

Automatisierte Erkennung von verschwörungstheoretischen Artikeln

Mit Methoden der natürlichen Sprachverarbeitung und des maschinellen
Lernens

David Fuhry

Arbeit zur Erlangung des akademischen Grades des
B.Sc. Digital Humanities

Betreuer: Jun.-Prof. Dr. Phil. Manuel Burghardt

Universität Leipzig
Institut für Informatik
Augustusplatz 10
04109 Leipzig

Inhaltsverzeichnis

1	Einleitung	1
2	Theorie	1
2.1	Stand der Forschung	1
2.2	Merkmale von verschwörungstheoretischen Texten	3
2.3	Textklassifizierung	5
3	Vorgehen	6
4	Umsetzung	6
4.1	Datenbasis	6
4.2	Vergleichskorpus	9
4.3	Datenvorverarbeitung	10
4.4	Features	11
4.5	Modellerstellung	12
5	Diskussion	14
6	Fazit	16

1 Einleitung

Verschwörungstheorien sind aktuell häufig im Zentrum der öffentlichen Aufmerksamkeit. Während dieser Aufmerksamkeitsschub vor allem auf die Covid-19 Pandemie zurückzuführen ist, erleben Verschwörungstheorien schon seit Anfang der 2000er Jahre einen Popularitätsanstieg. Dieser ist vor allem begründet in der zunehmenden Verbreitung des Internets und der sozialen Medien (vgl. Stano, 2020, S. 492).

Eine der Kernfragen in der wissenschaftlichen Auseinandersetzung mit Verschwörungstheorien ist die Frage was eine Verschwörungstheorie ist und was nicht. Eine theoretische Auseinandersetzung mit dieser Frage findet insbesondere in der Wissenssoziologie und der Philosophie statt.¹ Weitgehend unbeantwortet ist hingegen wie sich diese Frage quantitativ beantworten lässt. Die wenigen hierzu veröffentlichten Arbeiten nutzen meist verhältnismäßig spezifische und komplexe Methoden.²

Diese Arbeit stellt einen Ansatz vor, der versucht dieses Problem mit Methoden des Text-Mining und des maschinellen Lernens zu lösen. Der vorgestellte Ansatz ist dabei deutlich allgemeiner als die in der Literatur vertretenen und versucht verschwörungstheoretische Texte mittels sprachlicher und stilistischer Merkmale zu identifizieren.

Es wird dafür ein Korpus aus Artikeln von 7 Webportalen der deutschsprachigen, verschwörungstheoretischen Szene ausgewertet. Aus diesem werden Features extrahiert, die anschließend genutzt werden um ein Modell zu trainieren, welches diese Artikel von (wissenschafts-)journalistischen Artikeln unterscheiden kann. Dabei erfolgt ein Rückgriff auf in der bestehenden, überwiegend qualitativen Forschung gemachte Erkenntnisse und Beobachtung. Auch werden Methoden aus der natürlichen Sprachverarbeitung verwendet um Features aus Wortfrequenzen und Part-of-Speech Tags zu erstellen.

Das erstellte Modell kann präzise verschwörungstheoretische Artikel identifizieren und von denen im Vergleichskorpus unterscheiden. Die dabei gemachten Beobachtungen können dabei helfen in der qualitativen Forschung gemachte Beobachtungen quantitativ zu validieren. Auch könnte die hier vorgestellte Methode hilfreich sein um verschwörungstheoretische Inhalte in größeren Datenmengen zu identifizieren.

2 Theorie

2.1 Stand der Forschung

Das Erscheinen von Richard Hofstadter und Wilentz (1964/2008) Essay *The Paranoid Style in American Politics* wird heute üblicherweise als der Beginn der wissenschaftlichen Auseinandersetzung mit Verschwörungstheorien betrachtet.

¹Siehe z.B. Coady (2006) oder auch Uscinski und Parent (2014).

²Siehe etwa Samory und Mitra (2018) oder Shahsavari et al. (2020).

In seiner Arbeit setzt Hofstadter sich mit dem Einfluss von Verschwörungstheorien und Verschwörungsrhetorik (dem *Paranoid Style*) auf die US Amerikanische Politik auseinander. Ein Beispiel auf das Hofstadter eingeht sind etwa die Schriften von Senator McCarthy und die nach ihm benannte Praxis des McCarthyismus.

Das weitere Forschungsinteresse im restlichen 20. Jahrhundert war eher gering, die verbleibende Forschung fiel meist in eine von zwei Kategorien, wie Sunstein und Vermeule (2008) passend zusammenfassen:

The academic literature on conspiracy theories is thin, and most of it falls into one of two classes: (1) work by analytic philosophers, especially in epistemology and the philosophy of science, that asks what counts as a “conspiracy theory” and whether such theories are methodologically suspect; (2) a smattering of work in sociology and Freudian psychology on the causes of conspiracy theorizing. (Sunstein & Vermeule, 2008, S. 2)

In der jüngeren Vergangenheit haben vor allem drei Entwicklungen Verschwörungstheorien und damit auch der Forschung in diesem Bereich vorschub geleistet.

Dies war zum einen die Verbreitung des Internets um die Jahrtausendwende. Die Mehrheit der Autor:innen sieht darin ein erhebliches Hilfsmittel für die Verbreitung von Verschwörungstheorien, so schreibt etwa Stano (2020): “[...] The Internet, and in particular social networks, have proved fundamental to the spread and development of such [conspiracy] theories” (Stano, 2020, S. 492).³

In der jüngsten Vergangenheit haben insbesondere die Wahl Donald Trumps zum US-Präsidenten im Jahr 2016 und die aktuelle Covid-19 Pandemie Verschwörungstheorien in das allgemeine Bewusstsein gerückt. Während die wissenschaftliche Aufarbeitung, insbesondere der mit der Covid-19 Pandemie einhergehenden Welle an Verschwörungstheorien, noch am Anfang steht, sind dennoch bereits erste Arbeiten entstanden, die neue Ansätze zeigen, wie etwa Shabsavari et al. (2020).

Eines der Kernprobleme in der Arbeit mit Verschwörungstheorien, ist es diese verlässlich als solche zu identifizieren. Dies hat nicht nur eine fundamentale Rolle für theoretische Arbeiten, sondern ist auch von immenser praktischer Relevanz, in einer Zeit in der in sozialen Medien die Grenzen zwischen Nachrichten und Verschwörungstheorien immer schwerer auszumachen sind.

In der Vergangenheit war die wissenschaftliche Beantwortung dieser Frage meist eng mit der Frage danach verbunden, was eine Verschwörungstheorie ausmacht und anhand welcher Kriterien sich dies bestimmen lässt. Eine beispielhafte Arbeit ist hier Uscinski und Parent (2014). Die Autor:innen nutzen darin sechs verschiedene Tests um Verschwörungstheorien zu identifizieren. Als Beispiel seien Occam’s Razor, also die Frage nach der einfachsten Erklärung, und die Falsifizierbarkeit einer Theorie genannt. In der Arbeit wird mit diesen

³Es gilt jedoch anzumerken, dass auch die gegenteilige Auffassung in der Literatur vertreten ist, etwa bei Clarke (2007).

