Kunstmatige Intelligentie Impact Assessment WPI model onderzoekswaardigheid uitkeringsaanvragen

Kunstmatige Intelligentie (hierna KI): Slimme Check Bijstandsaanvragen, Gemeente Amsterdam
Directie: Cluster Sociaal
Afdeling: Werk, Participatie en Inkomen
Ambtelijke verwerkings verantwoordelijke: Renger Visser
Leverings verantwoordelijke: Renger Visser
Opdrachtgever: Renger Visser
Proceseigenaar: 5.1,2,e
Werkproces: Beoordelen aanvragen Levensonderhoud
Epic Owner: 5.1,2,e en 5.1,2,e
Product Owner: 5.1,2,e
Opgesteld door: 5.1,2,e en 5.1,2,e
5.1,2,e
Datum:12 augustus 2022
Versie: 0.5

Documenten waarnaar wordt verwezen in deze KIIA:

- DPIA aangaande de inzet van een ondersteunend algoritme voor het rechtmatigheidsonderzoek bij de aanvragen voor levensonderhoud
- Overzicht Verwerkte Data en Features
- Architectuurnotitie
- IAMA (Impact Assessment Mensenrechten en Algoritmes)[1]
- Weging betrokken (grond)rechten[2]
- Bewaartermijnen - algoritme onderzoekswaardigheid wpi
- Quickscan Informatiebeveiliging Directie WPI: Product WPI Risicomodel
- Documentatie bias analyse
- Beheerplan
- Meetplan

---

1 [1] Wij hebben meegedaan aan de pilotfase van IAMA. Daar was het format anders dan het definitieve format. Het nieuwe format is gevuld op basis van wat tijdens de pilot IAMA is opgeschreven.

2 [2] Deze weging zit ook in de definitieve versie van de IAMA. Wij hebben een pilotfase van de IAMA meegemaakt en toen was dit nog geen onderdeel van de IAMA. Wij hadden deze weging toen al wel gedaan en ook als input gegeven aan de ontwikkelaars van de IAMA. Daarom komt ons format niet helemaal overeen met het IAMA format maar de kern is hetzelfde.

Advies Participatieraad en reactie
CPA advies
Gebruikersinstructie;
(Pre)-Pilotplan
Pre-pilot resultaten
Doelstellingen en succescriteria
WPD bias analyse
Technische documentatie
Documentatie DASO
Documentatie Dataverkenner
Richtlijn voor hoge risico KI-toepassingen
Communicatieplan pilot informatiegestuurd handhaven
Richtlijn European Approach for Al (Proposal fora Regulation on artificial intelligence)

Inhoudsopgave:

1. Inleiding 4
2. Bepalen noodzaak voor het doen van een kunstmatige intelligentie impact assessment 6
3. KI-toepassing, verantwoordelijkheden en betreffende werkproces 13
4. Beschrijf de baten van de KI-toepassing 21
5. Architectuur 23
6. Data en data governance 24
7. Wettige en ethische toepassing 32
8. Bias analyse 42
9. Technische documentatie 43
10. Risicomanagement 44
11. Informatiebeveiliging 46
12. Beheer 51
13. Communicatie 52
14. Verantwoording 54
15. Afwegingen en conclusie 55

1. Inleiding - Aanleiding en probleemstelling

De afdeling **WPI** van de gemeente Amsterdam heeft als doel gesteld het aantal ten onrechte verstrekte bijstandsuitkeringen te verminderen. Dit voorkomt het ontstaan van bijstandsschulden als gevolg van een ten onrechte verstrekte uitkering. Dergelijke schulden hebben een grote impact op het leven van Amsterdammers, zeker wanneer hun bestaanszekerheid kwetsbaar is. Daarnaast moet ervoor worden gezorgd dat publiek geld wordt ingezet waar nodig en het sociale stelsel betaalbaar blijft.

Op dit moment worden aanvragen handmatig geselecteerd voor een extra onderzoek en deze processtap kan effectiever. Het algoritme draagt hieraan bij door elke nieuwe aanvraag te voorzien van een onderzoekswaardigheidsscore. Deze indicatie helpt een medewerker om te bepalen of een aanvraag nader moet worden onderzocht voordat een bijstandsuitkering wordt verstrekt.
Met de inzet van het model kan de onderzoekswaardigheid van een aanvraag beter worden vastgesteld, waardoor aanvragers die wel bijstandsgerechtigd zijn ook zo min mogelijk worden belast met een onderzoek. Dit leidt tot een meer proportionele inzet van de bevoegdheid om nader onderzoek uit te voeren. Dit verbetert ook de dienstverlening ten behoeve van de burger.

In het huidige aanvraagproces wordt 4% van de aanvragen handmatig door de inkomens consulent naar handhaving gestuurd voor nader onderzoek. Van deze selectie doorgestuurde aanvragen blijkt 48% achterafterecht geselecteerd. Het model onderzoekswaardigheid uitkeringsaanvragen zal dit proces ondersteunen. Als we kijken naar de voorlopige resultaten van het model, gebaseerd op de prestatie op historische data, dan zou 9% van de aanvragen nader onderzocht worden, met als resultaat dat 57% daadwerkelijk onderzoekswaardig is. Afhankelijk van de gekozen drempelwaarde voor het doorsturen van de aanvragen, zouden we ook minder dan 9% van de aanvragen kunnen doorzetten. Het percentage terecht onderzochte aanvragen zou dan verder stijgen.
Onafhankelijk van de gekozen drempelwaarde betekent dit dat we in verhouding met minder onterechte onderzoeken (vergeleken met het huidige proces) evenveel (of meer) onrechtmatige aanvragen vinden, en dus meer proportioneel werken.

De KIIA is een aanvulling op de reeds bestaande risicoanalyses, zoals de DPIA en het Security (ontrol Framework. In deze KIIAworden gerichtere vragen behandeld die meer inzicht geven in de werking van de Kl. Het bevat daarnaast de kern van eerder uitgevoerde analyses die zijn uitgevoerd door verschillende experts, zoals de architecten, security officer en data scientist. In dit document zal dan ook veelvuldig worden verwezen naar deze verschillende analyses, zoals de bias analyse, om dit document overzichtelijk te houden. De KIIA is gebaseerd op de Artificial lntelligence Impact Assesment, aangevuld met inzichten uit de Impact Assessment Mensenrechten en Algoritmes (IAMA) en nieuwe Europese wetgeving over regulatie van kunstmatige intelligentie.

2. Bepalen noodzaak voor het doen van een kunstmatige intelligentie impact assessment

2.3. ***Wordt de kunstmatige intelligentie (hierna: KI) toegepast in een nieuw (maatschappelijk) domein?***

Het werken met KI binnen het proces van WPI is nieuw. Het proces waarbinnen het algoritme wordt toegepast, is een bestaand werkproces. Namelijk het beoordelen en van een aanvraag voor levensonderhoud door Amsterdammers bij de gemeente Amsterdam.

2.4. ***Wordt gebruik gemaakt van een nieuwe vorm van KI-technologie?***

Nee, er is geen sprake van een nieuwe vorm van Kunstmatige Intelligentie.

2.5. ***Is er sprake van een verboden KI-toepassing zoals beschreven in art. 5 van de (concept) regelgeving European Approach/or Al van de Europese Commissie?***

Nee, toetsend aan art. 5 en bijlage 111 kan worden geconstateerd dat er geen sprake is van een verboden KI-toepassing.

2.6. ***Heeft de KI een hoge mate van autonomie?***

Nee. De KI heeft een lage mate van autonomie. Het geeft enkel een advies in de vorm van een onderzoekswaardigheidsscore voor een aanvraag aan de verantwoordelijke medewerkers. Na dit advies volgt nog een uitgebreid onderzoek dat door verschillende medewerkers wordt uitgevoerd. Eerst kijkt een medewerker of de aanvraag wel echt onderzoekswaardig is. Als dit het geval is vindt er een uitgebreid onderzoek plaats om de aanvraag te beoordelen. Er is dus geen geautomatiseerde besluitvorming en er is sprake van betekenisvolle menselijke tussenkomst voordat er een besluit wordt genomen.

2.7. ***Wordt de KI toegepast in een complexe omgeving?***

Nee, de KI wordt toegepast in een afgebakende omgeving die niet complex is. De input is vooraf bepaald en kan enkel wijzigen als er wordt gekozen voor (door)ontwikkeling van de Kl. Bij het implementeren van de KI in de bedrijfsvoering zal de input van de KI ook niet wijzigen.

2.8. ***Wordt de KI toegepast binnen een 'gevoelig'werkproces die impact heeft op burgers?***

Ja, het gaat om een gevoelig werkproces. Amsterdammers vragen een bijstandsuitkering aan op het moment dat zij in een afhankelijk situatie zitten en financiële ondersteuning nodig hebben. Het al dan niet verkrijgen van een bijstandsuitkering heeft daarom grote impact op deze burgers. Op basis van een grote hoeveelheid (persoonlijke) gegevens, die wordt getoetst aan het wettelijke kader en gemeentelijk beleid, zal worden vastgesteld of iemand al dan niet bijstandsgerechtigd is.

2.9. Heeft de KI eerder schade toegebracht aan fundamentele rechten van natuurlijke personen of groepen?

Nee, het is een nieuw ontworpen KI-toepassing specifiek voor het werkproces.

2.10. Zijn er zorgen over de KI en de potentiële schade en impact?

De KI wordt ingezet in een heel gevoelig werkproces zoals hierboven besproken. Daarnaast is de KI mede ontwikkeld op basis van persoonsgegevens[3]. In de basis zijn aandachtspunten:

1. Dat het werkproces gedurende de periode waarover de beschikbare data is gebruikt mogelijk op verschillende manieren is uitgevoerd door medewerkers;
2. Dat de gehanteerde business logica/ beslisregels niet doorlopend zijn getoetst op juridische houdbaarheid of zijn gewijzigd;
3. Dat dataverwerkingen historisch gezien mogelijk onjuist zijn geweest *(historica! bias);*
4. Dat er sprake is van onbewuste vooringenomenheid in de onderliggende dataset.

Vanwege voorgaande ontstaat, zonder aanvullende (risico)analyses en maatregelen, het gevaar dat:

De KI-toepassing onjuiste output geeft;
De KI-toepassing niet binnen de juridische grenzen blijft van de wettelijke bevoegdheid;
De KI-toepassing discrimineert bij het genereren van onderzoekswaardigheidsscores;
De KI-toepassing is ontwikkeld op basis van onrechtmatig verwerkte data.

Op deze onderdelen is onderzoek gedaan. De conclusie is dat deze risico's zijn gemitigeerd door het nemen van aanvullende maatregelen. Deze maatregelen zijn beschreven in de "Documentatie Bias Analyse"; "Overzicht Verwerkte Data en Features"; en het "Beheerplan". Daarnaast is een DPIA uitgevoerd om te beoordelen of de KI-toepassing wel binnen de juridische grenzen van de wettelijke bevoegdheid blijft. Daarnaast is de juridische documentatie voorgelegd aan de Commissie Persoonsgegevens Amsterdam. Dit adviesorgaan toetst ookop het niet overschrijden van juridische en ethische grenzen.

Daarnaast zijn er een aantal aanvullende risico's tijdens de pilotfase. Deze risico's en hun mitigatie is beschreven in Tabel 1.

Tabel 1: Risico's tijdens de pilotfase

Automation bias Medewerkers kunnen te veel gaan vertrouwen op het model, en daardoor te sterk beïnvloed worden door haar uitkomsten. Als het model	De interface is zoveel mogelijk ontworpen om automation bias tegen te gaan. Hoewel het model elke aanvraag een score geeft van 0 - 1, waarbij 1 het meest en 0 het minst

[3] Naast persoonsgegevens wordt data uit gemeentelijke processen (bijvoorbeeld over afhandeling van aanvragen uit het verleden) gebruikt.

het label 'onderzoekswaardig' geeft moet dit het startpunt zijn voor een onderzoek, zonder dat dit leidt tot vooringenomenheid bij de medewerkers. Hier sterk aan gelinkt is het begrip *deployment bias,* waarbij een algoritme in de praktijk om een andere wijze gebruikt wordt dan hoe het oorspronkelijk bedoeld was, hetgeen kan leiden tot oneerlijke/ ongelijke uitkomsten.	onderzoekswaardig is, ziet de medewerker enkel een label onderzoekswaardig ja/nee. Dit om te voorkomen dat een zeer hoge score leidt tot vooringenomenheid. Daarnaast ziet de medewerker bij onderzoekswaardige aanvragen altijd de drie belangrijkste features die hebben geleid tot dit label. Dit houdt de interface eenvoudig, maar geeft de medewerker wel inzicht in hoe het algoritme tot dit label is gekomen, hetgeen vooringenomenheid kan voorkomen (bovendien neemt na de eerste drie features de feature importance flink af). Ook staan er op meerdere plekken in de interface herinneringen dat het algoritme slechts een hulpmiddel is, en dat de medewerker altijd uit moet gaan van zijn/haar eigen expertise. Tot slot zal er in de werkinstructie en trainingen informatie worden opgenomen over hoe het algoritme ontworpen is, het gevaar van overmatig vertrouwen in het algoritme en het belang van de eigen expertise van de medewerkers.
Bias in het model Hoewel er geen gevoelige kenmerken zoals gender en nationaliteit worden gebruikt in het model, kan indirecte bias in het model toch voorkomen. Dit kan bijvoorbeeld gebeuren vanwege bias in de historische data *(historica! bias)* waarop het model getraind is, of doordat de werkelijke situatie veranderd waardoor het model onvoldoende aansluit de huidige werkelijkheid.	Er is een bias analyse uitgevoerd. Deze analyse, de resultaten en de vervolgstappen die zijn genomen zijn terug te vinden in de *Documentatie bias analyse.* Deze bias analyse zal tijdens de implementatiefase met regelmatig herhaald worden. Daarnaast zijn in het beheerplan monitoring maatregelen beschreven om bijvoorbeeld drift[4] te voorkomen.
Onvoldoende aanvragen Het kan zijn dat in de pilot periode onvoldoende aanvragen binnekomen om te kunnen bepalen of er een significante vooruitgang is in de effectiviteit van het handhavingsinzet.	Er is een pre-pilot uitgevoerd waardoor de hit-rate (en daarmee de benodigde sample size) van het model beter ingeschat kan worden. Daarnaast is er een mogelijkheid de pilot met beperkte tijd te verlengen totdat voldoende informatie is verzameld om te kunnen zeggen

[4] Drift vindt plaats wanneer de data waarop het model gebruikt wordt afwijkt van de data waarop het model getraind is, bijvoorbeeld doordat de actuele situatie veranderd is.

	of er een significante verbetering is in vergelijking met het huidige proces. Het is bij een dergelijke verlenging wel van belang dat scherp wordt vastgelegd hoelang de verlenging is, en welke informatie nog verzamelt dient te worden, zie "onduidelijk eindpunt pilot".
Onduidelijk eindpunt pilot Er kan een neiging ontstaan om een pilot steeds te blijven verlengen, waardoor het beslismoment van implementatie wordt uitgesteld- immers is er altijd meer informatie en meer onderzoek dat gedaan kan worden.	De scope van de pilot is vooraf helder vastgelegd, en tijdens de pilot zullen de tijdslijnen scherp bewaard worden. Daarnaast zijn de doelen van de pilot goed vastgelegd in het pilot plan.
Weerstand handhaving Als het model niet-onderzoekswaardige zaken aanmerkt als onderzoekswaardig, of regelmatig bugs en issues voorkomen in het model, bestaat het risico dat de handhavingsspecialisten weerstand opbouwen tegen het model, omdat het inefficiënt en tijdrovend is	De Medewerker Vooronderzoek zal worden ingezet om aanvragen die eenvoudig door de Inkomensconsulent kunnen worden afgewezen eruit te filteren, zodat de Handhavingsspecialisten deze aanvragen niet te zien krijgen. Ook blijven we de performance van het model goed monitoren, en zijn er duidelijke kanalen beschikbaar voor het geven feedback en het melden van bugs/ issues.
Veranderingen applicatielandschap Het applicatielandschap van WPI is aan verandering onderhevig, hierdoor kunnen problemen ontstaan in de data aanvoer naar het model	De cockpit van de WPI trein, bestaande uit de **1-**lead; Business lead en architect, houdt oog op het effect van veranderingen in het applicatielandschap van het model.
Negatieve opinievorming (extern/intern) Extern: als onvolledige of onjuiste informatie over het model en/of deze pilot in de media verschijnt, is er een risico dat het maatschappelijk draagvlak voor dit product afneemt. Daarnaast kan na implementatie door negatieve berichtgeving of wijzigingen in het politieke klimaat van Amsterdam het draagvlak ook in een later stadium afnemen. Intern: als binnen de organisatie onjuiste of onvolledige informatie rondgaat over deze pilot, is er een risico dat het draagvlak en vertrouwen in dit product intern afneemt.	Er is een zorgvuldig opgesteld mitigatieplan om deze risico's te mitigeren. Dit document heeft de titel *Communicatieplan Pilot Informatiegestuurd Handhaven.*

2.11. ***Wordt gebruik gemaakt van gevoelige gegevens over personen?***

Ja, er worden persoonsgegevens verwerkt van Amsterdammers die een aanvraag levensonderhoud hebben gedaan bij de Gemeente Amsterdam:

- *Amsterdammers die op grond van de Participatiewet een uitkering aanvragen*

Daarnaast kunnen persoonsgegevens verwerkt worden van de volgende categorieën:
- Partner van de klant op hetzelfde adres
- Kind(eren) van de klant op hetzelfde adres
- Kostendelers (personen die op hetzelfde adres wonen)

Ook worden in het hoofdproces persoonsgegevens verwerkt van de volgende categorieën betrokkenen:
- Uitwonend(e) kind(eren) ingeval van co-ouderschap
- Ex-partner(s) in geval van onderhoudsplicht
- De door de klant gemachtigde persoon

Zie *Overzicht Verwerkte Data en Features 1.0* voor een overzicht van de verwerkte gegevens. Dit document beschrijft ook de data die uiteindelijk noodzakelijk is om de KI de gewenste output te laten genereren.

