

## UITKERINGSFRAUDE PROJECT ANALYTICS – Pilot Prestatie010

### Verantwoording



Versie    **1.0**

Opdrachtgever	Cluster <b>Werk en inkomen - afdeling Toetsing &amp; Toezicht</b>
Auteur	<b>XXXXXXXXXX</b>
Datum	<b>16-11-2020</b>
Status	<b>Definitief</b>



# 1 Inleiding

Het doel van het project Analytics Uitkeringsfraude is om data te gebruiken voor het gericht opsporen van onrechtmatigheden in de uitkeringsverstrekking. Dit is een wettelijke taak van de gemeente Rotterdam op basis van de Participatiewet (2015). Om dit doel te bereiken wordt data gebruikt die is verzameld op basis van diezelfde Participatiewet, namelijk de operationele data uit het 'inkomen'- en 'werk'-proces van de gemeente zoals vastgelegd in de systemen Socrates en RMW/RAAK.

Het doel van deze data gedreven werkwijze is om een risicomodel te ontwikkelen dat onrechtmatigheid bij PW-uitkeringen voorspelt ten behoeve van het handavingsproces. Het beoogde resultaat is beter en sneller inzicht verkrijgen in onrechtmatigheidsgevallen bij de bijstandsverstrekking, waardoor de efficiëntie omhoog gaat:

- enerzijds door vermindering van gemeentelijk geld dat onterecht wordt uitgekeerd;
- anderzijds door een effectievere inzet van de beschikbare capaciteit voor handhaving, waardoor er een hogere pakkans is voor fraudeurs en compliante burgers minder worden belast.

De nevendoelestelling is tevens de kennisverwerving en vaardigheidsontwikkeling van (innovatieve) data-analysetechnieken binnen de gemeente Rotterdam.

In de Participatiewet spreekt men van onrechtmatigheid als na onderzoek tot één van de volgende conclusies is gekomen: 1) er is een (terug) te betalen bedrag bepaald, 2) de lopende uitkering moet naar aanleiding van het onderzoek beëindigd worden, 3) de lopende uitkering moet naar de toekomst toe aangepast worden.

Dit document geeft op hoog niveau inzicht in input, output en methodieken. Achterliggende theorie wordt niet toegelicht.

# 2 Model

Een model is een algoritme dat wordt gebruikt om van meerdere datapunten per ID (persoon) tot een enkel getal te komen, dat iets zegt over de mate waarin aan-/afwezigheid van uitkeringsonrechtmatigheid te verwachten is. Het ontwikkelen ("trainen") van een model gebeurt door in data van historische rechtmatigheidsonderzoeken naar patronen te zoeken en te testen of deze patronen in ongeziene data (een apart gehouden deel van de historische rechtmatigheidsonderzoeken) de correcte uitkomst voorspellen.

Bij eerdere iteraties van het model (in 2018 en 2019) viel op dat personen uit de Prestatie010-groep zeer weinig voorkwamen in de top van de risicoscores, terwijl deze groep ongeveer de helft van de doelpopulatie vormt en men niet de ervaring heeft dat in deze groep zoveel minder onrechtmatigheid voorkomt. Het vermoeden is dat deze groep kenmerken heeft die door de eerdere modellen niet voldoende herkend worden; patronen uit de rest van de populatie lijken een sterker signaal te bevatten, dat het signaal voor de Prestatie010-groep overstemt. Daarom is besloten om een separate pilot uit te voeren, specifiek voor de Prestatie010-groep. Er is een apart model getraind op onderzoeken van werkzoekenden die op het moment van onderzoek in de Prestatie010-groep zaten. Hieruit kwam een **gradient boosting algoritme** als best presterende naar voren (gemeten naar **accuracy**). Met het model hebben vervolgens ook weer alleen werkzoekenden uit de Prestatie010-groep een risicoscore gekregen.

# 3 Brondata

Er is gewerkt met gegevens die zich op het peilmoment (**23 juli 2020**) in de systemen Socrates en RMW/Raak bevinden. Daarin zitten:

- 1) de historische rechtmatigheidsonderzoeken (met de uitkomsten daarvan) van personen die op de meldingsdatum ondergebracht waren bij de "werk"-afdelingen Prestatie010;
- 2) de actuele werkzoekenden met een PW-uitkering die op het peilmoment ondergebracht zijn bij de "werk"-afdelingen Prestatie010;
- 3) allerlei gegevens die vastgelegd zijn in de processen "werk" en "inkomen" die de gemeente met de werkzoekenden in het kader van de Participatiewet doorloopt.

