Retouradres: Postbus 94652, 1090 GR Amsterdam

Gemeente Amsterdam

Commissie Persoonsgegevens

Aan: Directeur Wonen, de heer M. van Engen
C.c: Manager IVE Ruimte& Economie de heer 5.1, 2, e
C.c: De Stedelijk Directeur Ruimte& Economie, de heer H. Jagersma
C.c: CIO AI, de heer 5.1, 2, e
C.c: mevrouw 5.1, 2, e (privacy officer cluster R&E), de heer 5.1, 2, e (Business Analist cluster R&E)

Datum	16 februari 2020
Behandeld door	5.1, 2, e
E-mail	5.1, 2, e @amsterdam.nl
Bijlage	1
Onderwerp	Gebruik van een algoritme voor het prioriteren van meldingen t.b.v. Toezicht (Directie Wonen, Cluster Ruimte& Economie)

Geachte heer van Engen,

Op de vergadering van de Commissie Persoonsgegevens Amsterdam (CPA) van 23 januari 2020 waren bij het agendapunt "Gebruik van een algoritme voor het prioriteren van meldingen t.b.v. Toezicht" namens u aanwezig: mevrouw 5.1, 2, e (privacy officer cluster R&E) en de heer 5.1, 2, e (Business Analist cluster R&E).

Algemene opmerkingen algoritmen
De CPA maakt allereerst een aantal algemene opmerkingen met betrekking tot het gebruik van algoritmen.

1. Op de overheid rust een bijzondere verantwoordelijkheid als er met behulp van nieuwe technologieën gegevens worden verzameld en geanalyseerd. Die bijzondere verantwoordelijkheid houdt in dat dergelijke vormen van 'data gedreven' werken alleen zijn toegestaan als er wordt voorzien in voldoende waarborgen om aantasting van de persoonlijke levenssfeer van de betrokkenen en discriminatie te voorkomen.

2. Over de toepassing van algoritmen heeft de Autoriteit Persoonsgegevens op 17 februari 2020 een notitie gepubliceerd, waarnaar de CPA graag verwijst. De AP wijst erop dat de beginselen van 'rechtmatigheid', 'transparantie' en 'behoorlijkheid' een goed aangrijpingspunt vormen voor de toetsing door de verantwoordelijke voor de verwerking waarbij gebruik wordt gemaakt van algoritmes. Bijzonder aandacht verdient de verantwoordingsplicht (onder 3.2. in de notitie) ingevolge de AVG. De AP wijst erop dat een

organisatie die algoritmes gebruikt vooraf in kaart moet brengen welke risico's er hierdoor voor de rechten en vrijheden van personen ontstaan.

4. Concreet betekent dat vooraf goed over het ontwerp van het systeem nagedacht moet worden: is het gekozen algoritme passend bij het doel waarvoor het wordt ingezet? Ook betekent dit dat de werking van het algoritmen, gepaste maatregelen en waarborgen implementeren getest moet worden voordat het systeem gebruikt wordt. Het gaat daarbij bijvoorbeeld over wat er gebeurt wanneer algoritmes onbedoeld een verkeerde uitkomst geven en de vraag of door het gebruik van algoritmes misschien bepaalde groepen in de samenleving stelselmatig worden gediscrimineerd.

5. Daarnaast moet een verwerkingsverantwoordelijke voorafgaand aan een bepaalde verwerking een beoordeling uitvoeren van het effect hiervan op de bescherming van persoonsgegevens, een zogeheten gegevensbeschermingseffectbeoordeling, ook wel DPIA genoemd. Dit is vereist als de verwerking -gelet op de aard, de omvang, de context en de doeleinden-waarschijnlijk een hoog risico inhoudt voor de rechten en vrijheden van natuurlijke personen.

