan de hand van de variable importance plots, bijvoorbeeld met behulp van experts, bijvoorbeeld Algoritmegebruikers. Daarnaast zijn de gedocumenteerde variable importance plots verouderd. Het is niet gedocumenteerd of de Post Covid Algoritmes en het Pilot Algoritme vergelijkbare relaties vinden op basis van recentere, post COVID data, zoals ook beschreven in Sectie 4.3.4. Het risico ontstaat dat het Pilot Algoritme wordt gebruikt in de Pilot zonder dat de werking daarvan wordt begrepen of dat er goedkeuring wordt gegeven voor een verouderd Algoritme.

De aanbeveling van Deloitte is om de algoritmedocumentatie te verrijken met details over hoe de gemodelleerde verbanden zijn besproken met experts, bijvoorbeeld Algoritmegebruikers, en een bevestiging van deze experts dat de gemodelleerde verbanden niet onlogisch of onwenselijk zijn.



29 april 2024  
KD/wk/24-0725

| Nr. | Categorie    | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 14  | Documentatie | COA heeft de verbanden van het Algoritme niet onderbouwd in de algoritmedocumentatie. Daarnaast zijn de variable importance plots verouderd. Het is niet gedocumenteerd of de Post Covid Algoritmes en het Pilot Algoritme vergelijkbare relaties vinden op basis van recentere, post COVID data. Het risico ontstaat dat het Pilot Algoritme wordt gebruikt in de Pilot zonder dat de werking daarvan wordt begrepen of dat er goedkeuring wordt gegeven voor een verouderd Algoritme. | COA en IPL dienen samen de algoritmedocumentatie te verrijken met een onderbouwing over hoe de gemodelleerde verbanden in het Algoritme te verklaren zijn. COA kan de gemodelleerde verbanden verklaren door het gesprek aan te gaan met experts (bijv. algoritmegebruikers), deze betrokkenen kunnen bevestigen of de gemodelleerde verbanden niet onlogisch of onwenselijk zijn vanuit de praktijk. Hierbij kunnen bijvoorbeeld de variable importance plots als uitgangspunt worden genomen, en kan met experts worden getoetst of het belang van de verschillende attributen verklaarbaar is. |

4.6. Algoritme theorie

Als onderdeel van de analyse van de dimensie 'algoritme theorie' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van de toepasbaarheid van het gekozen Algoritme, hyperparameters, complexiteit en gestelde aannames met behulp van empirisch bewijs en theoretische onderbouwing.



29 april 2024  
KD/wk/24-0725

#### 4.6.1. Bevinding 15

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Tijdens de ontwikkeling van het Algoritme zijn bepaalde groepen, zoals mensen met harde criteria (bijvoorbeeld vergunningshouders met een arbeidscontract of een medische problemen), vergunningshouders onder het kinderpardon van 2019 en vergunningshouders onder de EU-Turkije deal, buiten scope gelaten. Het Algoritme is hierdoor niet getraind op deze groepen en werkt mogelijk niet optimaal voor deze groepen. Wanneer een dergelijke groep in de toekomst in scope is, moet het Algoritme opnieuw getraind worden. Deze beperking is niet beschreven in de algoritmedocumentatie.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met de beperking dat het Algoritme alleen toepasbaar is op de huidige scope en kalibreer het Algoritme opnieuw als de scope verandert.

##### Stap 1 van Analyse 2

In Stap 1 beschrijft COA in sectie 6.b van de Technische Documentatie 1 dat de in scope populatie wordt vastgesteld door het filteren van minderjarigen, de vaststelling door de Algoritmegebruiker dat er geen sprake is van harde criteria, en het feit dat de statushouder in aanmerking komt voor werk. Er wordt aangegeven dat dit specifiek opgenomen zal worden in de instructies voor werknemers en informatiebladen voor statushouders welke COA voor aanvang van de Pilot zal delen.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de werkinstructie.

##### Stap 2 van Analyse 2

In Stap 2 heeft COA de Werkinstructie toegevoegd aan de algoritmedocumentatie. De Werkinstructie beschrijft op pagina 2 dat het belangrijk is om als gebruiker rekening te houden dat het Algoritme uitsluitend gebruikt dient te worden voor de doelgroep die "geschikt is om te werken". Daarnaast wordt op pagina 1 beschreven welke groepen tijdens de ontwikkeling buiten beschouwing zijn gelaten. Deze groepen ontvangen geen advies vanuit het Algoritme.

COA heeft opvolging gegeven aan de aanbeveling met het toevoegen van bovenstaande informatie aan de Werkinstructie. De bevinding kan hiermee gesloten worden.

#### 4.6.2. Bevinding 16

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Essentiële onderdelen van het Algoritme, zoals het combineren van de uitkomsten van meerdere vergunningshouders in één casus, het stochastic online matching component, en de onderbouwing voor ontwerpkeuzes van deze onderdelen zijn niet opgenomen in de algoritmedocumentatie.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met alle onderdelen van het Algoritme inclusief de ontwerpcomponenten van het Algoritme zodat het Algoritme gereproduceerd kan worden door een onafhankelijke derde partij.

##### Stap 1 van Analyse 2

In de in Stap 1 beschrijft COA in secties 2, 3 en 4 van de Technische Documentatie 1 de drie onderdelen van het Algoritme. Voor aanvullende details verwijst COA naar wetenschappelijke publicaties en aanvullende documenten op deze publicaties.



29 april 2024  
KD/wk/24-0725

Verder beschrijft COA in de Technische Documentatie 1 dat er verschillende implementatieopties, zoals verschillende "mapping" functies en optimalisatie criteria, mogelijk zijn. Er is niet beschreven welke optimalisatie criteria COA heeft gekozen (bijvoorbeeld of "load balancing" wordt toegepast), waarom COA een specifieke "mapping" functie en optimalisatie criteria kiest en welke (ethische) gevolgen deze keuzes hebben op het Algoritme.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de algoritmedocumentatie.

### **Stap 1 van Analyse 2**

In Stap 2 beschrijft COA in sectie 4.c.ii van de Technische Documentatie 2 dat de load balancing optie niet wordt geïmplementeerd in het Algoritme. Echter, een onderbouwing voor de gekozen mapping functie en optimalisatiecriteria is niet toegevoegd.

De onderdelen van het Algoritme staan nu beschreven in de algoritmedocumentatie. COA heeft daarmee opvolging gegeven aan de aanbeveling. De bevinding is daarmee gesloten. Dat COA geen onderbouwing heeft gegeven voor de verschillende keuzes wordt verder opgevolgd onder bevinding 6 (zie Sectie 4.1.6).

## **4.7. Algoritme uitvoer**

Als onderdeel van de analyse van de dimensie 'algoritme uitvoer' is in Analyse 1 het volgende uitgevoerd:

- Analyseren van de performance statistieken en resultaten en deze vergelijken met vooropgezette minimum vereisten en key performance indicators (e.g. voor eerlijkheid en stabiliteit).

### **4.7.1. Bevinding 17**

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Tijdens het backtesten komt onvoldoende naar voren welke impact de inzet van het Algoritme heeft op subgroepen die niet worden geplaatst met ondersteuning van het Algoritme. Verder ontbreekt een analyse naar subgroepen die op basis van de leeftijdsriteria in scope zijn maar geen (of beperkt) deel zullen uitmaken van de arbeidsmarkt, zoals vergunningshouders met bijzondere behoeften, arbeidsongeschikten en senioren.
- **Aanbeveling:** Analyseer de impact van het Algoritme op subgroepen die niet worden geplaatst met het Algoritme en op subgroepen die door het Algoritme worden geplaatst maar geen onderdeel zijn van de arbeidsmarkt.

### **Stap 1 van Analyse 2**

COA heeft tijdens Stap 1 aangegeven dat voor subgroepen die niet worden geplaatst door het Algoritme niet altijd data beschikbaar is, en dat het niet mogelijk is een analyse uit te voeren op subgroepen waarvoor geen data beschikbaar is. Het COA geeft aan in de werkinstructie op te nemen over de beperkingen van het Algoritme binnen bepaalde target groepen. Merk op dat een deel van de out-of-scope populatie geplaatst wordt met voorrang in bepaalde arbeidsmarktregio's (e.g., door harde criteria), waardoor de verwachting is dat het Algoritme slechts beperkt invloed heeft op de arbeidsmarktregio's waar deze individuen geplaatst zullen worden.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de werkinstructie.



29 april 2024  
KD/wk/24-0725

## Stap 2 van Analyse 2

In Stap 2 heeft COA de Werkinstructie toegevoegd aan de algoritmedocumentatie. De Werkinstructie beschrijft op pagina 2 dat het belangrijk is om als gebruiker rekening te houden dat het Algoritme uitsluitend gebruikt dient te worden voor de doelgroep die "geschikt is om te werken". Het Algoritme dient niet ingezet te worden voor mensen die hier buiten vallen. Vergunningshouders die niet onderdeel zijn van de arbeidsmarkt worden daarmee niet geplaagd door het Algoritme.

Met bovenstaande beschrijving heeft COA opvolging gegeven aan de aanbeveling. De bevinding is hiermee gesloten.

### 4.7.2. Bevinding 18

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Bevinding:** De beschrijving en de accuraatheid van de impact op loon analyse is niet transparant. Dit heeft effect op de betrouwbaarheid van de impact op loon analyse en van de financiële voordelen analyse.
- **Aanbeveling:** Wees transparant over de accuraatheid, en daarmee de betrouwbaarheid, van de impact op loon analyse. Verder is het aanbevolen om niet actief te sturen op de uitkomsten van de impact op loon en financiële voordelen analyse voordat dit duidelijk is.

## Stap 1 van Analyse 2

In Stap 1 beschrijft COA in sectie 6.g.ii van de Technische Documentatie 1 de analyse waaruit de financiële voordelen van het Algoritme blijken verduidelijkt, waarbij op de aannames wordt gewezen en een betrouwbaarheidsinterval wordt gegeven. COA is voornemens dit op te nemen in een stakeholder voorstel.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van het stakeholder voorstel.

## Stap 2 van Analyse 2

Tijdens Stap 2 van de Analyse heeft COA aangegeven dat een stakeholder voorstel niet onderdeel zal zijn van de algoritmedocumentatie. Verder heeft COA aangegeven dat de analyse over de financiële voordelen van het Algoritme niet gecommuniceerd zal worden naar andere partijen of gebruikt zal worden in de afweging al dan niet met de Pilot van het Algoritme te zullen starten.

Desondanks staan de financiële voordelen beschreven in de Concept Invulijst Algoritmeregister en wordt de financiële analyse gebruikt als onderbouwing van het doel van het Algoritme. Verder worden de financiële voordelen ook als onderdeel van de argumentatie gebruikt voor het gebruiken van de persoonsgegevens land van herkomst en moedertaal (zie bevinding 20). Naast het ontbreken van een beschrijving over de betrouwbaarheid van de financiële voordelen, is dit getal ook gebaseerd op een gedateerde versie van het Algoritme zoals ook beschreven in Sectie 4.3.4.

Deloitte observeert dat over de financiële voordelen van het Algoritme gecommuniceerd wordt naar interne betrokkenen met mogelijk externe communicatie als gevolg terwijl de accuraatheid van dit getal onduidelijk en verouderd is. Het risico ontstaat hierdoor dat de financiële voordelen van het Algoritme worden gezien als rechtvaardiging voor het gebruik van het Algoritme, zonder dat hierbij wordt gewezen op de accuraatheid van dit getal.

De aanbeveling is om transparant te zijn over de accuraatheid, en daarmee de betrouwbaarheid, van de impact op loon analyse en de financiële voordelen van het Algoritme. Deze aanbeveling kan ook opgevolgd worden door de financiële voordelen van het Algoritme niet te communiceren en deze geen rol te laten spelen in de besluitvorming.



29 april 2024  
KD/wk/24-0725

| Nr. | Categorie    | Observatie                                                                                                                                                                                               | Aanbeveling                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|-----|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 18  | Documentatie | De financiële voordelen van het Algoritme worden gecommuniceerd naar interne betrokkenen met mogelijk externe communicatie als gevolg terwijl de accuraatheid van dit getal onduidelijk en verouderd is. | <p>Wees transparant over de accuraatheid, en daarmee de betrouwbaarheid, van de impact op loon analyse en de financiële voordelen van het Algoritme.</p> <p>Deze aanbeveling kan ook opgevolgd worden door de financiële voordelen van het Algoritme niet te communiceren, en deze dus ook niet op te nemen in het algoritmeregister en geen rol te laten spelen in besluitvorming rond het gebruik van persoonsgegevens (zie bevinding 20).</p> |

#### 4.8. AVG

Als onderdeel van de analyse van de dimensie 'AVG' is in Analyse 1 het volgende uitgevoerd:

- Er wordt bekeken of de ethische aspecten van 'Privacy' terug komen in de door het COA uitgevoerde AIIA en PIA.

##### 4.8.1. Bevinding 19

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Bevinding:** Het COA heeft geen inzage in de door IPL uitgevoerde analyses naar de integratiewinst van het gebruik van de bijzondere persoonsgegevens.
- **Aanbeveling:** Verrijk de algoritmedocumentatie met een onderbouwing van de noodzakelijkheid en evenredigheid van de afzonderlijke bijzondere persoonsgegevens (niet als een geheel). Hiervoor is inzage in de analyses van het IPL essentieel. Na afloop van de analyse heeft het COA aangegeven dat de volgende versie van het Algoritme geen gebruik zal maken van bijzondere persoonsgegevens.

