
BACHELORARBEIT

Herr
Tobias Münzner

**Sentimentanalyse in sozialen
Netzwerken im forensischen
Kontext**

Mittweida, 2017

BACHELORARBEIT

Sentimentanalyse in sozialen Netzwerken im forensischen Kontext

Autor:

Herr

Tobias Münzner

Studiengang:

Allgemeine und digitale Forensik

Seminargruppe:

FO14w1-B

Erstprüfer:

Professor Labudde

Zweitprüfer:

M.Sc. Michael Spranger

Einreichung:

Mittweida, 26.10.2017

Verteidigung/Bewertung:

Faculty < fakultätsbezeichner >

BACHELORTHESIS

sentiment analysis in social networks in forensic context

author:

Tobias Münzner

course of studies:

General and Digital Forensic Science

seminar group:

FO14w1-B

first examiner:

Professor Dirk Labudde

second examiner:

M.Sc. Michael Spranger

submission:

Mittweida, 26.10.2017

defence/ evaluation:

Bibliografische Beschreibung:

Münzner, Tobias

Sentimentanalyse in sozialen Netzwerken im forensischen Kontext- 2017 - 6, 68, 2

Mittweida, Hochschule Mittweida, Fakultät Computer- und Biowissenschaften, Bachelorarbeit, 2017

Referat:

Im Rahmen der Arbeit wird ein Überblick über aktuelle Verfahren der Sentimentanalyse gegeben, welche sich besonders für die Analyse von Nutzerdaten in sozialen Netzwerken eignen und auch auf die deutsche Sprache ausgerichtet sind bzw. auf diese übertragen werden können. Darüber hinaus sollen forensische Anwendungsgebiete der Sentimentanalyse aufgezeigt und diskutiert werden.

Inhalt

Inhalt I

Abbildungsverzeichnis	III
Abkürzungsverzeichnis	VI
1 Einleitung	1
2 Grundlegende Begriffe.....	3
2.1 <i>Opinion Mining</i>	3
2.1.1 Sentimentanalyse	5
2.1.2 Forschungsentwicklung	7
2.2 <i>Soziale Netzwerke</i>	9
2.2.1 Blog.....	10
2.2.2 Twitter	10
2.2.3 Facebook	12
2.3 <i>Forensik</i>	13
3 Ansätze der Sentimentanalyse	15
3.1 <i>Lernbasierter Ansatz</i>	16
3.1.1 Naive Bayes	19
3.1.2 Support Vector Machine (SVM)	20
3.1.3 Maximale Entropie.....	23
3.1.4 Cluster.....	23
3.2 <i>Lexikalischer Ansatz</i>	24
3.2.1 SentiWS	26
3.2.2 German Polarity Clues (GPC)	28
3.2.3 Sonstige	29
4 Implementierung	30
4.1 <i>Datenbeschaffung und Vorverarbeitung</i>	30
4.1.1 Twitter	31
4.1.2 Facebook	33
4.2 <i>Polaritätsbestimmung</i>	35
4.3 <i>Ergänzende Faktoren</i>	37
4.3.1 Ironie / Sarkasmus Erkennung	37

4.3.2	Berücksichtigung von Emotionsausdrücken (Emojis).....	38
4.3.3	Glaubwürdigkeit.....	39
4.3.4	Interaktion zwischen Nutzern.....	41
5	Evaluation	42
5.1	<i>Testdatensätze / Goldstandards</i>	<i>42</i>
5.1.1	Korpus nach Boland	43
5.1.2	MLSA	44
5.1.3	Sonstige	47
5.2	<i>Evaluationsmaße.....</i>	<i>48</i>
5.2.1	Konfusionsmatrix	48
5.2.2	Positiver Vorhersagewert (Precision).....	48
5.2.3	Sensitivität (Recall).....	49
5.2.4	Korrektklassifikationsrate (Accuracy)	49
5.2.5	F-Maß.....	49
5.3	<i>Evaluation vorgestellter Verfahren.....</i>	<i>50</i>
6	Forensische Anwendungsgebiete.....	54
7	Zusammenfassung und Fazit	64
Literatur		69
Anlagen		74
Selbstständigkeitserklärung.....		76

Abbildungsverzeichnis

Abbildung 1: Modell des Opinion Mining am Beispiel einer Kamera	4
Abbildung 2: Prozess des Opinion Mining	5
Abbildung 3: Anzahl der jährlichen Publikationen zum Thema OMSA	8
Abbildung 4: Anzahl der jährlich erscheinenden Publikationen sortiert nach Ländern.....	8
Abbildung 5: System der Kriminalwissenschaften nach Barthel.....	13
Abbildung 6: Formel der Euklidischen Distanz.....	17
Abbildung 7: Modell der Euklidischen Distanz	17
Abbildung 8: Formel des Naive Bayes Algorithmus	19
Abbildung 9: Prinzip des SVM Verfahrens.....	21
Abbildung 10: Problem bei der Klassifikation von Dokumenten	21
Abbildung 11: Überführen des Problems von SVM in einen höherdimensionalen Raum	22
Abbildung 12: KNN Klassifikation	23
Abbildung 13: Überblick über den Inhalt von SentiWS.....	26
Abbildung 14: Schema eines Eintrags in SentiWS.....	27
Abbildung 15: Verteilung der gewichteten Polarität und SentiWS	28
Abbildung 16: Verteilung der gewichteten Polarität und SentiWS	29
Abbildung 17: Ablauf einer Extraktion von Tweets.....	31
Abbildung 18: POS Tagging an einem Beispiel	33
Abbildung 19: Vorverarbeitungsschritte von Facebook Nachrichten	34
Abbildung 20: Algorithmus zur Bestimmung der semantischen Wortorientierung	36