2.12. ***Wordt een voorspelling gegenereerd op persoonlijke niveau op basis van persoonlijke kenmerken?***

Ja, er wordt een onderzoekswaardigheidsscore gegenereerd voor de mate waarin een aanvraag al dan niet in aanmerking komt voor nader onderzoek. De voorspelling is daarmee op zaa knivea u/persoonsniveau.

2.13. ***Neemt de KJ-beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?***

De KI-toepassing neemt geen zelfstandig besluit, maar heeft wel een wezenlijke invloed op de manier van het beoordelen van een aanvraag en het besluitvormingsproces. Het ontwikkelde algoritme genereert op basis van de informatie uit de aanvraag en andere beschikbare informatie over de aanvrager een score die de onderzoekswaardigheid van de aanvraag aangeeft. Deze onderzoekswaardigheidsscore wordt berekend voor nieuw binnenkomende aanvragen en het algoritme creëert dus geen "nieuwe" te onderzoeken zaken op basis van lopende uitkeringen. Daarnaast worden sowieso alle aanvragen beoordeeld op rechtmatigheid. Het model zorgt er dus niet voor dat sommige aanvragen niet meer bekeken worden, maar draagt slechts bij aan het bepalen van de noodzaak om een diepgaand nader onderzoek te doen in plaats van een minder diepgaand onderzoek. Daarnaast is de uitkomst van het diepgaandere onderzoek een advies aan de Inkomens Consulent. Die maakt uiteindelijk het definitieve besluit.

2.14. In hoeverre is de output van de KI van invloed op het besluitvormingsproces binnen de organisatie?

Zie antwoord bij vraag 2.11.

2.15. Is er, mede bezien vanuit de geconstateerde en voorzienbare risico's, sprake van een hoge risico KI-systeem?

Nee, de KI-toepassing valt niet onder de Al-systemen met een hoog risico in de zin van Bijlage 111 van de concept Richtlijn. Eén van de hoog-risico-systemen die genoemd is in Bijlage 111 van de Richtlijn is de volgende: *"Al-systemen die bedoeld zijn om door of namens overheidsinstanties te worden gebruikt om te beoordelen of natuurlijke personen in aanmerking komen voor overheidsuitkeringen en -diensten, evenals voor het verlenen, beperken, intrekken of terugvorderen van dergelijke uitkeringen en diensten,"* Het is evident dat hier sprake is van een Al-systeem in de zin van art. 3 van de Richtlijn, echter dit systeem wordt niet gebruikt om te beoordelen of natuurlijke personen in aanmerking komen voor overheidsuitkeringen en diensten. Het adviseert slechts of een aanvrager als onderzoekswaardig bestempeld moet worden. Bij onderzoekswaardigheid zal altijd sprake zijn van betekenisvolle menselijke tussenkomst.

2.16. Indien er sprake is van een (potentieel) hoge risico KI-toepassing, is de Autoriteit Persoonsgegevens geïnformeerd over de intentie om de KI-toepassing te ontwerpen en te ontwikkelen?

Nee. Dit is geen hoog-risico-systeem nu het systeem uitsluitend een beoordeling maakt op onderzoekswaardigheid en nooit beoordeelt of een natuurlijk persoon in aanmerking komt voor een overheidsuitkering. Er is sprake van betekenisvolle menselijke tussenkomst indien een natuurlijk persoon als onderzoekswaardig wordt bestempeld.

2.17. Is er een afweging gemaakt of het inzetten of ontwikkelen van de KI-toepassing proportioneel is ten opzichte van de kosten en impact op de bedrijfsvoering?

Aan het begin van het project is er een uitgebreide verkenning gedaan op de werkprocessen van handhaving, het doel van het proces en van de afdeling, de knelpunten en mogelijke oplossingsrichtingen. Bij die oplossingsrichtingen is er gekeken naar de impact, haalbaarheid en dat het hebben van een goede juridische onderbouwing (waarin proportionaliteit een belangrijk onderdeel is), noodzakelijk is bij iedere oplossing. De oplossingsrichtingen zijn toen ook besproken met een privacy officer. Hierbij is niet enkel gekeken naar een algoritme, maar ook naar andere oplossingsrichtingen. Er zijn kritische vragen gesteld om te voorkomen dat we in de *Solutionist trap* vallen, waarbij er teveel wordt gedacht vanuit de (technische) oplossing. Ook is er rekening gehouden met het subsidiariteitsprincipe. Subsidiariteit betekent datje het minst ingrijpende middel moet inzetten om een bepaald doel te bereiken. In dit geval is daarom gekozen voor een model binnen het aanvraagproces in plaats van bij al uitgegeven uitkeringen. Op deze wijze kunnen we zoveel mogelijk onrechtmatige uitkeringen aan de voorkant voorkomen, waardoor schulden voorkomen kunnen worden. Voorkomen is een lichter middel dan handhaven in een bestaande situatie. Wij hebben voor een algoritme gekozen omdat dit meer elementen meeneemt dan een simpele beslisregel en daardoor preciezer is waardoor relatief minder aanvragen onnodig

onderzocht te worden. Er is bewust gekozen voor een minder complex algoritme, met een hoge uitlegbaarheid. Meer informatie over deze keuze is te lezen in de technische documentatie. Na het maken van deze keuze is een intensiever juridisch en ethisch proces gestart om alle aspecten van het product goed in kaart te brengen en de juiste afweging te maken. Toen is ook de benodigde documentatie opgesteld zoals een DPIA en dit KIIA.

Er zijn grote kosten gemoeid met het ontwikkelen en in beheer nemen van de KI-toepassing. De KI-toepassing draagt eraan bij dat wordt voorkomen dat publieke middelen worden verstrekt aan Amsterdammers die daar niet toe gerechtigd zijn om zo schulden te voorkomen en het stelsel betaalbaar te houden. Vandaar dat wij denken dat de KI-toepassing proportioneel is ten opzichte van de kosten en impact op de bedrijfsvoering.

Na afloop van de pilot zal a.d.h.v. de resultaten een kosten-baten analyse worden gemaakt als onderdeel van de evaluatie. Dit zal worden opgenomen in een business case, die zal worden gebruikt bij het maken van de GO/ NO GO beslissing voor implementatie.

2.18. Conclusie (uitvoeren KJ-assessment of niet?)

De KI-assessment moet gezien de gevoeligheid van het werkproces, de verwerkte data, de potentiële impact op de burgers en medewerkers en de voorzienbare risico's, worden uitgevoerd.

3. KI-toepassing, verantwoordelijkheden en betreffende werkproces

3.1 Beschrijf de toepassing en het doel van de toepassing.

Het doel van handhaving is om te voorkomen dat bijstandsuitkeringen worden verstrekt aan Amsterdammers die daar geen recht op hebben en zo voorkomen dat Amsterdammers met schulden komen te zitten doordat zij de ten onrechte verstrekte uitkeringen (deels) moeten terugbetalen. De KI-toepassing kan hier een bijdrage aan leveren. Dit is een legitiem doel omdat sociale zekerheid één van de pijlers is van de Nederlandse maatschappij en sociale zekerheid tevens een grote bijdrage levert aan de welvaart. Het sociale zekerheidsstelsel kan slechts volledig functioneren als eenieder die geen aanspraak heeft op de voorzieningen, daar ook geen gebruik van maakt. Bovendien bestaat de bekostiging van het stelsel volledig uit gemeenschapsgeld, waardoor fraude het draagvlak van het sociale zekerheidsstelsel aantast.
Samengevat verwachten we dat de werkwijze een aantal voordelen heeft:

- Gelijkwaardigheid
  Hoewel alle medewerkers dezelfde werkinstructie volgen, zijn er interpretatieverschillen als twee personen dezelfde situatie beoordelen. Modellen kunnen vooringenomen zijn, maar personen ook. Uit een uitgevoerde vooringenomenheid (bias) analyse is gebleken dat op vrijwel alle gevoelige attributen (zoals leeftijd, geboorteland en nationaliteit) het ontwikkelde model verschillende groepen gelijkwaardiger behandeld dan in het huidige proces.
- Proportionaliteit
  Omdat het model beter kan inschatten welke aanvragen door handhaving bekeken dienen te worden, worden relatief minder aanvragen ten onrechte onderzocht door een Medewerker Fraudepreventie en Controle W&I. Dit leidt tot een meer proportionele inzet van Handhaving. Ook kunnen meer onrechtmatige uitkeringen voorkomen worden, hetgeen leidt tot een meer preventieve in plaats van repressieve aanpak.
- Effectiviteit
  Het model kan beter inschatten welke aanvragen door handhaving bekeken dienen te worden dan de Inkomensconsulent. Bij eenzelfde aantal onderzochte aanvragen (4% van het totaal) stuurt de Inkomensconsulent zo'n 46% van de totale aanvragen terecht door naar Handhaving. Bij het model is dit 56%.
  De geschatte additionele opbrengst van deze nieuwe aanpak op jaarbasis is als volgt:
  - 125 minder ten onrechte verstrekte uitkeringen;
  - 125 minder Amsterdammers die potentieel met terugvorderingen en schulden te maken krijgen;

- o € 2-4 miljoen[5] besparing op het BUIG-budget aan de uitkering van onrechtmatige aanvragen;
- o € 2-4 miljoen[5] minder aan ten onrechte uitgekeerde bijstandsuitkeringen, die potentieel in terugvordering en schulden voor de Amsterdammer kunnen worden gezet.

Voor een overzicht van alle doelen en succescriteria zie het document **'Doelstellingen en succescriteria'.**

3.2 Beschrijf het werkproces (BPMn uitwerking en toelichting).

Het werkproces bestaat uit een aantal stappen, genomen door Medewerkers Vooronderzoek, Handhavingsspecialisten en Inkomensconsulenten. Het proces is beschreven volgens BPMn notatie in Figuur 1.

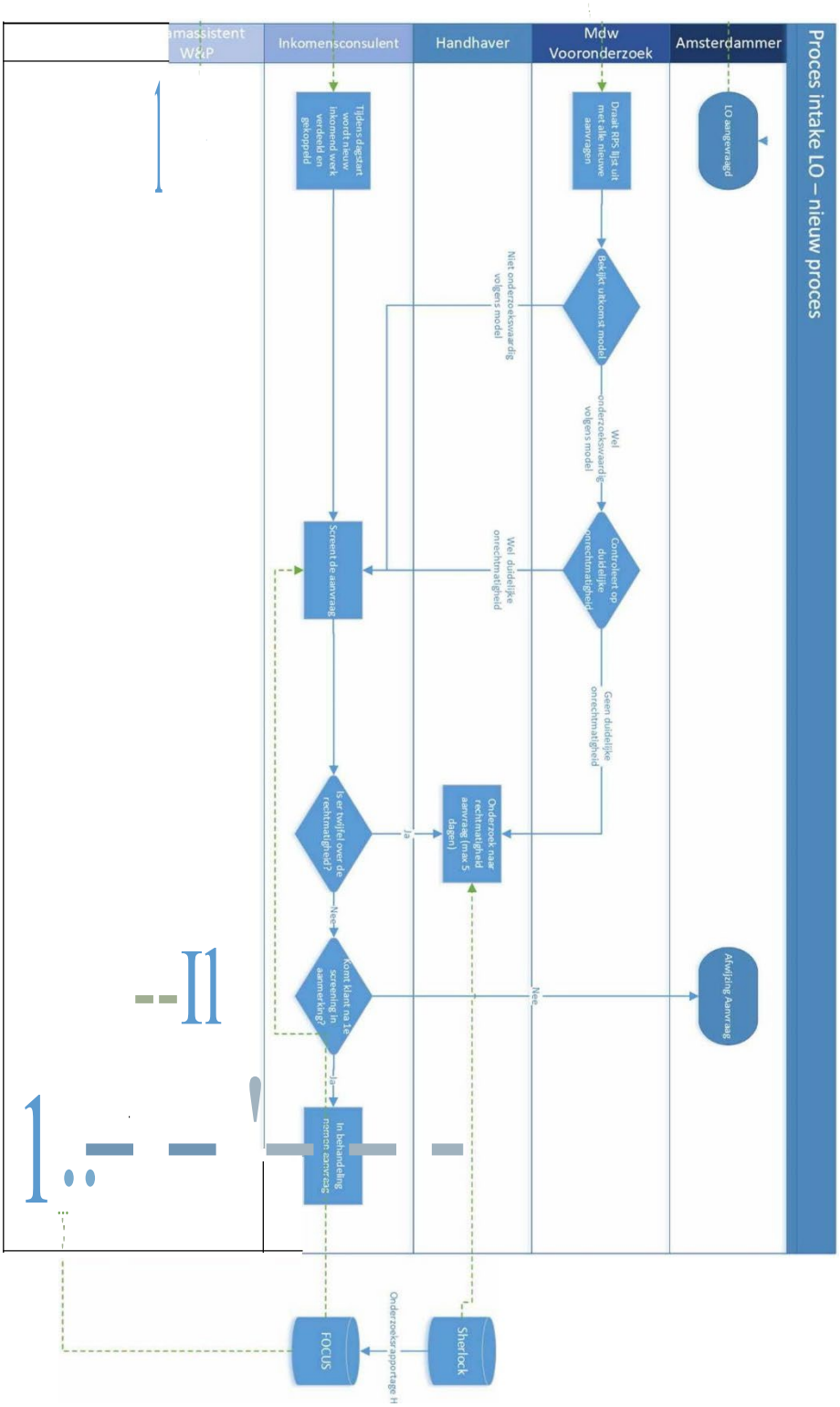
De Amsterdammer vraagt een uitkering aan via het aanvraagformulier: online via DigiD of fysiek via het stadsloket;
Het algoritme analyseert de onderzoekswaardigheid van de aanvraag;
De Medewerker Vooronderzoek draait via **RPS** (een management datasysteem) een lijst uit met daarop alle nieuwe aanvragen;
De Medewerker Vooronderzoek bekijkt de uitkomst van het model en de drie belangrijkste features die tot deze score hebben geleid in de Dataverkenner;
A. als de aanvraag niet onderzoekswaardig is volgens het model dan volgt stap 6; stap **Z** t/m 15 zijn dan niet van toepassing;
B. Als de aanvraag niet onderzoekswaardig is kan stap 6 worden overgeslagen en volgt

De Medewerker Vooronderzoek noteert op de **RPS** lijst dat de vraag niet onderzoekswaardig is volgens het model;
Medewerker Vooronderzoek controleert of er duidelijk sprake is van onrechtmatigheid, bijvoorbeeld omdat de aanvrager niet in Amsterdam woont, of het inkomen van de klant boven de norm is;
A. Indien duidelijk sprak is van onrechtmatigheid noteert de Medewerker Vooronderzoek op de lijst dat de aanvraag volgens het model onderzoekswaardig is, maar dat de aanvraag verder kan worden opgepakt door de Inkomensconsulent omdat die de aanvraag kan afwijzen- stap g t/m 14 kan dan worden overgeslagen
B. Indien geen sprake is van duidelijke onrechtmatigheid noteert de Medewerker Vooronderzoek dat de aanvraag volgens het model onderzoekswaardig is, en dat de aanvraag verder op kan worden gepakt door Handhaving - stap g t/m 14 volgen hierop;
Indien de aanvraag door de handhavingsspecialist dient te worden opgepakt voert de Medewerker Vooronderzoek een zaak op in Sherlock;
De Handhaver van de dag verdeelt het werk over de handhavingsspecialisten;
De toegewezen Handhavingsspecialist onderzoekt de aanvraag, waarbij de uitkomst van het model in de Dataverkenner het startpunt is;
Rapport en advies wordt overgedragen voor toetsing via Sherlock;
Een tweede Handhavingsspecialist toetst de behandeling en voert dit op in Sherlock;

---

5 Gebaseerd op SEOR (2016) *Effect Handhaving*

De tweede Handhavingsspecialist draagt dossier over aan inkomensconsulent met een consult;
Inkomensconsulent screent aanvraag- indien noodzakelijk wordt een nog niet door Handhaving bekeken aanvraag alsnog naar een Handhavingsspecialist doorgezet;
Inkomensconsulent bepaalt recht op uitkering
A. Als er recht is op uitkering, wordt de aanvraag toegewezen. Klant wordt overgedragen aan het beheerteam; vervolg participatieproces. Opslag in Socrates.
B. als er geen recht is op uitkering, wordt de uitkering afgewezen middels een gesprek. Einde aanvraagproces. Opslag in Socrates.



