

BACHELOR THESIS
Kristoffer Schaaf

Entwicklung einer Software zur Erkennung von Fake News auf Nachrichtenportalen

FAKULTÄT TECHNIK UND INFORMATIK
Department Informatik

Faculty of Engineering and Computer Science
Department Computer Science

Kristoffer Schaaf

Entwicklung einer Software zur Erkennung von Fake News auf Nachrichtenportalen

Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung
im Studiengang *Bachelor of Science Angewandte Informatik*
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Stefan Sarstedt
Zweitgutachter: Prof. Dr. Marina Tropmann-Frick

Eingereicht am: 01.04.2025

Kristoffer Schaaf

Thema der Arbeit

Entwicklung einer Software zur Erkennung von Fake News auf Nachrichtenportalen

Stichworte

Machinelles Lernen, Fake News, Nachrichtenportale,

Kurzzusammenfassung

Arthur Dents Reise in eine neue Zukunft ...

Kristoffer Schaaf

Title of Thesis

Development of a software for the detection of fake news on news portals

Keywords

Machine Learning, Fake News, Text Mining, Classification, NLP

Abstract

Arthur Dents travel to a new future ...

Inhaltsverzeichnis

Abbildungsverzeichnis	vi
Tabellenverzeichnis	vii
1 Einleitung	1
1.1 Hintergrund: Die zunehmende Verbreitung von Fake News und deren gesellschaftliche Auswirkungen	1
1.1.1 Wann entstanden Fake News	1
1.1.2 Wie definieren sich Fake News und wie sind sie aufgebaut	1
1.1.3 Aus welcher Motivation entstehen Fake News	3
1.1.4 Warum verbreiten sich Fake News	3
1.1.5 Probleme beim Erkennen von Fake News	3
1.1.6 Klassifizierungen	3
1.2 Wahl der Nachrichtenportale	4
1.3 Zielsetzung: Entwicklung einer Software zur automatisierten Fake-News-Erkennung	4
1.4 Aufbau der Arbeit	4
2 Grundlagen und Begriffsdefinitionen	5
2.1 Definition „Fake News“: Merkmale, Ziele, Beispiele	5
2.2 Kategorisierung der Fake News Detection-Ansätze	5
2.3 Warum der Fokus auf Machine Learning?	5
2.4 Überblick über relevante Plattformen und deren Rolle im Medienkonsum .	5
3 Maschinelles Lernen zur Fake News Erkennung	6
3.1 Grundlagen von ML	6
3.1.1 Überwachtes vs. unüberwachtes Lernen	6
3.1.2 Klassische ML-Modelle (Logistic Regression, Naive Bayes, SVM, Entscheidungsbäume)	6

3.2	Performance-Metriken: Accuracy, Precision, Recall, F1-Score	6
3.3	Herausforderungen im Kontext von Fake News	6
4	Natural Language Processing und Deep Learning	7
4.1	Text-Vorverarbeitung	7
4.2	Word Embeddings	7
4.3	Deep Learning-Modelle	7
4.4	Diskussion: White Box vs. Black Box-Ansätze zur Erklärbarkeit	7
5	Relevante Datensätze und Auswahlkriterien	8
5.1	Vorstellung verfügbarer deutscher Fake-News-Datensätze	8
5.2	Auswahl und Begründung des finalen Datensatzes	8
5.3	Herausforderungen bei deutschen Textdaten (Klassenverteilung, Quellen- vielfalt)	8
6	Konzeption der Softwarelösung	9
7	Umsetzung des Prototyps	10
8	Evaluation und Ergebnisse	11
9	Diskussion	12
10	Fazit und Ausblick	13
	Literaturverzeichnis	14
A	Anhang	15
A.1	Verwendete Hilfsmittel	15
	Selbstständigkeitserklärung	16

Abbildungsverzeichnis

Tabellenverzeichnis

A.1	Verwendete Hilfsmittel und Werkzeuge	15
-----	--	----

1 Einleitung

1.1 Hintergrund: Die zunehmende Verbreitung von Fake News und deren gesellschaftliche Auswirkungen

1.1.1 Wann entstanden Fake News

Fake News sind ein allgegenwärtiges Problem, doch hatten Sie Ihren ersten Auftritt bereits 44BC im römischen Reich [2]. Auch während des amerikanischen Bürgerkriegs 1779 wurden Sie als politischer Schachzug von Benjamin Franklin genutzt. Dieser schickte einen Brief an Captain Samuel Gerrish und schrieb in diesem über Grausamkeiten der Briten und deren Verbündeten. Diese Informationen wurde so veranschaulicht, dass sie die öffentliche Meinung bewusst beeinflussen sollten [4].

