Gemeente Rotterdam

accenture
High performance. Delivered.

Rapport project analytics uitkeringsfraude

Versie 1.99

Opdrachtgever	W&I Toetsing en Toezicht / BCO Onderzoek en BI
Auteurs	Projectgroep analytics uitkeringsfraude
Datum	23-08-2018
Status	Concept

1 Inhoud

Rapport project analytics uitkeringsfraude 1
1 Inhoud 3
2 Documentinformatie 5
2.1 Documenteigenschappen 5
2.2 Documenthistorie 5
2.3 Reviews 5
2.4 Gerelateerde documenten 5
3 Managementsamenvatting 6
4 Voorbereiding 7
4.1 Fasering 7
4.2 Verkenning en mobilisatie 7
4.3 Dataverzameling en -ontsluiting 7
4.4 Gebruikte analysesoftware en infrastructuur 8
5 Aanpak en gebruikte technieken 9
5.1 Een risicomodel 9
5.2 Target 10
5.3 Analytics Base Table en feature selectie 11
5.4 Text mining 13
5.5 Modelleren en testen van model 13
6 Resultaten labfase 15
6.1 Resultaten modellen op data Cognos-extracties 15
6.2 Socrates-extractie 16
6.3 Toevoegen voorgaande jaren en gestructureerde RAAK-data 16
6.4 Toevoegen data text mining 17
6.5 Overzicht en interpretatie van de resultaten 18
6.6 Toepasbaarheid 19
6.7 Volgende stappen 20
7 Aanpak en activiteiten pilotfase en voorbereiding uitrol 21
7.1 Ontwikkeling pilot-model 21
7.2 Onderzoeksopzet pilot 21
7.3 Herontwerp dataleveringen 22
7.4 Bestendigen van model en code, raamwerk toekomstige modellen 22
8 Resultaten en conclusies pilotfase 23
8.1 Hitrates: het model werkt 23
8.2 Selectie: veel afvallers door actualiteit en positieve uitstroom 24
8.3 Eigenschappen van de risicogroep 25
8.4 Verbanden tussen belangrijke features en risicogroep 25
9 Aanbevelingen doorontwikkeling en vervolgonderzoek 28
9.1 Aanknopingspunten voor vervolgonderzoek 32
9.2 Doorontwikkelmogelijkheden voor het risicomodel 32
9.3 Aanbevelingen rondom het selectieproces van T&T 34
10 Geleerde lessen 36
10.1 Juiste mix van omstandigheden 36
10.2 Data ontsluiten en gegevensmanagement 37
10.3 Lab-omgeving voor innovatie is noodzakelijk (maar niet ingewikkeld) 38
10.4 Privacy wordt handelbaar door PIA en goede fasering 39
10.5 Samen stappen zetten noodzakelijk voor leereffect en voortgang 40

10.6 Eigenaarschap, aansturing en verandermanagement bij innovaties . 40
10.7 Praktijkervaring en pionieren blijft nodig voor groei 41
BIJLAGE A – GEBRUIKTE DATA EN ABT's 42
BIJLAGE B – GEDETAILEERDE MODELRESULTATEN 45
BIJLAGE C – VERGELIJKING POPULATIES 46
Vergelijking van de onderzoekspopulatie met de gehele populatie uitkeringsgerechtigden 46
Zijn er significante verschillen tussen de populaties in de verschillende features? 46
Zijn de verschillende tussen de populaties betekenisvol? 46
Inzoemen om de features met een gemiddelde effectsize 46
Features 47
Conclusie 1 50
BIJLAGE D – PRIVACY IMPACT ASSESSMENT 56
Documenteigenschappen 56
Documenthistorie 56
Inleiding 56
Doelbinding 56
Proportionaliteit 56
Anonimisering en toegestane gegevensvastlegging 56
Informatiebeveiliging 57
Beantwoording vragenlijst Privacy Impact assessment Labfase Analytics Uitkeringsfraude 58

2 Documentinformatie

2.1 Documenteigenschappen

Titel	Rapport
Auteur	Projectgroep
Versie	1.99
Versie Datum	8-2-2018
Versie Status	Concept

2.2 Documenthistorie

Datum	Versie	Aangepast door	Opmerkingen
11-1-2017	0.5	[redacted]	Initiële versie in de vorm van verslag labfase
17-1-2017	0.6	[redacted]	Verder invullen H6 en 7 en verwerken review [redacted]
1-2-2017	1.0	[redacted]	Verwerken reviews [redacted]
14-8-2017	1.5	[redacted]	Start met verwerken pilotresultaten en ombouw tot Rapport van het project
27-9-2017	1.7	[redacted]	Vervolg aanvullen pilotresultaten, vullen H7,8
23-1-2018	1.9	[redacted]	Afgerond tot eindconcept eindrapport project
8-2-2018	1.99	[redacted]	Reviews verwerkt
	2.0		Eindversie ter bespreking in begeleidingsgroep

2.3 Reviews

Datum	Versie	gereviseerd door / goedgekeurd	Opmerkingen
16-1-2017	0.5	[redacted]	
25-1-2017	0.6	[redacted]	Diverse tekstverbeteringen en aanscherpingen
22-8-2017	1.5	Begeleidingsgroep, [redacted]	Mangementsamenvatting besproken in de begeleidingsgroep.
24-8-2018	1.9	[redacted]	

2.4 Gerelateerde documenten

Dit document doet verslag van de aanpak en resultaten van de pilot van het project Analytics uitkeringsfraude. Het is een uitbreiding van het verslag van de eerder voltooide labfase: de managementsamenvatting is aangepast aan de stand van zaken na de pilot en er zijn hoofdstukken toegeveogd over de aanpak en resultaten van de pilot. De volgende documenten liggen ten grondslag aan dit verslag:

- Verslag labfase uitkeringsfraude – 8 maart 2017
- Plan van Aanpak Pilotfase v1.1 – 13 februari 2017
- Privacy Impact Assesment pilot Project Uitkeringsfraude, v1.0 – *20 januari 2017*
- Technical Decisions Document Risicomodel Uitkeringsfraude – *Groeidocument*
- Verslagen bijeenkomsten begeleidingsgroep
- Plan van aanpak Informatiegestuurd Werken, juli 2016

3 Managementsamenvatting

Het doel van het Project Analytics Uitkeringsfraude (augustus 2016 – februari 2018) was om met analytics-technieken een risicomodel te ontwikkelen dat onrechtmatigheid bij gemeentelijke uitkeringen voorspelt ten behoeve van het handhavingsproces. Het beoogde resultaat is een beter en sneller inzicht in fraudegevallen bij bijstandsverstrekking, waardoor een verhoogde efficiency wordt bereikt:

- Enerzijds door vermindering van gemeentelijk geld dat onterecht wordt uitgekeerd aan bijstand
- 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 nevendoelstelling was kennisverwerving en vaardigheidsontwikkeling van (innovatieve) data-analysetechnieken binnen de gemeente.

De pilot heeft aangetoond dat rechtmatigheidsonderzoeken bij werkzoekenden die hoog scoren in het ontwikkelde risicomodel (de “top-groep”) in circa 40% van de gevallen tot een ‘hit’ (beëindiging of terugvordering van de uitkering) leiden tegenover 30% in de controlegroep. Ruim 22% van de 151 onderzochte uitkeringen van de top-groep werden beëindigd tegenover slechts 9% in de controlegroep. Daarmee is het aantal beëindigingen significant hoger voor de werkzoekenden met een hoge risicoscore dan bij willekeurig geselecteerde werkzoekenden, dit maakt het aannemelijk dat dit verschil niet op toeval berust maar een effect is van het ontwikkelde model. De totale hit-ratio lijken ook hoger te liggen dan andere eerder gebruikte selectiemethoden zoals woonprofielen, maar door de kleine aantallen is deze uitspraak nog niet statistisch te verifiëren.

Deze resultaten bevestigen de resultaten van de labfase, waarin al met een eerdere versie van het model op historische populaties (zonder daadwerkelijk nieuwe onderzoeken uit te voeren) betere voorspelkracht werd gezien dan met huidige werkwijzen. Bovendien verwacht het projectteam dat met verbetering van de onderliggende data en betere afstemming van de training van het model op het gebruik in de praktijk de hit-ratio nog omhoog kan in volgende versies van het model.

Met de gemeten verbeteringen levert een business case voor grootschaliger gebruik van dit risicomodel een mogelijke jaarlijkse besparing van circa EUR 2 miljoen per jaar door extra beëindigde en teruggevorderde uitkeringsbedragen. Dit bedrag is geldig wanneer ongeveer 3000 heronderzoeken worden uitgevoerd op basis van de hoogste risicoscores (de helft van de jaarlijkse doelstelling). De aannames en uitwerking bij deze business case zijn in een apart document te vinden.

Bij de afronding van dit project kan worden geconcludeerd dat met succes een model is ontwikkeld, getest en gevalideerd dat klaar is voor grootschalig gebruik om de doelstelling te bereiken, met een positieve business case bij voldoende gebruik. Het is aan W&I om het ontwikkelde product in te zetten en de ontwikkelkosten terug te verdienen. De ervaring leert dat voortdurende aandacht en hard werk nodig is om te voorkomen dat het bij een succesvol project blijft dat niet duurzaam landt in het primaire proces.

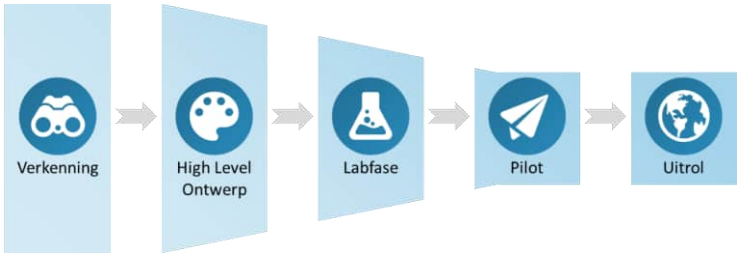
Daarnaast is het aan programma Datagedreven Werken om de opgedane ervaringen en geleerde lessen in bredere zin te borgen en in nieuwe trajecten te benutten. Het project analytics uitkeringsfraude is met dit machine learning model de eerste succesvolle toepassing van advanced analytics binnen de gemeente Rotterdam. Er is nu ook een raamwerk voor het uitvoeren van andere supervised machine learning projecten binnen de gemeente; de genomen stappen en een deel van de code zijn herbruikbaar. Dit project heeft naast bovenstaande conclusies over de onderzoeksvragen, waardevolle lessen opgeleverd over tal van onderwerpen op het gebied van informatiegestuurd werken: gegevensmanagement, privacy, infrastructuur en tooling, organisatie (zoals: het betrekken / commiteren van ‘de werkvloer’ en meer inzicht in voorwaarden voor het succesvol toepassen van analytics in het primaire proces). Een belangrijke conclusie is dat een project als dit echt pioniert binnen de gemeentelijke organisatie en veel hordes moet nemen op weg naar resultaten. Het nemen van die hordes lukt echter wel en de ervaring die zo wordt opgedaan draagt bij aan een volwassener organisatie op het gebied van IGW en GM. Momenteel zijn de gebrekkige infrastructuur en de mogelijkheden voor beheer nog onvoldoende om tot duurzame uitrol over te gaan, al is het model zelf daar wel bijna klaar voor. De geleerde lessen zijn in hoofdstuk 7 beschreven.

In de rest van dit rapport vindt u meer informatie over de gevolgde aanpak en details van de resultaten. Hierbij zijn aparte hoofdstukken gewijd aan de lab- en de pilotfase.

4 Voorbereiding

4.1 Fasering

Het project uitkeringsfraude is in fases opgeknipt om de beheersbaarheid te vergroten en periodiek te kunnen afwegen of de investering nog gerechtvaardigd is. Hiervoor wordt een vaste indeling gehanteerd die is weergegeven in Figuur 1.



Figuur 1: Innovatietrechter

Dit rapport beschrijft de uitvoering en resultaten totaan de uitrolfase. De labfase is uitgevoerd tussen augustus en december 2016, de pilot tussen januari en augustus 2017. Tot februari 2018 is het model verder versterkt met de lessen uit de pilot en zijn afspraken gemaakt over het gebruik, al kan vanwege beperkingen in de organisatie (o.a. infrastructuur, onvolwassen werkprocessen) duurzame uitrol nog niet volledig plaatsvinden.

4.2 Verkenning en mobilisatie

Voorafgaand aan de formele start van de labfase is een exploratie- en mobilisatiefase uitgevoerd. Daarin zijn afspraken gemaakt tussen BCO (OBI) en W&I (T&T) als gezamenlijke opdrachtgevers en Accenture en de Hogeschool Rotterdam als opdrachtnemers. Het projectteam is samengesteld en de aanpak van de labfase is vastgesteld. De gekozen aanpak gaat uit van een iteratieve en gefaseerde aanpak van dit analytics-project, zoals weergegeven in onderstaande figuur. De labfase had als doel eerste risicomodellen te bouwen en de onderzoeksvragen te beantwoorden teneinde een go/ no go-beslissing te kunnen nemen voor het uitvoeren van een pilot.

In de mobilistatiefase werd al duidelijk dat enkele belangrijke randvoorwaarden, voornamelijk op het gebied van de benodigde infrastructuur en op het gebied van de beschikbaarheid van data, nog onvoldoende waren ingevuld en dus als onderdeel van de labfase zouden moeten worden ingericht.

Onderdeel van de mobilisatie was het maken van afspraken met de privacy- en security officers van W&I, om akkoord te krijgen op het verwerken van de persoonsgegevens van uitkeringsgerechtigden ten behoeve van de analyses. Het doel van het project en de maatregelen om privacy en gegevensbeveiliging te beschermen zijn hiervoor vastgelegd in een Privacy Impact Assesment voor de labfase. Tevens is de afspraak gemaakt om deze PIA opnieuw te doen voor de volgende fasen van het project.

4.3 Dataverzameling en -ontsluiting

Na de start van de labfase was de eerste prioriteit om de benodigde data beschikbaar te krijgen voor het projectteam om de analyses te kunnen starten. Uitgangspunt hierbij waren door Toetsing & Toezicht ([redacted]) ter beschikking gestelde lijsten met de resultaten van rechtmatigheidsonderzoeken uit 2015, gegevens van de onderzochte personen en hun relaties (partners, kinderen etc.). Deze lijsten waren gemaakt door middel van een Cognos-extractie.

Daaraan moest in elk geval de ongestructureerde data worden toegevoegd uit gespreksverslagen, om de text mining uit te voeren die de tweede onderzoeksvraag kon beantwoorden. Daarnaast was de inschatting van het projectteam dat modellen die gebouwd werden op de Cognos-extractie onbetrouwbaar waren en/of niet voorspellend. Dit was omdat van een zeer groot deel van de kenmerken van uitkeringsgerechtigden in de dataset alleen de waardes bekend waren na afloop van het onderzoek; de gebouwde modellen zouden daarom ofwel ‘voorspellen met voorkennis’, wat de resultaten onbruikbaar maakt voor voorspellen van onrechtmatigheid wanneer die voorkennis niet beschikbaar is, ofwel de ‘voorkennis’-kenmerken moeten uit de data worden weggelaten, waarna er te weinig informatie overbleef om een model te maken dat beter scoort dan willekeurig trekken. Deze twee effecten zijn nader beschreven en onderbouwd in het hoofdstuk resultaten.

Om deze redenen was het noodzakelijk om aanvullende data-extracties op te vragen uit het datawarehouse van W&I. Het projectteam heeft hierbij nauw samengewerkt met het BI-team van W&I en de databasebeheerders ([redacted]). Zo zijn uiteindelijk in verschillende dataleveringen de benodigde datasets beschikbaar gekomen voor het projectteam. Deze bestonden uit 15GB aan data verdeeld in 7 Access-databases met in totaal 44 tabellen met tot een miljoen regels over circa 80.000 personen. De gebruikte datasets zijn beschreven in het overzicht in Bijlage A. Een moeilijkheid voor het team was dat een goede beschrijving van de datasets ontbrak.

Het verzamelen en ontsluiten van de data is een essentieel onderdeel van elk IGW-project en de opgedane ervaring biedt waardevolle lessen voor vervolgprojecten en voor de inrichting van gegevensmanagement. Meer hierover is te vinden in Hoofdstuk 7.

4.4 Gebruikte analysesoftware en infrastructuur

Voor het maken van een risicomodel is analysesoftware nodig die de data kan inlezen en de modelberekeningen kan doen. De gebruikte technieken gaan voorbij de mogelijkheden van de momenteel bij de gemeente beschikbare software. Spreadsheets als Excel of statistische programma's zoals SPSS zijn voor het beoogde doel niet geschikt. Er zijn vele programma's en pakketten in omloop die gebruikt kunnen worden voor analytics en machine learning. In de mobilisatiefase is afgesproken om als analysesoftware te kiezen voor 'R'. Omdat deze software niet beschikbaar is op de reguliere werkstations van de gemeente zijn de allereerste analysewerkzaamheden uitgevoerd op een reguliere laptop van de gemeente die niet met het gemeentelijke netwerk was verbonden. De rekenkracht van deze laptop was echter al snel onvoldoende en bovendien was samenwerken problematisch op losse laptops, zodat er moest worden gezocht naar alternatieven om analysesoftware en de te analyseren data bij elkaar te brengen.

Verschillende alternatieven zijn onderzocht en geprobeerd. Het pad om intern bij de gemeente een labaratoriumomgeving in te richten met de benodigde specificaties leidde niet (snel genoeg) tot resultaat. Uiteindelijk is daarom besloten om een tijdelijke omgeving in te richten in een beveiligde cloud via het Accenture Cloud Platform/ Amazon Web Services. Deze bestaat uit remote-desktop-toegang tot een beveiligde server waarop analysetool R studio, programmeertaal Java en databasesysteem SQL kunnen worden gebruikt (en op aanvraag andere software). Gebruikers die toegang nodig hebben tot de omgeving kunnen, na autorisatie en installatie, de server benaderen vanaf een met het internet verbonden computer.

Deze server wordt gehost door Amazon Web Services op een server in Europa (Frankfurt). De basis- server is volledig schaalbaar vanaf de basis van een zogeheten 'r3.xlarge'-server met 4 processoren, 30,5 GB RAM en 80 GB opslagruimte; als meer rekenkracht is vereist was kan deze tijdelijk twee, vier of acht keer zo krachtig worden gemaakt.

Aan het eind van het project is deze omgeving vervangen door een vergelijkbare omgeving onder duurzamer beheer. Dit is echter nog steeds een tijdelijke oplossing, in afwachting van het door de gemeente te ontwikkelen Advanced Analytics Platform dat onderdeel zal worden van de gemeentelijke infrastructuur in de loop van 2018.

De zoektocht naar een geschikte constructie biedt waardevolle lessen voor vervolgprojecten en de doorontwikkeling van de gemeentelijke infrastructuur. Hierover is meer te vinden in Hoofdstuk 7.

5 Aanpak en gebruikte technieken

5.1 Een risicomodel

Een risicomodel sorteert een populatie op kans (het risico) dat een vooraf gedefinieerde stelling waar is. Die stelling was voor dit project “Er is sprake van onrechtmatigheid bij de uitkering van deze persoon” (zie hieronder voor de gekozen technische definitie van het target). Het model sorteert dus de populatie zo, dat de personen met de hoogste kans op onrechtmatigheid bovenaan komen en degenen met de kleinste kans onderaan. De sortering komt simpel gezegd tot stand door kenmerken van nieuwe individuen te vergelijken met (zoveel mogelijk) historische voorbeelden waarvan al bekend is of de stelling waar of niet waar was. Het risicomodel zoekt dus naar patronen in de bekende historische gevallen om die te herkennen in nieuwe gevallen. Door het gebruik van zoveel mogelijk data en geavanceerde analytics technieken, kan de rekenkracht van computers worden gebruikt om patronen te vinden die door een mens niet zouden worden herkend.

Er zijn diverse modelleertechnieken waarmee een risicomodel kan worden gemaakt en in dit project zijn meerdere technieken toegepast en uitgeprobeerd. Alle modellen zijn gemaakt met het R-pakket “caret”:

Modeleertechniek	Uitleg	Gebruikt algoritme
Desicion tree	Een door de computer gegeneerde beslisboom. Het algoritme verdeelt de data in steeds kleinere homogene groepen. De verdeling wordt gemaakt op basis van de features. Welke feature er op welk moment gekozen worden om de data op te verdelen, wordt door het algoritme berekend. De feature die leidt tot de meest homogene groepen wordt gekozen. Dit proces wordt per “tak” van de boom herhaald.	CART classification tree. Split gebaseerd op de Gini index Tuning parameters: cp (complexity parameter)
Random Forest	Een random forest is een verzameling van vele bomen. Iedere boom wordt gemaakt op een subset van alle observaties en alle kolommen. Op deze subset wordt de beste boom gecreëerd. Bij het voorspellen van het target krijgt iedere boom een stem. De groep met de meeste stemmen wint en wordt de voorspelling voor die observatie.	Random forest met CART classification tree. Split gebaseerd op de Gini index. Tuning parameters: Mtry (hoeveelheid random features gebruiken per split)
Gradient Boosting	Met gradient boosting worden verschillende bomen die apart van elkaar werken als zwakke classifiers samen gecombineerd tot een sterk model. Iteratief wordt er een boom gegroeid op de data. De observaties die foutief zijn voorspeld krijgen in de volgende ronde een zwaardere weging. De bomen worden gewogen afhankelijk van de hoeveelheid data in de bomen.	Stochastic Gradient Boosting met CART classification trees. Split gebaseerd op Gini index. Tuning parameters: shrinkage, n.minobsinnode, n.trees, interaction.depth
Regressie	Methode om een lineair verband te vinden dat de sum-of-squared errors minimaliseert tussen het geobserveerde en voorspeelde uitkomsten. Methode wordt veelvuldig gebruikt in de statistiek.	Generalized linear model, family = binominal Tuning parameters: geen
Genetisch algoritme		Algoritme ontwikkeld door Hogeschool Rotterdam

5.2 Target

Zoals hierboven duidelijk werd gemaakt moet voor een risicomodel een target worden gedefinieerd. De definitie van het target bepaalt waarnaar het risicomodel zoekt en is dus erg belangrijk om het goede model te kunnen ontwikkelen.