Abbildung 21: Algorithmus zur Polarität eines Satzes.....	36
Abbildung 22: Einordnung von Emoji's	39
Abbildung 23: Verteilung von Produktkategorien im Korpus	43
Abbildung 24: Ausschnitt aus dem Mehrschichtensystem des MLSA	44
Abbildung 25: Verteilung der subjektiv bzw. objektiv Bewertungen in Schicht 1	45
Abbildung 26: Verteilung der Polarität in Schicht 1	45
Abbildung 27: Verteilung der Polarität in Schicht 2	45
Abbildung 28: Verteilung der Typen in Schicht 3	46
Abbildung 29: Verteilung der Kategorien in Schicht 3	46
Abbildung 30: Konfusionsmatrix	48
Abbildung 31: Formel für Precision.....	48
Abbildung 32: Formel für Recall	49
Abbildung 33: Accuracy als Formel	49
Abbildung 34: Formel für F-Maß.....	49
Abbildung 35: Evaluation der Resultate ohne Vorverarbeitung	50
Abbildung 36: Resultate mit Hinzufügen der Termfrequenz	50
Abbildung 37: Resultate mit Hinzufügen von TF-IDF	50
Abbildung 38: Ergebnisse ohne Berücksichtigung von 2-grams	51
Abbildung 39: Ergebnisse mit der Entfernung von Stop-Wörtern	51
Abbildung 40: Evaluation von SentiWS	52
Abbildung 41: Evaluation von englischsprachigen Features	52
Abbildung 42: Evaluation von deutschsprachigen Sentiment Wörterbüchern	53
Abbildung 43: Polarität einzelner URLs	55

Abbildung 44: Verteilung der Themen einer beobachteten Facebook Gruppe	55
Abbildung 45: Verteilung der Medien zum Thema Hass-Reden.....	55
Abbildung 46: Top Features zur Erkennung von Rechtsextremismus.....	56
Abbildung 47: Liste von Einträgen des SentiLEN mit Polaritätswert.....	57
Abbildung 48: Kategorien des DHS mit Beispielwörtern zum Thema Cyberkriminalität.....	58
Abbildung 49: Liste von Begriffen und deren Häufigkeit aus dem Ansar1 Datensatz	58
Abbildung 50: normalisierte und durchschnittliche Punktwerte von Begriffen aus dem SentiLEN	59
Abbildung 51: Landkarte mit Sterblichkeitsrisiko an AHD	61
Abbildung 52: mögliche Sentiment Klassen im medizinischen Kontext.....	62

Abkürzungsverzeichnis

API	Application Programming Interface
CDC	Centers for Disease Control and Prevention
DHS	Department of Homeland Security
GGD	Geneeskundige en Gezondheidsdienst
GPC	German Polarity Clues
HTML	Hypertext Markup Language
IGGSA	Interest Group for German Sentiment Analysis
JSON	JavaScript Object Notation
OSMA	Opinion Mining / Sentiment Analysis
PHP	Hypertext Preprocessor
POS	Part of Speech
PotTS	Potsdam Twitter Sentiment Corpus
SentiWS	Sentiment Wortschatz
SVM	Support Vector Machine
USAGE	Bielefeld University Sentiment Analysis Corpus for German and English
WEKA	Waikato Environment for Knowledge Analysis

1 Einleitung

Die Analyse von Nutzerdaten ist eine der wichtigsten Aufgaben aktuell im Bereich Big Data. Dafür sind besonders die sozialen Netzwerke prädestiniert, da dort die Nutzer oder Autoren oftmals ihre ehrlichen Meinungen und Emotionen zu einer Vielzahl von Themen ausdrücken. Gerade Mikroblogs wie beispielsweise Twitter sind für die Analyse sehr attraktiv, da sie aufgrund ihrer einfachen Handhabung und geringen Einschränkungen große Mengen an Nutzerdaten enthalten.¹

In den letzten Jahren sind, infolge der Entwicklung von Techniken zur Informationsextraktion, diverse Ansätze vorgestellt worden, um Meinungen in den verschiedenen sozialen Kanälen, wie Twitter, Facebook oder Blogs zu klassifizieren. Dieses Verfahren, bei dem Meinungen, mehr oder weniger, automatisiert in Kategorien wie positiv, negativ oder neutral eingestuft werden, wird als Sentimentanalyse bezeichnet. Die Mehrzahl der derzeitigen Sentimentanalyse Verfahren beziehen sich für gewöhnlich auf die englische Sprache. Mit der weltweiten Verbreitung des Internets und den damit verbundenen vielfältigen Sprachen stieg jedoch auch die Notwendigkeit die Sentimentanalyse multilingual weiterzuentwickeln. Ein grundlegendes Problem dabei war jedoch der Mangel an Ressourcen, da das Wissen von reichen Ländern und den damit verbundenen Sprachen, erst in ärmere Sprachregionen transferiert werden musste. Deshalb greifen die meisten Sentimentanalyse Techniken auf englischsprachige Daten und Ressourcen, wie das SentiWordNet, zurück oder übersetzen die Texte vorher maschinell aus anderen Sprachen ins Englische. Dabei treten jedoch häufig Probleme wie Lücken und sogenanntes Rauschen in den Daten auf oder wesentliche Teile des Textes werden gar nicht erst übersetzt, sodass keine vollständigen Sätze mehr entstehen. Somit war die Forschung gefordert alternative Ansätze zu entwickeln und inzwischen gibt es vermehrt semantische Wörterbücher oder Testdatensätze (Korpusse) in anderen Sprachen, wie Chinesisch, Japanisch oder auch Deutsch.²

