



# **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**  
**Betreuer**

04.10.2021 - 10.06.2022  
8268544, 6146532, INF19C  
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
<b>1 Einleitung</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Zielsetzung . . . . .	2
1.3 Aufbau der Arbeit . . . . .	3
<b>2 Stand der Technik</b>	<b>4</b>
2.1 Daten . . . . .	4
2.1.1 Datenqualität . . . . .	4
2.1.2 Bias . . . . .	4
2.2 Künstliche Intelligenz & Maschinelles Lernen . . . . .	4
2.2.1 Was ist Künstliche Intelligenz . . . . .	4
2.2.2 Teilgebiet Maschinelles Lernen . . . . .	4
2.2.3 Ethik in der Künstlichen Intelligenz . . . . .	4
2.3 Künstliche Intelligenz verbunden mit Bias . . . . .	4
2.3.1 Diskriminierung durch verzerrte Daten . . . . .	4
2.3.2 Gegenmaßnahmen . . . . .	4
<b>3 Praktischer Teil</b>	<b>5</b>
3.1 Szenarien . . . . .	5
3.1.1 Szenario 1 . . . . .	5
3.1.2 Szenario 2 . . . . .	6

3.2	Konzeption . . . . .	7
3.2.1	Grobkonzept . . . . .	7
3.2.2	Feinkonzept . . . . .	8
3.3	Umsetzung . . . . .	10
<b>4</b>	<b>Schluss</b>	<b>11</b>
4.1	Zusammenfassung . . . . .	11
4.2	Evaluierung . . . . .	11
4.3	Ausblick . . . . .	11

# Abkürzungsverzeichnis

<b>KI</b>	Künstliche Intelligenz
<b>bzw.</b>	beziehungsweise

# Abbildungsverzeichnis

3.1	Programmablaufplan der fünf Hauptschritte zur Generierung der Daten . .	7
3.2	Verbindungen zwischen den Attributen des zweiten Szenario . . . . .	9

# Tabellenverzeichnis

3.1	Tabelle für die Auswirkung der Attributen von Szenario 1 . . . . .	5
3.2	Tabelle der Attribute und Auswirkungen von Szenario 2 . . . . .	6



# 1 | Einleitung

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. Häufig ist hier die Rede von künstlicher Intelligenz. Dabei ist uns meist nicht einmal Bewusst, dass im Hintergrund mit künstlicher Intelligenz gearbeitet wird. Egal ob als intelligenten Routenplaner oder Sprachsteuerung, hinter all diese Anwendung steckt heute nicht mehr nur ein Optimierungsalgorithmus sondern Künstliche Intelligenz (KI).

Mit der Digitalisierung hat man begonnen große Datenmengen zu sammeln. Durch den technischen Fortschritt im Bereich von Big Data, werden diese Datenmengen heutzutage unvorstellbar groß. Mit dem Erfassen und Speichern von Daten ist man in der Lage seine Produkte stetig zu verbessern und sogar neue Geschäftsmodelle zu schaffen. Zu diesen neuen Geschäftsmodellen gehört auch die nicht mehr aus unserem Alltag wegzudenken KI. Sie ermöglicht es uns Entscheidungen zu treffen, wie sie auch ein Mensch treffen könnte, aber auch Vorhersagen zu machen, was zwar für den Menschen möglich ist, aber mit viel Aufwand verbunden ist. Egal ob eine Entscheidung oder eine Vorhersage von einer KI getroffen werden, dahinter stehen Daten die bereits gesammelt wurden und die Entscheidungsgrundlage für die KI bilden. Aus diesem Grund werden Daten eine wertvolle Ressource, sobald man anfängt die Daten zu verarbeiten und aktiv zu nutzen.

Für KI werden die Daten zum Lernen benutzt. Entscheidend für die Qualität der KI ist daher in den meisten Fällen die Datengrundlage auf der die KI basiert. Lernen bedeutet, dass Zusammenhänge und die dadurch abgebildeten Verhaltensweisen von der KI erkannt und sich selbst angeeignet werden. Durch diese Art des Lernens, wie auch wir Menschen unser Wissen erlernen, ergeben sich nicht nur Potentiale sondern auch Gefahren! Abhängig von der Datenqualität und Richtigkeit beziehungsweise (bzw.) Zuverlässigkeit der Daten werden zukünftige Entscheidungen und Vorhersagen getroffen. Eine KI betrachtet dabei die Daten vollkommen neutral ohne Hintergrundwissen über Richtigkeit und Zuverlässigkeit. Deshalb können Verzerrungen in den Daten durch die KI nicht erkannt werden. Diese Verzerrung wird auch Bias genannt und befindet sich in den Trainingsdaten mit denen die KI lernt. Die Folge daraus ist, dass sich KIs benachteiligende und diskriminierende Verhaltensweisen angeeignen und diese selbst in der Praxis ausüben.