#### Stap 1 van Analyse 2

In het Backtest Rapport geeft COA inzichten in de prestaties van het Algoritme met het gebruik van bijzondere persoonsgegevens (namelijk religie en etniciteit) en zonder bijzondere persoonsgegevens. COA heeft tijdens Stap 1 aangegeven dat het Algoritme niet de kenmerken religie en etniciteit zal gebruiken. Het Algoritme zal wel kenmerken als land van herkomst en moedertaal gaan gebruiken. De combinatie van deze kenmerken kan leiden tot het gebruik van een bijzonder persoonsgegeven. Dit wordt verder geanalyseerd onder bevinding 20.

COA heeft opvolging gegeven aan de aanbeveling door geen gebruik meer te maken van de kenmerken religie en etniciteit. De bevinding is daarmee gesloten.



29 april 2024  
KD/wk/24-0725

#### 4.8.2. Bevinding 20

Tijdens Analyse 1 van Deloitte is de volgende bevinding gemaakt:

- **Observatie:** Het COA heeft geen standpunt gevormd over het gebruik van de bijzondere persoonsgegevens.
- **Aanbeveling:** Werk het standpunt uit of het COA bijzondere persoonsgegevens mag en wil gebruiken. Toets het standpunt van het COA met relevante belanghebbenden. Na afloop van de analyse heeft het COA aangegeven dat de volgende versie van het Algoritme geen gebruik zal maken van bijzondere persoonsgegevens.

##### Stap 1 van Analyse 2

In de algoritmedocumentatie van Stap 1 worden de bijzondere persoonsgegevens religie en etniciteit niet langer gebruikt in het Algoritme. Kenmerken als land van herkomst en moedertaal worden wel gebruikt in het Algoritme. De Algemene verordening gegevensbescherming ("AVG") beschouwt persoonsgegevens waaruit ras of etnische afkomst blijkt als bijzondere persoonsgegevens. Het risico dat het gezamenlijke gebruik van kenmerken als land van herkomst en moedertaal leidt tot het gebruik van een bijzonder persoonsgegeven is onvoldoende onderzocht en gedocumenteerd door COA.

COA heeft aangegeven de privacy impact assessment ("PIA") in Stap 2 aan te leveren, waarin onderbouwing, proportionaliteit, evenredigheid en mogelijke wettelijke grondslag van de gebruikte attributen zal worden gegeven.

In de Tussentijdse Rapportage heeft Deloitte aangegeven dat deze bevinding open blijft en tijdens Stap 2 opnieuw geanalyseerd zal worden op basis van de PIA.

##### Stap 2 van Analyse 2

De PIA is tijdens Stap 2 niet door COA aangeleverd. COA heeft aangegeven in de Bevindingen Status Excel 2 dat op 25 maart (later uitgesteld naar april) de beslissing door de directie zal worden gemaakt welke persoonsgegevens door het Algoritme gebruikt zullen worden voordat de Pilot start. COA heeft destijds een voorlopige onderbouwing voor het gebruik van de persoonsgegevens land van herkomst en moedertaal gedeeld per mail<sup>2</sup>. COA heeft in andere algoritmedocumentatie geen onderbouwing gegeven waarom de attributen land van herkomst en moedertaal gebruikt kunnen worden. Voor nu gaat Deloitte er daarom vanuit dat de persoonsgegevens land van herkomst en moedertaal gebruikt zullen worden tijdens de Pilot.

Deloitte heeft een analyse gedaan naar het gebruik van deze kenmerken. In sommige gevallen kan uit de gevoelige persoonsgegevens land van herkomst en moedertaal gezamenlijk iemands ras of etniciteit blijken.

<sup>2</sup> Deze mail is verstuurd op 31-5-2023, met onderwerp 'Onderdeel 12 PIA m.b.t. Bijzonder persoonsgegevens (aanbeveling 20)'.



29 april 2024  
KD/wk/24-0725

De autoriteit persoonsgegevens kijkt in dit geval naar drie criteria (zie bijvoorbeeld *gebruik van camerabeelden*<sup>3</sup>):

- Het doeleinde van de verwerking is niet gericht op het verwerken van bijzondere persoonsgegevens, dan wel op het onderscheid maken op grond van een bijzonder persoonsgegeven.
- Het voor de verantwoordelijke redelijkerwijs niet voorzienbaar is dat de verwerking zal leiden tot het maken van onderscheid op grond van een bijzonder persoonsgegeven.
- De verwerking van die bijzondere persoonsgegevens onvermijdelijk is bij die verwerking.

De voorlopige onderbouwing van COA gaat in op punt 1 en 2, maar geeft geen onderbouwing voor punt 3:

- Voor punt 1 wordt aangegeven dat het doel van de verwerking niet is om onderscheid te maken naar ras of etniciteit. Het doel is juist om individuele bewoners op een gelijkwaardige, eerlijke en transparante wijze te positioneren zodat zij meer kans op succes hebben op arbeidsmarkt.
- Voor punt 2 wordt aangegeven dat het gebruik van land van herkomst en moedertaal niet tot een voor het COA voorzienbare wijze leidt tot onderscheid naar ras of etnisch afkomst. De onderbouwing hiervoor is dat er maatregelen worden getroffen om ongewenste uitkomsten van de AI-oplossing te voorkomen en, indien dergelijke uitkomsten voorkomen, de AI-oplossing bij te sturen. Daarnaast heeft de AI-oplossing niet als doel om individuele bewoners dan wel bewonersgroepen uit te sluiten, te stigmatiseren of te discrimineren.
- Voor punt 3 wordt hier niet specifiek op ingegaan.

In de onderbouwing van COA wordt niet ingegaan op de noodzakelijkheid om de gevoelige persoonsgegevens land van herkomst en moedertaal te gebruiken. Ook wordt niet ingegaan op de sterke correlatie tussen de combinatie van land van herkomst en moedertaal gezamenlijk en iemands ras of etniciteit. Uit een rapport van de autoriteit persoonsgegevens<sup>4</sup> blijkt dat iemands ras of etniciteit soms kan blijken uit een combinatie van persoonsgegevens: *"Indien nationaliteit wordt verwerkt in combinatie met bijvoorbeeld geboorteland, geboorteplaats, herkomst en/of pasfoto wordt in rechtspraak aangenomen dat er wel sprake is van gegevens waaruit het ras of de etnische afkomst blijkt."* Daarnaast wordt in de onderbouwing in de PIA niet verwezen naar de eerlijkeheidsanalyse die tijdens de backtest is uitgevoerd.

Deloitte observeert dat het standpunt van het COA over het gebruik van de persoonsgegevens land van herkomst en moedertaal onvoldoende is gedocumenteerd en uitgewerkt.

Het is niet gedocumenteerd of het gezamenlijke gebruik van kenmerken als land van herkomst en moedertaal leidt tot het gebruik van een bijzonder persoonsgegeven, i.e., ras of etnische afkomst. Het risico ontstaat dat er onverantwoord gebruik wordt gemaakt van persoonsgegevens door COA in de Pilot.

De aanbeveling is om het standpunt uit te werken of het COA deze persoonsgegevens mag en wil gebruiken en dit standpunt te toetsen met relevante belanghebbenden, zoals een functionaris gegevensbescherming en/of bedrijfsjurist van COA.

### Update naar aanleiding van nieuwe informatie aangeleverd op 10 april

COA heeft op 19 maart een voorstel aan haar Stuurgroep voorgelegd over het gebruik van de persoonsgegevens land van herkomst en moedertaal. De Stuurgroep heeft besloten dat het Algoritme wel gebruik zal maken van land van herkomst en moedertaal op basis van drie argumenten:

1. Een functioneel argument: COA geeft aan dat mensen die hetzelfde land van herkomst hebben en dezelfde taal spreken een natuurlijk netwerk vormen dat helpt bij integratie naar werk.

<sup>3</sup> [https://www.autoriteitpersoonsgegevens.nl/uploads/imported/beleidsregels\\_cameratoezicht-.pdf](https://www.autoriteitpersoonsgegevens.nl/uploads/imported/beleidsregels_cameratoezicht-.pdf)

<sup>4</sup> [https://autoriteitpersoonsgegevens.nl/uploads/imported/onderzoek\\_belastingdienst\\_kinderopvangtoeslag.pdf](https://autoriteitpersoonsgegevens.nl/uploads/imported/onderzoek_belastingdienst_kinderopvangtoeslag.pdf)



29 april 2024  
KD/wk/24-0725

2. Een statistisch argument: COA geeft aan dat het gebruik van land van herkomst en moedertaal zal leiden tot 4% meer integratie<sup>5</sup>, en zal zorgen voor 12.5% hogere besparingen voor de maatschappij<sup>6</sup>, ten opzichte van het niet gebruiken van deze gegevens.
3. Er wordt aangegeven dat er zorgen kunnen zijn dat land van herkomst en moedertaal gezamenlijk een indicator zijn van etniciteit. In dat geval mogen ze niet worden gebruikt. IPL geeft aan géén statistisch verband te vinden tussen land van herkomst, moedertaal en etniciteit.

Daarnaast is als mitigerende maatregel besloten dat IPL zal gaan monitoren op etniciteitsgroepen, zodat IPL kan bewaken dat er geen etnische groepen worden benadeeld. Hiervoor zal IPL toegang krijgen tot de etniciteitsdata.

COA heeft aangegeven niet te zullen sturen op de financiële voordelen van het Algoritme. Deloitte observeert dat bij het besluit om gebruik te maken van de persoonsgegevens land van herkomst en moedertaal wederom de verwachte besparingen die het Algoritme zal genereren zijn meegenomen in de argumentatie voor de besluitvorming (zie bevinding 18).

Verder observeert Deloitte dat de juridische onderbouwing voor het mogen gebruiken van de persoonsgegevens land van herkomst en moedertaal ontbreekt. Dit geldt ook voor de juridische onderbouwing voor het gebruik van etniciteit voor monitoringdoeleinden. COA heeft aangegeven dat deze juridische onderbouwing niet alleen voor dit Algoritme ontbreekt, maar een breder probleem is waar ze zich bewust van is en aan werkt.

Verder observeert Deloitte dat nog niet is uitgewerkt hoe de monitoring op etniciteit zal gaan plaatsvinden.

Als laatste observeert Deloitte dat, met betrekking tot het derde argument, details ontbreken over de door IPL uitgevoerde analyse waaruit blijkt dat er geen statistisch verband is tussen land van herkomst, moedertaal en etniciteit. Uit een publicatie van de autoriteit persoonsgegevens blijkt dat iemands ras of etniciteit soms kan blijken uit een combinatie van persoonsgegevens (zoals eerder besproken in dit hoofdstuk, zie ook voetnoot 4). Vanwege deze indicatie is meer detail gewenst over de door IPL uitgevoerde analyse.

Overkoepelend observeert Deloitte dat COA onderbouwd heeft dat het gebruik wil maken van de persoonsgegevens land van herkomst en moedertaal in het Algoritme, en dat IPL zal gaan monitoren dat er geen etnische groepen worden benadeeld. Echter, een juridische onderbouwing ontbreekt nog voor het gebruik van deze persoonsgegevens. COA heeft aangegeven dat deze juridische onderbouwing niet alleen voor dit Algoritme ontbreekt, maar een breder probleem is waar ze zich bewust van is en aan werkt. Ook is nog niet uitgewerkt hoe de monitoring op etniciteit zal gaan plaatsvinden. Als laatste ontbreken details over de door IPL uitgevoerde analyse waarmee wordt aangetoond dat er geen statistisch verband is tussen land van herkomst, moedertaal en etniciteit. Meer details over de uitgevoerde analyse zijn gewenst om de uitkomst te kunnen plaatsen.

---

<sup>5</sup> Uit de analyse blijkt dat de gemiddelde 'employment rate after year 1' met 4% zal toenemen van 15.33% tot 15.95%.

<sup>6</sup> Hierbij wordt gebruik gemaakt van de impact op loon analyse. Zie ook bevinding 18.



