

CASE 1: PREDICTIVE POLICING

Vandaag wordt vanuit criminologische hoek met veel interesse gekeken naar de toepassingsmogelijkheden van AI op vlak van criminaliteitsbeheersing- en bestrijding. In deze case wordt ingegaan op de mogelijkheden van *predictive policing*.

Kunnen we misdaad voorspellen en voorkomen? Het lijkt fictie, maar er wordt al volop mee geëxperimenteerd. Hoe gaat *predictive policing* in zijn werk? Wat zijn de voordelen en gevaren? Hoe veranderen nieuwe technologieën het werk van de politie?

In deze case staan we stil bij de uitdagingen in het omgaan met criminaliteitsvoorspellingen en welke impact deze ontwikkelingen kunnen hebben op de maatschappij. We zullen enerzijds bestuderen welke predictieve modellen het meest geschikt zijn om criminaliteit te voorspellen op basis van data uit verschillende organisaties, en welke technische problemen daarbij opduiken. We bekijken anderzijds ook de ethische vraag hoe wenselijk het is om dergelijke modellen in de praktijk te gebruiken.

Je kan hier (<https://dwars.be/nl/artikel/een-glazen-bol-voor-de-politie>) een populariserend artikel lezen van prof. Wim Hardyns over dit onderwerp en een Wetenschapscafé voor Wetenschap Centraal bekijken (<https://www.youtube.com/watch?v=KNc1vPXF5ig>).

Bronnen:

Rummens, A. & Hardyns, W. (2020). The effect of spatiotemporal resolution on predictive policing model performance. *International Journal of Forecasting* (online first).

<https://doi.org/10.1016/j.ijforecast.2020.03.006>

Rummens, A. & Hardyns, W. (2020). Comparison of near-repeat, machine learning and risk terrain modeling for making spatiotemporal predictions of crime. *Applied Spatial Analysis and Policy* (online first). <https://doi.org/10.1007/s12061-020-09339-2>

Hardyns, W. & Rummens, A. (2018). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research*, 24(3), 201-218. doi: 10.1007/s10610-017-9361-2

Rummens, A., Hardyns, W. & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. *Applied Geography*, 86, 255-261. doi: 10.1016/j.apgeog.2017.06.011

CASE 2: RACIALE BIAS IN GEZICHTSHERKENNING

Binnen het domein van “machinaal leren” is er de afgelopen jaren een grote vooruitgang geboekt, onder meer op vlak van object- en gezichtsherkenning. Bij gezichtsherkenning worden gezichten gescand om het geslacht van een persoon, het ras, emoties en zelfs de identiteit te bepalen. Jammer genoeg hangen de prestaties van deze technieken sterk af van het gezichtstype. Verschillende onderzoekers hebben een “raciale bias” vastgesteld bij gezichtsherkenningstechnieken: een studie van het MIT toont aan dat bij 3 algoritmes de nauwkeurigheid 99.2% bedraagt voor mensen met een lichte huidskleur, terwijl voor mensen met donkere huidskleur de nauwkeurigheid afneemt tot 80% en zelfs 67%.

Hoewel het onderzoek over eerlijkheid van machinale leertechnieken in het algemeen nog niet zo ver staat, worden gezichtsherkenningstechnieken al ruim ingezet door politiediensten om verdachten te identificeren. De prestaties van de algoritmes spelen dan een zeer belangrijke rol: als de resulterende herkenning een bias vertoont kan dit leiden tot verkeerde arrestaties en interventies. Bovendien kunnen deze interventies dan opgenomen worden in dezelfde datasets die men gebruikt om dergelijke technieken te ontwikkelen. Dit creëert negatieve feedbacklusen die bestaande biases uitvergrooten.

In januari 2020 heeft de politie van Detroit zo verkeerdelijk een willekeurige man gearresteerd die geïdentificeerd was op basis van gezichtsherkenning op camerabewakingsbeelden, dit terwijl zijn jonge kinderen toekeken. Hij werd vastgehouden voor 30 uren, en nadien vrijgelaten wegens gebrek aan bewijs. Het zogenaamde COMPAS-programma dat door het gerecht in de VS wordt gebruikt om risico op recidivisme te voorspellen, vertoont eveneens een grote bias: voor hetzelfde misdrijf onder vergelijkbare omstandigheden voorspelt het een veel hoger risico voor zwarte beklaagden dan voor witte beklaagden.

Taak

Deze case heeft als doel om een applicatie te ontwikkelen die aantoont dat raciale bias optreedt in gezichtsherkenningstechnieken, zelfs wanneer de datasets zorgvuldig worden samengesteld. Een bestaand gezichtsherkenning algoritme zal eerst getraind worden op een dataset van gezichten van personen. De applicatie zal enerzijds de mate waarin raciale bias voorkomt in de dataset kwantificeren en anderzijds ook voorbeelden geven van raciale bias (bijvoorbeeld, twee personen die het algoritme niet gemakkelijk van elkaar kan onderscheiden).

De studenten zullen verder ook verschillende vormen van bias bestuderen die de eerlijkheid van machinale leertechnieken bepalen (discriminatie op basis van invoer, representatie of predictie). Verder willen we nagaan of het mogelijk is om het herkenningssysteem bij te sturen om het biasprobleem te vermijden. De studenten zullen de assumpties van traditionele classificatie in vraag moeten stellen. Ze kunnen hiervoor andere paradigma's verkennen (zoals causale inferentie) of argumenteren dat het gebruik van geautomatiseerde systemen fundamenteel onwenselijk is in bepaalde contexten.

Interdisciplinaire samenwerking zal hier een grote rol spelen, aangezien in de literatuur veel verschillende definities van “eerlijkheid” worden gehanteerd die niet overeenstemmen met de manier waarop dit begrip wettelijk is vastgelegd. De correcte vertaling van juridische concepten naar technische specificaties is een uitdagend en belangrijk onderdeel van dit probleem.

