ACCOUNTABILITY, BIAS AND TRANSPARENCY

A Critical Look at the AMS Algorithm

Florian Cech
Centre for Informatics and Society
Centre for Informatics and Society @ TU Wien
Critical Algorithm Studies

- Research / Knowledge Centre
- General focus: *Digital Transformation*
  - ... interfacing technology and society
- Special Focus: *Critical Algorithm Studies*
  - Critical analysis of algorithmic systems
  - Issues include accountability, transparency, bias, discrimination & ethics
  - Practical focus on transparency and accountability improvements for algorithmic systems
AMS Algorithm - Timeline

- 2016: First iteration of project rejected by then Minister Alois Stöger (SPÖ)
- 2017: Client segmentation as analytic method
- 2018: Current government greenlights project
- 2019: First public statements & newspaper articles, Vienna among first locations
- 2020: Rollout to Production?
AMS Algorithm - Information Sources

- "Specification" document published by the Synthesis GmbH, a third party data analysis and research company
- Interviews and Statements by Johannes Kopf
- Internal AMS presentations
  - ... partly accidentally published?
- Reports by Austrian Court of Audit
- OECD Technical Workshop Presentation on Profiling Tools in the labor market
TECHNICAL BACKGROUND

„Machine Learning“ - AMS Algorithm

- Statistical profiling tool for client segmentation
- Logistic regression predicts job-seeker’s chances in the labor market based on prior observations
- Training dataset consists of AMS client’s PII
  - ... at least partially self-reported data!
- Prediction classifies into three categories (low, medium and high chances) with two target functions
  - Short-Term: 90 days of employment w/in 7 months
  - Long-Term: 180 days of employment w/in 24 months

Quelle:
XKCD by Randall Munroe
https://xkcd.com/1838/
Input Data - Client Information via AMS Data Warehouse

- Gender [m | f]
- Age [0-29 | 30-49 | 50+]
- Nationality [AT | EU | Non EWR]
- Education [mandatory school | vocational school | AHS, FH, University]
- Health impairments [y | n]
- Care obligations - only for women! [y | n]
- Type of occupation [production | service industry]
- Type of regional labor market [5 types - unclear!]

- Prior occupational career
  - Days of gainful employment within 4 years prior [>= 75% | < 75%]
  - Number of opened cases with AMS within 4 x 1 year prior intervals
    [0 cases | 1 case | 1+ cases in 2 intervals | 1+ cases in 3+ intervals ]
  - Duration of opened cases > 180 days [ 0 cases | 1+ cases ]
  - Support measures claimed [0 | 1+ supportive | 1+ educational | 1+ subsidized]
Impact coefficients of PII attributes - example short term model

- Impact of PII attributes based on observations in previous years

- Positive impacts: frequency of business cases, occupation in manufacturing/industry, education in vocational schools

- Multiple models for different populations

- Separation by PII and data quality / completeness

- Negative impacts: gender ‘female’, clients with health issues, background in immigration, obligations for care

Quelle: http://www.forschungsnetzwerk.at/downloadpub/arbeitsmarketchancen_methode_dokumentation.pdf
### Example Calculation

**Sample Persona 1**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.14</td>
</tr>
<tr>
<td>32 years old</td>
<td>-0.13</td>
</tr>
<tr>
<td>EU citizen</td>
<td>0.16</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>0.01</td>
</tr>
<tr>
<td>1 prior case &lt; 180 days</td>
<td>0.65</td>
</tr>
<tr>
<td>Responsibility f. Care</td>
<td>-0.15</td>
</tr>
<tr>
<td><strong>Total + 0.10</strong></td>
<td><strong>0.5</strong></td>
</tr>
</tbody>
</table>

**Logistic Transformation**

\[
f(x) = \frac{L}{1 + e^{-k \cdot (x-x_0)}}
\]

\[
f(0.5) = \frac{1}{1 + e^{-1 \cdot (0.5-0)}} = 60\%
\]

- **Group A:** 66% - 100%
- **Group B:** 25% - 66%
- **Group C:** 0% - 25%
In absolute numbers

Segmentation by Group

- Voll valide schätzbar, zu Beginn
- Männer, voll valide, zu Beginn
- Frauen, voll valide, zu Beginn
- - " - , nach 12 Monaten
- Kärnten, voll valide, zu Beginn
- Salzburg, voll valide, zu Beginn
- Mit Migrationshintergrund, nach 6 Monaten
- - " - , nach 24 Monaten
- Wien, »fragmentiert«, zu Beginn

