Themenblock Clustering: vom Sortieren bis zum Explorieren: 06_06Diskussion_EthikRecht

# Ethik und Recht

Erarbeitet von
Dr. Katarina Boland

*Die Inhalte dieses Videos stellen keinesfalls eine Rechtsberatung in irgendeiner Form dar oder rechtliche Leitlinien. Ziels dieses Videos ist es, ein Problembewusstsein im Umgang mit KI zu schaffen und für rechtliche Fragen in diesem Kontext zu sensibilisieren. Vor dem Einsatz von KI-Systemen im Rahmen deines Projekts oder deiner Arbeit wende dich an die jeweiligen Fachstellen deiner Universität oder deines Unternehmens, um die rechtlichen Rahmenbedingungen für den Einsatz von KI zu besprechen.*



## Lernziele

- Du kannst erläutern, wie Clustering die Benachteiligung bestimmter Gruppen oder Individuen verstärken kann
- Du kannst erklären, was „protected attributes" sind
- Du kannst Beispiele für „protected attributes" nennen
- Du kannst erklären, inwiefern ethische oder rechtliche Bedenken und effektive Forschungsdesigns manchmal gegeneinander aufgewogen werden müssen
- Du kannst vier Methoden aufzählen, mit denen sensible Daten bei der Anwendung von Clustering-Verfahren geschützt werden können



Projekt ai4all „KI für Alle"
HeiCAD –Heine Center for Artificial Intelligence and Data Science

# Inhalt

## Einstieg

Bei der Erhebung, Selektion und Verarbeitung von „found data" und von „designed data" gibt es vieles zu beachten, damit keine Verzerrungen entstehen und die Ergebnisse valide und repräsentativ ausfallen.
Gleichzeitig müssen auch für explorative Analysen mittels Clustering, Sentiment Analyse und weiteren Verfahren datenschutzrechtliche und ethische Aspekte dringend beachtet werden.

Denn die Einteilung von Personen in Gruppen, auch wenn diese datengetrieben und ohne vorgegebene Klassen erfolgt, kann bestimmte Gruppen benachteiligen. Selbst wenn der Algorithmus selbst nicht unfair ist, kann er in den Daten enthaltene Verzerrungen verstärken.
Beim Clustering kann dies beispielsweise geschehen, wenn sogenannte „geschützte Attribute" (protected attributes) wie Herkunft oder Geschlecht für die Gruppeneinteilungen verwendet werden und der Clusteringalgorithmus somit Gruppen hinsichtlich dieser Merkmale unterscheidet.

**Quelle [1]**

Dies kann tückischer Weise auch dann passieren, wenn diese Attribute explizit von der Analyse ausgeschlossen werden – nämlich dadurch, dass andere Attribute herangezogen werden, die mit den geschützten Attributen korrelieren, beispielsweise Körpergröße mit Geschlecht oder Wohnsitz mit ethnischer Herkunft.

**Quelle [1]**
**Quelle [2]**

Werden solche Cluster zur Entscheidungsfindung herangezogen, beispielsweise um die Eignung für einen Job oder die Kreditwürdigkeit von Personen zu bewerten, so kann dies zu einem Verstoß gegen das Allgemeine Gleichbehandlungsgesetz führen.

**Quelle [3]**

Im US-amerikanischen Raum spricht man hier von „disparate impact", der unbeabsichtigten Diskriminierung, die vom Supreme Court untersagt wurde.

**Quelle [2]**

In der EU müssen zudem, wie für andere KI-basierte Verfahren auch, anwendungsabhängig spezifische Anforderungen des KI-Grundgesetzes beachtet werden.

**Quelle [4]**

Auch bei der Analyse von Sentimenten und Emotionen ist Vorsicht geboten. Hiermit lassen sich nicht nur in bestimmten Fällen Meinungen analysieren, die nicht explizit geäußert wurden; die Informationen können potenziell auch zur Manipulation eingesetzt werden.

**Quelle [5]**

Ein Beispiel, das vor einigen Jahren Schlagzeilen machte, ist ein Experiment, das Facebook zusammen mit zwei US-amerikanischen Universitäten durchführte.