Insbesondere für durch Computer getroffene Entscheidungen und Vorhersagen spielt die Ethik daher eine große Rolle. Diese kann in der Regel nicht aus den Daten erlernt werden und hängt von uns Nutzern ab. So stellt sich die Frage wie sollen Menschen mit Entscheidungen durch KI umgehen und sich auf diese Verlassen. Diese fehlende Ethik sorgt für

nicht zu Vernachlässigende Verzerrungen und bildet so einen Bestandteil der "Dark side of KI". beziehungsweise (bzw.)

### 1.1 Motivation

Eine KI und deren Entscheidungen basieren stets auf Daten aus der Vergangenheit, den Trainingsdaten. Wenn diese Trainingsdaten durch einen Bias verzerrt sind, ist das nicht unbedingt bekannt. In den meisten Fällen ist eine solche Verzerrung verborgen und wird erst im produktiven Betrieb der KI festgestellt.

Diese Verzerrungen führen dann häufig zu Skandalen in der Medienwelt. Es wurde bereits diverse Male in der Presse darüber berichtet, dass bspw. in Unternehmen Bewerbungen durch ein KI vorsortiert werden und eine Diskriminierung in dem Muster der Auswahl erkennbar waren. Diese Diskriminierungen sind jedoch nicht zu vergessen immer auf Daten zurückzuführen und somit auch auf die Ersteller der Daten, also die Menschen dahinter.

- 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 [1]

## **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

Tabelle 3.1: Tabelle für die Auswirkung der Attributen von Szenario 1

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 in den generierten Daten die gewünschte Verzerrung auf einen gewissen Wert gelegt werden können. In diesem Szenario sind die möglichen Werte, welche durch eine Verzerrung und damit einem menschlichem Vorurteil eines Bewerter beeinflusst werden können, das Geschlecht und die Hautfarbe. Die anderen beiden Attribute, welche in der Tabelle 3.1 aufgeführt sind, wirken sich durch ihre Ausprägungen positiv oder negativ auf die Bewertung des Antrages aus. So wirkt z.B. eine Härte des Vergehens vom Niveau Leicht sich eher für eine positive Bewertung des Antrages aus, als eine mittlere Härte. Dasselbe gilt auch für die Laufende Strafe. So kann ein Bewerter dann anhand dieser beiden Werte eine Tendenz erhalten und dann über die Gestattung des Antrages entscheiden.

### 3.1.2 Szenario 2

Im zweiten Szenario wird das durch China populär gewordene sozial creditpoint System in einer lagenunabhängigen Version nachgebaut. Dafür werden Einträge zu Personen erstellt, nach welchen die Punktzahl der einzelnen Person zwischen 600 und 1400 Punkten bestimmt wird. Ein Eintrag zu einer Person beinhaltet die sieben in der folgenden Tabelle dargestellten Attribute mit den unterschiedlichen Ausprägungen. Die in Tabelle 3.2

Attribut	Ausprägungen
Name	Beliebig
Alter	20-79
Politische Orientierung	Links, Mitte, Rechts
Bildungsabschluss	Ausbildung, Fachschulabschluss, Bachelor, Master, Diplom, Promotion, ohne
Soziales	0-3
Wohnlage	Großstadt, Kleinstadt, Vorort, Ländlich
CO2-Fußabdruck	4-12

Tabelle 3.2: Tabelle der Attribute und Auswirkungen von Szenario 2

aufgeführten Ausprägungen haben ähnlich wie zu Szenario 1 unterschiedlich starke Auswirkungen auf den am Ende bestimmten social Score. Einzig allein der Name und das Alter sollen keine direkte Auswirkung auf den social Score haben. Die anderen Attribute wirken sich je nach Auswirkung positiv durch eine Erhöhung des Scores oder negativ durch eine Verringerung des Scores aus. Insgesamt werden so in diesem Szenario viele Einträge von Personen erstellt, welche alle unterschiedlichste Verteilungen der Ausprägungen besitzen und dadurch in der Bewertung einen individuellen social Score erzielen. Um nun eine gewünschte Verzerrung in die Daten zu bekommen können alle Attribute bis

auf den Namen, welcher rein als Füllwert dient, durch eine Verzerrung beeinflusst werden. So können z.B. Personen zwischen 20-30 Jahre negativ verzerrt werden, da ein oder zwei Bewerter etwas gegen junge Leute haben und diesen aus ihrer Überzeugung einen schlechteren Score geben. In diesem Szenario ist somit eine hohe Variabilität geboten inwieweit eine Verzerrung in die Daten gebracht wird. Zudem kann auch eine Verzerrung über mehrere Attribute eingebracht werden, da ein Bewerter z.B. auch etwas gegen eine Rechte Politische Orientierung und ein schlechtes Soziales Engagement von 0 haben kann.