Er is gekozen om als target 'onrechtmatigheid' te gebruiken: een persoon ontvangt meer uitkering dan waarop hij recht heeft. Opzettelijke fraude gebruiken als target geeft op dit moment te veel onzekerheid. Opzet kan met de in de labfase beschikbare gegevens niet altijd worden aangetoond. Wellicht wordt het in een vervolgfase mogelijk en wenselijk om andere keuzes te maken voor het target; het is bijvoorbeeld in principe mogelijk om verschillende typen onrechtmatigheid en fraude apart te gaan voorspellen met een model door een gedifferentieerd target te bouwen, of in de voorspelling rekening te houden met het verwachte terugvorderbedrag. Dit vergt echter complexere modellen en aanvullende data, vandaar dat voor deze fase is gekozen het target breed en eenvoudig te houden.

Van alle historisch onderzochte personen wordt bepaald of het target waar of onwaar is op basis van de volgende definitie:

Er is sprake van onrechtmatigheid ALS:

Het veld 'Bedrag' IS NIET *leeg* OF

Het veld 'Gevolgen' == "Beëindiging lopende uitkering" OF

Het veld 'Gevolgen' == "Beëindiging lopende uitkering ABO" OF

Het veld 'Gevolgen' == "Aanpassing lopende uitkering".

Met andere woorden: als de uitkomst van het onderzoek was dat er een bedrag is teruggevorderd of de uitkering is aangepast of beëindigd, wordt dit aangemerkt als onrechtmatigheid.

5.3 Tardis

Om dit model te maken is het noodzakelijk om de volgende vraag te kunnen beantwoorden:

"Wat wisten wij van persoon/uitkering x op het moment vlak voordat een historisch onrechtmatigheidsonderzoek werd gestart."

Wanneer we deze informatie voor historische onderzoeken hebben kunnen we - gecombineerd met het bekende resultaat van de historische onderzoeken - een model maken waarbij we op basis van de gereconstrueerde data de uitkomst van het onderzoek voorspellen. Het model dat we hiermee verkrijgen kunnen we vervolgen toepassen op de actuele populatie waarbij we eigen de volgende vraag beantwoorden: Wat is het te verwachten resultaat van een potentieel rechtmatigheidsonderzoek?

De twee systemen waar we informatie uit gebruiken voor dit project zijn RAAK en Socrates.

Dit document beschrijft voor beide systemen afzonderlijk de stappen die we hebben genomen om een historische status te reconstrueren. Hierbij wordt onderscheid gemaakt tussen de volgende 2 stappen:

1. Wat was de inhoud van de database op historische datum X?

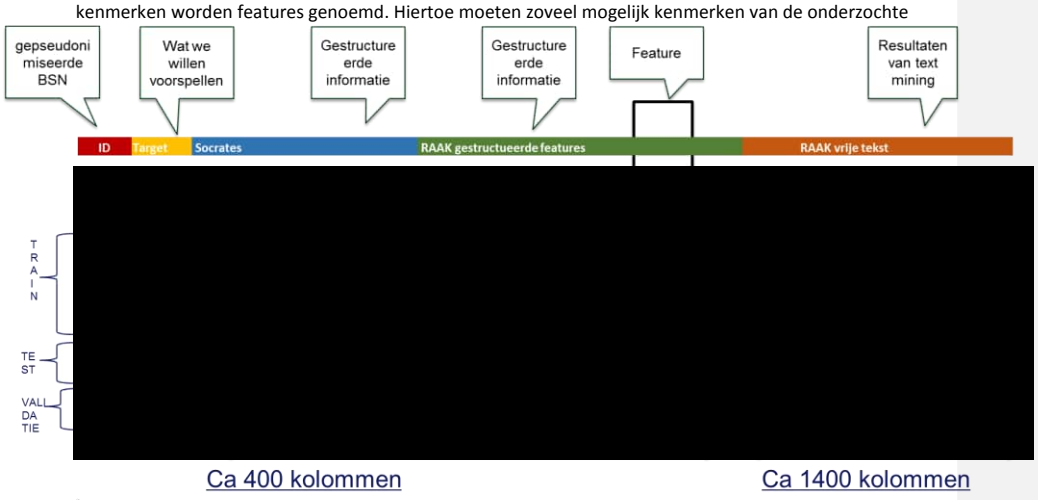
2. Wat was de actuele informatie op historische datum X?

Om het onderscheid tussen bovenstaande 2 te illustreren. Het kan zijn dat iemand een (psychische, sociale of fysieke) beperking had van 1-1-2015 t/m 31-12-2015. Wanneer we teruggaan naar de kennis van 1-7-2016 is er actueel geen sprake van een beperking (er was immers alleen in 2015 sprake van een beperking). Wel hadden we op 1-7-2016 de kennis dat er een historische beperking was.

5.4 Analytics Base Table en feature selectie

Het risicomodel zoekt zoals gezegd naar verbanden in de al onderzochte populatie, om deze verbanden vervolgens te kunnen terugvinden bij nog niet onderzochte personen. Een risicomodel wordt gemaakt door te leren van de data. Door dit leren maakt het algoritme een model specifiek voor een bepaald probleem.

Om een model te bouwen moet het algoritme zo veel mogelijk kenmerken hebben om van de leren. Deze kenmerken worden features genoemd. Hiertoe moeten zoveel mogelijk kenmerken van de onderzochte



Figuur 2 - schematische weergave van een ABT

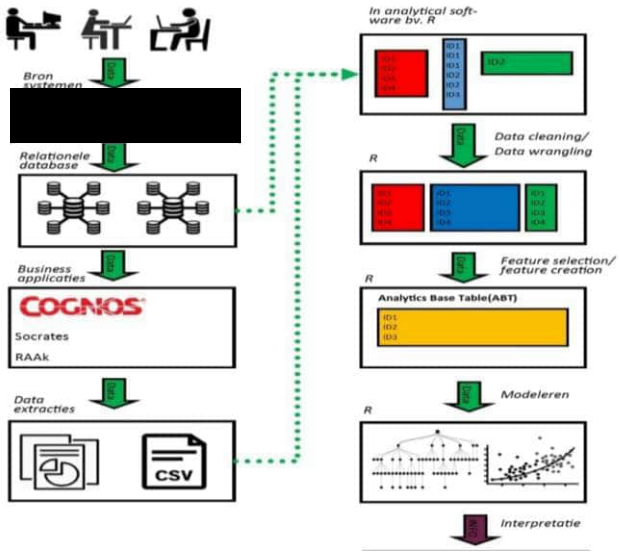
populatie, inclusief de waarde van het target, bijeengebracht worden in een grote tabel, die de Analytics Base Table wordt genoemd. In deze tabel is elke regel een persoon, en bevatten de honderden tot duizenden kolommen de kenmerken van die persoon; deze worden 'features' genoemd. Figuur 2 geeft een versimpeld voorbeeld van hoe deze tabel eruit ziet.

Het model is ontwikkeld, of 'getraind', op personen waarbij een onderzoek heeft aangetoond of het target waar of niet waar is. Dit is nodig zodat het model kan leren welke features een verband hebben met het target. Om te kunnen testen of het model goede voorspellingen doet is een deel van de populatie, de test-set, niet meegenomen bij het trainen zodat de voorspellingen die het model voor die personen doet kunnen worden vergeleken met de werkelijke uitslag van het onderzoek. Na de labfase wordt het model gevalideerd in een pilot: dan worden voorspellingen gedaan voor nog niet onderzochte personen en op basis van die voorspellingen onderzoeken gedaan.

Artikel 8.8 Woo jo artikel 65 Pw

Het bouwen van de ABT is de kern geweest van het werk in de labfase en is in stappen gebeurd waardoor verschillende, steeds uitgebreidere versies van de ABT zijn gemaakt. Om deze tabellen te maken zijn tientallen verschillende tabellen uit verschillende brondatabases aan elkaar gekoppeld. Uit die brontabellen worden naast vaste gegevens die rechtstreeks worden overgenomen, ook afgeleide features geconstrueerd waardoor vele bewerkingen op de oorspronkelijke data nodig zijn. Er wordt bijvoorbeeld uit alle losse regels van de 'belemmeringen' tabel geteld hoe vaak een persoon een belemmering heeft gehad, uit de tabel met werkervaring wordt uitgerekend hoe lang iemand gemiddeld een baan had en wordt er gekeken of er iets

Artikel 8.8 Woo jo. artikel 65 PW



genoteerd staat over de motivatie. Het proces van het opwerken van de data van bron tot ABT is schematisch weergegeven in onderstaande figuur.

Tabel 1 Beschrijving van de verschillende versies van de ABT (#rijen en features indicatief)

	Versie ABT	Data in ABT	# personen (regels)	# features (kolommen)
A	Basis	Elementaire Cognos-extractie tbv verslaglegging (2015)	8000	50
B	Basis geen onderzoek	Cognos-extractie minus directe uitkomsten onderzoek	8000	30
C	Basis gesnoeid	Cognos-extractie minus alle onbetrouwbare kolommen	8000	4
D	Basis plus 2015	Socrates-extractie van dezelfde velden om Time Travel problemen op te lossen (2015)	8000	50
E	Basis plus	Toevoegen alle jaren (2007-2015, 33.000 pers)	40 000	50
F	Advanced	Toevoegen RAAK tabellen alle jaren met aanvullende gestructureerde velden	40 000	400
G	Premium	Toevoegen RAAK-data Ongestructueerd (tekst)	40 000	1800

5.5 Text mining

Een van de expliciete vragen van W&I, vanuit de G12 waarin 12 gemeenten samenwerken aan betere handhaving, was of ongestructureerde tekst toegevoegde waarde heeft. Om een eerste antwoord op die vraag te vinden is in de labfase 'text mining' toegepast op de vrije tekstvelden in de applicaties, waarin onder andere gespreksverslagen van de onderzochte personen zijn opgeslagen. In deze gespreksverslagen wordt door medewerkers van W&I veel informatie vastgelegd over hun contacten met werkzoekenden, soms inclusief kopieën van e-mailwisselingen, waardoor de teksten soms duizenden woorden lang zijn. Veel teksten zijn echter ook heel kort. Text mining is een techniek om ongestructueerde teskt data te structueren. Er zijn vele vormen van text mining. Omdat het doel van dit project is het maken van een riscomodel, wordt hier text mining gebruikt om features te bouwen voor de ABT

In de labfase is alleen een eerste stap van text mining toegepast waar nog veel verfijningen en verbeteringen op mogelijk zijn. Deze techniek heet 'bag of words'; alle woorden zijn teruggebracht tot de stam en daarna is door een algoritme geteld hoe vaak elk woord voorkomt. De woorden zonder inhoudelijke betekenis, zoals lidwoorden, zijn eruit gefilterd en vervolgens is van circa 1400 veelvoorkomende woorden de frequentie als features toegevoegd aan de ABT. Dit is de stap van ABT-versie F naar G in bovenstaande tabel.

Er zijn nog heel veel verbeteringen denkbaar in dit deel van de analyse waardoor de text mining krachtiger kan worden. De reeds toegepaste technieken kunnen worden verbeterd en er zijn nog tal van aanvullende technieken die kunnen worden uitgeprobeerd.

5.6 Modelleren en testen van model

Zodra een ABT tot stand is gekomen kan het daadwerkelijke modelleren beginnen. Hiervoor wordt de ABT als input gegeven aan de verschillende modelleertechnieken die in 5.1 zijn beschreven. Het voordeel van het in stappen opbouwen van de ABT met steeds meer en verbeterde data is dat hierdoor ook het effect van het toevoegen en verbeteren van de data op de kwaliteit van het model duidelijker wordt; op elke versie van de ABT zijn meerdere modellen gemaakt met verschillende technieken. De kwaliteit van het beste model per versie van de ABT geeft een indicatie van hoeveel voorspelkracht er uit deze data te halen is. In hoofdstuk 6 worden deze resultaten gepresenteerd.

Er zijn vele statistische grootheden die de kwaliteit van een model beschrijven. Deze zijn te berekenen door het model uit te voeren op de test-set en de voorspellingen van het model te vergelijken met de werkelijkheid. Elke modelvoorspelling valt dan in een van vier categorieën:

- True positives: het aantal correct voorspelde gevallen van onrechtmatigheid: het model voorspelt dus dat er onrechtmatigheid is en dat blijkt te kloppen met de uitkomst van het onderzoek;
- True negatives: het aantal correct voorspelde gevallen zonder onrechtmatigheid: het model voorspelt dat er *geen* onrechtmatigheid is en blijkt te kloppen met de uitkomst van het onderzoek;
- False positives: het aantal onterecht als onrechtmatig aangemerkte gevallen: het model voorspelt dat er onrechtmatigheid is maar uit het onderzoek is gebleken dat dat niet het geval is;
- False negatives: het aantal onterecht als *niet onrechtmatig* aangemerkte gevallen: het model voorspelt dat er geen onrechtmatigheid is maar uit onderzoek is gebleken dat dat wel het geval is.

Het perfecte model heeft alleen voorspellingen in de eerste twee categorieën, maar dat is onmogelijk te bereiken. In de praktijk is het altijd een afweging, afhankelijk van het doel van het model, welke maat het meest belangrijk is; als je zoveel mogelijk van de onrechtmatigheidsgevallen wilt opsporen is het belangrijk dat het aantal false negatives zo laag mogelijk is. Het nadeel van een model dat daarop gericht is, is vaak dat het meer false positives geeft. In de context van Toetsing en Toezicht waar sprake is van een beperkte capaciteit is het van belang dat zo veel mogelijk van de als 'verdacht' aangemerkte gevallen ook echt onrechtmatig zijn, zodat de onderzoekscapaciteit zo efficiënt mogelijk wordt ingezet; het aantal false positives moet daarvoor zo klein mogelijk zijn, met weer als nadeel dat er waarschijnlijk ook meer false negatives zijn oftewel dat er onrechtmatigheden door de mazen van het model glippen en niet worden opgemerkt.

Een in de business erg bruikbare maat voor de kwaliteit die rekening houdt met deze afwegingen is de *hitrate bij een bepaald volume*. Het model geeft namelijk niet alleen een '1' (onrechtmatigheid) of '0' (geen onrechtmatigheid) maar een risicoscore tussen 0 en 1, waardoor een gerangschikte lijst kan worden gemaakt waarop de hoogste scores, dus het hoogste risico op onrechtmatigheid, bovenaan komen. De hitrate is dan gedefinieerd als het percentage true positives (correcte voorspellingen van onrectmatigheid) in een bepaald deel van de lijst, bijvoorbeeld de bovenste 100 of 1000 voorspellingen. Omdat de lijst van hoog naar laag risico is gesorteerd komen lager op de lijst dus steeds meer niet-onrechtmatige gevallen voor en zal de hitrate dalen naarmate het gekozen volume groter is. De resultaten zullen in het volgende hoofdstuk gepresenteerd worden door de hitrate af te zetten tegen het volume. De eigenschappen van de verschillende model-versies worden steeds beschreven met de volgende kenmerken:

Naam Model	De versie van de gebruikte ABT en het model dat daarop is gemaakt
Data in Model	Toelichting gebruikte data
%Ja	Het percentage onrechtmatigheid in de gebruikte populatie; dit is altijd veel hoger dan in de gehele populatie uitkeringsgerechtigden omdat de gebruikte populaties alleen de personen bevatten die zijn onderzocht, en waarbij er dus aanleiding was om dat onderzoek te starten.
Acc	De 'accuracy' van het model op basis van out-sample performance: Acc := (true positves + true negatives) / totaal # voorspellingen
Hitrate 500	De cumulatieve hitrate van het model in de eerste 500 voorspellingen
Hitrate 3000	De cumulatieve hitrate van het model in de eerste 3000 voorspellingen
Betrouwbaarheid	Kwalitatieve inschatting van hoe waarschijnlijk het is dat de resultaten ook in recente data standhouden.
Verbeter-potentieel	Mogelijkheden om op basis van dezelfde data het model nog verder te verbeteren
Verbetermogelijkheden op de toevoegde data	Specificatie van de verbetermogelijkheden

De gedetailleerde statistische scores van de belangrijkste model-versies zijn opgenomen in bijlage B.

6 Resultaten labfase

6.1 Resultaten modellen op data Cognos-extracties

De door T&T aangeleverde Cognos-extractie bevat kenmerken van de werkzoekenden én de resultaten van het onderzoek, vastgelegd op het moment van het maken van de extractie; dit is dus vaak langere tijd na afloop van het onderzoek, omdat de extracties zijn gemaakt in de loop van 2016 over de in 2015 uitgevoerde onderzoeken.

Op basis van die volledige extractie kan een perfecte 'voorspelling' worden gedaan van wel of geen onrechtmatigheid omdat de velden waaruit het target berekend wordt – de uitkomsten van het onderzoek – er gewoon in staan: het model zal dus bijvoorbeeld de simpele regel opleveren "als het veld 'terugvorderbedrag' niet 0 is of het veld 'gevolgen' gevuld is met 'beëindiging uitkering' of 'aanpassing uitkering', voorspel dan 'onrechtmatigheid', anders voorspel 'geen onrechtmatigheid'. Dit is uiteraard een zinloze exercitie aangezien je deze gegevens van niet onderzochte personen niet hebt en het model dan dus geen voorspelling kan doen. Voor de volledigheid is dit model wel gemaakt en opgenomen als versie A van de ABT in de overzichtstabel in 6.2; te zien is dat het model inderdaad een perfecte voorspelling doet (Accuracy en hitrates zijn 1,00), maar dat de betrouwbaarheid zeer slecht is: gezien bovenstaande redenen heb je in de praktijk niks aan dit model.

Voor een eerste serieus model zijn daarom alle kolommen die naar aanleiding van het onderzoek pas worden ingevuld verwijderd uit de ABT om tot versie B te komen. Het model krijgt dan dus niet meer de uitkomst van het onderzoek als input. Te zien is dat het model dan geen perfecte voorspelling meer kan doen, sterker nog, dat de modelresultaten matig worden. Hier komt een onderliggend probleem met de data aan het licht, namelijk dat deze is verzameld ruim na afloop van het onderzoek. Ook de velden die niet de directe uitkomst van het onderzoek bevatten kunnen daarom wel zijn aangepast met informatie uit het onderzoek, zonder dat dat evident is. Hierdoor kan het gebeuren dat het model patronen vindt die in een niet-onderzochte populatie niet aanwezig zijn. Een simpel voorbeeld hiervan in het model op ABT-versie B is dat het model de regel vond dat uitkeringen van mensen die als woonplaats een andere waarde hadden dan Rotterdam, bijvoorbeeld 'Suriname' of 'Breda', beduidend vaker onrechtmatig waren. Logisch natuurlijk aangezien niet-Rotterdammers per definitie geen recht hebben op een uitkering in deze stad, maar ook vreemd aangezien ze die uitkering nooit hadden kunnen krijgen als dit bij aanvraag al bekend is. In deze gevallen is dus hoogstwaarschijnlijk de daadwerkelijke woonplaats ingevuld die aan het licht kwam uit het onderzoek, terwijl voor het onderzoek nog gewoon 'Rotterdam' was ingevuld.

Dit probleem kan zichtbaar worden zoals in bovenstaand voorbeeld maar ook (in andere velden) verborgen zitten. Daarom wordt bij het maken van een risicomodel altijd alleen data gebruikt van vóór de start van ingrepen die tot aanpassingen kunnen leiden; in dit geval dus voor de start van het rechtmatigheidsonderzoek. Voor de volledigheid is nog een model gemaakt waarbij alle kolommen zijn verwijderd die in de loop van de tijd kunnen wijzigen (ABT-versie C); dan blijven alleen de kolommen geboortejaar, geslacht en geboorteland over. Zoals te verwachten blijkt dat een risicomodel op basis van zo weinig gegevens niet beter voorspelt dan willekeurig kiezen.

De resultaten van deze drie modelversies zijn als volgt:

Naam Model	Data in Model	%Ja	Acc	Hitrate 500	Hitrate 3000	Betrouw-baarheid	Verbeter-potentieel	Verbetermogelijkheden op de toevoegde data
A) Basis	Elementaire Cognos-extractie tbv verslaglegging (2015, ca 8000 pers))	0.61	1.00	1.00	1.00	zeer slecht	Niet	Geen
B) Basis geen onderzoek	Cognos-extractie minus directe uitkomsten onderzoek	0.61	0.70	0.82	NA	zeer slecht	Laag	Combinatie features maken
C) Basis gesnoeid	Cognos-extractie minus alle onbetrouwbare kolommen	0.61	0.63	0.65	NA	slecht	Niet	Geen

6.2 Socrates-extractie

Om de problemen met deze eerste datalevering op te lossen is historische data opgevraagd van dezelfde personen en zijn de waarden van alle features opnieuw vastgesteld zoals ze bekend waren voordat het onderzoek startte[1]. Dit levert ABT versie D op.

De modelresultaten op deze versie van de ABT zijn wat betrouwbaarder dan die van de eerdere modellen en er lijkt enige voorspelkracht in te zitten. De dataset is echter eigenlijk te klein voor een goed en stabiel risicomodel omdat het alleen gaat om de onderzoeken uit 2015. Dit levert problemen op in de modellen waardoor de betrouwbaarheid als slecht wordt ingeschat, ondanks redelijke statistische scores.

