
Sollten Algorithmen im Gericht benutzt werden, um die Rückfälligkeitswahrscheinlichkeit von Angeklagten zu berechnen?

Gruppe: 10

Journalist: Jonas Opitz

Chefredakteur: Frank Eric Mbouga

March 29, 2019

1 Einleitung

Welche Folgen kann es für wen geben, wenn man einen Algorithmus zum Bestimmen der Rückfälligkeitswahrscheinlichkeit von Angeklagten, wie OCEAN, im Gericht einführt?

Nachdem es im vorherigen Blogeintrag darum ging, wie Algorithmen wie Northpointes OCEAN funktionieren, soll es nun um die Konsequenzen der Einführung dieser gehen.

Um diese Frage zu betrachten wird zunächst das soziale System, d.h. die betroffenen Akteure, betrachtet, in das diese Algorithmen eingefügt werden, und ob es sich hier nach Kienle und Kunau um ein sozio-technisches System handelt. Danach wird noch ein Mal auf die Funktionsweise von OCEAN eingegangen und diskutiert, woher die in [1] beschriebenen rassistischen Tendenzen her kommen könnten.

Da Algorithmen zum Bestimmen der Rückfälligkeitswahrscheinlichkeit bereits in den USA angewandt werden, und sich die meisten Studien auf die USA beziehen, wird sich dieser Blogeintrag ebenfalls auf die USA beziehen.

diese unterstützt werden soll.

- Die Angeklagten, für die diese Algorithmen entscheiden, ob sie vor der Gerichtsverhandlung in das Gefängnis gehen bzw. Bail zahlen müssen.

Jedoch lässt sich die letzte Gruppe von Menschen, die Angeklagten, in zwei Subgruppen unterteilen:

- Die, die historisch im Gericht bevorteilt sind (d.h. weiße Amerikaner).
- Die, die historisch im Gericht benachteiligt sind (besonders Afro-Amerikaner).

Dass diese Unterteilung sinnvoll ist folgt aus Studien wie [4], in denen untersucht, und bestätigt [4, Kapitel 7], wurde, ob es eine Korrelation zwischen schwereren Gerichtsurteilen und der Ethnizität der/des Angeklagten gibt.

Im Folgenden wird primär betrachtet, wie das Einführen von Algorithmen wie OCEAN die Gruppe von historisch benachteiligten Angeklagten beeinflusst.

2 Das interagierende soziale System

Auf erstem Blick scheint es vier Akteure zu geben, die von Algorithmen wie OCEAN betroffen sind:

- Die Privatunternehmen, die solche Algorithmen entwickeln, vermieten und somit von ihnen profitieren.
- Der Gesetzgeber, der gegebenenfalls den Einsatz dieser Algorithmen regulieren muss.
- Die Richterschaft, für die diese Algorithmen überhaupt gemacht werden, und deren Arbeit durch

3 Handelt es sich um ein sozio-informatisches System?

Um ein sozio-informatisches System nach Kienle/Kanau handelt es sich hier nicht, da die dritte Bedingung, "das technische System findet Eingang in die Selbstbeschreibung des sozialen Systems" [5], nicht erfüllt ist - das Verwenden von Algorithmen wie OCEAN dient lediglich der Unterstützung von Richtern und führt zu keiner Neuheit, die in die Selbstbeschreibung des sozialen Systems einhergehen würde.

4 Gibt es rassistische Tendenzen in der Bewertung von OCEAN?

Im folgenden werden falsche Positive/Negative betrachtet, um auf eine generelle Tendenz zu schließen. In diesem Kontext bedeutet ein falsches Positiv, dass ein Straftäter in den nächsten 2 Jahren nach dessen Freilassung keine weiteren Straftaten begangen hat, obwohl OCEAN ein hohes Rückfallrisiko berechnet hat, und ein falsches Negativ das umgekehrte [3].