## 3.2 Konzeption

In diesem Kapitel wird die erarbeitete Konzeption für die Umsetzung der beiden im Kapitel 3.1 aufgeführten Szenarien erläutert. Dabei wird in ein Grobkonzept zur allgemeinen Generierung der Daten und darauf in ein Feinkonzept für jedes Szenario unterteilt.

### 3.2.1 Grobkonzept

Das Grobkonzept beinhaltet die Überlegungen, wie die Programme/Notebooks für die beiden Szenarien generell aufgebaut sein sollen. Der Ablauf der Programme von der Eingabe der Parameter bis hin zu den fertig generierten Daten wird in fünf Schritten durchgeführt. Der Ablauf der Schritte ist in folgendem Programmablaufplan dargestellt.

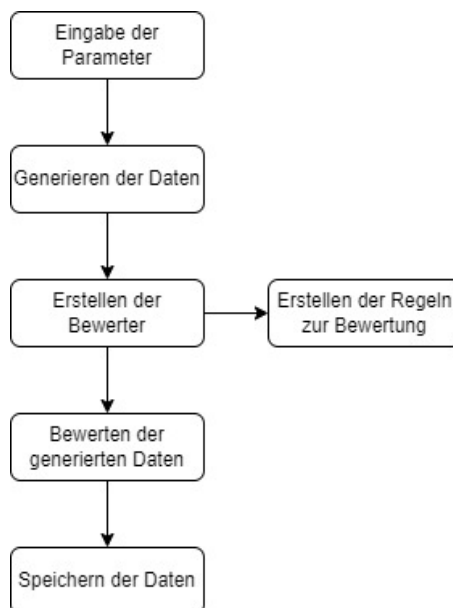


Abbildung 3.1: Programmablaufplan der fünf Hauptschritte zur Generierung der Daten

Die in der Abbildung 3.1 dargestellten Hauptschritte des Programmablaufs lauten: Parametereingabe, Generieren der Daten, Regeln aufstellen, Bewerten und Speichern der Daten. Im ersten Schritt der Parametereingabe, wird den Benutzenden die Möglichkeit gegeben die Parameter für die Generierung der Daten einzugeben, wie z.B. die Anzahl der Daten oder Bewertende welche generiert werden sollen. Im folge Schritt werden daraufhin die passende Anzahl an Daten für das jeweilige Szenario generiert. Dabei sollen die Daten möglichst an Verhältnissen aus der Realität angepasst und auf dieser Grundlage generiert werden. Es soll jedoch eine gewisse Zufälligkeit in der Generierung vorhanden sein, sodass bei mehrfach Generierung unterschiedliche Datensätze auf Basis der definierten Verteilungen entstehen. Nach Abschluss der Generierung wird der fertige Datensatz zwischengespeichert, um diesen später bewerten zu können. Für die Bewertung des Datensatzes muss im folgenden 3. Schritt die Regeln nach welchen bewertet wird aufgestellt werden. Dafür soll zuerst die in den Parametern gewünschte Anzahl an Bewertenden erstellt werden, da durch diese die Regeln erstellt werden. Unter den Erstellten Bewertenden müssen zudem noch die angegebene Anzahl an diskriminierenden Bewertenden in solche umgewandelt werden. Daraufhin können dann die Regeln für die Bewertenden erzeugt werden. Somit ist der 3. Schritt abgeschlossen und alle Vorbereitungen getroffen für die Bewertung. In der Bewertung bekommen die Bewertenden alle Anträge des Datensatzes vorgelegt, welche Sie basierend auf den Regeln bewerten. Die Bewertung des jeweiligen Antrages wird diesem in den Daten hinzugefügt. Damit kann zum letzten Prozess übergegangen werden. In diesem wird der Ursprungsdatensatz und der bewertete Datensatz abgespeichert und damit auch außerhalb von dem Programm zugänglich gemacht. Somit ist das Grobkonzept der Szenarien abgeschlossen und ein Grundgerüst konnte entworfen werden. Im weiteren kann nun auf die detaillierte Feinkonzeption der einzelnen Szenarien eingegangen werden.