**Quelle [6]**
**Quelle [7]**
**Quelle [8]**
**Quelle [9]**

Ohne die Nutzenden um Erlaubnis zu fragen oder sie auch nur vorab darüber zu informieren, manipulierte Facebook den Newsfeed zig tausender Nutzer*innen. Sie bekamen, je nach Gruppenzuweisung, entweder vornehmlich positive oder vornehmlich negative Beiträge ihrer Freund*innen angezeigt. Ziel war es, herauszufinden, inwiefern die Stimmungen der angezeigten Nachrichten die Stimmungen der Nutzer*innen beeinflussen. Und tatsächlich: Es zeigte sich, dass die Nutzenden, die mehr positive Beiträge angezeigt bekamen, selbst auch stärker dazu neigten, positive Beiträge zu veröffentlichen. Nutzende, die mehr negative Beiträge zu sehen bekamen, neigten stärker zur Veröffentlichung negativer Inhalte.

Eine Diskussion der Aussagekraft dieser Studie mal außen vor gelassen:
Die Empörung über dieses Vorgehen war groß. Nicht nur, dass, wie Medien berichteten, sich die Nutzenden als „Versuchskaninchen missbraucht" fühlten: Die Forschenden schienen ja davon auszugehen, dass es möglicherweise einen negativen Effekt auf die Stimmung der Nutzenden geben könnte, wenn deren Newsfeed in negativer Weise verzerrt würde. Und trotzdem setzten sie tausende Personen vorsätzlich dieser negativen Verzerrung aus.
Ist ein solches Vorgehen gesetzlich erlaubt?

Die Forscher*innen weisen in ihrer Publikation darauf hin, dass die Datenverwendungsregeln von Facebook ein derartiges Vorgehen zulassen.

**Quelle [10]**

Und diese Regeln wurden schließlich von allen Facebook-Nutzenden akzeptiert.
Gleichzeitig gibt es für Experimente mit menschlichen Proband*innen spezielle Vorgaben. Hier wird eine „informierte Einwilligung" in die Datenverarbeitung zum Studienzweck vorausgesetzt, die über eine solche Regelung hinausgehen. Die Publisher des wissenschaftlichen Journals, das die Publikation veröffentlichte, argumentieren, dass Facebook als privatwirtschaftliches Unternehmen nach geltendem US-amerikanischem Recht nicht zur Einhaltung dieser Vorgaben verpflichtet war, als es die Daten sammelte, und die resultierende Forschung deshalb zur Veröffentlichung freigegeben wurde.

**Quelle [11]**

Aber auch sie räumen ein, dass die nicht-Einhaltung der Vorgaben problematisch sei.
Wie man sieht, ist die rechtliche Einordnung also kompliziert.
Und selbst wenn das Vorgehen rechtmäßig ist: Die ethische Fragwürdigkeit bleibt allemal.

Das Problem ist hier nur: Wie lässt sich das vermeiden?
Angenommen, Facebook hätte die Nutzenden vorab direkt um Erlaubnis gefragt, ihren Newsfeed zum Zweck der Studie manipulieren zu dürfen. Selbst wenn genug Personen eingewilligt hätten, ist davon auszugehen, dass allein das Wissen um die Intervention den Effekt und das Verhalten der Nutzenden beeinflusst hätte.

**Vgl. Quelle [12]**

Die Forschung hätte also so gar nicht stattfinden können.

Dieses Beispiel illustriert sehr gut, dass ethisch-rechtliche Anforderungen und ein möglichst effektives Forschungsdesign sich manchmal gegenseitig im Weg stehen und sorgfältig, auch hinsichtlich des zu erwartenden Nutzens der Forschungsergebnisse, gegeneinander abgewogen werden müssen.

**Quelle [12]**

Dies trifft auch zu, wenn es um das Teilen von sensiblen Nutzendendaten geht:

**Quelle [13]**

Die General Data Protection Regulation (GDPR) der EU und der Health Insurance Portability and Accountability Act (HIPAA) der USA verbieten es Unternehmen, derartige Daten zu teilen. Gleichzeitig kann die Kombination von Daten verschiedener Quellen ihren Nutzen deutlich erhöhen, beispielsweise wenn Daten aus verschiedenen Krankenhäusern zusammengeführt werden. Auch ist es manchmal notwendig, Clusteringverfahren, die auf großen Datenmengen arbeiten müssen, auf Hochleistungs-Cloudserver auszulagern. Um den Datenschutz trotzdem gewährleisten zu können, gibt es spezielle Ansätze (Stichwort: privacy-preserving clustering). Hierzu gehören:

1. Anonymisierung
2. Verschlüsselung
3. Datenpartitionierung
4. Perturbationen

**Quelle [13]**
**Quelle [14]**

Auch bei diesen gilt es wieder, Abwägungen zu treffen.
Anonymisierung von Daten kann auf Kosten ihrer Granularität oder Genauigkeit gehen, denn hier kann es sein, dass sensible Informationen aggregiert, ersetzt oder vollständig entfernt werden müssen; Verschlüsselungsmechanismen und ausgefeilte

