

Arbeitsblatt: KI-Blögeintrag zum Bias im Maschinellen

Kurzbeschreibung

Nach der Bearbeitung des Teachable Machine-Projekts und der Diskussion des TED-Talks von Joy Buolamwini lesen und diskutieren Sie noch einen Blog-Eintrag zu diesem Thema.

Hinweis: es gibt zu diesem Arbeitsblatt Hinweise für Lehrkräfte.

Vorbereitung

Sie finden den Blögeintrag unter diesem Link (er ist auch in Ihrem Moodle-Kurs verlinkt):

<https://www.lernen-wie-maschinen.ai/ki-pedia/was-ist-algorithmische-voreingenommenheit-algorithmic-bias/>

Aufgaben

Frage 1

Im ersten Teil des Eintrags werden außer dem Beispiel von Joy Buolamwini noch zwei weitere Beispiele von schlagzeilenträchtiger Diskriminierung aufgrund von Bias genannt – welche sind das, und worin lag jeweils die Diskriminierung?



Frage 2

Anschließend nennt der Artikel fünf technische Mechanismen, die zu Bias führen können. Diskutieren Sie Beispiele für alle fünf:

1. Subjektive Definition von Zielvariablen

Als Beispiel wird angeführt, dass „beste/r Mitarbeiter/in“, also konkret die Variable „Qualität des Mitarbeiters/der Mitarbeiterin“, subjektiv ist. Wie ist das gemeint – wie könnte man Qualität bei Mitarbeitenden messen, und wer würde jeweils benachteiligt?

2. Falscher Umgang mit Trainingsdaten

Die ersten beiden Punkte sollten hier klar sein. Überlegen Sie sich, bei welcher Anwendung historische Verzerrung zum Tragen kommen könnte.

3. Ungenaue Feature Selection

Übrigens: Was hier als Attribut bezeichnet wird, haben wir bei uns im Kurs meist Dimension oder Feature genannt.

Wir verdeutlichen das Problem mit einem ausführlicheren Beispiel:

Stellen Sie sich vor, Sie wollen Daten zusammenstellen, in denen die Umsätze von in einem Möbelladen arbeitenden Personen festgehalten werden. Die Filiale erfasst dafür, welche Kunden von wem beraten wurden und berechnet so die Umsätze nach Mitarbeitenden. Folgende Features haben Sie zur Auswahl:

- Anzahl der von der Person beratenen Kund/innen
- Wochenarbeitszeit der Person
- Abteilung, in der die Person arbeitet (Kellermöbel, Haushaltswaren, Sonderanfertigungen, Wohnzimmermöbel, Schlafzimmermöbel, Kinderzimmer)
- Anzahl der Waren, die die von der Person beratenden Kund/innen gekauft haben
- Tageszeit, zu der die Verkäufe stattgefunden haben

Welche dieser Features könnten zu Diskriminierung führen, wenn Sie von einem auf diesen Daten trainierten Modell weggelassen werden?

Frage 3

Worin sieht der Blog-Eintrag den Hauptgrund, dass voreingenommene Datensätze so verbreitet sind?