Figuur 1: Aanvraagproces in BPMn

3.3 Welke processtap/ taak wordt ondersteund met de KI?

Voor wat betreft de taak om aanvragen te selecteren voor al dan niet een nader onderzoek (geldt dat alle aanvragen door een medewerker worden bekeken, maar dat bij de door de KI geselecteerde aanvragen aanvullend onderzoek gedaan wordt door een handhavingsspecialist). Voor het volledige proces zie 3.2.

3.4 Worden bij deze processtap/taak beslisregels toegepast door medewerkers of systemen?

Ja, momenteel worden een aantal beslisregels toegepast om te bepalen of een aanvraag nader wordt onderzocht. Wanneer wordt overgegaan op de KI-toepassing komen deze beslisregels te vervallen.

Twee van deze beslisregels worden geautomatiseerd toegepast. Dit houdt in dat GALO analyseert wat per aanvraag de uitkomst van deze beslisregels, hetgeen daarna door een medewerker gecontroleerd wordt. Deze zijn:

> Als er medebewoners worden gesignaleerd op het adres en is er mogelijk sprake van een gezamenlijke huishouding of medebewoning, dan is dit een signaal voor onderzoekswaardigheid;
>
> Als de persoonsgegevens van de aanvrager niet overeen met de **BRP,** dan is dit een signaal voor onderzoekswaardigheid.

Een derde beslisregel wordt niet geautomatiseerd, maar wel handmatig toegepast en als aanvulling op de voorgaande business rules. Deze is:

> Als één of meerdere van de onderstaande omstandigheden worden geconstateerd, dan is een zaak onderzoekswaardiger:

- Klant is eerder afgewezen vanwege de woon- en leefsituatie
- Klant heeft geen huurcontract, hoeft geen huur/ geen kostgeld te betalen
- Klant heeft een postadres/ correspondentieadres (niet zijnde bewindvoerder)
- Klant geeft aan niet te verblijven op het opgegeven uitkeringsadres
- Woning heeft minder slaapkamers dan er volwassenen geregistreerd staan
- Klant vermeldt niet alle geregistreerde bewoners
- **(NB:** sluit wel uit dat het niet om de ouders gaat/ dan wel dat klant in het algemene invulveld heeft aangegeven wie er op het adres wonen!)
- Klant had voor deze aanvraag een Daklozenuitkering
- Twee maanden voor aanvraag is sprake van uitschrijvingen van het adres
- Klant woont samen in met iemand anders in een 1-kamerwoning
- Er is sprake van een gezamenlijk huurcontract
- Partners zijn uit elkaar, echter partner woont nog op het adres
- Klant geeft op geen eigen slaapkamer te hebben en/of op de bank te slapen
- Klant geeft aan op verschillende adressen te slapen
- Twee volwassenen wonen op een adres:

- Mogelijk is sprake van een gezamenlijke huishouding.
- (NB: Uitzonderingen: - 1e graads en - zorgfunctie)
- De woning heeft een kantoorfunctie in plaats van een woonfunctie
- Klant geeft zich op zich niet op het uitkeringsadres te mogen inschrijven

Daarnaast zijn er een aantal beslisregels die handmatig worden toegepast door de Medewerkers Vooronderzoek, welke bovenstaande beslisregels overschrijven en waardoor de aanvraag als niet onderzoekswaardig beschouwd wordt:

- Alleenstaande, alleen ingeschreven op adres
- 1e graads inwonenden, alleenstaande ouder
- Klant is eerder met werk uit de uitkering gestroomd en klant vraagt weer aan omdat klant geen werk/inkomsten meer heeft
- Gezin met kinderen
- Verblijf in inrichting
- Begeleid wonen
- BD klanten
- IOAW/IOAZ klanten
- Ex-gedetineerden

Op het moment dat op basis van de beslisregels geen aanleiding bestaat voor een nader onderzoek, kan een medewerker (inkomensconsulent) er te allen tijde zelfstandig voor kiezen om een aanvraag door te sturen naar de afdeling Handhaving wanneer hier behoefte aan is. Dit kan bijvoorbeeld het geval zijn wanneer er een complexe situatie is waarbij aanvullend onderzoek vereist is.

Op het moment dat door toepassing van de beslisregels een aanvraag als onderzoekswaardig wordt geclassificeerd, dan controleert een medewerker vooronderzoek of dit een correcte conclusie is vóórdat een aanvraag voor nader onderzoek door de handhavingsspecialist wordt geselecteerd.

3.5 Welke verantwoordelijkheden zijn er als het gaat om het verwerken van de gegevens?

Tabel 2: Overzicht verantwoordelijkheden gegevensverwerking

Verwerkersverantwoordel ij ke (persoons)gegevens en eigenaar van de KI	Gemandateerd Directeur Cluster Sociaal: Renger Visser
Gegevensverstrekker data uit (zaak)systemen aan gegevensverwerker	Gemandateerd Directeur Cluster Sociaal: Renger Visser
Gemeentelijke verstrekker Basisregistraties aan Cluster Sociaal in zaaksysteem	Voorheen directeur Basisinformatie: Dignar van Onzenoort, na reorganisatie directeur directie Data: Marleen Gebraad
Gemeentelijke verstrekker Basisinformatie (BRP-gegevens) aan Team Advanced Analytics ten behoeve van onderzoek naar	Directeur Basisinformatie: Dignar van Onzenoort

onbewuste vooringenomenheid (bias analyse)	
Gegevensverwerker (ontwikkeling product) in de hoedanigheid van intern leverancier	Directeur Digitalisering en Innovatie: Ger Baron

3.6 Wie zijn betrokken bij de ontwikkeling van de KI?

Zie bijlage VII van de DPIA met overzicht rollen en verantwoordelijkheden.

3.7 *Welke verantwoordelijkheden en taken heeft ieder rol?*

Zie bijlage VII van de DPIA met overzicht rollen en verantwoordelijkheden.

3.8 *Beschrijf welke actoren een rol spelen in de toepassing.*

Zie bijlage VII van de DPIA met overzicht rollen en verantwoordelijkheden.

3.9 Beschrijf welke KI-technologie wordt ingezet om het doel te bereiken.

Er wordt gebruik gemaakt van een KI die is ontwikkeld om de onderzoekswaardigheid van een uitkeringsaanvraag in te schatten. Er is gekozen om een Explainable Boosting Machine algoritme te gebruiken en deze is "voorspellend" van aard.

Een Explainable Boosting Machine is een machine learning algoritme voor classificatie- en regressievraagstukken. Het kan niet-lineaire relaties in data beschrijven en maakt daarvoor gebruik van een "tree-based supervised learning" techniek. Daarmee is het algoritme redelijk complex van aard en geschikt om voorspellingen te doen, kansen op gebeurtenissen te schatten, verdachte gebeurtenissen te detecteren of veranderingen in een dataset te ontdekken. Het algoritme is geen "black box", want het is transparant hoe de features tot een bepaalde voorspelling leiden. Voor alle features wordt op dezelfde wijze op basis van de data bepaald welk effect de feature heeft.

Het ontwikkelde algoritme genereert op basis van de informatie uit de aanvraag en andere beschikbare informatie over de aanvrager een score die de onderzoekswaardigheid van de aanvraag aangeeft. Deze onderzoekswaardigheidsscore wordt slechts berekend voor nieuw binnenkomende aanvragen en het algoritme creëert dus geen "nieuwe" te onderzoeken zaken op basis van lopende uitkeringen. Daarnaast worden sowieso alle aanvragen beoordeeld. Het model zorgt er dus niet voor dat sommige aanvragen niet meer bekeken worden, maar draagt slechts bij aan het bepalen van de noodzaak om een diepgaand nader onderzoek te doen in plaats van een minder diepgaande screening.

De onderzoekswaardigheidsscore is het gemiddelde van een zeer groot aantal berekeningen (middels beslisbomen) dat het model heeft uitgevoerd. Deze score kan worden gebruikt om te bepalen of een aanvraag beter kan worden beoordeeld met een uitvoerig onderzoek of met een minder diepgravende screening. Het algoritme geeft ook uitleg over hoe de score tot stand is

gekomen. Deze informatie is om meerdere redenen wenselijk, maar onder andere om in het onderzoek gericht te kijken naar de door het algoritme geïdentificeerde belangrijke factoren.

Zie voor meer technische details over het algoritme de technische documentatie. Informatie over de verwerkte data en de bronsystemen zie "Overzicht data die op Azure verwerkt wordt".

3.10 Is de inzet van de KI-toepassing een effectief middel om *het doel te bereiken?*

Betoogd kan worden dat het algoritme een effectief middel is om te ondersteunen bij het bepalen van de onderzoekswaardigheid van ingekomen aanvragen. Het beoogde resultaat met het ontwikkelen van de KI is dat er aan het begin van het proces, aan de hand van historische data, een effectieve, efficiëntere werkwijze wordt gehanteerd, waarbij een inschatting kan worden gemaakt welke klanten snel een uitkering kunnen ontvangen zonder de inzet van meer capaciteit binnen de organisatie. Daarnaast kan het algoritme ook helpen een gelijke behandeling te bevorderen en het proces transparanter te maken. Wanneer verschillende mensen naar soortgelijke zaken kijken, kan het zijn dat de één een hier anders naar kijkt dan de ander. Daarbij komt dat niet alleen algoritmes, maar ook mensen bepaalde onbewuste bias kunnen hebben. Door de inzet van een algoritme dat goed getest is op mogelijke bias kan het proces dus juist eerlijker en gelijkwaardig worden. We kunnen het expliciet hebben over de variabelen op basis waarvan we de onderzoekswaardigheid willen bepalen. En op basis van de historische data worden er waarden gegeven aan die variabelen. Ook is het bij iedere aanvraag duidelijk hoe de score op onderzoekswaardigheid tot stand is gekomen. Zo krijgen we minder willekeur in het proces en worden gelijke gevallen gelijk behandeld.

Door de inzet van het model kan aan de hand van de reeds beschikbare informatie en data binnen de organisatie gerichter onderzoek verricht worden. (Zie het document "Overzicht dat adie op Azure verwerkt wordt wordt".) Dit draagt bij aan het hoofddoel namelijk alleen die klanten een uitkering te verstrekken die hier daadwerkelijk ook recht op hebben, zoals een goed werkend sociaal zekerheidsstelsel betaamd. Uit heersende jurisprudentie blijkt dat de effectiviteit van een instrument als deze niet vooraf vast hoeft te staan om te voldoen aan de eis van een 'pressing social need'. Dit betekent echter niet dat gemeente Amsterdam de effectiviteit van het model niet wil onderbouwen.

De effectiviteit van het model is gemeten tijdens de pre-pilot. Hier zagen we dat het huidige proces een hit-rate heeft 46% (bij 4% doorgestuurde aanvragen), terwijl het model met de in de pre-pilot gekozen drempelwaarde een hit-rate heeft van 57% (bij 9% doorgestuurde aanvragen. Dit is dus een verbetering van tien procentpunt in de hit-rate, ook terwijl er meer aanvragen door handhaving onderzocht worden.[6] Dit betekent een verlaging in het aantal ten onrechte verstrekte uitkeringen.Na de officiële pilotfase wordt de daadwerkelijke effectiviteit van het model in de praktijk gemeten aan de hand van een validatieplan.

[6] Voor meer informatie zie de pre-pilot resultaten

4. Beschrijf de baten van de KI-toepassing

4.1. Watzijn de baten voor de organisatie?

In het aanvraagproces worden alleen die aanvragen nader onderzocht als daar ook een concrete aanleiding toe is. Dat levert een efficiëntere inzet van de beschikbare capaciteit op;

Het aanvraagproces wordt sneller doorlopen want onderzoekswaardige aanvragen hoeven niet eerst langs de inkomensconsulent en dan pas naar de handhavingsspecialist;

Er ontstaat meer kwaliteit bij de uitvoering van het werkproces doordat medewerkers meer informatie tot zich kunnen nemen en het werkproces meer wordt gestandaardiseerd.

4.2. Watzijn de baten voor het individu?

- Door een betere toewijzing voorkomen we het ontstaan van schulden als gevolg van een ten onrechte verstrekte uitkering;
- Het proces wordt transparanter en beter uitlegbaar. We kunnen het expliciet hebben over de variabelen op basis waarvan we de onderzoekswaardigheid willen bepalen. En op basis van de historische data worden er waarden gegeven aan die variabelen. Ook is het bij iedere aanvraag duidelijk hoe de score op onderzoekswaardigheid tot stand is gekomen;
- Er ontstaat minder willekeur in het proces;
- Gelijke gevallen worden gelijk behandeld;
- Er kan een biasanalyse worden uitgevoerd om vooringenomenheid te voorkomen;
- Een Amsterdammer wordt in mindere mate 'onnodig' belast met nader onderzoek, vragen en bijbehorende stress;
- Doorlooptijd van het behandelen van de aanvraag wordt verkort door een aanvraag direct door de juiste partij te laten onderzoeken;
- Een hogere onderzoekswaardigheidsscore kan er ook op duiden dat iemand meer hulp nodig heeft dan alleen een aanvraag levensonderhoud. Er kan daardoor sneller worden geschakeld met de ketenpartners.

4.3. Wat zijn de baten voor de maatschappij als geheel?

Door een rechtmatige toekenning van gemeenschapsgeld wordt het sociale zekerheidsstelsel niet aangetast en komt het geld terecht waar het terecht zou moeten komen;

Daardoor houden we ook maatschappelijke steun voor deze voorziening;

Omdat het om grote aantallen (aanvragen bijstandsuitkering) gaat binnen de gemeente Amsterdam, is geconcludeerd dat er naar verwachting een gezonde input/output ratio is. Elke aanvraag voor levensonderhoud die terecht wordt geweigerd zorgt ervoor dat een grote hoeveelheid financiële middelen en capaciteit wordt bespaard.

5. Architectuur

5.1. Hoe ziet de architectuur eruit voor het gebruik van de KI?

Voor informatie over de architectuur zie de architectuurnotitie.

5.2. Wat zijn de belangrijkste aandachtspunten die zijn gesignaleerd bij de beoogde doelarchitectuur?

Voor informatie over de architectuur zie de architectuurnotitie.