Kriterien und der Methode der Inhaltsanalyse sowie einer Vielzahl von studentischen Kodierer:innen ein Korpus aus Briefen an US-Redaktionen ausgewertet (Uscinski & Parent, 2014, 54ff). Während die Autor:innen überzeugend dafür argumentieren, alle der aufgestellten Kriterien für eine Klassifizierung einzusetzen (Uscinski & Parent, 2014, 52f), ist keines der eingeführten Kriterien, zumindest mit den aktuellen technischen Möglichkeiten, wirklich geeignet um automatisierte Klassifikationen vorzunehmen.

In der jüngeren Vergangenheit sind demgegenüber erste Arbeiten entstanden, die sich mit automatisierten Verfahren rund um Verschwörungstheorien beschäftigen. Eine solche Arbeit ist die von Samory und Mitra (2018), in der die Autor:innen automatisiert Triplets bestehend aus *agent*, *action* und *target* aus verschwörungstheoretischen Online Beiträgen extrahieren. Als Datengrundlage nutzen sie dazu Beiträge aus dem Reddit Subforum r/conspiracy. Diese werten sie mittels einer NLP-Pipeline aus, die unter anderem Topic Modeling, Dependency Parser und Wortvektoren nutzt, aber auch vereinzelt menschliche Expertise einbringt (Samory & Mitra, 2018, 6ff). Der Ansatz stellt einen wichtigen Beitrag zu einer vollständig automatisierten Analyse dar, ist aber auch verhältnismäßig komplex und benötigt Eingriffe durch menschliche Expert:innen.

Eine weitere Arbeit die sich mit automatisierten Auswertungen beschäftigt, ist die von Shahsavari et al. (2020). Die Autor:innen werten dabei Online-Diskussionen in Foren wie Reddit und 4Chan aus, die sich mit der Covid-19 Pandemie beschäftigen. Die Auswahl der Beiträge mit verschwörungstheoretischem Inhalt, wurde dabei von einem Expert:innengremium getroffen (Shahsavari et al., 2020, 284f). Es werden mittels Methoden aus der natürlichen Sprachverarbeitung narrative Netzwerke ausgewertet und so Verbindungen zu den Inhalten von Nachrichten und den aktuell populärsten Verschwörungstheorien aufgezeigt.

Näher an der Problemstellung dieser Arbeit ist die Methode die von Pott-hast et al. (2018) vorgestellt wird. Darin stellen die Autor:innen Methodik vor um mittels Verfahren des maschinellen Lernens und stilistischen Features Fake-News und politisch extreme Artikel zu identifizieren. Sie greifen dabei auf einen von professionellen Journalist:innen kategorisierten Korpus zurück und können extremistische Artikel mittels dieser Merkmale von anderen unterscheiden. Da Verschwörungstheorien eine Schnittmenge mit Fake-News und insbesondere der rechtsextremen Szene haben (vgl. Römer & Stumpf, 2019), könnte dieser Ansatz auch für die identifizierung verschwörungstheoretischer Inhalte relevant sein.

2.2 Merkmale von verschwörungstheoretischen Texten

Um Verschwörungstheorien zuverlässig zu identifizieren, soll zunächst ein Blick in die Literatur erfolgen und in die in der qualitativen und quantitativen Forschung gemachten Beobachtungen zu sprachlichen Merkmalen von Verschwörungstheorien. Eine erschöpfende Auflistung aller in der Literatur zu findenden sprachlichen Merkmale von Verschwörungstheorien ist insbesondere mit dem erstarkten Forschungsinteresse in der jüngsten Vergangenheit kaum möglich noch wäre es zielführend. Es sollen daher hier vor allem solche Merkmale genannt werden, die sich für eine quantitative, automatisierte Erfassung eignen und in

dieser Arbeit angewandt werden.

Eine in der Literatur häufig gemachte Beobachtung ist die Emotionalität der Argumentation in Verschwörungstheorien (Vgl. Miller, 2002, S. 10). Dieses Argument findet sich auch bei Butter (2018, 93ff), der noch spezifischer darauf eingeht, dass es vor allem die vermeintlichen Verschwörer sind denen mit "metaphorisch aufgeladener, bisweilen apokalyptischer Sprache ausschließlich negative Eigenschaften zugeschrieben [werden]" (Butter, 2018, 93f).

Während die emotionalität spezifischer Aspekte der Argumentation eine häufig gemachte Beobachtung ist, steht dazu im Gegensatz die Beobachtung über den allgemeinen Stil der Verschwörungstheoretiker:innen. So spricht ebenfalls Butter (2018, S. 61) davon, dass sich Verschwörungstheoretiker traditionell um eine seriöse Darstellung bemühen und sich der verwendete Stil an dem der Wissenschaft anlehne. Diese Feststellung findet sich bereits bei Hofstadter und Wilentz (1964/2008):

The higher paranoid scholarship is nothing if not coherent—in fact the paranoid mind is far more coherent than the real world. It is nothing if not scholarly in technique. McCarthy’s 96-page pamphlet, McCarthyism, contains no less than 313 footnote references, and Mr. Welch’s incredible assault on Eisenhower, The Politician, has one hundred pages of bibliography and notes. (Hofstadter & Wilentz, 1964/2008, S. 37)

Seit der Erstveröffentlichung von Hofstadters Aufsatz hat vor allem die Existenz des Internets zu Veränderungen in der Erzählart von Verschwörungstheorien geführt. Die Obsession an Belegen und Referenzen existiert zwar weiterhin, hat sich aber insoweit verändert, als das nicht mehr unbedingt Fussnoten das Mittel der Wahl sind, vielmehr wird sich der Mittel des Internets bedient, Inhalte werden direkt eingebunden oder verlinkt. So schreibt etwa Soukup (2008) der sich spezifisch mit Verschwörungstheorien um den 11. September im Internet befasst, diese müssten als "digital, hypertextual, and multimedial experience" (Soukup, 2008, S. 10) verstanden werden.

Erst in der jüngeren Vergangenheit finden sich Arbeiten die sich spezifisch mit den linguistischen Merkmalen von Verschwörungstheorien auseinandersetzen. Schäfer (2018) etwa analysiert einen Korpus aus Kommentaren zu Verschwörungstheoretischen YouTube-Videos. Sie stellt dabei unter anderem eine gehäufte Verwendung von Ironie und *scare quotes* fest, wenn es darum geht die Gegenseite negativ darzustellen (Schäfer, 2018, S. 235). Ähnlich der bereits erörterten Literatur findet die Autorin auch eine hohe Dichte von Belegen und Referenzen die die Kompetenzen der Verschwörungstheoretiker:innen unterstreichen sollen, stellt aber zusätzlich heraus, dass diese sich häufig in Form von Nennungen von Namen und Zahlenangaben ausdrücken (Schäfer, 2018, S. 234).

Eine ähnliche Beobachtung macht auch Filatkina (2018), die feststellt:

In einem Beitrag oft mehrfach angeführte Verweise auf offizielle Zahlen [...] bzw. das wörtliche Zitieren der VertreterInnen der

Wissenschaft und Politik sollen die Glaubwürdigkeit herstellen. (Filatkina, 2018, S. 208)

Ebenfalls nennt die Autorin eine übliche Konstruktion in der durch offene und geschlossene Fragen die Aufmerksamkeit der Leser:innen auf bestimmte Aspekte gelenkt werden soll (Filatkina, 2018, S. 205). Als letztes stilistisches Merkmal sei hier noch die häufige Nutzung von Negationswörtern genannt, wie sie etwa von Römer und Stumpf (2019, S. 149) beobachtet wird.