6. De transparantie krijgt vorm door hetgeen in de AVG is bepaald over de informatie die – proactief of op verzoek -moet worden gegeven over het verwerken van persoonsgegevens in algoritmes aan degenen op wie deze persoonsgegevens betrekking hebben (de 'betrokkene').Zo is het onder andere verplicht om, als er sprake is van profilering en/of geautomatiseerde besluitvorming, dit kenbaar te maken. Daarbij moet in bepaalde gevallen ook nuttige informatie worden gegeven over de onderliggende logica van het algoritme, alsmede het belang en de verwachte gevolgen van die verwerking voor de betrokkene. In het geval van profilering heeft de betrokkene het recht om daartegen bezwaar te maken. Een algoritme beoordelen op behoorlijkheid (fairness) is op zichzelf complex maar mogelijk met de juiste middelen en kennis. De vraag of een algoritme of een uitkomst van een algoritme 'fair' of behoorlijk is, is immers sterk verweven met de omstandigheden van het geval, maar daarnaast ook met mogelijk subjectieve opvattingen over rechtvaardigheid. Een verwerkingsverantwoordelijke zal zelf actief moeten verantwoorden en motiveren waarom een algoritme fair is en het gebruik van het gekozen algoritme niet leidt tot onbehoorlijke uitkomsten. In de uitspraak van het SyRI proces lijkt als waarborgen in ieder geval voldoende transparantie en onafhankelijke toetsing vooraf te beschouwen. Op grond van het EVRM moet er worden voldaan aan 'fair balance', de redelijke verhouding tussen het maatschappelijke belang dat de wetgeving dient en de inbreuk op het privéleven die de wetgeving oplevert om te kunnen spreken over een voldoende gerechtvaardigde inbreuk op het privéleven. Daarbij komen de fundamentele beginselen die ten grondslag liggen aan de bescherming van gegevens op grond van het Unierecht (het Handvest en de AVG) in aanmerking, in het bijzonder de beginselen van transparantie, het doelbindingsbeginsel en het beginsel van dataminimalisatie.

Het transparantiebeginsel vergt toegankelijke en begrijpelijke informatie, communicatie en eenvoudig taalgebruik, en informatieverstrekking aan betrokkenen over de identiteit van de verwerkingsverantwoordelijke en de doeleinden van de verwerking. Los daarvan moet op grond van dit beginsel actief verdere informatie worden verschaft om te zorgen voor een behoorlijke en transparante verwerking van gegevens en moeten natuurlijke personen bewust worden gemaakt van de risico's, regels, waarborgen en rechten in verband met de

verwerking van persoonsgegevens en de wijze waarop zij hun rechten met betrekking tot de verwerking kunnen uitoefenen.

Het doelbindingsbeginsel houdt in dat persoonsgegevens voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden moeten worden verzameld en dat zij vervolgens niet verder op een met die doeleinden onverenigbare wijze mogen worden verwerkt.

Het beginsel van minimale gegevensverwerking vereist dat persoonsgegevens toereikend, ter zake dienend en beperkt zijn tot wat noodzakelijk is voor de doeleinden waarvoor zij worden verwerkt. Er mogen, zoals ook volgt uit het beginsel van opslagbeperking dat de AVG daarnaast kent, niet meer gegevens worden verwerkt dan noodzakelijk is voor het doel van de verwerking.

De AVG bevat ook bepalingen met betrekking tot profilering en een verbod op geautomatiseerde individuele besluitvorming, waaronder profilering. In artikel 4 onderdeel 4 AVG is profilering gedefinieerd als iedere vorm van geautomatiseerde verwerking van persoonsgegevens waarbij aan de hand van persoonsgegevens bepaalde persoonlijke aspecten van een natuurlijk persoon worden geëvalueerd, met name met de bedoeling zijn beroepsprestaties, economische situatie, gezondheid, persoonlijke voorkeuren, interesses, betrouwbaarheid, gedrag, locatie of verplaatsingen te analyseren of te voorspellen. Op grond van artikel 22 AVG geldt een algemeen verbod op volledig geautomatiseerde individuele besluitvorming, met inbegrip van profilering, waaraan voor de betrokkene juridische of anderszins aanmerkelijke gevolgen zijn verbonden

***Proces gebruik algoritme voor* het prioriteren van meldingen t.b.v. toezicht vakantieverhuur**

Op basis van eigen gegevens, openbare informatie en meldingen kan per adres een statistische voorspelling (hierna: voorspelling) van illegale vakantieverhuur worden gedaan. Het uitgangspunt is prioriteit aan te brengen in de reguliere meldingen.

Het algoritme berekent eerst een kans. Dit is een gemiddelde van een zeer groot aantal berekeningen die worden uitgevoerd volgens het principe van een kansboom. Met deze kans kan een voorspelling worden gegenereerd van woonfraude op een adres. De voorspelling wordt gedaan op basis van een reeds ontvangen melding door bijvoorbeeld een burger. Het algoritme creëert geen nieuwe te onderzoeken zaken. Cluster Wonen probeert met het algoritme transparant te zijn in de 'redeneringen'. Met deze algoritme kan worden uitgelegd hoe het algoritme werkt, welke data er wordt gebruikt, welke verbanden belangrijk zijn voor het algoritme en welke methodiek toegepast is die het inzichtelijk maakt hoe het algoritme heeft geredeneerd om tot een voorspelling te komen.

