

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

<b>Abbildungsverzeichnis</b>	<b>vi</b>
<b>Tabellenverzeichnis</b>	<b>vii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Hintergrund: Die zunehmende Verbreitung von Fake News und deren gesellschaftliche Auswirkungen . . . . .	1
1.1.1 Klassifizierungen . . . . .	2
1.2 Wahl der Nachrichtenportale . . . . .	2
1.3 Zielsetzung: Entwicklung einer Software zur automatisierten Fake-News-Erkennung . . . . .	2
1.4 Aufbau der Arbeit . . . . .	2
<b>2 Grundlagen und Begriffsdefinitionen</b>	<b>3</b>
2.1 Definition „Fake News“: Merkmale, Ziele, Beispiele . . . . .	3
2.2 Kategorisierung der Fake News Detection-Ansätze . . . . .	3
2.3 Warum der Fokus auf Machine Learning? . . . . .	3
2.4 Überblick über relevante Plattformen und deren Rolle im Medienkonsum .	3
<b>3 Maschinelles Lernen zur Fake News Erkennung</b>	<b>4</b>
3.1 Grundlagen von ML . . . . .	4
3.1.1 Überwachtes vs. unüberwachtes Lernen . . . . .	4
3.1.2 Klassische ML-Modelle (Logistic Regression, Naive Bayes, SVM, Entscheidungsbäume) . . . . .	4
3.2 Performance-Metriken: Accuracy, Precision, Recall, F1-Score . . . . .	4
3.3 Herausforderungen im Kontext von Fake News . . . . .	4
<b>4 Natural Language Processing und Deep Learning</b>	<b>5</b>
4.1 Text-Vorverarbeitung . . . . .	5
4.2 Word Embeddings . . . . .	5

4.3	Deep Learning-Modelle . . . . .	5
4.4	Diskussion: White Box vs. Black Box-Ansätze zur Erklärbarkeit . . . . .	5
<b>5</b>	<b>Relevante Datensätze und Auswahlkriterien</b>	<b>6</b>
5.1	Vorstellung verfügbarer deutscher Fake-News-Datensätze . . . . .	6
5.2	Auswahl und Begründung des finalen Datensatzes . . . . .	6
5.3	Herausforderungen bei deutschen Textdaten (Klassenverteilung, Quellen- vielfalt) . . . . .	6
<b>6</b>	<b>Konzeption der Softwarelösung</b>	<b>7</b>
<b>7</b>	<b>Umsetzung des Prototyps</b>	<b>8</b>
<b>8</b>	<b>Evaluation und Ergebnisse</b>	<b>9</b>
<b>9</b>	<b>Diskussion</b>	<b>10</b>
<b>10</b>	<b>Fazit und Ausblick</b>	<b>11</b>
	<b>Literaturverzeichnis</b>	<b>12</b>
<b>A</b>	<b>Anhang</b>	<b>13</b>
A.1	Verwendete Hilfsmittel . . . . .	13
	<b>Selbstständigkeitserklärung</b>	<b>14</b>

# Abbildungsverzeichnis

# Tabellenverzeichnis

A.1	Verwendete Hilfsmittel und Werkzeuge . . . . .	13
-----	--	----

# 1 Einleitung

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

Fake News sind ein allgegenwärtiges Problem. Als politischer Schachzug bereits im Amerikanischen Bürgerkrieg 1779 genutzt, hat der Term auch in der US-Wahl 2016 wieder eine besondere Rolle gespielt [3].

Fake News aufgeteilt in Satire, Clickbait, Rumor, Stance News, Propaganda, Large Scale Hoaxes [3]-p2

Fake News ist die größte Drohung zu unserer angeblich freien Presse [3]-p3

Fake News als Begriff politischer Kampfbegriff aus Trump-Wahlkämpfen -> stattdessen "Desinformationen"[2] Propaganda arbeitet mit Desinformationen. [2]

Nachrichteninhalt: Quelle, Titel-erzielt die Aufmerksamkeit der Lesers, Text-enthält die eigentliche Information, Media (Bilder/Videos)

Sie 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].

Support Vector Machines sind die mit am meisten genutzten Kategorisierungs Algorithmen [3]p12 -> Hier gibt es eine Aufzählung an Algorithmen die für das ML Model genutzt werden könnten! (auch deep learning im späteren Teil)

### 1.1.1 Probleme beim Erkennen von Fake News

Fake News können erst erkannt werden, nachdem diese erstellt und im Internet verbreitet wurden. [3]



### **1.1.2 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) [5]

Text Tokenisierung [4]

Aufteilung der Features in Syntactic, Semantic, Sentiment, Lexical, Style-based [3]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] 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
- [3] 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
- [4] 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
- [5] 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) - SECRIPT 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 <sup>A</sup> T <sub>E</sub> 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