Bisher wird die Sentimentanalyse hauptsächlich in Bereichen wie der Bewertung von Produkten oder Dienstleistungen verwendet. Zunehmend rücken aber auch weitere Forschungsfelder

¹ Adarsh M J und Pushpa Ravikumar 2015, S. 34

² Dashtipour et al. 2016, S. 757–758

in den Vordergrund, welche unmittelbar oder entfernt mit der Forensik zu tun haben. Dazu zählen Beispiele wie die Auswertung von Profilen oder Nachrichten aus den sozialen Netzwerken, um Missbrauch oder potentiell gefährliche Nutzer frühzeitig zu erkennen. Aufgrund der zunehmenden Digitalisierung und der damit verbundenen Cyberkriminalität ist davon auszugehen, dass die Sentimentanalyse in den diversen Teilgebieten der Forensik zukünftig eine immer wichtigere Rolle spielt.

Ziel der vorliegenden Arbeit soll es sein, einen Überblick über aktuelle Verfahren der Sentimentanalyse zu geben, welche sich besonders für die Analyse von Nutzerdaten in sozialen Netzwerken eignen und auch auf die deutsche Sprache ausgerichtet sind bzw. auf diese übertragen werden können. Darüber hinaus sollen forensische Anwendungsgebiete der Sentimentanalyse aufgezeigt und diskutiert werden.

Zunächst sollen dazu in Kapitel 2 grundlegende Begriffe aus den Themengebieten des Opinion Mining und der sozialen Netzwerke definiert und erläutert werden. Zusätzlich wird auch die Forensik definiert und als Wissenschaftsgebiet eingeordnet. Daraufhin werden in Kapitel 3 die verschiedenen Ansätze der Sentimentanalyse mit den jeweiligen Vertretern vorgestellt. In Kapitel 4 wird dann der Prozess der Implementierung betrachtet und auch auf ergänzende Faktoren, welche die Sentimentanalyse unterstützen können, eingegangen. Das nachfolgende Kapitel 5 widmet sich der Evaluation bereits bestehender Techniken und dabei werden auch die dafür verwendeten Maße vorgestellt. In Kapitel 6 werden Anwendungsgebiete der Sentimentanalyse in der Forensik vorgestellt und diskutiert. Abschließend werden dann in Kapitel 7 noch einmal die wesentlichen Erkenntnisse der Arbeit in Form eines Fazits zusammengefasst und ein Ausblick auf zukünftigen Forschungsbedarf in diesem Bereich gegeben.

2 Grundlegende Begriffe

2.1 Opinion Mining

Opinion Mining beschäftigt sich im Gegensatz zum klassischen Data Mining, was auf Fakten beruht, mit subjektiven, personengebunden Meinungen. Gerade seit der Entwicklung des sogenannten Web 2.0 findet diese Art der Analyse immer mehr Beachtung und Verbreitung. Insbesondere Unternehmen haben ein großes Interesse an Meinungen von Kunden zu Produkten und Dienstleistungen, da diese die Kaufentscheidungen und Markenbindung anderer Menschen beeinflussen. Im Gegensatz zu früher, wo diese Meinungen erst durch Fragebögen oder ähnliches in Erfahrungen gebracht werden mussten, sind diese heute bereits in großer Anzahl im Internet und dort insbesondere in den sozialen Netzwerken vorhanden. Die ungeheure Menge vorhandener Daten hat aber auch zur Folge, dass eine reine manuelle Analyse und Bewertung nicht mehr realisierbar ist und daher auf maschinelle Hilfe zurückgegriffen werden muss. Dabei wird in Kauf genommen, dass ein Algorithmus nie ganz die Qualität einer manuellen Bewertung erreichen kann, aber er erlaubt durch die gestiegene Quantität eine repräsentativere Gesamtaussage. Hier kommt die Forschungsdisziplin des Opinion Mining ins Spiel, welche versucht mithilfe von Algorithmen diese Meinungen auszuwerten.

Obwohl es als Opinion Mining bezeichnet wird, ist es eigentlich eher eine Sammlung von Tätigkeiten, die in einer bestimmten Reihenfolge erledigt werden. Dazu gehören die Erkennung von subjektiven Daten, die Einschätzung der Relevanz der Daten, die Vorhersage der Polarität (Sentimentanalyse), die Extraktion von Produktfeatures oder die Zusammenfassung von meinungstragenden Texten.³

Opinion Mining wird auch oft synonym zu Sentimentanalyse verwendet, doch diese stellt nach überwiegender Meinung lediglich einen Teilbereich des Opinion Mining dar, wenn auch den vermutlich zentralsten. Opinion Mining kann vielmehr als Modell verschiedener Elemente, wie Objekten, Features, Meinungsäußerung, Meinungseigentümer und semantischer Orientierung zusammengefasst werden (*siehe Abbildung 1*). Ein Objekt stellt dabei eine Entität dar, auf die sich die Meinungsäußerungen beziehen, diese können im Einzelnen Personen, Gegenstände oder auch Firmen sein. Die einzelnen Komponenten eines Objekts samt seiner Eigenschaften