De historische rechtmatigheidsonderzoeken (1) worden gebruikt om modellen te "trainen" en te "testen"; we gebruiken hiervan alleen de onderzoeken die vanaf **01-01-2014** afgerond zijn. De actuele werkzoekenden (2) zijn de mensen van wie we met het model een risicoscore gaan bepalen. Van de overige gegevens (3) worden "features" afgeleid: gegevens over de persoon en het proces die het model gebruikt om de risicoscore te bepalen. In de bijlage is in meer detail gespecificeerd welke Socrates- en RMW/Raak-tabellen precies worden gebruikt.

Om organisatorische redenen kon het team HerOnderzoeken pas na enkele weken met de scores aan de slag. Door tussentijdse mutaties in het werkzoekendenbestand was de lijst niet meer actueel. Daarom is het algoritme ververst met data van het peilmoment **29 september 2020**. De keuze voor gradient boosting is dus vastgesteld met data set 23 juli; de feitelijke scores zijn met de data set van 29 september bepaald.

## 4 Output

De output van het data-analyseproject is een lijst met alle gepseudonimiseerde PW-werkzoekenden en hun risicoscore. Dat is een getal tussen 0 en 1, dat in relatieve zin aangeeft hoe aannemelijk het is dat er uitkeringsonrechtmatigheid te verwachten is. De scores zijn niet absoluut, ze kunnen niet als "kans" worden geïnterpreteerd. De scorelijst wordt door een ander persoon, die geen toegang heeft tot de omgeving waarin het risicomodel ontwikkeld wordt, vertaald naar kenbare werkzoekenden en doorgestuurd naar de afdeling Toetsing & Toezicht. De afdeling T&T gebruikt die lijst om (mede) te bepalen welke werkzoekenden met bijstand daadwerkelijk voor een rechtmatigheidsonderzoek door het team HerOnderzoeken worden uitgenodigd.

### Doelgroep scope

Bepaalde persoonsprofielen worden bewust niet onderzocht, omdat dat onmogelijk of ongewenst is. Daarom wordt de uitkeringontvangende populatie, door de afdeling Toetsing & Toezicht, gefilterd aan de hand van business rules. De belangrijkste onderzoeksuitsluitingsregels worden hieronder weergegeven:

BUSINESS RULE
Er mag geen lopend rechtmatigheidsonderzoek door uBO, pHO en tHO zijn
De uitkering is niet opgeschort
De uitkering heeft het kenmerk Participatiewet
De uitkeringsontvanger is jonger dan 64 jaar
De uitkering loopt langer dan 6 maanden
De ontvanger is 2 jaar lang niet onderzocht door uBO, pHO en tHO

## 5 Privacy

### 5.1 Pseudonimiseren

De brondata uit Socrates en RMW/Raak is gepseudonimiseerd en versimpeld, zodat de data niet direct terug te herleiden is naar individuen. BSN-nummers zijn vervangen voor unieke gepseudonimiseerde ID's. De geboortedatums zijn beperkt tot jaar en maand. Van postcodes worden alleen cijfers meegenomen. Nationaliteit en geboorteland zijn niet beschikbaar in de data set.

Na de data-analyse moeten de pseudoniemen voor de personen waar het model een score heeft opgeleverd, worden terughergeleid naar de natuurlijke personen (ten behoeve van selectie en rechtmatigheidsonderzoek). De ontsleuteling wordt onder strenge voorwaarden uitgevoerd door geselecteerde personen, die niet bij de data-analyse en niet bij een vervolgonderzoek betrokken zijn.

### 5.2 Informatiebeveiliging

Alle data blijft binnen het netwerk van de gemeente en wordt verwerkt op gemeentelijke hardware met bijbehorende beveiligingsmaatregelen. Alleen direct bij het project betrokken personen krijgen voor het voor hen relevante onderdeel toegang tot de gegevens. Autorisaties voor de projectlocatie op het netwerk en eventueel later afgeschermd delen van databases worden door de opdrachtgever beheerd. Daarbij geldt dat alleen die personen autorisaties krijgen voor wie toegang noodzakelijk is voor de uitvoering van het project.

## Bijlage

In onderstaand overzicht staat uit welke Socrates en RMW/Raak tabellen de gebruikte gegevens afkomstig zijn. Om de gegevens te verkrijgen zijn altijd koppelingen tussen verschillende tabellen nodig; de belangrijkste zijn genoemd in kolommen 3 en 4. In alle gevallen wordt een koppeling gemaakt met de Socrates tabel PERSOON en de RMW/Raak tabel KVS\_KLANTEN. In de 5<sup>e</sup> en 6<sup>e</sup> kolom staan de belangrijkste velden, met voorbeelden van de inhoud ter illustratie. In kolom 2 staat voor welk doel de gegevens in het proces zijn gebruikt (zoals beschreven in paragraaf 3 Brondata). De voorbeelddata (rechterkolom) is niet uitputtend gevuld.

