



Starter Kit Fair Algorithms

ONS-UNECE Machine Learning 2021 Group

Sigrid van Hoek (Statistics Netherlands)

March 22, 2021

Impact with AI

Partners: Ministry of Internal Affairs, municipality of Amsterdam, VNG, CodeForNL

Fair Algorithms project:

- Research projects
 - Master thesis Fairness by Archiving August 2019
 - CBS Working paper Fair Algorithms in Context April 2020
 - Article Improving Fair Predictions Using Variational Inference In Causal Models August 2020
- Starter kit login (starter, I3dcRb2) will be hosted by Utrecht Data School
- Dashboard (tab 'Dashboard' in Starter Kit)

Study on fair algorithms for policy-making

14/06/2019 10:00 / Author: Miriam van der Sangen



© Tijdschrift van der Rijck Fotografie

In Dutch society, more and more decisions are being made by automated systems which are based on algorithms. But how fair is the use of algorithms? This question was studied on behalf of the City of Amsterdam by Nik Kievelwegen, researcher at the University of Amsterdam (UvA). He used various databases of Statistics Netherlands (CBS) in his research.

Data-driven approach

A large Dutch municipality recently made the news because it deploys algorithms to detect social welfare scheme fraud. This caused protests among dozens of citizens as it occurred in neighbourhoods with a relatively high migrant population, which was considered to be

Need for fair algorithms

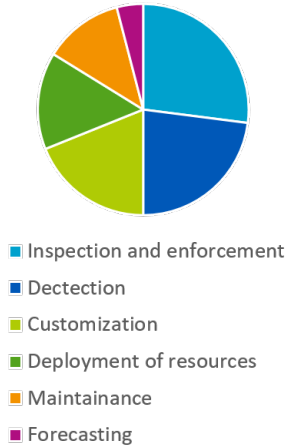


Figure: Veenstra et al. (2019). *Quick Scan AI in de publieke dienstverlening*. TNO



Fair data

Types of biases

1. Sample bias: model trained on white males
2. Exclusion bias: unable to measure important information
3. Measurement bias: e.g. multiple camera types for image recognition
4. Recall bias: wrong labels (type of measurement bias)
5. Observer bias: confirmation bias
6. Association bias: feedback-loops

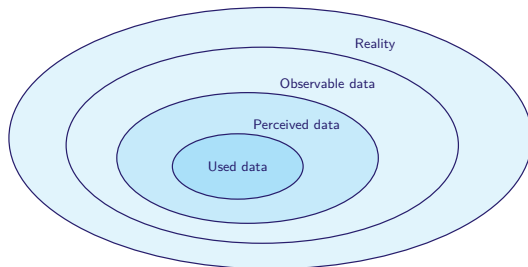
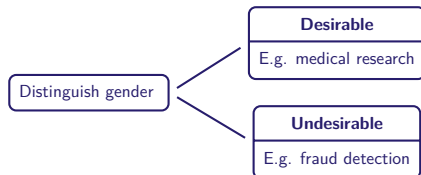


Figure: Context of fair data



Racial bias: problem?



Fairness: *results should not be dependent of sensitive attributes.*

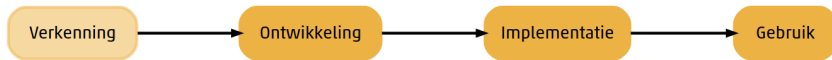
Fairness criteria for binary classification:

Let \hat{Y} be the predicted label, Y the true label, X features, A sensitive attribute, $C = f(X, A)$.

- Equalized odds: $\mathbb{P}_0[C = c|Y = y] = \mathbb{P}_1[C = c|Y = y] \quad \forall c, y \in \{0, 1\}$
- Predictive rate parity: $\mathbb{P}_0[Y = y|C = c] = \mathbb{P}_1[Y = y|C = c]$
- Counterfactual Fairness: $\mathbb{P}[C_{A \leftarrow 0} = c|X, A = a] = \mathbb{P}[C_{A \leftarrow 1} = c|X, A = a]$



Selecteer een proces



Technische vragen

Verkenning

In het verkennende proces staan de legimiteit en voorwaarden van het project centraal. In deze fase wordt onder andere in kaart gebracht of er data is van voldoende kwaliteit, wat de risico's zijn en of het project voldoet aan wet- en regelgeving. Zowel de bestuurder, de project manager, de data analyst als de privacy en security officer zijn betrokken bij dit proces.

Context

Wordt het ontwikkelen van het algoritme extern uitbesteed?

- Nee
 Ja

Is er een verwerkersovereenkomst (VVO) opgesteld?

- Nee
 Ja

Antwoord: Op basis van de Algemene Verordening Gegevensbescherming (AVG) is het wettelijk verplicht een VVO af te sluiten met alle opdrachtnemers die namens hen persoonsgegevens verwerken. Er is een standaard VVO beschikbaar via o.a. de website van VNG.

Is er voldoende aandacht besteed aan de beveiliging van (persoons-)gegevens?

