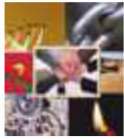




Gemeente Rotterdam

Concern Auditing



Concern Auditing Advies

Concept
4 februari 2021
Zaaknummer: 3461046

Aan [REDACTED],
directeur Inkomen W&I

**Advies Analytics
Uitkeringson-
rechtmatigheid**



Advies Analytics Uitkeringsonrechtmatigheid

Aanleiding, context

W&I past een slim algoritme ('Analytics Uitkeringsonrechtmatigheid') toe als risicomodel in het proces Heronderzoeken (HO). Doelstelling van het algoritme is, naast meer effectiviteit van HO, ook verbeterde efficiëntie (minder HO's nodig) en daarmee onder meer minder belasting van de burger.

Artificiële Intelligentie staat maatschappelijk onder kritische aandacht met vaak negatieve beelden op het gebied van privacy, ongelijkheid en transparantie. Bij de ontwikkeling en het gebruik van AI is 'ethiek' (dan ook) een belangrijke factor, met aandacht bijvoorbeeld voor ongelijke behandeling en transparantie. Hiervoor zijn weliswaar nationale en internationale richtlijnen, maar de concrete uitwerking daarvan is nog in ontwikkeling. Mede als gevolg van de kritische maatschappelijke aandacht komt de vraag op of voor dit algoritme de juiste 'ethische' regels zijn/worden gevolgd.

W&I heeft CA gevraagd op korte termijn te kijken naar de 'ethische regels' met betrekking tot non-bias, -discriminatie en profiling (kortom ongelijke en oneerlijke behandeling, gemakshalve gebruiken we de term 'vooringenomenheid') en transparantie (uitlegbaarheid, communicatie). We gaan hierbij uit van onder meer nationale en EU-richtlijnen, met de aantekening dat de concrete uitwerking van deze richtlijnen ook binnen Rotterdam nog in ontwikkeling zijn. Naast het beeld dat wij hierover schetsen, geven we waar mogelijk handvatten/advies ter verbetering; beleid en regels zijn immers nog in ontwikkeling; dit geldt ook voor instrumenten om het algoritme op de ethische aspecten te kunnen toetsen.

Algemeen beeld

Bij de ontwikkeling van het algoritme en de werking daarvan is goed nagedacht over keuzes en maatregelen om vooringenomenheid te voorkomen. We zien echter ook, dat deze niet altijd expliciet zijn beschreven, vaak in lastig te begrijpen technische taal. Het toetsen van mogelijke vooringenomen gebeurt op dit moment aan de 'voorkant' (de input) onder meer op basis van keuzes ten aanzien van brondata en aan de 'achterkant', met de resultaten van de heronderzoeken die op basis van het algoritme worden uitgevoerd ('outcome'). Naast het trainen en testen van het algoritme (dat ook iets kan zeggen over mogelijke vooringenomenheid) vindt het toetsen van het algoritme zelf nog niet plaats. Er wordt daarom nu ook nagedacht over de inzet van een tool die meer technisch de werking van het algoritme toetst op eerlijkheid en vooringenomenheid. De maatregelen die we bij de 'input' en 'outcome' in opzet zien, geven geen belemmeringen voor het algoritme zoals het is toegepast. Daarbij is de uiteindelijke vaststelling op onrechtmatigheid 'mensenwerk' en het past in het reguliere proces van W&I. Om het 'compleet' te maken, is een 'technische' toets op het algoritme heel zinvol.

Voor het aandachtsgebied transparantie (uitlegbaarheid, communicatie) moeten nog veel stappen genomen worden. Het gaat hierbij om de proactieve communicatie naar burgers en de betrokkenen over functie en gebruik van het algoritme. De uitwerking van de principes hiervoor zijn onder meer nationaal (Ministeries BZK en V&J) in ontwikkeling. De AVG en ook uitspraken van de Raad van State geven aan, dat 'openheid' en 'uitlegbaarheid' over het gebruik van algoritmes essentieel is. Op dit moment is er echter geen duidelijk plan op de wijze waarop burgers en betrokkenen worden geïnformeerd. De huidige beschikbare documentatie is weliswaar nog niet geschikt voor externe communicatie (niet actueel, technisch), maar bevat wel geschikte onderdelen. Wat op dit moment vanuit dit oogpunt de risico's zijn voor de huidige bezwaarcasus, is moeilijk in te schatten.