De CPA begrijpt dat de algoritme ervoor moet zorgen dat de medewerkers IT-ondersteuning krijgen om te bepalen welke meldingen het meest waarschijnlijk het waard zijn met voorrang opgevolgd te worden. De beslissing blijft aan de medewerkers. Zij bepalen op welke adressen de volgende dag nader onderzoek gaat plaatsvinden. Daarnaast is er gedurende het proces de gelegenheid voor de bewoners om hun zienswijze te geven voordat een daadwerkelijk besluit genomen wordt. Er is dus geen sprake van een geautomatiseerde besluitvorming.

Wettelijke grondslag
De CPA begrijpt dat de volgende grondslagen gebruikt worden voor het toepassen van de algoritme:

- Artikel 174a Gemeentewet;
- Artikel 14 Woningwet;
- Artikel 8, 21, 32 en 35 Huisvestingswet;
- Artikel 3.1.2 lid 5, 4.2.1 en 4.2.2 van de Huisvestingsverordening; en
- Artikel 6 lid 1 sub e AVG.

Schriftelijke vragen
Na de vergadering van 23 januari zijn er schriftelijke vragen opgestuurd naar directie Wonen. De vragen en antwoorden zijn te vinden in bijlage 1. De CPA heeft de schriftelijke vragen meegenomen in haar advies.

Advies
Allereerst dankt de CPA de aanwezigen voor hun toelichting. De CPA adviseert positief op de pilot van "*Gebruik van een algoritme voor het prioriteren van meldingen t.b.v. Toezicht"*. De commissie adviseert wel om de volgende punten in acht te nemen:

- Het bepalen van welke grondslag gebruikt kan worden voor de dataverzameling, voor het trainen van het algoritme en voor het onderzoek gedeelte van het proces;
- Per gegeven duidelijk toelichten waarom de gegeven nodig is om in het algoritme te verwerken in plaats van alle dataverzameling koppelen aan de algemene handhavingstaak;
- Het maken van een overzicht waarin per maatregel aangegeven wordt voor welk risico die maatregel bedoeld is en daarbij duidelijk vermelden wat er gebeurt als er voor een bepaald risico geen maatregel te treffen valt;
- Het opsturen van de DPIA voor de afdeling Toezicht en Handhaving in Q2;
- Het aanwijzen van een autorisatieverantwoordelijke in het proces;
- In de processtukken het autorisatiegebruik nader toelichten; en
- Raadplegen van de website van de Autoriteit Persoonsgegevens met betrekking tot informatie rondom geautomatiseerde besluitvorming.

De CPA vraagt de aanwezigen over een tijd weer langs te komen voor een terugkoppeling over de voortgang. Verder verwijs ik u naar de notulen van de vergadering over het betreffende agendapunt.

Hoogachtend,

De heer mr. dr. U. van de Pol
Voorzitter Commissie Persoonsgegevens Amsterdam

Bijlage 1
Vragen en antwoorden na de vergadering 23 januari 2020

1) Vindt het leren van het algoritme plaats op een aparte data set (waar bijvoorbeeld ook alle historische dossiers inzitten en dossiers van verschillende teams?). Zo ja, hoe komt de data daar is daar een (eenmalige of dagelijkse of ..) integratie voor? Op welke manier wordt vervolgens de resulterende score/waarschijnlijkheid gedeeld met Toezicht en Handhaving? (5.2, Deel 1, p2: "De geprioriteerde lijst wordt teruggezonden naar de afdeling Handhaving van Wonen"

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering. Uit het zaaksysteem BWV van Toezicht en Handhaving worden een aantal kolommen met informatie geselecteerd van woonfraude zaken van de laatste 5 jaar. De geselecteerde kolommen zijn terug te vinden in bijlage 5a. Deze gegevens worden handmatig in de VAO geplaatst. In de VAO is het algoritme op deze data ontwikkeld en een voorspelling gegenereerd per adres.

In februari wordt gewerkt aan een zogenaamde "webservice". Deze webservice maakt het mogelijk om aan het ontwikkelde algoritme te "vragen" wat de voorspelling is van een nieuwe melding van een woonfraude vermoeden. Deze voorspelling wordt samen met de belangrijkste "variabelen" doorgestuurd naar de planningtool van Toezicht en Handhaving. Medewerkers kunnen zo zien per te onderzoek adres wat de voorspelling is en wat de belangrijkste variabelen hiervoor zijn geweest. Zie hiervoor bijlage 1.

2) Welke andere alternatieven zijn overwogen? En waarom is juist voor een algoritme gekozen? De afgelopen jaren heeft de Directie Wonen, Toezicht en Handhaving de beschikbare capaciteit opgehoogd van 30 medewerkers naar ongeveer 120 medewerkers. Ook met deze uitbreiding is het aantal te onderzoeken adressen groter dan de beschikbare capaciteit.