- Nee
 Ja

Heeft u de ethische aspecten van het project in kaart gebracht. Nieuwsgierig met hulp van De Ethische Data Assistent

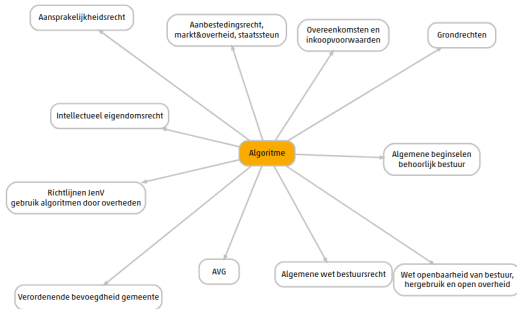
Notities

voortgang verkenning

9%

Ruimte voor eventuele notities

Selecteer een juridisch element



Wet- en regelgeving

Selecteer hieronder een juridisch aspect voor meer informatie:

AVG

Algemene verordening gegevensbescherming (AVG)

⚠ Dit is een voorzet naar een juridisch kader voor het gebruik van algoritmen die persoonsgegevens verwerken. Het is (nog) niet volledig. Volgens art. 2 AVG is de hele AVG van toepassing indien er sprake is van geheel of gedeeltelijke geautomatiseerde verwerkingen. Onderwerpen als aansprakelijkheid en bewaartermijnen zijn hierin niet meegenomen. Overigens heeft de Nederlandse wetgever in art. 34 AVG de schriftelijke beslissing van een bestuursorgaan op een verzoek als bedoeld in de artikelen 15 t/m 22 AVG, gelijkgesteld met een besluit in de zin van art 1:3 Awb.

Worden er met het algoritme op enige wijze persoonlijke gegevens verwerkt?

Nee

Ja



Wet- en regelgeving

Selecteer hieronder een juridisch aspect voor meer informatie:

AVG

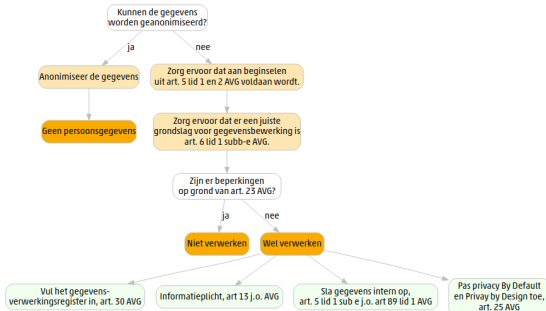
Algemene verordening gegevensbescherming (AVG)

! Dit is een voorzet naar een juridisch kader voor het gebruik van algoritmen die persoonsgegevens verwerken. Het is (nog) niet volledig. Volgens art. 2 AVG is de hele AVG van toepassing indien er sprake is van geheel of gedeeltelijke geautomatiseerde verwerkingen. Onderwerpen als aansprakelijkheid en bewaartermijnen zijn hierin niet meegenomen. Overigens heeft de Nederlandse wetgever in art. 34 AVG de schriftelijke beslissing van een bestuursorgaan op een verzoek als bedoeld in de artikelen 15 t/m 22 AVG, gelijkgesteld met een besluit in de zin van art 1:3 Awb.

Worden er met het algoritme op enige wijze persoonlijke gegevens verwerkt?

Nee

Ja



Informatie

Externe documenten

Er zijn verschillende ethische richtlijnen opgesteld voor betrouwbare en eerlijke algoritmen. Deze richtlijnen, en vooral de richtlijnen van het ministerie van Justitie en Veiligheid hebben we verwerkt in de vragen in het tabblad techniek en beleid. Ter informatie vind u de links naar verschillende richtlijnen hieronder:

Bijlage bij brief over waarborgen tegen risico's van data analyses door de overheid

Beschrijving

Deze bijlage bij de kamerbrief van 8 oktober 2018 bevat onder meer richtlijnen voor de toepassing van algoritmen door de overheid. De richtlijnen zijn relevant voor data-analyses in brede zin en bevatten concrete en op de stand van de technologie afgestemde aanwijzingen, bedoeld om het inzicht in, de transparantie en de kwaliteit van algoritmen en data-analyses door overheidsinstanties te vergroten.

Organisatie

Ministerie van Justitie en Veiligheid

Referentie

<https://www.rijksoverheid.nl/documenten/rapporten/2019/10/08/tk-bijlage-over-waarborgen-tegen-risico-s-van-data-analyses-door-de-overheid>

Ethical Guidelines for Thrustworthy AI

AI Government Procurement Guidelines

Tevens zijn er verschillende worksheets, handleidingen en vragenlijsten beschikbaar. Deze bevatting nuttige informatie over bijvoorbeeld de socio-economische impact van AI. Verschillende bronnen hieronder zijn ook genoemd in de vragen van het tabblad techniek.

De Ethische Data Assistent (DEDA)

Guidelines on Data Protection Impact Assessment (DPIA)

Handreiking Standaard Verwerkersovereenkomst Gemeenten (VWO)

Feedback


De wet- en regelgeving omtrend algoritmen is continu in ontwikkeling. Het is daarom belangrijk om de starterskit niet als vaststaand product te zien. Wij zijn dan ook erg benieuwd naar uw feedback: wat is uw impressie van de starterskit en welke vragen wilt u graag terugzien?

Vul hier uw algemene feedback in over de vorm en inhoud van de starterskit:

Vul hier een vraag in die u graag in de starterskit ziet terugkomen:

Aan welk onderwerp mag volgens u meer aandacht besteed worden?

- Introductie: bevat o.a. doel en gebruik van de Starterskit
- Techniek: verkenning, ontwikkeling, implementatie en/of gebruik
- Beleid: o.a. juridische aspecten
- Informatie: literatuur en/of definities
- Download
- Feedback

 Verstuur feedback

Conclusion

Most important elements:

- Data awareness (quality, context, observation methods)
- Domain knowledge
- Criteria for (individual) fairness
- Performance trade-off: racial bias versus utility

Future research:

- Communication internal and external
- Interaction with policy
- Examples of ethical data projects
- Continue research into fair AI



Contact information

Sigrid van Hoek (Data Scientist, st.vanhoek@cbs.nl)

Anna Mitriaieva (Project Leader, a.mitriaieva@cbs.nl)

Barteld Braaksma (Innovation Manager, b.braaksma@cbs.nl)

