www-ftm-nl.translate.goog

Zo leerde een Rotterdams fraudealgoritme kwetsbare groepen te verdenken

Tom Claessens

16–20 minutes

An algorithm that the municipality of Rotterdam used to predict welfare fraud for years counted young mothers and people who speak poor Dutch among the highest risk groups. They had the greatest chance of strict control by the municipality. This is evident from research by Lighthouse Reports, Argos, Vers Beton and Follow the Money, in which journalists got their hands on a complete fraud algorithm for the first time.

This piece in 1 minute

- **What's the news?**

  For years, Rotterdam has used an algorithm to predict fraud and errors in social assistance benefits. Vulnerable groups, such as people who do not speak Dutch well and young mothers with problems, came into focus more quickly and were therefore checked more often.

- **Why is this relevant?**

  The municipality of Rotterdam is known for its strict checks on social security fraud. It was unknown how the municipal algorithm determined exactly who qualifies for such an in-depth check. And what the risk of discrimination was.

- **How has this been researched?**

  Journalists from Lighthouse Reports, *Argos* , Vers Beton and Follow the Money managed to get their hands on the Rotterdam algorithm and the data with which this mathematical model calculates the risk of welfare fraud. This shows that age, gender and language skills weigh heavily. But what is that based on?

[read more]

The letter says that you must come by, with all your papers and bank statements. Once at the office, a fire of questions follows. Why, Adriana, do you take cash for groceries? (You couldn't pin, on the market.)

Where does that credit of 17 euros come from? (You had sold a computer game via Marktplaats.) You get questions about undeclared work. (You work as a volunteer at a school, you have not yet succeeded in getting paid work.) They want to know everything about you.

A small irregularity can have major consequences for your social assistance benefit. Rotterdam is known for being very strict. The municipality even announces an investigation into possible possessions abroad, while it must be clear that you have no money for any luxury whatsoever.

She is not told why Adriana in particular was selected for an in-depth check. It's not the first time she's had to show up either. Is it coincidence? Did the neighbors click? Does Adriana herself do something that provokes suspicion?

What the official on the other side of the table certainly does not tell her is that a complex algorithm has calculated that the chance of fraud, errors and mistakes would be higher for Adriana than for others (her 'risk score' is 0.683 on a scale from 0 to 1). The calculation was made on the basis of more than three hundred different characteristics that the municipality has recorded about Adriana and her life.

From her age (30 years old) and relationships (married for three years, one son) to her psychological condition (Adriana has been through a lot, but it is going pretty well), where she lives (since a year in North Rotterdam), her language skills (she now speaks enough Dutch), how a social service employee estimates her abilities to find a job and whether she has a representative appearance (no negative comments about this, according to this official).

Since according to the system, Adriana has important characteristics in common with people who have been caught in the past for tampering with their benefits, intentionally or unintentionally, Adriana is also suspicious.

Minority Report in Rotterdam

It seems like a polder version of *Minority Report* , the science fiction film made by Steven Spielberg in which the police department helps *precrime* predict and prevent future murders. In Rotterdam, a complex algorithm calculated for years which of the tens of thousands of Rotterdammers in thewelfare could tamper with his or her benefits. Every year, hundreds of residents with the highest risk scores based on this mathematical model could count on a check.

Lighthouse Reports, Vers Beton, Follow the Money and *Argos* got their hands on the Rotterdam algorithm – with the unimaginative name *analytics benefit fraud* – after a series of Woo requests. For the first time, outsiders are mapping out how an advanced government data model predicts fraud: what data about people it contains, how the computer code calculates with this and who gets the highest risk scores.

Spoiler : Those are vulnerable groups among the already poorest people in the city. Think of people entitled to social assistance who do not speak Dutch well, young people, single women with children

who come from a long relationship and people with financial problems.

Concerns about the Rotterdam forecasts have existed for some time. In 2021, the Court of Audit asked Rotterdamattention to the ethical risks of the welfare fraud algorithm, which was co-developed by consultancy firm Accenture. Characteristics such as language, for example, can lead to discrimination just like nationality. According to the calculation, this was not taken into account.

