



Datengenerator für Daten mit Bias als Grundlage für Data Science Projekte

Studienarbeit

für die Prüfung zum
Bachelor of Science

des Studiengangs Informatik
an der Dualen Hochschule Baden-Württemberg Stuttgart

von
Simon Jess, Timo Zaoral

Juni 2022

Bearbeitungszeitraum
Matrikelnummer, Kurs
Ausbildungsfirma
Betreuer

04.10.2021 - 10.06.2022
8268544, 6146532, INF19C
TRUMPF GmbH & Co KG, Ditzingen
Prof. Dr. Monika Kochanowski

Erklärung

Wir versicherern hiermit, dass wir die vorliegende Studienarbeit mit dem Thema: *Datengenerator für Daten mit Bias als Grundlage für Data Science Projekte* selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt haben. Wir versichern zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

Stuttgart, Juni 2022

Simon Jess

Timo Zaoral

Abstract

Fasst die Aufgabenstellung und Ergebnisse kompakt und übersichtlich in wenigen Zeilen zusammen (4-7 Zeilen).

Inhaltsverzeichnis

Abkürzungsverzeichnis	V
Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	2
1.3 Aufbau der Arbeit	2
2 Stand der Technik	3
2.1 Daten	3
2.1.1 Datenqualität	3
2.1.2 Bias	3
2.2 Künstliche Intelligenz & Maschinelles Lernen	3
2.2.1 Was ist Künstliche Intelligenz	3
2.2.2 Teilgebiet Maschinelles Lernen	3
2.2.3 Ethik in der Künstlichen Intelligenz	3
2.3 Künstliche Intelligenz verbunden mit Bias	3
2.3.1 Diskriminierung durch verzerrte Daten	3
2.3.2 Gegenmaßnahmen	3
3 Praktischer Teil	4
3.1 Szenarien	4
3.1.1 Szenario 1	4
3.1.2 Szenario 2	5

3.2	Konzeption	5
3.2.1	Grobkonzet	5
3.2.2	Feinkonzept	5
3.3	Umsetzung	5
4	Schluss	6
4.1	Zusammenfassung	6
4.2	Evaluierung	6
4.3	Ausblick	6

Abkürzungsverzeichnis

KI Künstliche Intelligenz

Abbildungsverzeichnis

Tabellenverzeichnis

1 | Einleitung

Idee: Digitalisierung -> Digitalisierung durch KI -> KI durch Daten -> Daten sind Öl der Zukunft -> Was sind Daten

Die fortschreitende Digitalisierung ist kaum noch aus unserem Alltag wegzudenken. Durch immer mehr Programme die einem den Alltag erleichtern sollen, nutzen wir die Errungenschaften der Digitalisierung täglich. Am häufigsten wird dabei von künstlicher Intelligenz geredet und meist ist uns nicht einmal Bewusst, dass im Hintergrund mit künstlicher Intelligenz gearbeitet wird. Egal ob als intelligenten Routenplaner oder Sprachsteuerung, all diese Anwendung basieren auf Künstliche Intelligenz (KI)

Alter Text:

Die Digitalisierung ist aus unserem Leben nicht mehr wegzudenken. Viele Kleinigkeiten aus unserem Alltag werden bereits durch die Digitalisierung vereinfacht. Teils ist uns dies nicht einmal Bewusst. Neue Möglichkeiten die nur durch künstliche Intelligenz möglich sind, sind in unserem Alltag überall auffindbar. Egal ob bei der Nutzung von Navigationssystemen wie Google Maps und einer optimierten Routenplanung, oder

Die Digitalisierung ist jedoch nur eine Folge daraus, dass Daten gesammelt und analysiert werden.

Die Digitalisierung ist aus unserem Leben nicht mehr wegzudenken. Sie beeinflusst nahezu unseren ganzen Alltag, egal ob

Alter Text:

Das Arbeiten mit Daten wird immer relevanter und so werden auch in den Hochschulen und Universitäten immer mehr Vorlesungen im Bezug auf das Arbeiten mit Daten gehalten. Durch zunehmende Internetnutzung und neue Möglichkeiten Daten zu sammeln und zu speichern, wächst die Datenmenge stetig weiter an. Daten allein sind jedoch relativ wertlos, was sie wertvoll macht, ist die Arbeit mit den Daten. Bei dieser Arbeit handelt es sich in den meisten Fällen um die Entwicklung von Maschinellen Lernmodellen oder das Konzipieren und Trainieren einer KI, die durch die gesammelten Daten trainiert werden kann. Durch die Verwendung der Daten als Lerndaten für eine KI, können Entscheidungen aber auch Vorhersagen auf Basis der in den Lerndaten enthaltenen Zusammenhänge treffen. Da sich letztendlich alles um Zusammenhänge in Daten dreht, ist auch ein wichtiges Themenfeld im Bereich der KI die Ethik. Da Fragen gestellt werden wie "Inwieweit darf die KI Entscheidungen aufgrund von Datengrundlagen treffen". [1]

1.1 Motivation

- Anonymisierung/Pseudonymisierung bei besonders großen Datensätzen ist schwierig
- Nachvollziehbarkeit von Bias verzerrten Daten

- Veranschaulichung von Bias in Daten für die Allgemeinheit, um auf das Problem im Bereich ML aufmerksam zu machen

1.2 Zielsetzung

- Datengenerator für Bias verzerrte Daten
- Visualisierung von Bias in Lerndaten für ML
- Gesamt Produkt zur Erstellung von Daten und derer Bias Visualisierung für die Lehre
- 1 Satz, was sollen wir machen -> Stichwortliste mit Anforderungen
- Maschinelles Lernen hängt von den Trainingsdaten ab.
- Trainingsdaten können einen Bias Data enthalten.