Der eigentliche Begriff "Fake News" wurde erst viele Jahre später durch Donald Trump im amerikanischen Wahlkampf 2016 bekannt [1].

Unter anderem ist Fake News auch ein Teil von Propaganda [3], welche schon lange als Mittel zur Meinungsmanipulation eines Volkes genutzt wird.

Heute ist Fake News die größte Drohung zu unserer angeblich freien Presse [4].

1.1.2 Wie definieren sich Fake News und wie sind sie aufgebaut

Fake News sind bewusst erstellte Online-Falschmeldungen, die teilweise oder vollständig unwahre Inhalte verbreiten, um Leser*innen gezielt zu täuschen oder zu manipulieren. Sie imitieren klassische Nachrichtenformate, nutzen auffällige Titel, emotionale Bilder und strategisch gestaltete Inhalte, um Glaubwürdigkeit zu erzeugen und Aufmerksamkeit zu gewinnen. Ziel ist es, durch das Verbreiten dieser Inhalte Klicks, Reichweite und damit finanzielle oder ideologische Vorteile zu erzielen. [Baptista2020].

Fake News fallen in die Kategorien Satire, Clickbait, Gerüchte, Stance News, Propaganda und Large Scale Hoaxes [4].

- **Satire:** ist eine humorvolle oder übertriebene Darstellung gesellschaftlicher oder politischer Themen, die Kritik üben soll.
- **Clickbait:** bezeichnet reißerische Überschriften oder Vorschaubilder, die Neugier wecken und zum Anklicken eines Inhalts verleiten sollen, oft ohne den Erwartungen gerecht zu werden.
- **Gerüchte:** sind unbestätigte Informationen, die sich schnell verbreiten und oft falsch oder irreführend sind.
- **Stance News:** sind Nachrichten, die eine klare Meinung oder politische Haltung einnehmen, statt neutral zu berichten.
- **Propaganda:** ist die gezielte Verbreitung von Informationen oder Meinungen, um das Denken und Handeln von Menschen zu beeinflussen, meist im Interesse einer bestimmten Gruppe oder Ideologie.
- **Large Scale Hoaxes:** sind absichtlich erfundene Falschmeldungen oder Täuschungen, die weit verbreitet werden und viele Menschen täuschen sollen.

Die eigentliche Nachricht ist aufgebaut in folgende Teile:

- **Quelle:** gibt den Ersteller der Nachricht an.
- **Titel:** erzielt die Aufmerksamkeit der Lesenden.
- **Text:** enthält die eigentliche Information der Nachricht.
- **Medien:** in Form von Bildern oder Videos.

Fake News können die Form von Text, Fotos, Filmen oder Audio annehmen und sind dementsprechend auf jeder Plattform auffindbar, die die Verbreitung nicht unterbindet. Die momentan populärste Plattform zum Teilen der Fake News ist WhatsApp [1].

1.1.3 Aus welcher Motivation entstehen Fake News

Das Hauptinteresse der Ersteller der Fake News ist das Verdienen von Geld. Auf die Artikel wird Werbung geschaltet und anhand einer entsprechenden Reichweite ergibt sich der verdiente Betrag. Je mehr Reichweite, desto mehr Verdienst für die Ersteller [2].

1.1.4 Warum verbreiten sich Fake News

In sozialen Medien neigen Nutzer aufgrund von FOMO (Fear of Missing Out) dazu, Fake News zu teilen, um Anerkennung zu gewinnen und soziale Zugehörigkeit zu erfahren. Besonders häufig werden kontroverse, überraschende oder bizarre Inhalte verbreitet – insbesondere dann, wenn sie starke Emotionen wie Freude, Wut oder Aufregung hervorrufen. Das Teilen solcher Inhalte stärkt das eigene Ansehen, da es signalisiert, über neue und relevante Informationen zu verfügen. Fake News bestehen meist aus eindrucksvoll präsentierten Falschinformationen. [2]

1.1.5 Probleme beim Erkennen von Fake News

Fake News können erst erkannt werden, nachdem diese erstellt und im Internet verbreitet wurden. [4]