³ Kumar und T 2015, S. 15–16

werden als Features bezeichnet. Unter den Begriffen Meinungsäußerung versteht man eine subjektive Bewertung einer Person zu einem Objekt bzw. dessen Features und diese Person wird dann als Meinungseigentümer bezeichnet. Die Ausrichtung der Meinungsäußerung wird semantische Orientierung oder Polarität genannt, also die Einstufung der Aussage in positiv, negativ oder neutral und damit ist letztendlich die Sentimentanalyse gemeint.⁴

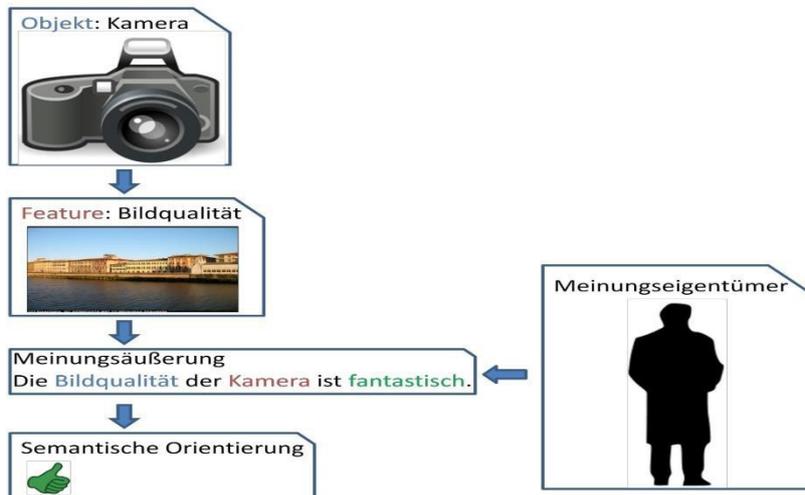


Abbildung 1: Modell des Opinion Mining am Beispiel einer Kamera⁵

Der Ablauf der Opinion Mining Analyse ist ähnlich zu dem des Text Mining und gliedert sich in Datenbeschaffung, Vorverarbeitung und Wissenserwerb (siehe Abbildung 2). Die Datenbeschaffung erfolgt dabei üblicherweise mithilfe von Web Crawlern, HTML-Parsern oder speziellen APIs, die sich auf einen bestimmten Dienst, wie beispielsweise Facebook oder Twitter spezialisiert haben. Eine weitere Parallele zum Textmining ist die Analyse von Fließtexten, welche vermehrt unstrukturiert, aber auch mit Formatierungsinformationen, wie HTML Tags, vorliegen. Die Entfernung dieser unnützen Information, welche die spätere Analyse in ihrer Performance beeinflussen, ist Aufgabe der Vorverarbeitung. In diesen Teil des Prozessablaufs fallen auch klassische Text Mining Werkzeuge, wie Tokenisierung und POS-tagging. Ist die Vorverarbeitung abgeschlossen kann zu dem eigentlichen Wissenserwerb übergegangen werden, wo besonders die Klassifikation der Meinungsausrichtung, also die Sentimentanalyse im

⁴ Valerius 2012, S. 12–14

⁵ Valerius 2012, S. 13

Mittelpunkt steht.⁶ Auf die Phasen im Einzelnen und deren praktische Umsetzung wird dann noch einmal in Kapitel 4 genauer eingegangen.