6. Data en data governance

6.1. Beschrijf welke brondata worden gebruikt binnen het werkproces en in het kader van de toepassing.

Socrates is het financiële klantvolgsysteem dat als zaaksysteem wordt gebruikt door de Inkomens Consulenten. Socrates is gekoppeld aan de BRP. Aan het gebruik van de BRP-data ligt een transparantiedocument ten grondslag.

Sherlock is het systeem waarin Handhaving de onderzoeken naar onrechtmatig verstrekken van de uitkering/fraude verwerkt (Rechtmatigheid).

Raak is het klantvolgsysteem inzake Doelmatigheid. In dit systeem worden alle onderzoeken in het kader van re-integratie verwerkt door de klantmanagers en is inzichtelijk voor handhaving als zijnde leesautorisatie.

Galo is het aanvraagportaal, daar komen alle digitale aanvragen van de klant binnen. In Galo wordt beoordeeld middels een eenvoudige beslisboom welk spoor de uitkering gaat lopen, groen oranje of rood. Groen betekent niet onderzoekswaardig, oranje betekent onduidelijk of onvolledig en rood betekent onderzoekswaardig. 95% van de aanvragen loopt het oranje spoor, omdat de aanvragen onduidelijk en onvolledig zijn. Galo wordt echter uitgefaseerd. Het wordt op dit moment nog minimaal gebruikt en deels vervangen door dit model.

Voor een volledig overzicht van de verwerkte data zie het Overzicht Verwerkte Data en Features.

6.2. Heeft de gemeentelijke organisatie volledige controle en beheersing (eigenaarschap) over de gebruikte data voor de KI?

Ja, de gemeente heeft de volledige controle en beheersing over de gebruikte data voor het model en het model zelf. Het model is door de gemeente zelf ontwikkeld. Externe medewerkers hebben ondersteund bij de ontwikkeling door de gemeente. Zij hebben een geheimhoudingsverklaring getekend. Interne medewerkers van de gemeente hebben de regie over de ontwikkeling van het model gehad. Ook alle gebruikte data komen uit bronnen binnen de gemeente.

6.3. Beschrijf de reikwijdte van de gegevens in de gegevenssets die worden onderzocht (bijv. of ze persoonsgegevens bevatten)?

Zie antwoorden bij vraag 2.9 en 6.1 en het Overzicht Verwerkte Data en Features.

6.4. Wat is de bewaartermijn van de gegevens en voor verwerking ten behoeve van de KI-toepassing?

De vakgroep informatiebeheer heeft een werkgroep Bewaartermijnen en Algoritmes opgetuigd. Deze werkgroep is geconsulteerd en heeft op 27 juni 2022 een advies uitgebracht aangaande de omgang met bewaartermijnen van data en algoritmes. Voor de gedetailleerde informatie over de toepassing van de bewaartermijnen zie het document 'Bewaartermijnen - algoritme onderzoekswaardigheid wpi'.

Voor de ontwerpdocumentatie geldt een onbeperkte bewaartermijn in het stadsarchief. Dit vloeit voort uit het cultureel/historisch belang van het algoritme als een van de eerste algoritmes dat door Amsterdam is ontwikkeld. Voor de beheerdocumentatie en andere documentatie geldt een bewaartermijn van 5 jaar. Dit vloeit voort uit hoofdstuk 15.1.1. van de selectielijst. Logging dient minimaal 6 maanden bewaard te blijven op grond van hoofdstuk 19.1.23 van de Selectielijst. Logging kan ook langer bewaard worden op basis van risicoanalyses.

lnputdata (o.a. trainingsdata) heeft dezelfde bewaartermijn als de ontwerpdocumentatie, namelijk blijvend bewaren. Hetzelfde geldt voor de logica (code+ datamodel). De output moet bewaard worden zo lang als het klantdossier ofwel 10 jaar na laatste klantcontact. Hieraan gekoppeld dient ook de gebruikte versie van het algoritme voor deze tijd bewaard te worden.

6.5. Hoe is de data geleverd en verzameld?

De verwerkte data zijn afkomstig uit het werkproces (zie 6.1). Er is geen data apart verzameld ten behoeve van dit project. Er is eerst met verschillende mensen binnen het WPl-proces gepraat die uitvoerend werk doen, in het management zitten of die juist veel kennis hebben van de data in het WPl-proces. Op basis hiervan is een dataset gemaakt om mee te experimenteren. Er is toen onderscheid gemaakt over gegevens die we sowieso niet mogen gebruiken (zoals nationaliteit) en dus ook niet onderzocht hoeven te worden, gevoelige gegevens waarvan we de wenselijkheid van gebruik verder moeten onderzoeken en gegevens die in orde waren om te gebruiken. Tijdens het ontwikkelproces zijn uit de laatste twee groepen nog veel gegevens afgevallen, omdat wij het niet wenselijk vonden om te gebruiken of omdat het geen waarde had voor het model. Het uiteindelijke resultaat van gebruikte data is te vinden in het Overzicht Verwerkte Data en Features.

Voor het trainen van de KI is gebruik gemaakt van eenmalige, gedeeltelijke data dumps van de bronsystemen Socrates, Sherlock, Raak en Galo. Deze dumps zijn op de VAO geplaatst en daar verder verwerkt. In een later stadium zijn dezelfde dumps verplaatst naar de Analyse Service van de gemeente Amsterdam op Azure in het kader van de migratie naar Azure. Meer informatie hierover is te vinden in het document "Overzicht van data die op Azure verwerkt wordt".

Bij inzet van de KI wordt de benodigde data opgehaald uit een speciaal voor dit doeleinde gecreëerde SOL-database op Azure. Deze database wordt dagelijks gevoed met de benodigde data, dus de nieuwe aanvragen en bijbehorende informatie uit het data warehouse (DWH) van WPI.

6.6. *Welke afspraken zijn er gemaakt over de levering en verzameling van de data?*

De ruwe data wordt eerst verwerkt door een data engineer, voordat de data scientist deze te zien krijgt. Dit zijn twee verschillende personen. De verwerking bestaat uit het hashen van gevoelige

data volgens SHA-256 methodiek (zie document "Overzicht data die op Azure verwerkt wordt"). Alleen de gehashte data is beschikbaar gemaakt voor de data scientists. De data wordt daarnaast alleen gebruikt ten behoeve van het ontwikkelen en inzetten van de in dit document beschreven **KI.**

Daarnaast is er voor de data levering al een selectie gemaakt over welke data we sowieso niet gaan gebruiken. Die data is ook niet geleverd. Daarna is nog, tijdens het onderzoek, veel data afgevallen omdat het niet wenselijk was om te gebruiken te worden omdat het bijvoorbeeld t te gevoelige ofte subjectieve data is. Of omdat de data niet veel toevoegt aan de prestaties van het model.

6.7. *Wat is de kwaliteit en kwantiteit van de beschikbare data?*

Dit is aan de start van het proces uitvoerig onderzocht. Zie de technische documentatie.

6.8. Is de data passend voor de KI-toepassing? Zo ja, waarom?

Ja, de data die is verwerkt is noodzakelijk om het werkproces uit te voeren. Hierop is toegezien door de gebruikte features kritisch te bespreken in een aantal feature validatie sessies. We hebben ongeveer tien sessies gehouden om tot de finale feature set te komen. Bij deze sessies waren twee handhavingsspecialisten uit de werkgroep aanwezig, een jurist, twee privacy officers en een adviseur gegevensbescherming. In deze sessies hebben we ons gefocust op (in willekeurige volgorde) (i) business logica; (ii) wet- en regelgeving; (iii) verbanden zoals aangetoond in de data; en (iv) privacy. Tijdens deze sessies is onder andere gebruik gemaakt van Partial Depence Plots (PDPs), waarin in detail wordt weergegeven wat de bijdrage voor elke mogelijke waarde van een feature is aan de onderzoekwaardigheid. Deze PDPs en het overzicht van de features zijn gedocumenteerd in het *Overzicht Verwerkt Data en Features.* De deelnemers aan de feature validatiesessies hebben er met behulp van de PDPs op toegezien dat gemodelleerde verbanden niet onlogisch of onwenselijk zijn. Ook hebben we de features in een aparte sessie voorgelegd besproken met twee CPA leden, en de finale set besproken met de gehele CPA. De finale feature set is tot slot formeel vastgesteld door de Business Owner. Er is in sommige gevallen voor gekozen om een deel van de beschikbare data niet te verwerken vanwege een gebrek aan noodzakelijkheid en passendheid. Om dezelfde reden is bovendien niet alle data opgevraagd die in de systemen staat.

6.9. Heeft de data passende 'statistische' eigenschappen? Dit geldt zowel voor de individuele dataset alsook voor een combinatie van datasets.

Ja, want het model is in staat om onderzoekswaardige zaken te onderscheiden van niet-onderzoekswaardige zaken. De exacte effectiviteit zal nog uit de pilot blijken. Daarnaast is de datakwaliteit gecontroleerd, zoals beschreven is in de technische documentatie.

6.10. Zijn er relevante aannames, met betrekking tot de informatie die de gegevens geacht worden te meten en te duiden?

We nemen aan dat de gebruikte data correct is verzameld en representatief is voor toekomstige uitkeringsaanvragen in Amsterdam. De trainingsdata komt uit hetzelfde proces als waar het

algoritme ingezet wordt en er vinden geen subselecties op plaats die de representativiteit (representation bias) beïnvloeden.

In het label wordt er verder vanuit gegaan dat zaken die in het verleden niet onderzoekswaardig gevonden werden door de handhavingsspecialisten, dat daadwerkelijk niet waren. Deze aanname is hoogstwaarschijnlijk niet 100% juist, want er zullen altijd per ongeluk onrechtmatige aanvragen worden toegekend, maar het is een noodzakelijke aanname om het model te kunnen ontwikkelen. Doordat het onderzoek door de handhavingsspecialist het meest uitgebreide onderzoek is dat binnen de gemeentelijke processen op aanvragen levensonderhoud gedaan wordt vinden we dit een redelijke aanname.

Een andere belangrijke aanname is dat het door ons gecreëerde label voldoende het begrip 'onderzoekswaardigheid' beschrijft. Dit wordt in de technische documentatie in detail beschreven.

Een punt van aandacht is dat de trainingsdata alleen aanvragen omvat die al onderzocht zijn, omdat voor die aanvragen labels beschikbaar zijn, terwijl het algoritme ingezet zal worden op alle nieuwe aanvragen. Het is de verwachting dat deze groepen iets van elkaar afwijken, in het bijzonder is te verwachten dat de trainingsdata gemiddeld genomen 'onderzoekswaardiger' is dan een gemiddelde aanvraag. Tijdens de pre-pilot van de handhavingsspecialisten hebben we gezien dat het model alsnog goed presteert op deze aanvragen. Bij een gelijkwaardige drempelwaarde had het model in deze pre-pilot een hit-rate van 58% op aanvragen die in het oude proces ook al door handhaving werden onderzocht (en dus overeenkomen met de trainingsdata), terwijl het een hit-rate van 64% had op aanvragen die in het verleden niet als onderzoekswaardig werden beschouwt.

We nemen aan dat de trainings-, validatie- en testdata representatief is voor hoe keuzes en beslissingen in de toekomst worden genomen. Met andere woorden, we gaan ervan uit dat de correlaties tussen de features en het onderzoekswaardigheidslabel op grote lijnen gelijk blijft. Zie hiervoor ook 6.15.

Daarnaast nemen we aan dat de input data correct is, en dat de features op de juiste wijze geformuleerd zijn. Voor onderbouwing van deze aannames, zie de technische documentatie 4.3 met betrekking tot de datakwaliteit.

6.11. Zijn er 'data gaps' of tekortkomingen in de data? Zo ja, kunnen er maatregelen worden troffen om dit te corrigeren? Welke maatregelen zijn dit?

Er is veel data die in het werkproces wordt geraadpleegd, die niet uit de systemen komt die tot dusver zijn ontsloten, bijvoorbeeld informatie van Suwinet. Om de logica uit het werkproces volledigere te kunnen verwerken, zou het nodig zijn dit soort bronnen te ontsluiten. Hiervoor is (nog) niet gekozen, omdat we zo min mogelijk data willen gebruiken om de gewenste waarde te creëren. Daarnaast kost het ook veel werk om nieuwe bronnen te ontsluiten. De waarde van het model eerst getest zal worden in een pilot, voordat men begint aan het ontsluiten van andere databronnen.

6.12. Zijn passende maatregelen genomen om categorieën persoonlijke data te beschermen? Zoja, welke maatregelen zijn die (bv: encryptie, pseudonimisatie, aggregatie, anonimisatie)

Ja, de data wordt gepseudonimiseerd. Zie voor details het document "Overzicht data die op Azure verwerkt wordt" kolommen 'Pseudonimiseren' en 'Pseudonimiseren details'. Dit geld voor zowel in de ontwikkel als in de productieomgeving.

6.13. Welke datavoorbereiding- en verwerkingsmethodieken zijn toegepast (annotatie, labelling, cleaning, aggregatie)

De data is opgeschoond o.a. door duplicaten te verwijderen, datums op te schonen, hoofdletters te verwijderen uit stringvelden en missende waardes op te vullen. Verder zijn er filteringen en aanpassingen gedaan om de datakwaliteitsproblemen, beschreven in de technische documentatie, zoveel mogelijk op te lossen.

Daarnaast zijn alle features gecreëerd op aanvraagniveau. Hiervoor was het nodig om bijvoorbeeld meerdere rijen met informatie over het inkomen van de aanvrager te aggregeren naar het aanvraagniveau door ze bij elkaar op te tellen. Vergelijkbare transformaties zijn uitgevoerd op andere informatie. De volledige lijst van features is te vinden in het Overzicht Verwerkte Data en Features. Zie de technische documentatie voor een beschrijving van het label en meer details over de dataverwerking.

6.14. Zijn er afspraken gemaakt over de bewaartermijnen van de verwerkte data en de verwijdering/archivering van de data als de termijn is verlopen of de data niet meer noodzakelijk is voor de ontwikkeling van de KI? Zo ja, wat zijn deze afspraken?

Ja er zijn afspraken gemaakt over bewaartermijnen. Zie voor het antwoord vraag 6.4.

6.15. Is de training-, validatie- en testdata relevant, representatief,foutloos en compleet?

Van de relevante soorten uitkeringen bevat de data alle aanvragen die gedaan zijn in de periode die gebruikt wordt, dus ook van kwetsbare groepen. In die zin is de data representatief voor alle groepen (hetgeen *representation bias* voorkomt). Het is wel zo dat alleen aanvragen konden worden gelabeld die in het verleden door handhaving zijn onderzocht. Het is mogelijk dat dit de representativiteit van de data voor sommige groepen vermindert, omdat de beslissing om te gaan onderzoeken is genomen door een inkomensconsulent en dus niet 100% objectief is. Er is daarom een afweging gemaakt tussen de betrouwbaarheid van de labels voor aanvragen (en het voorkomen van *measurement bias)* die niet door handhaving zijn onderzocht, en de mogelijk betere representativiteit als deze aanvragen meegenomen worden. Omdat de beslissing om te onderzoeken ondersteund wordt door schriftelijke werkinstructies met als doel standaardisatie en omdat er weinig vertrouwen was in de labels, is ervoor gekozen om niet-onderzochte aanvragen weg te laten. We zijn ervan overtuigd dat de data desondanks representatief genoeg is, met als extra check daarop ook nog de bias analyse. Naar aanleiding van de bias analyse is een hertraining van het model gedaan waarbij de diverse subgroepen op eigen wijze gewogen werden (zie bias

analyse documentatie voor details), onder andere om het risico op *aggregation bias* en *learning bias* te minimaliseren.

Zie de technische documentatie voor een beschrijving van (keuzes rondom) de datakwaliteit. We gaan ervan uit dat de training-, validatie- en testdata hierdoor voldoende compleet en foutloos is om het algoritme mee te ontwikkelen.

De betekenis en de juistheid van de data is daarnaast doorlopend gecontroleerd met medewerkers van de afdeling WPI in de vorm van een werkgroep, bestaande uit de projectleider, handhavingsspecialisten en inkomensconsulenten. De relevantie van de features is gecontroleerd in een beoordelingsgroep voor de features, bestaande uit onder andere juristen, handhavers, inkomensconsulenten, data scientists en beleidsmakers.