Naam Model	Data in Model	%Ja	Acc	Hitrate 500	Hitrate 3000	Betrouw-baarheid	Verbeter-potentieel	Verbetermogelijkheden op de toevoegde data
D) Basis plus 2015	Socrates-extractie van dezelfde velden om Time Travel problemen op te lossen (2015)	0.61	0.75	0.90	NA	slecht	Redelijk	Historische features maken, combinatie features maken, model tunen, expert features toevoegen, feature verbetering

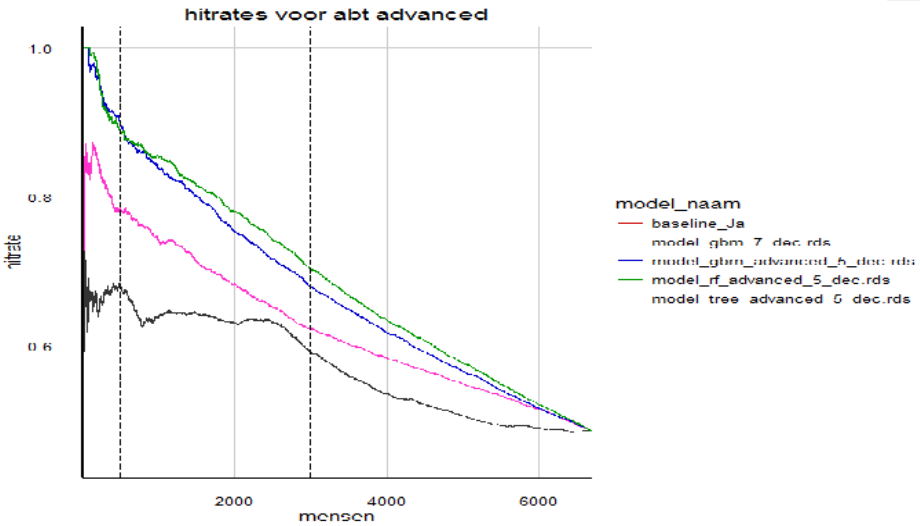
6.3 Toevoegen voorgaande jaren en gestructureerde RAAK-data

In de volgende stap is een grote hoeveelheid aanvullende data toegevoegd, namelijk de onderzoeken uit alle beschikbare jaren vóór 2015 sinds start van de vastlegging in 2007. Het gaat hier om circa 33 000 personen (ABT-versie E). Daarnaast is een grote hoeveelheid kolommen toegevoegd op basis van gestructureerde RAAK-data (ABT-versie F).

Naam Model	Data in Model	%Ja	Acc	Hitrate 500	Hitrate 3000	Betrouw-baarheid	Verbeter-potentieel	Verbetermogelijkheden op de toevoegde data
E) Basis plus	Toevoegen alle jaren (2007-2015, 33.000 pers)	0.49	0.63	0.78	0.62	voldoende	Hoog	Historische features maken, combinatiefeatures maken, model tunen, expertfeatures toevoegen, feature verbetering
F) Advanced	Toevoegen RAAK tabellen alle jaren met aanvullende gestructureerde velden	0.49	0.68	0.90	0.68	voldoende	Hoog	Historische features maken, combinatiefeatures maken, model tunen, expertfeatures toevoegen, feature verbetering, dimensiereductie

[1] Voor personen waarbij meerdere onderzoeken hebben plaatsgevonden is het moment gekozen voorafgaand aan het eerste pósitieve onderzoek (het eerste onderzoek dat als uitkomst onrechtmatigheid had).

Van vier verschillende modellen op ABT-versie F zijn hieronder de hitrate – volume plots weergegeven. Deze laten zien dat verschillende modeltechnieken op dezelfde ABT-versie sterk verschillende resultaten geven. De horizontale rode lijn geeft de verwachte hitrate bij willekeurig trekken aan.



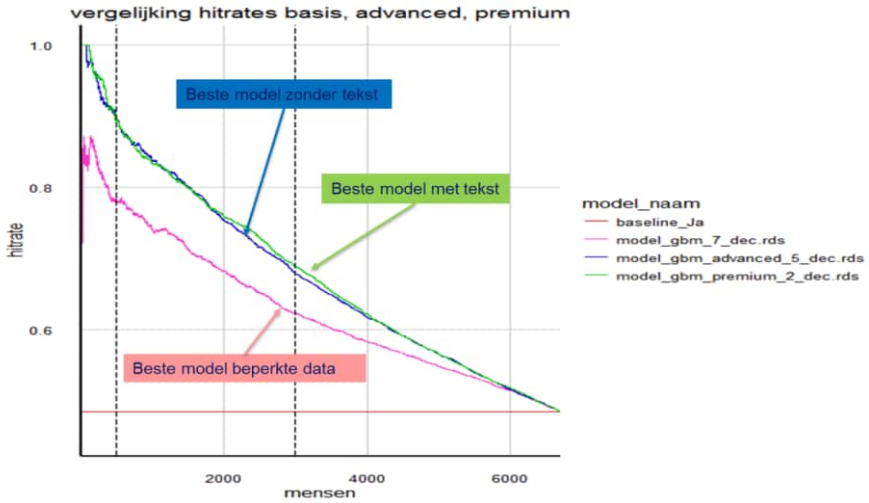
6.4 Toevoegen data text mining

De laatste stap was het toevoegen van de resultaten van de bag-of-words text mining. In deze stap zijn als features tellingen toegevoegd van circa 1.400 'gestemde' (tot de stam teruggebrachte) woorden die niet als stopwoord waren aangemerkt.

De modellen geven vergelijkbare resultaten als de beste modellen zónder de text mining, wat aangeeft dat er met deze eerste versie van text mining nog weinig extra voorspelkracht kan worden gehaald ten opzichte van het meest uitgebreide gebruik van gestructureerde data. Dit is niet onverwacht gezien het feit dat er nog vele mogelijkheden zijn tot verbetering van de text mining ten opzichte van deze eerste pogingen. Deze pasten echter niet in de tijd en scope van de labfase.

Wel is bij gedetailleerde analyse van de modellen zichtbaar dat de text-features wel worden gebruikt door het model; dat geeft aan dat de data in elk geval niet waardeloos is, alleen dat er op deze manier nog niet méér waarde uit te halen is dan uit de beschikbare gestructureerde data.

Naam Model	Data in Model	%Ja	Acc	Hitrate 500	Hitrate 3000	Betrouw-baarheid	Verbeter-potentieel	Verbetermogelijkheden op de toevoegde data
Premium	Toevoegen RAAK-data Ongestructueerd (tekst)	0.49	0.68	0.89	0,69	voldoende	Zeer hoog	Stemming, stopwoorden verbeteren, synoniemenlijst, spellingscheck, labels creëren, expert features maken, model tunen, dimensiereductie, historische features, combinatie features.



6.5 Overzicht en interpretatie van de resultaten

De labfase had twee expliciete onderszoeksvragen:

1. Kan een analytics-risicomodel onrechtmatigheid beter voorspellen dan de huidige werkwijze(n)?
2. Wat is de toegevoegde voorspellende waarde van ongestructureerde (tekst)data ten opzichte van gestructureerde data?

De hierboven gepresenteerde resultaten laten zien dat het mogelijk is om een risicomodel te maken dat onrechtmatigheid bij bijstandsuitkeringen voorspelt. Hiervoor is meer data nodig dan beschikbaar was in de initiële Cognos-extractie maar deze data is wel beschikbaar in de bron-databases van Cognos en RAAK.

Op basis van de resultaten zijn de onderzoeksvragen als volgt te beantwoorden:

Vraag 1: De nieuw ontwikkelde risicomodellen kunnen in de onderzochte historische populatie onrechtmatigheid beter voorspellen dan een groot deel van de huidige werkwijzen. Het toevoegen van meer en betere data verbetert de voorspellende waarde van de modelresultaten. Omdat de in de labfase gebruikte populatie per definitie niet representatief is voor de gehele populatie uitkeringsgerechtigden, omdat deze alleen daadwerkelijk onderzochte gevallen uit het verleden betreft, biedt dit resultaat geen zekerheid de voorspellingen ook in de praktijk waardevol blijven. Hiervoor is een praktijkonderzoek (een pilot) nodig die als volgende fase wordt uitgevoerd.

Vraag 2: Met de gebruikte rudimentaire vorm van text mining op de ongestructureerde data uit gespreksverslagen is nog geen significant verschil in voorspelkracht te zien tussen het beste model met alleen gestructureerde data en de modellen met ook ongestructureerde data. Daaruit kan niet worden geconcludeerd dat tekstdata geen toegevoegde waarde heeft. Wel lijkt het vanuit business-perspectief en efficiëntieoverwegingen het aanbevelen waard om eerst de voorspelkracht van de gestructureerde data te gaan benutten voordat grote investeringen worden gedaan in betere modellen met text mining.

De onderstaande tabel vat de resultaten samen van de beste modellen aan per versie van de ABT.

Naam Model	Data in Model	%Ja	Acc	Hitrate 500	Hitrate 3000	Betrouw-baarheid	Verbeter-potentieel	Verbetermogelijkheden op de toevoegde data
Basis	Elementaire Cognos-extractie tbv verslaglegging (2015, ca 8000 pers))	0.61	1.00	1.00	1.00	zeer slecht	Niet	Geen
Basis geen onderzoek	Cognos-extractie minus directe uitkomsten onderzoek	0.61	0.70	0.82	NA	zeer slecht	Laag	Combinatie features maken
Basis gesnoeid	Cognos-extractie minus alle onbetrouwbare kolommen	0.61	0.63	0.65	NA	slecht	Niet	Geen
Basis plus 2015	Socrates-extractie van dezelfde velden om Time Travel problemen op te lossen (2015)	0.61	0.75	0.90	NA	slecht	Redelijk	Historische features maken, combinatie features maken, model tunen, expert features toevoegen, feature verbetering
Basis plus	Toevoegen alle jaren (2007-2015, 33.000 pers)	0.49	0.63	0.78	0.62	voldoende	Hoog	Historische features maken, combinatiefeatures maken, model tunen, expertfeatures toevoegen, feature verbetering
Advanced	Toevoegen RAAK tabellen alle jaren met aanvullende gestructureerde velden	0.49	0.68	0.90	0.68	voldoende	Hoog	Historische features maken, combinatiefeatures maken, model tunen, expertfeatures toevoegen, feature verbetering, dimensiereductie
Premium	Toevoegen RAAK-data Ongestructueerd (tekst)	0.49	0.68	0.89	0,69	voldoende	Zeer hoog	Stemming, stopwoorden verbeteren, synoniemenlijst, spellingscheck, labels creëren, expert features maken, model tunen, dimensiereductie, historische features, combinatie features.

6.6 Toepasbaarheid

Al met al bieden de resultaten van de labfase een goed perspectief op het ontwikkelen van een risicomodel dat in de praktijk van het handhavingsproces kan worden gebruikt. Hiertoe zal het model echter moeten worden doorontwikkeld en gevalideerd. Dit is wat typisch gebeurt in een pilot op basis van de labfase-resultaten.

De belangrijkste kanttekening bij de resultaten van de labfase-modellen is dat deze gemaakt zijn op populaties van in het verleden uitgevoerde onderzoeken. Deze populaties zullen om twee redenen niet representatief zijn voor de populatie (huidige uitkeringsgerechtigden) waarvoor het model uiteindelijk voorspellingen zal moeten doen om te kunnen worden toepast:

- In de lab-populaties is per definitie sprake van onderzoeken die al zijn uitgevoerd. In de handhavingspraktijk worden vrijwel alle onderzoeken uitgevoerd naar aanleiding van een signaal, zoals een tip van een burger of een melding van wit inkomen. Om deze reden bevat de onderzochte populatie veel vaker onrechtmatigheid dan de totale populatie
- De populatie ontwikkelt zich elk jaar; er ontstaan nieuwe manieren van fraude, de vulling van de bronsystemen verandert door evolutie in de werkprocessen etc. Hierdoor zal een model gemaakt op historische gegevens nooit volledig representatief zijn.

De effecten van deze niet-representativiteit zullen merkbaar worden bij het valideren van het model: het uitvoeren van rechtmatigheidsonderzoeken op basis van voorspellingen op de huidige populatie. Een voorbeeld van dit effect is dat de 'baseline' – het percentage onrechtmatigheid bij willekeurig kiezen – in lab-populatie 50 tot 60% is, terwijl uit eerdere aselecte steekproeven bekend is dat in de volledige populatie dit percentage tussen de 10 en 25% ligt. De hitrates van de lab-modellen zijn dus zeker niet representatief voor de te verwachten resultaten in de praktijk.

Een statistische analyse van de verschillen tussen de lab- en volledige populatie laat zien dat voor een groot deel van de gebruikte features statistisch significante verschillen bestaan. Wanneer echter wordt onderzocht welke van die verschillen groot genoeg zijn om sterke invloed op de modelresultaten te hebben blijven enkele tientallen features over. De details van deze analyse zijn opgenomen als bijlage C.

Het is daarom reëel om te verwachten dat de lab-modellen ook op de echte populatie nuttige voorspellingen kunnen doen. De kwaliteit van die voorspellingen kan echter pas na echte validatie worden vastgesteld.

6.7 Volgende stappen

In de labfase heeft de focus gelegen op zo snel mogelijk bruikbare resultaten genereren om aan te tonen dat het mogelijk is een risicomodel te maken. Hierbij was niet het doel om deze resultaten ook te optimaliseren om de maximale voorspelkracht uit de data te halen. Daarvoor was geen tijd en capaciteit. Het is wel duidelijk dat er nog vele verbetermogelijkheden zijn. Enkele voorbeelden:

- Meer gestructureerde features gebruiken: momenteel wordt slechts een deel van de beschikbare data gebruikt in de ABT. De gebouwde features zijn de meest voor de hand liggende kenmerken die uit de brongegevens te extraheren zijn.
- Expert features toevoegen (kennis van domeinexperts gebruiken als input voor features)
- Volgende stappen in tekst mining nemen: van woordentelling (bag-of-words) naar groepen woorden (n-grams) etc.

Bij het maken van een risicomodel is er altijd een afweging te maken tussen kwaliteit en snelheid van resultaat. Het is absoluut mogelijk om nog zes maanden vol door te ontwikkelen aan het model en in die tijd kan het model waarschijnlijk sterk worden verbeterd. De business-waarde wordt echter pas duidelijk bij validatie van een versie van het model. Om deze reden is na de labfase besloten te starten met een pilotfase waarin zowel het model is doorontwikkeld als een praktijkvalidatie is uitgevoerd door daardwerkelijke rechtmatigheidsonderzoeken uit te voeren. In de volgende hoofdstukken zijn de aanpak en resultaten daarvan te vinden.

7 Aanpak en activiteiten pilotfase en voorbereiding uitrol

7.1 Ontwikkeling pilot-model

Na de veelbelovende resultaten van de labfase is in een tweede iteratie een nieuw model ontwikkeld. In essentie zijn de gebruikte technieken daarbij dezelfde als voor het eerste model, maar op verschillende vlakken zijn verbeteringen doorgevoerd:

- Actuele data zijn ingeladen vanuit het datawarehouse van W&I;
- Uit delen van de data die nog niet werden gebruikt zijn aanvullende features gebouwd om het model te verrijken, en de feature selectie is aangescherpt en verbeterd. Zo is meer werk besteed aan het verminderen van multicolineariteit en near zero variance in de features om te voorkomen dat onnodig veel features worden meegenomen.
- De scope (de selectie van historische onderzoeken waarmee het model getraind wordt) is aangepast aan het beoogde gebruik; zo zijn detentiesignalen dit keer niet meegenomen omdat deze al zeer betrouwbaar worden opgevolgd.
- De scripts zijn verbeterd, beter leesbaar en onderhoudbaar gemaakt.

Deze stappen hebben geleid tot een '1.0' versie van het model dat klaar was om in de praktijk te testen.

7.2 Onderzoeksopzet pilot

De essentie van een pilot is om het ontwikkelde product op kleine schaal in de praktijk te testen om de werking te valideren en verbeterpunten te vinden. Omdat het gebouwde risicomodel bedoeld is om gevallen van onrechtmatigheid betrouwbaarder te voorspellen, is deze test uitgevoerd door daadwerkelijke rechtmatigheidsonderzoeken uit te voeren naar aanleiding van de voorspellingen van het model. Om een zo betrouwbaar resultaat te bereiken is daarbij de volgende opzet gekozen:

- Met behulp van het model zijn twee groepen uitkeringsgerechtigden geselecteerd:
  - Een groep met hoog risico op onrechtmatigheid. Deze groep bestaat uit een willekeurige selectie uit de duizend personen met de hoogste risicoscores. (top-groep)
  - Een controlegroep. Deze groep is aselect gekozen uit alle uitkeringsgerechtigden. (random-groep)

- Van beide groepen is de rechtmatigheid van de uitkeringen onderzocht op dezelfde manier waarop alle rechtmatigheidsonderzoeken van de laatste jaren zijn uitgevoerd: door het project heronderzoeken (PHO). Op deze manier wordt invloed van de onderzoeksmethode op de resultaten geminimaliseerd. Een schematische weergave van het onderzoeksproces is in figuur x te vinden.

- De resultaten van de onderzoeken zijn gemeten door in elke groep bij te houden hoeveel uitkeringen werden beëindigd, hoeveel terugvorderingen er werden gedaan en hoeveel ongewijzigd zijn voortgezet. De pilot is geslaagd als de door het model als hoog-risico aangewezen top-groep tot een hoger percentage 'hits' (beëindigingen of terugvorderingen) leidt dan de aselecte controlegroep.

Voor de betrouwbaarheid en de statistische significantie van de resultaten is een zo groot mogelijk aantal onderzoeken aan te bevelen. Het projectteam adviseerde daarom minimaal 300 onderzoeken uit te voeren, verdeeld over beide groepen. Vanwege beperkingen in de capaciteit en vanwege de planning heeft de begeleidingsgroep besloten de pilot te beperken tot 250 onderzoeken, waarvan 150 in de top-groep en 100 in de random-groep.

Begin maart is een selectie in twee groepen door het projectteam overgedragen aan Toetsing en Toezicht, waarna enkele controles zijn uitgevoerd naar of de geselecteerde personen voor een onderzoek in aanmerking kwamen. Uit de selectie van het projectteam zijn zo de top- en randomgroep van respectievelijk 150 en 100 personen samengesteld. Deze groepen zijn half maart gemengd overgedragen

aan het project heronderzoeken voor het uitvoeren van de rechtmatigheidsonderzoeken. De onderzoekers wisten niet wat tot welke groep de te onderzoeken personen behoorden om ze niet te beïnvloeden.

De in totaal 250 onderzoeken zijn uitgevoerd tussen half maart en eind juni.

7.3 Herontwerp dataleveringen

Bij het ontwikkelen van het labfase-model en daarna het pilotmodel werden steeds meer de tekortkomingen van de aangeleverde data zichtbaar. Deze werden geleverd via meerdere handmatige tussenstappen uit het datawarehouse van W&I, en het projectteam ontdekte verschillende problemen. Zo bleken achtereenvolgende leveringen niet consistent (de tabellen hadden soms een andere vorm gekregen of data die in vorige leveringen zat ontbrak in nieuwe leveringen), en rezen vraagtekens bij de meegeleverde historie. Van de zogenoemde 'kubussen' waarin data over uitkeringen en fraude werden geleverd, bleken de totstandkoming en ontwerpkeuzes onbekend en ongedocumenteerd, waardoor onduidelijk was of deze gegevens betrouwbaar waren en hoe ze precies moesten worden geïnterpreteerd. Dit probleem heeft bredere implicaties dan alleen voor het risicomodel, omdat juist deze kubussen de basis zijn van de managementinformatie van W&I. Dit probleem is aangekaart in de begeleidingsgroep.

Voor het ontwikkelen van het pilotmodel was weinig tijd, zodat gewerkt moest worden met de beschikbare data. Tijdens het uitvoeren van de onderzoeken door PHO heeft het projectteam de tijd genomen om de problemen met de data verder te onderzoeken, en samen met het datawarehouse-team van W&I gewerkt aan verbeteringen.

Gezien de aard en ernst van de problemen heeft het datawarehouse-team besloten de gehele levering te herontwerpen en voortaan onder te brengen in het concern datawarehouse, als stap op weg naar het uitfaseren van het W&I-datawarehouse. Hiertoe is een Cognosontwikkelaar samen met de oorspronkelijke dataleverancier secuur alle data nagegaan om de gegevens juist in het concern datawarehouse te krijgen. Hierbij zijn ook de handmatige stappen en Access-stappen verwijderd, wat de consistentie heeft vergroot. Bovendien is uitgebreidere informatie over de historie toegevoegd zodat situaties uit het verleden betrouwbaarder kunnen worden gereconstrueerd. In paragraaf 11.2 zijn de geleerde lessen over dit onderwerp verder beschreven.

7.4 Bestendigen van model en code, raamwerk toekomstige modellen

Al bovenstaande verbeteringen in de datalevering konden nog niet worden verwerkt in het pilot-model, maar zijn de basis van een derde iteratie (een '2.0'-versie), die na het verwerken van de pilotresultaten is begonnen en aan het eind van het project is opgeleverd.

In deze versie is bovendien veel werk verzet om het model gemakkelijk te kunnen beheren; de code is verder verbeterd door het toevoegen van foutcontroles en betere structurering. Dit heeft onder andere als resultaat gehad dat het opgeleverde model ook als framework kan dienen voor andere unsupervised machine learning modellen: de stappen na het verwerken van de data en het bouwen van de ABT zijn immers generiek; de code om verschillende modelleertechnieken toe te passen op de ABT en daaruit de beste te selecteren is grotendeels herbruikbaar.