Het enkel ophogen van de beschikbare capaciteit zorgt er niet voor dat er ook effectiever kan worden gehandhaafd met betere inzichten. Met het beoogde woonfraude algoritme ontstaan extra / nieuwe inzichten waardoor er een stap wordt gezet om effectiever toezicht en handhaving in te zetten.

3) Voor welk specifiek proces is de DPIA geschreven? Baken dit af in de inleiding. Waar begint/eindigt het proces c.q. de verwerking? Hoe verhoudt dat zich tot andere verwerkingen/dpia's?

Er is een DPIA opgesteld voor de afdeling Toezicht en Handhaving. Deze is in concept en wordt in april 2020 afgerond. Er is geen aparte DPIA opgesteld voor het woonfraude algoritme. Er is wel een privacy assessment, kunstmatige intelligentie impact assessment opgesteld en een risico-inventarisatie gemaakt.

4) Op welke wijze worden gegevens vernietigd? Wie, wanneer, hoe? Hoelang mogen/moeten ze bewaard worden voor dit doel? Staaf dit door termijnen uit (basis)selectielijsten c.q. archiefwet!
De data waarmee het model is getraind is alleen voor het trainen van de KI relevant. Het algoritme leert van de trainingsdata welk gewicht aan bepaalde "features" mee moet worden gegeven. Na het trainen wordt alleen het algoritme met de gewichten opgeslagen. De trainingsdata is daarom niet meer nodig en kan worden verwijderd uit de database van de VAO.

Als ervoor wordt gekozen om de KI door te ontwikkelen (bv. met meer data), dan is het wenselijk om de data te blijven gebruiken. Het kan namelijk zijn dat het patroon zal veranderen voor vakantieverhuur. Het algoritme zal verbeteren als er meer data is om op te (her)trainen. De data zal niet langer worden gebruik dan is toegestaan op grond van onderstaande wettelijke bezwaartermijnen.

De wettelijke bewaartermijnen voor gegevens van Toezicht en Handhaving zijn:
- Toezichtsgegevens, onder daarop volgend handhavingsbesluit = 5 jaar
- Bestuurlijkereperatoire sancties (last onder dwangsom/bestuursdwang) = 10 jaar
- Bestuurlijke boete = 7 jaar

Of dit volledig overeenkomst zal nadere moeten worden onderzocht. Bovenstaande termijnen zijn gevalideerd bij de VNG.

5) Autorisaties. Is het echt nodig om alle data scientists toegang te geven tot alle gegevens? Zelfde voor medewerkers uit de bedrijfsvoering? Hoe zit het met beheerders?

Niet alle data scientist of data analisten hebben toegang tot de VAO (zie toelichting bij vraag 6). Dit aantal is beperkt tot:

- Data engineer (1x) heeft toegang tot database
- Data scientisten (2x) en data analist (1x) voeren analyses uit.

Eindgebruikers zien enkel het resultaat (kans op woonfraude en de belangrijkste variabelen). De onderliggende data waarop het algoritme is ontwikkeld mogen en moeten zij ook kunnen inzien vanuit hun wettelijke taak. Het gaat om de volgende rollen:

- Projectmedewerkers / planner
- Toezichthouders
- Projecthandhavers

Na de pilot en afhankelijk van de eindevaluatie wordt vastgesteld hoe het algoritme in beheer moet worden genomen. Tussentijds blijven de data engineer en data scientisten verantwoordelijk voor het beheer.

6) In welke omgeving train je het model? Bijv. eigen pc, virtuele desktop, extra beveiligde omgeving? En hoe doe je de voorbereiding voor het trainen? Hoe ziet het trainingproces er stapsgewijs eruit?

Het algoritme is ontwikkeld in een veilige analyse omgeving (hierna: VAO). In deze VAO is data samengebracht, geanalyseerd en vervolgens een algoritme op ontwikkeld. Binnen de VAO zijn extra beveiligingsmaatregelen getroffen om de gebruikte informatie (persoonsgegevens) te beschermen. Hier heeft een privacy assessment aan ten grondslag gelegen. Er is besloten om deze persoonsgegevens niet te anonimiseren of te pseudonimiseren omdat deze informatie aantoonbaar noodzakelijk is voor het prioriteren van bezoeken voor toezicht. Deze beveiligingsmaatregel is inmiddels wel ontwikkeld en zal worden toegepast binnen het algoritme. Dat zal naar verwachting in Q2 plaatsvinden.