6.16. Zijn gedurende de training van de KI-toepassing de uitkomsten en relevante ontwikkelkeuzes geëvalueerd en gedocumenteerd?

Tijdens het ontwikkelen van het model is de gangbare manier van werken gebruikt dat na iedere verandering aan het model en bij iedere nieuwe feature is gekeken of dit het model verbetert. Veranderingen zijn uiteindelijk alleen doorgevoerd als dit het geval bleek. De ontwikkel keuzes zijn verder gedocumenteerd in de technische documentatie.

Daarnaast is ook in de gehele ontwikkelperiode reviews geweest met een werkgroep. Hierin zaten verschillende mensen uit het proces en management. Veel ontwerpkeuzes zijn uitvoerig met hen besproken om echt de juiste keuze te maken.

6.17. Zijn er duidelijke criteria/parameters vastgesteld voor het correct functioneren van de KI?

Ja, er zijn succescriteria vastgesteld. Zie hiervoor het pilotplan en het meetplan. Dit wordt gedurende de pilot met regelmaat gemonitord. Voor de ontwikkeling van het model zijn er doelstellingen en succescriteria opgesteld met de werkgroep. Dit heeft de richting bepaald van de ontwikkeling van het model.

6.18. Welke methoden worden gebruikt om te verifiëren of de KI binnen de gestelde parameters blijft?

Op dit moment wordt niet continu gemonitord of de KI binnen de gestelde parameters blijft. Het is gepland dat dit geïmplementeerd wordt, zodra het model daadwerkelijk in productie genomen wordt. Dat zouden bijvoorbeeld automatische notificaties kunnen zijn wanneer een van de waardes uit vraag 6.17 zich sterk verandert. Zie voor meer details het beheerplan.

Tijdens de pilotfase zal wel periodieke monitoring plaatsvinden. Dit bestaat onder andere uit een tussentijdse evaluatie waarin de prestaties van het model tot dat moment gemeten worden, en een eindevaluatie voor de GO/NO GO besluitvorming. Meer informatie hierover kan worden gevonden in het pilot plan.

6.19. Is de output van de KI consequent?

Onder consequent verstaan we dat bij gelijke inputgegevens dezelfde output wordt gegeven door het model.

Er zijn twee maatregelen genomen om ervoor te zorgen dat het model consequent handelt. De eerste maatregel is het implementeren van een zogenaamde 'random seed'. Dit betekent dat iedereen die het model traint met dezelfde dataset het model op dezelfde manier traint en op dezelfde waarden uitkomt. Verder wordt het getrainde algoritme opgeslagen in een zogenaamde 'pickle' file. Dit bestandstype zorgt ervoor dat het getrainde model in de praktijk dezelfde resultaten en handeling verricht. Daarnaast kan het model hierdoor overgedragen worden naar andere systemen en daar hetzelfde handelen.

6.20. *Welke output data genereert de KI?*

De KI geneert een onderzoekswaardigheidsscore tussen oen 1. Deze score hoort bij een uitkeringsaanvraag. Naast de score wordt ook het effect van iedere feature meegeleverd, dus of ze de score verhogen of verlagen. Hierdoor iste zien wat de belangrijkste features waren. De gebruikers zien echter niet de score, maar enkel of de aanvraag gelabeld wordt als onderzoekswaardig ja/nee. De drempelwaarde (voor de pilot ingesteld op 0.56) vanaf welke een aanvraag als onderzoekswaardig wordt gelabeld wordt bepaald in samenwerking met de business. Daarnaast zien de gebruikers ook de belangrijkste drie features voor die aanvraag.

6.21. Is het mechanisme van peer review toegepast? Zo ja, welke afspraken zijn gemaakt?

Ja, alle wijzigingen aan de code zijn door tenminste twee paar ogen gezien. Het proces werkt als volgt:

De schrijver van de code maakt een *merge request* **(MR)** aan en wijst deze toe aan een *assignee* en eventueel een of meerdere *reviewers.* De *assignee* is verantwoordelijk voor het reviewen en uiteindelijk mergen (of afwijzen) van de **MR.** Dit review proces wordt ook technisch afgedwongen in de Azure omgeving. De *reviewers* worden gevraagd om een review te geven. De maker van **MR** is verantwoordelijk voor het doen van aanpassingen op basis van de gegeven reviews. Reviews mogen gedaan worden vanuit Azure (of in het verleden Gitlab), de reviewer hoeft de code niet zelf te draaien. Notebooks hoeven niet gereviewd te worden.

6.22. *Welke relevante designkeuzes voor de KI-toepassing zijn gemaakt?*

Voor informatie over technische designkeuzes, zie de technische documentatie.

Naast het technische design is veel aandacht besteed aan het beoordelen van de passendheid van de features. De uiteindelijke selectie is opgenomen in het Overzicht Verwerkte Data en Features. Deze selectie is tot stand gekomen na een beoordelingsproces bestaande uit vele sessies verspreid over meerdere maanden. In deze sessies zijn met een diverse groep alle potentiële features besproken. Vertegenwoordigd waren o.a. de rollen handhavingsspecialist, inkomensconsulent, data scientist, business analist, jurist, privacy officer en adviseur gegevensmanagement.

Er is gekeken of de features herkenbaar zijn vanuit het huidige proces, wat hun belangrijkheid voor het model is, of er in wet- en regelgeving of het gemeentelijk beleid aanknopingspunten zijn voor het gebruik, of ze ethisch verantwoord zijn en in hoeverre ze een bias kunnen veroorzaken. Op basis hiervan zijn sommige features afgekeurd voor gebruik, andere zijn in principe goedgekeurd, maar vallen af vanwege een gebrek aan belangrijkheid of om andere technische redenen. De overgebleven, goedgekeurde features zijn opgenomen in het model.

6.23. ***Is er onderzoek gedaan naar onbewuste vooringenomenheid in de dataset en de KI-toepassing? Zo ja, wat zijn de uitkomsten hiervan en welke maatregelen zijn getroffen?***

Ja, zie de bias analyse documentatie.

7. Wettige en ethische toepassing

7.1. Watis de wettelijke grondslag voor de beoogde verwerking van de gegevens ten behoeve van het ontwikkelen, trainen en inzetten van de KI?

De grondslag voor deze verwerking is gestoeld op de noodzaak voor de vervulling van een taak van algemeen belang of van een taak in het kader van de uitoefening van het openbaar gezag dat aan de verwerkingsverantwoordelijke is opgedragen. Dit is artikel 6 lid 1 sub e van de AVG. Het betreft hier een taak van algemeen belang. Deze taak vloeit voort uit de Participatiewet en de Wet Suwi:

Artikel 7, 53a en 64 Participatiewet,
Artikel 62 wet SUWI

Zie de DPIA voor meer informatie.

7.2. *Welke waarden en belangen spelen een rol in de context van de KI-toepassing?*

Zie bijlage weging betrokken (grond)rechten en waarden bij inzet KI en de IAMA.
Daarnaast spelen de Tada waarden een belangrijke rol bij het ontwikkelen van nieuwe technologie binnen de Gemeente Amsterdam. De zes Tada waarden zijn:

- Legitiem en gecontroleerd: Amsterdammers, bezoekers en gebruikers hebben zeggenschap over de vormgeving van onze digitale stad.
- Open en transparant: Transparantie en openheid is de grondhouding van de Amsterdamse-datawaarden.
- Van en voor iedereen: Data die overheden, bedrijven en andere organisaties uit de stad en over de stad verzamelen, zijn voor gemeenschappelijk gebruik.
- Inclusief: Data dienen bij te dragen aan een inclusieve samenleving.
- Zeggenschap: Zeggenschap over data vergroot de invloed en vrijheid van Amsterdammers en bezoekers.
- Menselijke maat: Data dragen bij aan menselijke waarden en hebben nooit het laatste woord.

Meer informatie is te vinden op de website van Tada (www.tada.city).

7.3. *Hoe worden de waarden en belangen van betrokken actoren geborgd?*

Zie bijlage weging betrokken (grond)rechten en waarden bij inzet **KI.**

De hierboven beschreven Tada waarden zijn als volgt meegenomen in de ontwikkeling en in gebruik name van het algoritme:

- Legitiem en gecontroleerd: In het ontwikkelproces is advies opgehaald bij de Commissie Persoonsgegevens Amsterdam (CPA) en de participatieraad.

De gemeente Amsterdam hecht grote waarde aan het zorgvuldig omgaan met persoonsgegevens. Om dat te waarborgen heeft het gemeentebestuur een onafhankelijke commissie ingesteld, de Commissie Persoonsgegevens Amsterdam (CPA). De CPA adviseert over het privacy beleid van de gemeente en de uitvoering daarvan. Daarnaast adviseert de commissie de ambtelijke organisatie bij complexe en/of politiek gevoelige kwesties rondom persoonsgegevens. De CPA signaleert wat er binnen de gemeente speelt op het gebied van persoonsgegevens en brengt dit onder de aandacht van het gemeentebestuur en het management van de ambtelijke organisatie. In de Commissie Persoonsgegevens Amsterdam hebben personen van buiten de gemeentelijke organisatie zitting op persoonlijke titel. Er wordt gestreefd naar een goede balans tussen mannen en vrouwen, jong en oud, overheid en bedrijfsleven, en achtergrond (medisch, onderwijs, journalistiek, juridisch, technisch, etc.). Dit helpt om de rol van maatschappelijk klankbord goed te in te vullen. Er is gedurende het gehele ontwikkelproces meerdere keren contact geweest met de CPA. De eerste vergadering was op 12 september 2019 en de laatste is gepland op 25 augustus 2022. Tussendoor zijn er nog drie andere sessies geweest. De CPA is meegenomen door alle fases van de ontwikkeling en in de belangrijke besluiten en heeft daarop geadviseerd. Deze adviezen zijn voor ons zeer waardevol geweest en hebben bijgedragen aan de kwaliteit van het product.

De Participatieraad is bevoegd om gevraagd en ongevraagd de gemeente te adviseren over onderwerpen die cliënten van Werk Participatie en Inkomen (WPI) en Pantar betreffen. Het gaat om de voorbereiding, de uitvoering, de controle en de evaluatie van het gemeentelijk beleid op het gebied van werk, participatie en inkomen. De Participatieraad heeft 15 leden, waarvan 8 als vertegenwoordigers van belangenorganisaties en 7 op persoonlijke titel. Op 21 september 2021 hebben wij een eerste presentatie gegeven aan de participatieraad. Op dat moment hadden wij wel al een idee hoe het product eruit zou komen te zien maar er was ook nog veel ruimte om bij te sturen als dat nodig was. Wel was het na de eerste sessie direct duidelijk dat er meer tijd nodig was omdat het lastig was de complexe materie begrijpelijk over te brengen. Daardoor zijn er nog twee andere sessies geweest (op 28 september 2021 en 10 november 2021) van meerdere uren om hier verder over te spreken. Op 10 maart 2022 heeft de Participatieraad uiteindelijk een negatief advies uitgebracht. Op 18 juli 2022 is er een tussentijdse reactie naar de raad gestuurd waarin zij worden bedankt voor hun advies en wordt uitgelegd dat we, mede naar aanleiding van hun advies, een externe partij onderzoek willen laten doen naar het door ons ontwikkelde algoritme. Deze partij, Deloitte, zal adviseren over ofwe met dit model ethisch verantwoord een pilot kunnen starten.
Het advies van de Participatieraad heeft ons geholpen om het algoritme te verbeteren. Zo hebben wij een extra test gedaan voorafgaand aan de pilot, hebben wij de variabele 'geboortejaar' verwijderd en hebben we een validatie laten uitvoeren door een extern bureau. Daarnaast zijn wij van mening dat het advies op sommige punten niet in overeenstemming is met de daadwerkelijke situatie (zoals het gebruik van de data of

diensten van het Inlichtingenbureau). Graag zouden wij met de Participatieraad verder in gesprek gaan maar op dit verzoek is de Participatieraad tot op heden niet ingegaan.

Daarnaast zijn gedurende het gehele proces werkgroepsessies georganiseerd met een groep eindgebruikers, waaronder Medewerkers Vooronderzoek; Handhavingsspecialisten; en Inkomensconsulenten. Tijdens deze sessies is informatie opgehaald over de werkprocessen en over de wensen van de medewerkers m.b.t. het algoritme. Daarnaast zijn tweewekelijkse reviewsessies georganiseerd met een bredere groep stakeholders binnen de gemeente, waarin een update werd gegeven van de stand van zaken en belangrijke/ ethische besluitvorming samen gedaan werd.

- Open en transparant: Omdat het hier gaat om een handhavingsproces worden de onderliggende features en code niet openbaar gemaakt. Wel wordt het algoritme toegevoegd aan het algoritmeregister waar veel informatie wordt gegeven en wordt aanvullende informatie verstrekt in de privacyverklaring. Ook wordt er een O&A beschikbaar gesteld op www.amsterdam.nl/veelgevraagd, en worden de medewerkers van het Contact Center Amsterdam (CCA) goed geïnformeerd over dit onderwerpen zodat ze binnenkomende vragen van Amsterdammers zo volledig mogelijk kunnen beantwoorden. Meer informatie over zowel interne als externe communicatie is te vinden in het *Communicatieplan Pilot Informatiegestuurd Handhaven.* Rechterlijke en toezichthoudende instanties kunnen deze informatie desgevraagd wel krijgen. Ook zijn de features gedeeld met de CPA en de externe partij. Op geaggregeerd niveau zijn de features gedeeld met de Participatieraad.

- Van en voor iedereen: Omdat het hier gaat om privacygevoelige persoonsgebonden informatie zal de ontsloten data niet openbaar worden gemaakt. Rechterlijke en toezichthoudende instanties kunnen deze informatie desgevraagd wel krijgen.

- Inclusief: In het bestaande proces (voor toevoeging van het algoritme) wordt de onderzoekswaardigheid bepaald door de medewerker, met de hulp van een tweetal beslisregels. Hoewel er werkinstructies zijn voor het bepalen van de onderzoekswaardigheid valt een zekere willekeur hierbij niet te voorkomen. Het algoritme beoordeelt elke aanvraag op dezelfde wijze, en maakt het proces daarmee gelijkwaardiger. Daarnaast is een bias analyse uitgevoerd op het model om er zeker van te zijn dat het model geen nieuwe bias in het proces introduceert. Meer informatie hierover is te vinden de *Bias analyse documentatie.* Ook hebben wij geprobeerd dit product met een zo divers mogelijke groep te ontwikkelen zodat meerdere perspectieven ingebracht worden.

- Zeggenschap: Wij vinden het belangrijk om burgers en belangengroepen te betrekken bij de ontwikkelingen. Wel zijn wij nog heel erg zoekende in hoe dit op de juiste manier te doen. Binnen dit project hebben wij vooral de CPA betrokken en ook geprobeerd de Participatieraad op de juiste manier te betrekken. Wij denken dat de samenwerking met

de CPA goed is verlopen. Bij de Participatieraad zijn wij nog wel zoekende of dit, voor de burger en andere belanghebbende, de meest waardevolle vorm is van zeggenschap. Wij vonden het lastig om alle belangrijke aspecten van dit complexe product duidelijk over te brengen. Daarnaast was er, begrijpelijk, ook veel wantrouwen tegenover het inzetten van een algoritme door voorbeelden zoals de toeslagenaffaire en SyRI. In vervolgprojecten willen wij verder opzoek naar hoe wij het belang van betrokkenen en burgers zo goed mogelijk kunnen borgen binnen het ontwikkelen van dataproducten. Daarnaast blijven wij naar goede manieren zoeken om burgerparticipatie te waarborgen voor de verdere ontwikkeling van dit product.

- Menselijke maat: Het model geeft enkel een advies in de vorm van een onderzoekswaardigheidsscore voor een aanvraag aan de verantwoordelijke medewerkers. Er is geen sprake van geautomatiseerde besluitvorming. Zie 3.2.6.

7.4. Zijn de waarden en belangen geconcretiseerd in wet- en regelgeving?

Zie bijlage weging betrokken (grond)rechten en waarden bij inzet Kl.

7.5. Hoe wordt gedurende het ontwikkelproces getoetst aan deze waarden en belangen wat betreft de werking van de KI-toepassing?