Een supervised machine learning model kan gebruikt worden om een uitkomst te voorspellen, in het geval van dit project is deze uitkomst onrechtmatigheid. W&I wil weten welke uitkeringen onrechtmatig zijn voordat het onrechtmatigheidsonderzoek is afgenomen en heeft veel historische voorbeelden van eerdere onrechtmatigheidsonderzoeken. Deze werkwijze kan toegepast worden op elke situatie waar je een uitkomst wilt weten in de toekomst en waar er veel historische voorbeelden zijn van deze uitkomst. Een aantal hypothetische voorbeelden zijn; vervangingsmoment infrastructuur, toekennen van bezwaren tegen boetes, uitstroom uit de uitkering naar werk.

8 Resultaten en conclusies pilotfase

8.1 Hitrates: het model werkt

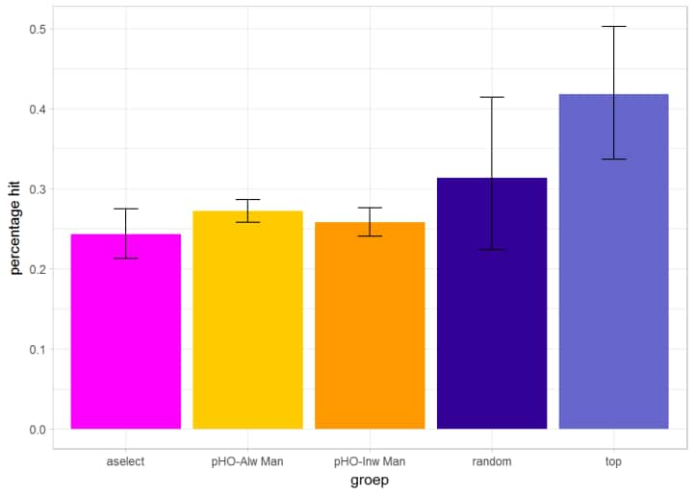
De pilot laat zien dat rechtmatigheidsonderzoeken bij werkzoekenden die hoog scoren in het ontwikkelde risicomodel (de "top-groep") in circa 40% van de gevallen tot een 'hit' (beëindiging of terugvordering van de uitkering) leiden tegenover 30% in de controlegroep (chi-square(1)= 2.33, p = 0.13). Ruim 22% van de 151 onderzochte uitkeringen van de top-groep werden beëindigd tegenover slechts 9% in de controlegroep. Daarmee is het aantal beëindigingen significant hoger (chi-square(1) = 7.27**, p <.01) voor de werkzoekenden met een hoge risicoscore dan bij willekeurig geselecteerde werkzoekenden, dit maakt het aannemelijk dat dit verschil niet op toeval berust maar een effect is van het ontwikkelde model. De totale hit-ratio lijken ook hoger te liggen dan andere eerder gebruikte selectiemethoden zoals woonprofielen, maar door de kleine aantallen is deze uitspraak nog niet statistisch te verifiëren.

Onderstaande tabel zet de resultaten af tegen eerdere onderzoeksgroepen van PHO:

	Onderzocht #	**# einde**	**# TV**	**% einde**	**% TV**	**gemiddelde** terugvordering	**% hit**
Selectie							
pilot top	151	34	36	22,5	23,8	2.576,39	40,4
Pilot random	101	9	25	8,9	24,8	3.187,14	30,7
1.000 op Zuid	912	281	88	30,8	9,6	1.978,87	31,9
pHO-AIO	32	18	1	56,3	3,1	1.442,03	59,4
pHO-Alw Man	3.957	742	430	18,8	10,9	3.548,13	27,2
pHO-Inw Man	2.401	565	191	23,5	8,0	1.308,07	25,8
pHO-Inw Vrouw	1.013	207	62	20,4	6,1	1.740,66	25,0
pHO-Matching	139	73	13	52,5	9,4	3.225,27	56,1
pHO-MO	16	11	1	68,8	6,3	12.402,84	75,0
pilot Alw Man	107	19	10	17,8	9,3	1.016,93	20,6
pilot aselect	771	97	100	12,6	13,0	2.306,79	24,3
pilot Inw Vrouw	100	31	6	31,0	6,0	1.857,46	32,0

Te zien is dat het percentage beeïndigingen van de topgroep duidelijk hoger is dan dat van de random-groep en vergelijkbaar met een aantal van de eerdere onderzoeksselecties zoals inwonende mannen en vrouwen, en alleenwonende mannen. De groepen met flink hogere beëindigingspercentages zoals AIO, Matching en MO betreffen kleine zeer specifieke groepen. Ook valt op dat het aantal terugvorderingen in zowel de top- als de randomgroep erg hoog is vergeleken met eerdere selecties. Mogelijk heeft dit te maken met het hoge percentage kostendelers en/of veranderingen in de keuzes van de onderzoekers.

Door het relatief kleine aantal onderzoeken is de onzekerheidsmarge in de hitrates bij de pilotgroepen vrij groot. Deze onzekerheidsmarges zijn in onderstaande figuur opgenomen.



Figuur 3 – hitrates (percentage beëindiging en/of terugvordering) van belangrijkste onderzoeksgroepen PHO met rechts de twee pilotgroepen.

Conclusie op basis van deze cijfers is dus dat de pilot geslaagd is: personen die door het model een hoge risicoscore krijgen toegewezen blijkt vaker onrechtmatig een uitkering hebben dan de controlegroep.

8.2 Selectie: veel afvallers door actualiteit en positieve uitstroom

Zoals in het vorige hoofdstuk werd vermeld zijn de top- en randomgroep samengesteld in twee stappen: het project heeft twee lijsten met (gepseudonimiseerde) BSNs opgeleverd met meer dan het uit te nodigen aantal personen. . Bij Toetsing en Toezicht zijn uit die lijsten vervolgens de te onderzoeken personen gekozen door personen te schrappen waarbij een onderzoek naar het oordeel van T&T niet mogelijk/wenselijk was.. In deze tweede stap zijn opvallend veel personen afgevallen: 242 van de 700. Het grootste deel van deze afvallers viel af doordat de werkzoekende inmiddels geen uitkering meer had (85) of de uitkering was opgeschort (85). Een andere grote groep was al recent onderzocht (31).

Een deel van de verklaring hiervoor ligt in een actualiteitsprobleem met de data die werd gebruikt voor het risicomodel; door een fout in de datalevering waren de uitkeringsgegevens een half jaar oud op het moment van selecteren, en waren er dus in de tussentijd mensen uitgestroomd. Dit kon het effect echter niet geheel verklaren, en nader onderzoek wees uit dat bovengemiddeld veel mensen met hoge risicoscore's lijken te zijn uitgestroomd naar werk of studie (topgroep = 8%, randomgroep = 3%, chi-square(1) = 6.17*, p <.05). Mogelijk voorspelt het model onbedoeld niet alleen onrechtmatigheid, maar ook *positieve* uitstroom.

Dit heeft als nadeel dat deze mensen onnodig een onrechtmatigheidsonderzoek krijgen zij zullen uitstromen zonder tussenkomst van een fraudeonderzoek. Daarnaast zijn dit niet de mensen die je negatief wilt benaderen, maar juist wilt stimuleren om uit te stromen naar een baan. Verder kan het targetten van juist deze succesverhalen de draagkracht voor het model bij de werkconsulenten verkleinen. In een doorontwikkelfase kan dit probleem aangepakt worden. Door bijvoorbeeld uitstroom expliciet mee te nemen in het model. Het model zou dan twee scores op kunnen leveren: een voor onrechtmatigheid en een voor de kans op positieve uitstroom. Deze laatste score biedt dan juist weer kansen in de vorm van een voorspelling dat iemand waarschijnlijk positief kan uitstromen en wellicht nog een laatste zetje kan gebruiken.

[redacted]

Artikel 5.2 lid 1 Woo

8.3 Eigenschappen van de risicogroep

De pilot laat zien dat in de top-groep vaker onrechtmatigheid voorkomt dan bij aselect gekozen personen. Het is interessant om te onderzoeken welke verschillen er verder zijn tussen de top- en random groep. Onderstaande tabel geeft daar een eerste indruk van door de waardes van een aantal kenmerken te vergelijken tussen verschillende groepen.

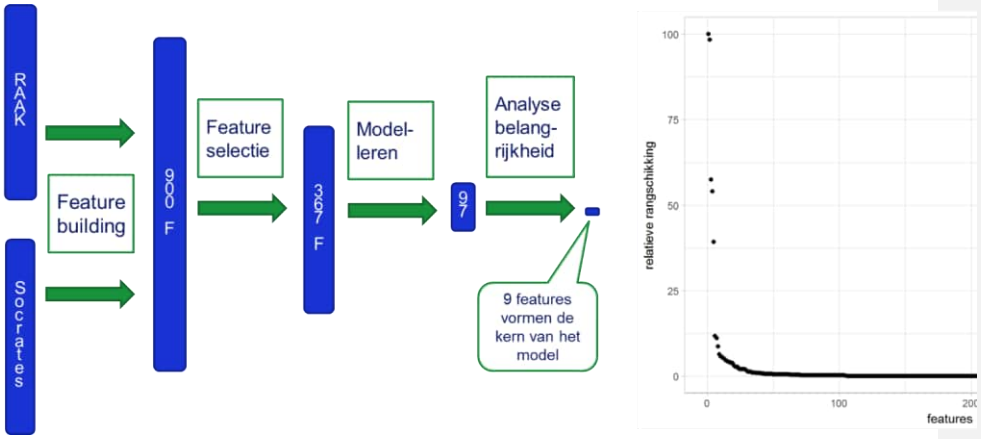
Hieruit is op te maken dat de personen in de topgroep ruim tien jaar jonger zijn dan gemiddeld, wat vaker man en relatief kort een uitkering hebben. Opvallend is dat een groot deel kostendeler is en dat alle personen in de top-groep eerder waren onderzocht.

Het is belangrijk om te realiseren dat deze statistieken beschrijvend zijn en niet verklarend of voorspellend en achteraf zijn gemaakt. Dit betekent dat de onderstaande tabel niet gelezen kan worden als kenmerken van mensen die vaker frauderen maar om een beeld te geven van welke groep is onderzocht. Inzichten zoals deze kunnen gebruiken worden om de hitrate te interpreteren, werkproces aan te passen en te controleren of het model een wenselijke groep uitwerpt.

Groep Feature	Top	random	Aselect 2015	Hele populatie	Woon-profielen PHO	PHO overig	Trainset
aantal_onderzoeken	151	101	759	36764	4745	3492	25926
Per_hit	40,4%	30,7%	24,3%	-	20 – 27 %		50%
gem_leeftijd	34.7	47.4	43.0	46.2	44.0	36.4	39.9
percentage_vrouw	43.7	58.4	66.7	53.4	18.5	8.10	48.2
gem_duur_uitkering	1036	3200	2687	2670	2109	1131	1436
aantal_bewindvoerders	4	4	0	843	0	0	0
per_bewindvoerders	2.64%	3.96%	0%	2.29%	0%	0%	0%
aantal_kostendelers	85	16	52	3908	19	9	515
per_kostendelers	56.2%	15.8%	6.85%	10.63%	0.40%	0.25%	1.98%
per_mensen_onderzocht	100%	30.6%	28.9%	41.4%	34.4%	31.8%	20.9%
per_mensen_onderzocht_negatief	88.1%	25.7%	22.1%	32.4%	25.6%	23.2%	19.0%

8.4 Verbanden tussen belangrijke features en risicogroep

Zoals in hoofdstuk 5 werd beschreven kiest een modelleeralgoritme uit alle honderden features zelf welke features in welke mate gebruikt worden voor de voorspelling. In het best presterende model bleken uiteindelijk negen features dominant in de voorspelling. Dat betekent niet dat in de andere features geen voorspelkracht zit; er zijn vaak vele andere manieren om tot een vergelijkbaar goede voorspelling te komen.



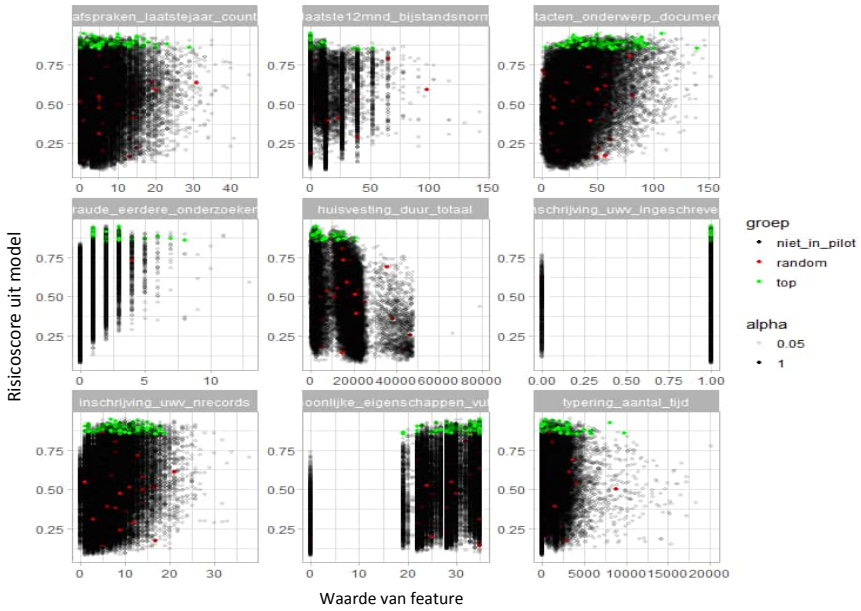
Figuur 4. *Data komt uit RAAK en socrates en maken daarna verschillende stappen door. Het model finale model heeft van 367 features kunnen leren en heeft 97 features gebruikt waarvan 9 feature de kern vormen. Tijdens de feature selectie worden andere algoritmes gebruikt om tot de juiste selectie van features te komen. Figuur 4 rechts: Op de y-as staat relatieve belangrijkheid van de feature in het model. Het is te zien dat de belangrijkheid zeer snel afloopt.*

Opvallend is dat de belangrijkste features (hieronder weergegeven) vrijwel allemaal te maken hebben met het registratieproces van de gemeentelijke organisatie: het zijn features die activiteit rondom de werkzoekende in de gemeentelijke organisatie en interactie tussen de werkzoekende en de gemeente weergeven, in plaats van echte kenmerken van de werkzoekende zelf. Dat is te interpreteren als indicatie dat het model erin slaagt de kennis die zit verborgen in alle losse interacties en registraties te lezen en het 'systeembewustzijn' om te zetten in een risicoscore. Een individuele registratie van een verstuurde brief of een afgezegde afspraak bevat te weinig informatie om een conclusie uit te trekken, maar het model vindt patronen in duizenden afzonderlijke registraties van activiteit en gebruikt die om tot een sortering te komen. Het model zoekt naar de kortste manier op te voorspellen. Dit zegt niks over de verklarende waarde van deze features.

Rang	feature	top	random	scoring
1	persoonlijke_eigenschappen_vulling	31.4	26.55	26.75
2	inschrijving_uwv_nrecords	9.32	6.44	6.32
3	afspraken_laatstejaar_count	6.57	4.47	4.86
4	component_aantal_laatste12mnd_bijstandsnorm_alleenstaande_21	1.53	8.34	8.58
5	fraude_eerdere_onderzoeken	1.65	0.35	0.51
6	inschrijving_uwv_ingeschreven	1	0.98	0.98
7	huisvesting_duur_totaal	4804.44	11176.11	9529.33
8	typering_aantal_tijd	1429.34	894.77	857.97
9	contacten_onderwerp_document__	51.19	27.46	27.77

In bovenstaande tabel worden de negen belangrijkste features getoond met daarachter de gemiddelde waarde van die feature in drie verschillende groepen. Dit laat zien dat hoewel de waardes van de features

soms flink verschillen tussen groepen, het verband niet eenvoudig is te omschrijven in de vorm van bijvoorbeeld “personen die meer dan 5 afspraken gehad hebben in het laatste jaar hebben een hoger risico”; de verbanden zijn complex en niet lineair. Dat is nog beter te zien in Figuur 5, die visueel weergeeft wat de verbanden tussen deze negen features en de risicocores zijn.



Figuur 5 - verband met risicoscore van de 9 belangrijkste features: elk puntje in de grafieken stelt een werkzoekende voor met op de horizontale as de waarde van de feature, op de verticale as de risicoscore. De groene en rode puntjes zijn de geselecteerde personen voor respectievelijk de top- en de random-groep, de zwarte puntjes zijn wel gescoord maar niet onderzocht in de pilot.

Deze figuur geeft een beeld van de complexiteit van de werking van het model en maakt inzichtelijk waarom de voorspellingen niet met een simpel verband te reproduceren zijn. Een risicomodel vindt hoogcomplexe, niet-lineaire voorwaardelijke verbanden.

9 Aanpak en activiteiten uitrol fase

9.1 Ontwikkeling finale model

Na de succesvolle pilot is in een finale iteratie gedaan en is een nieuw model ontwikkeld. Het doel van de uitrol fase is om een end-to-end oplossing af te leveren dat in de lijnorganisatie opgenomen kan worden. N.a.v. de pilot zijn model-technische wijzigingen doorgevoerd en is ook het proces en de daarbij behorende verantwoordelijken vastgesteld. Nieuwe validatie data is gebruikt om van het verbeterde model een indicatie te geven van de verwachte prestatie in de praktijk. De gebruikte modelleer-technieken zijn niet radicaal veranderd, maar op enkele vlakken zijn incrementele verbeteringen doorgevoerd:

Data en model:

- Actuele data zijn ingeladen vanuit het datawarehouse van W&I;
- Een nieuwe validatie set is gemaakt, op basis van het aselecte onderzoek in het NPRZ gebied.
- De scripts zijn herschreven zodat ze met minimale handmatige input gedraaid kunnen worden en de output overzichtelijk wordt weggeschreven.

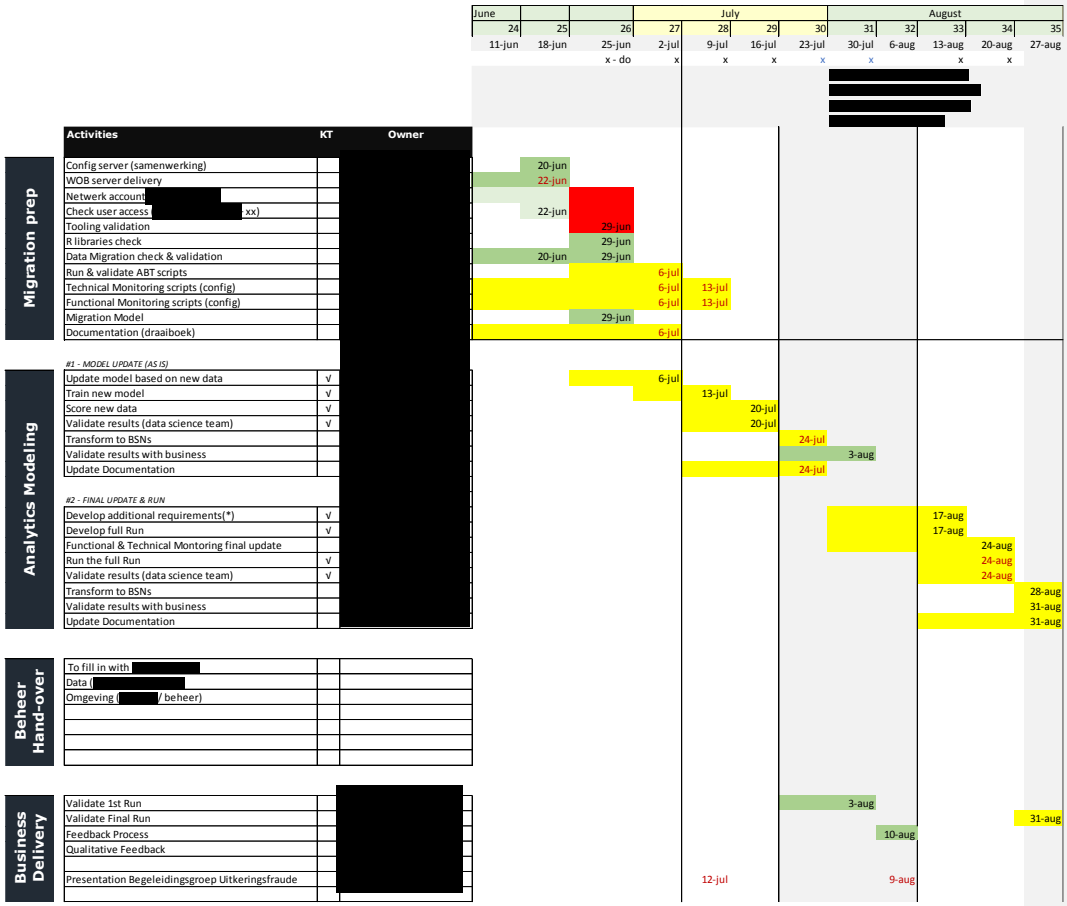
Business proces:

- Alle rollen in het end-to-end proces zijn gedefinieerd, en personen/afdelingen zijn daar aan toegeschreven.
- Het formaat, frequentie en wijze waarop de voorspelde lijst wordt geleverd is vastgesteld.
- De impact van de cut-points van de business rules is bestudeerd a.d.h.v. de data.
- Begeleidende documentatie is geschreven, dan wel geüpdatet.

Deze stappen hebben geleid tot een end-to-end oplossing die klaar is om in de lijnorganisatie op te worden genomen.

