Deutsche Stiftung für Recht und Informatik

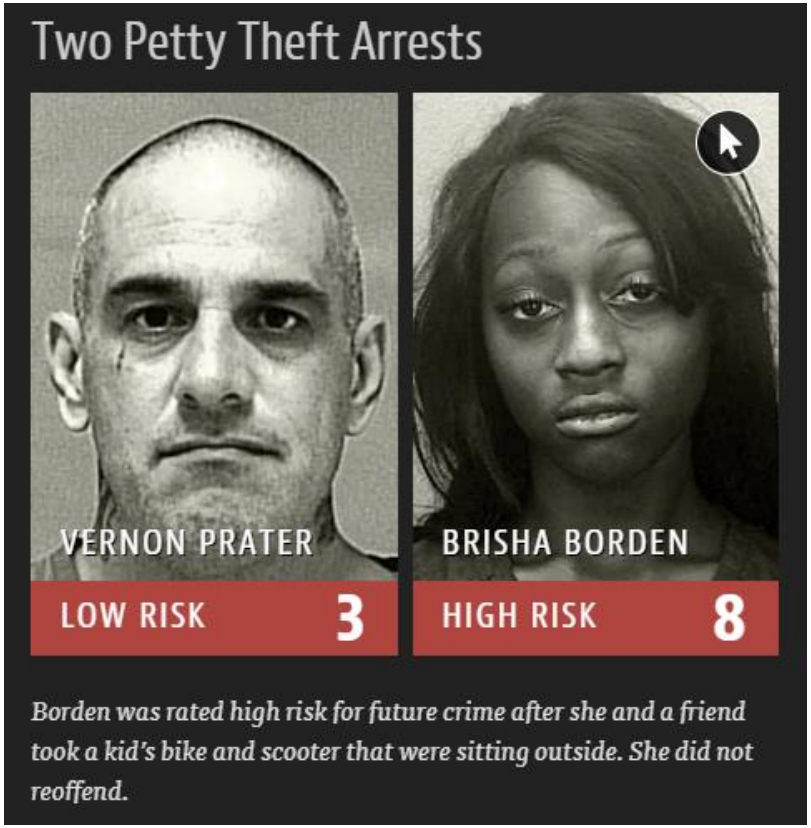
Fairnessmetriken bei algorithmischen Entscheidungen aus juristischer Perspektive

Caroline Gentgen-Barg

Institut für Rechtsinformatik, LUH

Herbstakademie 2022

COMPAS und die Diskussion um Fairnessmetriken



Evaluierung von COMPAS

- Northpointe (Hersteller)

  PPV (positive predictive value) =

  Anzahl der Personen, die, nachdem sie als rückfällig vorhergesagt wurden, tatsächlich rückfällig werden

  PPV = TP/{TP+FP}

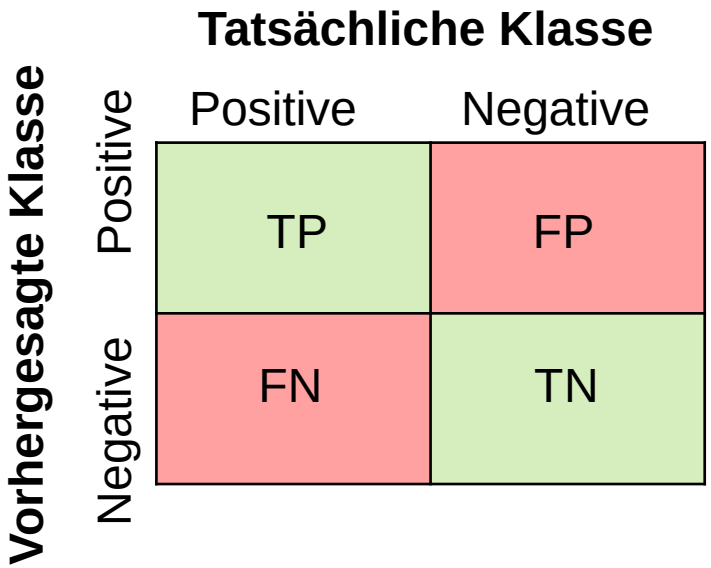
- ProPublica

  FPR (predictive equality) =

  Anzahl der Personen, die nicht wieder straffällig werden, bei denen fälschlicherweise vorhergesagt wurde, dass sie rückfällig werden

