

# Standortbestimmung mittels ADS-B Sensordaten und maschinellem Lernen

David Fuhry<sup>1</sup>

Universität Leipzig  
Machine Learning Group  
Leipzig, Germany  
`df15nocu@studserv.uni-leipzig.de`

**Zusammenfassung.** Es wird ein Verfahren vorgestellt um mit ADS-B Sensordaten den Breitengrad eines Flugzeugstandortes zu ermitteln. Zum Einsatz kommen Regressionsverfahren des maschinellen Lernens. Der Breitengrad kann mit einer mittleren Genauigkeit von unter 20km bestimmt werden. Für eine produktive Anwendung in der Luftraumüberwachung müssen die eingesetzten Verfahren weiter optimiert werden.

**Schlüsselwörter:** Maschinelles Lernen, ADS-B, Regression, Positionsbestimmung

## 1 Einleitung

Verschwörungstheorien sind aktuell häufig im Zentrum der Aufmerksamkeit. Während dieser, aktuelle Aufmerksamkeitsschub inhaltlich auf die aktuelle Pandemie zurückzuführen ist, bietet das Internet und die Verbreitung sozialer Medien das technische Fundament für die schnelle Verbreitung wie sie aktuell stattfindet. Das zunehmende Vernetzung solchen Theorien vorschub leistet ist keine neue Entwicklung, sondern wird schon länger im wissenschaftlichen Diskurs über Verschwörungstheorien beobachtet. Die wissenschaftliche Auseinandersetzung mit Verschwörungstheorien ist trotz des verstärkten Interesses in der jüngsten Vergangenheit bis heute eher verhalten.

Die komplexe Natur von Verschwörungstheorien stellt die wissenschaftliche Auseinandersetzung mit ihnen vor einige Grundlegende Probleme. Das offensichtlichste davon ist sicherlich die Frage was überhaupt eine Verschwörungstheorie ist. Hierzu existiert vor allem in der Wissenssoziologie eine Vielzahl von Literatur, ein klarer Konsens über eine Definition von Verschwörungstheorien fehlt jedoch bislang. Insbesondere für die Arbeit mit Beiträgen aus sozialen Netzen und anderen Online-Quellen schließt sich die Frage an, wie Verschwörungstheoretische Beiträge zuverlässig identifiziert werden können. Der Nutzen solcher Techniken nicht nur innerhalb der Wissenschaft ist naheliegend, so versuchen auch die Betreiber von sozialen Netzwerken daran verschwörungstheoretische Inhalte automatisiert zu erkennen (Beispiel google/youztube). In der Forschung wird dieses Problem klassischerweise mit der Hilfe einer Vielzahl von Hilfskräften gelöst, die die Daten manuell bearbeiten. (Studie mit codierern)

Diese Arbeit stellt einen Ansatz vor, dieses Problem mit Technologien der automatischen Sprachverarbeitung und des maschinellen Lernens zu lösen. Primär werden dabei stilistische Merkmale der untersuchten Texte genutzt und nur sekundär inhaltliche Informationen wie das Verwenden von Schlüsselwörtern etc. Durch das Verwenden solcher Technologien kann zum einen der Ressourcenaufwand drastisch reduziert werden, aber auch die Abhängigkeit von komplexen und damit fehleranfälligen Kriterien für die Klassifizierung lässt sich so umgehen. Auch ist der hier vorgestellte Ansatz durch den Fokus auf stilistische Merkmale verhältnismäßig robust gegenüber neuen inhaltlichen Entwicklungen der Verschwörungstheorien und somit potentiell zukunftsicherer als vergleichbare Ansätze.