9.2 Planning

In de uitrol fase zijn we begonnen met de migratie van de data en code naar de nieuwe (WOB) server. Daar is getest of alles nog werkt. Vervolgens is een nieuwe data extractie gedaan (inclusies nieuwe validatie data) waarop de nieuwe modellen worden ontwikkeld. De resulterende lijst is gevalideerd bij W&I en feedback is meegenomen voor verdere verbetering van het model. Ten slotte zijn de scripts herschreven en is de final run neergezet, om vervolgens dé scoring lijst op te leveren. Naast het data science werk is ook met business aan het operationele model gewerkt, zodat deze oplossing werkbaar blijft na afsluiting van de projectfase. De uitkomst hiervan wordt toegelicht in het business document(.pptx).



9.3 Data

9.3.1 Actualiteit

Door een fout in de datalevering tijdens de pilotfase waren de uitkeringsgegevens een half jaar oud op het moment van selecteren, en waren er dus in de tussentijd mensen uitgestroomd. De fout in het extraheren van de data en het creëren van de ABT is gecorrigeerd. De data is maximaal één dag oud op moment van extractie.

Uiteraard duurt het enige tijd voordat de lijst W&I / THO bereikt. Daarom wordt de lijst ook nog op actualiteit gecontroleerd vlak voordat hij naar team Heronderzoeken (tHO) gestuurd wordt.

9.3.2 Tardis

In de ABT wil je alleen data hebben die bekend was voordat iemand onderzocht werd, en niet de data die resulteert uit een dergelijk onderzoek. De meest recente data meenemen is dus niet altijd wenselijk. De recentere data werd voorheen inactief gemaakt, maar nu is een zogenaamde Tardis gemaakt. Dit script

haalt de data op zoals die vlak vóór het onderzoek bekend was. Het dient dus als een soort tijdsmachine. In bijlage D wordt de Tardis in detail besproken.

9.3.3 Nieuwe validatie set voor model

In de pilotfase is een relatief kleine validatie set gebruikt. Deze set is in deze uitrolfase in de traindata toegevoegd, omdat er op dit moment een nieuw – grootschaliger – aselect onderzoek bezig is: Aselect NPRZ. Dit bestaat uit 3000 onderzoeken waarvan er nu ongeveer 1000 zijn afgerond, met een hitrate van 16%. 500 observaties zijn voor de validatie set gebruikt en de overige zijn in de traindata opgenomen. Het aselecte onderzoek wordt als benchmark gebruikt om iets te kunnen zeggen over de verwachtte prestatie van het model in de praktijk.

9.4 Model

De modellen en resultaten werden voorheen weggeschreven in aparte mappen en in losse files. Om dit overzichtelijker te maken worden ze nu centraal opgeslagen in tibbles. Hierin staan alle model versies én hun performance metrics in één file opgeslagen. De model evaluatie wordt automatisch gegenereerd en het notebook wordt opgeslagen als een html bestand.

De modellen worden getraind op drie performance metrics: 1) accuracy, 2) Area under ROC, 3) Kappa. Daarnaast worden de volgende vier methodes toegepast: 1) Random Forest, 2) Gradient Boosting Models, 3) XGBoost, 4) Decision Tree. Dit resulteert in twaalf initiële modellen, waar verder mee getraind wordt door met de tuning parameters te spelen (b.v. de grid size en het aantal iteraties).

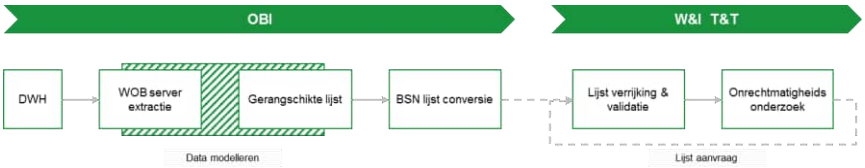
Verder is er afgesproken dat de business rules voor de doelgroep scope ná het genereren van de lijst worden toegepast. Op die manier kunnen we inzicht geven in welke (en waarom) personen worden weg gefilterd.

9.5 Business proces

Er is overeengekomen dat het end-to-end proces uit twee zelfstandige processen bestaat. Dit is gedaan zodat de OBI en W&I betrokkenen hun eigen planning kunnen aanhouden. OBI pusht vier keer per jaar een scoringslijst naar de server, waarop W&I de lijst kan ophalen en de actualiteit kan verversen wanneer tHO daar om vraagt. Er zal uiteraard wel communicatie tussen OBI en W&I zijn, omdat het data science proces afhankelijk is van kennis uit de business en vice versa.

De output vanuit OBI naar W&I is een lijst met gescoorde uitkeringsontvangers. Deze is gesorteerd van hoog naar laag risico. De kolommen zijn als volgt: volgnummer, ID, risicoscore. De risicoscore is puur ter documentatie. De scores moeten beschikbaar zijn voor wanneer er dieper inzicht in het uitkeringsfraude project moet worden gegeven. Deze scores zijn moeilijk te interpreteren en dienen daarom niet voor verdere analyse of communicatie te worden gebruikt!

In dit project is er voor gekozen om inhoudelijke keuzes zo veel mogelijk bij W&I te houden. Op die manier is OBI ‘slechts’ de facilitator van het data science proces. Er is uiteraard continue samenwerking, omdat dat aan de basis staat van een relevante en bruikbare data science output. De business rules die in het modelleer proces worden gebruikt, worden bijvoorbeeld door W&I bepaald. OBI kan op basis van de data suggesties doen, maar is vooral de partij die de rules toepast op de data. De uiteindelijke eigenaar van de ontwikkelde oplossing is W&I. Een versimpelde illustratie van het end to end proces is weergegeven in het figuur hieronder.



Ten slotte zijn er verschillende documenten geschreven om het kennis en inzichten overdraagbaar te maken. Er zijn zowel algemene als data science specifieke documenten geschreven, en zowel mede eenmalige als doorlopende documenten.

EENMALIG

BUSINESS OVERVIEW iedereen

Geeft overzicht in de wat, hoe, en wie omtrent dit product. Hierin wordt gespecificeerd waar de verantwoordelijkheden liggen en wat de mogelijkheden en beperkingen van het model zijn.

DRAAIBOEK data scientist

Stap voor stap handleiding om 1) een nieuwe lijst te genereren, 2) modellen te hertrainen.

PROJECT DOCUMENT project team

Dit document. Proces document dat het project van begin tot eind beschrijft. Verklaart gemaakte keuzes en stappen.

(Incl: privacy impact assesment)

PERIODIEK

LEGAL DOCUMENT product owner, data scientist

Geeft inzicht in welke data types en bronnen zijn gebruikt. Wordt bij elke lijst gegenereerd – en meegestuurd naar W&I – op basis van de gebruikte data voor specifiek die versie van de gegenereerde lijst.

MODEL BESCHRIJVING data scientist

Technische uitleg van het voorspellende model. Wordt geupdated na hertrainen en beschrijft dus altijd de meest recente versie van het model.

(Incl: technical decisions document)

10 Aanbevelingen doorontwikkeling en vervolgonderzoek

Met de modelversie die bij het afronden van het project is opgeleverd is een toepasbaar en volwassen product beschikbaar gekomen dat op grotere schaal kan worden toegepast door W&I. Dat betekent echter allerminst dat het model is uitontwikkeld en dat al het onderste uit de kan is gehaald. Op twee vlakken zijn nog heel veel kansen ontstaan: verdere verbetering van het model teneinde het nog effectiever te maken (hogere hitrates) en nader onderzoek ter beantwoording van kennisvragen naar aanleiding van inzichten uit het project.

10.1 Aanknopingspunten voor vervolgonderzoek

In de voorgaande paragrafen zijn al een aantal interessante observaties genoemd die nader onderzoek verdienen. In het algemeen is duidelijk geworden dat hoewel het bouwen van een risicomodel niet als primaire doel heeft om kennis op te doen en verklaringen te geven voor effecten, er als 'bijvangst' wel degelijk erg veel kennis vergaard wordt en vooral veel aanknopingspunten voor nader onderzoek uit voortkomen.

Hieronder is samengevat welke onderwerpen en welke onderzoeksvragen volgens het projectteam het belangrijkst zijn. Er zijn een aantal initiëlee analyses gedaan door het team die apart beschikbaar zijn. In de beperkte beschikbare capaciteit en tijd, en om de scope van het project te bewaken, was het helaas niet mogelijk deze vragen nog verder uit te diepen.Het team beveelt dan ook met klem aan om deze vragen verder te onderzoeken.

- **Meer contact (brieven, bellen, afspraken) leidt tot hogere risicoscore**
    - Wat zijn dit voor mensen en contacten? Hypothese: probleemgevallen waar lang omheen gedraaid wordt?
    - Hebben deze mensen een hoger risico om te frauderen? Of is er slechts een hoger risico dat zij betrapt worden?
- **Mensen met eerdere (negatieve) 'onderzoeken' scoren heel hoog**
    - Hoe is recidivisme te voorkomen?
    - Is er onderscheid tussen UBO – PHO – BI? Hypothese: onderzoeken worden vaak niet afgerond bij gebrek aan aanknopingspunten, maar wel degelijk onrechtmatigheid.
    - Zijn er betere onderzoeksmethoden te vinden die 'hits' beter opsporen?
    - Willen we ook in een andere vijver gaan vissen?
- **Veel positieve uitstroom (naar werk/studie) onder hoge risicoscores**
    - Kan het zijn dat model als bijvangst heeft mensen die op het punt staan uit te stromen? Hypothese: veel contact voorspelt zowel positieve als negatieve uitstroom, maar model heeft alleen van negatief geleerd.
    - Kan dit verband ten goede worden ingezet?
- **Overig: hoge aantal terugvorderingen, tragere behandeling, vulling data in proces etc etc**

10.2 Doorontwikkelmogelijkheden voor het risicomodel

Tijdens het maken van het model zijn we verschillende mogelijkheden tegenkomen om het model verder te ontwikkelen. Met het model verder ontwikkelen bedoelen we de risicoscores accurater of specifieker maken. Voor een deel zijn deze gerelateerd aan de bovengenoemde onderwerpen. Een aantal van deze mogelijkheden zijn voorgesteld aan Toetsing en Toezicht. Er was toen wel interesse in deze mogelijkheden maar het was niet het moment om deze op te pakken.

Performance metrics focussen op meer dan we geinteresseerd in zijn

De modellen zijn tot nu toe getraind op verschillende metrics die zowel true positives en true negatives over de gehele populatie in acht nemen. Echter, wij zijn slechts geïnteresseerd om in de top X personen zoveel mogelijk true positives te krijgen. Het maakt daarbij niet uit of we de niet-fraudegevallen goed kunnen voorspellen. Voor verdere ontwikkeling is het daarom interessant om een eigen performance metric te schrijven aan de hand waarvan tijdens het trainen louter op de fraudegevallen in de top X wordt gefocussed.

Werkzoekende stroomt vaak uit naar arbeid

We zien dat er relatief veel mensen uitstromen naar arbeid binnen de pilotgroep met een hoog risico.

Dit heeft 3 belangrijke nadelen bij gebruik van het risicomodel voor selectie van een onderzoeksgroep (naast de hierboven genoemde kansen tot kennisontwikkeling die het ook biedt):

- Onderzoek naar deze mensen is niet nodig want deze zullen uitstromen naar werk. Er is geen tussenkomst van een fraude onderzoek nodig om deze mensen uit de uitkering te laten gaan. Dus ook geen reden deze mensen uit te werpen met een hoog risico score
- Dit zijn niet de mensen die je lastig wilt vallen met een negatief onderzoek. Mensen die op het punt staan uit te stromen naar werk wil je juist stimuleren dit te doen.
- Dit verkleint het draagvlak voor pHO, het kan zijn dat werkconsulenten het niet goed vinden als hun beste klanten opgeroepen worden door pHO.

Er zijn concrete ideeën om dit op te lossen in het model, bijvoorbeeld door ook positieve uitstroom te voorspellen met hetzelfde mode (zie vorige paragraaf). Voordeel van deze oplossing is dat er een model gemaakt kan worden voor uitstroom in plaats van onrechtmatigheid. Dit model zou dan een voorspelling geven voor negatieve uitstroom (die moet opgepakt worden door pho) en positieve uitstroom (dit kan opgepakt worden door werkconsulenten, in de toekomst misschien).

Een ander idee om dit op te pakken is door een tweede model te bouwen dat positieve uitstroom voorspelt. Dit model kan naast het onrechtmatigheidsmodel gebruikt worden om vast te stellen of iemand en een hoge kans op onrechtmatigheid heeft en een hoge kans om uit te stromen naar arbeid.

Onderzoek naar beide zal nodig zijn om te kijken welke oplossing als beste past en of het mogelijk is dit op te nemen in het model.

Mensen zijn al vaker eerder onderzocht

We zien in de pilotgroep dat iedereen die in de topgroep zit al een keer eerder onderzocht is (dat wil zeggen, eerder voorkwamen in de fraudemodule; dat kan variëren van slechts een melding tot een echt onderzoek) terwijl niet de gehele populatie eerder is onderzocht. Het lijkt er echter wel op dat mensen die eerder zijn onderzocht een hogere kans hebben op onrechtmatigheid ongeacht de uitkomst van het onderzoek.

Het blijven aanleveren van mensen die eerder zijn onderzocht, en daarmee op termijn het aanleveren van steeds dezelfde mensen heeft een aantal belangrijke nadelen:

- Je blijft veel aandacht geven aan de mensen die al een keer eerder verdacht zijn gevonden. De werkzoekende komt moeilijk van het stempel af;
- Het wordt moeilijker nieuwe onrechtmatigheid te vinden. Groepen die nu buiten beeld zijn, blijven buiten beeld;
- Er kan een zichzelf versterkende loop in het model ontstaan.

Een idee om dit op te lossen is door twee modellen te maken. Een voor veelplegers en een mensen die nog nooit fraude hebben gepleegd. Aandacht van onrechtmatigheidsonderzoeken kan beter verdeeld worden over alle werkzoekende en de kans is kleiner dat iemand buiten beeld blijft, maar zo hoef je niet aselect te selecteren en als het model werkt, heeft het ook een hogere hitrate dan aselect.

Met 2 modellen kunnen er sneller nieuwe groepen aan het licht komen dan met het huidige risicomodel en hierdoor kan de algemene hitrate stijgen.

Het moet natuurlijk eerst onderzocht worden of er twee verschillende modellen gemaakt kunnen worden. Een probleem zou kunnen zijn dat veelplegers en nieuwe onrechtmatigheid er in de data hetzelfde uit zien en dat deze moeilijk uit elkaar te trekken zijn.

Specifiek advies welke methode (UBO, PHO of BI) toegepast moet worden

Momenteel wordt er alleen een risicoscore voor onrechtmatigheid berekend. Het zou onderzocht kunnen worden of dit verder uitgebreid kan worden.

Bijvoorbeeld kan er voorspeld worden welke onderzoeksmethode (UBO onderzoek, PHO onderzoek of BI onderzoek) er het beste toegepast kan worden op welke persoon. Dit kan ingezet worden door alleen de zaken met een grote pakkans door pho uit te werpen of als het process daar klaar voor is, dat een zaak van een bepaalde werkzoekende naar aanleiding van de situatie van die werkzoekende toegewezen wordt aan een bepaalde afdeling en dat er tijdens het onderzoek de zaak doorgezet kan worden naar een andere afdeling.

Een ander idee is dat er een voorspelling wordt gedaan van de soort onrechtmatigheid zoals inkomnensonrechtmatigheid, woonfraude etc. Dit heeft dan dezelfde uitwerking. In plaats van een enkele score op onrechtmatigheid, geeft dit een beter idee waar ernaar gezocht moet worden en welke methode hiervoor gebruikt moet worden.

Het voordeel van een dergelijk advies is dat er passend op de situatie de juiste onderzoeksmethodes worden ingezet en hiermee de hitrate verhoogd kan worden. Bijvoorbeeld partnerfraude wordt minder makkelijk opgespoord met de pho-methode.

Hybride model

Momenteel staat het model in het werkproces van T&T naast alle andere regels en profielen die er zijn om werkzoekenden te selecteren. Dit is niet nodig. Er kan gekeken worden of het risicomodel samengevoegd kan worden met andere regels en profielen tot een hybride model. In een dergelijk model leidt een bepaalde regel of een bepaald profiel tot een bepaald risico. Dit risico wordt vervolgens afgezet tegen de risicoscores.

Met selectiemodules zijn in dit geval de huidige methodes zoals de woonprofielen, maar kunnen ook businessregels van BI zijn, of andere signalen van bijvoorbeeld andere gemeentes.

De selectiemodules en het risicomodel samen geven een bepaald risico aan. Aan de hand van dit risico wordt de capaciteit verdeeld.

10.3 Aanbevelingen rondom het selectieproces van T&T

De werkprocessen bij W&I zijn sterk afhankelijk van handwerk, losse excel-bestanden en kennis en ervaring van enkele individuen. Dit project heeft een mogelijkheid gecreëerd om een gedeelte daarvan te automatiseren en verbeteren. Voor een volgende stap ik informatiegestuurd werken zijn ook dit soort ontwikkelingen essentieel. Voor een beter reproduceerbare, objectievere en minder arbeidsintensieve werkwijze zijn de volgende verbeteringen mogenlijk:

- Objectificeren en definiëren van selectiecriteria profielen; momenteel vinden selecties plaats door middel van handmatig filteren in een excel-extractie uit cognos. Het is tot nu toe onmogelijk gebleken die criteria eenduidig te reproduceren.
- Hetzelfde geldt voor de uitsluitingscriteria; er wordt bij selecties bijvoorbeeld een regel gehanteerd dat iemand in het laatste jaar niet eerder onderzocht mag worden, maar deze regel wordt niet eenduidig gehandhaafd.

- Het vaststellen van de resultaten van onderzoeken gebeurt momenteel ook nog met veel handwerk; deels moeten de vastgelegde resultaten in het systeem handmatig worden gecontroleerd aan de hand van de daadwerkelijk verzonden beschikkingen, omdat in de registratie fouten voorkomen die een scheef beeld veroorzaken. Dit heeft tot gevolg dat resultaten van verschilllende selectiemethodes niet altijd eerlijk met elkaar vergeleken kunnen worden.

Deels hebben deze moeilijkheden te maken met de ontoereikendheid van de gegevens in het datawarehouse, waardoor interpretatie en correctie noodzakelijk blijft. Verbetering van de vastlegging is daarom ook nodig. Zolang echter handmatig wordt gecorrigeerd zal aan de oppervlakte ook de noodzaak ontbreken van deze verbeteringen, zodat het probleem zichzelf in stand houdt.

11 Geleerde lessen

Het project uitkeringsfraude is de eerste ervaring met advanced analytics binnen de gemeente Rotterdam. Het was een pionierproject, waarin niet alleen voor het eerst risicomodellen zijn ontwikkeld, maar ook ervaring wordt opgedaan met tal van randvoorwaardelijke processen. Het leereffect voor de organisatie was ook een expliciet doel van het project, en de visie van de opdrachtgevers is dat de ervaringen uit het project leiden tot vergroting van de volwassenheid van de organisatie op het gebied van Informatiegestuurd Werken.

Zo diende het project niet alleen een inhoudelijk doel voor de business: het ontwikkelen van een product dat de handhavingspraktijk verbetert, maar komt er ook waarde uit voort op vele andere terreinen. De belangrijkste lessen staan samengevat in onderstaande figuur en worden in dit hoofdstuk beschreven.

11.1 Juiste mix van omstandigheden

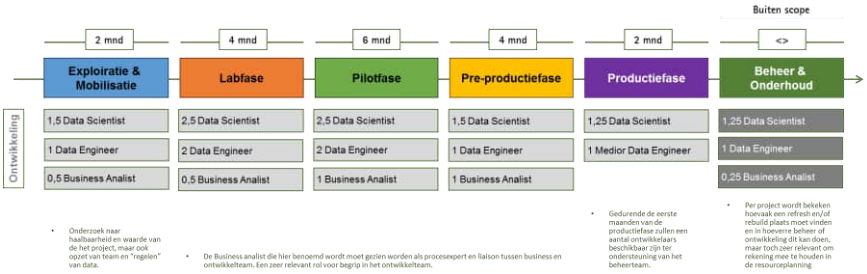
De totstandkoming van het project illustreert dat er verschillende stoplichten op groen moeten staan om een initiatief daadwerkelijk van de grond te laten komen, en dat niet altijd voorspelbaar is welke initiatieven tot succes gaan leiden. Zeker vier elementen kwamen samen die ertoe leidden dat op het gebied van uitkeringsfraude iets tot stand kwam:

a) Er was een duidelijke **vraag en opdrachtgeverschap** vanuit de business (het primaire proces). In dit geval werd deze vraag nog extra urgent door de toezegging die de gemeente Rotterdam in de G12 had gedaan om de mogelijkheden van text mining op basis van eigen gegevens te onderzoeken en voor 1-1-2017 met resultaten te komen. Dit zorgde bij betrokkenen voor de urgentie om echte stappen te zetten.
b) Er werd een **team** opgericht met capaciteit en de juiste vaardigheden om het werk uit te voeren. Er zitten honderden mens-uren in het ontwikkelen van de gewenste producten, en het is dus niet reëel om te verwachten dat resultaten tot stand komen als medewerkers het project naast hun reguliere werk moeten uitvoeren. De benodigde vaardigheden zitten op zowel het gebied van data science als op het gebied van projectleiderschap; zeker bij een pionieerproject moeten voortdurend obstakels worden opgelost en keuzes worden gemaakt om voortgang te verzekeren.
c) De **data** die nodig is voor de gewenste resultaten bestaat en is beschikbaar. Het opvragen en ontsluiten van de juiste data speelt in dit project een grote rol en neemt veel tijd in beslag, maar in elk geval is duidelijk dat er wel voldoende is. Zie ook de volgende paragraaf.
d) Het **product** wordt helder gedefinieerd en afgestemd met de business. In dit geval zou een risicomodel worden gemaakt dat uitkeringsgerechtigden rangschikt van hoge naar lage kans op

onrechtmatigheid. Een duidelijk en gedeeld beeld van het eindresultaat en het goed definiëren daarvan zorgt ervoor dat alle betrokkenen hetzelfde doel voor ogen hebben en de verwachtingen kloppen.