Bij de totstandkoming en uitvoering van het algoritme zijn naast W&I (UBO), het DWI (OBI) en het team Advanced Analytics betrokken. Het algoritme is in projectvorm (programma DGW, OBI (team Advance Analytics) en W&I) ontwikkeld. Het is niet duidelijk wat nu de status van het algoritme is (pilot, productie?) en het is niet altijd helder hoe de verschillende verantwoordelijkheden zijn belegd. Ook is niet duidelijk hoe verdere ontwikkeling van het algoritme eruitziet.

In het onderzoek hebben we gekeken naar de 'algoritme-keten' en hiervoor gesproken met de betrokkenen in de ontwikkeling en uitvoering hiervan en medewerkers van JD. We hebben desgevraagd ook de 'DT-documenten' beoordeeld. Hierna gaan we meer in detail in op beelden over en handvatten voor de 'algoritme-keten', de 'DT-documentatie' en onder meer 'governance'.

We adviseren voor het verdere gebruik van het algoritme onder meer:

- duidelijke beschrijving van alle (kwaliteits-)maatregelen inzake vooringenomenheid;
- actualisering documentatie;
- communicatieplan extern;
- inzet tool 'eerlijke algoritme';
- visie en plan verdere ontwikkeling algoritme.

Algoritme-keten (en vooringenomenheid)

Het algoritme levert een lijst op (gebaseerd op een risicoscore), dat in het proces HO wordt toegepast. Daarvoor analyseert het algoritme op basis van historische data in welke mate er een kans is op onrechtmatigheid op de actuele populatie.

De kwaliteit van het algoritme staat niet op zichzelf, maar wordt gevormd door de maatregelen in de gehele 'keten' van dat algoritme. Uit de gesprekken en de documentatie hebben wij het beeld, dat er goed is nagedacht over 'vooringenomenheid' (non-bias, -discriminatie en profiling) en zijn er in opzet maatregelen hiervoor getroffen in de 'algoritme-keten'.

- De 'input' vanuit de bronsystemen Socrates en Raak (W&I, Wigo4it): de velden van de brondata bevatten geen gevoelige items, zoals nationaliteit e.d. (behalve 'taalniveau'). Mogelijke vraag is wel hoe de inputdata vanuit het proces HO tot stand zijn gekomen.
- Bewerkingen op de brondata en het aanleveren (vanuit het DWH) aan het team Advanced Analytics (tAA): Er vindt pseudonimisering (DWH) plaats op de data waarop het algoritme draait (vanwege privacy).
- Werking van het algoritmemodel (tAA): trainen, testen en uitvoeren en het resultaat van het algoritme: de scorelijst ('output'): het algoritmemodel (methode 'random forest') levert de variabelen op een wijze die 'wij' niet kunnen reproduceren en berekent in welke mate er een kans is (score tussen 0 en 1) dat er sprake is van onrechtmatigheid. Er is nog geen methode/tool ingezet dat het algoritme zelf toetst op vooringenomenheid.
- Screening van de scorelijst (UBO) en het gebruik van de scorelijst door het team 'Heronderzoeken' (tHO) en het resultaat daarvan ('outcome'): de door het algoritme opgeleverde scorelijst wordt gescreend om te voorkomen dat er onnodige heronderzoeken plaatsvinden. Het algoritme geeft een risicoscore aan, feitelijke vaststelling of er wel of niet sprake is van onrechtmatigheid gebeurt door de medewerkers in het proces HO. Naast de selectie vanuit het algoritme werkt het tHO ook met andere selecties ('thema' en 'random'). De resultaten van alle selecties worden geanalyseerd door Unit Bijzondere Onderzoeken (UBO) op effectiviteit ('hitrates') en word gekeken naar eventuele over- en of ondervertegenwoordiging vanuit bepaalde categorieën, ook ten aanzien van etnische achtergrond.
- Analyse en rapportage van de resultaten (UBO, tAA).