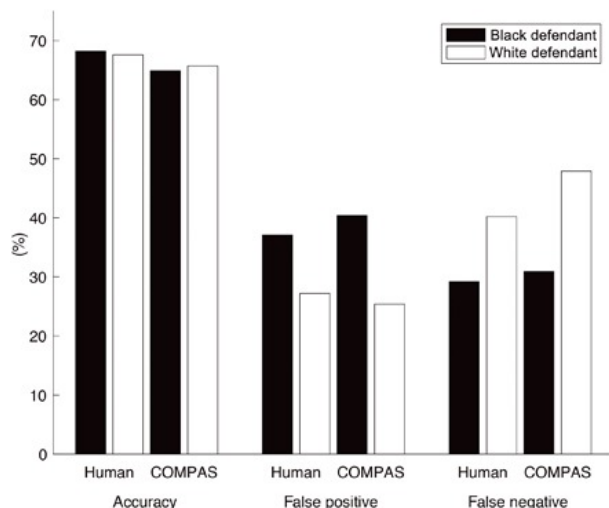


Figure 1: Die Ergebnisse der Studie [3], in der der COMPAS-Algorithmus mit zufällig gewählten Probanden bezüglich Aussagekraft der vorhergesehenen Rückfälligkeitswahrscheinlichkeit von Kriminellen verglichen wurde.

Die x-Achse beschreibt die Wahrscheinlichkeit für die auf der y-Achse beschriebenen Fälle.

Das erste paar Balken bezieht sich auf die Prognosen der Probanden, das zweite auf die Prognosen von COMPAS.

Die Farbe der Balken (schwarz oder weiß) entspricht der entsprechenden Ethnizität der Kriminellen.

Quelle: [3]

Wenn man die Ergebnisse aus Studie [3] betrachtet (Bild 1), springt sofort die recht große Diskrepanz zwischen weißen und Afro-Amerikanern in falschen Positiven/Negativen ins Auge: bei OCEAN ist die Rate an falschen Positiven für Afro-Amerikaner ungefähr 60% höher als die von Weißen (ca. 40% vs. ca. 25%), und die Rate an falschen Negativen ungefähr 38% niedriger (ca. 30% vs. ca. 48%) [3, Tabelle 1].

Diese große Differenz ist jedoch nicht nur im Algorithmus zu finden; auch die Probanden, denen die Ethnizität der Person nicht gegeben wurde, bewerteten in der Studie Afro-Amerikaner schlechter als Weiße, wenn auch nicht stark. Hier ist die Rate an falschen Positiven für Afro-Amerikaner ungefähr 37% höher (ca. 37% vs. ca. 27%) und die an falschen Negativen ungefähr 28% niedriger (ca. 29% vs. ca. 40%) [3, Tabelle

1].

Dass diese Raten dennoch schlechter in OCEAN sind impliziert, dass der Algorithmus im Durchschnitt Afro-Amerikaner strenger bewertet als Weiße. Dies ist problematisch, da Afro-Amerikaner ohnehin schon im Justizsystem in Amerika benachteiligt sind [4], und dieses Problem durch die Einführung von Algorithmen somit noch schlimmer werden könnte.

5 Mögliche Gründe für diese Resultate

Nun muss man sich fragen, wo diese signifikante Diskrepanz an falschen Positiven/Negativen her kommen könnte - der Fragenkatalog, der von OCEAN ausgewertet wird [6] enthält nämlich keine Frage dazu, welcher Ethnizität der oder die Angeklagte angehört.

5.1 Die Wahrscheinlichkeit, bei einigen Fragen schlecht ab zu schneiden ist für Afro-Amerikaner höher

Zuerst ein paar generelle Fakten über Florida, der US-Staat, in dem Broward County liegt (der Bezirk, auf den sich die Studien [1] und [3] beziehen):

- Der Anteil von afro-amerikanischen Insassen in Floridas Gefängnissen höher als der von Weißen (ca. 48% vs. ca. 40%) [7, Seite 16]
- Die überwiegende Mehrheit der Population in Florida ist weiß (ca. 77%) [8]. Nur ca. 17% der Menschen in Florida ist afro-amerikanischer Ethnizität.