Ter controle op het datamodel is de eerste 80% van de beschikbare data in het algoritme verwerkt om het te "trainen" en te leren welke "variabelen" in verband staan met woonfraude. Het overige gedeelte (20%) is gebruikt om het model te valideren. Bij deze validatie heeft het algoritme wel de features (historische zaken waarvan we de uitkomst weten) ontvangen, maar niet de daadwerkelijke uitkomsten van deze zaken. De uitkomst die de KI genereert is vergeleken met de daadwerkelijke resultaten. Het resultaat van deze validatie is dat de prestaties van het algoritme beter is dan de gestelde benchmark.

7) Is er een autorisatieverantwoordelijke? Is er voorzien in een proces waarbij periodiek wordt getoetst of autorisaties overeenkomen met need to know? Hoevaak wordt getoetst? Wie doet toetsing? Hoe wordt dit geregistreerd? En wat doen we als dit niet geregeld is?

Op dit moment is er nog geen autorisatieverantwoordelijke aangewezen. Vanuit OIS/Datapunt is kritisch gekeken wie toegang mag krijgen tot de VAO en vanuit het verantwoordelijk afdelingshoofd, privacy officers en security officers is goedkeuring gegeven voor de aangewezen rollen . Tijdens de pilot en afhankelijk van de resultaten zal dit punt nader moeten worden uitgewerkt.

8) Hoe werkt het algoritme ongeveer en hoe zijn de 'denk'stappen die zijn gezegd achteraf te volgen door een mens, wordt dit ergens vastgelegd? Zo ja, hoe, wie mag daar bij? Is dat in te zien naast de score bijvoorbeeld? (5.1, par 3.2 Het ontwikkelde algoritme, p. 2)

Het algoritme is ontwikkeld om een antwoord op de vraag te kunnen formuleren of woonfraude kan worden voorspeld. Er is gekozen om een random forest regressie algoritme te ontwikkelen en deze is voorspellend van aard. Het algoritme valt in de categorie "gebied gebonden analyses" die een "risicotaxatie" geeft op adresniveau . Een random forest is een machine learning algoritme dat wordt toegepast op classificatie- en regressievraagstukken. Een random forest 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, het kunnen schatten van kansen op gebeurtenissen, het detecteren van verdachte gebeurtenissen of veranderingen in een dataset te ontdekken.

Het ontwikkelde algoritme genereert op basis van de omstandigheden van het geval een risico-percentage voor woonfraude op een bepaald adres. Deze voorspelling wordt berekend voor een melding die is ontvangen door een burger. Het algoritme creëert dus geen "nieuwe" te onderzoeken zaken, maar geeft een voorspelling van woonfraude voor een

reeds ontvangen melding. Het algoritme berekent eerst een "kans", wat een gemiddelde is van een zeer groot aantal berekeningen (middels kansbomen) dat het model heeft uitgevoerd. Het risico-percentage die daar vervolgens statistisch mee wordt berekend kan worden gebruikt voor het prioriteren van zaken. Het algoritme geeft ook uitleg over hoe die kans tot stand is gekomen zoals hiervoor al beschreven.

Er zijn slechts enkele data scientists / engineers bevoegd om het algoritme en de onderliggende data aan te passen en in te zien.

9) Waarom kan het algoritme de waarschijnlijkheid niet berekenen met gepseudonimiseerde persoonsgegevens? Waarom is bijvoorbeeld geslacht nodig?

Deze vraag uitvoerig besproken tijdens de CPA-vergadering.

10) In het werkproces vakantieverhuur wordt beschreven dat de medewerkers op dit moment zelf business rules toepassen om te prioriteren. Hoe gaat dit in zijn werk? Is dit volledig handmatig (checklijstje?) of zit hier al enige vorm van automatiseren of workflow in? In hoeverre wijken de business rules af van de logica van het algoritme?

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering. Hier is een set business rules (zie bijlage 1) besproken. Deze business rules zijn voor een deel opgenomen in het algoritme. Dat zorgt ervoor dat het algoritme een betere weergave is van de werkprocessen van Toezicht en Handhaving en mogelijk een betere voorspelling genereert.

11) Noodzaak data voor het bereiken van het doel ontbreekt. Dit ziet op de betrouwbaarheid van het algoritme. Kan dit worden onderbouwd aan de hand van use cases? Is er getoetst met behulp van trainingsdata?

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering. De handhavingscapaciteit is geringer dan het aantal te onderzoeken adressen. De afdeling Toezicht en Handhaving is de afgelopen jaar fors uitgebreid. De behoefte blijft daarentegen bestaan om efficiënter te kunnen werken. Een woonfraude algoritme kan daaraan bijdragen.