Vanwege deze mix van omstandigheden is het project uitkeringsfraude als eerste uit de startblokken gekomen, terwijl andere geïdentificeerde kansrijke projecten minder snel gaan.

Wat betreft het team is in het programmaplan DARE reeds beschreven welke vaardigheden en capaciteit in een typisch data science project nodig zijn. Figuur 6 geeft dit weer per fase. Daarin is te zien dat voor een volledig team in de labfase circa 5 fte nodig zijn. In dit project is gewerkt met een minimale variant; gedurende de labfase was er uiteindelijk 1,5 fte aan data-science capaciteit beschikbaar en 1 fte aan gedeeld projectleiderschap/business analist. Dit is verre van optimaal en met een volledig team waren minder concessies nodig geweest.



Figuur 6 – teamsamenstelling per fase, exclusief projectleiderschap. *Bron – Programmaplan DARE*

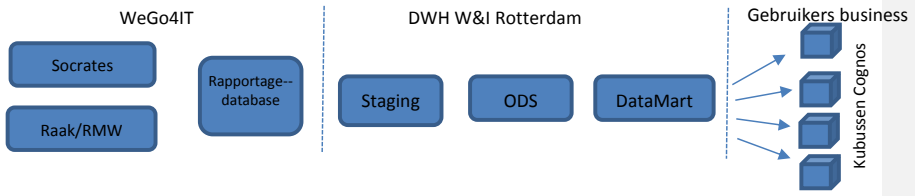
Aan het eind van het project is toepassing op grote schaal nog steeds niet gegarandeerd. Dit heeft te maken met verschillende belangen en bestuurlijke prioriteiten. Goede bestuurlijke betrokkenheid vanaf het begin is dan ook zeer aan te bevelen.

Advies: controleer bij het opdrachtgeven voor een project of aan de juiste randvoorwaarden is voldaan. Vooral voldoende capaciteit en projectleiderschap met een pro-actieve en probleemoplossend vermogen zijn essentieel. Voor succesvolle inbedding in het primaire proces is goede betrokkenheid en steun tot op bestuurlijk niveau noodzakelijk.

11.2 Data ontsluiten en gegevensmanagement

De eerste drie maanden van het project zijn voor een groot deel in beslag genomen door het opvragen en ontsluiten van data. Er was toegang nodig tot de gegevens (waaronder ongestructureerde tekst) uit RAAK. Het werd bovendien al snel duidelijk dat de Cognos-extractie (dit was een via Cognos gemaakte extractie uit de 'Kubus' fraude van het datawarehouse van W&I) waarmee werd begonnen onvoldoende data bevatte voor het bouwen van een goed risicomodel. (Zie 6.1). Het zoeken naar de juiste data en deze ontsluiten in de juiste vorm is vrijwel altijd een tijdrovend onderdeel van data-science projecten en wordt makkelijk onderschat. Een van de lessen op dit gebied is dus dat hiervoor voldoende tijd moet worden gereserveerd en capaciteit nodig is zowel in het projectteam als bij de teams die de databases beheren. Daarnaast zijn uit de projectervaringen lessen te trekken voor het inrichten van gegevensmanagement, waarvoor de gemeente net een programma is gestart.

Er is binnen de gemeente / W&I nog geen duidelijke structuur van data-eigenaarschap en -beheer. Vanaf de bronsystemen RAAK en Socrates worden vele stappen en bewerkingen op de data gedaan en er zijn verschillende ontsluitingen naast de systemen zelf, onder andere via een zogenoemde Kubus. Verschillende Gebruikers hebben daardoor verschillende perspectieven op de data wat communicatie en eenduidige datalevering niet altijd makkelijk maakt; er is vaak nog geen 'single version of the truth'.



Figuur 1 – Schematische weergave van de opwerkingsstappen van de data. Het project startte met data uit een van de kubussen, en heeft later meer onbewerkte data ontvangen uit de stap 'staging'. Dit schema is pas aan het eind van de labfase duidelijk geworden.

Het kostte het projectteam al enige tijd om de personen te lokaliseren die het beheer voerden over de benodigde data. Door versnippering van het gegevensbeheer en gebrek aan documentatie was er bovendien geen duidelijk overzicht van wat er precies beschikbaar was en wie daar toegang toe had. Dit heeft geleid tot moeilijkheden bij het specificeren van de vraag, en leidt soms tot problemen met de kwaliteit van de data of het verifiëren daarvan. In het project was bovendien behoefte aan data met historie. Hoewel alle mutaties door het systeem worden vastgelegd en de historie dus wel beschikbaar is, ontstond de indruk dat hier tot nu toe zelden naar wordt gevraagd en dat deze historie moeilijker beschikbaar te maken is. Het team had geluk met zeer competente en meewerkende databasebeheerders, waardoor het lukte om buiten de gebaande paden toegang te organiseren tot voldoende data.

Na de pilotfase is naar aanleiding van de ervaringen met vereende krachten gebouwd aan een gedeeltelijke verbetering van de beheers- en reproduceerbaarheid van de dataleveringen, door de brondata op te nemen in het concern datawarehouse en de data via een extratieschema toegankelijk te maken, zie ook paragraaf 7.3. Dit was de start van een breder project vanuit het datawarehouseteam W&I om een stap naar een volgend volwassenheidsniveau te nemen.

De ervaringen uit het project kunnen waardevolle input zijn bij het inrichten van een volwassener gegevensbeheer. Tijdens de looptijd van het project heeft het programma gegevensmanagement zich nog op een hoger abstractieniveau begeven waardoor nog weinig echte kruisbestuiving heeft kunnen plaatsvinden. Voor echte verbetering zal een inspanning nodig zijn op het gebied van metadata- en masterdatamanagement als onderdeel van gegevensmanagement. Er zal soms een sterke link naar architectuur zijn (keuzes en richtlijnen voor de plaats waar we welk soort gegevens opslaan). Met deze inspanningen zal het gegevenslandschap en het beheer daarvan zich steeds verder in kunnen ontwikkelen naar een volwassener niveau. Als daarin goede stappen gemaakt kunnen worden zou dat volgende projecten tijd schelen en de betrouwbaarheid van data sterk vergroten, en het is een randvoorwaarde voor het betrouwbaar uitrollen van ontwikkelde producten.

Advies: betrek de ervaringen uit het project uitkeringsfraude bij het lopende programma gegevensmanagement; zo kan de top-down inspanning worden gevoed met bottom-up praktische ervaringen om tot een inrichting van data diensten te komen die directe meerwaarde heeft voor data-projecten.

11.3 Lab-omgeving voor innovatie is noodzakelijk (maar niet ingewikkeld)

Een grote complicatie voor dit project lag in een eigenlijk heel simpel probleem: er was een omgeving nodig waarin zowel de juiste analysesoftware (voor dit project 'R') als de data beschikbaar waren, met voldoende rekenkracht en werkgeheugen. De software is echter niet beschikbaar op de gemeentelijke (Citrix) werkstations, en de vertrouwelijke data mag daar vanwege beveiligingsredenen niet zomaar vanaf.

Er zijn verschillende routes geprobeerd om hiervoor oplossingen te vinden. Als tijdelijke noodoplossing heeft het team met gemeentelijke stand-alone laptops gewerkt waarop de software wel geïnstalleerd kon worden en waarop de data werd gekopieerd. Dit had echter grote nadelen, zowel voor het samenwerken (er was geen centrale plek waarop meerdere mensen te gelijk konden werken), als voor de rekenkracht; de laptops hadden te weinig geheugen voor de gevraagde toepassingen.

Een tijd lang heeft het team geprobeerd via de reguliere IT-kanalen een oplossing te vinden: ofwel 'R' beschikbaar krijgen op het gemeentelijke netwerk ofwel een aparte server in te richten waarop dit kon. Hoewel de technische vraag eigenlijk heel simpel is was het proces hiervoor belemmerend; dit is niet ingericht op snelheid en flexibiliteit, maar eerder op robuustheid en lange-termijnoplossingen. Dat is een goede attitude voor bedrijfsoplossingen die door honderden of duizenden medewerkers in de gehele gemeente jarenlang gebruikt gaan worden, maar onwerkbaar voor innovatie. Dit spoor is dan ook doodgelopen met de nodige vertraging voor het project tot gevolg. De les voor de organisatie is dat voor dit soort projecten mogelijkheden moeten worden ingericht om snel en flexibel tijdelijk de juiste IT-voorzieningen te organiseren. Dat kan zo simpel zijn als een of twee mensen met de juiste kennis en houding die bij dit soort vragen als aanspreekpunt fungeren.

Uiteindelijk is als oplossing voor het project een externe cloudomgeving ingericht via Amazon Web Services, na uitgebreid overleg met de privacy- en security-verantwoordelijken. Deze bestaat uit remote-desktop-toegang tot een beveiligde server waarop analysetool R studio, programmeertaal Java en databasesysteem SQL kunnen worden gebruikt (en op aanvraag andere software). Gebruikers die toegang nodig hebben tot de omgeving kunnen, na autorisatie en installatie, de server benaderen vanaf een met het internet verbonden computer. (zie ook paragraaf 4.4)

Het inrichten van deze oplossing kostte alsnog de nodige tijd en afstemming, en kent af en toe nog problemen. Sowieso is beheer en technische expertise nodig om deze te onderhouden. Gedurende de laatste maanden van het project is een nieuwe, robuustere externe cloudomgeving neergezet. Het proces hier naartoe was draconisch vergeleken bij de technische complexiteit. Dit werd onder meer veroorzaakt door veranderde houding en processen bij IT-security, en processen die niet zijn afgestemd op dit soort vraagstukken. Uiteindelijk is met vereende krachten een resultaat geboekt waarmee het product voorlopig beheerd kan worden, maar de benodigde tijd (meer dan 6 maanden) en inspanning (vele werkweken van meerdere in- en externe medewerkers) staat in geen verhouding tot de technische complexiteit.

Hoewel het project nu hiermee tijdelijk uit de brand is, is een oplossing binnen de gemeente toch noodzakelijk voor duurzaam beheer; hiervoor wordt een Advanced Analytics Platform ontwikkeld.

Advies: richt een proces in dat geschikt is voor het opleveren van flexibele IT-diensten voor innovatie (Damhof-kwadrant IV), los van de reguliere beheerde IT-kanalen (Damhof-kwadrant II). Houd de infrastructuur voorlopig zo eenvoudig en klein als voor de gewenste projecten nodig. Geleidelijk opschalen is hierin verstandiger dan in één keer een grote omgeving neer te zetten.

11.4 Privacy wordt handelbaar door PIA en goede fasering

Privacy is een 'hot topic' in overheidsland en met nieuwe Europese regels is er sprake van een grote alertheid en voorzichtigheid, mede vanwege de dreiging van hoge boetes. Dit heeft in de gemeentelijke context soms een verlammend effect omdat er nog beperkt duidelijkheid is over wat officieel wel en niet mag. Dit wordt nog eens extra gecompliceerd door de hierboven genoemde problemen in het gegevensmanagement, waardoor vaak niet makkelijk goed te specificeren is welke data exact wordt gebruikt en welke privacygevoelige elementen daar al dan niet in zitten.

In het project is een heldere aanpak gekozen van duidelijk specificeren van doel en scope van de projectfase, zodat de vraag om toestemming van de privacy-officer zo smal mogelijk kon worden gemaakt. Er is alleen toestemming gevraagd voor datalevering ten behoeve van de labfase, met daarbij de afspraak dat voor vervolgfasen een nieuwe PIA zal worden gedaan, waarin dan gebruik kan worden gemaakt van de kennis en onderbouwing uit de vorige fase. Er is een Privacy Impact Assessment (PIA) opgesteld waarin de risico's en mitigerende maatregelen zijn beschreven. Deze PIA is bijgevoegd als bijlage E.

Advies: zie privacy niet bij voorbaat als een blokkerend onderwerp, maar beschrijf doel, scope en te gebruiken data van het project helder en ga vroeg in het project in gesprek met de privacy en security officers. Deel het project op in fases en vraag per fase akkoord. Schrijf de

privacy- en beveiligingsrisico's en -maatregelen uit in een Privacy Impact Assessment volgens een vast template om deze vast te leggen.

11.5 Samen stappen zetten noodzakelijk voor leereffect en voortgang

In een data-science project worden dagelijks kleine en grotere beslissingen genomen over hoe om te gaan met de data en hoe deze te interpreteren. De grote lijnen van de aanpak en de resultaten kunnen goed worden samengevat en overgebracht in een presentatie of rapport, maar het echte leren door de organisatie treedt pas op als interne mensen het hele proces van zo dicht mogelijk bij meemaken. Van een afstand meekijken werkt daarbij maar beperkt en ook documentatie kan daar slechts een beperkt niveau van diepgang bereiken.

Hiervoor is dus vaste interne betrokkenheid voor meerdere dagen in de week sterk gewenst van medewerkers die motivatie hebben om in de betreffende richting door te groeien. Er zijn ook meerdere specialismen binnen een data-science team die vaak verschillende medewerkers vereisen. In de labfase is pas in de laatste fase echte structurele ondersteuning aangesloten vanuit OBI. In de pilotfase is het team uitgebreid met nog twee OBI-medewerkers en zijn er twee vaste teamwerkdagen ingericht waarop het gehele team bij elkaar op kantoor zat. Er is een team nodig en een bepaalde team dynamiek om te werken aan een innovatietraject als dit. Je hebt meerdere mensen nodig om elkaar scherp te houden, maar ook bepaalde oplevermomenten om de vaart erin te houden. Gedurende het project heeft dit op sommige momenten goed gewerkt en op andere minder. Duidelijke rolverdeling en opdrachtgeverschap zijn essentieel hierin.

In de projectfase bleek dat er in de organisatie maar beperkt mensen waren die echt mee konden leren van het project. De afdeling OBI, die als ambitie had om door te groeien tot expertisecentrum IGW, was slechts beperkt betrokken en geïnteresseerd, hoewel wel meerdere presentaties en deep dives zijn georganiseerd. Vooralsnog is in de gemeente Rotterdam zeer beperkt expertise aanwezig op dit gebied en om de beschreven ambities waar te maken zullen flinke ingrepen nodig zijn.

Advies: zorg bij projecten met een leerdoelstelling voor een goede mix van gemotiveerde interne en externe medewerkers, maak de interne medewerkers ook echt zoveel als mogelijk vrij om deel te nemen aan de dagelijkse projectpraktijk.

11.6 Eigenaarschap, aansturing en verandermanagement bij innovaties

Zoals al in 10.1 beschreven was duidelijk opdrachtgeverschap noodzakelijk om het project van de grond te laten komen. Ook betrokkenheid van het primaire proces (de business) waarin de innovatie veranderingen zou laten plaatsvinden, was essentieel. Tegen het eind van het project zijn hierover de volgende reflecties te geven:

- Hoewel het gezamenlijk oprachtgeverschap grote betrokkenheid heeft laten zien, is de kennis en ervaring met dit soort trajecten nog gering, waardoor echte sturing voor de opdrachtgevers vaak lastig was. Echt richting geven vond maar beperkt plaats en veel beslissingen kwamen uiteindelijk op het uitvoerende team neer. Dit is begrijpelijk in deze fase van volwassenheid en zal met de tijd verbeteren. Kennis en ervaring zijn echter dus zowel op werkvloerniveau als op managementniveau noodzakelijk.
- Afstemming op hoog bestuurlijk niveau is ook noodzakelijk bij ingrepen in het primaire proces. Op grote schaal toepassen van het ontwikkelde product bleek aan het eind van het proces nog op weerstand en verschillende belangen te stuiten. Eerder betrekken van alle beslissers is hierbij belangrijk.
- Het proces waarin de veranderingen moeten landen heeft zelf ook een lage volwassenheid; veel handmatige acties, verschillende zelfgemaakte excelbestanden en weinig definities en vastgelegde procedures. Dit werd al beschreven in paragraaf 10.3. Dit heeft zijn weerslag op het gemak van implementatie. Het belang van goed verandermanagement is doorslaggevend in of een goed en

gevalideerd ook daadwerkelijk gebruikt gaat worden. Hier moet expliciet aandacht en expertise voor beschikbaar zijn; verandermanagement kan vaak niet door dezelfde mensen worden begeleid die het technische werk uitvoeren.

-

Advies: zet in op ontwikkeling van kennis en ervaring bij managers die met innovatietrajecten in aanraking komen om aansturing te verbeteren. Neem in teams en projectaanpak expliciet ruimte voor verandermanagement op; een innovatie is pas geslaagd als hij leidt tot daadwerkelijke verandering van het primaire proces. Veranderbereidheid is makkelijk te overschatten.

11.7 Praktijkervaring en pionieren blijft nodig voor groei

Dit project laat zien dat het mogelijk is om data science projecten en advanced analytics succesvol uit te voeren binnen de gemeente Rotterdam, als aan een aantal randvoorwaarden wordt voldaan. Het laat zien dat er tal van obstakels zijn voor succes, maar dat deze te overwinnen zijn met een pragmatische houding en doorzettingsvermogen.

Het geloof is echter dat deze obstakels niet keer op keer opnieuw hoeven te worden overwonnen; dit soort pionier-projecten kan de weg effenen voor vervolgprojecten en leiden tot standaarden en werkprocessen voor veelvoorkomende moeilijkheden. Bovendien creëren ze kennis en begrip in de organisatie over data science en wat erbij komt kijken.

Hiermee is het uitvoeren van echte projecten een onmisbare input voor de bredere ontwikkelingen die zijn gericht op informatiegestuurd werken, gegevensmanagement, infrastructuur en architectuur.

Advies: blijf innovatieve projecten van verschillende aard starten en uitvoeren om kennis en ervaring te genereren voor de bredere ontwikkeling. Laat juist de medewerkers die verantwoordelijk zijn voor het inrichten van gegevensmanagement, architectuur en infrastructuur, intensief kennismaken met deze projecten om te borgen dat de lessen worden omgezet in echte verandering.

BIJLAGE A – GEBRUIKTE DATA EN ABT's

ABT naam	modelnaam	brontabellen	abt buidling	ABT	modelling	results
basis	code/7.modellen/basis/model_fit_tree1.rds	Brontabellen/kopie van 1-JV-2015-jaarporductie versie DEF.xlsx Brontabellen/Kopie van 2-Algemene informatie WZ versie DEF.xlsx Brontabellen/Kopie van 3-Algemene informatie Wzrelatie versie DEF.xlsx	Code/4.feature building gestructureerd/abt_basis.R	Code/2.data/abt_basis_6_dec.rds	Code/4.feature building gestructureerd/abt_basis.R	Code/4.feature building gestructureerd/abt_basis.R Code/12.Resulaten/plot_hitrates.R
basis_geen_onderzoek	code/7.modellen/basis/model_fit_gbm3.rds	Brontabellen/kopie van 1-JV-2015-jaarporductie versie DEF.xlsx Brontabellen/Kopie van 2-Algemene informatie WZ versie DEF.xlsx Brontabellen/Kopie van 3-Algemene informatie Wzrelatie versie DEF.xlsx	Code/4.feature building gestructureerd/abt_basis.R	Code/2.data/abt_basis_6_dec.rds	Code/4.feature building gestructureerd/abt_basis.R	Code/4.feature building gestructureerd/abt_basis.R Code/12.Resulaten/plot_hitrates.R
basis_gesnoeid	code/7.modellen/basis/model_fit_tree4.rds	Brontabellen/kopie van 1-JV-2015-jaarporductie versie DEF.xlsx Brontabellen/Kopie van 2-Algemene informatie WZ versie DEF.xlsx Brontabellen/Kopie van 3-Algemene informatie Wzrelatie versie DEF.xlsx	Code/4.feature building gestructureerd/abt_basis.R	Code/2.data/abt_basis_6_dec.rds	Code/4.feature building gestructureerd/abt_basis.R	Code/4.feature building gestructureerd/abt_basis.R Code/12.Resulaten/plot_hitrates.R

basis_plus_2015	code/7.modellen/basis_plus_2015/model_gbm_7_dec.rds	brontabellen/igw_09_99_export_01.mdb brontabellen/igw_09_99_export_02.mdb brontabellen/igw_09_99_export_04.mdb brontabellen/igw_09_99_export_06.mdb brontabellen/igw_09_99_export_component.mdb	Code/4.feature building gestructureerd/abt_basis_plus.R	Code/2.data/abt_basis_plus_5_dec.rds	Code/6.Modelleren/run_models_over_abts_basis_plus_2015.R	Code/6.Modelleren/run_models_over_abts_basis_plus_2015.R Code/12.Resultaten/plot_hitrates.R
basis_plus	code/7.modellen/basis_plus/model_gbm_7_dec.rds	brontabellen/igw_09_99_export_01.mdb brontabellen/igw_09_99_export_02.mdb brontabellen/igw_09_99_export_04.mdb brontabellen/igw_09_99_export_06.mdb brontabellen/igw_09_99_export_component.mdb	Code/4.feature building gestructureerd/abt_basis_plus.R	Code/2.data/abt_basis_plus_5_dec.rds	Code/6.Modelleren/run_models_over_abts_basis_plus.R	Code/6.Modelleren/run_models_over_abts_basis_plus.R Code/12/Resultaten/plot_hitrates.R
advanced	code/7.modellen/model_gbm_advanced_5_dec.rds	brontabellen/igw_09_99_export_01.mdb brontabellen/igw_09_99_export_02.mdb brontabellen/igw_09_99_export_04.mdb brontabellen/igw_09_99_export_06.mdb brontabellen/igw_09_99_export_component.mdb brontabellen/igw_09_99_export_03_2007_2013.mdb brontabellen/igw_09_99_export__032014_2016.mdb	Code/4.feature building gestructureerd/abt_gestructueerd_advanced.R	Code/2.data/abt_advanced_5_dec.rds	Code/6.Modelleren/run_models_over_abts_advanced.R	Code/6.Modelleren/run_models_over_abts_advanced.R Code/12/Resultaten/plot_hitrates.R

premium	code/7.modellen/model_gbm_premium_14_dec.rds	brontabellen/igw_09_99_export_01.mdb brontabellen/igw_09_99_export_02.mdb brontabellen/igw_09_99_export_04.mdb brontabellen/igw_09_99_export_06.mdb brontabellen/igw_09_99_export_component.mdb brontabellen/igw_09_99_export_03_2007_2013.mdb brontabellen/igw_09_99_export__032014_2016.mdb	Code/4.feature building gestructureerd/abt_gestructueerd_advanced.R Code/5.feature building tekst/tm_contacten_parallel_freq_words_ID_modular.R	Code/2.data/abt_premium_14_dec.rds	Code/6.Modelleren/run_models_over_abts_premium.R	Code/6.Modelleren/run_models_over_abts_premium.R Code/12.Resulaten/plot_hitrates.R