There was also criticism in the city council. 'I am fundamentally against the use of personal characteristics,' said PvdA councilor Duygu Yildirim, for example. According to her, the municipality cannot predict on an individual level who is inclined to commit fraud and who is not on the basis of personal characteristics. "That gentleman with an addiction has nothing to do with the fact that, statistically speaking, people from that group may be more likely to fail their information duties."

From figures previously requestedArgos and Lighthouse Reports also showed that a striking number of women were examined on the basis of the algorithm: in the period from 2018 to 2020, no less than 2179 women compared to 933 men .Rotterdam , which is the 'social assistance capital' of the Netherlands with about thirty thousand benefit claimants, argued that there were logical explanations for this.

For example, with other selection methods many men would previously have been examined, who were then removed from the results of the algorithm because they had already had their turn. Checks would also have shown 'no under- or over-representation of certain groups of Rotterdammers', according to Richard Moti, then the responsible alderman. In short: there was no bias.

> Rotterdam continues to believe in the promise of this technology: a higher chance of being caught, more efficient deployment of inspectors and less unjustly paid assistance

The system was nevertheless shut down at the end of 2021 as a precaution. The municipality wants to develop a new version, without elements that could discriminate. Because Rotterdam continues to believe in the promise of this technology: a higher chance of being caught, more efficient deployment of inspectors and less unjustified assistance.

It is not possible to say exactly how much the 'risk assessment model' will yield. 'The total recovery amount based on a re-examination averages 2.5 million euros per year,' according to the municipality. A fraction of that amount was credited to the algorithm. The municipality uses several methods to select people for research.

Rotterdam points out that the purpose of an investigation is not necessarily to find fraud, but that it is also in the interest of those entitled to social assistance to rectify mistakes as quickly as

possible and to prevent someone from getting into trouble because too much assistance is required. is received.

Which people has the system learned to suspect?

Is the municipality's claim that there was no bias correct? Vulnerable groups indeed have nothing to fear from this system, as they had from the infamous benefits algorithm, for examplefrom the tax authorities?

What exactly is the influence of personal characteristics that people cannot change, such as origin, age and gender, on their risk score? How does sensitive data weigh up like language and money and addiction problems? In other words: which people has the system learned to suspect, and why?

These questions could only be answered by testing the algorithm extensively, with data on real citizens. The experiments make it clear that the algorithm estimates the risk of fraud much higher for one Rotterdammer than for another Rotterdammer.

In particular with Rotterdammers who do not speak the Dutch language well, are female, very young, have children, come from a longer relationship, share the costs of a household with others and have financial or addiction problems. They are much more likely to be among the very highest risk scores than others, especially if they also fall into more than one of these categories. It means that they are at the top of the list for scrutiny.

The experiments also examined the influence of a single characteristic on the level of risk scores. It has been determined in this waythat women actually score higher *because* they are women. And that people who do not meet the language requirement are considered to be at greater risk simply because they have this characteristic.

Chances of control are nil for George

The differences at an individual level are large. An 'average Rotterdammer' on welfare is, for example, a 30-year-old man, George. He does not exist, but among the Rotterdam welfare recipients there are many real people with the same characteristics as George. He lives in North Rotterdam, is sporty, single and has no children.

With these characteristics, the algorithm calculates George's risk score at 0.50 and places him at 20,973 out of 30,000 on the risk list. This means that the chance of inspection for him is nil.

If George had been a woman, she would immediately have been thousands of places higher. If she does have a child and a partner, with whom she is on welfare and shares the household costs, then George has changed into Adriana, from the beginning of this story, with a score of 0.683. She is then among the highest risk scores and a check seems almost guaranteed.

That is calculated without language, the most sensitive feature in the Rotterdam welfare fraud algorithm. A total of twenty different variables in the Rotterdam algorithm are related to this: from someone's spoken language to writing skills and the language requirement for social assistance.