Daraus folgt, dass die Wahrscheinlichkeit, dass ein Elternteil oder Freund/Partner (unter der Annahme, dass Menschen einer bestimmten Ethnizität tendenziell eher Freund-/Partnerschaften mit Personen gleicher Ethnizität bilden) Zeit im Gefängnis verbracht hat, deutlich höher für Afro-Amerikaner als für Weiße ist. Damit zeigt sich auch, warum OCEAN Afro-Amerikaner im Durchschnitt schlechter bewerten könnte: die Fragen 33 bis 36 und 38 bis 40 beziehen sich direkt darauf, ob diese jemals verhaftet wurden oder im Gefängnis waren.

Viele weitere Fragen mit ähnlicher Begebenheit lassen sich in dem Fragebogen finden (z.B. Fragen über Gangmitgliedschaft, Probleme mit Drogen-/Alkoholbenutzung, Stabilität der Lebenssituation, Sicherheit der sozialen Umgebung oder Bildung), die Analyse dieser würde den Umfang dieses Blogbeitrags jedoch sprengen.

5.2 Die Daten, auf den die Algorithmen basieren könnten, sind rassistisch

Unter der Annahme, dass OCEAN und ähnliche Algorithmen ihre Auswertungsweise durch Machine Learn-

ing entwickeln - eine Annahme, die momentan weder bewiesen, noch widerlegt werden kann, da wir, wie im ersten Blogeintrag erwähnt, nicht wissen, wie diese Algorithmen arbeiten - gibt es noch eine weitere Erklärung der obigen Ergebnisse: OCEAN hat "gelernt", rassistisch zu entscheiden, bzw. Attribute, die mit einer bestimmten Ethnizität korrelieren, als risikofördernd ein zu stufen.

Dies kommt daher, dass Machine Learning darauf basiert, aus gegebenen Daten einen Algorithmus zu entwickeln. D.h. benutzt man rassistisch beeinflusste Daten, wie es Rechtsprechungen sind [4], ist der daraus resultierende Algorithmus ebenfalls rassistisch.

6 Schluss

Zusammenfassend lässt sich sagen, dass die Einführung von Algorithmen zur Vorhersage der Rückfallswahrscheinlichkeit eines Angeklagten die in diesem Rechtssystem vorhandenen Probleme (hier Rassismus) zu amplifizieren, oder wenigstens nicht zu mindern, scheint, und diese Algorithmen dazu nicht einmal im Allgemeinen besser als Laien zu performieren scheinen.

Es wirkt nicht so, als wäre die resultierende Entlastung der Richter diese Konsequenz wert.

7 Quellen

- [1]: Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner (ProPublica): "*Machine Bias*", 2016
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (abgerufen am 19.03.2019)
- [2]: Northpointe Inc.: "*Practitioner's Guide to COMPAS Core*", 2015
<https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf> (abgerufen am 19.03.2019)
- [3]: Julia Dressel, Hany Farid: "*The accuracy, fairness, and limits of predicting recidivism*", Publiziert 2018 in *Science Advances*, Vol. 4, No. 1
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5777393/> (abgerufen am 19.03.2019)
- [4]: David S. Abrams, Marianne Bertrand, Sendhil Mullainathan: "*Do Judges Vary in Their Treatment of Race?*", Publiziert 2012 in *The Journal of Legal Studies*, Vol. 41, No. 2
<https://www.povertyactionlab.org/sites/default/files/publications/210%20Do%20Judges%20Vary%20Sept%202010.pdf> (abgerufen am 20.03.2019)
- [5]: Andrea Kienle, Gabriele Kunau: "*Informatik und Gesellschaft - eine sozio-technische Perspektive*", 2014
- [6]: Fragenkatalog, der von OCEAN ausgewertet wird, Northpointe Inc, 2011
<https://assets.documentcloud.org/documents/2702103/Sample-Risk-Assessment-COMPAS-CORE.pdf> (abgerufen am 29.03.2019)
- [7]: Jährlicher Bericht vom Florida Department of Corrections für das Haushaltsjahr 2016 - 2017
http://www.dc.state.fl.us/pub/annual/1617/FDC_AR2016-17.pdf (abgerufen am 29.03.2019)
- [8]: United States Census Bureau: "*Quick Facts*",
<https://www.census.gov/quickfacts/fact/table/fl/PST045218> (abgerufen am 29.03.2019)