2.3 Textklassifizierung

Die Aufgabe das so gesammelte Wissen zu operationalisieren fällt in den Bereich der Textklassifizierung bzw. -kategorisierung. Dieses Feld geht auf Maron (1961) zurück, der Dokumente aufgrund sogenannter *clue words* und einem probabilistischen Ansatz Kategorien zuordnete und dabei schon viele Methoden die in modernen Verfahren genutzt werden, wie etwa das Filtern von Stopwörtern, implementierte.

Das Problem lässt sich grundlegend Formalisieren als die Suche nach einer Funktion F mit $F : D \times C \rightarrow 0, 1$. Es wird also eine Funktion gesucht, die alle Dokumente D jeder Kategorie C entweder positiv oder negativ zuordnet (vgl. Feldman & Sanger, 2006, 66f). Im hier vorliegenden Fall ist nur eine Kategorie gesucht, so dass es sich um ein binäres Klassifizierungsproblem handelt. Da komplette Textdateien für Computer nur sehr bedingt zu Verarbeiten sind, werden die Dokumente D durch sogenannte Feature Vektoren repräsentiert.

Die Erstellung dieser Vektoren ist der wesentliche Schritt um eine gute Klassifizierungsleistung zu erhalten. Vermutlich am häufigsten als Repräsentation genutzt werden Wortfrequenzen, meist in der Form term-frequency inverse-document-frequency (*tf-idf*). Dieser Wert soll Ausdrücken wie wichtig ein Wort für ein spezifisches Dokument ist, indem er die Häufigkeit des Auftauchens in einem Dokument ins Verhältnis zur Häufigkeit des Vorkommens in allen Dokumenten setzt. Entscheidend ist auch welche Wörter als Features genutzt werden, während es häufig ausreicht die Häufigsten n -Prozent der Wörter⁴ zu nutzen (Feldman & Sanger, 2006, S. 68), gibt es hier eine Vielzahl von komplexeren und spezialisierten Verfahren. Für einen Überblick über solche sei auf die entsprechende Literatur verwiesen, etwa Yang und Pedersen (1997).

Neben reinen Wortfrequenzen sind für diese Arbeit auch Part-of-Speech Tags relevant. Diese Annotationen der Wortarten werden zum Teil für Textklassifikation benutzt, werden aber insbesondere für stilistische Fragestellungen häufig genutzt (vgl. Jimenez et al., 2020). Da in dieser Arbeit der, in Anlehnung an Hoffstadter, Paranoide Stil Untersuchungsgegenstand ist, könnten diese also einen potentiell wichtigen Beitrag zur Klassifizierungsgenauigkeit leisten.

⁴Nachdem Wörter ohne Informationswert wie etwa Stopwörter gefiltert wurden.

3 Vorgehen

Als Grundlage dieser Arbeit dient ein Korpus aus Texten von deutschsprachigen Internetportalen, die der verschwörungstheoretischen, bzw. Truther Szene zuzuordnen sind. Diese wurde im Verlauf der letzten Jahre über Webcrawling zusammengestellt.

Um einen strukturierte Analyse mit Methoden des maschinellen Lernens zu ermöglichen wird ein Vergleichskorpus mit Artikeln die dem Untersuchungskorpus von Form und Thematik möglich ähnlich gelagert sind, ohne aber dessen verschwörungstheoretischen Aspekte zu teilen.

Anschließend werden die Korpora gereinigt und mit Verfahren aus der natürlichen Sprachverarbeitung Features extrahiert auf denen Klassifizierungsverfahren arbeiten können. Zum Einsatz kommen dabei sowohl Features die klassischerweise für Textklassifikation herangezogen werden wie Wortfrequenzen, Features die typischerweise eher für stilometrische Analysen verwendet werden wie Part-of-Speech Tags, sowie auf der vorhandenen Literatur zu Verschwörungstheorien basierende Features. Die Implementierung aller Funktionen dieser Arbeit erfolgt in der Programmiersprache *R* (R Core Team, 2021)

Mit den so erstellten Daten soll abschließend ein Modell trainiert werden, dass zwischen den verschwörungstheoretischen Texten und dem Vergleichskorpus unterscheiden kann. Um die Effektivität der auf der Literatur basierten Features einschätzen zu können wird ein Datensatz erstellt der nur diese Features enthält, sowie ein weiterer Datensatz der alle Features enthält. Für jeden dieser Datensätze wird zunächst ein einfacher Entscheidungsbaum als Baseline Modell trainiert. Für das endgültige Modell kommt die *LightGBM* Bibliothek (Ke, 2019), eine Implementierung von Gradient Boosting Decision Trees (2002).

Dieses Verfahren bietet neben der typischerweise relativ guten Präzision den Vorteil, dass die Ergebnisse, ähnlich einfachen Entscheidungsbäumen, verhältnismäßig gut interpretierbar sind. Dies soll es Erlauben den Einfluss einzelner Features auf die Vorhersagegenauigkeit zu bestimmen, insbesondere auch der aus der Literatur abgeleiteten Features.

4 Umsetzung

4.1 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 das R-Paket *rvest* (Wickham, 2020) und extrahieren mittels für jede Seite separat erstellten CSS-Selektoren und XPATH-Queries den jeweiligen Artikeltext. Während dieses Prozesses wurden auch 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 jedem Artikel

das (angegebene) Veröffentlichungsdatum, der Artikeltitel sowie die Rubrik auf der Website, so diese Angabe vorhanden war.

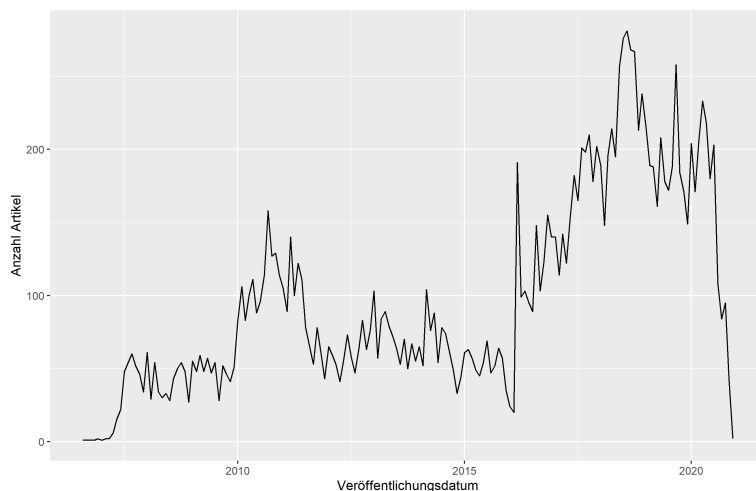


Abbildung 1: Anzahl Artikel nach Veröffentlichungsdatum (nach Monat gruppiert)

Der so erstellte Korpus umfasst insgesamt 16836 Texte und deckt beginnend in 2006 einen Zeitraum von 14 Jahren ab. Die zeitliche Verteilung der Artikel im Korpus ist Grafik 1 zu entnehmen, wie dort ersichtlich ist steigt beginnend in 2016 die Zahl der Artikel deutlich an. Dies ist zum einen damit zu erklären, dass einzelne der gecrawlten Websites ihre Veröffentlichungsfrequenz zu dieser Zeit erhöht haben, überwiegend ist diese Entwicklung aber darauf zurückzuführen, dass insbesondere das Angebot von *Watergate.tv* erst im Jahr 2016 seinen Betrieb aufgenommen hat und eine hohe Veröffentlichungsfrequenz aufweist. Genauere Informationen zu den von den einzelnen Angeboten abgedeckten Zeiträumen sowie zu Artikelzahl und Länge finden sich in Tabelle 1.