Zoals is aangegeven, zijn er keuzes gemaakt om vooringenomenheid te voorkomen. Deze zijn niet altijd duidelijk (impliciet of te technisch) beschreven.

- Wij adviseren deze keuzes en (kwaliteits-)maatregelen - van alle schakels in de keten - helder te beschrijven (met ook een beschrijving van de verantwoordelijkheden, wie voert uit en wat zijn de onderlinge afspraken). Zie ook hierna, de 'DT-documenten' bevatten zeker ingrediënten hiervoor.
- Werk de bestaande infographic(s) uit voor de gehele algoritme-keten.
- Maak gebruik van de mogelijkheid om samen met het team Advanced Analytics met een CBS-tool ('eerlijke algoritmes') het algoritme zelf te toetsen op 'vooringenomenheid'. Zo'n instrument zou standaard in de (beoogde) kwaliteitscontrole van tDS/OBI passen.
- Gebruik (UBO en tAA) de analyses van de resultaten (waar mogelijk) om het algoritme te verbeteren zowel op effectiviteit als op eventuele vooringenomenheid. Deze 'terugkoppeling' moet een vast onderdeel zijn van de ontwikkel- en beheercyclus van het algoritme.
- Ga na wat het risico is, dat het HO-proces uit het verleden (waarvan de resultaten de input vormt om het algoritme) elementen bevat met betrekking tot vooringenomenheid en zo onbedoelde uitkomsten van het algoritme kan beïnvloeden.

Transparantie / DT documentatie

Voor intern en externe communicatie zijn drie (concept-)documenten opgesteld: "... omdat binnen de gemeente vragen rijzen over de heronderzoeksystematiek en de werkwijze". Op dit moment betreft dit vooral het behandelen van bezwaar- en rechtszaken. Mogelijk kunnen WOB-verzoeken hiertoe ook aanleiding geven. In het algemeen is het de wens te voorkomen dat er een (veelal dan ook hardnekkig) negatief beeld ontstaat over het gebruik van algoritmes. Het gaat om de volgende documenten:

1. Heronderzoeken Werk en Inkomen (algemeen)
2. Rapport Project Analytics Uitkeringsfraude (Big Data)
3. Leeswijzer rapport Project Analytics Uitkeringsfraude (Big Data)

Deze beschrijven "de systematiek van het Heronderzoek en het gebruik van Big Data binnen de rechtmatigheidsonderzoeken".

Algemeen beeld

Transparantie, communicatie en uitlegbaarheid over het algoritme is niet alleen voor juridische aangelegenheden of raadvragen, maar ook om betrokkenen (onder meer burgers) te informeren over gebruik en werking daarvan. Alhoewel niet altijd even scherp hoe, geven de richtlijnen (EU, nationaal, AVG) wel aan dat over het gebruik en de werking van algoritmes pro-actief gecommuniceerd moet worden naar de betrokkenen. Er is in Rotterdam nog geen vastgestelde visie/beleid voor de (actieve) communicatie naar burgers en betrokkenen inzake het gebruik van algoritmes.

Alhoewel de documenten een goede aanzet geven, vinden we ze in deze vorm niet geschikt voor externe communicatie. De documenten hebben op dit moment vooral de praktische insteek voor het behandelen van bezwaar- en rechtszaken, ook voor dit doel hebben wij kanttekeningen.

Projectverslag

Het projectverslag geeft een helder en goed beeld van de ontwikkeling en werking van het algoritmisch model in de pilotfase (status 2018). Het document is in deze vorm waardevol voor intern gebruik met belangrijke aanbevelingen en lessons learned voor de verdere ontwikkeling van het algoritme, zowel technisch als organisatorisch/beheersmatig.