Measure Names:
- Category A - High Chances
- Category B - Medium Chances
- Category C - Low Chances
TECHNICAL BACKGROUND

Error rate

Category „A“
High Chance

Category „B“
Medium Chance

Category „C“
Low Chance

= 10,000 clients
440,000 in total
TECHNICAL BACKGROUND

Error rate

= 10,000 clients
440,000 in total

Category „A“
High Chance

Category „B“
Medium Chance

Category „C“
Low Chance

Falsely classified:

2,640 „C“
28,160 „A“
?????? „B“
AMS Algorithm as prime example for ‘Mathwashing’

„Models are opinions embedded in mathematics.“

Cathy O’Neil,
„Weapons of Math Destruction“

- Seemingly „neutral“ mathematics / statistics hide...
  - ... human decisions & prejudice
  - ... discrimination & bias
- Pretense of objectivity allows delegation of responsibility towards automated decision making systems
AMS & Accountability

- Responsibility/Liability in case of errors unclear
- ... once occurred it’s mostly too late!
- No transparency of operative rules and guidelines
- Promise of ‘explanations’ for classification
  - Most likely by showing largest impact factors
- Regulatory oversight of algorithmic systems and their use generally lacking

Algorithmic System trains on and replicates bias

Discriminating Labor Market

Error rate, algorithmic illiteracy, stigmatisation & implicit biases affect discriminated groups disproportionately

Decisions are based on biased results

Feedback Loops

Other Critical Issues

- Calculations and training data are opaque
- Results not traceable / explainable
  - ... almost certainly not for AMS clients and workers!
- As of December 2018: staff reported no training or guidance on the system
- Unclear procedural consequences for clients

Claim:
- „The algorithm doesn’t decide - AMS workers can ignore the suggestions!“

Reality:
- Lack of algorithmic literacy limits understanding for data biases & error rates
- Suggestions of algorithmic systems tend to be taken with little question - „Automation Bias“
- Pressure to explain when disagreeing with the algorithm
### Statistical Profiling across OECD Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Outcome (probability of)</th>
<th>Data derived from</th>
<th>Type of data sources (in addition to socioeconomic info)</th>
<th>Statistical model</th>
<th>Accuracy</th>
<th>Compulsory/voluntary use by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Labour market history</td>
<td>Job readiness</td>
<td>Motivation</td>
<td>Opportunities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hard skills</td>
<td>Soft skills</td>
<td>Jobseekers’ behaviour</td>
<td>Regional labour market info</td>
</tr>
<tr>
<td>Australia</td>
<td>Long-term unemployed (12 months)</td>
<td>Personal interview; online trial ongoing</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Austria</td>
<td>LM integration probability of unemployment</td>
<td>Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Belgium (Flanders)</td>
<td>Long-term (&gt;6 months) unemployed</td>
<td>Administrative data; “click” data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Denmark</td>
<td>Long-term (&gt;26 weeks) unemployed</td>
<td>Online questionnaire; Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ireland</td>
<td>Probability of exit to employment within 12 months</td>
<td>Questionnaire as part of benefit claim process, administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Italy</td>
<td>Long-term unemployed (12 months)</td>
<td>Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Latvia</td>
<td>Long term unemployed (12 months)</td>
<td>Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Netherlands</td>
<td>Long-term unemployed (12 months)</td>
<td>Online questionnaire</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Lifetime income support costs (LET), change in lifetime income support and staff costs from receiving a case management service (SEM)</td>
<td>SEM/LET are based on administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sweden</td>
<td>LTU (6 months)</td>
<td>Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>US</td>
<td>Exhausting the 26-week entitlement to UI benefits</td>
<td>Online questionnaire; Administrative data</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Moral of the story

Hard Problems + Algorithms != Easy Problems

OUR FIELD HAS BEEN STRUGGLING WITH THIS PROBLEM FOR YEARS.

STRUGGLE NO MORE! I'M HERE TO SOLVE IT WITH ALGORITHMS!

SIX MONTHS LATER:
WOW, THIS PROBLEM IS REALLY HARD.
YOU DON'T SAY.

Quelle: XKCD by Randall Munroe
https://xkcd.com/1831/