29 april 2024  
KD/wk/24-0725

De aanbeveling is om de juridische onderbouwing te finaliseren voor het gebruik van persoonsgegevens, waaronder de persoonsgegevens land van herkomst, moedertaal en etniciteit (voor bias monitoringdoeleinden), en deze onderbouwing vast te leggen in de DPIA. Daarnaast is de aanbeveling meer details te geven over de door IPL uitgevoerde analyse, en uit te werken hoe de monitoring op etniciteitsgroepen zal plaatsvinden.

| Nr. | Categorie    | Observatie                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Aanbeveling                                                                                                                                                                                                                                                                                                                                                              |
|-----|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20  | Documentatie | <p>COA heeft onderbouwd dat het gebruik wil maken van de persoonsgegevens land van herkomst en moedertaal in het Algoritme, en dat IPL zal gaan monitoren dat er geen etnische groepen worden benadeeld. Echter, een juridische onderbouwing ontbreekt nog voor het mogen gebruiken van deze persoonsgegevens.</p> <p>COA heeft aangegeven dat deze juridische onderbouwing niet alleen voor dit Algoritme ontbreekt, maar een breder probleem is waar ze zich bewust van is en aan werkt. Ook is nog niet uitgewerkt hoe de monitoring op etniciteit zal gaan plaatsvinden. Als laatste ontbreken details over de door IPL uitgevoerde analyse waarmee wordt aangetoond dat er geen statistisch verband is tussen land van herkomst, moedertaal en etniciteit. Meer details over de uitgevoerde analyse zijn gewenst om de uitkomst te kunnen plaatsen.</p> | <p>Finaliseer de juridische onderbouwing voor het gebruik van persoonsgegevens, waaronder de persoonsgegevens land van herkomst, moedertaal en etniciteit (voor bias monitoringdoeleinden), en leg dit vast in de DPIA.</p> <p>Geef meer details over de door IPL uitgevoerde analyse. Werk ook verder uit hoe de monitoring op etniciteitsgroepen zal plaatsvinden.</p> |



29 april 2024  
KD/wk/24-0725

## 5. Disclaimers

De algoritme analyse is niet uitgevoerd in het kader van een assurance-opdracht zoals gedefinieerd in het International Framework for Assurance Engagements van de International Federation of Accountants ("IFAC"). Het is de verantwoordelijkheid van het COA om te beoordelen of de resultaten van de algoritme analyse in het perspectief van het geheel van de hen ter beschikking staande informatie en hun risicoperceptie aan de door hen te stellen eisen voldoen.

Uitsluitend het COA is verantwoordelijk voor onder meer: (a) het nemen van alle managementbeslissingen, het uitoefenen van alle managementfuncties en het dragen van alle managementverantwoordelijkheden, (b) het evalueren van de toereikendheid en de resultaten van de algoritme analyse, (c) het accepteren van de verantwoordelijkheid voor het gebruik van de resultaten van de algoritme analyse, en (d) het inrichten en onderhouden van een intern beheersingssysteem, inclusief het toezicht op de lopende activiteiten.

Deloitte wil benadrukken dat de opgeleverde deliverables (zoals het rapport met daarin de observaties en aanbevelingen) slechts bedoeld zijn voor intern gebruik door het COA ten behoeve van het omschreven doel in de opdrachtbrief met referentie KD/js/23-0546. Verder mag het rapport ook gedeeld worden met de algoritme ontwikkelaars (in dit geval het IPL) en het CBS.

Zonder uitdrukkelijke en voorafgaande schriftelijke toestemming van Deloitte is het niet toegestaan deliverables, dan wel delen daaruit, te gebruiken voor andere doeleinden dan overeengekomen, aan derden te verspreiden of openbaar te maken, aan deliverables te refereren of uit deliverables te citeren.

Op dit rapport zijn de Algemene Voorwaarden Dienstverlening Deloitte Nederland, januari 2020, van toepassing waarvan u bijgaand een exemplaar aantreft.



29 april 2024  
KD/wk/24-0725

## 6. Referenties

- [1] CBS, „Analyse mogelijke risico's inputdata Kansrijke Koppeling,” 24 maart 2022.
- [2] Deloitte, „Q&A Algoritme Analyse,” februari, maart en april 2024.
- [3] Algemene Rekenkamer, „Aandacht voor Algoritmes,” 26 januari 2021.



## PROJECT INITIATION DOCUMENT- SAMENVATTING

| Project Initiation Document (PID) | Samenvatting                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Projectdefinitie</b>           | "Kansrijke Koppeling met behulp van AI" is ontworpen om de overgang van statushouders naar de Nederlandse samenleving te vergemakkelijken door hen te koppelen aan regio's waar hun vaardigheden en ervaring het meest waardevol en nodig zijn. Dit wordt bereikt door een algoritme dat rekening houdt met een reeks variabelen, waaronder de individuele vaardigheden, kwalificaties en voorkeuren van de statushouders, evenals de specifieke behoeften en kansen van verschillende arbeidsmarktregio's. |
| <b>Impact Meerjarenstrategie</b>  | De inzet van artificiële intelligentie kan bijdragen aan de realisatie van diverse tactische (t), operationele (o) en dwarsdoorsnijdende doelstellingen / thema's van de COA Meerjarenstrategie 2020-2025, die in het PID nader worden toegelicht.                                                                                                                                                                                                                                                          |
| <b>Samenvatting projectkosten</b> | Onderzoeksfase € 5.1.1.c euro<br>Pre-pilotfase € 5.1.1.c euro<br>Pilotfase € 5.1.1.c (5.1.1.c x 3 jaar)                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Projectorganisatie</b>         | 5.1.2.e 5.1.2.e 5.1.2.e<br>5.1.2.e 5.1.2.e 5.1.2.e<br>5.1.2.e 5.1.2.e                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <b>Initiële projectplanning</b>   | De planning van het project loopt van Q4 2022 t/m Q3 2025, waarvan het eerste half jaar (Q4 2022 – Q1 2023) gebruikt wordt voor de prepilot en de resterende tijd (Q2 2023 – Q3 2025) van in totaal 3 jaar gebruikt wordt voor (een deel van) de pilot-uitvoering.                                                                                                                                                                                                                                          |
| <b>Beheersingsmechanismen</b>     | Stuurgroep en hoofdpuntenrapportage (om de maand)                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b>Project risico's</b>           | Het projectteam is zich zeer bewust van de verschillende risico's binnen dit project, verdeeld over de volgende categorieën: externe leverancier, draagvlak en cultuur, veranderende organisatie, datakwaliteit en data governance.                                                                                                                                                                                                                                                                         |
| <b>Borging</b>                    | In het project wordt het nader te bepalen (beleidsteam / Informatiediensten) team opgeleid om het project te borgen na implementatie.                                                                                                                                                                                                                                                                                                                                                                       |

## Versiebeheer

|                   |                    |            |                                                                                                                     |
|-------------------|--------------------|------------|---------------------------------------------------------------------------------------------------------------------|
| <b>Versie 0.1</b> | 5.1.2.e<br>5.1.2.e | 22-7-2022  | Initiele opzet tbv Amif-subsidie                                                                                    |
| <b>Versie 0.2</b> | 5.1.2.e            | 13-10-2023 | Nieuw PID                                                                                                           |
| <b>Versie 1.0</b> | 5.1.2.e            | 12-12-2023 | Verwerken redactie                                                                                                  |
| <b>Versie 1.1</b> | 5.1.2.e<br>5.1.2.e | 30-4-2024  | <ul style="list-style-type: none"> <li>• Pilotlocaties aangepast</li> <li>• Mijlpalenplanning bijgewerkt</li> </ul> |



## Inhoud

|           |                                                                    |           |
|-----------|--------------------------------------------------------------------|-----------|
| <b>1.</b> | <b>INLEIDING .....</b>                                             | <b>4</b>  |
| <b>2.</b> | <b>ACHTERGROND &amp; AANLEIDING PROJECT.....</b>                   | <b>5</b>  |
| 2.1       | Achtergrond project.....                                           | 5         |
| 2.2       | Business Case.....                                                 | 5         |
| 2.3       | Doelstellingen en Verwachtingen.....                               | 7         |
| 2.4       | Samenwerkingspartners .....                                        | 7         |
| <b>3.</b> | <b>DEFINITIE.....</b>                                              | <b>8</b>  |
| 3.1       | Projectgrenzen.....                                                | 8         |
| 3.2       | Buiten scope.....                                                  | 9         |
| 3.3       | Doelgroep.....                                                     | 9         |
| 3.4       | Pilotlocaties.....                                                 | 10        |
| 3.5       | Eindresultaat .....                                                | 10        |
| 3.6       | Relaties met andere projecten / beleidsthema's .....               | 11        |
| <b>4.</b> | <b>PROJECTAANPAK.....</b>                                          | <b>12</b> |
| 4.1       | Methodologie .....                                                 | 12        |
| 4.2       | Strategieën .....                                                  | 12        |
| 4.3       | Technieken.....                                                    | 15        |
| 4.4       | Planning .....                                                     | 16        |
| <b>5.</b> | <b>RISICO'S.....</b>                                               | <b>17</b> |
| <b>6.</b> | <b>BUDGET, FINANCIERINGSBRONNEN EN FINANCIEEL MANAGEMENT .....</b> | <b>20</b> |
| 6.1       | Aangevraagde financiering .....                                    | 20        |
| 6.2       | Financieringsbronnen .....                                         | 21        |
| 6.3       | Financieel Managementstrategie .....                               | 21        |
| <b>7.</b> | <b>PROJECTORGANISATIE EN ROLLEN .....</b>                          | <b>22</b> |
| <b>8.</b> | <b>BORGING.....</b>                                                | <b>24</b> |



## 1. Inleiding

Dit Project Initiatie Document (PID) is opgesteld om een gedetailleerd overzicht te bieden van het project "Kansrijke Koppeling met behulp van AI", dat tot doel heeft een algoritme te implementeren voor het effectief koppelen van statushouders aan arbeidsmarktregio's. Het document dient als een gids voor alle stakeholders, met inbegrip van projectteamleden, het management en directie van het COA, en externe partners, en biedt uitgebreide informatie over de planning, uitvoering, en monitoring van het project. Het PID is gestructureerd in verschillende secties om een georganiseerd overzicht van het project te bieden:

1. **Achtergrond en inleiding:** Een overzicht van de projectachtergrond, doelstellingen, en de business case, waarbij de waarde en het belang van het project worden benadrukt.
2. **Definitie:** Een gedetailleerde beschrijving van de projectgrenzen, deliverables, en uitsluitingen, om duidelijke verwachtingen te stellen.
3. **Projectaanpak:** Een overzicht van de methodologieën, strategieën, en technieken die zullen worden gebruikt voor de implementatie van het project. Ook de planning komt in dit hoofdstuk terug.
4. **Risico's:** Een analyse van potentiële risico's, uitdagingen, en de strategieën voor mitigatie en management.
5. **Kosten:** Een gedetailleerde begroting, met een overzicht van de kosten.
6. **Projectteam en Rollen:** Een presentatie van het projectteam, inclusief de rollen, verantwoordelijkheden van elk lid.
7. **Stakeholderanalyse:** Een overzicht van de belangrijkste stakeholders, hun belangen, en strategieën voor verbinding en communicatie.
8. **Borging:** Een plan voor het borgen van de deliverables van het project.

Dit PID is ontworpen om transparantie, duidelijkheid, en richting te bieden gedurende de gehele levenscyclus van het project. Het kan worden herzien en bijgewerkt vanwege veranderende omstandigheden en nieuwe informatie. Het dient als een levend document dat de basis vormt voor besluitvorming, coördinatie, en communicatie, en waarborgt dat het project op koers blijft om zijn doelstellingen te bereiken. Het PID wordt door de stuurgroep geaccordeerd.

## 2. Achtergrond & aanleiding project

### 2.1 Achtergrond project

In het licht van de voortdurende inspanningen om de integratie en participatie van statushouders in Nederland te bevorderen, initieert het Centraal Orgaan opvang Asielzoekers (COA) een project genaamd "Kansrijke Koppeling met behulp van AI". Dit project is gewijd aan het ontwikkelen en implementeren van een innovatief algoritme dat tot doel heeft statushouders effectief te koppelen aan passende arbeidsmarktregio's.

#### Projectbeschrijving

"Kansrijke Koppeling met behulp van AI" is ontworpen om de overgang van statushouders naar de Nederlandse samenleving te vergemakkelijken door hen te koppelen aan regio's waar hun vaardigheden en ervaring het meest waardevol en nodig zijn. Dit wordt bereikt door een algoritme dat rekening houdt met een reeks variabelen, waaronder de individuele vaardigheden, kwalificaties en voorkeuren van de statushouders, evenals de specifieke behoeften en kansen van verschillende arbeidsmarktregio's.

### 2.2 Business Case

De implementatie van dit algoritme is gebaseerd op de overtuiging dat een doelgerichte plaatsing van statushouders niet alleen hun individuele kansen op succesvolle integratie en economische zelfstandigheid zal vergroten, maar ook zal bijdragen aan de diversiteit en vitaliteit van de Nederlandse arbeidsmarkt.

De business case voor "Kansrijke Koppeling met behulp van AI" is gebaseerd op de volgende overwegingen:

- **Effectiviteit:** Door statushouders te koppelen aan regio's waar hun vaardigheden het meest nodig zijn, wordt de kans op succesvolle integratie en participatie vergroot.
- **Economische impact:** Het project zal bijdragen aan het verminderen van de werkloosheid onder statushouders en hun economische bijdrage aan de samenleving vergroten.
- **Sociale cohesie:** Door de integratie en participatie van statushouders te bevorderen, wordt de sociale cohesie en diversiteit binnen de Nederlandse samenleving versterkt.

De business case wordt versterkt door de volgende aanvullende overwegingen:

*Objectievere Besluitvorming*



- **Data-Gedreven inzichten:** Het algoritme zal het COA in staat stellen om beslissingen te nemen op basis van objectievere en kwantitatieve data. Dit vermindert de afhankelijkheid van subjectieve beoordelingen en intuïtie, wat leidt tot meer geïnformeerde en betrouwbare beslissingen.

#### *Organisatorische intelligentie*

- **Leer- en verbetercultuur:** De integratie van het algoritme bevordert een cultuur van continu leren en verbeteren binnen het COA. Het biedt mogelijkheden voor feedback en aanpassingen, wat de organisatorische deskundigheid en lerende cultuur verbetert.
- **Kennisopbouw:** Het verzamelen en analyseren van data draagt bij aan de opbouw van een kennisbasis. Dit stelt het COA in staat om trends en patronen te identificeren, toekomstige behoeften te voorspellen en proactieve strategieën te ontwikkelen.