Analyseren werkproces en informatiegebruik

Het werkproces van Toezicht en Handhaving is inzichtelijk gemaakt door het opstellen van aparte procesbeschrijvingen (BPMn) en door het uitvoeren van een DPIA. Dit zorgt voor een breder inzicht op het mogelijke effect bij de inzet van een algoritme en helpt om strategische keuze te maken.

Selecteren brondata

De data scientist heeft samen met de Functioneel Beheerder onderzocht welke data beschikbaar is binnen de beschikbare zaaksystemen. Daaruit is gebleken welke informatie op welke momenten uit welke bronnen worden geraadpleegd of wordt ontvangen. Deze data is inzichtelijk gemaakt en per kenmerk is aangegeven of deze al dan niet mag worden verwerkt ten behoeve van dit project. Dit heeft tot dataminimalisatie geleidt voor het ontwikkelen van het model, bijvoorbeeld door ervoor te kiezen kenmerken als nationaliteit, verblijfsstatus en geboortejaar niet te verwerken. Vervolgens is het WPD doorlopen en zijn de privacy en security officers van WPI gevraagd te adviseren op de te verwerken data. Dit heeft ertoe geleid dat uiteindelijk een groot aantal kenmerken als geschikt zijn beschouwd om het model mee te ontwikkelen. Zie hiervoor het document *Overzicht Verwerkte Data en Features.*

De 'input features' voor het algoritme

Nadat de brondata is geselecteerd, kunnen op basis van deze brondata zogenaamde features worden gemaakt. Deze features moeten passend zijn bij de processtap die moet worden ondersteund met het model.

Bepalen grenswaarde

Niet elke feature is even belangrijk voor de KI. Beredeneerd vanuit de beginselen van dataminimalisatie (noodzakelijkheidsbeginsel) wordt er gebruik gemaakt van *forward feature selection,* waarbij de beschikbare features tijdens het trainen van het model een voor een worden toegevoegd op volgorde van de grootste verbetering in het model, tot de volgende feature niet meer tot een (substantiële) verbetering leidt (dit voorkomt *omitted variable bias).* Daarnaast wordt met de betrokken Product Owner een waarde vastgesteld voor de minimale cumulatieve feature importance waaraan de feature bij moet dragen, waarmee het model in zekere zin wordt 'afgeslankt'. Dit levert de situatie op dat de KI naar minder features 'kijkt' en tegelijkertijd statistisch gezien een relevante onderzoekswaardigheidsscore blijft genereren. Hier moet wel worden opgemerkt dat, vanwege het machine learning karakter van het model, in de toekomst de relevantie van de features kunnen wijzigen en mogelijk andere features relevanter worden voor het model. Features die nu wel binnen de grens vallen kunnen in de toekomst dus buiten de grenswaarde kunnen vallen.

Valideren van de features (ontwikkeltraject)
Tijdens dit ontwikkeltraject gaat de data scientist onderzoeken welke features op zichzelf en in samenhang met elkaar het sterkst mogelijk model vormen en de meest betrouwbare onderzoekswaardigheidsscore geeft. Dit wordt gemeten aan de hand van de statistisch gezien meest 'betrouwbare' onderzoekswaardigheidsscore. Tijdens dit proces worden features toegevoegd of verwijderd. Het resultaat van dit ontwikkeltraject zijn de opgesomde features het document *Overzicht Verwerkte Data en Features.*

Per feature is 'statistisch' inzichtelijk gemaakt in hoeverre deze van invloed is op de voorspellende waarde van het model. Hieruit kan worden afgeleid of een bepaalde features, bijvoorbeeld gebaseerd op een business rule, statistisch gezien van meerwaarde is en dus of het specifiek kijken naar bepaalde zaken in een onderzoek iets toevoegt aan de uitkomst van een onderzoek. Dit kan worden meegenomen in de afweging om bepaalde features al dan niet te gebruiken.

Medewerkers worden betrokken bij de vraag in hoeverre een feature passend is bij de uitvoering van de processtap. Wordt deze feature al min of meer toegepast binnen het werkproces of is het een nieuw 'inzicht' die statistisch gezien relevant is. Er wordt per feature genoteerd in hoeverre hier sprake van is. Voor meer informatie zie antwoord 6.22.

Uitvoeren bias analyse
Zie hiervoor onderdeel 9.

Bespreken features, juridische/ ethische toetsing en vaststellen welke features worden verwerkt.
Nadat het model is ontwikkeld, is er een weging gemaakt met de medewerkers of de features, mede op basis van het resultaat bias analyse, mogen worden gebruikt. Voor een deel zijn de input features een weergave van hoe het werkproces wordt uitgevoerd, maar voor een deel bestaat dit ook uit afwegingen die een data scientist heeft gemaakt. Dit moet aandachtig worden bekeken en wordt telkens herhaald. Deze uitwerking is weergeven in het document *Overzicht Verwerkte Data en Features.*

Tot slot wordt door de eindverantwoordelijk (5.1,2,e) vastgesteld wat de definitief te gebruiken features zijn. Daarmee wordt ook de eindversie van het model bepaald.

Valideren tijdens pilot
Het algoritme moet worden gevalideerd om te kunnen concluderen of de gegenereerde onderzoekswaardigheidsscores "correct" en daarmee bruikbaar zijn. Dat is al gedaan door een deel van de historische data te verwerken in de KI en de uitkomsten te vergelijken. De resultaten hiervan zijn positief. Nu moet het algoritme ook worden getoetst in de praktijk. Dit wordt gedaan in de vorm van een pilot en een pre pilot. Zie het pilot plan en het pre-pilot plan voor de uitwerking hiervan.

7.6. Is er, in de gebruikssituaties waarin er mogelijk sprake kan zijn van negatieve gevolgen voor de grondrechten, een effectbeoordeling op het gebied van grondrechten uitgevoerd?

Ja, er is een IAMA uitgevoerd en aanvullend een diepgaande weging van betrokken (grond)rechten en waarden bij inzet van de KI-toepassing gemaakt.

7.7. In hoeverre hebben betrokkene een kwetsbare positie ten opzichte van de gebruiker van de KI (bezien vanuit macht, kennis, economische en sociale omstandigheden of leeftijd)?

Betrokkenen hebben in dit proces veelal een kwetsbare positie ten opzichte van de gemeente Amsterdam. Aanvragers van een bijstandsuitkering verkeren vaak in een financiële en mentaal kwetsbare situatie waarin er regelmatig ook al een afhankelijkheidspositie bestaat richting de gemeente. Het vertrekpunt zal dan ook altijd zijn dat het voor aanvragers complex is om aanvragen correct en volledig in te dienen. In veel gevallen is nadere informatie noodzakelijk voordat een besluit kan worden genomen en zal de aanvrager hierom moeten worden gevraagd (brief aanvullende gegevens).

De KI gaat een medewerker ondersteunen bij de taak om de onderzoekswaardigheid van een aanvraag te bepalen. De inhoud van de aanvraag en historische data vormen de basis om de onderzoekswaardigheid te bepalen. Een aanvrager heeft daar met het indienen van de aanvraag invloed op, maar vrijwel niet voor zover het gaat om historische data die in de loop der jaren is vergaard door de gemeente. Goede communicatie vooraf over deze gegevensverwerking is hierdoor noodzakelijk. Betrokkenen moeten weten wat hun rechten zijn, zodat ze inzicht kunnen krijgen in persoonsgegevens, bezwaar kunnen maken of onjuiste gegevens kunnen rectificeren.

Ten opzichte van de gebruiker van de KI (medewerker vooronderzoek en de handhavingsspecialist) bestaat er, ook zonder de KI, een kwetsbare situatie voor de aanvrager. Deze rollen zullen namelijk bepalen of er nader onderzoek moet worden gedaan en zullen dit ook uitvoeren. Dat heeft impact op de burger. De KI draagt er wel aan bij dat deze processtap meer gestandaardiseerd wordt uitgevoerd. Dat werkt in het voordeel van de burger, mits de KI ook goed is onderzocht op juridisch en ethische gronden. Dit hebben wij grondig gedaan.

7.8. Wordt door de KI bepaalde personen of groepen schade toegebracht of ongunstig behandeld wat niet gerechtvaardigd is en in verhouding staat tot hun sociale gedrag of de ernst ervan?

Nee, er wordt onderscheid gemaakt tussen onderzoekswaardige en niet-onderzoekswaardige aanvragen. Dit onderscheid wordt gemaakt door een medewerker die dit gaat doen mede op basis van de onderzoekswaardigheidsscore van de Kl. Hetgeen waar de KI naar 'kijkt' of het onderscheid op maakt om deze onderzoekswaardigheidsscore te geven moet de juridische en ethische toets doorstaan. Na het inwinnen van interne en externe adviezen kan worden geconcludeerd dat het onderscheid gerechtvaardigd is omdat de input variabelen/features passend zijn bij het werkproces en het model is onderzocht op discriminerende werking.

7.9. Wordt door keuze architectuur of de gebruikersinterface van de KI menselijk gedrag, meningen of beslissingen beïnvloed?

Ja, beslissing van de gebruiker kunnen worden beïnvloed omdat de output van de KI een onderzoekswaardigheidsscore is waarin ook wordt toegelicht wat de reden is voor de hoge of lage indicatie. Mede op basis hiervan zal een medewerker bepalen welke zaak wordt geselecteerd voor een nader onderzoek. De output van de KI wordt zo neutraal mogelijk gepresenteerd aan de gebruiker, zodat een medewerker altijd zelfstandig de keuze maakt welke zaken wel of niet nader moeten worden onderzocht. De gebruikers zullen met workshops en de gebruikersinstructie worden gewezen op onbewuste protocollering en dus het risico dat medewerkers enkel afgaan op de onderzoekswaardigheidsscore en niet meer hun eigen oordeelsvorming inzetten. Zo zullen de gebruikers niet de score zien die het model geeft, maar enkel of een aanvraag als onderzoekswaardig wordt gelabeld ja/nee. Zo voorkomen we dat gebruikers zich te zeer laten leiden door zeer hoge of zeer lage scores, en de KI zien als een hulpmiddel bij het nemen van een eigen beslissing. Er wordt ook geen gebruik gemaakt van groene of rode kleuren die als sturend kunnen worden ervaren. Daarnaast wordt de gebruiker in de interface eraan herinnerd geen directe conclusies te trekken uit de indicatoren en hun eigen expertise te gebruiken om de aanvraag te behandelen. In 2.14 staat dit verder toegelicht.

7.10. Versterkt of vergroot het KI-systeem menselijke capaciteiten?

Het model geeft mede op basis van historische data een onderzoekswaardigheidsscore per aanvraag. De analyse van de historische data kan het model vele malen sneller en nauwkeuriger dan een medewerker. De medewerker gebruikt deze resultaten als hulpmiddel om te bepalen of de aanvraag al dan niet dieper onderzocht dient te worden. De menselijke capaciteiten worden dus vergroot.

7.11. Wie is de "human in control" en wat zijn de momenten of hulpmiddelen voor menselijke interventie?

Er wordt een menselijk toezichthouder aangewezen die volgens een instructie de werking van het model monitort. Zodra de output van het model niet meer werk conform verwachting, dan zal deze via Topdesk een melding kunnen maken waarna de verantwoordelijke beheerder kan gaan onderzoeken wat het probleem is.

Daarnaast neemt de gebruiker, op basis van de onderzoekswaardigheidsscore, enkel een beslissing om een zaak nader te onderzoeken en geen besluit om een bijstandsuitkering te verlenen. Dat is een wezenlijk verschil, omdat na nader onderzoek ook alsnog een bijstandsuitkering kan worden verleend. Er is daarmee voldoende betekenisvolle menselijke interventie.

7.12. In hoeverre is het mogelijk om een persoon 'uit' de KI te halen (opt-out?)

Als een persoon niet door het model wil worden onderzocht kan de onderzoekswaardigheid ook handmatig door een medewerker worden bepaald. Deze optie wordt niet standaard aangeboden maar is wel een optie als iemand bezwaar indient.

7.13. Zijn er voorzorgsmaatregelen genomen om overmatig vertrouwen in of op het KI-systeem bij werkprocessen te voorkomen?

Ja, er zijn en worden workshops en gebruikersinstructies gegeven/opgesteld aan de gebruikers waarin zij zijn gewezen op het risico van overmatig vertrouwen in de KI-toepassing. Daarnaast is er ook op verschillende manieren in het design van de interface rekening mee gehouden zoals bij vraag 7.9 wordt beschreven.

7.14. In hoeverre kan de output van de KI worden teruggedraaid, gecorrigeerd of buiten beschouwing worden gelaten?

De output van het model kan volledig buiten beschouwing worden gelaten. Een medewerker zal zelf mogen bepalen of deze de output van het model al dan niet gebruikt. Uiteraard is het model ontwikkeld om een waardevolle toevoeging te zijn voor medewerkers. Het ligt in lijn der verwachting dat de output van het model wordt gebruikt.

7.15. Is er getoetst in hoeverre er sprake is van profilering en in hoeverre dat is toegestaan?

In artikel 4, punt 4), AVG wordt profilering als volgt gedefinieerd: *"Elke vorm van geautomatiseerde verwerking van persoonsgegevens waarbij aan de hand van persoonsgegevens bepaalde persoonlijke aspecten van een natuurlijke persoon worden geëvalueerd, met name met de bedoeling zijn beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsing te analyseren of te voorspel/en".* Dat bij dit product hiervan sprake is, is evident. Er worden namelijk analyses gemaakt op de economische situatie van betrokkenen. Daarnaast kan betoogd worden dat de betrouwbaarheid wordt geanalyseerd, doordat een van de doelen is om onrechtmatigheid te verminderen. Uit stukken van de AP[7] blijkt dat profilering bestaat uit drie componenten namelijk:

- **Het moet een geautomatiseerde vorm van verwerking zijn;**
- **Het moet betrekking hebben op persoonsgegevens; en**
- **Het doel van de profilering moet het evalueren van persoonlijke aspecten van een natuurlijke persoon zijn.**

---

[7] https://www.a utoriteitpe rsoo nsgegeve ns.n1/sites/defa u lt/fi l es/ato ms/fi l es/wp251revo1_n1. pdf

Het betreft eengeautomatiseerde vorm van verwerking, nu er een algoritme wordt gebouwd om het de onderzoekswaardigheid van bijstandsuitkeringsaanvraag vast te stellen. Het heeft betrekking op persoonsgegevens. Immers, het algoritme bestaat uit verschillende features. De features hebben betrekking op persoonsgegevens. Daarnaast kan betoogd worden dat profilering het evalueren[8] van persoonlijke aspecten van natuurlijke personen omvat om zo de onderzoekswaardigheid te kunnen bepalen.

Uit stukken van de Autoriteit Persoonsgegevens komt mede naar voren dat van profilering sprake is, mits er bepaalde voorspellingen of conclusies worden getrokken over een bepaald persoon[9]. Hier is i.c. sprake van nu er bepaalde conclusies worden getrokken op basis van een analyse die gemaakt is door het algoritme.

7.16. Is er sprake van geautomatiseerde besluitvorming

Er wordt uiteindelijk een beslissing door het bestuursorgaan genomen, op basis van profilering, in dit geval door middel van een model, om een aanvraag al dan niet verder te onderzoeken; de medewerker vooronderzoek besluit uiteindelijk of een aanvraag wel of niet nader onderzocht moet worden. Het is dus een advies en geen automatische beslissing.

Er is dan nog geen sprake van een genomen besluit. De aanvraag is namelijk niet verleend, geweigerd of buiten behandeling gesteld. Er is nog geen sprake van een rechtsgevolg. Pas nádat nader onderzoek is uitgevoerd door een specialist en interne bespreking heeft plaatsgevonden met de inkomensconsulent, zal een besluit worden genomen in de zin van art. 1:3 Awb door de inkomensconsulent. Er is dus sprake van voldoende menselijke tussenkomst. Om van voldoende menselijke tussenkomst te kunnen spreken moet de tussenkomst betekenisvol zijn. Dit betekent meer dan slechts een symbolische rol. Er moet dus geen sprake zijn van 'rubber stamping': handelingen die in principe ook automatisch hadden kunnen plaatsvinden. De tussenkomst moet uitgevoerd worden door iemand die bekwaam en bevoegd is en deze moet alle relevante gegevens in zijn analyse betrekken[10]. In dit proces zal het besluit altijd genomen worden door een bekwaam persoon in functie. Deze is ook bevoegd het besluit op te maken en zal alle relevante gegevens in zijn analyse betrekken. De uitkomst van het model biedt een ondersteuning, maar de medewerker zal een verder onderzoek uitzetten en zal meer gegevens uit de verschillende zaaksystemen bij zijn analyse betrekken om vast te stellen of er al dan niet sprake is van onrechtmatigheid. Er worden werkinstructies opgesteld die het doel hebben om te veel vertrouwen in de uitkomst van het model te voorkomen. Daarnaast volgen de medewerkers training.