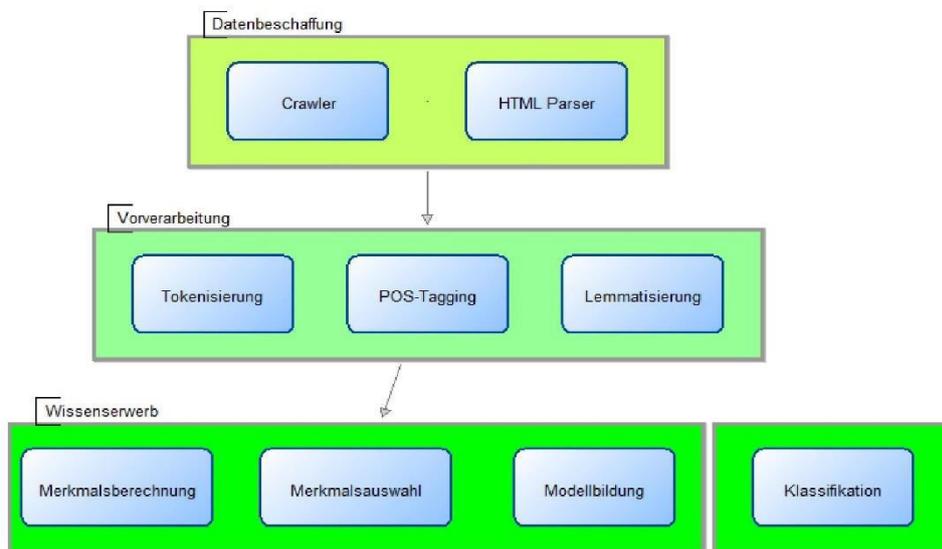


Abbildung 2: Prozess des Opinion Mining⁷

2.1.1 Sentimentanalyse

Die Sentimentanalyse ist, wie bereits erwähnt, für die Einordnung von Aussagen in Stimmungskategorien wie positiv, negativ oder neutral zuständig. Grundsätzlich kann man dabei zwischen semantischen bzw. lexikalischen und lernbasierten Ansätzen unterscheiden. Bei den semantischen Techniken werden Wörter anhand eines Lexikons ihrer Polarität zugeordnet und bei lernbasierten Ansätzen geschieht dies durch manuell annotierte Trainingsdatensätze (Korpusse), deren Muster auf unstrukturierte Daten übertragen werden. Die Zuordnung zu einer Polarität kann dabei entweder eindeutig oder skaliert geschehen. Als Sentimentträger, also Wörter, die eine Meinung verkörpern, kommen dabei Substantive, Adjektive, Verben oder auch Adverbien in Frage.⁸

⁶ Valerius 2012, S. 14

⁷ Valerius 2012, S. 14

⁸ Valerius 2012, S. 36

Sentimentanalyse ist eine effektive Methode, um öffentliche Meinungen zu ermitteln. Als Datengrundlage eignen sich dafür besonders die sozialen Netzwerke, da hier im Gegensatz zu einer klassischen Umfrage eine größere Zielgruppe erreicht werden kann und auch die Wahrscheinlichkeit größer ist, dass die Nutzer ihre ehrliche Meinung äußern. Die Ergebnisse können dann beispielsweise von Unternehmen zu Marketingzwecken eingesetzt werden um ihre Aktivitäten auf Bereiche zu fokussieren, die bisher auf ein eher negatives Meinungsbild stoben.⁹

Die Sentimentanalyse lässt sich auf verschiedenen Ebenen durchführen. Die erste ist die Dokumentenebene, dabei wird der gesamte Text im Bezug auf eine Entität klassifiziert. Da sich jedoch nicht jeder Satz eines Dokuments auf diese Entität bezieht, ist es notwendig, die Sätze in subjektiv bzw. objektiv zu unterteilen, um die irrelevanten Teile auszusortieren. Für diesen Zweck werden oftmals überwachte Lernmethoden, wie Naive Bayes oder SVM verwendet (siehe Kapitel 3.1). Als Trainingsdaten eignen sich dabei einzelne Wörter und deren Häufigkeit, bestimmte Wortformen, wie Adjektive, ganze Phrasen oder auch Verneinungen. Zusätzlich ist bei der Sentimentanalyse auf Dokumentenebene auch noch eine Adaption der Wörter auf bestimmte Domänen zu beachten, da sich deren Bedeutung je nach Kontext unterscheiden kann. Beispielsweise ist das Adjektiv gruselig per se eher negativ annotiert, kann aber bei der Bewertung von Horrorfilmen auch ein positives Werturteil darstellen.¹⁰

Der Sentimentanalyse auf Dokumentenebene liegt, wie bereits erwähnt, die Hypothese zugrunde, dass ein vollständiges Dokument nur eine Entität und einen Autor besitzt. Jedoch ist diese Annahme in der Realität häufig nicht zutreffend, da gerade Texte aus sozialen Medien oftmals mehrere Meinungen unterschiedlicher Autoren beinhalten. Ein weiteres Problem bei dieser Art der Klassifikation ist, dass die Betrachtung der Polarität einzelner Schlüsselwörter nicht ausreicht, da eine einfache Aufsummierung dieser Stimmungen oftmals nicht mit der Gesamtaussage des Dokuments übereinstimmt.

Dementsprechend greift die Forschung häufiger auf die Klassifikation auf Satzniveau zurück, weil das für viele Anwendungsfälle von Vorteil ist. Jedoch liegt auch hierbei die Annahme einer einheitlichen Meinung innerhalb eines Satzes zugrunde, welche in der Praxis nicht immer zutreffend ist, aber zumindest häufiger wie auf Dokumentenebene.¹¹ Bei dieser Methode wird

⁹ Anwar Hridoy et al. 2015, S. 1

¹⁰ Patel et al. 2015, S. 24–25

¹¹ Sprejz, S. 19–20

jeder Satz einzeln analysiert und bewertet. Dazu wird zunächst ein Satz ebenfalls als subjektiv oder objektiv eingestuft, um zu prüfen, ob er überhaupt eine Meinung enthält und danach kann dann seine Polarität ermittelt werden.

Für die Fälle, in denen unterschiedliche Meinungen bzw. sowohl Meinungen als auch Fakten, innerhalb eines Satzes existieren, ist es sinnvoll, diesen noch in seine Entitäten zu zerlegen. Dazu werden diese einzeln extrahiert und die entsprechenden Aspekte (Features) oder Meinungen in Relation gesetzt. Dies hat auch den Vorteil, dass gegebenenfalls Konflikte zwischen Polaritäten innerhalb eines Satzes aufgelöst werden können.¹² Dafür werden oftmals lexikalische Ansätze herangezogen, die jedem Wort einen Punktwert zuordnen, welcher dann für eine bestimmte Entität aufsummiert wird. Dabei sind auch Verneinungen oder gegenteilige Wörter zu berücksichtigen.¹³