  FPR = FP/{FP+TN}

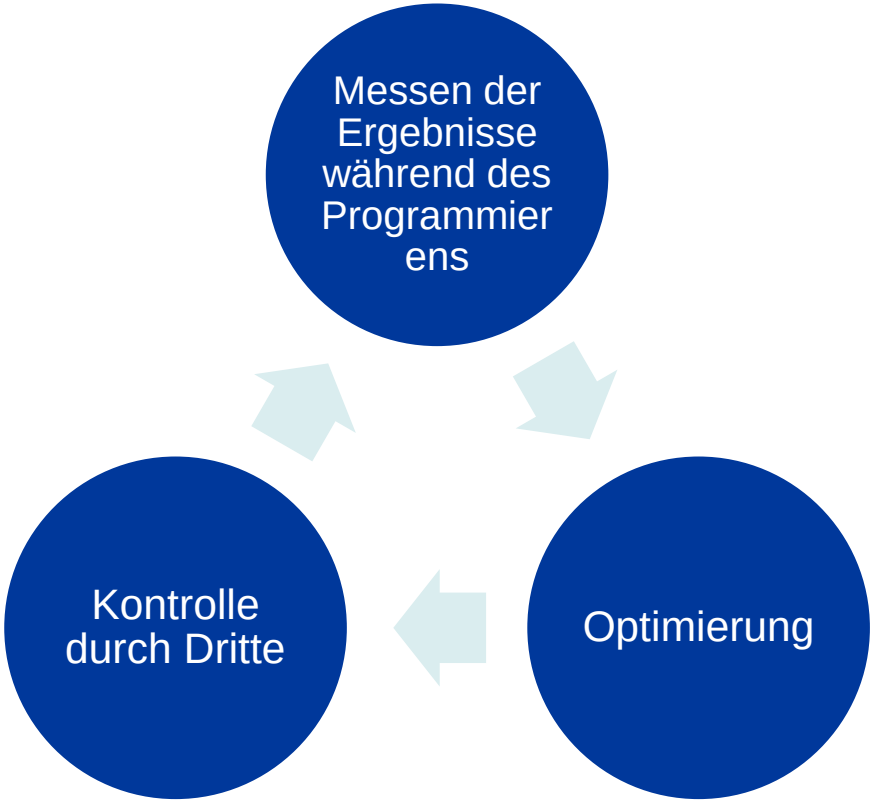
Statistisch Fairness messen



- ProPublica =
  FPR = FP/{FP+TN}
- Northpointe (Hersteller) =
  PPV = TP/{TP+FP}

Der Einsatz von Fairnessmetriken – Rechtliche Fragestellungen

- Verstößt der Anbieter des Algorithmus gegen das Antidiskriminierungsrecht und kann/soll er anhand einer Fairnessmetrik optimieren?
- Verstößt die Anwendung einer Fairnessmetrik gegen das Antidiskriminierungsrecht?



Rechtsrahmen: Anwendungsbereich des Antidiskriminierungsrechts

- Für Diskriminierungen durch die öffentliche Hand **Art. 3 GG**
- Insb. im Privatrecht ist der Schutz vor Diskriminierungen stark kontextabhängig
  - sachlicher und persönlicher Anwendungsbereich des **AGG**
  - Diskriminierung teils erlaubt

Der Diskriminierungsbegriff

- Art. 3 GG: unmittelbare und mittelbare Diskriminierung
  - Umstritten, inwieweit mittelbare Diskriminierung außerhalb des Geschlechts vom Schutzbereich umfasst ist
- § 3 Abs. 1 und Abs. 2 AGG: direkte und indirekte Benachteiligung
- & Spezialgesetze

Problemkreise der indirekten Benachteiligung

„Eine mittelbare Benachteiligung liegt vor, wenn dem Anschein nach neutrale Vorschriften, Kriterien oder Verfahren Personen wegen eines in § 1 genannten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können, es sei denn, die betreffenden Vorschriften, Kriterien oder Verfahren sind durch ein rechtmäßiges Ziel sachlich gerechtfertigt und die Mittel sind zur Erreichung dieses Ziels angemessen und erforderlich."

Problemkreise der indirekten Benachteiligung

(P) Vergleichs-gruppen

(P) Quantifizier-barkeit?

„Eine mittelbare Ben[illegible]egt vor, wenn dem Ansch[illegible]n nach neutrale Vorsc[illegible]n oder Verfahren Personen wegen eines in § 1 ge[illegible]nten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können, es sei denn, die betreffenden Vorschriften, Kriterien oder Verfahren sind durch ein rechtmäßiges Ziel sachlich gerechtfertigt und die Mittel sind zur Erreichung dieses Ziels angemessen und erforderlich."