If all language variables are set in such a way that they indicate poor Dutch language skills, this ensures that these people are more than twice as likely to appear in the highest risk scores than people who have Dutch as their mother tongue.

Large differences also apply to other groups. Single mothers are 40 percent more likely to be among the highest risk scores than single women without children. And people who have been struggling with financial problems for several years get a high score 21 percent more often than people entitled to social assistance without these problems.

The results of the experiments have been presented to the municipality of Rotterdam, which describes the findings in a very extensive response as 'interesting, instructive and partly recognisable'. “Over time, we found that the risk assessment model could never be 100 percent free from bias or the appearance of it. This situation is undesirable, especially when it comes to variables that carry a risk of bias based on discriminatory grounds such as age, nationality or gender. Your findings also demonstrate these risks.'

Black box broken open

The joint research provides an insight into what used to be a *black box* . It is usually unclear how risk models work, because governments say they fear that citizens will adjust their behavior to avoid fraud checks. Rotterdam has opted for extensive transparency. In 2021, following Woo requests from Lighthouse and *Argos* , the municipality released a list of more than three hundred risk indicators and even the computer code.

After new Woo requests, extensive technical information was added. 'Rotterdam considers it very important that not only we ourselves, but also other governments and organizations are aware of the risks of algorithms,' says the municipality . For privacy reasons, the municipality wanted the data on thousands of citizens who had been investigated for fraud and with which the algorithm was 'trained' to make its predictions.not provide.

In 'histograms' sent along about the 315 variables in the algorithm, these data nevertheless appeared to be present in the background: more than 12,700 *records* , stripped of directly identifiable data such as names, citizen service numbers and contact details, but originating from real Rotterdammers who were checked at some point. . For journalistic reasons, these data were used for the investigation. A very select number of people had access. After publication of the findings, the data will be destroyed. According to

the municipality, the data should not have been released. 'We have reported this to our *privacy officer* .'

Read all about the research into the algorithm and the methods used here.

And the extensive response of the municipality of Rotterdam.

read more Collapse

The research into the operation and results of the algorithm are part of the story. At least as important are the data on which the algorithm has 'learned' to make its predictions. In the case of Rotterdam, these are the data of 12,700 welfare recipients that have been checked previously.

In order to be able to predict with a model who has a high risk of fraud and who has a small risk, these data must correspond to reality. There are countless examples where this goes wrong: from automatic facial recognition that does not work well with people with a dark skin color, to an algorithm of a job site that disadvantages women compared to men. In such cases, the data is often not a reflection of reality, for example because the algorithm is mainly trained on people with a light skin color, which means that the system will often be wrong.

> 'The Rotterdam algorithm does not perform well, it actually guesses randomly'

There are also question marks in Rotterdam. For example, it is not clear how fraudsters were selected from previous investigations. That may have been a random selection, but also another method that may not have been free of bias. For example, data may originate from previous thematic checks, in which groups were researched in advance, such as people with a specific living situation or household composition. It is striking in the Rotterdam data that very few young people appear in it, while age has the greatest influence on the level of the risk score.

According to Rotterdam, the algorithm was ultimately more effective than random checks. The algorithm 'scored' about 39 cases of fraud or other forms of irregularity on a hundred checks, according to figures provided by the municipality. In random checks it was 25 times out of a hundred.

But the renowned American computer scientist Margaret Mitchell points to evaluations of the Rotterdam algorithm: these show that the model does not perform well and 'actually guesses randomly'. Mitchell specializes in artificial intelligence, ethics and bias. On request, she looked at the investigation. Assessing whether people can pose a risk is always a human job, says Mitchell. According to her, computer models will never make a good prediction of the actual risk that people pose, because 'all lives are different'.

A mathematical model never takes into account all the factors that play a role in each individual case, says Mitchell. She thinks that the developers of the Rotterdam algorithm did not have enough or

not the right information to make a good model. 'All things considered, you have a recipe for a model that doesn't give an accurate picture of reality, based on what it has learned. That means it's not usable in the real world."