2.1.2 Forschungsentwicklung

In einer Studie von Piryani et al. wurde die Entwicklung des Forschungsfeldes Opinion Mining / Sentiment Analysis (OSMA) in den letzten 16 Jahren (2000-2015) betrachtet. Dies entspricht nahezu auch der gesamten Zeitdauer der OSMA Forschung vom Ursprung bis zur Gegenwart. Als Datengrundlage wurden Artikel, Paper, Kommentare, Buch Zusammenfassungen o.ä. gewählt, welche in englischer Sprache verfasst wurden. Insgesamt erhielt man dabei 697 Ergebnisse, von denen sich, nach manueller Prüfung, 488 direkt auf das Forschungsfeld OMSA bezogen. Die übrig gebliebenen Arbeiten wurden dann mit statistischen Verfahren manueller und computergestützter Natur ausgewertet und anhand verschiedener Parameter aufbereitet.¹⁴

¹² Sprejz, S. 19–20

¹³ Patel et al. 2015, S. 24–25

¹⁴ Piryani et al. 2017, S. 123–126

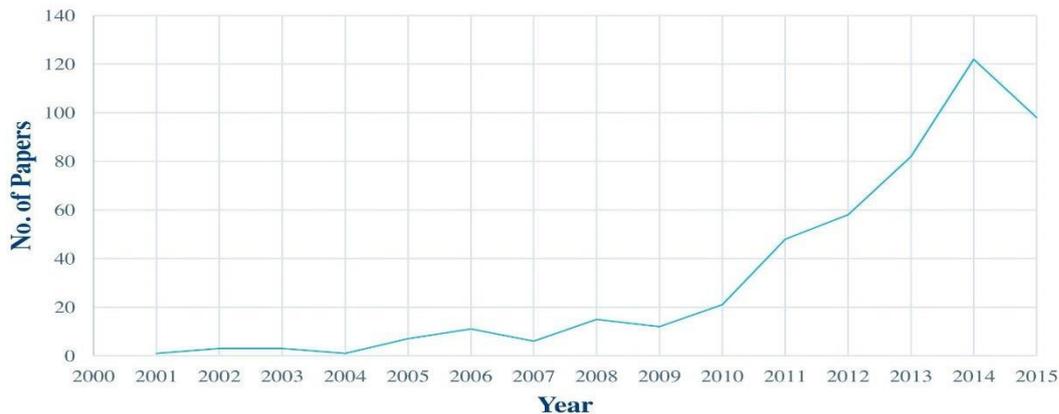


Abbildung 3: Anzahl der jährlichen Publikationen zum Thema OMSA¹⁵

Betrachtet man die Kurve zu der Anzahl der jährlichen Publikationen (*siehe Abbildung 3*), sieht man, dass sie bis ins Jahr 2009 mehr oder weniger flach verläuft, mit einer leichten Tendenz nach oben und danach steil ansteigt. Von 2010-2014 erhöht sich die Anzahl der Publikationen etwa um das sechsfache und es wird deutlich, dass sich OMSA in den letzten Jahren zu einer wichtigen Forschungsdisziplin entwickelt hat, welche in verschiedenen Bereichen Anwendung findet. Der Knick im Jahr 2015 lässt sich dadurch erklären, dass viele Arbeiten aus diesem Jahr zum Zeitpunkt der Studie noch nicht indexiert waren und deshalb in der Statistik nicht auftauchen.¹⁶

S. No	Country/year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
1	China	0	0	0	0	3	2	0	4	6	7	21	10	21	38	33	145
2	USA	1	3	2	0	1	4	4	7	3	6	11	11	19	20	17	109
3	UK	0	0	0	1	1	1	0	2	2	0	5	10	10	4	9	45
4	Spain	0	0	0	0	0	0	0	1	0	0	2	10	10	12	8	43
5	Italy	0	0	0	0	0	2	0	1	0	0	2	5	6	7	8	31
6	Germany	0	0	0	0	0	0	0	0	0	1	5	1	7	3	6	23
7	Japan	0	0	0	0	1	2	0	2	1	0	3	3	3	4	3	22
8	Singapore	0	0	0	0	1	0	0	0	2	1	1	5	3	7	2	22
9	Taiwan	0	0	0	0	0	0	1	0	0	2	2	1	7	5	3	21
10	France	0	0	0	0	0	0	0	0	0	1	2	1	3	6	7	20
11	Canada	0	0	0	0	1	1	0	0	1	1	1	4	3	6	1	19
12	South Korea	0	0	0	0	0	0	0	0	0	1	2	3	2	5	4	17
13	India	0	0	0	0	0	0	0	0	1	0	1	0	1	2	8	13
14	Netherlands	0	0	1	0	0	0	0	0	0	1	0	4	3	2	2	13
15	Australia	0	0	0	0	0	0	1	1	0	0	1	2	1	3	2	11

Abbildung 4: Anzahl der jährlich erscheinenden Publikationen sortiert nach Ländern¹⁷

Betrachtet man die Verteilung der Publikationen nach Ländern wird deutlich, dass die ersten Arbeiten zu OSMA in den USA erschienen sind und diese Forschungsdisziplin in den anderen

¹⁵ Piryani et al. 2017, S. 126

¹⁶ Piryani et al. 2017, S. 126

¹⁷ Piryani et al. 2017, S. 127

Ländern erst nach und nach an Bedeutung gewonnen hat (*siehe Abbildung 4*). Aktuell ist es so, dass die höchste Anzahl an Arbeiten zu diesem Thema in China veröffentlicht wird, was etwa 30 Prozent aller Veröffentlichungen im Zeitraum von 2000-2015 ausmacht. Danach folgt die USA mit etwa 22 Prozent und Deutschland befindet sich aktuell auf Platz 6 (Stand 2016).¹⁸