#### *Verhoogde transparantie*

- **Stakeholdercommunicatie:** Het gebruik van objectieve data verhoogt de transparantie in de communicatie met stakeholders zoals gemeentes. Het biedt duidelijke en meetbare indicatoren in het matchen van statushouders aan gemeentes, wat het vertrouwen en de geloofwaardigheid van het COA versterkt.

#### *Duurzame integratie*

- **Aangepaste ondersteuning:** Het algoritme kan het COA op langere termijn in staat stellen om de ondersteuning aan statushouders te personaliseren op basis van hun specifieke behoeften, vaardigheden en voorkeuren. Dit bevordert een duurzame integratie en participatie.
- **Innovatie en groei:** De implementatie van het algoritme markeert een belangrijke stap in de digitale transformatie van het COA. Het opent deuren voor verdere innovaties en draagt bij aan de groei en ontwikkeling van de organisatie.

#### *Arbeidsmarktparticipatie*

- Het algoritme wordt als positief en bruikbaar gezien als er een significante stijging van 4%-punt toename arbeidsparticipatie statushouders in het eerste jaar na uitstroom naar een gemeente. T.o.v. 12% in de huidige situatie, is dat een relatieve verbetering van 33% op arbeidsmarktparticipatie is ten opzichte van de testgroep die geen match beschikbaar krijgt. Dit percentage is afkomstig uit het [vooronderzoek van IPL](#) van mei 2022.

In de context van deze business case wordt het duidelijk dat "Kansrijke Koppeling met behulp van AI" niet alleen een instrument is voor betere plaatsing van statushouders, maar ook een vliegwiel voor organisatorische ontwikkeling, innovatie en een data gedreven cultuur.

### 2.3 Doelstellingen en Verwachtingen

Het primaire doel van dit project is het ontwikkelen en implementeren van het algoritme om de koppeling van statushouders aan arbeidsmarktregio's te optimaliseren. We streven ernaar om door middel van dit initiatief de toekomstige levenskwaliteit van statushouders te verbeteren, de integratieprocessen te versnellen en bij te dragen aan de arbeidsparticipatie in Nederland.

Statushouders hebben op dit moment een minder zekere positie op de arbeidsmarkt en hun arbeidsparticipatie is relatief laag.<sup>1</sup> De arbeidsparticipatie van vergunninghouders verhogen aan de hand van artificiële intelligentie (AI). Het betreft hier vergunninghouders die met ondersteuning van een algoritme worden gekoppeld aan een zo kansrijk mogelijke arbeidsmarktregio op basis van hun kenmerken.<sup>2</sup> De verhoging in arbeidsparticipatie wordt gerealiseerd (en gemonitord) wanneer de vergunninghouders al zijn uitgestroomd uit de COA-opvang.

Dit project kan bijdragen aan de realisatie van diverse tactische (t) en operationele (o) doelstellingen van de COA Meerjarenstrategie 2020-2025.

- Kansrijke koppeling nieuwe statushouders (o)
- Creëren van gezamenlijke oplossingen voor nationale en lokale maatschappelijke vraagstukken (t)
- Resultaten en meer draagvlak door goede samenwerking met betrokken partners (o)
- Begrip, draagvlak en reputatie door open proactieve communicatie (o) en door integratie van de statushouder in de Nederlandse maatschappij.

### 2.4 Samenwerkingspartners

In dit project werken we samen met meerdere organisaties. Dit zijn onder andere:

- Immigrant Policy Lab (IPL)  
IPL is een onderzoeksbureau gespecialiseerd in de ontwikkeling van geavanceerde algoritmen, om ons te assisteren bij het ontwerpen, testen en implementeren van de 'Kansrijke Koppeling' oplossing. IPL is ontstaan vanuit een samenwerking tussen de Stanford University in San Francisco en de Eidgenössische Technische Hochschule Zürich (ETH) in Zurich.
- Centraal Bureau voor de Statistiek (CBS)  
Het CBS helpt ons project met twee belangrijke onderdelen. We hebben CBS gevraagd te helpen met het identificeren van mogelijke risico's bij toepassing van artificiële intelligentie in het plaatsingsproces van statushouders. Daarnaast levert CBS data en trendinformatie, waarmee de COA-data verrijkt wordt)

---

<sup>1</sup> <https://www.cbs.nl/nl-nl/nieuws/2021/15/arbeidsparticipatie-statushouders-stagneert>

<sup>2</sup> <https://www.cpb.nl/sites/default/files/omnidownload/CPB-Policy-Brief-Investeren-in-de-arbeidsmarktintegratie-van-statushouders.pdf>

- Een externe auditkantoor. Voor 2022 en 2023 is Deloitte aangesloten vanuit de mantelpartij om de technische en ethische elementen van het algoritme te auditen. Voor eventuele toekomstige of vervolgaudits worden de aanbestedingsregels vanuit de Rijksoverheid gevolgd.

### 3. Definitie

#### 3.1 Projectgrenzen

##### *2.1.1 Ruimtelijke en Thematische Grenzen*

Het "Kansrijke Koppeling met behulp van AI" project is specifiek gericht op de Nederlandse context, met een nadruk op het koppelen van statushouders aan arbeidsmarktregio's binnen het land. Het algoritme wordt ontwikkeld om rekening te houden met de unieke sociaal-economische, culturele en geografische kenmerken van de verschillende (arbeidsmarkt) regio's in Nederland.

##### *2.1.2 Projectfasering*

Het project kent drie fases, een onderzoeksfase, een pre-pilotfase en een pilotfase. In de onderzoeksfase staat de haalbaarheid van het project centraal. Aan welke technische, organisatorische en juridische onderwerpen moeten we voldoen om een succesvolle pilot te kunnen opzetten? Er wordt een vooronderzoek uitgevoerd om te bepalen of een algoritme in staat is om een stijging in arbeidsmarktparticipatie te kunnen verzorgen. In deze fase wordt ook het research design opgesteld om de effectiviteit, nauwkeurigheid en controleerbaarheid van het algoritme te kunnen controleren.

De pilot is opgedeeld in een pre-pilotfase en een pilotfase. In de pre-pilotfase wordt de pilot voorbereid, de eindgebruikers betrokken, trainingsmaterialen gemaakt en vindt de implementatie van het algoritme binnen de COA-applicaties plaats. De pilotfase is de periode waarin we met een zestal locaties gaan starten met gebruik van het algoritme. Dit is ook de fase waarin er sprake is van evaluatie zowel van gebruik door eindgebruikers als van werking.

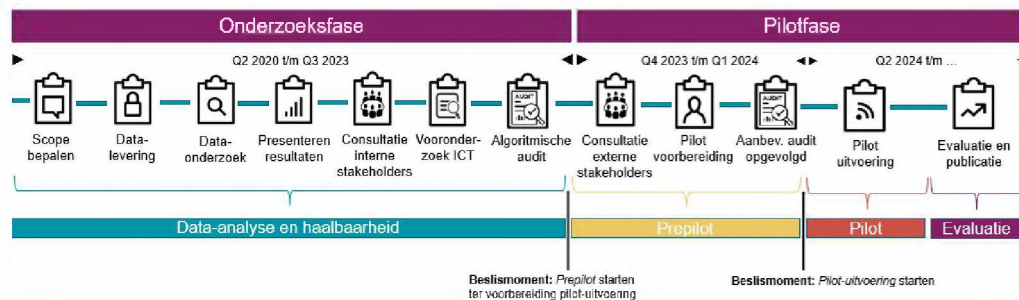
##### *2.1.3 Tijdsgrenzen*

Het project bestaat uit twee delen. Een onderzoeksdeel en een pilotdeel.

Het onderzoeksdeel is gepland van het tweede kwartaal (Q2) 2020 tot en met het derde kwartaal (Q3) 2023.

Het pilotproject is opgedeeld in een pre-pilotfase en een pilotfase. De pre-pilot is gepland van Q4 2023 tot en met Q1 2024. De pilotfase is gepland van Q2 2024 tot en met Q2 2025. Alle projectactiviteiten, inclusief de ontwikkeling, implementatie en evaluatie van het algoritme, zullen binnen deze periode worden voltooid.

## Projectplanning



### 3.2 Buiten scope

1. **Buitenlandse implementatie:** Hoewel het algoritme potentieel aanpasbaar is voor andere contexten, is de huidige scope van het project uitsluitend gericht op Nederland.
2. **Directe plaatsing:** Het project omvat de ontwikkeling en implementatie van het algoritme als beslissingsondersteunend instrument, maar koppeling van statushouders aan de arbeidsmarktregio's blijft een taak van de regievoerder en valt buiten de projectscope.
3. **Onderhoud en updates:** Het project omvat de initiële implementatie van het algoritme. Toekomstig onderhoud, updates en verbeteringen vallen buiten de huidige projectscope en zullen worden aangepakt in vervolgprojecten of onderhoudscontracten.

### 3.3 Doelgroep

De doelgroep van dit project bestaat uit medewerkers van het COA die betrokken zijn bij het proces van screening en matching, waarbij statushouders aan gemeenten worden gekoppeld. Dit proces is met name relevant voor programmabegeleiders en regievoerders.

Aan het bestaande proces dat statushouders in COA-opvangcentra doorlopen, wordt een beslissingsondersteunende applicatie toegevoegd. Statushouders worden ervan op de hoogte gesteld dat ze gekoppeld worden. Bij de koppeling wordt gebruikgemaakt van zowel zachte als harde criteria. De applicatie geeft echter alleen advies aan de programmabegeleider als er geen harde criteria van toepassing zijn, die als volgt zijn:

1. Familie in de eerste graad: Er bestaat een ouder-kindrelatie.

2. Arbeid: Er is sprake van een arbeidscontract voor minimaal 8 uur per week gedurende het verblijf in de opvang.
3. Studie: De betrokkene is toegelaten tot een onderwijsinstelling.
4. Medisch of psychosociaal: De vergunninghouder ondergaat een specifieke, niet-overdraagbare medische behandeling in een bepaalde gemeente waarvoor minimaal vier keer per jaar een bezoek aan de behandelaar noodzakelijk is.

Daarnaast zijn nareizigers uitgesloten. Statushouders die gekenmerkt worden als "nareiziger" zijn naar Nederland gekomen voor gezinshereniging en zullen herenigd worden met de referent.

### 3.4 Pilotlocaties

Het project zal in de pilotfase worden uitgevoerd op pilotlocaties, geselecteerd uit de groep referentielocaties die binnen het COA aangewezen zijn om met ICT-toepassingen te experimenteren. Het project start op 6 azc- en 2 pol locaties te weten:

- Azc Zutphen
- Azc Apeldoorn
- Azc Zeist
- Azc Almelo
- Pol Ter Apel
- Azc Grave
- Azc Maastricht
- Pol Wageningen

De medewerkers die met de applicatie gaan werken zijn zowel programmabegeleiders, die verantwoordelijk zijn voor het opstellen van het regio-advies, als ook regievoerders die verantwoordelijk zijn voor de definitieve plaatsing van de statushouder in de arbeidsmarktregio.

### 3.5 Eindresultaat

In elk project, en in het bijzonder in het "Kansrijke Koppeling met behulp van AI" project, is het cruciaal om het beoogde eindresultaat duidelijk en concreet te definiëren. Een goed gedefinieerd eindresultaat stelt alle betrokkenen in staat om met een eenduidige focus en doelgerichtheid te werken, en zorgt voor meetbaarheid en verantwoording gedurende het gehele project.

Dit project kent drie doelstellingen.

1. Dit project heeft een positief eindresultaat als er in de pilotfase minimaal 20 programmabegeleiders en 10 regievoerders training is gevolgd om de GeoMatch-applicatie te gebruiken. Binnen deze pilotfase is 500 keer gebruik gemaakt van de applicatie ter ondersteuning van de match tussen een

statushouders aan een arbeidsmarktregio. Op het moment dat deze doelstelling is bereikt, zullen we het algoritme gaan aanbieden aan een grotere groep binnen het COA. Deze groep kent dezelfde verhouding qua ratio gebruik / succes als de pilotgroep om het succesvol te noemen.

2. Het gebruik van (algoritmische) data als beslissingsondersteuning wordt gezien als een cultuurverandering binnen het COA. Uit het vragenlijstonderzoek zal 1) een gestegen waardering voor (algoritmische) data als beslissingsondersteunend instrument moeten blijken en 2) een gestegen waardering voor meer datagedreven werken.
3. Het project initieert een nieuwe manier van werken, en vraagt samenwerking tussen veel afdelingen (ict, projectbureau, finance, juridische zaken, beleid, uitvoering) Het project is geslaagd als we geslaagd zijn om vanuit deze verschillende afdelingen in staat zijn om gezamenlijk een nieuw product in productie te kunnen nemen.

### 3.6 Relaties met andere projecten / beleidsthema's

Met de volgende projecten binnen het COA heeft dit project raakvlakken en daar zal tijdens de uitvoer van het project rekening mee gehouden moeten worden.