1.3 Aufbau der Arbeit

2 | Stand der Technik

2.1 Daten

2.1.1 Datenqualität

2.1.2 Bias

- Begriffserklärung: Data Bias vs Bias Verzerrung (zu viel/zu wenig lernen im ml)
 - Arten von Bias:
 - Bias durch Abwesenheit - Wenn eine Info fehlt, kann das zu Diskriminierung führen.
 - Diskriminierung durch Menschen.
- Arten von Bias: Cognitive, Social, Perceptual und Motivational Bias [2]

2.2 Künstliche Intelligenz & Maschinelles Lernen

2.2.1 Was ist Künstliche Intelligenz

2.2.2 Teilgebiet Maschinelles Lernen

- Supervised learning
- Unsupervised learning

2.2.3 Ethik in der Künstlichen Intelligenz

2.3 Künstliche Intelligenz verbunden mit Bias

2.3.1 Diskriminierung durch verzerrte Daten

2.3.2 Gegenmaßnahmen

- Wenn der Parameter mit dem Bias entfernt wird, wird das Ergebnis erstmal schlechter.

3 | Praktischer Teil

In diesem Teil der Arbeit werden zuerst die beiden Szenarien erläutert und daraufhin die Konzeption und Umsetzung derer in Python beschrieben.

3.1 Szenarien

Für das generieren von Daten wurden zwei möglichst reale Szenarien ausgewählt. Zum einen das Szenario eines Bewährungsantrages, für welches 5 verschiedene Attribute und eine endgültige Bewertung mit stattgegeben oder nicht generiert werden. Zum anderen das zweite Szenario des social creditpoint system, für welches pro Person 7 Attribute zu generieren sind und eine numerische Bewertung zwischen 600 und 1400 creditpoints erstellt wird. Diese beiden Szenarien werden im folgenden genauer erläutert.

3.1.1 Szenario 1

In Szenario 1 soll ein Bewährungsantrag einer Person Bewertet werden. Ein Antrag besteht dabei aus dem Namen der Person, dessen Geschlecht, Hautfarbe und den entscheidenden Attributen der laufenden Strafe in Jahre und der Härte des Vergehens. Basierend auf diesen Attributen soll ein Bewerter beurteilen, ob der Antrag genehmigt oder abgelehnt wird. Das Geschlecht wird in „Männlich“ und „Weiblich“ angegeben. Die Hautfarbe der Person wird als „Schwarz“ oder „Weiß“ festgehalten. Die noch laufende Strafe des Gefangenen wird in Jahren von als Ganzzahlen von 1-5 angegeben. Da hier definiert wird ein Bewährungsantrag kann erst ab maximal 5 Jahren noch offene Strafe gestellt werden. Die Härte des Vergehens wird einfachheitshalber in den Gruppen „Leicht“, „Mittel“ oder „Hart“ festgehalten.

Für die Beurteilung des Antrags von dem Bewerter werden folgende Regeln definiert:

Attribut	Positive Auswirkung	Negative Auswirkung
Laufende Strafe	1-3	4-5
Härte des Vergehens	Leicht, Mittel	Hart

Das Geschlecht und die Hautfarbe werden hierbei nicht direkt aufgelistet, da diese in der Regel keine Auswirkung auf die Bewertung haben sollten. Diese können jedoch durch einen konkreten Bias Aussagekraft bekommen. Damit soll dann in den generierten Daten ein Bias auf ein gewissen Wert gelegt werden können. Somit sind in diesem Szenario die möglichen Werte, welche durch einen Bias beeinflusst werden können, das Geschlecht und die Hautfarbe.

3.1.2 Szenario 2

3.2 Konzeption

3.2.1 Grobkonzept

3.2.2 Feinkonzept

3.3 Umsetzung

4 | Schluss

4.1 Zusammenfassung

- Fazit ziehen!!!

4.2 Evaluierung

4.3 Ausblick

Literaturverzeichnis

- [1] T. Dullien, “Maschinelles lernen und künstliche intelligenz in der informationssicherheit,” *Datenschutz und Datensicherheit-DuD*, vol. 42, no. 10, pp. 618–622, 2018.
- [2] P. A., A. Jawaid, S. Dev, and V. M.S., “The patterns that don’t exist : Study on the effects of psychological human biases in data analysis and decision making,” in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp. 193–197, 2018.