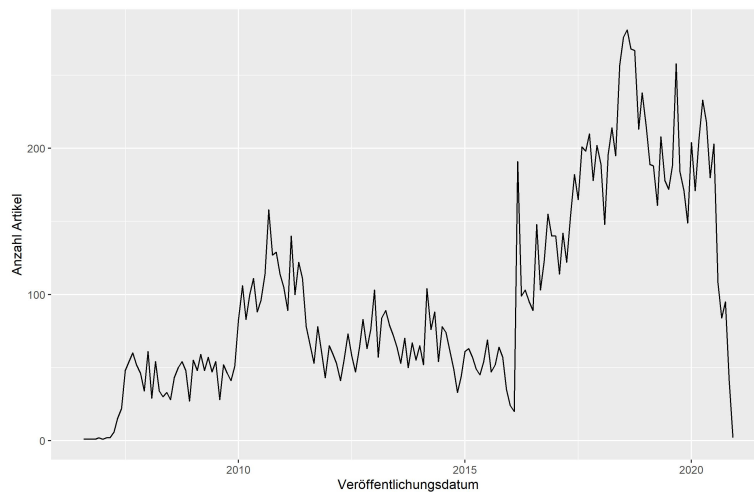
Im folgenden soll zunächst ein kurzer Abriss über den aktuellen Stand der wissenschaftlichen Auseinandersetzung mit Verschwörungstheorien allgemein, sowie der Forschung zu linguistischen Merkmalen im speziellen, gegeben werden. Anschließend soll kurz auf die verwendeten Technologien sowie die Datengrundlage dieser Arbeit gegeben werden, bevor konkrete Ergebnisse diskutiert werden.

## 2 Umsetzung

### 2.1 Datenbasis

Als Grundlage dieser Arbeit wurde ein Korpus aus Texten von insgesamt 7 deutschsprachigen, der verschwörungstheoretiker Szene zuzurechnenden Internetangeboten erstellt. Die Skripte zum automatischen Abruf der Texte nutzen dafür das R-Paket *rvest* (Wickham, 2020) erstellt und extrahieren mittels für jede Seite separat erstellten CSS-Selektoren und XPATH-Querys den jeweiligen Artikeltext. In diesem Vorgang wurden noch kleinere Bereinigungen an den Texten vorgenommen um wiederkehrende, nicht zum Artikeltext gehörende Elemente wie Werbung oder Spendenaufrufe zu entfernen, die Texte selbst wurden jedoch soweit möglich nicht weiter bearbeitet. Weiterhin erfasst wurde zu den Artikeln das (angegebene) Veröffentlichungsdatum, der Artikeltitel sowie die Rubrik so diese Angabe vorhanden waren.

R zitieren so nicht schon geschehen



**Abb. 1.** Anzahl Artikel nach Veröffentlichungsdatum (nach Monat gruppiert)

Der so erstellte Korpus umfasst insgesamt 16836 Texte und deckt beginnend in 2006 einen Zeitraum von 14 Jahren ab. Wie in Grafik 1 zu erkennen ist, steigt beginnend in 2016 die Zahl der Artikel im Korpus deutlich an. Dies liegt zum einen daran, dass einzelne der betrachteten Angebote die Veröffentlichungsfrequenz erhöht haben, ist aber überwiegend darauf zurückzuführen, dass nicht alle Internetangebote über den gesamten Zeitraum existiert haben sondern erst später aktiv wurden. Genauere Informationen zu den von den einzelnen Angeboten abgedeckten Zeiträumen sowie zu Artikelzahl und Länge finden sich in Tabelle 1.

Wirklich?

Wie dort auch ersichtlich ist, unterscheidet sich nicht nur die absolute Zahl der Artikel nach Quelle stark, auch das Veröffentlichungsintervall schwankt von weniger als einem Artikel den Monat bis über 100. Es besteht weiterhin ein negativer Zusammenhang zwischen der Artikellänge und dem Veröffentlichungsintervall ( $r = -0.72, p = 0.065$ ).

r richtiges Zeichen?