- **Reguliere beleidsthema kansrijke koppeling**  
De kansrijke koppeling wordt structureel doorontwikkeld door het beleidsteam Vroege Integratie en Participatie (VrIP). Gedurende het project kunnen er thema's naar voren komen die buiten scope van het project vallen maar wel kunnen worden meegenomen in de aanpassing van de kansrijke koppeling door VrIP. Er is een werkgroep "processen" waarin de <sup>5.1.2.e</sup> <sup>5.1.2.e</sup> structureel is aangehaakt bij het project en het project kan meenemen in relevante ontwikkelingen binnen het beleidsthema zoals de interne evaluatie van het werkproces die op dit moment plaatsvindt.
- **Flexibilisering asielketen**  
Het programma Flexibilisering Asielketen ondersteunt en faciliteert de organisatie bij de ontwikkeling naar een flexibele en effectieve organisatie. De wijzigingen die worden doorgevoerd binnen dit programma kunnen invloed hebben op het werkproces van de kansrijke koppeling en specifiek de aanpassing van het werkproces door de verrijking met AI. Bijvoorbeeld is er vanuit Flexibilisering en Kansrijke koppeling mbv AI beide de wens om de beperking tot advies binnen de eigen uitplaatsingsregio te verwijderen en n de koppeling (en eventuele doorplaatsing indien capaciteit dit toelaat) zoveel mogelijk in lijn met het regio-advies uit te voeren. Er is desgewenst overleg tussen de (deel) <sup>5.1.2.e</sup> <sup>5.1.2.e</sup> <sup>5.1.2.e</sup>

- *Deelinrichtingsplan Informatiediensten en kwartiermaker gegevensmanagement*  
De huidige afdeling Informatie & Diensten, binnen de directie Bedrijfsvoering van COA, krijgt een nieuwe inrichting op basis van het deelinrichtingsplan afdeling Informatiediensten. Het kerndoel van deze verandering is een versterking van de taken op gebied van informatievoorziening en -management van het COA. Daartoe komt er ook in de nieuwe afdeling een CIO Office waarin het eigenaarschap rondom beleid, advisering, kaderstelling, toepassing en toetsing van de IV-beleidsterreinen architectuur, informatiemanagement, gegevensmanagement, portfoliomanagement & informatieplanning en informatiebeveiliging & privacy wordt gepositioneerd. Omdat gegevensmanagement (inclusief kaders op gebied van algoritmes en AI) een zeer actueel onderwerp is met urgentie is vooroplopend aan de definitieve besluitvorming op het deelinrichtingsplan al een kwartiermaker de opdracht gegeven om gegevensmanagement (beleid, advisering, kaderstelling, toepassing en toetsing daarvan) binnen het COA in te richten. Naast het nieuwe CIO Office valt binnen de afdeling ook het Management Informatie Team (MIT) dat informatieproducten zoals management rapportages en dashboards ontwikkelt en levert.

## **4. Projectaanpak**

### **4.1 Methodologie**

#### **Agile projectmanagement**

Het "Kansrijke Koppeling" project zal de Agile projectmanagementmethodologie gebruiken om flexibiliteit, aanpassingsvermogen en continue verbetering te waarborgen. De Agile aanpak stelt het projectteam in staat om snel te reageren op veranderingen en nieuwe informatie, en om incrementele en iteratieve ontwikkeling te bevorderen.

Ook onze externe ontwikkelpartner, IPL, zal Agile ontwikkelingsmethodieken toepassen om iteratieve en incrementele verbeteringen aan het algoritme te waarborgen. Ze zullen ook zorgen voor uitgebreide testprotocollen om de functionaliteit, betrouwbaarheid en veiligheid van de oplossing te verifiëren.

We werken met JIRA als platform om het Agile projectmanagement. Tevens wordt er conform werkwijze binne het COA gebruik gemaakt van elementen van Prince II, namelijk het gebruik van een stuurgroep en structurele rapportage.

### **4.2 Strategieën**

#### **Samenwerkende werkgroepen**

Het project maakt gebruik van verschillende werkgroepen waarin op een specifiek thema wordt samengewerkt. Deze zijn ingedeeld op thema:

| <i>Naam werkgroep</i> | <i>Onder leiding van</i> | <i>Deliverables</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Processen             | 5.1.2.e                  | <p><b>Personadocument</b><br/>Document met vijf persona's die betrokken zijn bij het project welke als basis worden gebruikt voor de</p> <p><b>Procesbeschrijving</b><br/>Het bestaande proces screening&amp;matching in Mavim wordt aangepast met het gebruik van de GeoMatch-applicatie.</p> <p><b>Requirementanalyse</b><br/>Document met daarin de eisen van gebruikers, voortkomend uit gesprekken met regievoerders en programmabegeleiders.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Adoptie               | 5.1.2.e 5.1.2.e          | <p><b>Werkinstructie</b><br/>Een uitgebreide handleiding voor het gebruik van het algoritme, inclusief stapsgewijze instructies, FAQ's en probleemoplossingsrichtlijnen.</p> <p><b>Trainingssessies</b><br/>Georganiseerde trainingssessies voor het COA-personeel en andere relevante stakeholders om hen vertrouwd te maken met het algoritme en de bijbehorende processen. Er wordt gebruik gemaakt van een flyer, op deze flyer moet in overzichtelijke en logische taal de werking van het algoritme beschreven worden.</p> <p><b>Informatiemateriaal</b><br/>Middels de aankondigingsbrief beëindigen opvang wordt de vergunninghouder tijdens het huisvestingsgesprek geïnformeerd dat COA gebruikt maakt van een algoritme binnen het proces screening&amp;matching. Dit wordt in het huidige proces aangepast en bijgewerkt.</p> <p><b>Opleiding business owner</b><br/>Het project eindigt met een werkend algoritme in productie. Onderdeel van het project is dat de</p> |

|                        |         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                        |         | business owner opgeleid zal worden om het onderhoud en doorontwikkeling aan te sturen.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| ICT                    | 5.1.2.e | <ul style="list-style-type: none"> <li>-Nieuwe infrastructuur voor de GeoMatch software</li> <li>-Test/acceptatie server met Docker met mogelijkheid om containers vanuit IPL te ontvangen en te installeren</li> <li>-Productieserver met Docker</li> <li>-Aanpassingen in IBIS</li> <li>-Knop GeoMatch en Informatie in scherm Screening &amp; matching</li> <li>-Gegevens kunnen verzenden en inlezen naar en vanuit GeoMatch</li> <li>-Gegevens (regioadvies) van GeoMatch vastleggen in IBIS database en tonen op het scherm</li> <li>-Advies GeoMatch overnemen of niet met vermelden van reden</li> <li>-Regioadvies GeoMatch opnemen in VGH dump</li> <li>-Gekozen regioadvies uit TVS opnemen in IBIS database</li> <li>-Aanpassingen in TVS</li> <li>-Regioadvies GeoMatch uit VGH dump verwerken en tonen in TVS</li> <li>-Gekozen arbeidsregio vastleggen en opnemen in export naar IBIS</li> <li>-Data uitwisseling met IPL via ESB (oplossing om periodiek, geautomatiseerd gegevens naar IPL te overzenden vanuit IBIS)</li> <li>-Data uitwisseling met CBS via ESB (proces inrichten om incidenteel, 2-4 keer per jaar, handmatig specifieke gegevens uit IBIS naar CBS portaal te uploaden)</li> <li>-Documentatie benodigd voor functioneel beheer van de GeoMatch servers</li> </ul> |
| Contracten en risico's | 5.1.2.e | <ul style="list-style-type: none"> <li>-De Data Protection Impact Assessment (DPIA),</li> <li>-De verwerkingsovereenkomst tussen COA en de hostende partij,</li> <li>-De verwerkingsovereenkomst (bij een verwerker relatie) of gegevensleveringsovereenkomst (bij gezamenlijke verantwoordelijkheid) tussen COA &amp; CBS, danwel andere partijen waarmee persoonsgegevens gedeeld worden,</li> <li>-De Artificial Intelligence Impact Assessment</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

|                   |                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Communicatie & PR | <div>5.1.2.e</div> <div>5.1.2.e</div> <div>5.1.2.e &amp; 5.1.2.e</div> <div>5.1.2.e</div> <div>5.1.2.e</div> | <p><b>Communicatieplan</b></p> <p>In het communicatieplan wordt beschreven welke doelgroepen op welke manier worden geïnformeerd over het project.</p> <p><b>Evaluatie:</b> Er wordt door middel van een vragenlijst onder de eindgebruikers gemeten hoe de adoptie en het gebruik van (algoritmische) data ervaren wordt als beslissingsondersteunend instrument.</p> |
|-------------------|--------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

De werkgroepen voeren zelfstandig taken uit en rapporteren structureel via een wekelijks projectgroepoverleg aan de 5.1.2.e over de voortgang van het project.

### 4.3 Technieken

#### Data-analyse

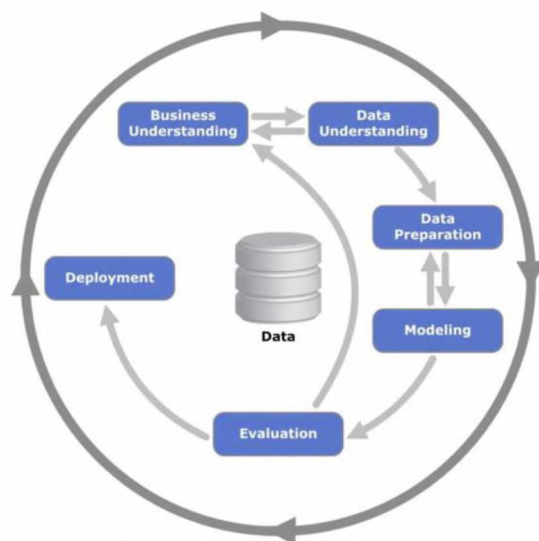
Geavanceerde data-analysetechnieken zullen worden gebruikt om de data te verzamelen en te analyseren, waaronder de vaardigheden en voorkeuren van statushouders, en de behoeften en kansen van arbeidsmarktregio's.

#### Machine Learning

Machine learning-algoritmen zullen worden geïmplementeerd om patronen en trends in de data te identificeren en om voorspellende modellen te ontwikkelen die de koppeling van statushouders aan arbeidsmarktregio's optimaliseren.

#### Continue Evaluatie

Het project zal een cyclus van continue evaluatie en verbetering volgen, waarbij het algoritme regelmatig wordt getest, geëvalueerd en verfijnd om de nauwkeurigheid, relevantie en impact te maximaliseren. We hanteren hierbij de CRISP-DM methodiek.



#### 4.4 Planning

Dit hoofdstuk biedt een gedetailleerde tijdsplanning voor het "Kansrijke Koppeling met behulp van AI" project, inclusief de belangrijkste mijlpalen, activiteiten, en deadlines. De tijdsplanning is ontworpen om het projectteam, de externe leverancier en alle stakeholders een duidelijk overzicht te geven van de projectfasen en -tijdslijnen.

##### Pre-pilot

###### Fase 1: Voorbereiding

- Voorbereiding – 1 september 2023 t/m 31 december 2023
- Start projectgroepen
- PID vaststelling in stuurgroep
- Contracten opstellen voor vervolg
- Gebruikerswensen ophalen
- Persona's vaststellen
- Inrichten server Solvinity

###### Fase 2: Ontwikkeling – 1 januari t/m 1 februari 2024

- Bepaling van proceswijziging
- Gebruikerstrainingen maken
- DPIA en AIIA afronden
- Wijzigingen releasen binnen ontwikkel en testomgeving van IBIS en TVS
- API maken

###### Fase 3: Testen – 1 februari t/m 1 maart 2024

- Uitvoeren van uitgebreide tests en validatieprocessen.
- Verzamelen van feedback en maken van noodzakelijke aanpassingen

### **Pilotfase**

Fase 4: Implementatie pilot locaties - 1 maart t/m 1 mei 2024

- Implementatie van het algoritme in de operationele omgeving.
- Training en capaciteitsopbouw voor het COA-personeel en relevante stakeholders.
- Officiële lancering van het "Kansrijke Koppeling" algoritme.

Fase 5: Evaluatie en Afsluiting - 1 mei t/m 1 juni 2024

- Verzamelen van feedback en documenteren van lessen en inzichten.
- Afsluiting van het project en overdracht naar de operationele fase.
- Opstellen van het eindrapport en delen met alle stakeholders.

### **Mijlpalen**

- **Mijlpaal 1:** Datum procestest (processen op orde) - 1 januari 2024
- **Mijlpaal 2:** Datum technische acceptatie en een functionele acceptietesten – nnb
- **Mijlpaal 3:** Definitieve go vanuit DT - 27 mei 2024
- **Mijlpaal 4:** Datum livegang – 1 juni 2024
- **Mijlpaal 5:** Uitbreiding livegang naar alle locaties – nnb

### **Conclusie**

Deze tijdsplanning is een leidraad voor de uitvoering van het "Kansrijke Koppeling" project. Het projectteam en de externe leverancier zullen nauw samenwerken om de gestelde deadlines te halen, terwijl ze flexibel blijven om aanpassingen te maken op basis van real-time feedback en veranderende omstandigheden.

## **5. Risico's**

Risicomanagement biedt een uitgebreide analyse van de potentiële risico's die het "Kansrijke Koppeling met behulp van AI" project kunnen beïnvloeden. Het identificeert specifieke risico's, evalueert hun potentiële impact en stelt mitigatiestrategieën voor om de risico's te beheersen en te minimaliseren.