Om deze reden is er **geen sprake** van geautomatiseerde besluitvorming door de inzet van het model.

---

[8] In de AVG wordt profilering omschreven als geautomatiseerde verwerking van persoonsgegevens voor het evalueren van persoonlijke aspecten, met name om zaken over personen te analyseren ofte voorspellen. Het gebruik van het woord "evalueren" suggereert dat bij profilering een zekere beoordeling over een persoon wordt gemaakt.

[9] Zie voetnoot 1.

[10] Richtsnoeren inzake geautomatiseerde individuele besluitvorming en profilering voor de toepassing van Verordening (EU) 2016/679, februari 2018, p. 24.

7.17. Is er een menselijke toezichthouder aangewezen om de werking van de KI te monitoren en om effectieve maatregelen als de KI niet naar behoren presteert?

Deze wordt aangewezen en krijgt aparte instructies die zijn opgenomen in de gebruikershandleiding.

7.18. Is de onderliggende logica in de KI-toepassing passend bij de wettelijke taak die moet worden uitgevoerd met de inzet hiervan?

De logica is intern en extern getoetst. De features die uiteindelijk worden gebruikt worden als passend beschouwd. Zie hiervoor het document *Overzicht Verwerkte Data en Features.*

7.19. Is de KI-toepassing voldoende transparant en controleerbaar? Zo ja, waarom?

Ja, er is veel informatie gedocumenteerd. De inputvariabelen zijn allemaal beoordeeld en worden ook getoond voor zover die relevant waren bij een onderzoekswaardigheidsscore van de KI-toepassing. Daardoor is het voor een medewerker én een belanghebbende mogelijk om te achterhalen wat de output van het model is en hoe dit moet worden geïnterpreteerd. Daarnaast informeert de wethouder het college (hoe dit gebeurt wordt verder beschreven in vraag 13.7), wordt er een privacyverklaring gepubliceerd en komt het algoritme te staan op het algoritmeregister. Daarnaast kunnen we de onderliggende documentatie waar nodig overhandigen aan toezichthoudende en/of rechterlijke instanties.

7.20. ***Is de ontwikkelde KI-toepassing wettig en ethisch verantwoord? Zo ja, waarom?***

De impact van de persoonlijke levenssfeer van de betrokkene staat in redelijke verhouding tot het doel. Niet miskend wordt dat hier sprake is van een zeer gevoelige gegevensverwerking. Ook is de Gemeente Amsterdam zich er terdege van bewust dat er een bijzondere verantwoordelijkheid op haar rust nu er gebruik gemaakt gaat worden van nieuwe technologieën. Er worden echter maatregelen genomen om de negatieve impact zo klein mogelijk te houden. Er wordt onder andere een limitatieve opsomming gegeven welke persoonsgegevens er verwerkt gaan worden. Persoonsgegevens worden gepseudonimiseerd en er wordt een bias analyse uitgevoerd om het model te toetsen op vooringenomenheid. Ook wordt er een DPIA opgesteld om de risico's van de inzet van het model in te kunnen zien en waar nodig te mitigeren. Er worden ook maatregelen genomen om aan het beginsel van transparantie te voldoen. Zo wordt er een aparte privacyverklaring opgesteld voor deze gegevensverwerking en wordt de gegevensverwerking opgenomen in het algoritmeregister.

Zo denken wij dat kijkend naar de baten, de risico's en de mitigerende maatregelen bij die risico's, het inzetten van het ontwikkelde product proportioneel en ethisch verantwoord is. In de DPIA staat een uitvoerende onderbouwing van de wettigheid van dit product, vooral in punt 52 en 53.

8. Bias analyse

Informatie over de bias analyse, de conclusies hieruit en de getroffen maatregelen is in een apart document uitgewerkt. Zie hiervoor de bias analyse documentatie.

9. Technische documentatie[11]

Technische informatie over het model is in een apart document uitgewerkt. Zie hiervoor de technische documentatie.

[11] Zie Richtlijn European Approach for AI voor alle documentatievereisten.

10. Risicomanagement

10.1. Welke risicomanagement systeem wordt toegepast en welke instrumenten worden daarbij ingezet?

Voor dit product zijn wij uitgegaan van het Kader van de Algemene Rekenkamer. Dit kader is vastgesteld als tijdelijk kader totdat de Gemeente Amsterdam een ander kader vaststelt door het Directie Team van het cluster Digitalisering, Innovatie en Informatie. Deze zal waarschijnlijk wel gebaseerd zijn om het kader van de Algemene Rekenkamer. Daarnaast zijn de belangrijkste risicomanagement systemen de DPIA, KIIA en de BIO Ouickscan. Daarnaast wordt het Security (ontrol Framework (SCF) en het Privacy (ontrol Framework (PCF) gebruikt als instrumenten om risico's te mitigeren.

Instrumenten risicomitigatie:

- SCF controls overzicht voor Team Advanced Analytics
- Architectuurnotitie
- Weging betrokken (grond)rechten en maatregelen (op basis van IAMA)
- Overzicht Verwerkte Data en Features
- Bias analyse documentatie
- IAMA
- Beheerplan
- Technische documentatie

10.2. Is er een risicoanalyse uitgevoerd over de toepassing van de KI?

Ja, zie vraag 10.1voor de ingezette instrumenten en hoofdstuk 2 KIIA.

10.3. Is er een overzicht van de geïdentificeerde risico's?

Ja, zie uitkomst DPIA, Ouickscan en overzicht 'betrokken (grond)rechten toetsing'. Ook in sectie 2.10 worden een aantal belangrijke risico's verder toegelicht.

10.4. Kunnen deze risico's worden geëlimineerd of worden gereduceerd?

Zie uitkomst DPIA, Ouickscan, Architectuurnotitie en overzicht 'betrokken (grond)rechten toetsing. Hier is aangegeven in hoeverre de risico's kunnen worden geëlimineerd of gereduceerd. Maatregelen uit het SCF (BIO) en PCF (AVG) zorgen ervoor dat de risico's worden geëlimineerd of worden gereduceerd.

10.5. Is er een overzicht opgesteld van de getroffen of nog te treffen maatregelen?

Zie uitkomst DPIA, Ouickscan, architectuurnotitie, overzicht 'betrokken (grond)rechten toetsing, Daarnaast is er een overzicht van de SCF controls voor Team AA en een beheerplan waarin maatregelen zijn uitgewerkt. Hier is aangegeven in hoeverre de risico's kunnen worden geëlimineerd, gereduceerd of welke procedure moet worden doorlopen.

10.6. Zijn potentiële veiligheidsrisico's vastgesteld van (andere) voorzienbare toepassingen van de technologie, met inbegrip van onbedoeld of kwaadwillig misbruik ervan? Is er een plan om deze risico's te beperken of te beheren?

De KI geeft enkel een onderzoekswaardigheidsscore bij ontvangen aanvragen. Een risico is ongeautoriseerde wijziging van de code waardoor de KI een foutieve score geeft ten nadele van burgers. Toekennen en beheren van rollen en rechten is een belangrijk aandachtspunt en het beveiligingsniveau van de Azure omgeving en de gebruikte componenten zal BBN2 moeten zijn. Er is een procedure voor het toekennen van rollen en rechten en het SCF is toegepast op de Azure omgeving, de bouwblokken en op workload niveau.

10.7. Is het acceptabel, gezien de risico's en de beoogde maatregelen, om de KI-toepassing in gebruik te nemen?

Gezien de getroffen maatregelen en de werking van de KI-toepassing is het acceptabel om de KI-toepassing in gebruik te nemen gedurende de pilot. Aandachtspunt zal zijn de consequente toepassing van de gedefinieerde maatregelen, bijv. het opslaan van logs bij beveiligingsincidenten en het actueel houden van autorisatiematrix.

10.8. Wordt het risicomanagementsysteem gedurende de levenscyclus van de KI-toepassing toegepast?

De verschillende analyses worden gedurende bepaalde momenten in het ontwikkelproces uitgevoerd (bijvoorbeeld bij een grote wijziging). Er zijn op alle disciplines experts betrokken. Zij worden gevraagd om advies mondeling en via het uitvoeren van de risicoanalyses te geven, zoals de DPIA en het toepassen van het SCF.

10.9. Wordt het risicomanagementsysteem periodiek geëvalueerd?

De instrumenten worden zelfstandig periodiek geëvalueerd. Dit wordt gedaan door de directie Digitale Strategie en Informatie (voorheen CIO-office). Deze instrumenten worden tevens voorgelegd ter vaststelling.

11. Informatiebeveiliging

11.1. Is er sprake van security by design? Zo ja, hoe wordt dit toegepast?

Ja, vanaf het begin is er een informatiebeveiligingsadviseur betrokken geweest en die heeft geadviseerd over meerdere vraagstukken. Hierbij kan worden gedacht aan het verplaatsen van data uit de VAO naar de Azure omgeving, het leveren van de data naar de VAO ten behoeve van de bias analyse en hoe we zo direct in productie ervoor kunnen zorgen dat alles veilig is.

Het model zal uiteindelijk worden gebruikt in de Azure-omgeving. Het Security Controle Framework (hierna: SCF) is een vertaling van vereisten uit de Baseline Informatiebeveiliging Overheid (hierna: BIO). De BIO en dus het SCF bevat vereisten waaraan moet worden voldaan om gegevens te beschermen.

Samen met DASO heeft Team AA het SCF doorlopen, maatregelen getroffen en dit is middels Peer Review door het Cloud Competence Center (hierna: CCC) als akkoord bevonden. Uitgangspunt is dat de DevOps-team/workload eigenaar zelf verantwoordelijk zijn voor het toepassen van de juiste maatregelen.

11.2. Is een Ouickscan uitgevoerd?

Ja, met de betrokken ISO is afgesproken om de Ouickscan over de KI uit te voeren in plaats van over het gehele werkproces. Deze KI maakt onderdeel uit van het hoofdproces "Bepalen en beheren recht op een uitkering". Voor dit hoofdproces is reeds een Ouickscan uitgevoerd. Het KI deel zal worden ondergebracht in de OuickScan van dit hoofdproces.

11.3. Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de KI te borgen?

Zie 11.6.

11.4. Welke maatregelen zijn genomen om de veiligheid van de KI te borgen?

Er is voldaan aan de veiligheidsmaatregelen uit het SCF.

11.5. Welke maatregelen zijn genomen om de vertrouwelijkheid te borgen?

De van toepassing zijnde maatregelen uit het SCF zijn genomen, in het bijzonder het inregelen van het autorisatieproces zie: 11.12 t/m11.20 en11.23.

11.6. Welke maatregelen zijn genomen om de integriteit te borgen?

Een selecte groep medewerkers heeft toegang tot de data en de output van de Kl. Daarnaast geldt een '4-ogen-principe' voor alle wijzigingen aan de code. Beide maatregelen verlagen het risico dat de integriteit van de data wordt geschaad. Er zijn daarnaast op meerdere plekken in de KI en de

omringende componenten checks ingebouwd op de validiteit van de data en de gegenereerde output om onbedoelde wijzigingen te detecteren.

11.7. Welke maatregelen zijn genomen om de beschikbaarheid te borgen?

De van toepassing zijnde maatregelen uit het SCF zijn genomen, in het bijzonder het automatisch maken van back-ups van de KI binnen de Azure omgeving (zie: 11.24). Daarnaast is er een Service Level Agreement opgesteld met de afdeling WPI, waarin afspraken zijn gemaakt over de termijn waarbinnen storingen opgelost dienen te worden en hoeveel downtime acceptabel is. Er zal een 1e versie van de SLA zijn bij aanvang van de pilot. Als de pilot succesvol is en we willen door met dit product zal de SLA verder worden uitgebreid.

11.8. Welke maatregelen zijn er genomen om de traceerbaarheid te borgen?

De van toepassing zijnde maatregelen uit het SCF zijn genomen, in het bijzonder het inregelen van logging en de opzet van monitoring (zie: 11.10 en 11.11). Er wordt vastgelegd welke voorspelling wanneer is gemaakt, op basis van welke feature waardes en met welke versie van het model. Daarnaast is in de ontwikkelomgeving, Azure Machine Learning, tot in detail te traceren welke data, code en experimenten ten grondslag liggen aan een specifieke versie van het model. Meer informatie over logging is te vinden in de technische en beheerdocumentatie.

11.9. Welke maatregelen zijn genomen om de verklaarbaarheid te borgen?

Ten eerste is gekozen voor een relatief eenvoudig te begrijpen algoritme. De manier waarop dit algoritme voorspellingen maakt zorgt ervoor dat een individuele voorspelling direct kan worden verklaard op basis van de features. De belangrijkste features worden getoond aan de eindgebruikers wanneer zij kennisnemen van de uitkomst van het algoritme.

Daarnaast is de interface ontworpen om verklaarbaarheid te maximaliseren en automation bias te minimaliseren. Voor meer informatie zie 2.10 en 7.9.

11.10. Wordt loginformatie over de werking van de KI (geautomatiseerd) gegeneerd, bewaard en toegankelijk gemaakt?

Ja, er wordt voldaan aan de logstandaarden uit het SCF voor wat betreft de infrastructuur waarop de KI draait. Daarnaast worden extra zaken gelogd over de werking van het model en de API. Zie voor details hierover de technische documentatie.

11.11. Door wie worden de logs bewaard?

De logs worden bewaard binnen Azure met Log Analytics Workspace. Er zijn enkel leesrechten toegekend. Logs kunnen niet worden gewijzigd. In het geval van een beveiligingsincident worden de relevante logs langer bewaard op een aparte plek, namelijk een Azure storage account. Voor meer details hierover, zie het beheerplan.

11.12. Wordt gecontroleerd of toegangsrechten up-to-date zijn met betrekking tot de omgeving waarin de KI functioneert?

Ja, er is een procedure opgesteld waarin de toegangsrechten middels een autorisatiematrix zijn geregeld. Hierin is vastgelegd welke rol tot welke componenten en welke logs toegang heeft. Dit is beschreven in het beheerplan.

11.13. Worden toegangsrechten aangepast zodra er een uitdiensttreding of functiewijziging van een werknemer plaatsvindt?

Ja, er is een procedure opgesteld om dit te regelen volgens de richtlijnen van het SCF. Zie hiervoor het beheerplan.

11.14. Worden toegangsrechten uitgegeven door daarvoor bevoegde personen?

Ja, er is een procedure opgesteld om dit te regelen volgens de richtlijnen van het SCF. Zie hiervoor het beheerplan.

11.15. Wordt functievermenging voorkomen bij de toegang van gebruikers tot de KI?

Ja. De rollen van data scientist en data engineer zijn gescheiden.

De data engineer (DASO) heeft toegang tot de ruwe data en maakte deze in een database beschikbaar voor verwerking. Daarnaast beheert hij/zij de infrastructuur waarop de KI en de data pipelines draaien. Hij/zij kan daarom onder andere:

1. Infrastructuur opzetten
2. De database en data pipelines beheren
3. De input data API beheren
4. Logs en monitoring betreffende de infrastructuur, database, data pipelines en input data API inzien

De data scientist (team AA) is verantwoordelijk voor het ontwikkelen, deployen en beheren van de Kl. Hij/zij kan daarom onder andere:

Aanpassingen maken aan het model
Docker images bouwen rondom het model en deze deployen
Logs en monitoring betreffende het model inrichten
Logs en monitoring betreffende het model inzien

Dit zijn de enige twee autorisaties waarvoor functiescheiding van toepassing is. iedere medewerker in een van deze twee rollen heeft één eigen account waarmee ze toegang hebben tot Azure.