Ter controle op het datamodel is de eerste 80% van de beschikbare data in het algoritme verwerkt om het te "trainen" en te leren welke "variabelen" in verband staan met woonfraude. Het overige gedeelte (20%) is gebruikt om het model te valideren. Bij deze validatie heeft het algoritme wel de features (historische zaken waarvan we de uitkomst weten) ontvangen, maar niet de daadwerkelijke uitkomsten van deze zaken. De uitkomst die de KI genereert is vergeleken met de daadwerkelijke resultaten. Het resultaat van deze validatie is dat de prestaties van het algoritme beter is dan de gestelde benchmark.

12) Er is sprake van een juridische analyse die is/wordt uitgevoerd door DJZ en/of Pels Rijcken. Is deze analyse beschikbaar?

DJZ heeft een juridische toets gedaan op basis van de opgestelde stukken en akkoord gegeven. DJZ heeft geen eigen analyse opgesteld. De juridische kaders zijn in de "Kunstmatige Intelligentie Impact Assessment" (bijlage 3) uitgewerkt. Pels Rijcken is niet meer verzocht om een juridisch advies.

13) Er lijkt sprake te zijn van het verwerken van gegevens voor een ander doeleinde dan waarvoor de gegevens zijn verzameld. In dat geval is een toets conform artikel 6 lid 4 AVG noodzakelijk.

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering

14) In WPD document wordt als grondslag art. 6 lid 1 e en f genoemd. Dit ziet op de handhaving door gemeente resp. corporatie maar niet op het gebruik van het algoritme. Zie ook onder 3.

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering

15) Worden betrokkenen geïnformeerd over de pilot? Zo ja, wanneer en op welke wijze?

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering. Betrokken worden op verschillende manier geïnformeerd. Dit zal plaatsvinden door:

- Een persbericht van de gemeente Amsterdam na het collegebesluit;
- Een artikel van een gerenommeerde krant wordt gepubliceerd na het collegebesluit;
- Er wordt een animatie / filmpje opgesteld waarin uitleg wordt gegeven over de Pilot;
- De privacy statement wordt aangepast en er zal op de gemeentelijk website extra informatie worden gegeven over de pilot fase. Deze is te zien na publicatie van het besluit.

16) In het WPD document staat dat advies is ingewonnen bij de fg en dat dit ertoe heeft geleid dat er extra maatregelen zijn getroffen op het gebied van informatiebeveiliging in de VAO. Dat is ons niet bekend, hoe is dat in z'n werk gegaan?

Dit heeft plaatsgevonden in maart 2019. De CPA is hierover niet geïnformeerd.

17) Valkuil is idd dat steeds hetzelfde soort zaken wordt aangepakt, dat zou nog verder uitgewerkt kunnen worden.

Deze vraag is uitvoerig besproken tijdens de CPA-vergadering. Er zullen op termijn ook zaken worden onderzocht met een lage voorspelling van woonfraude op een adres. In de eerste fase van de pilot zullen hoofdzakelijk de zaken met een hoge voorspelling van woonfraude worden onderzocht. Dit wordt gedaan om het werkproces zoveel mogelijk hetzelfde te houden / zo min mogelijk te beïnvloeden zodat een goede vergelijking kan worden gemaakt met eerdere periodes.

18) Er staat in de stukken dat geselecteerde adressen in Word geplakt worden, en dan in excel en dan in de agenda's en op TamTam. Is TamTam wel een goed middel hiervoor en hoe zit het met de BIV rating van deze informatie?

TamTam wordt momenteel gebruikt in de bedrijfsvoering van Toezicht en Handhaving. Naar verwachting zal vanaf Q3 TamTam niet meer worden gebruikt, vanwege het feit dat gegevens worden opgeslagen op een Amerikaanse server. TamTam zal verdwijnen ten gevolge van de ontwikkelingen binnen het Programma Moderniseren Wonen. Binnen het programma wordt gewerkt aan een nieuwe planningstool en zaaksysteem. Daarmee is

TamTam overbodig geworden. De privacy en security officer zijn nauw betrokken bij deze ontwikkelingen.

19) Kunnen, op basis van de gebruikte data en dit type algoritme, personen ten onrechte in verband worden gebracht met woonfraude door deze vorm van gegevensverwerking?

Het model voorspelt de hitkans op vakantieverhuur per adres (woonfraude). Hiervoor neemt het model onder andere data mee over de personen die op het adres wonen op het moment dat er woonfraude wordt geconstateerd of dat hier een signaal op binnen kwam. Als een persoon verhuist naar een nieuw adres en daar komt een melding op, dan zal het model meenemen dat deze persoon op zijn/haar vorige adres met woonfraude in aanraking is geweest. De reden dat het model een hogere hitkans toe kent als een persoon eerder in contact is geweest met woonfraude, komt voort uit het feit dat dit gegeven historisch gezien tot een hogere kans op woonfraude leidt. Deze bevinding is ook bevestigd door de projectmedewerkers van de afdeling Wonen. Zij gebruiken dit feit ook als een reden om een nieuw signaal wel of niet te onderzoeken.