### **Risico-evaluatie en -monitoring**

Elk geïdentificeerd risico zal worden geëvalueerd op basis van de waarschijnlijkheid van optreden en de potentiële impact op het project. Een risicoregister zal worden ontwikkeld en regelmatig worden bijgewerkt om de status en de behandeling van elk risico te volgen. De volgende tabel geeft een overzicht van de tot nu toe onderkende bedreigingen ten aanzien van het project met voorgestelde tegenmaatregelen, de kans van optreden en de mate van negatief effect op het project (aangegeven op een schaal

van 1 tot 5)<sup>3</sup>. De laatste kolom geeft het risico aan (kans\*impact). Op deze wijze kan gefundeerd worden afgewogen welke bedreigingen de meeste aandacht of hulpbronnen verdienen.

Toevoeging van bedreigingen of andere wijzigingen worden vermeld in het eerstvolgende Highlight Report.

| Onderwerp               | Bedreiging                                                                                                                                              | Tegenmaatregel                                                                                                                                                                                                                                                                                         | Eigenaar risico                          | Kans<br>(impact<br>x kans) |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|----------------------------|
| Draagvlak<br>en Cultuur | <b>Weerstand tegen Verandering:</b><br>Mogelijke weerstand binnen de organisatie en onder stakeholders tegen de implementatie van een nieuwe algoritme. | <b>Communicatie:</b> Kies voor een communicatiestrategie waar de nadruk op de bruikbaarheid en ondersteuningsmogelijkheid ligt.<br><br><b>Training en Ondersteuning:</b> Bied training en ondersteuning aan om de overgang naar het nieuwe systeem te vergemakkelijken en competenties te ontwikkelen. | 5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e | (3x3)<br>9                 |
| Draagvlak<br>en Cultuur | <b>Onvoldoende Betrokkenheid:</b><br>Gebrek aan actieve participatie en betrokkenheid van sleutelpersonen en -groepen.                                  | <b>Draagvlak</b>                                                                                                                                                                                                                                                                                       | 5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e | (2x4)<br>8                 |
| Externe leverancier     | <b>Kwaliteitsproblemen:</b><br>Het geleverde algoritme voldoet niet aan de verwachte kwaliteitsstandaarden of specificaties.                            | <b>Regelmatische Updates:</b><br>Implementeer een systeem voor regelmatische updates en voortgangsrapportages om de prestaties van de leverancier te monitoren.                                                                                                                                        | 5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e | (2x5)<br>10                |

<sup>3</sup> 1 = kleine kans of gering effect, 5 = grote kans of groot effect

|                          |                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                               |                                                             |             |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|-------------|
| Externe leverancier      | <b>Vertragingen:</b><br>Vertragingen in de ontwikkeling, het testen of de levering van het algoritme.                                                                                                                                                 | <b>Contractuele Afspraken:</b><br>Zorg voor duidelijke contractuele afspraken met betrekking tot kwaliteit, tijdlijnen en verantwoordelijkheden.                                                                                                              | 5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e                    | (4x2)<br>8  |
| Externe leverancier      | <b>Relatie opdrachtgever en leverancier.</b><br>Het COA gaat een samenwerking aan met een organisatie die een ander niveau van professionaliteit heeft.                                                                                               | <b>Contractuele Afspraken:</b><br>Zorg voor duidelijke contractuele afspraken met betrekking tot kwaliteit, tijdlijnen en verantwoordelijkheden.                                                                                                              | 5.1.2.e<br>5.1.2.e<br>5.1.2.e                               | (3x4)<br>12 |
| Veranderende organisatie | <b>Integratieproblemen</b><br>Uitdagingen bij het integreren van het nieuwe algoritme binnen de bestaande ambtelijke structuren en systemen.                                                                                                          | <b>Voorbereiding:</b> Zorg voor een gedetailleerd integratieplan om de overgang soepel en efficiënt te laten verlopen.<br><br><b>Ondersteuning:</b> Bied aanvullende ondersteuning en middelen aan om eventuele operationele verstoringen te minimaliseren.   | 5.1.2.e<br>5.1.2.e                                          | (5x2)<br>10 |
| Datakwaliteit            | <b>Onnauwkeurige Data:</b><br>Het risico van onnauwkeurige, onvolledige of verouderde data die de effectiviteit van het algoritme kunnen beïnvloeden. Inconsistentie in dataformaten en -structuren kunnen leiden tot integratie- en analyseproblemen | <b>Data Auditing:</b><br>Regelmatige audits om de nauwkeurigheid en volledigheid van de data te waarborgen.<br><br><b>Standaardisatie:</b><br>Implementatie van standaarden en protocollen voor dataverzameling en -verwerking om consistentie te waarborgen. | 5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e<br>5.1.2.e 5.1.2.e | (5x3)<br>15 |

|                 |                                                                                                                       |                                                                                                                                                       |                    |            |
|-----------------|-----------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|------------|
|                 |                                                                                                                       | <b>Data Auditing:</b><br>Regelmatige audits om de nauwkeurigheid en volledigheid van de data te waarborgen.                                           |                    |            |
| Data governance | <b>Privacy en Beveiliging</b><br>Risico's gerelateerd aan de privacy en beveiliging van gevoelige data.               | <b>Beveiligingsprotocollen:</b><br>Implementatie van strikte beveiligingsprotocollen om data te beschermen tegen ongeautoriseerde toegang en gebruik. | 5.1.2.e en 5.1.2.e | (1x5)<br>5 |
| Data governance | <b>Compliance:</b><br>Uitdagingen om te voldoen aan wettelijke en regelgevende vereisten voor databeheer en -gebruik. | <b>Compliance Framework:</b><br>Ontwikkeling van een compliance framework om te zorgen voor naleving van alle relevante wetten en regelgeving.        | 5.1.2.e en 5.1.2.e | (1x5)<br>5 |

### Conclusie

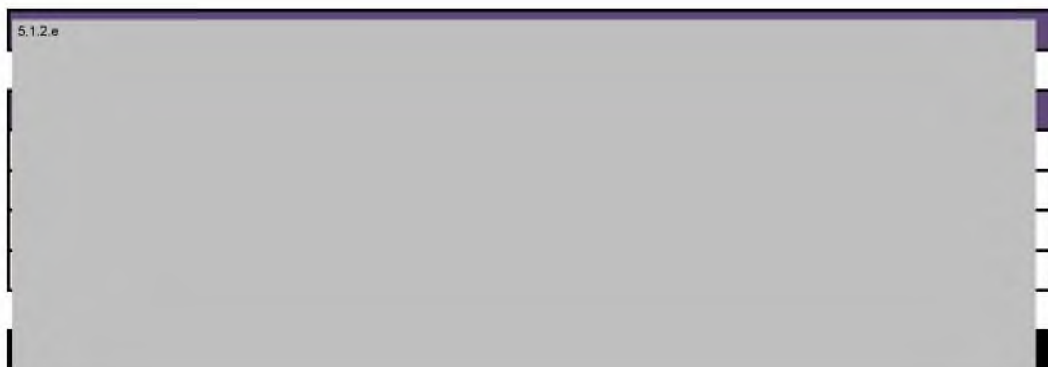
Risicomanagement is een cruciaal aspect van het project. Het identificeren, evalueren en beheren van risico's, inclusief die gerelateerd aan datakwaliteit en data governance, is essentieel om de integriteit, betrouwbaarheid en effectiviteit van het algoritme te waarborgen. Een proactieve en systematische benadering van risicomanagement zal bijdragen aan het succes en de duurzaamheid van het project.

## 6. Budget, financieringsbronnen en financieel management

In dit hoofdstuk wordt het financiële plan voor het "Kansrijke Koppeling" project uiteengezet.

### 6.1 Aangevraagde financiering

Voor de onderzoeksfase is de volgende financiering aangevraagd en gehonoreerd: € 5.110 Dit bedrag is opgesplitst in:



Medio 2023 is het volgende financiële besluit goedgekeurd door de directie van het COA.

- Het aanvullend beschikbaar stellen van € 5.1.1.c voor 2023 voor het voortzetten van het project Kansrijke Koppeling met behulp van AI (*korte termijn tot april 2024*) én
- Het jaarlijks beschikbaar stellen, voor ten minste drie jaar, van € 5.1.1.c (in totaal € 5.1.1.c over drie jaar) voor het project Kansrijke koppeling m.b.v. AI. In te zetten als de prepilot fase positief wordt afgerond (*middellange termijn*)

## 6.2 Financieringsbronnen

Er is medio 2022 een AMIF subsidie aangevraagd en deze aanvraag is afgewezen. De financiering is goedgekeurd door de COA-directie en onderdeel van de lopende P&L van het COA.

## 6.3 Financieel Managementstrategie

Budgetbewaking en kostenbeheersing wordt uitgevoerd door

- **Proces:** Continue bewaking van de uitgaven om te zorgen dat ze binnen het vastgestelde budget blijven.
- **Tools:** Gebruik van SAP om de kosten op het project te bewaken.

Kostenbeheersing

- **Proces:** Implementatie van kostenbeheersingsmaatregelen om onnodige uitgaven te voorkomen en te beperken.
- **Initiatieven:** Periodieke financiële reviews en aanpassingen van het budget op basis van de projectvoortgang en -behoeften.

## 7. Projectorganisatie en rollen

| Naam            | Rol                                                 |
|-----------------|-----------------------------------------------------|
| Joeri Kapteijns | Portefeuillehouder thema VrIP / kansrijke koppeling |

| Stuurgroep business |                                     |
|---------------------|-------------------------------------|
| Naam                | Rol                                 |
| 5.1.2.e             | 5.1.2.e 5.1.2.e ( 5.1.2.e )         |
| 5.1.2.e             | 5.1.2.e 5.1.2.e ( 5.1.2.e 5.1.2.e ) |
| 5.1.2.e             | 5.1.2.e 5.1.2.e ( 5.1.2.e 5.1.2.e ) |
| 5.1.2.e             | 5.1.2.e 5.1.2.e ( 5.1.2.e )         |

De stuurgroep is eindverantwoordelijk voor het overall eindresultaat van het project en voor het ter beschikking stellen van mensen en middelen. De stuurgroep komt formeel alleen bij elkaar voor het bespreken van faseplanningen en uitzonderings-rapportages.

*Project borging* vergewist zich ervan dat het project conform de afspraken wordt beheerst en uitgevoerd.

| Projectteam     |                 |
|-----------------|-----------------|
| Naam            | Rol             |
| 5.1.2.e         | 5.1.2.e         |
| 5.1.2.e         | 5.1.2.e         |
| 5.1.2.e 5.1.2.e | 5.1.2.e 5.1.2.e |
| 5.1.2.e         | 5.1.2.e         |
| 5.1.2.e 5.1.2.e | 5.1.2.e 5.1.2.e |
| 5.1.2.e         | 5.1.2.e         |

De 5.1.2.e krijgt van de *Stuurgroep* de verantwoordelijkheid en bevoegdheid voor de dagelijkse gang van zaken in het project. De 5.1.2.e kan beslissingen nemen binnen de grenzen zoals aangegeven door de Stuurgroep (tolerantie). De belangrijkste verantwoordelijkheid van de 5.1.2.e is het zekerstellen dat het project de gewenste producten oplevert, binnen tijd, binnen budget en volgens de afgesproken kwaliteit. De 5.1.2.e rapporteert aan de *Stuurgroep*.

5.1.2.e ondersteunt de 5.1.2.e technisch en administratief. Het projectarchief wordt beheerd door de 5.1.2.e eventueel gedelegeerd aan 5.1.2.e

De 5.1.2.e is verantwoordelijke contactpersoon voor de gebruikers tijdens de pilot-uitvoering.

De 5.1.2.e is o.a. verantwoordelijk voor het opstellen van het communicatieplan. Overige activiteiten worden separaat afgestemd met de 5.1.2.e (hier in het PID vastleggen). Ditzelfde geldt voor de nader te bepalen functies. FA is verantwoordelijk voor de financiële administratie.

| Ondersteunende diensten                   |                                                                |
|-------------------------------------------|----------------------------------------------------------------|
| Nm                                        | Rol                                                            |
| <span>5.1.2.e</span>                      | <span>5.1.2.e</span> <span>5.1.2.e</span> <span>5.1.2.e</span> |
| <span>5.1.2.e</span> <span>5.1.2.e</span> | <span>5.1.2.e</span> <span>5.1.2.e</span>                      |
| <span>5.1.2.e</span>                      | <span>5.1.2.e</span> <span>5.1.2.e</span>                      |

De externe leverancier, IPL, zal verantwoordelijk zijn voor de technische ontwikkeling, het testen en de levering van het algoritme. Met hun uitgebreide ervaring en technische capaciteiten zullen zij een cruciale rol spelen in het waarborgen van de effectiviteit en efficiëntie van de oplossing.

Tevens is er binnen het COA een ICT project opgestart waarin verschillende aanpassingen aan de technische infrastructuur van het COA worden gerealiseerd, waaronder aan IBIS en TVS. In het [Scrum Initiation Document](#) staat beschreven om welke wijzigingen het gaat, de middelen en planning. 5.1.2.e 5.1.2.e 5.1.2.e rapporteert aan het projectteam over de voortgang van dit deelproject.