Daarnaast gaat de eindgebruiker de informatie consumeren via de Dataverkenner. Er hebben twee functies toegang tot deze informatie (medewerker vooronderzoek en handhavingsspecialist) en deze rollen zijn apart beschreven in de autorisatiematrix van de Dataverkenner. Meer informatie over hun beveiliging is te vinden in hun documentatie.

11.16. Wordt er gebruikgemaakt van generieke beheeraccounts? Staat het aantal beheeraccounts in logische verhouding tot de beheerders?

Nee, iedereen van team AA en DASO logt in met zijn/haar AD credentials.

11.17. Wordt er bij het inrichten van toegangsrechten van verschillende gebruikersgroepen/rollen gebruikgemaakt van naamgevingsconventies en systematiek?

Ja, dit wordt door DASO uitgevoerd volgens de richtlijnen van het SCF. Zie voor meer details de documentatie van DASO omtrent de implementatie van het SCF.

11.18. Voeren beheerders werkzaamheden als beheerder en werkzaamheden als gewone gebruiker uit onder 2 verschillende gebruikersnamen?

Het onderscheid tussen beheerwerkzaamheden en werkzaamheden als gewone gebruiker is niet te maken. In plaats daarvan is er onderscheid tussen de werkzaamheden van de data scientist en die van de data engineer. Zie voor de beschrijving van deze rollen vraag 11.15.

11.19. Hebben gebruikersaccounts (geen) directe toegang tot onderliggende componenten?

Nee, enkel als deze hiervoor geautoriseerd zijn. Niet elk gebruikersaccount heeft toegang nodig tot elke component. In de autorisatiematrix wordt nauwkeurig beschreven welke rol precies wat kan doen. Dit geldt zowel voor inhoudelijk toegang tot de database (op data niveau) als voor de Azure componenten.

11.20. Bestaat er een functiescheiding tussen aanvragen, autoriseren en verwerken van wijzigingen in gebruikersaccounts en toegangsrechten?

Ja, het aanvragen van toegangsrechten wordt gedaan door de gebruiker zelf, het autoriseren van de aanvraag door de product owner van team AA of DASO en de verwerking van de wijziging gaat automatisch na autorisatie.

11.21. Is het wachtwoordbeheer interactief en zijn de wachtwoorden van geschikte kwaliteit?

Ja, de persoonlijke wachtwoorden zijn de AD-wachtwoorden die een gebruiker voor de hele gemeente gebruikt en voldoen dus aan de eisen die daar gesteld worden. De generieke 'secrets', die nodig zijn om bijvoorbeeld twee componenten met elkaar te verbinden, worden door DASO automatisch gegenereerd op basis van OpenSSL base64 met 24 karakters.

11.22. Worden wijzigingen in de code van de KI op een gecontroleerde wijze uitgevoerd? Denk aan het testen en accorderen/autoriseren van wijzigingen.

Ja, wijzigingen/verschillende versies van de code van de KI worden automatisch opgeslagen in de repository op Azure. Hierdoor is versiebeheer geregeld. Alleen de Data Scientist kan wijzigingen in de code van KI doorvoeren. Deze wijzigingen worden altijd door een andere Data Scientist gecontroleerd. Dit wordt technisch afgedwongen.

11.23. Is de KI beveiligd, zodat er geen risico is op ongeautoriseerde toegang, wijziging, beschadiging en/of dataverlies?

Ja, binnen Azure hebben alleen bevoegde rollen toegang tot de KI (code).

11.24. Worden er back-ups van de KI gemaakt en kan de KI en de data hersteld worden?

Ja, de trainingsdata kan (afhankelijk van verwijdering gegevens op basis van de Archiefwet) middels een query opnieuw worden bevraagd. Er worden automatisch back-ups van de KI gemaakt binnen de Azure omgeving.

12. Beheer

12.1. Is er een beheerplan opgesteld?

Ja, er is een beheerplan opgesteld op hoofdlijnen. Tijdens de pilotfase zal dit plan verder worden uitgewerkt. Daarnaast is een apart beheerplan opgesteld voor de pilotfase.

12.2. Is er voldoende capaciteit om de KI-toepassing operationeel te houden en wijzigingen door te voeren?

Ja, dit zal worden gedaan door het team dat het model ook ontwikkeld heeft. Voor meer details, zie het beheerplan.

12.3. Is er voldoende kennis aanwezig om de KI-toepassing operationeel te houden en wijzigingen door te voeren?

Ja, in het team zitten mensen met de juiste kennis om dit te kunnen doen. Voor meer details, zie het beheerplan.

12.4. Worden de werking van de KI-toepassing periodiek geëvalueerd?

Ja, een groep specialisten komt periodiek samen om te evalueren of het model nog naar behoren functioneert. Dit team bestaat ten minste uit het functioneel beheerteam; applicatie beheerteam en de Epic Owner. Daarnaast zal zoveel mogelijk externe expertise zoals de CPA worden aangehaakt. Ook monitoren de beide beheerteams continue de prestaties en technische performance van het model. Voor meer informatie, zie het beheerplan.

13. Communicatie

Zie voor meer informatie met betrekking tot de communicatie het *Communicatieplan Pilot Informatiegestuurd Handhaven.*

13.1. Is er een gebruikersinstructie opgesteld?

Nog niet, dit wordt opgesteld in de aanloop naar de pilot.

13.2. Zijn mechanismen ingesteld om gebruikers in kennis te stellen van de redenen en criteria achter de resultaten van het KI-systeem?

Er worden een aantal maatregelen getroffen:

Er worden workshops worden gegeven en gebruikersinstructies worden opgesteld om medewerker bewust te maken van de kansen en risico's van het gebruik van **KI.** Daarbij wordt toegelicht hoe de uitkomst van het algoritme moet worden geïnterpreteerd.
De output van de **KI** is zichtbaar voor de medewerker. Hoewel het model elke aanvraag een score geeft, ziet de medewerker enkel of het onderzoekswaardig is ja/nee. Dit om automation bias te voorkomen bij een zeer hoge onderzoekswaardigheidsscore.
Daarnaast worden voor elke onderzoekswaardige aanvraag de drie belangrijkste features weergegeven en toegelicht.
De onderbouwing wat bepaalde features betekenen en waarom deze mogen worden gebruikt kan worden nagelezen in het document *Overzicht Verwerkte Data en Features V1.0.*

13.3. Is duidelijk gemaakt richting gebruikers en betrokkenen wat het doel van het KI-systeem is en wie of wat er profijt kan hebben van het product of de dienst?

Ja, er worden workshops gegeven en werkinstructies opgesteld om dit toe te lichten. De doelstellingen van het product worden hierbij toegelicht aan de gebruikers. Voor Amsterdammers is een toelichting beschikbaar in de privacyverklaring en het algoritmeregister. Daarnaast zal er een O&A publiekelijk beschikbaar worden gemaakt via amsterdam.nl/veelgevraagd. Hierin zullen zowel de baten als risico's worden meegenomen. Deze O&A zal ook beschikbaar worden gesteld aan de medewerkers van het Contact Center Amsterdam (CCA), zodat vragen vanuit Amsterdammers ook persoonlijk beantwoord kunnen worden.

13.4. Is inzichtelijk gemaakt welke kenmerken, beperkingen en potentiële tekortkomingen van het KI-systeem duidelijk kenbaar gemaakt?

Ja, er worden workshops gegeven en werkinstructies opgesteld om dit inzichtelijk te maken. In zowel de werkinstructie als de workshop wordt het risico op automation bias en het belang van het eigen expertise van de medewerker benadrukt. Daarnaast zijn met de Product Owner en gebruikers de opgestelde features besproken. Daardoor is er kennis aanwezig over de kenmerken, beperkingen en potentiële tekortkomingen van de KI-toepassing.

13.5. Heeft de verwerkingsverantwoordelijke akkoord gegeven voor het in gebruik nemen van de KI-toepassing?

Nog niet, dit moet plaatsvinden voorafgaand aan de pilot. Eerst zal er nog een extern advies worden ingewonnen en dan zal de verwerkingsverantwoordelijke, op basis van alle informatie, een besluit nemen.

13.6. Is de verantwoordelijke woordvoerder en wethouder geïnformeerd over het in gebruik nemen van de KI-toepassing?

Ja, deze zijn geïnformeerd. Het ontwikkelen van dit model gebeurt naar aanleiding van een opdracht van de wethouder.

13.7. Is er een collegebesluit genomen over het in gebruik nemen van de KI-toepassing?

Er is nog geen collegebesluit genomen over het in gebruik nemen van de KI-toepassing.
Vooropgesteld: wij weten niet zeker hoe de Wethouder dit wil gaan doen. De bestuursadviseur denkt echter dat hij het ter beslissing wil voorleggen, sowieso aan het college. Hij vermoedt dat dit nog wel breed besproken moet worden (dus: raad en niet alleen in de commissie).
Hij denkt dat als we dit ter kennisname aanbieden aan de commissie (en dat is het minste wat we hebben toegezegd) de commissie dit zal 'piepen' naar de Raad in de vorm van een bespreekpunt. Dit raakt immers niet alleen de portefeuille sociale zekerheid maar ook digitalisering en dergelijke.

Dus wat hij verwacht:

Met de bevindingen van Deloitte naar de staf wethouder inclusief planning pilot

Dan kabinet naar het college voor akkoord

Dan streven naar ter kennisname commissie ervan uitgaand dat het punt waarschijnlijk gepiept wordt naar bespreken in de Raad. Hij kan zich in dit geval echter voorstellen dat het college het direct ter bespreking aan de raad wil voorleggen.

14. Verantwoording

14.1. Hebt u mechanismen ingesteld die de controleerbaarheid van het systeem door interne en/of externe onafhankelijke actoren mogelijk maken, bijvoorbeeld in de vorm van waarborging van de traceerbaarheid en het bijhouden van de processen en resultaten van het KI-systeem?

In het pilot en beheerplan staan verschillende mechanismen voor het controleren van het systeem. Denk hierbij aan een duidelijk meetplan om de resultaten te monitoren en het opslaan van logs om de traceerbaarheid te waarborgen. Voor meer informatie over de logs zie het antwoord op vraag 11.10.

Ook wordt er bij iedere voorspelling ook informatie opgeslagen om de controleerbaarheid te waarborgen. Kijk voor meer informatie hierover naar het document 'Bewaartermijnen - algoritme onderzoekswaardigheid wpi'

14.2. Is er, naast de interne initiatieven of kaders voor het toezicht op ethiek en verantwoording, ook sprake van een vorm van externe begeleiding of zijn er ook controleprocessen ingesteld?

De Participatieraad en de Commissie Persoonsgegevens Amsterdam hebben beide geadviseerd, waaronder ook op de ethische aspecten. Hun adviezen hebben wij zo veel mogelijk probeert te verwerken. Daarnaast gaat er ook door Deloitte een uitvoerende analyse gedaan worden op ofwe het algoritme ethisch verantwoord kunnen inzetten en welke mogelijke extra maatregelen we nog moeten treffen.

14.3. Hebt u een mechanisme ingesteld om relevante belangen en waarden die door het KI-systeem in het gedrang komen, alsook potentiële afwegingen daartussen, vast te stellen?

Ja, er zijn meerdere (risico)analyses uitgevoerd om deze belangen en waarden in kaart te brengen. Om deze belangen en waarden te eerbiedigen is onderzocht welke maatregelen kunnen worden getroffen. Zie hiervoor het overzicht 'weging betrokken (grond)rechten', de DPIA en de IAMA.

15. Afwegingen en conclusie

15.1. Is de toepassing proportioneel?

De toepassing van de KI valt binnen de bestuursrechtelijke publieke taak en bevoegdheden, namelijk het zo goed mogelijk behandelen van ontvangen aanvragen voor levensonderhoud. Hierbij is rekenschap gegeven van rechten als eerbiediging van de persoonlijke levenssfeer, recht op informatie over overheidsoptreden, transparantie omtrent verwerking van persoonsgegevens en dataminimalisatie. Tevens zijn de noodzakelijke technische en organisatorische maatregelen genomen ter beveiliging van de verwerkte persoonsgegevens bij het ontwikkelen van het algoritme. De impact op het individu en de maatschappij wordt met voldoende technische en organisatorische waarborgen beschermd en de beperking van enkele grondrechten valt binnen de (in de rechtspraak) ontwikkelde grenzen als vrijheid van informatie en recht op een eerlijk proces. Er is onderzoek gedaan naar een mogelijk discriminerend effect van de KI-toepassing op aanvragers. Er kan worden geconcludeerd dat er geen sprake is van een discriminerende werking van de KI-toepassing.

Daarnaast is het met het model zo dat onafhankelijk van de gekozen drempelwaarde in verhouding met minder onterechte onderzoeken (vergeleken met het huidige proces) evenveel (of meer) onrechtmatige aanvragen worden gevonden. Daarmee is deze werkwijze dus meer proportioneel dan het huidige proces (zie hoofdstuk 1: *Inleiding -Aanleiding en probleemstelling* voor cijfermatige onderbouwing).

Hierbij dient te worden vermeld dat een van de aanleidingen voor het ontwikkelen van het algoritme is dat het percentage aanvragen dat in het huidige proces naar handhaving wordt doorgestuurd sinds 2015 flink is afgenomen (van 15% naar 4%). Een van de doelstellingen van de pilot is dan ook het vergroten van de controle die **WPI** heeft op het percentage aanvragen dat bij handhaving landt. In de praktijk betekent dit dat het aandeel van de aanvragen dat door handhaving onderzocht wordt zal toenemen. Hoewel er dus in absolute zin meer aanvragen onderzocht zullen worden, zal het aandeel hiervan wat onterecht onderzocht wordt afnemen. Dit maakt het proces meer proportioneel.

15.2. Is met minder ingrijpende middelen hetzelfde doel te verwezenlijken?

De KI-toepassing is de beste methode om mede op basis van historische gegevens een snelle en accurate onderzoekswaardigheidsscore te geven. Medewerkers zijn niet zelfstandig in staat om deze afweging op een even effectieve wijze en binnen hetzelfde tijdsbestek als de KI toepassing te doen. Om tot de beoogde effectiviteit te komen, lijkt de KI-toepassing het meest geschikte middel te zijn. Dit moet uiteraard worden gevalideerd tijdens de Pilot.

15.3. Gaat het om positive sum in plaats van zero sum?

Het betreft een positive sum. Bij een juist werkende KI-toepassing zal de capaciteit van de beoordelaars aanvragen effectiever worden benut. Er zijn verschillende positieve effecten voor de aanvrager/burger zoals dat deze alleen aan een nader onderzoek wordt onderworpen als daar een reden voor is. De hit-rate van het algoritme is namelijk hoger dan van het huidige proces, hetgeen ertoe leidt dat voor elke terecht onderzochte aanvraag minder aanvragen onterecht onderzocht hoeven te worden. De KI-toepassing helpt om deze reden inzichtelijk te maken. Uit de pilot zal uiteraard moeten blijken of de doelstelling en de voordelen voor de aanvrager en burger ook daadwerkelijk wordt behaald.

15.4. Zijn er rest-risico's?

Er zijn risico's van maatschappelijke en politieke aard. Het zou kunnen dat de Amsterdamse bevolking grote bezwaren heeft met de inzet van een KI-toepassing wat de onderzoekswaardigheid van een aanvraag levensonderhoud kan voorspellen. Het publieke sentiment is vaak negatief door toedoen van de toeslagen affaire en fraude-opsporingsinstrument Systeem Risico-Indicatie (SyRI) dat door de Gemeente Rotterdam werd gebruikt. Dit bezwaar zal zich kunnen manifesteren via media-aandacht, de politiek of bezwaar- en beroep procedures. Het is daarom van belang om transparant te zijn richting betrokkenen, de wethouder en het college van B&W volledig te informeren over de materie. Daarnaast is het daarom ook van belang om veel externe advisering te organiseren zoals het advies van de CPA en een extern adviesbureau.

15.5. Hoe wordt verantwoordelijkheid genomen voor verder gebruik?

Als er sprake is van een succesvolle pilot, dan zal de Directeur WPI een besluit moeten nemen om tot implementatie van de KI-toepassing over te gaan. De Wethouder zal hier eveneens akkoord voor moeten geven en zal de raad en het college informeren en wellicht ook bespreken.