Data	Gebruik	Brontabel(len) Socrates	Brontabel(len) RMW/Raak	Belangrijkste velden	Voorbeeld data (niet uitputtend)
ADRES	Features	ALG_ADRES		DTAANVANG, DTEINDE	
				POSTCODENUM	3085, 3014, 1318
				BUURTNAAM, WIJKNAAM, PLAATS	"ZUIDWIJK", "CHARLOIS", "ROTTERDAM"
AFSPRAAK	Features		KVS_AFSpraakRESULTATEN, KVS_AFSpraken_JN, KVS_AFSpraakTYPES	DATUM_AFGEHANDELD	
				TEKST	"Klant is aangemeld of reeds gestart op activiteit RTB WOF 01 DLN.", "Herkeuring AOB", "De geplande einddatum van trajectplan Werk en Inburgering is bereikt."
				RESULTAAT_NAAM	"Beëindigd, wél Participatieplaats", "Check correcte registratie controlegroep", "Geen ZP, werk of onderwijs"
				AFSPRAAK_NAAM	"Beëindigd, wél Participatieplaats", "Check correcte registratie"

					controlegroep", "Geen ZP, werk of onderwijs"
				SYSTEEMTEKST	"Aan de klant is een mondelinge waarschuwing opgelegd. Onderzoek naar verlenging van de waarschuwing.", "Aanmelding ABQ RIO 01 aangemeld op 01-09-2010 afgesloten, kwaliteitsoordeel nog invullen."
BELEMMERING	Features		KVS_KLA_BELEMMERINGEN_JN, SGR_BELEMMERINGWERKAA NVAARDING	AANVANG, EINDE	
				OMSCHRIJVING_DWI	"Psychische problemen", "Niet computervaardig", "Volledig ontbrekende verdien capaciteit"
				TOELICHTING	"PTSS. In behandeling bij psycholoog van Pionier te Amsterdam.", "Dit gaat tijdens het gesprek op 9-2-2015 beter. Ze geeft aan dat haar dochter haar helpt.", "meneer kampt met schulden, klant heeft een aantal regeling lopen."
BESCHIKBAARHEID	Features		KVS_KLA_BESCHIKBAARHEID_JN, KVS_BESCHIKBAARHEID	AANVANG, EINDE	
				BESCHIKBAARHEID_NAAM	"Afwijkend wegens sociaal/maatschappelijke situatie", "Beschikbaarheid afwijkend wegens zorg kind < 5 jaar"
				BESCHIKBAARHEID_TOELICHTING	"Zware lichamelijke en psychische klachten", "zie rapp.", "Onder behandeling BAVO, last van chronische psychose"

COMPETENTIE	Features		KVS_KLA_COMPETENTIES, KVS_COMPETENTIES	DATUM_AANVANG, DATUM_EINDE	
				COMP_OMSCHRIJVING	"Inlevingsvermogen", "Beslissen en activiteiten initiëren", "Stressbestendig"
				COMP_TOELICHTING	"behandeling psycholoog ivm psych klachten", "ADHD, maakt rare taalfouten, komt niet geloofwaardig over.", "analfabeet, niet aanspreekbaar in Ned.taal."
CONTACT	Features		KVS_CONTACTEN_JN, ALG_CONTACT_ONDERWERP EN	DATUM_VASTLEGGING	
				SOORT_CONTACT_OMSCHRIJVING	"E-mail (uitgaand)", "Gesprek", "Document (uitgaand)"
				ONDERWERP	"Inspanningstoets", "Traject", "(Arbeids)motivatie"
DEELNAME_ACTIVITEITEN	Features		SAS_DEELNAMES_JN, SAS_REDENEN	DATUM_START, DATUM_EINDE	
				REINTEGRATIELADDER	"Activering", "Ondersteunende instrumenten", "Werk/Re-integratie"
				VOORTIJDIG_AFGEBROKEN	"Betaald werk gevonden", "Disfunctioneren", "Verhuizing buiten gemeente"
INSTRUMENT	Features		SAS_CONTRACTEN, SAS_PROJECTEN, SAS_PRJ_CSA, SAS_CATEGORIE_SDA,	DATUM_AANMELDING, DTEINDE	
				REDEN_BEEINDIGING	"Matchbaar", "Doelstelling bereikt, uitstroom naar regulier werk", "Overdracht naar Prematching"