## 8. Borging

Borgen betekent dat je er voor zorgt dat de producten en resultaten die voortkomen uit het project op een vanzelfsprekende manier worden geïntegreerd in de staande organisatie.

Na een periode met een beperkt aantal pilotlocties te hebben getest wordt er een evaluatierapport gemaakt met daar in beschreven:

- Ervaringen van gebruikers n.a.v. een vragenlijst
- Een rapport van een externe auditor die
- Opleiding van beheerorganisatie (nader te bepalen binnen COA)
- Advies van de 512e om al dan niet de applicatie in gebruik te nemen voor de rest van de organisatie na de pilotfase en op welke manier dit gerealiseerd kan worden.

Besluitvorming vindt plaats door de stuurgroep, DT en bestuur.

De 512e is verantwoordelijk voor het opstellen van de rapportages en wordt hierbij ondersteund door de 512e en 512e 512e 512e

Bovenstaande kan leiden tot het opstellen van een Herziene aanvraag met alternatieven voor de aanpak van het vervolg van het project en aangepaste planningen.

# Werkinstructie GeoMatch pilot - Vroege screening voor vergunningverlening op de pol

Versienummer: 0.9

Laatste mutatie: 29-04-24

Doelgroep FAQ: intern gebruik COA

## 1. Leeswijzer

Deze werkinstructie (WI) is bedoeld voor het proces [Vroege screening voor vergunningverlening op de POL en heeft](#) als doel dat de pilot-locaties aan de hand van deze beschrijving gebruik kunnen maken van GeoMatch.

De MAVIM-versie en de uiteindelijke WI kunnen qua vorm afwijken.

Deze werkinstructie is van toepassing op het GeoMatch algoritme versie 1.0 maart 2024 en Ibis versie 6.96 en TVS versie 4.52

## 2. Het Vroege Screeningsgesprek en GeoMatch

### 2.1 GeoMatch

GeoMatch is een tool in Ibis die ons ondersteunt in het proces van screenen en matchen bij het vinden van de beste arbeidsmarktregio om werk te vinden. GeoMatch werkt met een algoritme, wat een reeks stap-voor-stap instructies is om een bepaald doel te bereiken. In dit geval helpt het algoritme om te bepalen welke arbeidsmarktregio's het meest geschikt zijn voor de bewoner om werk te vinden op basis van zijn of haar kenmerken.

#### Hoe werkt GeoMatch?

Hoe doet GeoMatch dit? Het Centraal Bureau voor de statistiek (CBS) heeft de afgelopen jaren informatie verzameld over kansrijke asielzoekers en vergunninghouders die eerder naar Nederland zijn gekomen en in asielzoekerscentra hebben gewoond. Deze informatie bevat gegevens zoals waar ze eerder hebben gewerkt, wat voor opleiding ze hebben gedaan in land van herkomst en wat hun persoonlijke kenmerken zijn zoals leeftijd en moedertaal. GeoMatch legt patronen en toont hoe succesvol deze mensen waren in het vinden van werk in verschillende delen van Nederland. Met al deze informatie kan GeoMatch voorspellen hoe waarschijnlijk het is dat iemand werk zal vinden in de verschillende regio's, gebaseerd op hun unieke achtergrond en de (historische) werkgelegenheidstrends in die regio's. Tevens houdt GeoMatch rekening met de taakstelling van de regio.. Het resultaat is een advies van GeoMatch die de regio's met de hoogste werkgelegenheidskansen laat zien voor de plaatsingseenheid (PLE) waarbij er ook rekening gehouden wordt met toekomstige koppelingen in de regio. Voor meer informatie over het algoritme van GeoMatch kun je bijlage 1 raadplegen.

#### Wanneer wordt het gebruikt?

GeoMatch zal worden toegepast op personen die in aanmerking komen voor het werkproces [vroege screening voor vergunningverlening op de POL](#) en wanneer er geen sprake is van zogenaamde harde criteria conform de procedure.

### **Voor welke doelgroep?**

Voor de gebruiker is het belangrijk te weten dat het algoritme ontwikkeld is voor de doelgroep die “geschikt is om te werken”. Dit betekent dat het algoritme niet ingezet kan worden voor minderjarigen en bewoners die arbeidsongeschikt of pensioengerechtigd zijn.. Het doel van GeoMatch is om te identificeren waar de meeste werkgelegenheidskansen zijn voor bewoners, daarom kun je geen advies genereren voor mensen die niet geschikt zijn om te werken. Het is niet de taak van de <sup>5.1.2e</sup> om actief buiten deze criteria te beoordelen of de betrokkene “arbeidsgeschikt” is. Deze taak ligt bij de begeleiding van de gemeente bij uitstroom naar de gemeente.

### **Wanneer wordt het advies getoond?**

Het advies wordt getoond, nadat je de knop GeoMatch in Ibis hebt aangeklikt.

Let op: het kan zijn dat de knop in Ibis niet actief is, dit kun je zien doordat de knop grijs is. Dit betekent niet dat er een storing is. Tijdens de pilot vindt er een wetenschappelijk onderzoek plaats naar de effecten van het GeoMatch-advies. Op deze manier kan het succes van GeoMatch beter gemeten worden. Willekeurig (en dit is gerandomiseerd) is het niet mogelijk om een advies op te vragen. De knop in Ibis is dan grijs.

### **Harde en zachte criteria**

De kansrijke koppeling gaat over de aanwezigheid van zachte criteria. Wanneer er geen sprake is van harde criteria komt de optie beschikbaar in Ibis voor het opvragen van een regioadvies.

De zachte criteria zijn:

- Opleidingsachtergrond
- Werkervaring
- Ambities
- De aanwezigheid van een sociaal netwerk<sup>1</sup>

Wanneer er sprake is van een hard criterium zal GeoMatch niet te raadplegen zijn.

De harde criteria zijn:

- Betaald werk
- Familie eerste graad regulier gehuisvest
- Sociaal medisch advies
- Studie

Na het afronden van het screeningsgesprek beoordeel je of er gebruik gemaakt wordt van het algoritme. Het doel van GeoMatch is om kansrijke asielzoekers te voorzien van een advies voor een arbeidsmarktregio, gebaseerd op zorgvuldige analyse van informatie van zowel het CBS en UWV als de zachte criteria en kenmerken van bewoners. Dit advies toont de top 3 arbeidsmarktregio's waar de PLE de meeste werkgelegenheidskansen heeft. Hierbij houdt GeoMatch rekening met alle toekomstige koppelingen in de regio.

Het regioadvies is zowel gericht aan de afdeling Plaatsing In- en Doorstroom, als aan de regievoerder om zoveel mogelijk maatwerk te leveren bij zowel de doorplaatsing naar het azc/arbeidsmarktregio

---

<sup>1</sup> Zie voor meer informatie de werkinstructie “[Vroege screening voor vergunningverlening op de POL](#)”

als de koppeling aan de gemeente. De bewoner wordt doorgeplaatst naar een regio en binnen deze regio door de regievoerder gekoppeld aan een gemeente waar de kansen op (arbeids)integratie het grootst zijn.. Registreer alle zachte criteria voordat je het advies opvraagt via GeoMatch. Het algoritme functioneert beter wanneer de zachte criteria volledig zijn geregistreerd.

### 3.1.1 Stappenplan opvragen regioadvies via GeoMatch

Aangezien de bewoners die in aanmerking komen voor Vroege Screening nog geen vergunninghouder zijn, komen zij ook nog niet voor in de vergunninghouders module in Ibis Groepshandelingen. De invulvelden voor Screening en Matching kunnen voor deze bewoners op de volgende manier worden bereikt:

Ibis > Bewoner > Opvang > Groepshistorie. Selecteer hier 'PLE'. Voor bewoners die binnen de doelgroep van de Vroege Screening vallen is nu het tabblad Screening & Matching beschikbaar. Bij bewoners buiten deze doelgroep is dit tabblad niet beschikbaar (de knop in Ibis is uitgedijdsd).

Allereerst rond je de volgende stappen af tijdens de registratie in Ibis. Onderstaande onderdelen zijn te vinden in de [Vroege screening voor vergunningverlening op de POL](#).

- "Registratie zachte criteria in het scherm Screening en Matching voor vergunningverlening volgens de instructies in hoofdstuk 2.6."
- "Registratie harde criteria via Groepshistorie volgens de instructies in hoofdstuk 2.6."
- Doorloop alle stappen in hoofdstuk 2.6 tot aan het onderdeel "Het regioadvies".

Na het afronden van de bovenstaande stappen, ga je als volgt te werk:

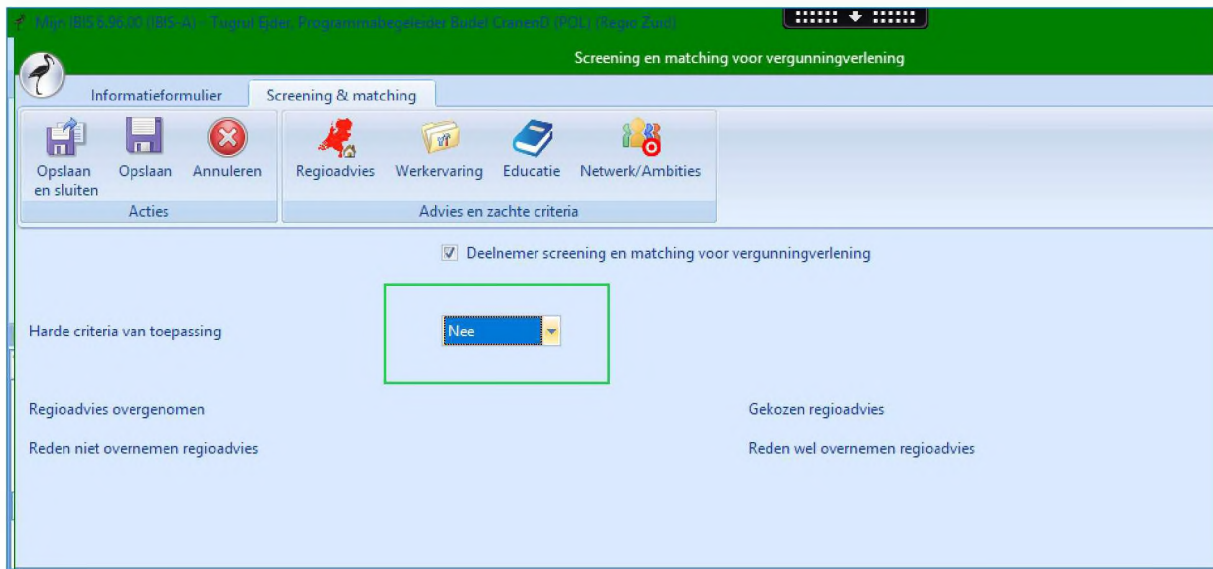
#### Harde criteria van toepassing?

The screenshot shows the 'Screening & matching' tab in the Ibis application. The 'Harde criteria van toepassing' field is highlighted in yellow with a 'Ja' button. The 'Arbeidsregio's' table is also highlighted in yellow.

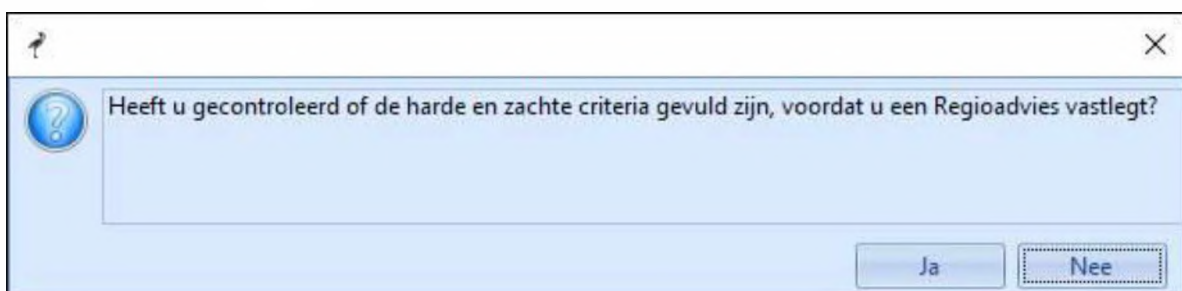
| Uitplaatsingsregio | Arbeidsregio |
|--------------------|--------------|
| Blanco             |              |
| Regio 1 FR/GR/DR   | Drenthe      |
| Regio 1 FR/GR/DR   | Friesland    |
| Regio 1 FR/GR/DR   | Groningen    |

Indien er sprake is van harde criteria, selecteer je “Ja” in het scherm hierboven. De keuzemogelijkheid voor het opvragen van een regioadvies via GeoMatch vervalt en de ‘Advies’-knop is niet actief. GeoMatch zal voor deze PLE niet te raadplegen zijn. Indien het antwoord “Nee” is ga je door naar de volgende stap. Zie hiervoor het scherm hieronder.

Als er geen harde criteria van toepassing zijn, is GeoMatch te raadplegen. In het tabblad “Screening & matching” selecteer je ‘Nee’ in het veld ‘Harde criteria van toepassing’. Klik vervolgens op de knop “Regioadvies” zoals hieronder afgebeeld.