1.1.6 Klassifizierungen

Linguistische Features werden von Textmaterial auf verschiedenen Leveln gesammelt, z.B. Buchstaben, Wörter, Sätze und Features auf dem Satzlevel (Häufigkeit von Funktionswörtern? und Sätzen) [6]

Text Tokenisierung [5]

Aufteilung der Features in Syntactic, Semantic, Sentiment, Lexical, Style-based [4] p7

1.2 Wahl der Nachrichtenportale

1.3 Zielsetzung: Entwicklung einer Software zur automatisierten Fake-News-Erkennung

1.4 Aufbau der Arbeit

2 Grundlagen und Begriffsdefinitionen

2.1 Definition „Fake News“: Merkmale, Ziele, Beispiele

2.2 Kategorisierung der Fake News Detection-Ansätze

2.3 Warum der Fokus auf Machine Learning?

2.4 Überblick über relevante Plattformen und deren Rolle
im Medienkonsum

3 Maschinelles Lernen zur Fake News Erkennung

3.1 Grundlagen von ML

3.1.1 Überwachtes vs. unüberwachtes Lernen

3.1.2 Klassische ML-Modelle (Logistic Regression, Naive Bayes, SVM, Entscheidungsbäume)

3.2 Performance-Metriken: Accuracy, Precision, Recall, F1-Score

3.3 Herausforderungen im Kontext von Fake News

4 Natural Language Processing und Deep Learning

4.1 Text-Vorverarbeitung

4.2 Word Embeddings

4.3 Deep Learning-Modelle

4.4 Diskussion: White Box vs. Black Box-Ansätze zur Erklärbarkeit

5 Relevante Datensätze und Auswahlkriterien

5.1 Vorstellung verfügbarer deutscher Fake-News-Datensätze

5.2 Auswahl und Begründung des finalen Datensatzes

5.3 Herausforderungen bei deutschen Textdaten (Klassenverteilung, Quellenvielfalt)

6 Konzeption der Softwarelösung

7 Umsetzung des Prototyps

8 Evaluation und Ergebnisse

9 Diskussion

10 Fazit und Ausblick

Literaturverzeichnis

- [1] ASHISH ; SONIA ; ARORA, Monika ; HEMRAJ ; RANA, Anurag ; GUPTA, Gaurav: An Analysis and Identification of Fake News using Machine Learning Techniques. In: *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2024, S. 634–638
- [2] BAPTISTA, Joao P. ; GRADIM, Anabela: Understanding Fake News Consumption: A Review. In: *Social Sciences* 9 (2020), Nr. 10. – URL <https://www.mdpi.com/2076-0760/9/10/185>. – ISSN 2076-0760
- [3] BÜRKER, Michael: *Fake-News, Propaganda & Co: Wie behalte ich den Überblick?* 2022. – URL <https://www.haw-landshut.de/aktuelles/beitrag/fake-news-propaganda-co-wie-behalte-ich-den-ueberblick>. – Zugriffsdatum: 2025-04-23. – Interview geführt von EINFALLSreich, Hochschule Landshut
- [4] SHARMA, Upasna ; SINGH, Jaswinder: A comprehensive overview of fake news detection on social networks. In: *Social Network Analysis and Mining* 14 (2024), Nr. 1, S. 120. – URL <https://doi.org/10.1007/s13278-024-01280-3>. ISBN 1869-5469
- [5] WAGNER, Wiebke: Steven Bird, Ewan Klein and Edward Loper: Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit. In: *Language Resources and Evaluation* 44 (2010), Nr. 4, S. 421–424. – URL <https://doi.org/10.1007/s10579-010-9124-x>. ISBN 1574-0218
- [6] ZHANG, Xichen ; HABIBI LASHKARI, Arash ; A. GHORBANI, Ali: A Lightweight Online Advertising Classification System using Lexical-based Features. In: *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - SECRYPT INSTICC* (Veranst.), SciTePress, 2017, S. 486–494. – ISBN 978-989-758-259-2

A Anhang

A.1 Verwendete Hilfsmittel

In der Tabelle A.1 sind die im Rahmen der Bearbeitung des Themas der Bachelorarbeit verwendeten Werkzeuge und Hilfsmittel aufgelistet.

Tabelle A.1: Verwendete Hilfsmittel und Werkzeuge

Tool	Verwendung
L ^A T _E X	Textsatz- und Layout-Werkzeug verwendet zur Erstellung dieses Dokuments

Erklärung zur selbständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort

Datum

Unterschrift im Original