Voor extern gebruik is het in deze vorm niet geschikt:

- het is niet actueel, en geeft geen beeld van de huidige status van het algoritme (pilot, productie?) en de huidige werkwijze;
- we zien goede elementen inzake 'non bias, -discriminatie, profiling'. Deze zijn vaak impliciet en in technische taal beschreven;



- stevige uitspaken over de kwaliteit van proces en beheer (onvolwassen);
- de dPIA is uit 2016 en wordt op dit moment geactualiseerd.

Dit zijn ook de kanttekeningen voor de behandeling van de bezwaar- en rechtzaken.

- Gebruik de relevante onderdelen van dit document voor het beschrijven van de actuele situatie rond het gebruik en beheer van het algoritme (de aanzet hiervoor staat in het projectverslag, hoofdstuk 9).
- Geef in een aparte paragraaf aan wat de keuzes en (kwaliteits-)maatregelen zijn inzake non-bias, -discriminatie en profiling (dit is ook van belang voor de dPIA).
- Stel vast (W&I en OBI) op welke wijze de aanbevelingen en lessons learned (hoofdstukken 10 en 11) worden aangepakt. Het gaat hierbij niet alleen om risico's van het algoritme, maar vooral ook om de potentiële verdere mogelijkheden daarvan.

Heronderzoeken W&I

Het document 'Heronderzoeken' geeft in algemene termen duidelijk de achtergrond en werkwijze aan van tHO. De rol van het algoritme (Big Data) wordt in algemene termen benoemd, en verwijst voor de details naar het 'projectverslag. Aangegeven wordt 'dat'

W&I met een kritisch blik kijkt naar: de brondata (eventuele aanpassingen hierop), het algoritme/risicomodel (nieuwe data nodig, of wijziging in het model) en de uitkomsten van de heronderzoeken (verdeling leeftijden, wijken, woonsituatie en eventuele bijzondere uitschieters). In het algemeen doet W&I 'continu monitoren met een menselijke blik'. We zien echter niet of en hoe dit zich verhoudt ten aanzien van non-bias, -discriminatie en profiling.

- Geef meer duidelijkheid over het 'hoe' kritisch naar brondata wordt gekeken, alsmede het continue monitoren met een menselijke blik. Relateer dit concreet aan 'vooringenomenheid/discriminatie'.
- Geef ook aan dat en op welke wijze de resultaten van de onderzoeken (thema, 'random', 'algoritme') met elkaar worden vergeleken en hoe rapportage plaatsvindt.

De Leeswijzer

Deze is niet meer nodig met een duidelijke beschrijving van het proces en de actuele situatie, inclusief dPIA/Deda. Teksten hieruit kunnen wel als input dienen.

Adviezen t.a.v. transparantie

- Stel een strategie vast voor externe communicatie over het algoritme. Maak op grond hiervan bewuste keuzes in de wijze waarop burgers en betrokkenen worden geïnformeerd over het gebruik en de werking van het algoritme.
- Maak op basis van de huidige documenten en informatie een heldere beschrijving in duidelijke en begrijpbare taal over:
 - doelstelling (rechtvaardiging) en functie van het algoritme in het proces. Wat doet het algoritme en hoe wordt het in het proces toegepast met welk resultaat. Geef aan wat de rechtvaardiging is voor het gebruik van het algoritme en wat dat betekent voor de gemeente en de burger;
 - input (data die gebruikt wordt door het algoritme; de keuzes en de bewerkingen daarop). Geef aan op welke wijze de 'input' voldoet aan de ethische regels;
 - algoritme (het model, trainen, testen en werking → het resultaat is risico scorelijst ('output')). Geef aan hoe het algoritme zelf niet tot vooringenomenheid leidt (eventueel ook dat dit wordt gecontroleerd);
 - de resultaten van het heronderzoek ('outcome') → analyses van de resultaten (ook die van de andere selecties) en de terugkoppeling daarvan naar/met het tDS;
 - leren en verbeteren van het algoritme.

(Let op, maak wel onderscheid tussen 1) de technische documentatie bedoeld voor ontwikkeling, beheer en toepassing van het algoritme en 2) de informatie voor de externe communicatie).