Problemkreise der indirekten Benachteiligung

(P) Hierarchie der Merkmale & multi-fairness

„Eine mittelbare Benac[illegible]or, wenn dem Anschein nach neutrale Vorschrifte[illegible]ntenen oder Verfahren Personen wegen eines in § 1 genannten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können, es sei denn, die betreffenden Vorschriften, Kriterien oder Verfahren sind durch ein rechtmäßiges Ziel sachlich gerechtfertigt und die Mittel sind zur Erreichung dieses Ziels angemessen und erforderlich."

Problemkreise der indirekten Benachteiligung

„Eine mittelbare Benachteiligung liegt vor, wenn dem Anschein nach neutrale Vorschriften, Kriterien oder Verfahren Personen wegen eines in § 1 genannten Grundes gegenüber anderen Personen in besonderer Weise benachteiligen können, es sei denn, die betreffenden Vorschriften, Kriterien oder Verfahren sind durch ein rechtmäßiges Ziel sachlich gerechtfertigt und die Mittel sind zur Erreichung dieses Ziels angemessen und erforderlich."

(P) Rechtferti-gung

Fairnessmetriken als „Positive Maßnahmen“

- Die Optimierung anhand von Fairnessmetriken kann eine direkte Diskriminierung darstellen

- Die Zulässigkeit von positiven Maßnahmen ist umstritten, § 5 AGG enthält eine ausdrückliche Regelung

  „Ungeachtet der in den § § 8 bis 10 sowie in § 20 benannten Gründe ist eine unterschiedliche Behandlung auch zulässig, wenn durch geeignete und angemessene Maßnahmen bestehende Nachteile wegen eines in § 1 genannten Grundes verhindert oder ausgeglichen werden sollen.“

Die “richtige“ Fairnessmetrik

- Kontextabhängige Antidiskriminierungsverbote im Kontrast zu standardisierten statistischen Messungen

- Es sind verschiedene Methoden gegeneinander abzuwägen - dabei sind die verschiedenen Interessen, die im Rahmen der Rechtfertigungsprüfung Berücksichtigung finden, wertend in Ausgleich zu bringen:

  - Gruppen- sowie Individualinteressen
  - Kosten der Fehlurteile
  - Genauigkeit der Vorhersage
  - wirtschaftliche Erwägungen

KI-VO-Entwurf & Fairnessmetriken

- Diskriminierungsfreiheit und Schutz vor Diskriminierungen als Ziele des KI-VO-Entwurfs
- KI-VO-Entwurf enthält keine Vorschriften über die Wahl der Mittel zur Erreichung dieses Ziels

- spezielle Anforderungen an Hoch-Risiko-KI in den Art. 8 ff. KI-VO-Entwurf geregelt
  - Risikomanagement-Maßnahmen Art. 9 Abs. 2 lit. d) KI-VO-Entwurf
  - Qualitätsstandards in Art. 10 Abs. 2 lit. f) KI-VO-Entwurf (erfasst nicht das algorithmische Ergebnis)
  - Dokumentationspflichten, Art. 11 KI-VO-Entwurf; insb. hinsichtlich Parametern, die zur Messung potenziell diskriminierender Auswirkungen verwendet werden; u.a. Genauigkeitsgrad für bestimmte Personen und Personengruppen

Verantwortlichkeit der Anbieter

- Dokumentation der Parameter und möglicher Kompromisse zwingt zur sorgfältigen Abwägung
- Problem der Rechtsunsicherheit: Unterstützung der Anbieter durch „harmonisierte Standards, Orientierungshilfen und Instrumente zur Einhaltung der Vorschriften“
- Ausgestaltung der Kontrolle: Konformitätsbewertungsverfahren, in vielen Fällen als interne Kontrolle, ausreichend?

Fazit

- Statistische Fairness steht im Kontrast zur Einzelfallbezogenheit des Antidiskriminierungsrechts
- KI-VO-Entwurf sieht den Anbieter in der Verantwortung, unterstützt durch Standards & Orientierungshilfen

- Fairnessmetriken können nur einen Beitrag zu diskriminierungsfreier KI liefern; Entscheidungsprozess und Transparenzvoraussetzungen bleiben ebenfalls für die rechtliche Beurteilung entscheidend

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:
gentgen@iri.uni-hannover.de