Enkel het feit dat een persoon op een adres heeft gewoond waar woonfraude heeft gepleegd, zal geen reden zijn voor een hoge hitkans. Hiervoor zullen meerdere factoren, welke in het verleden in relatie stonden tot woonfraude, samen moeten komen.

20) Als een bewoner wijzigt van adres, is dit dan wel of niet zichtbaar in het algoritme?

In de dataset is de historie van een persoon meegenomen. Het model kan daarom leren van het feit dat iemand vroeger op een adres woonde waar woonfraude is geconstateerd door de gemeente Amsterdam. Het model neemt niet elke verhuizing van alle inwoners mee.

Commissie Persoonsgegevens Amsterdam
Uittreksel van het verslag van het openbare agendapunt van de vergadering op donderdag 23 januari 2020

1. Gebruik van een algoritme voor het prioriteren van meldingen t.b.v. Toezicht (Cluster R&E, directie Wonen)

De voorzitter zegt te verwachten dat niet alle vragen in deze sessie zullen kunnen worden beantwoord, maar dat resterende vragen naderhand schriftelijk worden afgehandeld.

De indieners hebben deze kwestie met veel documenten aan de CPA voorgelegd omdat het gebruik van algoritmes in toenemende mate ter discussie staat in de media en in de Europese Commissie. De EC vraagt om een assessment van de impact van kunstmatige intelligentie (KIIA), maar daar zijn nog geen standaardvragen voor vastgesteld. Daarom heeft men voor dit project een toets gemaakt op basis van kennis die via openbare bronnen is te achterhalen. Die KIIA wordt nu aan de commissie voorgelegd voor advies met daarbij de vraag welke bestuurlijke procedure doorlopen zou moeten worden om dit proces goed af te wikkelen.
Op zich bevat het algoritme niet echt nieuwe dingen. In de huidige praktijk komen alle meldingen over mogelijke woonfraude terecht bij de medewerkers van Handhaving en Toezicht die vervolgens op grond van hun ervaring en inzichten komen tot een prioritering. Dat laatste is die nodig omdat het aantal meldingen de toezichtscapaciteit overtreft. Men verwacht dat het algoritme kan helpen om tot een objectievere en effectievere prioritering te komen. Er is geen sprake van echte profilering en als dat al gebeurt, vindt dat plaats naar adressen en niet naar personen. Het gaat erom te komen tot prioritering, niet tot voorspellingen. In het algoritme zelf zijn echter wel degelijk persoonsgegevens verwerkt. De pseudonimisering in het algoritme moet nog geregeld worden.
Ter aanvulling op de verstrekte documenten wordt ter plekke nog een update aangeleverd met onder meer de 17 business rules voor prioritering bij verschillende typen meldingen. Het algoritme moet gaan zorgen dat de medewerkers IT-ondersteuning krijgen om te bepalen welke meldingen het meest waarschijnlijk het waard zijn met voorrang opgevolgd te worden. De beslissing blijft aan de medewerkers. Zij bepalen op welke adressen de volgende dag nader onderzoek gaat plaatsvinden.

De CPA vraagt welke waarborgen er zijn dat de handhaver nog tot een autonoom oordeel kan komen als het algoritme een bepaald adres al bestempeld heeft en of er niet het gevaar bestaat dat men vervalt in protocollisering.

De selectie van een adres is de eerste stap in het handhavingsproces. Daarna gaan een professionals die het veld en de wijk kennen, ter plekke op onderzoek uit. Zij tekenen op wat zij daar waarnemen zonder er een oordeel aan te verbinden. Er treedt dan een proces in werking met veel menselijke toetsmomenten en er is gedurende het proces gelegenheid voor de bewoners om hun zienswijze te geven voordat een daadwerkelijk besluit genomen wordt. De handhavers houden hun eigen verantwoordelijkheid en worden getraind om risico's voor vertekende waarnemingen te onderkennen. Het wordt zeer zeker geen blindvaren op de uitkomsten van het algoritme. Gedurende de pilot zal er voortdurend gemonitord worden en worden bevindingen tweewekelijks met de PO en het data-science-

team besproken. Dan gaat het er niet alleen om of het algoritme klopt, maar ook om hoe de medewerkers het werken met het algoritme ervaren. Er zal ook doelgericht gekeken worden of een groep meldingen met een lage prioriteit in de praktijk terecht die lage score gekregen heeft. Voor de proef wordt nog niet gewerkt met een controlegroep zonder algoritme, maar dat zal daarna waarschijnlijk wel een keer gebeuren.