BIJLAGE B – GEDETAILEERDE MODELRESULTATEN

	model naam	TP	FN	FP	TN	Baseline	Accuracy	Kappa	Sensitivity	Specificity	hit_volume50	hit_volume100	hit_volume500	hit_volume1000	hit_volume3000
basis	model_fit_tree1	936,00	0,00	0,00	588,00	0,61	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,94	NA
basis_geen_onderzoek	model_fit_tree3	778,00	158,00	357,00	231,00	0,61	0,66	0,24	0,83	0,39	0,72	0,67	0,65	0,69	NA
basis_geen_onderzoek	model_fit_gbm3	783,00	153,00	298,00	290,00	0,61	0,70	0,35	0,84	0,49	0,88	0,89	0,82	0,74	NA
basis_geen_onderzoek	model_fit_rf3	738,00	198,00	279,00	309,00	0,61	0,69	0,32	0,79	0,53	0,70	0,78	0,83	0,73	NA
basis_gesnoeid	model_fit_tree4	776,00	160,00	407,00	181,00	0,61	0,63	0,15	0,83	0,31	0,64	0,63	0,65	0,67	NA
basis_plus_2015	model_gbm_7_dec	946,00	102,00	282,00	191,00	0,61	0,75	0,34	0,90	0,40	0,96	0,97	0,90	0,83	NA
basis_plus_2015	model_glm_7_dec	939,00	109,00	285,00	188,00	0,61	0,74	0,33	0,90	0,40	0,94	0,95	0,88	0,82	NA
basis_plus_2015	model_rpart_7_dec	961,00	87,00	323,00	150,00	0,61	0,73	0,27	0,92	0,32	0,86	0,88	0,86	0,79	NA
basis_plus_2015	model_rf_7_dec	917,00	131,00	262,00	211,00	0,61	0,74	0,35	0,88	0,45	0,96	0,92	0,91	0,82	NA
basis_plus	model_gbm_7_dec.rds	1881,00	1366,00	1140,00	2309,00	0,51	0,63	0,25	0,58	0,67	0,86	0,83	0,78	0,74	0,62
basis_plus	model_glm_7_dec.rds	1857,00	1390,00	1187,00	2262,00	0,51	0,62	0,23	0,57	0,66	0,88	0,86	0,78	0,72	0,61
basis_plus	model_rpart_7_dec.rds	1711,00	1536,00	1164,00	2285,00	0,51	0,60	0,19	0,53	0,66	0,80	0,77	0,63	0,57	0,59
basis_plus	model_rf_7_dec.rds	1916,00	1331,00	1186,00	2263,00	0,51	0,62	0,25	0,59	0,66	0,70	0,81	0,75	0,74	0,62
advanced	model_gbm_advanced_5_dec.rds	1912,00	1335,00	813,00	2636,00	0,51	0,68	0,35	0,59	0,76	1,00	0,97	0,90	0,84	0,68
advanced	model_tree_advanced_5_dec.rds	1640,00	1607,00	973,00	2476,00	0,51	0,61	0,22	0,51	0,72	0,68	0,67	0,68	0,64	0,59
advanced	model_rf_advanced_5_dec.rds	2030,00	1217,00	795,00	2654,00	0,51	0,70	0,40	0,63	0,77	1,00	1,00	0,89	0,85	0,70
advanced	model_gbm_advanced_spec_23_jan.r	1503,00	1744,00	702,00	2747,00	0,51	0,63	0,26	0,46	0,80	0,98	0,99	0,84	0,77	0,64
advanced	model_gbm_advanced_Sens_24_jan.r	2059,00	1188,00	964,00	2485,00	0,51	0,68	0,36	0,63	0,72	0,98	0,92	0,88	0,84	0,68
advanced	model_gbm_advanced_AUC_25_jan.r	1966,00	1281,00	813,00	2636,00	0,51	0,69	0,37	0,61	0,76	1,00	1,00	0,91	0,85	0,69
premium	model_gbm_premium_2_dec.rds	1954,00	1293,00	818,00	2631,00	0,51	0,68	0,37	0,60	0,76	1,00	1,00	0,89	0,83	0,69
premium	model_rf_premium_2_dec.rds	2026,00	1221,00	831,00	2618,00	0,51	0,69	0,38	0,62	0,76	1,00	0,97	0,87	0,83	0,70
premium	model_gbm_premium_14_dec.rds	1934,00	1313,00	781,00	2668,00	0,51	0,69	0,37	0,60	0,77	1,00	0,99	0,91	0,84	0,69
premium	model_tree_premium_14_dec.rds	1737,00	1510,00	1046,00	2403,00	0,51	0,62	0,23	0,53	0,70	0,98	0,97	0,63	0,64	0,60
premium	model_logreg_premium_14_dec.rds	1912,00	1335,00	936,00	2513,00	0,51	0,66	0,32	0,59	0,73	0,98	0,96	0,89	0,82	0,66

BIJLAGE C – VERGELIJKING POPULATIES

[redacted]

January 12, 2017

Vergelijking van de onderzoekspopulatie met de gehele populatie uitkeringsgerechtigden

Alle de modellen zijn gebouwd op de onderzoekspopulatie. De modellen konden enkel gebouwd worden op de onderzoekspopulatie omdat er voor deze populatie labels beschikbaar waren voor het target. Oftewel, omdat deze mensen onderzocht zijn weten we of deze onrechtmatigheid plegen met de uitkering. Er zijn redenen waarom mensen onderzocht en andere niet. Daarom hebben de mensen in deze set een verhoogd risico op onrechtmatigheid. Dit maakt deze set gebiased en waarschijnlijk niet representatief voor de gehele populatie. Het volgende onderzoek is uitgevoerd op te kijken hoe veel deze populaties van elkaar verschillen en bij welke features dit met name naar voren komt.

```
results <- readRDS(paste(datafolder,"resultaten_vergelijk_populaties_130117.rds",
sep = "/"))
abt <- readRDS(paste(datafolder, "abt_advanced_all_3_jan.rds", sep = "/"))
```

Voor een goede vergelijking met de abt voor het modelleren zijn ook hier alle NA's vervangen door 0

Zijn er significante verschillen tussen de populaties in de verschillende features?

Voor de numerieke features is dit getest met een t-test. Voor de categorische features is dit getest met een chi-square test. Significantie level is vastgesteld op 0.05. Dit significantie level in met gecorrigeerd met bonferroni correctie.

Hieruit volgt dat van de 436 features, 327 features significant verschillen tussen de populaties

Dit is veel en betekent dat de onderzoekspopulatie op veel vlakken verschilt van de echte populatie. Er moet wel rekenen mee gehouden worden dat hier met grote volumes aan data worden gewerkt. Hierdoor wordt een klein verschil sneller een significant verschil. Het is daarom van belang om te weten over dit verschil betekenisvol is voor deze data.

Zijn de verschillende tussen de populaties betekenisvol?

Om dit te berekenen is gebruik gemaakt van de effect size.

Voor numerieke features is CohensD gebruikt. Een effectsize vanaf .3 wordt gezien als een gemiddeld effect. Effectsize vanaf .5 wordt gezien als een hoog effectsize. Voor de categorische features is gebruik gemaakt voor CramersV. Voor de interpretatie van CramersV is gebruik gemaakt van de richtlijnen van Cramer 1988

Een telling van de features in de low medium en high classes

```
##
##    low medium
##    395     33
```

Zoals te zien is in de resultaten zijn er geen features met een hoge effectsize en slecht NA features met een gemiddelde effect size. Dit betekend dat ondanks dat de populaties significant verschillend zijn, zijn deze verschillen niet van grote betekenis te zijn voor de meeste features. Dit maakt het reeler dat de onderzoekspopulatie als referentie voor de gehele populatie kan dienen.

Inzoemen om de features met een gemiddelde effectsize

Als we nu kijken welke features dit zijn komen we uit op het lijstje hieronder

Features

target_per
koppelcodedef
relatie_algemene_relatie
relatie_subject
uitkering_actief_systeem2
soort_uitk_participatiewet
dgn_in_huidige_uitkering
persoonlijke_eigenschappen_nl_spreken
persoonlijke_eigenschappen_nl_lezen
persoonlijke_eigenschappen_nl_schrijven
persoonlijke_eigenschappen_nl_begrijpen
comp__bijstandsnorm_alleenstaande_21
comp_ind_component
deelname_project
reintegratieladder_werkreintegratie
contacten_onderwerp_arbeidsmotivatie
contacten_onderwerp_diagnosegesprek
contacten_onderwerp_document
contacten_onderwerp_documenten_innemen
contacten_onderwerp_inkomen
contacten_onderwerp_maatregel_overweging
contacten_onderwerp_match__werk
contacten_onderwerp_overige
contacten_onderwerp_terugbelverzoek
contacten_onderwerp_traject
contacten_onderwerp_uitnodiging
contacten_opmerkingen_totaaltelling
contacten_opmerkingen_aantal_afspraken
contacten_onderwerp_afg_jaar_document
contacten_onderwerp_afg_jaar_traject
contacten_opmerkingen_afg_jaar_totaaltelling
contacten_opmerkingen_afg_jaar_aantal_afspraken onderzoekspopulatie

Wat hier direct aan opvalt is dat de grootste hoeveelheid aan features, features uit contacten tabel zijn

Om naar de richting van dit effect te kijken kunnen we kijken naar de gemiddelden

features	**onderzoekspopulatie**	**gehele_populatie**	**verschil**
koppelcodedef	59867.58	60332.55	-464.97
relatie_algemene_relatie	0.77	0.70	0.07
relatie_subject	1.49	1.44	0.05
dgn_in_huidige_uitkering	1863.33 days	1800.32 days	NA
persoonlijke_eigenschappen_nl_spreken	1.09	0.66	0.43
persoonlijke_eigenschappen_nl_lezen	1.09	0.65	0.44
persoonlijke_eigenschappen_nl_schrijven	1.11	0.66	0.45
persoonlijke_eigenschappen_nl_begrijpen	1.08	0.65	0.43
deelname_project	2.04	1.57	0.47

features	onderzoekspopulatie	gehele_populatie	verschil
reintegratieladder_werkreintegratie	1.18	0.80	0.38
contacten_onderwerp_arbeidsmotivatie	1.04	0.33	0.71
contacten_onderwerp_diagnosegesprek	0.34	0.15	0.19
contacten_onderwerp_document	16.46	5.82	10.64
contacten_onderwerp_documenten_innemen	1.07	0.55	0.52
contacten_onderwerp_inkomen	0.57	0.17	0.40
contacten_onderwerp_maatregel_overweging	0.17	0.03	0.14
contacten_onderwerp_match__werk	0.18	0.03	0.15
contacten_onderwerp_overige	5.57	1.75	3.82
contacten_onderwerp_terugbelverzoek	0.44	0.13	0.31
contacten_onderwerp_traject	4.11	1.24	2.87
contacten_onderwerp_uitnodiging	0.60	0.21	0.39
contacten_opmerkingen_totaaltelling	1768.13	588.08	1180.05
contacten_opmerkingen_aantal_afspraken	29.47	10.63	18.84
contacten_onderwerp_afg_jaar_document	6.05	3.15	2.90
contacten_onderwerp_afg_jaar_traject	1.54	0.63	0.91
contacten_opmerkingen_afg_jaar_totaaltelling	834.02	372.12	461.90
contacten_opmerkingen_afg_jaar_aantal_afspraken	13.47	6.77	6.70

Er is te zien dat voor elke features het verschil positief is. Dit betekent dat er voor elke features meer van te vinden zijn bij de onderzoekspopulatie dan in de gehele populatie. Dit lijkt aan te geven dat er meer contact is geweest met de onderzoekspopulatie dan met de gemiddelde uitkeringsgerechtigden.

```
CrossTable(abt$soort_uitk_participatiewet, abt$onderzoekspopulatie, chisq = T,
expected = T, format = "SPSS")
##
##    Cell Contents
## |-------------------------|
## |                   Count |
## |         Expected Values |
## | Chi-square contribution |
## |             Row Percent |
## |          Column Percent |
## |           Total Percent |
## |-------------------------|
##
## Total Observations in Table:  57455
##
##                              | abt$onderzoekspopulatie
## abt$soort_uitk_participatiewet |        0  |        1  | Row Total |
## -------------------------------|-----------|-----------|-----------|
##                              0 |     2918  |      640  |     3558  |
##                                | 2087.488  | 1470.512  |           |
##                                |  330.421  |  469.054  |           |
##                                |   82.012% |   17.988% |    6.193% |
```

```
##                                |    8.656% |    2.695% |           |
##                                |    5.079% |    1.114% |           |
## -------------------------------|-----------|-----------|-----------|
##                              1 |    30791  |    23106  |    53897  |
##                                | 31621.512 | 22275.488 |           |
##                                |   21.813  |   30.965  |           |
##                                |   57.129% |   42.871% |   93.807% |
##                                |   91.344% |   97.305% |           |
##                                |   53.592% |   40.216% |           |
## -------------------------------|-----------|-----------|-----------|
##                   Column Total |    33709  |    23746  |    57455  |
##                                |   58.670% |   41.330% |           |
## -------------------------------|-----------|-----------|-----------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  852.2528     d.f. =  1     p =  2.353928e-187
##
## Pearson's Chi-squared test with Yates' continuity correction
## ------------------------------------------------------------
## Chi^2 =  851.227     d.f. =  1     p =  3.933858e-187
##
##
##        Minimum expected frequency: 1470.512
CrossTable(abt$comp__bijstandsnorm_alleenstaande_21, abt$onderzoekspopulatie,
chisq = T, expected = T, format = "SPSS")
## Warning in chisq.test(t, correct = FALSE, ...): Chi-squared approximation
## may be incorrect
##
##    Cell Contents
## |-------------------------|
## |                   Count |
## |         Expected Values |
## | Chi-square contribution |
## |             Row Percent |
## |          Column Percent |
## |           Total Percent |
## |-------------------------|
##
## Total Observations in Table:  55020
##
##                                          | abt$onderzoekspopulatie
## abt$comp__bijstandsnorm_alleenstaande_21 |        0  |        1  | Row Total |
## -----------------------------------------|-----------|-----------|-----------|
##                                        0 |    15231  |     8754  |    23985  |
##                                          | 13479.465 | 10505.535 |           |
##                                          |  227.596  |  292.025  |           |
##                                          |   63.502% |   36.498% |   43.593% |
##                                          |   49.258% |   36.325% |           |
##                                          |   27.683% |   15.911% |           |
## -----------------------------------------|-----------|-----------|-----------|
##                                        1 |    15689  |    15343  |    31032  |
##                                          | 17439.849 | 13592.151 |           |
##                                          |  175.774  |  225.532  |           |
##                                          |   50.557% |   49.443% |   56.401% |
##                                          |   50.739% |   63.667% |           |
##                                          |   28.515% |   27.886% |           |
## -----------------------------------------|-----------|-----------|-----------|
##                                        2 |        1  |        2  |        3  |
##                                          |    1.686  |    1.314  |           |
```

```
##                                          |    0.279  |    0.358  |           |
##                                          |   33.333% |   66.667% |    0.005% |
##                                          |    0.003% |    0.008% |           |
##                                          |    0.002% |    0.004% |           |
## -----------------------------------------|-----------|-----------|-----------|
##                              Column Total |    30921  |    24099  |    55020  |
##                                          |   56.200% |   43.800% |           |
## -----------------------------------------|-----------|-----------|-----------|
##
##
## Statistics for All Table Factors
##
##
## Pearson's Chi-squared test
## ------------------------------------------------------------
## Chi^2 =  921.5642     d.f. =  2     p =  7.671448e-201
##
##
##
##        Minimum expected frequency: 1.314013
## Cells with Expected Frequency < 5: 2 of 6 (33.33333%)
```

Hetzelfde is te zien voor de categorische variables.

Conclusie 1

De resultaten lijken aan te geven dat er meer contact is geweest met mensen die binnen de onderzoekspopulatie vallen dan met de standaard uitkeringsgerechtigeden. Dit impliceert dat iemand eerst op de radar moet komen voordat er een onderzoek naar wordt gestart. Dit zou kunnen komen doordat veel tips vanuit de gemeente zelf komen. Er is dan veel contact met iemand en tijdens deze contacten komt er een vermoede van fraude op. Verder laten deze resultaten zijn dat, zoals verwacht, de onderzoekspopulatie niet gelijk is aan de gehele populatie. Echter is dit nog geen reden om te stoppen met het gebruiken van de onderzoekspopulatie als basis voor de modellen omdat de verschillen klein lijken te zijn. Om verder te kunnen beslissen of onderzoekspopulatie representief is (en daarmee of een model gemaakt op de onderzoekspopulatie zal werken op de hele populatie) moet het model uitgetest worden op een representieve steekproef van de gehele populatie. Zo'n dergelijke steekproef is te vinden in de aselecte steekproef van 2015.

BIJLAGE D – Tardis

Om dit model te maken is het noodzakelijk om de volgende vraag te kunnen beantwoorden:

"Wat wisten wij van persoon/uitkering x op het moment vlak voordat een historisch onrechtmatigheidsonderzoek werd gestart."

Wanneer we deze informatie voor historische onderzoeken hebben kunnen we - gecombineerd met het bekende resultaat van de historische onderzoeken - een model maken waarbij we op basis van de gereconstrueerde data de uitkomst van het onderzoek voorspellen. Het model dat we hiermee verkrijgen kunnen we vervolgen toepassen op de actuele populatie waarbij we eigen de volgende vraag beantwoorden: Wat is het te verwachten resultaat van een potentieel rechtmatigheidsonderzoek?

De twee systemen waar we informatie uit gebruiken voor dit project zijn RAAK en Socrates.

Dit document beschrijft voor beide systemen afzonderlijk de stappen die we hebben genomen om een historische status te reconstrueren. Hierbij wordt onderscheid gemaakt tussen de volgende 2 stappen:

1. Wat was de inhoud van de database op historische datum X?

2. Wat was de actuele informatie op historische datum X?

Om het onderscheid tussen bovenstaande 2 te illustreren. Het kan zijn dat iemand een (psychische, sociale of fysieke) beperking had van 1-1-2015 t/m 31-12-2015. Wanneer we teruggaan naar de kennis van 1-7-2016 is er actueel geen sprake van een beperking (er was immers alleen in 2015 sprake van een beperking). Wel hadden we op 1-7-2016 de kennis dat er een historische beperking was.

RAAK

Een typische tabel (bv persoonlijke eigenschappen) met historie uit RMW/RAAK heeft de volgende vorm:

rownum	jn_nr	Jn datetime	Jn operation	Hob-bies sport	Diag-nose	advies	Zelf-standig-heid opm	Spreek taal
[redacted]								

Artikel 8.8 Woo, jo artikel 65 PW

[redacted]

Artikel 8.8 Woo, jo artikel 65 PW

Het rijnummer (eerste kolom) is geen onderdeel van RAAK, maar is toegevoegd om verwijzing vanuit de tekst naar de tabel mogelijk te maken. Ook zijn veel kolommen niet getoond omdat dit te veel zou zijn.

Belangrijkste eigenschappen uit de tabellen uit RNW/RAAM zijn:

- Elk logisch record krijgt een eigen unieke journaalnummer (jn_nr).
- Er wordt nooit data overschreven. Wanneer een logisch record wordt gewijzigd wordt er technisch een record toegevoegd met hetzelfde journaalnummer (jn_nr) met een actuelere journaaldatum (jn_datetime).
- Ook wanneer het logische record wordt gewist wordt er technisch een record toegevoegd met waarde "DEL" in veld jn_operation.

Historische status tabel
Om de historische status van een tabel te reconstrueren is het alleen maar nodig om de journaalmutatiedatums na de gewenste datum te verwijderen. Zo was na 19-7-2016 regels 1 t/m 8 beschikbaar. Tussen 21-9-2015 en 19-8-2016 waren regels 1 t/m 7 beschikbaar. Tussen 18-5-2015 en 21-9-2017 waren regels 1 t/m 6 beschikbaar. Enz.