Jedoch ist Deutsch die am meisten gesprochene Sprache in Europa und inzwischen fokussieren sich auch diverse Forschungsaktivitäten auf deutschsprachige Sentimentanalyse. Darunter ist auch die Interest Group for German Sentiment Analysis (IGGSA), eine europäische Vereinigung von Forschern, die sich mit diesem Thema beschäftigen. Die IGGSA veranstaltet unter anderem Workshops, verteilt Projekte und veröffentlicht eine Reihe von Forschungsarbeiten auf ihrer Website.¹⁹

2.2 Soziale Netzwerke

Die Bezeichnung soziale Medien, wird oftmals als Sammelbegriff für jede Art von nutzer-generiertem Inhalt (user-generated content) verwendet. Eine der wichtigsten Vertreter sind Facebook, Twitter und Blogs, welche sich jedoch in diversen Aspekten unterscheiden. Facebook und Twitter werden im Gegensatz zu Blogs zusätzlich als soziale Netzwerke klassifiziert, da sie es dem Nutzer erlauben, ein Profil anzulegen, sich mit anderen Nutzer in Verbindung zu setzen und auch (teilweise eingeschränkt) die Beziehungen anderer Nutzer zu betrachten. Blogs hingegen sind oftmals nur von informativer Natur, dabei werden von den Nutzern viel weniger Diskussionen über persönliche Belange geführt oder Fotos ausgetauscht wie etwa bei Facebook.²⁰

¹⁸ Piryani et al. 2017, S. 127–128

¹⁹ Mark Cieliebak et al. 2017, S. 46

²⁰ Johnson und Kaye 2015, S. 545

2.2.1 Blog

Ein Blog stellt, wie bereits erwähnt, kein soziales Netzwerk dar und ist somit für diese Arbeit von untergeordneter Bedeutung, jedoch lassen sich seine Eigenschaften auch bei Twitter, in Form eines Mikro Blogs, wiederfinden.

Im Unterschied zu traditionellen Medien hat der Autor beim Verfassen eines Blogs, dem sogenannten Blogging, vollkommene Kontrolle über den Inhalt und das Layout seines Textes. Dies hat aber auch zur Folge, dass man in einen Blog alles, ungeachtet seines Wahrheitsgehaltes, veröffentlichen kann, jedoch auch mit seinem Namen für den Inhalt geradestehen muss. Ein weiterer Unterschied zu textgebundenen Medien besteht darin, dass in einem Blog Links zu anderen Webseiten verwendet und Inhalte in Echtzeit veröffentlicht werden können.²¹

Der Mikro Blog gilt als eine Unterform des regulären Blogs und fungiert eher als soziales Netzwerk, zudem ist er auf 140 bis 200 Zeichen begrenzt. Indem User die Nachrichtenkanäle anderer Nutzer abonnieren können, entsteht eine direkte Verbindung zwischen Autor und Leser. Da ein solches Abonnieren fremder Nachrichten prinzipiell jedem möglich ist, kann durch diese Art des Blogs schnell eine öffentliche Diskussion zu einem Thema ausgelöst werden. Ein weiterer Vorteil eines Mikro Blogs gegenüber einem normalen Blog ist die Geschwindigkeit, mit der Nachrichten versendet werden können. Aufgrund ihrer geringen Länge können sehr viele Nachrichten in kurzer Zeit verbreitet werden, was auch noch durch die mobile Nutzung von Mikro Blogs als Smartphone App gefördert wird. Aufgrund der bereits beschriebenen Eigenschaften eignen sich Mikro Blogs jedoch nicht für ausführliche Erläuterungen, sondern dienen eher dazu kurze Informationen, wie Meinungen, Gedanken oder Verweise zu veröffentlichen. Der Ursprung von Mikro Blogs liegt in der Status Update Funktion von Facebook, welche 2006 eingeführt wurde um seinen Nutzern die Möglichkeit zu geben, das zu schreiben, was sie gerade tun.²²

2.2.2 Twitter

Twitter ist der derzeit größte Anbieter für Mikro Blogging weltweit. Für das Versenden von Nachrichten bei Twitter stehen den Nutzern 140 Zeichen zur Verfügung. Gegründet wurde das

²¹ Valerius 2012, S. 16

²² Valerius 2012, S. 17

Unternehmen im Jahr 2006 als Forschungsprojekt von Jack Dorsey, Biz Stone und Evan Williams bei der Firma Odeo. Grundgedanke war es Textnachrichten über mobile Geräte zu verteilen, wobei zunächst SMS als Übertragungsart genutzt wurde, was die Limitierung des Zeichensatzes auf 140 Zeichen begründete. Die Grundstruktur von Twitter entspricht der eines sozialen Netzwerks, wobei Verbindungen im Gegensatz zu Facebook nur unidirektional möglich sind. Das bedeutet, dass ein Nutzer die Möglichkeit hat, einem anderen Nutzer zu folgen (Following) und damit alle seine Nachrichten zugesendet bekommt, was aber nicht zwangsläufig auf Gegenseitigkeit beruht. Studien haben sogar ergeben, dass eine gegenseitige Kommunikation nur in ca. 22% aller Fälle erfolgt. Zudem werden die Twitter Nachrichten chronologisch geordnet, was eine zeitliche Nachvollziehbarkeit erlaubt und auch ihre Funktion als Status Updates unterstützt.