Bronnen:

"Facial recognition to 'predict criminals' sparks row over AI bias", BBC News, June 24, 2020, <https://www.bbc.com/news/technology-53165286>

Allyn, Bobby. "The Computer Got It Wrong': How Facial Recognition Led To False Arrest Of Black Man." June 24, 2020. <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig>.

Du, Mengnan, et al. "Fairness in deep learning: A computational perspective." IEEE Intelligent Systems (2020).

"Machine Bias", ProPublica, May 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

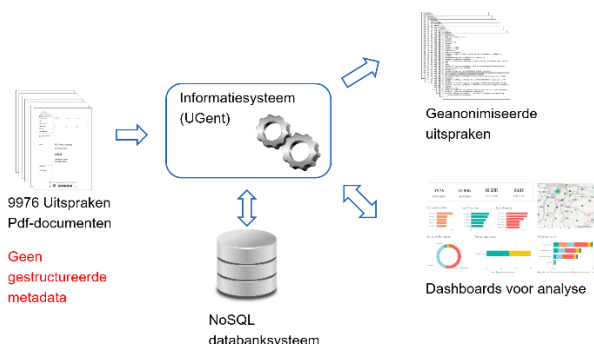
O'Neil, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

Barocas, Solon et al. Fairness and Machine Learning. <https://fairmlbook.org/>

CASE 3: TEKSTANONIMISERING VAN VONNISSEN

Tekstanonimisering is een soort datasanering, die bedoeld is voor privacybescherming. Het is een proces waarbij alle persoonsgegevens die voorkomen in een tekst zodanig worden gewijzigd dat het onmogelijk wordt om de betrokkenen te identificeren.

Via een samenwerking tussen de faculteit Ingenieurswetenschappen en Architectuur en de faculteit Recht en Criminologie wordt momenteel gewerkt aan een informatiesysteem om arresten van de Vlaamse Hoven van Beroep (semi)automatisch te anonimiseren. Een dergelijk informatiesysteem is heel belangrijk wanneer men arresten en vonnissen publiek toegankelijk wil maken.



Het beoogde informatiesysteem wordt gevoed met arresten in pdf-formaat en verwerkt de inhoud van deze pdf-documenten op een semantische manier. De geïdentificeerde semantische tekstcomponenten worden bewaard in een NoSQL databanksysteem en geven onder meer aan welke tekstcomponenten duiden op personen, bedrijfsnamen en adressen. Persoonsgebonden informatie wordt vervolgens geannoteerd met een label dat de rol van een persoon binnen het arrest aangeeft. Bv. 'magistraat_1', 'geïntimeerde_1', 'advocaat_1_geïntimeerde_1', 'griffier_1', 'derdepartij_4', etc. Deze labels worden dan tenslotte gebruikt om de tekst te anonimiseren en daarbij toch nog een deel van de leesbaarheid te garanderen.

Een probleem bij het annoteren is dat namen en woorden door schrijffouten en afkortingen soms op een andere wijze worden herhaald in de tekst en daardoor onterecht aan een nieuwe persoon worden toegeschreven. Bv. 'Verdonck' en 'Verdonk' worden in het geval van een tyffout als twee verschillende personen gezien. Dit probleem is in de literatuur omschreven als een *coreferentieprobleem*.

De bedoeling van het werk van de studenten is om na te gaan hoe bestaande coreferentiedetectietechnieken kunnen worden aangewend in het anonimiseringsproces om zo tot betere resultaten te komen. Het werk omvat een korte literatuurstudie, het experimenteren met code die voorhanden is in de onderzoeksgroep en het ontwerpen en implementeren van een mogelijke oplossing.

Bronnen:

<https://journalist.be/2021/02/openbaarheid-gerechtelijke-uitspraken-nood-aan-een-debat>

Keereman A., 'Ooit zullen algoritmes een voorstel van uitspraak doen, maar nu nog niet'. Juristenkrant 360 (2017), 8-9. Zie <https://www.jurisquare.be/nl/journal/djk/2017-360/> (toegankelijk via UGent netwerk of via VPN verbinding).

CASE 4: KLEINSCHALIGE DETENTIEHUIZEN

We voorzien volgende drie thema's:

1. Op zoek naar locaties voor detentiehuisen in het Gentse: we ontwikkelen een geavanceerde beslissingstool of -matrix waarmee we voor leegstaande locaties kunnen kijken of ze al dan niet in aanmerking kunnen komen voor de ontwikkeling van een detentiehuis.
2. Ontwikkelen van een virtuele tool die buurtbewoners toelaat een detentiehuis en de werking ervan digitaal te verkennen.
3. Conflict kleinschaligheid versus grootschaligheid: de detentieproblematiek kan abstract herleid worden tot een keuze tussen kleinschaligheid en grootschaligheid. Analoge problemen stellen zich ook voor bijvoorbeeld woonzorgcentra, waar grotere schaal vaak ook voor grotere problemen zorgde tijdens de coronacrisis. We bestuderen de economische randvoorwaarden voor zowel woonzorgcentra als voor detentiehuisen vanuit deze invalshoek.

Voor deze case zijn we in nauw contact met vzw De Huizen en met Stad Gent (Schepenen van Sociaal Beleid en Armoedebestrijding) die dit initiatief erg ondersteunt en ons onder andere een lijst van leegstaande panden zal bezorgen. Dit initiatief behoort ook tot het nieuwe regeerakkoord.

Bronnen:

From Prison to Detention House, a story of civilisation, Hans Claus, TEDxAntwerp
<https://youtu.be/sRcbVduOoE8>