Overige aandachtspunten (en governance)

De status van het algoritme is, zo lijkt het, in de fase tussen 'pilot' en 'productie'. De 'pilot' is als project afgerond. De huidige werkwijze en toepassing van het algoritme is (grotendeels) hetzelfde als die van de pilot. We zien nog geen bewuste 'formalisering' naar de productiefase met afspraken over eigenaarschap, verantwoordelijkheden en kwaliteitsmaatregelen in een vastgesteld en beschreven proces (met daarbij duidelijk de verantwoordelijkheden in het algemeen en die van de schakels).

- Voor zover dit nog niet is gebeurd, stel vast wat nu de stappen zijn om het algoritme formeel in de productiefase te laten landen; maak hierbij duidelijk wie voor wat verantwoordelijk is en wat de afspraken daarbij zijn voor de kwaliteitsmaatregelen (grotendeels W&I (UBO, HO) en OBI/DWH, Advanced Analytics).
- Geef duidelijk weer waar in de actuele situatie het eigenaarschap en de verantwoordelijkheden liggen in de 'algoritme-keten' (ook in termen van beheer en productie). Waar ligt het 'overall' eigenaarschap over de ethische aspecten in de algoritme-keten? Is dat de proceseigenaar verantwoordelijk voor de toepassing effecten en impact van het algoritme? Is dat Advanced Analytics verantwoordelijk voor generieke algoritmes die verscheidene toepassingen kent? Wat is de mogelijke rol van een concernbrede 'Ethische toetsingscommissie' voor de borging van de ethische randvoorwaarden.

Deze status-situatie geldt ook voor afspraken en (de visie) voor het beheer en de verdere ontwikkeling van het algoritme. Op dit moment zien we nog geen concrete plannen voor de verdere ontwikkeling van het algoritme specifiek voor de toepassing voor W&I of voor breder generiek gebruik binnen de gemeente.

- Geef concreet vorm aan de visie over en plannen voor de verdere ontwikkeling van het algoritme 1) specifiek voor W&I en 2) als generiek AI-instrument. Dit laatste is zeker ook de verantwoordelijkheid van tDS/IIFO. Het gaat hierbij ook om de verdere mogelijkheden van deze nieuwe ontwikkelingen te verkennen.

Zoals eerder is aangegeven, zijn beleid en policies inzake 'vooringenomenheid' en 'transparantie' in ontwikkeling. Dit zijn aandachtspunten die ook op concernniveau spelen. De AI-Visie (IIFO/DGW) geeft aan dat 'non-bias' en 'transparantie' elementen zijn waarop AI (Artificial Intelligence) getoetst moet worden. Concrete handvatten hiervoor zijn er nog niet. Het meest duidelijk hiervoor zijn wel de richtlijnen vanuit de AVG hierover.

Werk (concern-)policies/principes verder uit voor het treffen van kwaliteitsmaatregelen inzake ethiek. Vanuit IIFO (DGW) zijn hier al de eerste stappen ondernomen (waarbij 'corona' wel voor vertraging heeft gezorgd).

- ➔ Rotterdam is regievoerder van het consortium 'publieke controle op algoritmes'. Vanuit dit consortium wordt een format voor een openbaar algoritmeregister opgeleverd met de bedoeling dit in het Rotterdamse beleid inzake algoritmes op te nemen.
- ➔ Er wordt gewerkt aan een kwaliteitsmonitor die ervoor zorgt, dat algoritmes werken zoals beoogd en conform ethische, privacy en business grenzen/doelstellingen 'handelen'.

JD is betrokken bij het algoritme vanwege de bezwaarzaak met onder meer de focus op vooringenomenheid. Niet vanwege eventuele andere risico's die juridische consequenties kunnen hebben (bijvoorbeeld inzake transparantie).

- Denk na over de mogelijkheid om vanuit juridische risico's (breed) de 'does and the don'ts' inzake ethiek te benoemen bij de ontwikkeling en het gebruik van algoritmes. JD moet idealiter aan het begin van de ontwikkeling/ontwerp van het algoritme betrokken zijn.