De CPA merkt op dat de focus niet alleen op het adres is, omdat bij de business rules staat opgenomen dat een melding een hogere prioriteit dient te krijgen als er een persoon staat ingeschreven met een voorgeschiedenis van woonfraude..

Het gaat volgens de indieners uiteindelijk wel om het adres. De Huisvestingswet biedt voldoende wettelijke grondslag om de fraudehistorie van een bewoner over de vijf jaar vóór de melding, op dit of een ander adres, mee te wegen in de prioritering.
De bedoeling is de prioritering verder ook door te voeren en te laten terugkomen in het werkproces rond handhaving zodat die ook helpt de voorraadplanning beter te organiseren. In het net aangeleverde document valt op pagina 4 te zien welke karakteristieken tot nu toe belangrijk zijn gebleken om te duiden of een melding daadwerkelijk op een woonfraude wijst. Gaandeweg zal men kunnen bepalen welke factoren en gegevens echt relevant zijn voor het model en welke niet. Die laatste kan men dan verder weglaten in de dataverzameling. Op dit moment heeft men bijvoorbeeld uit ervaringen opgemaakt dat de kans dat er bij een huis met erg jonge kinderen sprake is van woonfraude, erg gering is. Het algoritme zou kunnen uitwijzen dat die aanname niet terecht is. De primaire database komt uit de BWV. Het algoritme zal patronen en verbanden gaan ontdekken en zichzelf daar ook maandelijks op trainen.

De CPA adviseert nog eens goed te kijken naar de grondslag voor de dataverzameling vanuit de BWV en voor het trainen van het model.

De dataverzameling hoort bij de algemene wettelijke taak van handhaving en toezicht. De BWV dateert echter nog uit het pre-digitale tijdperk en biedt waarschijnlijk net te weinig houvast voor dit grensgeval van onderzoek en handhaving. Wellicht kan de Archiefwet op dat punt uitkomst bieden. Het gaat erom precies te bepalen welke grondslag er is voor het element van onderzoek. Dat moet nog gebeuren.

De CPA adviseert per gegeven duidelijk toe te lichten waarom men dat in het algoritme wil verwerken in plaats van zoals nu alle dataverzameling te koppelen aan de algemene handhavingstaak. Verder adviseert men een overzicht te maken waarin per maatregel aangegeven wordt voor welk risico die bedoeld is en wat er gebeurt als er voor een bepaald risico geen maatregel te treffen valt.
De CPA adviseert ook nog eens goed te kijken naar de website van de AP over veel gestelde vragen rond geautomatiseerde besluitvorming.

Uit het aangeleverde validatieplan blijkt dat tussen de 20 en 30% van de meldingen van vakantieverhuur daadwerkelijk tot een handhavingsactie leidt. De vraag is of die resterende 70% dan wel met die 17 business rules gedekt wordt.

Dat is in de praktijk lastig vast te stellen omdat het in veel gevallen onmogelijk blijkt met drie bezoeken ter plekke tot een feitelijke constatering te komen omdat het niet lukt in de

betreffende woning binnen te treden. Wel wordt genoteerd als er volgens eigen waarneming van buitenaf geen sprake is van fraude. In het werkproces wordt voorzien in drie pogingen van de handhaver om iemand thuis aan te treffen. Daarna wordt mede vanwege capaciteitsoverwegingen verder afgezien van die aanpak en wordt genoteerd dat er geen feitelijke constatering mogelijk was. De cijfers uit het validatieplan zijn te zien als een nulmeting. De proef zal uitwijzen of het hanteren van het algoritme leidt tot stijgingen en dalingen en of sommige business rules wel zin hebben.

De voorzitter merkt op dat de schriftelijke vragen uit het ambtelijk vooroverleg nog met de secretaris doorgenomen moeten worden buiten deze vergadering. Hij is benieuwd hoe de planning voor de pilot nu verder gaat.

Met de wethouder en het MT is afgesproken dat men pas aan de slag mag gaan als er een positief advies van de CPA ligt. Vanwege de ervaringen met SyRI gaat men pas daadwerkelijk aan de slag als er een collegebesluit genomen is. Het CPA-advies wordt bij het collegevoorstel gevoegd. Er wordt daarna voorzien in communicatie naar de burgers via de website, filmpjes en persberichten.

De voorzitter concludeert dat er nog vragen afgewikkeld moeten worden voor het schriftelijke advies. Hij bedankt voor de toelichting en vraagt over een tijd weer eens een terugkoppeling te komen geven over de voortgang.