### 3.2.2 Feinkonzept

Die im Grobkonzept beschriebenen fünf Hauptschritte der Programme sind für beide Szenarien gleich. Jedoch unterscheiden sich die Schritte im Detail bei beiden Szenarien. Daher wird für jedes Szenario ein eigenes Feinkonzept zur Füllung des selben vorhandenen Grundgerüst entwickelt.

#### **Parametereingabe**

Im ersten Schritt der Parametereingabe unterscheiden sich die Szenarien nicht, da beide einen Parameter für die gewünschte Diskriminierung, die Anzahl der Daten, die Anzahl der Bewertenden, die Anzahl der diskriminierenden Bewertenden und der Stärke der Auswirkung von der Diskriminierung benötigen. Diese Parameter können von den Benutzenden



in beiden Fällen in einer finalen Zelle editiert werden.

### Daten generieren

In diesem Prozess unterscheiden sich beide Szenarien stark, da zum einen für das erste Szenario nur fünf anstelle von sieben Attribute bei Szenario 2 generiert werden müssen und zum anderen existieren deutlich weniger Verbindungen zwischen den Attributen in Szenario 1. Bei dem ersten Szenario basiert nur die Härte der Strafe und die Hautfarbe auf dem Geschlecht, dies bedeutet die Wahrscheinlichkeiten für diese Attribute soll dem Geschlecht entsprechend angepasst werden. So haben zum Beispiel weibliche Personen eine eher seltener eine Harte Strafe als männliche Personen. Alle anderen Attribute in diesem Szenario haben eine feste Wahrscheinlichkeitsverteilung. Damit sind die Zusammenhänge in diesem Szenario sehr klein gehalten und überschaubar.

In Szenario 2 müssen sieben Attribute generiert werden und es sollen deutlich mehr Verbindungen zwischen diesen existieren. Um diese verständlich darzustellen wurde ein Diagramm für das Feinkonzept entworfen, welches nachfolgend dargestellt ist.



Abbildung 3.2: Verbindungen zwischen den Attributen des zweiten Szenario

In der Abbildung 3.2 sind alle Attribute, bis auf das Attribut Name, des zweiten Szenario als abgerundete Rechtecke und die Verbindungen zwischen diesen mit Pfeilen dargestellt. Insgesamt existieren wie zu sehen fünf Verbindungen zwischen den Attributen. Diese sollen so aufgebaut werden, um möglichst Realitätsnahe Daten generieren zu können. Die Verbindung zwischen dem Alter und dem Bildungsabschluss existiert, da zum Beispiel die Wahrscheinlichkeit, dass eine Person mit 20 schon eine Promotion besitzt nicht so hoch ist wie bei einer Person im Alter von 50. Zudem hat das Alter einen Einfluss auf die Politische Orientierung einer Person, da junge Leute sicherlich andere Orientierungen haben als Personen im Alter von 50 zum Beispiel, siehe Aktionen wie „FridaysforFuture“. Durch die Politische Orientierung einer Person wird in diesem Fall auch mit einer Beeinflussung auf das soziale Verhalten gerechnet und es besteht daher hier auch eine Verbindung. Im Falle des CO2-Fußabdruck wird in dieser Arbeit mit einer Auswirkung der Wohnlage und einer Auswirkung des Sozialen gerechnet. Durch das Soziale und die Vorverbindung wird die Politische Orientierung und das Alter ebenfalls indirekt darauf mit ein Bezogen. Dies

ist hier der Fall, da die Berechnung CO<sub>2</sub>-Fußabdruck sich aus vielen Faktoren zusammensetzt und daher auch dieser in diesem Szenario durch viel beeinflusst werden soll. Durch diese Zusammenhänge sollen dann möglichst realitätsnahe Daten entstehen.

Insgesamt für beide Szenarien müssen dann später bei der Umsetzung die Wahrscheinlichkeiten der Ausprägungen der Attribute, wo es möglich ist, durch Statistiken bestimmt werden und die Verbindungen dadurch ebenfalls bestätigt werden. Dies wird im Kapitel 3.3 im Detail erläutert.

**Regeln aufstellen**

### 3.3 Umsetzung

Hier kommt die Umsetzung der Szenarien mit Code snippets usw.

## 4 | Schluss

### 4.1 Zusammenfassung

- Fazit ziehen!!!

### 4.2 Evaluierung

### 4.3 Ausblick

## Literaturverzeichnis

- [1] P. A., A. Jawaaid, 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.