Einfluss der Datenbasis

Bennet Bleßmann

15. November 2018

Inhaltsverzeichnis

Einführung

Was ist die Datenbasis?

Probleme

-

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Alte Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Keine Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Verzerrte Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Falsche Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Nicht repräsentative Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Zu wenige Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Probleme

Zu viele Daten

- Alte Daten
- Keine Daten
- Verzerrte Daten
- Falsche Daten
- Nicht Repräsentative Daten
- Zu wenig Daten
- Zu viele Daten

Lösung

Alte Daten

Lösung

Alte Daten

- Daten in einem Format Speichern welches nicht Veraltet, z.B. Geburtsjahr statt Alter

Lösung

Alte Daten

- Daten in einem Format Speichern welches nicht Veraltet, z.B. Geburtsjahr statt Alter
- Regelmäßig neue Daten sammeln und alte Daten löschen

Lösung

Keine Daten/Zu wenig Daten

Lösung

Keine Daten/Zu wenig Daten

- Mehr Daten Sammeln

Lösung

Keine Daten/Zu wenig Daten

- Mehr Daten Sammeln
- Mehr Arten von Daten nutzen/sammeln

Lösung

Verzerrte Daten

Lösung

Verzerrte Daten

- Bei Credit-Scoring nicht nur Daten über negative Ereignisse sammeln sondern auch über positive Ereignisse

Lösung

Falsche Daten

Lösung

Falsche Daten

- Einsicht\Transparenz und Korrektur zulassen

Lösung

Falsche Daten

- Einsicht\Transparenz und Korrektur zulassen
- Mit anderen Quellen vergleichen

Lösung

Falsche Daten

- Einsicht\Transparenz und Korrektur zulassen
- Mit anderen Quellen vergleichen
- Fehlerquellen minimieren

Lösung

Nicht repräsentative Daten

Lösung

Nicht repräsentative Daten

- Statt allgemeine Daten zu nutzen, speziell für einen Zweck Daten Sammeln

Lösung

Nicht repräsentative Daten

- Statt allgemeine Daten zu nutzen, speziell für einen Zweck Daten Sammeln
- Maßnahmen gegen fehlende, zu wenige und verzerrte Daten

Lösung

Zu viele Daten

Ethik

Diskussion

- Was wäre wenn?
- Alltags Beispiele
- Szenario

Diskussion

Was wäre wenn?

Angenommen die Zulassung und die Auswahl des Studienganges würde von einem Algorithmus abhängen würde.
Wer von den Anwesenden könnte sich denken heute nicht hier zu sein (und warum)?

Diskussion

Alltags Beispiele

Diskussion

Alltags Beispiele

Warum ist dies ein Beispiel für den Einfluss der Datenbasis?

Diskussion

Alltags Beispiele

Kennt jemand hier ein vergleichbares Beispiel?

Diskussion

Szenario

Einführung

Diskussion

Szenario

- Besuchen von vielen Seiten zum Thema **ISIS**
- Besuchen von vielen Seiten zu **Explosiven Gemischen**
- Kauf einer Angelweste
- Kauf von Rohren
- Kauf verschiedener Haushaltsmittel
- Last-Minute Kauf von Karten für einen großen Kongress

Diskussion

Szenario

Was soll der Algorithmus entscheiden?

Diskussion

Szenario

- Nichts Unternehmen, keine Gefahr ausgehend!
- Kavallerie zum aufhalten schicken, Bombenanschlag wahrscheinlich!
- Eskalieren und einen Menschen entscheiden lassen!

Diskussion

Szenario

Neue Fakten:

- Studiert Chemie mit Nebenfach Gesellschaftswissenschaftlich
- Hat Angeln als Hobby
- Hat defektes Rohr selbst ersetzt statt den Handwerker zu rufen.

Diskussion

Szenario

Reflexion:

- War die ursprüngliche Entscheidung korrekt?
- Würde eine neuen Entscheidung anders ausfallen?
- Hatten wir zu wenige oder zu viele Daten?

References

(2017). Special Issue Informatik Spektrum , 40(4) . Retrieved from https://link.springer.com/journal/287/40/4/page/1

ACM US Public Policy Council. (12 January 2017). Statement on Algorithmic Transparency and Accountability. Retrieved from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Baeza-Yates, R. (June 2018). Bias on the Web. *Communications of the ACM , 61(6)*, 54-61. doi:10.1145/3209581 Retrieved from https://cacm.acm.org/magazines/2018/6/228035-bias-on-the-web/fulltext

Executive Office of the President. (May 2016). Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Maynes, R., & Everdell, I. (16 May 2018). The evolution of Google search results pages & their effect on user behaviour. *Mediative* . Retrieved from http://www.mediative.com/whitepaper-the-evolution-of-googles-search-results-pages-effects-on-user-behaviour/

Mercer, A., Deane, C., & McGeeney, K. (9 November 2016). Why 2016 election polls missed their mark. *Pew Research Center* . Retrieved from http://www.pewresearch.org/fact-tank/2016/11/09/why-2016-election-polls-missed-their-mark/

Oltenau, A., Castillo, C., Diaz, F., & Kiciman, E. (20 December 2016). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Retrieved from https://dx.doi.org/10.2139/ssrn.2886526

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., ... Nosek, B. A. (25 August 2018). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Pew Research Center* . doi:10.31234/osf.io/qkwst Retrieved from https://psyarxiv.com/qkwst/

Vincent, J. (24 March 2018). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge* . Retrieved from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Vielen Dank

Vielen Dank fürs Zuhören und Teilnehmen