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

Auch eine inhaltliche Betrachtung zeigt deutliche Unterschiede zwischen den einzelnen Quellen. So ist *Alles Schall und Rauch (ASuR)* als ältestes vertretenes Angebot auch einer der "klassischsten" Vertreter der Verschwörungswbsites. Hier werden die Besucher:innen direkt auf der Startseite mit Fragen konfrontiert wie ‚Was geschau wirklich am 11. September?‘ oder ‚Was passiert tatsächlich mit dem Klima?‘ (Freeman, 2021). Hier lässt sich bereits die von Filatkina (2018, S. 205) beobachtete Konstruktion des offenen Fragen finden, die die Aufmerksamkeit der Leser:innen auf sich ziehen soll.

	Von	Bis	Mittlere Zei- chen- zahl	Anzahl Artikel	Artikel/ Monat
Alles Schall und Rauch	23.08.06	02.08.20	5000	5349	32.0
conrebbi	05.09.12	20.10.14	5610	24	1.0
deutschlandpranger	29.10.16	30.12.20	7527	118	2.4
fm-tv	31.07.08	02.11.18	9158	97	0.8
hinterderfichte	16.01.10	31.05.18	4221	1083	10.8
recentr	03.08.07	05.08.20	5071	4762	30.5
Watergate.tv	20.05.16	16.11.20	2810	5403	101.9

Tabelle 1: Kennzahlen der einzelnen Quellen des Korpus

Die Seite bedient sich klassischer Verschwörungstheorien und wird im allgemeinen der Truther Szene zugeordnet („Alles Schall und Rauch“, 2020). Sie hat laut Eigenaussage im Schnitt über 50.000 Zugriffe täglich.⁵

Inhaltlich finden sich sowohl Essay-artige Artikel zu Themen wie 9/11, den Bilderberger-Treffen oder dem Klimawandel auf der Seite, aber 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. Hier lässt sich Bereits die von Soukup (2008) beobachtete Multimedialität von Verschwörungstheorien beobachten.

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

Von den bisherigen Quellen abheben tun sich dagegen die Angebote *fm-tv* sowie *Deutschlandpranger*. Während sich beim *Deutschlandpranger* etwa inhaltlich recht klassische Themen wie die Leugnung des Klimawandels (vgl. 2020) finden, sind die dazugehörigen Text überdurchschnittlich lang und voller Werbung für Bücher zu z.T. komplett unabhängigen Themen. Insbesondere aber sind die Texte selbst z.T. sehr wirr, die (sehr ausführlichen) AGB der Website enthalten etwa einen Abschnitt zu Strafzahlungen bei "Übersenden eines Statements anstatt einer echten Rechnung (True Bill) des wahren Haftungsgläubigers" (2016) und weitere ähnlich obskure Paragraphen.

Die verbleibenden beiden Quellen im Korpus *recentr* und *Watergate.tv* sind beide der rechten truther Szene zuzuordnen und verbreiten Verschwörungstheo-

⁵Der Besucherzähler der Website ist inzwischen nicht mehr funktional, Angabe übernommen aus Anonymous (2020)

rien in diese Richtung. Beide zeichnet ein eher journalistischer Stil aus, es werden kurze bis mittellange Meldungen veröffentlicht, häufig begleitet von eigenen oder fremden Videos. *Recentr* wurde 2006 als deutscher Ableger von *Inforwars.com* gegründet und bedient sich noch heute einem ähnlichen Konzept. Ähnlich dem Vorbild wird massiv Werbung für den eigenen Shop gemacht in dem diverse Produkte zur Vorbereitung auf apokalyptische Szenarien sowie fragwürdige Nahrungsergänzungsmittel verkauft werden. Inhaltlich sticht vor allem heraus, dass sich die Seite (bzw. der Hauptautor Alexander Benesch) z.T. sehr explizit von anderen Teilen der rechten/verschwörungstheoretischen Szene abgrenzt.⁶ So kommt es durchaus vor, dass in einem Artikel zunächst von der implausiblen Sichtweise, verschiedenste Regierungen hätten eine Fake-Pandemie orchestriert (Benesch, 2021a) gesprochen wird, nur um im nächsten Satz zu erklären:

In Wirklichkeit sind die Methoden der globalen Bevölkerungsreduktion fast ausschließlich legal, simpel und werden schrittweise angewandt [...]. Das Konzept ist so entworfen, um mit möglichst wenig Zwang auszukommen [...]. (Benesch, 2021a)

Die Website *Watergate.tv* (inzwischen Teil von *NEOPresse.com*) ist ein Nachrichten und Videoportal, das sich verschwörungstheoretischer Narrative bedient. Auf dem Portal finden sich klassische Themen wie Bilderberger-Treffen (vgl. Redaktion, 2018) aber auch Fake-News und Berichterstattung mit starker prorussischen Tendenz. In die Öffentlichkeit rückte das Portal als sich das Neo-Magazin-Royale wegen seiner Mitarbeit bei *Watergate.tv* von Hans Meiser trennte (Siehe z.B. 2017).

Insgesamt deckt der hier verwendete Korpus einen signifikanten Teil der eher klassischen deutschsprachigen Truthther Szene ab. Es muss allerdings angemerkt werden, dass insbesondere in den letzten Jahren eine Bewegung in der Szene Richtung Social Media zu Portalen wie Facebook und Telegram stattfand. Diese zu Untersuchen ist allerdings nicht Teil der Fragestellung dieser Arbeit.

Eventuell noch ne Zitation reinwamsen.

4.2 Vergleichskorpus

Um einen möglichst guten Vergleich zwischen den Texten im Korpus und dem Vergleichskorpus zu ermöglichen war es das Ziel Medien zu finden die in Form und inhaltlicher Ausrichtung dem Korpus möglichst ähnlich sind, ohne aber deren verschwörungstheretische Aspekte zu teilen. Da insbesondere die letzten beiden besprochenen Angebote sich inhaltlich sehr an journalistische Angebote anlehnen, lag es nah ebensolche heranzuziehen. Aus Gründen der Vollständigkeit (die Angebote sollten möglichst den kompletten Zeitraum des Korpus abdecken) sowie der Zugänglichkeit wurden die Angebote von *Spiegel Online* sowie der *Frankfurter Rundschau* ausgewählt. Es wurde für beide Onlineangebote zunächst ein Index aller Artikel die in den selben Zeitraum wie der Korpus fallen

⁶So grenzt sich das Portal in Beiträgen inzwischen auch von *Inforwars* Gründer Alex Jones ab (2018) oder spekuliert darüber, ob die Alternative für Deutschland vom britischen Geheimdienst gegründet wurde (Benesch, 2021b)

	Von	Bis	Mittlere Zei- chen- zahl	Anzahl Artikel	Artikel/ Monat
Frankfurter Rundschau	03.08.06	04.08.20	3670	9959	59.3
scienceblogs	14.02.07	04.08.20	3261	8765	54.4
scilogs	29.12.00	03.08.20	5206	9533	40.6
Spiegel Online	01.08.06	07.01.20	4687	8665	53.8

Tabelle 2: Kennzahlen der einzelnen Quellen des Vergleichskorporuses

erstellt und aus diesem zufällig eine Auswahl von 10000 Artikeln pro Anbieter gezogen.