Datenpartitionierungsansätze, bei denen Datensätze in verschiedene Partitionen unterteilt und getrennt verarbeitet werden, können die Komplexität erhöhen, da hier zusätzliche Berechnungen durchgeführt werden müssen. Perturbationen, also das absichtliche Hinzufügen von Rauschen/Verzerrungen durch die Änderung oder Vertauschung von Werten, können sensible Originaldaten verschleiern und verhindern, dass sie rekonstruiert werden können. Allerdings muss darauf geachtet werden, dass relevante statistische Eigenschaften und Relationen (z. B. Durchschnittswerte, Varianz und Korrelationen) erhalten bleiben, da andernfalls die Genauigkeit der Clusteringverfahren negativ beeinflusst werden kann.

**Quelle [14]**

Du siehst also, dass ethische und rechtliche Fragestellungen auch für explorative Datenanalyse mit Verfahren wie Clustering und Sentiment Analyse nicht einfach zu beantworten sind.
Du solltest dich daher im Regelfall bereits beim Design von Studien und Experimenten von Ethikkommissionen und Rechtsexpert*innen beraten lassen. Ansprechpartner*innen hierfür findest du in den verschiedenen Fakultäten der Universität.

## Quellen


Quelle [1] Chierichetti, F., Kumar, R., Lattanzi, S., & Vassilvitskii, S. (2017). Fair clustering through fairlets. Proceedings of the 31st International Conference on Neural Information Processing Systems, 5036–5044.

Quelle [2] Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and Removing Disparate Impact. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 259–268. https://doi.org/10.1145/2783258.2783311

Quelle [3] Antidiskriminierungsstelle des Bundes, Allgemeines Gleichbehandlungsgesetz (AGG), Abgerufen 27. August 2024, von https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/AGG/agg_gleichbehandlungsgesetz.pdf?__blob=publicationFile

Quelle [4] KI-Gesetz: Erste Regulierung der künstlichen Intelligenz. (2023, Juni 8). Themen | Europäisches Parlament. Abgerufen 27. August 2024, von https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz

Quelle [5] Mohammad, S. M. (2022). Ethics Sheet for Automatic Emotion Recognition and Sentiment Analysis. Computational Linguistics, 48(2), 239–278. https://doi.org/10.1162/coli_a_00433

Quelle [6] Facebook manipulierte Hunderttausende Newsfeeds. (2014, Juni 29). Süddeutsche.de. Abgerufen 27. August 2024, von

https://www.sueddeutsche.de/wirtschaft/psychologisches-experiment-hunderttausende-manipulierte-newsfeeds-bei-facebook-1.2022006

Quelle [7] Facebook-Experiment: Ärger um manipulierte Newsfeeds. (2014, Juni 29). Der Spiegel. Abgerufen 27. August 2024, von https://www.spiegel.de/netzwelt/web/facebook-experiment-aerger-um-manipulierte-newsfeeds-a-978147.html

Quelle [8] Booth, R. (2014, Juni 29). Facebook reveals news feed experiment to control emotions. The Guardian. Abgerufen 27. August 2024, von https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds

Quelle [9] Morin, R. (2014, Juli 2). Facebook's experiment causes a lot of fuss for little result. Pew Research Center. https://www.pewresearch.org/short-reads/2014/07/02/facebooks-experiment-is-just-the-latest-to-manipulate-you-in-the-name-of-research

Quelle [10] Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Quelle [11] Editorial Expression of Concern: Experimental evidence of massivescale emotional contagion through social networks. (2014). Proceedings of the National Academy of Sciences, 111(29), 10779–10779. https://doi.org/10.1073/pnas.1412469111

Quelle [12] Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. Public Opinion Quarterly, 85(S1), 399–422. https://doi.org/10.1093/poq/nfab018

Quelle [13] Hegde, A., Möllering, H., Schneider, T., & Yalame, H. (2021). SoK: Efficient Privacy-preserving Clustering. Proceedings on Privacy Enhancing Technologies, 2021 (4), 225–248. https://eprint.iacr.org/2021/809, https://eprint.iacr.org/2021/809, https://eprint.iacr.org/2021/809.pdf

Quelle [14] Wie können Sie sensible Daten beim Clustering und Mining schützen? (o. J.). Abgerufen 27. August 2024, von https://de.linkedin.com/advice/0/how-can-you-protect-sensitive-data-during-clustering?lang=de

## Disclaimer

Transkript zu dem Video „Clustering: vom Sortieren bis zum Explorieren: Ethik und Recht“, Katarina Boland.