Een reminder pop-up verschijnt met de vraag of je de harde en zachte criteria ingevuld hebt. Ben je wat vergeten, druk dan op ‘Nee’ en registreer dit alsnog voordat je weer terugkomt bij deze stap. Heb je alle gegevens gecontroleerd, selecteer dan “ja”. Deze pop-up is bedoeld om je eraan te herinneren dat het algoritme optimaal functioneert wanneer alle harde en zachter criteria zijn geregistreerd voor de PLE.



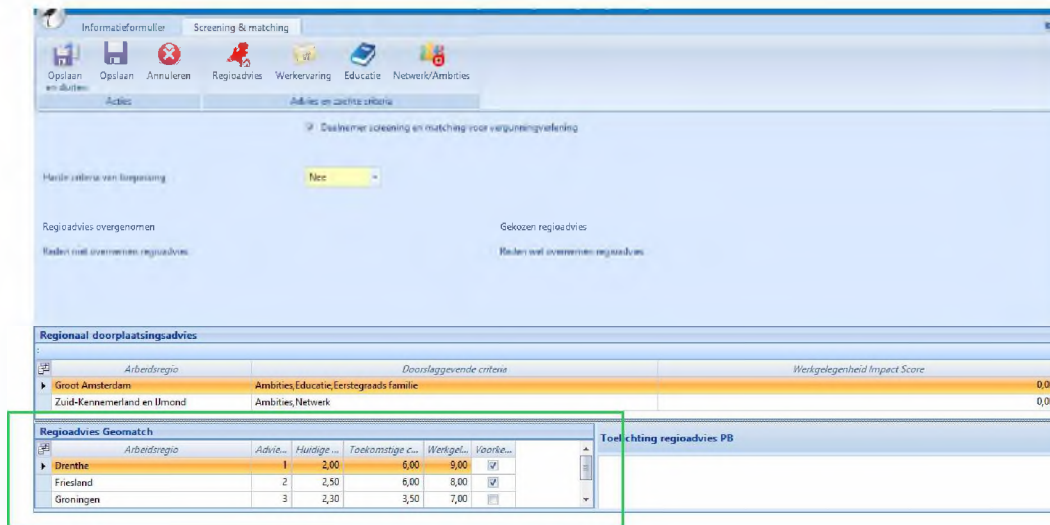
In het volgende scherm kun je de keuze maken voor de arbeidsregio's.

| Uitplaatsingsregio                                   | Arbeidsregio                       | Doorslaggevende criteria |
|------------------------------------------------------|------------------------------------|--------------------------|
| <input type="checkbox"/> Blanco                      |                                    |                          |
| <input type="checkbox"/> Regio 1 FR/GR/DR            | Drenthe                            |                          |
| <input checked="" type="checkbox"/> Regio 1 FR/GR/DR | Friesland                          | Ambities                 |
| <input type="checkbox"/> Regio 1 FR/GR/DR            | Groningen                          |                          |
| <input checked="" type="checkbox"/> Regio 2 OV/GLD   | Achterhoek                         | Netwerk                  |
| <input type="checkbox"/> Regio 2 OV/GLD              | Food Valley                        |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Midden-Gelderland                  |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Regio Zwolle                       |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Rijk van Nijmegen                  |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Rivierenland                       |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Stedendriehoek en Noordwest Veluwe |                          |
| <input type="checkbox"/> Regio 2 OV/GLD              | Twente                             |                          |
| <input type="checkbox"/> Regio 3 NH/UTR/FL           | Amersfoort                         |                          |

In het tabblad "Arbeidsregio's" selecteer je een eventueel regioadvies op basis van harde en zachte criteria. Dit doe je volgens de instructies in hoofdstuk 2.6 het onderdeel "Het Regioadvies".

Dit is het regioadvies wat je zelf indient op basis van het gesprek. Nadat je de invoervelden hebt ingevuld, selecteer je "GeoMatch" in het tabblad ernaast, zoals afgebeeld hierboven.

In het tabblad 'GeoMatch' kun je met een druk op de knop 'Advies' een regioadvies ontvangen vanuit GeoMatch. Na enkele seconden zal dit scherm sluiten en de adviezen worden getoond in het volgende scherm.



De adviezen van de tool GeoMatch zullen zoals afgebeeld in bovenstaand scherm worden getoond in de tabel onder het kopje "Regioadvies GeoMatch".

Het algoritme van GeoMatch zal een top 3 geven van arbeidsmarktregio's in het hele land voor zowel de doorplaatsing naar het azc als de koppeling aan een gemeente.

### Wat is de Werkgelegenheid Impact Score (WIS)?

. De Werkgelegenheid Impact Score, of WIS, geeft aan hoe geschikt een bepaalde arbeidsmarktregio is voor iemand om werk te vinden, waarbij de score een balans is van de huidige PLE en de toekomstige PLE. Deze score is bedoeld om advies te geven over waar de kansrijke asielzoekers het beste kunnen gaan wonen en werken, op basis van hun kansen op het vinden van een baan in die regio.

### Hoe werkt het?

1. **Evaluatie van arbeidsmarktregio's:** Alle arbeidsmarktregio's worden beoordeeld om te zien waar de kansen op werk het best zijn voor mensen die op zoek zijn naar een baan. Er wordt gekeken naar de huidige werkgelegenheidssituatie en de toekomstige verwachtingen.
2. **Toekenning van scores:** Elke arbeidsmarktregio krijgt een score van 0 tot 10. Een score van 10 betekent dat de regio uitstekend is - de best mogelijke plek om werk te vinden. Een score van 0 geeft aan dat het volgens de berekeningen de minst ideale arbeidsmarktregio is. In bovenstaand voorbeeld komt Drenthe als meest geschikte regio naar voren met de waardering 9.
3. **Vergelijking van arbeidsmarktregio's:** De arbeidsmarktregio met de beste kans op werk krijgt de hoogste WIS score. Andere regio's krijgen een lagere score, afhankelijk van hoeveel minder kans op werk ze bieden vergeleken met de best scorende regio.

### Wat betekent dit voor kansrijke asielzoekers?

Door gebruik te maken van de WIS worden kansrijke asielzoekers naar de arbeidsmarktregio's geleid waar ze de beste kans hebben om werk te vinden.

## Adviesrang

Op basis van de WIS wordt een adviesrang opgesteld voor de drie meest kansrijke regio's. Adviesrang 1 heeft de hoogste WIS, Rang 2 iets lager enzovoorts.

Voor het adviseren volgens de WIS, is het bovenstaande voldoende om tot een eindadvies te komen. Wil je meer uitleg over hoe de WIS wordt berekend en wat de kolommen 'huidige casus' en 'toekomstige casus' betekenen, dan kun je deze informatie raadplegen via bijlage 1.

| Adviesrang | Huidige casus | Toekomstige casus | WIS  |
|------------|---------------|-------------------|------|
| 1          | 2,00          | 6,00              | 8,00 |
| 2          | 2,50          | 6,00              | 8,00 |
| 3          | 2,30          | 1,50              | 7,00 |

## Regioadvies van de 5.1.2.e

Op basis van het regioadvies GeoMatch en je eigen regioadvies kom je tot een eindadvies. Dit betekent dat je in het scherm hierboven de voorkeur voor het regioadvies kan opgeven. Door het aanvinken van de kolom 'voorkeur PB' geef je als 5.1.2.e de voorkeur aan. Dit doe je in de tabel van GeoMatch voor één of meer van de drie adviesregio's. Extra toelichting op je regioadvies is mogelijk in het open tekstvak 'Toelichting regioadvies PB'. Hier kan je nuance aanbrengen in je eigen regioadvies, de toelichting is dan zichtbaar zijn voor de regievoerder.

Zowel je eigen regioadvies als de GeoMatch adviezen en de toelichting worden verstuurd naar TVS en dit wordt na vergunningverlening zichtbaar voor de regievoerder. De regievoerder beslist uiteindelijk of het advies wordt overgenomen en komt dan met een koppeling met een gemeente.

## Terugkoppeling aan Plaatsing

Na het verkrijgen van bovenstaande adviezen, kan dit in zijn geheel doorgegeven worden aan Plaatsing In- en Doorstroom en kan het genoteerd worden in het logboek. Hierbij kun je te werk gaan volgens hoofdstuk 2.4 uit de bestaande [werkinstructie](#).

## Traject na vergunningverlening

Alle adviezen van de 5.1.2.e en de adviezen van GeoMatch komen via TVS in beeld bij de regievoerder. Het traject zal verlopen volgens hoofdstuk 2.5 uit de bestaande [werkinstructie](#)

## Het eindadvies door de 5.1.2.e

GeoMatch is een beslissing-ondersteund algoritme dat helpt bij het formuleren van een regioadvies. Als 5.1.2.e heb je een breder beeld dan alleen de arbeidsmarkt. Je weegt dit af samen met zachte criteria, zoals ambitie, netwerk en opleidingsachtergrond. GeoMatch geeft extra informatie over arbeidsmarktkansen. Het eindadvies is gebaseerd op het totale beeld dat je als 5.1.2.e hebt gekregen en het is daarom niet erg als er van het GeoMatch advies wordt afgeweken.

## **Bijlage 1: De WIS, toekomstige casus en huidige casus**

### **De WIS**

De Werkgelegenheid Impact Score (WIS) geeft het niveau van de verwachte werkgelegenheidswinsten weer bij het gebruik van GeoMatch. Hierbij worden de voordelen voor zowel huidige als toekomstige situaties afgewogen. De WIS wordt op een schaal van 0 tot 10 gerangschikt, waarbij 0 de minst ideale locatie volgens GeoMatch vertegenwoordigt en 10 de meest ideale locatie aangeeft. De WIS voor de beste locatie zal altijd 10 zijn, wat aangeeft dat dit volgens het algoritme de optimale aanbeveling is: een locatie die de werkgelegenheidsresultaten maximaliseert voor zowel de huidige als toekomstige situaties. De WIS neemt af naarmate de werkgelegenheidsvooruitzichten voor zowel de huidige als toekomstige situaties voor een locatie afnemen in vergelijking met de best gerangschikte locatie.

De Werkgelegenheid Impact Score (WIS) voor locatie X wordt als volgt berekend:

1. Het totaal van de geschatte kans voor de huidige casus om werk te vinden op locatie X wordt opgeteld bij het aantal banen dat volgens GeoMatch-schattingen door toekomstige casussen zou worden gevonden.
2. Een WIS van 10 wordt toegekend aan de locatie met de hoogste berekening. Dit wordt beschouwd als de best gerangschikte locatie.
3. De WIS van andere locaties wordt verlaagd met 1 punt voor elke 0,05 banen die naar verwachting "verloren" zullen gaan in vergelijking met de best gerangschikte locatie. Dit verwachte "banenverlies" kan bestaan uit voorspelde vermindering van kansen op werk voor de huidige casus en vermindering van kansen op werk voor toekomstige casussen. Elke WIS die op basis van deze methode onder 0 valt, wordt afgerond naar 0.

### **Huidige casus**

Het huidige casusniveau geeft de waarschijnlijkheid van werkgelegenheid aan voor de huidige casus (zonder rekening te houden met voordelen voor toekomstige gevallen) als deze regio wordt gekozen. Voor de locatie met de hoogste kans op werk voor de huidige casus zou dit niveau 10 zijn. Andere locatieniveaus worden met 1 punt verminderd voor elke vermindering van 5 procentpunten in de waarschijnlijkheid dat de huidige casus werk vindt. Als locatie A bijvoorbeeld een niveau 10 heeft en locatie C een niveau 6, betekent dit dat de kans op het vinden van werk voor de huidige casus op locatie C 20 procentpunten lager is dan op locatie A.

### **Toekomstige casus**

Het niveau van toekomstige casussen geeft de verwachte werkgelegenheidswinsten aan voor toekomstige casussen die in Nederland zullen arriveren, als deze locatie wordt gekozen voor de huidige casus, zonder rekening te houden met de voordelen voor de huidige casus.

De locatie die naar verwachting het grootste aantal banen voor toekomstige casussen zal genereren, krijgt niveau 10. Voor andere locaties wordt het niveau met 1 punt verlaagd voor elke afname van 0,05 in het verwachte aantal banen dat toekomstige casussen kunnen verkrijgen. Bijvoorbeeld, als locatie A een niveau van 10 heeft en locatie C een niveau van 6, betekent dit dat als de huidige casus

op locatie C wordt geplaatst, toekomstige casussen naar verwachting 0,2 banen minder zullen vinden dan als de huidige casus op locatie A wordt geplaatst.

Om het voorspelde aantal toekomstige werkende casussen te berekenen wanneer de huidige casus op deze locatie wordt geplaatst, moeten we eerst schatten welke casussen waarschijnlijk in de toekomst zullen arriveren. Hiervoor worden patronen in historische gegevens gebruikt om toekomstige casussen te simuleren. Vervolgens gebruikt het algoritme wiskundige optimalisatie om de meest voordelige plaatsing voor alle casussen te vinden (zowel de huidige als de verwachte toekomstige casussen). Het totale aantal verwachte banen wordt berekend op basis van de voorspelde kansen op werk. Om de onzekerheid in toekomstige casussen weer te geven, wordt dit proces meerdere malen herhaald, waarbij telkens een nieuwe steekproef wordt genomen uit de historische gegevens om toekomstige casussen te simuleren en het verwachte totale aantal verkregen banen te berekenen. Het voorspelde aantal toekomstige werkenden is het gemiddelde werkgelegenheidsniveau van de toekomstige casussen over deze simulaties.