Da die restlichen im Korpus enthaltenen Seiten alle mehr oder weniger in der Form eines Blogs gestaltet sind, sollten Angebote vergleichbarer Form herangezogen werden. Um auch dem in der Literatur häufig genannten wissenschaftlichen Stil von verschwörungstheoretischen Texten Rechnung zu tragen wurde eine Reihe von Wissenschaftsblogs als zweite Komponente im Vergleichskorpus gewählt. Dafür wurden von den beiden Plattformen *scienceblogs* und *scilogs*⁷ ähnlich den journalistischen Angeboten aus dem gesamten Index der Einträge eine Auswahl von jeweils 10000 Artikeln ausgewählt. Genauere Kennwerte zum Vergleichskorpus sind Tabelle 2 zu entnehmen.

4.3 Datenvorverarbeitung

Zunächst wurden einige Süberungsverfahren auf den Korpus angewendet, um zu verhindern, dass fehlerhafte oder irrelevante Daten mit für die Feature Erstellung herangezogen werden.

Es wurde zunächst Unicode Normalisierung für alle Texte durchgeführt, anschließend wurde mittels des *Google Compact Language Detector 3* (Ooms, 2021) die Sprache aller Text bestimmt und alle überwiegend Fremdsprachigen Artikel wurden aus dem Korpus entfernt. Da im Korpus an einigen Stellen nicht zum Artikeltext gehörige Komponenten aus vielen Sonderzeichen vorhanden waren, wurden alle Wörter entfernt, die nur aus Zeichen bestanden, die unter 1000 Mal im gesamten Korpus vorhanden waren. Weiterhin wurden noch kleinere Korrekturen vorgenommen, wie das Zusammenfassen multipler, aufeinanderfolgender Leerzeichen und Unterstriche (die gerne als Trennlinien in den Text benutzt wurden).

Abschließend wurden noch alle Text unter 100 Zeichen Länge aus dem Korpus entfernt, da diese für die Klassifizierungsaufgabe nur wenig Informationen enthalten können. Der finale Korpus umfasste 53758 Texte, mit einem Anteil der positiven Klasse (sprich von verschwörungstheoretischen Texten) von etwa

⁷Beide Plattformen sind Anbieter bei denen eine Vielzahl verschiedener Blogs zu finden sind.

32%.

4.4 Features

Zunächst wurden Features basierend auf den in der Literatur beobachteten linguistischen Eigenschaften von verschwörungstheoretischen Texten erstellt. Die etwa von Miller (2002, S. 10) festgestellte Emotionalität wird mit der Methode der Sentimentanalyse bearbeitet. Dafür wurde das Sentimentwörterbuch *SentiWS* (Remus et al., 2010) genutzt. Dieses umfasst positive und negative Sentimentwerte für über 3000 Grundformen deutscher Wörter. Während komplexere Ansätze wie etwa der von Neethu und Rajasree (2013) eine höhere Präzision versprechen,⁸ scheint die genutzte Methode für den vorliegenden Kontext ausreichende Ergebnisse zu liefern. Es wurde sowohl die Summe der positiven und negativen Sentimente erfasst, also auch die Summe sowie der Betrag nach Text insgesamt. Hiermit soll insbesondere auf solche Fälle eingegangen werden, in denen ein Text zwar viele emotional behaftete Wörter verwendet, sich die Sentimentsumme aber durch einen Wechsel an positiven und negativen Wörtern ausgleicht.

Als weiteres in der Literatur anzutreffendes Merkmal soll auch die Menge an Zitationen betrachtet werden. Um diese zu operationalisieren wurden zum einen die Anzahl der Anführungszeichen im Text gezählt. Eine genauere Betrachtung der Verwendung von Fremdbelegen von denen etwa Schäfer (2018, S. 235) spricht, wird ermöglicht indem als zusätzliche Features erfasst werden:

1. Die Zahl der äußersten Zitate⁹
2. Die durchschnittliche Länge der Zitate
3. Der Anteil von zitiertem Text am gesamten Artikeltext

Die Menge an *scare quotes* wird ebenfalls erfasst, wobei ein *scare quote* für diese Arbeit definiert wird als ein Zitat in dem keine Leerzeichen (oder andere Formen von whitespaces) enthalten sind.

Die Menge der Fragen, wie sie etwa Filatkina (2018, S. 205) als Merkmal von verschwörungstheoretischen Texten feststellt, wird durch ein simples Zählen der im Text vorkommenden Fragezeichen abgebildet. Ähnlich lässt sich auch die Annahme über die Häufung von Zahlenangaben wie sie sich bei Schäfer (2018, S. 234) findet, durch einfaches Zählen von Zahlengruppen operationalisieren. Um die bei Römer und Stumpf (2019, S. 149) gemachte Beobachtung der häufigen Verwendung von Negationswörtern abzubilden, wurde händisch eine Liste von 15 Negationswörtern erstellt und deren Häufigkeit in den Texten gemessen.

Die Multimedialität und Verknüpftheit von Verschwörungstheoretischen Online Texten wie sie Soukup (2008, S. 10) anspricht soll mit einer Reihe von Informationen die aus den dem Korpus zugrunde liegenden HTML Dateien extrahiert wurden abgebildet werden. Um die Multimedialität zu erfassen wurde

⁸So ist der hier vorgestellte Ansatz nicht in der Lage korrekt mit Negationen umzugehen.

⁹Also Zitate die nicht Teil anderer Zitate sind

Feature	Korpus		Vergleichskorpus		W	p
	\bar{X}	σ	\bar{X}	σ		
Textlänge	4309	4429	4208	4186	3.01e8	< 0.01
Tweets	0.0006	0.0058	0.0004	0.013	3.13e8	< 0.01
Youtube-Videos	0.016	0.088	0.004	0.037	3.45e8	< 0.01
Interne Links	0.0094	0.066	0.077	0.274	2.31e8	< 0.01
Externe Links	0.081	0.2	0.12	0.434	3.16e8	< 0.01
Negationen	0.158	0.11	0.151	0.109	3.23e8	< 0.01
Anteil Zitation	9.351	18.062	7.373	13.984	2.91e8	< 0.01
Sentimente Betrag	8.778	9.615	7.649	8.039	3.15e8	< 0.01

Tabelle 3: Statistische Kennzahlen zu ausgewählten Features

die Anzahl der eingebundenen Bilder, Tweets, Youtube-Videos sowie sonstiger eingebetteter Elemente im Artikel gezählt. Für eine Messung der Verknüpftheit wurde die Anzahl von in den Artikeln enthaltenen Links, aufgeteilt nach internen und externen Zielen, erfasst. Ebenso wurde als stilistisches Merkmal die Menge an fettgedruckten und kursiven Text erfasst.