	Von	Bis	Mittlere Zeichenzahl	Anzahl Artikel	Artikel/Monat
Alles Schall und Rauch	23.08.2006	02.08.2020	5000	5349	32.0
conrebbi	05.09.2012	20.10.2014	5610	24	1.0
deutschlandpranger	29.10.2016	30.12.2020	7527	118	2.4
fm-tv	31.07.2008	02.11.2018	9158	97	0.8
hinterderfichte	16.01.2010	31.05.2018	4221	1083	10.8
recentr	03.08.2007	05.08.2020	5071	4762	30.5
Watergate.tv	20.05.2016	16.11.2020	2810	5403	101.9

**Tabelle 1.** Kennzahlen der einzelnen Quellen des Korpus

Auch eine inhaltliche Betrachtung zeigt deutliche Unterschiede zwischen den einzelnen Angeboten. So ist etwa *Alles Schall und Rauch (ASuR)* als ältestes vertretenes Angebot auch einer der "klassischsten" Vertreter des Genres. Hier wird der Nutzer direkt auf der Startseite mit Fragen konfrontiert wie "Was geschau wirklich am 11. September?oder "Was passiert tatsächlich mit dem Klima?"(Freeman, 2021).

Das muss anders

Eventuell zitation zu häufigen fragen.

Die Seite bedient sich klassischer Verschwörungstheorien und wird im allgemeinen der Truthther Szene zugeordnet („Alles Schall und Rauch“, 2020). Sie hat laut eigenaussage im Schnitt über 50.000 Zugriffe täglich<sup>1</sup>.

Inhaltlich finden sich sowohl Essay-artige Artikel zu Themen wie 9/11, Bilderberger oder Klimawandel auf der Seite, als auch kürzer gehaltene Meldungen zu aktuellen Ereignissen. Es werden, wie bei fast allen anderen Seiten im Korpus, häufig Inhalte Dritter eingebunden sowie die Artikel durch Bilder oder Grafiken ergänzt.

Ähnlich gelagert sind die Angebote *Hinter der Fichte (HdF)* und *conrebbi*, beide können der Truthther Szene zugerechnet werden (siehe etwa „Conrebbi“, 2020). In beiden Angeboten werden aktuelle Ereignisse kommentiert und verschwörungstheoretisch interpretiert, sowie längere Essays zu typischen Themen verfasst. Im Vergleich zu *ASuR* werden bei beiden Angeboten deutlich mehr Inhalte in Form von Videos und Fremdquellen in Bild und Textform genutzt.

Etwas aus der Reihe fallen die Angebote *fm-tv* sowie *Deutschlandpranger*. Inhaltlich finden sich etwa beim *Deutschlandpranger* recht klassische Themen

<sup>1</sup> Der Besucherzähler der Website ist inzwischen nicht mehr funktional, Angabe übernommen aus (Anonymous, 2020)

wie die Leugnung des Klimawandels (vgl. 2018), sind die dazugehörigen Text überdurchschnittlich lang und voller Werbung für Bücher zu z.T. komplett unabhängigen Themen. Auch sind die Texte selbst teilweise relativ wirr, so enthalten beispielsweise die (sehr ausführlichen) AGB einen Abschnitt zu Strafzahlungen bei "Übersenden eines Statements anstatt einer echten Rechnung (True Bill) des wahren Haftungsgläubigers"(2018).

## Software

Wickham, H. (2020). *rvest: Easily Harvest (Scrape) Web Pages* [R package version 0.3.6]. R package version 0.3.6. <https://CRAN.R-project.org/package=rvest>

## Software

- Alles Schall und Rauch*. (2020). Psiram. Verfügbar 23. Mai 2021 unter [https://www.psiram.com/de/index.php/Alles\\_Schall\\_und\\_Rauch](https://www.psiram.com/de/index.php/Alles_Schall_und_Rauch)
- Anonymous. (2020). *Mein Vater ist der Verschwörungstheoretiker hinter Alles Schall und Rauch*. Vice. Verfügbar 23. Mai 2021 unter <https://www.vice.com/de/article/gyx8aq/mein-vater-der-bekannte-verschwoerungstheoretiker>
- Conrebbi*. (2020). Psiram. Verfügbar 23. Mai 2021 unter <https://www.psiram.com/de/index.php/Conrebbi>
- Freeman. (2021). *Homepage*. Verfügbar 23. Mai 2021 unter <http://alles-schallundrauch.blogspot.com/>
- Verbal-Fäkalien: Das Treibhausgas*. (2018). Deutschland - Pranger. Verfügbar 23. Mai 2021 unter <http://deutschland-pranger.de/b2evolution/index.php/treibhauseffekt-jpg>