Actuele informatie tabel

Na reconstructie kan de actueel (voor de historische datum) geldende informatie teruggevonden worden in het meest recente record per journaalnummer (jn_nr). Zo was na 19-7-2016 regel 8 actueel. Tussen 21-9-2015 en 19-8-2016 was regel 7 actueel. Enz. Indien het meest actuele record binnen een journaalnummer (jn_nr) waarde "DEL" heeft in veld jn_operation dan bestaat het logische record niet.

Socrates

Een typische tabel (bv relatie) met historie uit Socrates heeft de volgende vorm:

Row-num	Subjectnr-relatie	geldig	status	Dt-opvoer	Dt- afvoer	Dt-aanvang	dteinde

[redacted]

Artikel 8.8 Woo, jo artikel 65 PW

Artikel 8.8 Woo, jo artikel 65 PW

Het rijnummer (eerste kolom) is geen onderdeel van Socrates, maar is toegevoegd om verwijzing vanuit de tekst naar de tabel mogelijk te maken. Ook zijn veel kolommen niet getoond omdat dit te veel zou zijn.

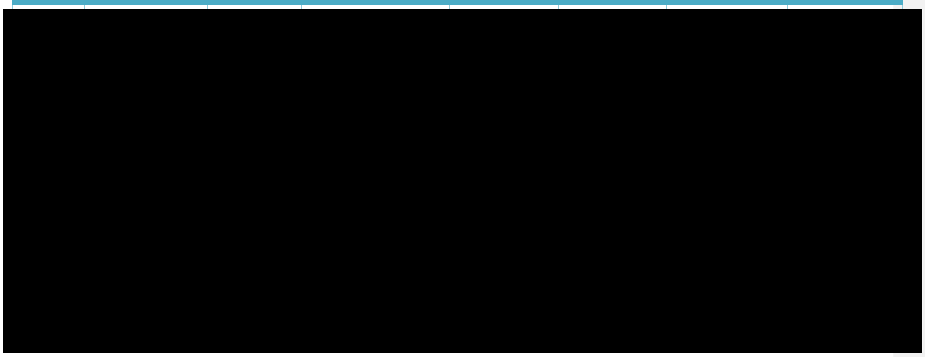
Belangrijkste eigenschappen uit de tabellen uit Socrates zijn:

- Records zijn voorzien van een datum opvoer, afvoer, aanvang en einde. De datum opvoer en afvoer geven de tijdsperiode aan wanneer de informatie in het record geldig is en de datums aanvang en einde geven aan op welke periode de informatie betrekking heeft. (Voorbeeldinterpretatie regel 1: Tussen 20-8-2010 en 20-9-2010 hadden we de kennis dat kind met subjectrelatienummer [redacted] vanaf 7-8-1998 (zonder einddatum) bij de cliënt inwoonde.)
- Vrijgegeven (actuele) records hebben geen dtafvoer. Indien er een update plaatsvindt wordt het oude record van een dtafvoer voorzien en verandert de status van "Vrijgegeven" in "Overschreven".
- Bij bovenstaande mutatie worden twee nieuwe records aangemaakt waarvan een waarde "GELDIG" heeft in veld 'geldig' en een waarde "NIET GELDIG" heeft in veld 'geldig'. Deze laatste is een technisch artefact en bevat geen informatie.

Historische status tabel

Om de historische status van een tabel te reconstrueren moeten alle records met dtopvoer na de gewenste datum verwijderd worden. Indien datum afvoer (dtafvoer) in een bepaald record een waarde na de gewenste datum heeft dient veld dtafvoer NA/NULL gemaakt te worden en dient de status gewijzigd te worden van "Overschreven" naar "Vrijgegeven". De status van bovenstaande tabel op 1-1-2011 was derhalve:

Row-num	Subjectnr-relatie	geldig	status	Dt-opvoer	Dt-afvoer	Dt-aanvang	Dt-einde



Artikel 8.8 Woo, jo artikel 65 PW

Actuele informatie tabel

Wanneer we een Socrates-tabel interpreteren beperken we ons tot de geldige records (waarde "GELDIG" voor veld 'geldig'). Vervolgens moeten we voor de actuele situatie op zoek gaan naar de records met vrijgegeven status (waarde "Vrijgegeven" in veld 'status'). Op 1-1-2011 was derhalve bekend dat kind met subjectrelatie [redacted] tussen 7-8-1988 en 6-9-2010 woonachtig was bij cliënt. (NB. Het kind is dus per 7-9-2010 niet meer woonachtig bij cliënt).

Nieuwe invoer

In enkele tientallen gevallen komt het voor dat records in Socrates status "Ingevoerd" of "Oud" hebben. Aangezien deze wijzigingen soms (of vaak?) compleet en in de database niet traceerbaar verwijderd worden lijkt het erop dat dit voorgenomen wijzigingen zijn die (nog) niet doorgevoerd zijn. Het kan dus nog twee kanten op:

1. De wijzigingen worden doorgevoerd waarna de status "Oud" veranderd in "Overschreven" en de status "Ingevoerd" veranderd in "Vrijgegeven"
2. De wijzigingen worden teruggedraaid en de records met status "Ingevoerd" worden gewist en de records met status "Oud" worden teruggezet naar status "Vrijgegeven" met een NA/NULL bij dtafvoer.

In onze interpretatie van de inhoud van de tabellen wordt de mutatie onder 2. doorgevoerd en we beschouwen de voorgenomen wijzigingen als niet-bestaand.

Hieronder volgt een voorbeeld:

Row-num	Subjectnr-relatie	geldig	status	Dt-opvoer	Dt-afvoer	Dt-aanvang	dteinde
[redacted]							

Artikel 8.8 Woo, jo artikel 65 PW

BIJLAGE E – PRIVACY IMPACT ASSESSMENT

Documenteigenschappen

Titel	Project Analytics Uitkeringsfraude: Privacy Impact Assesment
Auteur	Projectgroep
Versie	1.2
Versie Datum	7-9-2016

Documenthistorie

Datum	Versie	Aangepast door	Opmerkingen
16-8-2016	0.5	[redacted]	PIA-vragenlijst ingevuld
17-8-2016	1.0	[redacted]	Algemene info, overwegingen en maatregelen toegevoegd
7-9-2016	1.2	[redacted]	Opmerkingen [redacted] (MO) verwerkt

Inleiding

In dit document zijn de privacy-aspecten van het te starten project beschreven, en zijn de vragen uit de modelvragenlijst van NOREA voor een Privacy Impact Assesment beantwoord. Dit document is een bijlage bij de 'Projectopzet project Analytics Uitkeringsfraude'.

Doelbinding

Het uiteindelijke doel van het project (na de uitrolfase) is om data te gebruiken voor het beter opsporen van onrechtmatigheden in de uitkeringsverstrekking. Dit is een wettelijke taak van de gemeente op basis van de Participatiewet (2015). Om dit doel te bereiken wordt vooralsnog alleen 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. Als in de toekomst sprake is van het koppelen van andere gegevens zal daaraan een privacy-toets voorafgaan.

Proportionaliteit

Fraude en misbruik van gemeenschapsvoorzieningen zijn zware vergrijpen die een stevige handhaving van de uitvoerder rechtvaardigen. Bovendien kan een betrouwbaarder voorspelling leiden tot minder belasting van compliante burgers door minder onterechte onderzoeken. Desalniettemin wordt in het project veel aandacht besteed aan borgen van proportionaliteit. Analytics en machine learning gebruiken in beginsel vaak zoveel mogelijk data juist om onvermoede verbanden te kunnen opsporen. Voor de labfase gebruiken we veel meer velden (namelijk zoveel als mogelijk) juist om te onderzoeken in welke van die velden de voorspellende waarde verborgen zit. Als dat bekend is kunnen we in de pilotfase (als we echt de resultaten gaan gebruiken in de praktijk) met veel minder informatie toe; Voor de pilot fase worden uiteindelijk alleen die velden gebruikt waarvan gebleken is dat ze een voorspellende waarde hebben. Zo bewaken we proportionaliteit.

Anonimiseren en toegestane gegevensvastlegging

Voor de doelstelling van de labfase is het in orde om met zoveel mogelijk geanonimiseerde data te werken, omdat we in die fase nog geen actie koppelen aan de voorspellingen maar alleen een model ontwikkelen. Het zal praktisch niet mogelijk zijn de data volledig te anonimiseren, vooral vanwege de vrije tekstvelden met gespreksverslagen die we willen onderzoeken. Daarin zit mogelijk zoveel informatie verborgen dat hij met wat moeite tot identificatie kan leiden. Bovendien bestaat er een geringe kans dat hierin gegevens

vastgelegd zijn die niet direct verband houden met de uitvoering van de participatiewet en dus niet verzameld/vastgelegd hadden mogen zijn. In de labfase zal in het proces van 'feature selection' gelet worden op deze effecten. De resultaten van de tekstanalyse worden vastgelegd en besproken met privacy-experts om vast te stellen welke gegevens in de vervolgfases wel en niet gebruikt mogen worden.

In de pilot- en realisatiefase zullen de pseudoniemen voor die personen waar het model een gefundeerd vermoeden van fraude oplevert, weer moeten worden terugherleid naar de natuurlijke personen. Dit proces zal gecontroleerd en onder strenge voorwaarden worden uitgevoerd. Voor het starten van de Pilotfase (en opnieuw voorafgaand aan eventuele uitrol) wordt om deze redenen een nieuwe PIA uitgevoerd.

Informatiebeveiliging

Alle data blijft op gemeentelocaties en wordt verwerkt op gemeentelijke hardware met bijbehorende beveiligingsmaatregelen. Alleen direct bij het project betrokken personen krijgen toegang tot de gegevens. Autorisaties voor de projectlocatie op het netwerk en eventueel later afgeschermde delen van databases worden door de opdrachtgever beheerd. Daarbij geldt het principe dat alleen personen waarbij toegang noodzakelijk is voor het project autorisaties krijgen.

Beantwoording vragenlijst Privacy Impact assessment Labfase Analytics Uitkeringsfraude

Hieronder worden de vragen voor een Privacy Impact Assement beantwoord. Dit gebeurt specifiek voor de labfase van het project. Voor elke vervolgfase zal deze vragenlijst opnieuw worden ingevuld en beoordeeld.

1 Het type project

1.1 Is er sprake van het verwerken van persoonsgegevens?

Ja. Er wordt zoveel als mogelijk geanonimiseerd (en gepseudonimiseerd) direct bij de bron; BSNs worden niet meegezonden (maar gepseudonimiseerd), geboortedata worden op geboortejaar afgekort, van postcodes worden alleen de cijfers meegegeven etc. Echter, de vrije tekstvelden met gespreksverslagen bevatten grote hoeveelheden gegevens waarin mogelijk ook persoonsgegevens zoals e-mailadressen en mogelijk BSNs verwerkt zijn.

1.2 Is het duidelijk wie verantwoordelijk is voor de verwerking van gegevens?

Ja. Het project wordt uitgevoerd in opdracht van [redacted], manager W&I Toetsing en Toezicht. Het project wordt begeleid door een begeleidingsgroep met de opdrachtgever, de afdelingsmanager van OBI en externe experts van de Hogeschool Rotterdam en Accenture.

1.3 Verwerkt uw organisatie de persoonsgegevens in opdracht en onder verantwoordelijkheid van een andere organisatie? Ofwel treedt uw organisatie op als bewerker?

Nee.

1.4 Is het duidelijk wie na afloop van het project verantwoordelijk is voor het in stand houden en evalueren van de getroffen maatregelen?

Ja. Na afloop van de labfase worden de resultaten geëvalueerd door de begeleidingsgroep. De opdrachtgever zal in overleg met de afdelingsmanager OBI beslissen of het project wordt voortgezet in een pilotfase. Voor deze pilotfase zal een nieuwe PIA worden uitgevoerd die wordt beoordeeld door de privacyfunctionaris van W&I.

1.5 Is het doel van de verwerking van persoonsgegevens binnen het project voldoende SMART omschreven?

Ja. In de labfase van dit project worden de gegevens verwerkt teneinde vast te stellen of er voorspellende waarde voor fraude uit is te halen met behulp van analytics en tekst mining. In de periode tot 1 januari 2017 wordt een eerste model ontwikkeld en getest, en wordt de vraag beantwoord of vrije tekstvelden voor het detecteren van fraude en/of onrechtmatigheden bij bijstandsuitkeringen toegevoegde waarde hebben.

1.6 Is er sprake van:

a. Gebruik van nieuwe technologie?

Nee

b. Gebruik van technologie die bij het publiek vragen of weerstand op kan roepen?

Mogelijk; geautomatiseerde analyse van grote hoeveelheden gegevens kan discussie oproepen.

c. De invoering van bestaande technologie in een nieuwe context?

Mogelijk. Hoewel door de nationale overheid zeer vergelijkbare analyses al enige tijd worden uitgevoerd, is het relatief nieuw voor een overheid.

d. (Andere) grote verschuivingen in de werkwijze van de organisatie, de manier waarop persoonsgegevens worden verwerkt en/of de technologie die daarbij gebruikt wordt?

Nee, eventuele verschuivingen in de werkwijze zullen op zijn vroegst plaatsvinden na afloop van een succesvolle pilot (twee projectfases verder).

e. Een nieuwe verwerking van persoonsgegevens

Nee, de gegevens werden al verwerkt.

f. Het verzamelen van meer of andere persoonsgegevens dan voorheen of een nieuwe manier van verzamelen.

Nee, er wordt gebruik gemaakt van gegevens die reeds verzameld zijn.

g. Gebruik van al verzamelde gegevens voor een nieuw doel of een nieuwe manier van gebruiken.

Nee

1.7 Heeft u op alle bovenstaande (a t/m g) nee geantwoord?

Nee

1.8 Is er (naast de Wbp) veel wet-en regelgeving ten aanzien van persoonsgegevens waar het project mee te maken heeft?

Ja, de Participatiewet regelt de Bijstand

1.9 Zijn er veel maatschappelijke belanghebbenden?

Nee

1.10 Zijn er bij veel partijen betrokken de uitvoering van het project?

Nee; de Hogeschool Rotterdam en Accenture werken mee aan het project, maar als medewerkers van de Gemeente.

1.11 Is er een geschillenregeling of een partij waar de betrokkene terecht kan bij vragen of klachten?

Ja, reguliere regeling Gemeente Rotterdam.

2 De Gegevens

2.1 Zijn alle gegevens nodig om het doel te bereiken (worden er zo min mogelijk gegevens verzameld)?

Ja, er worden zo min mogelijk gegevens verzameld voor deze fase: de labfase dient mede om vast te stellen welke gegevens voorspellende waarde hebben teneinde in een eventuele vervolgfase een groot deel van de gegevens niet meer te hoeven gebruiken. Voor de labfase is nog een brede selectie gegevens nodig om alle eventuele onvermoede verbanden te kunnen vaststellen. Dit heeft mede als doel om gegevens die niet relevant blijken te zijn in latere fases te kunnen uitsluiten.

2.2 Kan het doel met geanonimiseerde of gepseudonimiseerde gegevens worden bereikt (terwijl daar op dit moment geen gebruik van wordt gemaakt)?

De gegevens worden al maximaal gepseudonimiseerd; de in de ongestructureerde tekstvelden verborgen persoonsgegevens worden op het eerst mogelijke moment uitgefilterd en niet verder gebruikt in de analyse.

2.3 Kunnen de gegevens gebruikt worden om het gedrag, de aanwezigheid of prestaties van mensen in kaart te brengen en/of te beoordelen (ook al is dit niet het doel)

Ja, dit is ook het doel van het project.

2.4 Is sprake van het verwerken van:

a. Bijzondere persoonsgegevens?
Nee

b. Uniek identificerende gegevens?

Nee

c. Wettelijk voorgeschreven persoonsnummers?

Nee (gepseudonimiseerd)

d. Andere gegevens dan hiervoor beschreven waarvoor geldt dat sprake is van een (gepercipieerde) verhoogde gevoeligheid?

Aantekeningen van gesprekken met gemeentemedewerkers en soms e-mailwisselingen.

2.4.1 Bij een van bovenstaande Ja: kan het doel met andere gegevens worden bereikt die een verminderd risico op misbruik met zich mee brengen?

Nee

2.5 Verwerkt u gegevens over kwetsbare groepen of personen?

Nee

2.6 Hebben de gegevens betrekking op de gehele of grote delen van de bevolking?

De gehele populatie van bijstandsgerechtigden

3 Betrokken partijen

3.1 Zijn er (na afronding van het project) bij het verzamelen en verder verwerken van de gegevens meerdere interne partijen betrokken?

Nee

3.2 Zijn er (na afronding van het project) bij het verzamelen en verder verwerken van de gegevens meerdere externe partijen betrokken?

Nee, wanneer het project wordt voortgezet zullen eventuele externe medewerkers wederom als gemeentelijk medewerkers worden ingezet.

3.3 Zijn er partijen betrokken (in het project of bij de verwerking) die zich niet aan een met Nederland vergelijkbare privacywetgeving hoeven te houden?

Nee.

3.4 Is de verstrekking van de gegevens aan derde partijen in lijn met het doel waarvoor de gegevens oorspronkelijk zijn verzameld?

n.v.t.

3.5 Worden de gegevens verkocht aan derde partijen?

Nee

4 Verzamelen van gegevens

4.1 Kan de manier waarop de gegevens worden verzameld worden opgevat als privacy gevoelig?

Nee

4.2 Is het doel van het verzamelen van de gegevens publiekelijk bekend of kan het publiekelijk bekend gemaakt worden?

Ja, reguliere proces rond aanvraag en beheer van uitkering.

4.3 Verzamelt u de gegevens op basis van een van de wettelijke grondslagen?

Ja, op basis van de participatiewet

4.4 Is duidelijk of u de gegevens verzamelt op basis van opt-in (verzameling uitsluitend als de betrokkene daarvoor toestemming heeft gegeven) of op basis van opt-out (verzameling tenzij de betrokkene daartegen bezwaar heeft gemaakt)?

Ja, verplichte verzameling o.b.v. de wet.

4.4.1 Indien u toestemming aan de betrokkene vraagt (opt-in), kunnen de betrokkenen de toestemming op een later tijdstip intrekken (opt-out)?

Ja, de kandidaat kan altijd afzien van de uitkering.

4.4.2 Is de impact van het intrekken van de toestemming groot voor de betrokkene?

Ja; verlies van uitkering

4.5 Meldt u de betrokkene dat de gegevens worden verzameld?

Ja, regulier proces uitkeringen

4.5.1 n.v.t.

4.5.2 Bij Ja (op vraag 4.5): meldt u de betrokkene waarom de gegevens worden verzameld (wat u er mee gaat doen)?

Ja

4.5.3 Bij Ja: (op vraag 4.5): meldt u de betrokkene aan wie de gegevens worden verstrekt (daar waar dit geen wettelijke verplichting is)?

Ja.

4.6 Zou de betrokkene kunnen worden verrast door de verwerking (op het moment dat hij daarover wordt geïnformeerd)?

Nee

5 Gebruik van gegevens

5.1 Is het gebruik van de gegevens verenigbaar (in lijn) met het doel van het verzamelen?

Ja, de gegevens worden gebruikt om te onderzoeken of de wettelijke taak beter kan worden uitgeoefend door deze nieuwe toepassing.

5.2 Worden de gegevens gebruikt voor andere bedrijfsprocessen of doelen dan waar ze oorspronkelijk voor verzameld zijn?

Nee, ga verder met vraag 5.3

5.2 N.v.t.

5.2.1 N.v.t

5.3 Is de kwaliteit van de gegevens gewaarborgd, dat wil zeggen: zijn de gegevens actueel, juist en volledig?

Nee; veel velden zijn niet of niet volledig gevuld, mogelijk niet altijd actueel.

5.4 Worden op basis van de gegevens beslissingen genomen over de betrokkenen?

Nee.

5.4.1 Bij Ja: leveren de gegevens een volledig en actueel beeld van de betrokkenen op?

N.v.t.

5.5 Is sprake van koppeling, verrijking of vergelijking van gegevens uit verschillende bronnen?

Ja; Socrates en RMW/RAAk

5.6 Worden de gegevens breed verspreid binnen de organisatie?

Nee

5.7 Worden de gegevens breed verspreid buiten de organisatie?

Nee

5.7.1 Is het doorgeven van de gegevens aan partijen buiten de organisatie in lijn met de verwachtingen van het individu?

n.v.t.

5.8 Stelt uw organisatie profielen op van de betrokkenen, al dan niet geanonimiseerd?

Ja.

5.8.1 Indien profielen worden opgesteld, kan het profiel tot uitsluiting of stigmatisering leiden?

Nee

5.9 Kunnen de betrokkenen hun gegevens inzien of daarom vragen?

Ja, conform reguliere proces gemeente.

5.10 Kunnen de betrokkenen hun gegevens corrigeren of daarom vragen (verbeteren, aanvullen)?

Ja, conform reguliere proces gemeente

5.11 Kunnen de betrokkenen hun gegevens verwijderen of daarom vragen?

Ja, conform reguliere proces gemeente

6 Bewaren en vernietigen

6.1 Is een bewaartermijn voor de gegevens vastgesteld?

Ja, regulier proces gemeente

6.2 Kunnen de gegevens na afloop van de bewaartermijn fysiek worden verwijderd (uit een bestand) of vernietigd (papier)?

Ja, regulier proces gemeente

6.2.1 Zo ja (op vraag 6.2), worden de gegevens na verstrijken van de bewaartermijn op zo'n manier vernietigd of verwijderd dat ze niet meer te benaderen en te gebruiken zijn?

Ja, regulier proces gemeente

7 Beveiligen

7.1 Is sprake van intern geformuleerd beleid over het beveiligen van informatie?

Ja, dit is belegd bij de afdeling Bestuurs- en Concernondersteuning IM MO/W&I in de functie van de adviseur informatiebeveiliging, dhr. [redacted]