Um weitere Features aus dem Bereich der Textklassifikation zu nutzen, wurde der Corpus zunächst mithilfe von *spaCy* (Honnibal et al., 2020) annotiert um Lemmata und Part-of-Speech Tags zu generieren. Es wurden anschließend die Anzahl jedes Part-of-Speech Tags sowie der Anteil an großgeschriebenen Text und Sonderzeichen als Feature übernommen. Ebenso wurde die Wortdichte für jeden Text bestimmt, also die Anzahl an Wörtern im Verhältnis zur Textlänge.

Zur Ermittlung von Wortfrequenzen wurden zunächst Stopwörter aus dem Corpus entfernt¹⁰, Wörter auf ihre Lemmata reduziert, einige Sonderfälle aussortiert¹¹ und anschließend die 400 häufigsten Wörter als Features übernommen.¹² Alle Features abgesehen von den sowieso als *tf-idf* oder als relative Häufigkeit vorliegenden wurden mit der Textlänge skaliert.

4.5 Modelerstellung

Zunächst wurde ein Datensatz erstellt der alle Beobachtungen, aber nur solche Features enthielt die aus der Literatur zu Verschwörungstheorien abgeleitet wurden. Auf diesem Datensatz wurde zunächst ein einfacher Entscheidungsbaum aus dem Paket *rpart* (Therneau & Atkinson, 2019) als Baseline Klassifizierer trainiert. Hierfür wurde ein Train/Test Split im Verhältnis 2/1 auf den Da-

¹⁰Zum Einsatz kam die Stopwort Liste aus dem *tm* Paket (Feinerer et al., 2008)

¹¹So wurde beispielsweise das Wort *dpa* nicht übernommen, da es praktisch exklusiv in Artikeln der Frankfurter Rundschau und von Spiegel Online vorkommt.

¹²Es ist durchaus üblich mehr Wörter als Features zu übernehmen (siehe etwa 2006, S. 68), da dies aber die Dimensionalität der Daten und damit die Komplexität der Klassifizierung deutlich erhöhen würde und die Auswahl von 400 Wörtern bereits zufriedenstellende Ergebnisse liefert wurde hier darauf verzichtet.

Modell	Accuracy	Sensitivity	Specificity	AUC
dtree	0.7534	0.7761	0.7428	0.829445
lightGBM	0.8745	0.8700	0.8766	.0.950817

Tabelle 4: Ergebnisse der Klassifizierer nur auf literaturbasierten Features

Modell	Accuracy	Sensitivity	Specificity	AUC
dtree	0.8668	0.7845	0.9052	0.881973
lightGBM ¹³	0.9784	0.9770	0.9791	N/A

Tabelle 5: Ergebnisse der Klassifizierer auf dem vollständigen Datensatz

ten durchgeführt. Das auf den Trainingsdaten trainierte Modell erreichte eine Präzision von 75.3%.

Mit dem gleichen Daten wurde anschließend ein Modell mit der *LightGBM* Bibliothek(Ke, 2019) trainiert. Da Gradienten-Boosting Verfahren schnell zu Overfitting neigen wurde dafür zunächst mittels Kreuzvalidierung auf den Trainingsdaten die optimale Iterationstiefe bestimmt. Anschließend wurde das Modell mit einer Lernrate von 0.1 und einer maximalen Blattzahl von 50 trainiert. Das so trainierte Modell erreichte eine Präzision von 87.4%.

In Tabelle 4 sind die Leistungen der beiden Modelle auf den Testdaten zusammengefasst. Wie zu erkennen ist, ist selbst der einfache Entscheidungsbaum deutlich Präziser als es ein bloßes Raten im Klassenverhältnis wäre ($p < 0.01$). Auch ist ersichtlich, dass das komplexere *LightGBM* Modell eine in allen Bereichen deutlich bessere Klassifizierungsleistung erbringt.

Im weiteren wurde ein Modell mit allen vorhandenen Features erstellt. Hierfür wurde ebenfalls zunächst ein Baseline Entscheidungsbaum trainiert. Dieser erreichte eine Präzision von 87.7% und konnte damit den Entscheidungsbaum der auf dem kleineren Korpus trainiert wurde in Genauigkeit deutlich übertreffen. Interessant ist vor allem die verhältnismäßig geringe Sensitivität.

Bevor das finale Modell trainiert wurde, wurde zunächst mittels eines Tune-grids eine Optimierung der Hyperparameter der Lernrate sowie der Anzahl der äußersten Blätter durchgeführt. Getestete Werte waren dabei für die Lernrate 0.01, 0.05, 0.10, 0.20 und für die Blätterzahl 5, 50, 127 (Der Standardwert von lightGBM), 200. Das beste Ergebnis wurde mit einer Lernrate von 0.05 und einer Blätterzahl von 50 erzielt. Mit diesen Werten wurde erneut die optimale Iterationstiefe mittels Kreuzvalidierung bestimmt. Das mit diesen Parametern erstellte, finale Modell wurde schließlich mittels einer 10-Fold Kreuzvalidierung validiert und erreichte insgesamt eine Präzision von 98.14%.

¹³10-Fold Kreuzvalidiert

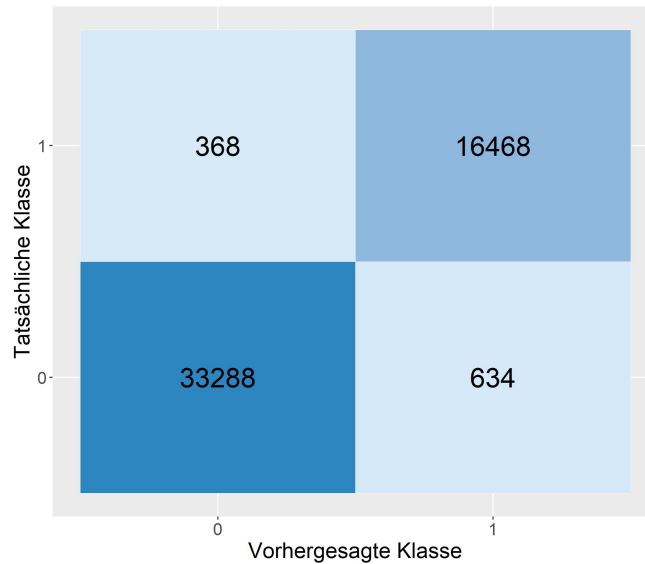


Abbildung 2: Confusion-Matrix der 10-Fold Kreuzvalidierung des finalen Modells

5 Diskussion

Das kleinere, ausschließlich mit literaturbasierten Features trainierte Modell erreichte mit 87% bereits eine relativ hohe Präzision. Dies kann dahingehend interpretiert werden, dass die genutzten Features grundsätzlich geeignet sind um verschwörungstheoretische Texte von anderen zu unterscheiden. Die verwendete *lightGBM* Software erlaubt es die Bedeutung einzelner Features für das erstellte Modell abzuschätzen. Das Features mit dem höchsten Informationszuwachs (Gain) für das Modell ist die Anzahl der internen Links (Gain = 0.12). Die Anzahl der externen Links hingegen bot einen deutlich geringeren Gain (0.076) für das Modell. Dies ist insofern bemerkenswert, als dass die in der Literatur gemachten Beobachtungen wie etwa die von Soukup (2008, S. 10) hätten vermuten lassen, dass die Verknüpfung mit externen Quellen ein wichtigeres Merkmal ist als die mit internen.