			SAS_AANMELDINGEN_JN, SAS_REDENEN	REINTEGRATIELADDER	"Werk/Re-integratie", "Ondersteunende instrumenten", "Activering"
ONTHEFFING	Features		KVS_KLA_ONTHEFFINGEN_J N, KVS_KLA_ONTHEFFINGEN	AANVANG, EINDE	
				ONTHEFFING	"Sociale gronden", "Tijdelijke ontheffing arbeidsverpl. en tegenprestatie", "Medische gronden"
PERS_EIGENSCHAPPEN	Features		KVS_PERSOON_EIGENSCH_J N_VW	MOTIVATIE_OPM, FLEXIBILITEIT_OPM, DOORZETTINGSVERMOGEN_O PM, ZELFSTANDIGHEID_OPM, INITIATIEF_OPM, COMMUNICATIE_OPM, PRESENTATIE_OPM, UITERLIJKE_VERZORGING_OPM , LEERGIERIGHEID_OPM, HOUDING_OPM, HOBBIES_SPORT, OPSTELLING, SOLL_GEDRAG, OVERIGE_OPMERKINGEN, SPREEKTAAL_ANDERS, NL_SPREKEN, NL_LEZEN, NL_SCHRIJVEN, NL_BEGRIJPEN, SPREEKTAAL, IND_VERKL_GOED_GEDRAG, IND_REGULIER_ARBEIDSRIJME, IND_BUITEN_KANTOORTIJDEN, IND_ACTIVERING_TRAJECT, UITSTROOM_VERW_VLGS_KLA NT, UITSTROOM_VERW_VLGS_KM, TAALEIS_VOLDAAN,	

				TAALEIS_SPREEKV_OK, TAALEIS_LUISTERV_OK, TAALEIS_GESPREKSV_OK, TAALEIS_SCHRIJFV_OK, TAALEIS_LEESV_OK	
PERSOON	Features	PERSOON	KVS_KLANTEN	GESLACHT, GEBJAARMAAND	
PLA	Features		KVS_TRAJECTPLANNEN_JN, KVS_KLANTEN, KVS_REDENEN_TRAJECTPLA N_EIND, KVS_TRN_TYPEN	DATUM_BEGIN, RTE_DATUM, DATUM_ONDERTEKENING	
				PLA_NAAM	"Bemiddeling", "Diagnose", "Sociale activering"
				REDEN_EINDE_NAAM	"Doelstelling bereikt, overdracht", "Klant ziet af van aanbod", "Uitstroom als zelfstandige"
				DOELSTELLING_TKST	"Uitstroom naar regulier werk door middel van korte (werkgerelateerde) trajecten en/of regelmatige afspraken.", "Maatschappelijke participatie door middel van activiteiten die gericht zijn op het actief positioneren in de samenleving."
RELATIE (KIND/PARTNER/OVERIG)	Features	RELATIE		DTAANVANG, DTEINDE	
				SOORT	"ouder -> kind", "verhuurder -> huurder", "partner - partner (gehuwd)"
				CATEGORIE	"kindrelatie", "medebewoner", "partnerrelatie"
TYPERING	Features		KVS_KLA_TYPERINGEN_JN, KVS_KLA_TYPERINGEN	DATUM_INGANG, DATUM_EINDE	

				TYPERING_NAAM	"Sector Transport & Logistiek", "Vluchteling", "2e profiel: Snel, Precies & Fysiek fit"
UITKERING	Referentie (voor bepaling te scoren populatie)	STOPZETTING, REDEN, DIENSTREDEN, DIENST, SRTDIENST, PRODUCT, SRTVERSTREKKING		DTBEGINDIENST, DTEINDEDIENST, DTBEGINSTOPZETTING	
				SRTDIENST_OMSCHRIJVING	"Participatiewet", "WIJ", "loaw"
				VERSTREKKING_OMSCHRIJVING	"Levensonderhoud", "Zelfstandige Bbz Binnenschippers", "Beginnende zelfstandige (Bbz)"
				SOORT_UITKERING_OMSCHRIJVING	"Participatiewet", "WIJ", "WWB (BBZ)"
				STOPZETTING_OMSCHRIJVING	"Geen inlichtingen/bewijsstukken (rechtmatigheid)", "Klant is overleden", "Door medewerker zelf (noodrem)"
				UITSTROOM_OMSCHRIJVING	"Bereiken leeftijd van 27 jaar", "Detentie", "Inkomsten uit uitkering arbeidsongeschiktheid"
				BESLUIT	"Principe toekenning", "Toekennen"
WERKOPDRACHT	Referentie (voor targetbepaling, training, testing)	COMPONENT, DIENSTFRAUDESUBJECT, WERKOPDRACHT, SRTWERKOPDRACHT, AFDELING, FRAUDESUBJECT, FRAUDE, PROJECT		FRAUDE_BEDRAG	
				AFDELING_KLM_OP_DTMELDING	"Intake Werk & Prematching Noord", "Locatie Delfshaven", "Jongerenloket"
				GEVOLGEN	"Aanpassing lopende uitkering", "Ongewijzigde voortzetting", "Beëindiging lopende uitkering"

				OMS_PROJECT	"HO2018_Aselect NPRZ", "BO2018_Horeca", "PHO17_Matching"