Die nächstwichtigen Features sind die Anzahl der eingebundenen Bilder (Gain = 0.1), die Summe der Sentimente (0.1) und der Anteil an zitierten Text (0.9). Interessant ist hier noch, dass anders als die Summe an Sentimentscores der Betrag der Sentimentscores für die Klassifizierung relativ unbedeutend ist (Gain = 0.018). Dies kann zum einen daran liegen, dass diese Werte nicht vollständig unabhängig voneinander sind und der Trainingsalgorithmus deshalb u.U. nur einen der Werte berücksichtigt. Zum anderen kann es aber auch ein Indiz sein, dass Verschwörungstheoretische Texte nicht notwendigerweise allgemein emotionaler sind, aber im Schnitt deutlich negativere Sentimente transportieren wie sich auch an den Durchschnittswerten der Sentimentsumme zeigt

($\bar{X} = -4.22$ gegenüber $\bar{X} = -1.35$).

Die meisten anderen Features lieferten dem Modell einen mittelgroßen Informationszuwachs, besonders wenig nützlich waren vor allem Features die mehr oder weniger mit anderen Features zusammenhängen (die Summe positiver bzw. negativer Sentimente, Zahl der direkten Zitate, etc.), die Menge an Negationen, an Zahlenangaben sowie allgemein an eingebetteten Medien (Twitter, Youtube, Sonstige). Eine Erklärung für letztere könnte dabei aber auch schlicht sein, dass herkömmliche Medien inzwischen solche Medien auch häufiger einbinden und in diesem stilistischen Merkmal sich den Verschwörungstheorien angenähert haben.

Es bleibt anzumerken, dass Verfahren des Maschinellen Lernens, wie hier Gradientenboosting, komplexe Systeme sind und selbst wenn wie hier der Einfluss einzelner Features auf das Modell abgeschätzt werden kann, die Interpretation solcher Ergebnisse immer schwierig und zu einem gewissen Teil spekulativ ist.

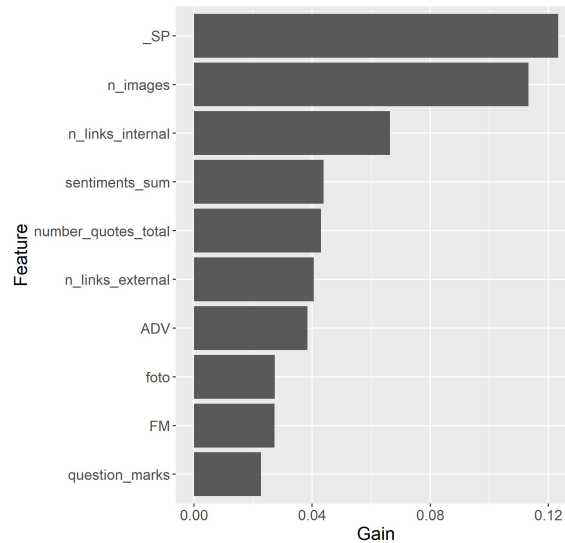


Abbildung 3: Top-10 Features nach Bedeutung für das Modell. Namen in Großbuchstaben = Part-of-Speech Tag; Namen in Kleinbuchstaben = tf-idf Worthäufigkeit; Benannte Features = auf Literatur basierende Statistiken

Das vollständige Modell erreichte mit fast 98% eine sehr gute Präzision. Es kann zuverlässig zwischen den Verschwörungstheoretischen Texten und dem Vergleichskorpus unterscheiden. Eine Betrachtung der wichtigsten Features aus Grafik 3 zeigt zunächst, dass viele der auch im ersten Modell wichtigen Features wieder von großer Bedeutung sind (Zahl der Bilder, Zahl der internen Links, Summe der Sentimentscores und Anteil von zitiertem Text). Einige der Wortfrequenzen mit dem meisten Einfluss auf die Klassifizierungen sind foto, krieg, merkel und regierung. Die meisten davon sind durchaus erwartbare Ergebnisse, die Verwendung von Worten aus solchen Themenbereichen ist in der Literatur

gut belegt, etwa bei Römer und Stumpf (2019, S. 150) oder Uscinski und Parent (2014, S. 25). Unter den wichtigen POS Tags sind Leerzeichen, Adverben, für fremdsprachigen Text (meist englischer Text, häufig zitiert) und für Coordinating Conjunctions (also Wörtern wie und, oder, aber, etc.(vgl. 2003)). Interessant ist auch, dass die Zahl der Fragezeichen im kleineren Modell zwar kaum relevant ist, im größeren Modell aber unter den 10 einflussreichsten Features.

Eine mögliche Erklärung für die überraschend gute Leistung des trainierten Modells ist, dass die relativ geringe Vielfalt insbesondere des Ausgangskorporates einen Teil zu der guten Leistung beigetragen hat. Auch wenn etwa für *Watergate.tv* mehrere Autor:innen schreiben, ist zu vermuten dass die totale Anzahl an Autor:innen im Korpus im niedrigen zweistelligen Bereich liegt. Da für die Modellerstellung mit POS-Tags auch eher stilistische Features enthalten sind, ist es durchaus möglich, dass ein Teil der Leistung des Modells weniger auf verschwörungstheoretischen Merkmalen beruht als auf dem Stil der individuellen Autor:innen.

Dieser Faktor dürfte in Kombination mit der relativ geringen Quellenzahl des Korpus als auch des Vergleichskorporates die Wahrscheinlichkeit, dass das hier trainierte Modell gut generalisiert deutlich verringern.

6 Fazit

Es wurde basierend auf einem Korpus aus Texten deutschsprachiger Verschwörungstheoretischer Texte und einem Vergleichskorpus aus (wissenschafts-)journalistischen Texten ein Modell trainiert, dass zuverlässig die verschwörungstheoretischen Texte identifizieren kann. In der Literatur gemachte Beobachtungen zur literarischen Form von Verschwörungstheorien wurden dabei aufgegriffen und als Features in das Modell eingebracht. Diese leisteten einen wichtigen Beitrag für die Funktion des trainierten Modells, es konnte so gezeigt werden, dass einige dieser qualitativen Beobachtungen auch für eine qualitative Analyse von verschwörungstheorien geeignet sind. Es wurden weiterhin Wortfrequenzen und Part-of-Speech Tags als zusätzliche Features in das Modell eingebracht. Das so entstandene Modell arbeitet deutlich genauer.

Der vorgestellte Ansatz konnte eine gute Leistung erbringen, ist dabei aber deutlich weniger komplex als die bisherigen Verfahren zur automatischen Verarbeitung von Verschwörungstheorien. Er ist dafür allerdings u.U. deutlich unschärfer, es wird keine spezifische Verschwörungstheoretische Aussage identifiziert wie dies etwa Samory und Mitra (2018) tut, vielmehr wird ein allgemeiner verschwörungstheoretischer Stil identifiziert.

Ein mögliches Problem des vorgestellten Verfahrens könnte auch die Generalisierbarkeit des trainierten Modells sein. Ein möglicher Ansatzpunkt um hier Verbesserungen herbeizuführen wäre die Zusammenstellung eines bedeutend größeren und diverseren Korpus, sowie auch eines entsprechend Vergleichskorporates vonnöten. Auch bedarf es deutlich mehr an quantitativer Grundlagenarbeit zu sprachlichen Eigenschaften von Verschwörungstheorien um in der Zukunft bessere Systeme zu deren Erkennung zu entwickeln.

Literatur

- Butter, M. (2018). *Nichts ist, wie es scheint: über Verschwörungstheorien* (Originalausgabe, Erste Auflage). Suhrkamp.
- Clarke, S. (2007). Conspiracy Theories and the Internet: Controlled Demolition and Arrested Development. *Episteme*, 4(2), 167–180. <https://doi.org/10.3366/epi.2007.4.2.167>
- Coady, D. (Hrsg.). (2006). *Conspiracy Theories*. Routledge. <https://doi.org/10.4324/9781315259574>
- Feldman, R. & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511546914>
- Filatkina, N. (2018). Der Anfang vom Ende Europas. Verschwörungstheoretisches Sprechen im Migrationsdiskurs (D. Römer & S. Stumpf, Hrsg.). *aptum*, 3 (2018), 197–220.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/s0167-9473\(01\)00065-2](https://doi.org/10.1016/s0167-9473(01)00065-2)
- Hofstadter, R. & Wilentz, S. (2008). *The Paranoid Style in American Politics*. Random House. <https://books.google.de/books?id=XcLSoljnmBcC> (Original erschienen 1964)
- Jimenez, S., Avila, Y., Dueñas, G. & Gelbukh, A. (2020). Automatic prediction of citability of scientific articles by stylometry of their titles and abstracts. *Scientometrics*, 125(3), 3187–3232. <https://doi.org/10.1007/s11192-020-03526-1>
- Maron, M. E. (1961). Automatic Indexing: An Experimental Inquiry. *J. ACM*, 8(3), 404–417. <https://doi.org/10.1145/321075.321084>
- Miller, S. (2002). Conspiracy Theories: Public Arguments as Coded Social Critiques: a Rhetorical Analysis of the Twa Flight 800 Conspiracy Theories. *Argumentation and Advocacy*, 39(1), 40–56. <https://doi.org/10.1080/00028533.2002.11821576>
- Neethu, M. S. & Rajasree, R. (2013). Sentiment analysis in twitter using machine learning techniques. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–5. <https://doi.org/10.1109/ICCCNT.2013.6726818>
- Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J. & Stein, B. (2018). A Stylo-metric Inquiry into Hyperpartisan and Fake News. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 231–240. <https://doi.org/10.18653/v1/P18-1022>
- Remus, R., Quasthoff, U. & Heyer, G. (2010). SentiWS - A Publicly Available German-language Resource for Sentiment Analysis. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf

- Römer, D. & Stumpf, S. (2019). Populismus und Verschwörungstheorien aus linguistischer Perspektive (D. Römer & C. Spieß, Hrsg.). *Osnabrücker Beiträge zur Sprachtheorie*, 95, 129–158.
- Schäfer, P. (2018). Verschwörungstheorien – Ein Fall für die linguistische Vertrauensforschung (D. Römer & S. Stumpf, Hrsg.). *Aptum*, 3 (2018), 221–238.
- Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R. & Roychowdhury, V. (2020). Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. *Journal of Computational Social Science*, 3(2), 279–317. <https://doi.org/10.1007/s42001-020-00086-5>
- Smith, G. et al. (2003). A brief introduction to the TIGER treebank, version 1. *Tiger Projektbericht*, Univ. Potsdam.
- Soukup, C. (2008). 9/11 conspiracy theories on the World Wide Web: Digital rhetoric and alternative epistemology. *Journal of Literacy and Technology*, 9(3), 2–25.
- Stano, S. (2020). THE INTERNET AND THE SPREAD OF CONSPIRACY CONTENT. In M. Butter & P. Knight (Hrsg.), *Routledge Handbook of Conspiracy Theories* (S. 483–496). Routledge.
- Sunstein, C. R. & Vermeule, A. (2008). Conspiracy Theories. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1084585>
- Uscinski, J. E. & Parent, J. M. (2014). *American Conspiracy Theories*. Oxford University Press.
- Yang, Y. & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. *Icml*, 97(412-420), 35.

Software

- Feinerer, I., Hornik, K. & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54. <https://www.jstatsoft.org/v25/i05/>
- Honnibal, M., Montani, I., Van Landeghem, S. & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. Zenodo. <https://doi.org/10.5281/zenodo.1212303>
- Ke, G. (2019). *lightgbm: Light Gradient Boosting Machine* [R package version 2.3.2]. <https://github.com/Microsoft/LightGBM>
- Ooms, J. (2021). *cld3: Google’s Compact Language Detector 3* [R package version 1.4.1]. <https://CRAN.R-project.org/package=cld3>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Therneau, T. & Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees* [R package version 4.1-15]. <https://CRAN.R-project.org/package=rpart>

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

Online

- Alles Schall und Rauch*. (2020). Psiram. Verfügbar 23. Mai 2021 unter 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>
- Benesch, A. (2018). *Alex Jones: Internierungslager und REX-84 sind jetzt die Rettung Amerikas*. Recentr. Verfügbar 23. Mai 2021 unter <http://recontr.com/2018/06/26/alex-jones-internierungslager-und-rex-84-sind-jetzt-die-rettung-amerikas/>
- Benesch, A. (2021a). *Bevölkerungsreduktion wird schrittweise, mit fast ausschließlich legalen Mitteln betrieben, anders als viele sich das vorstellen*. Recentr. Verfügbar 23. Mai 2021 unter <http://recontr.com/2021/05/10/bevoelkerungsreduktion-wird-schrittweise-mit-fast-ausschliesslich-legalen-mitteln-betrieben-anders-als-viele-sich-das-vorstellen/>
- Benesch, A. (2021b). *Ist die AfD eine britische Schöpfung?* Recentr. Verfügbar 23. Mai 2021 unter <http://recontr.com/2021/06/29/ist-die-afd-eine-britische-schoepfung/>
- Blog AGB's*. (2016). Deutschland - Pranger. Verfügbar 23. Mai 2021 unter <http://deutschland-pranger.de/b2evolution/index.php/ueber-diese-seite>
- Conrebbi*. (2020). Psiram. Verfügbar 23. Mai 2021 unter <https://www.psiram.com/de/index.php/Conrebbi>
- 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/>
- Redaktion, W. (2018). *+++ Schulz tritt sofort zurück – Bilderberger Scholz gibt jetzt den Ton an +++*. Watergate.tv. Verfügbar 23. Mai 2021 unter <https://wg.neopresse.com/schulz-tritt-sofort-zurueck-bilderberger-scholz-gibt-jetzt-den-ton-an/>
- ZDF bestätigt Trennung von Hans Meiser*. (2017). Spiegel Online. Verfügbar 23. Mai 2021 unter <https://www.spiegel.de/kultur/tv/jan-boehmermann-neo-magazin-royale-kuenftig-ohne-hans-meiser-a-1148048.html>