Die Nachrichten bei Twitter werden Tweets genannt und enthalten ähnlich wie ein Blog eine Reihe von Markierungen, die Hashtags genannt werden. Um Nachrichten zu verschicken, muss man sich bei Twitter registrieren und dann kann man selbst Mitteilungen versenden oder die Tweets anderer Nutzer abonnieren. Die Nutzer, die die Nachrichten anderer User abonniert haben, nennt man Follower und der Verfasser wird dann als Followee bezeichnet. Alle Tweets von Nutzern, den man folgt, werden automatisch in der so genannten Timeline angezeigt. Die Nutzer können untereinander interagieren, indem sie auf Tweets antworten oder sich in ihrer Nachricht auf einen bestimmten Nutzer beziehen. In beiden Fällen geschieht das durch Einbeziehung des @ Symbols in den Tweet. Eine weitere Form der Interaktion ist das Weiterleiten von Tweets, was Retweeting genannt wird. Retweets entsprechen dabei genau dem Original, jedoch wird noch der Name des Nutzers angehängt, der die Nachricht weitergeleitet hat.²³

Wie bereits erwähnt, fördert der Netzwerkcharakter von Twitter auch öffentliche Diskussionen und die Verbreitung von Nachrichten von Gruppen. Möchte man über bereits bestehende Themen diskutieren, lässt sich das mit Hilfe von Hashtags realisieren, welche in Form einer Raute (#) dem Thema vorangestellt werden. Außerdem bietet Twitter auch eine API an, dies es erlaubt, Fremdanwendungen zu integrieren, was zu einer enormen Verbreitung des Service geführt hat. Für den Bereich der Sentimentanalyse ist besonders interessant, dass man diese API auch in Programmiersprachen wie Java einbinden kann, um möglichst viele Nachrichten

²³ Alvanaki 2014, S. 6–7

für die spätere Untersuchung zu sammeln. Inzwischen nutzen nicht nur Privatpersonen Twitter, sondern auch Nachrichtensender, Vereine oder Politiker, wie beispielsweise Barack Obama oder Donald Trump.²⁴

2.2.3 Facebook

Facebook wurde 2004 gegründet und ist inzwischen das beliebteste soziale Netzwerk weltweit. Bereits im Oktober 2012 erreichte es die Schallmauer von einer Milliarde Nutzer (Anzahl Nutzer mit Zugriff innerhalb eines Monats).²⁵

Im Gegensatz zu Twitter sind die Beziehungen zwischen den Nutzern bei Facebook bidirektional angelegt. Unter bidirektional versteht man, dass zunächst eine Freundschaftsanfrage gesendet und diese auch akzeptiert werden muss, bevor eine Nachricht gesendet werden kann. Diese Bestätigung wird bei Twitter nicht benötigt, sodass man jedem folgen kann. Ein weiterer Unterschied besteht darin, dass die Beziehungen bei Facebook häufig zwischen reellen Freunden oder Bekannten geführt werden, wo hingegen die Mehrzahl der Verbindungen bei Twitter zu Fremden bestehen, die zum Teil ihre wahre Identität gar nicht preisgeben. Die Tatsache, dass die Beziehungen bei Facebook untereinander weniger professionell, sondern vielmehr freundschaftliche Natur sind, hat für die Sentimentanalyse den entscheidenden Vorteil, dass damit auch die Nachrichten spontaner und emotionaler verfasst werden.

Diese Nachrichten werden bei Facebook Posts genannt und man unterscheidet vier verschiedene Inhalte. Das sind Statusmeldungen, hochgeladene Fotos, der derzeitige Aufenthaltsort und aktuelle Ereignisse aus dem Leben des Nutzers.²⁶

Der Raum, wo ein Nutzer seine eigenen Nachrichten und Inhalte veröffentlicht, wird dabei "Wand" genannt und diese ist typischerweise für seine Freunde sichtbar, welche dann wiederum Kommentare und Likes hinterlassen können.²⁷

²⁴ Valerius 2012, S. 18–19

²⁵ Johnson und Kaye 2015, S. 545

²⁶ Kuo et al. 2016, S. 402

²⁷ Ortigosa et al. 2014, S. 528

Daran sieht man, dass es bei Facebook möglich ist, Beschränkungen einzuführen, um das eigene Profil nur teilweise für andere sichtbar zu machen, während es bei Twitter komplett öffentlich ist. Facebook Nutzer sind somit stärker sozial verbunden als Twitter Nutzer und sie unterscheiden zu zudem in Persönlichkeitsmerkmalen, Geschlecht oder Ausbildung.²⁸

2.3 Forensik

Im folgenden Abschnitt wird die Forschungsdisziplin der Forensik kurz definiert und in das System der Wissenschaften eigenordnet, um sie später als Anwendungsbereich für die Sentimentanalyse zu betrachten. Die Forensik beschäftigt sich im Allgemeinen mit der Anwendung von wissenschaftlichen Methoden auf Fragen des Rechtssystems. Die Bezeichnung Forensik ist vom lateinischen Wort "forensis" abgeleitet, was so viel bedeutet, wie auf dem Marktplatz. Das bezieht sich auf den historischen Hintergrund, wo Gerichtsverfahren oftmals auf dem Marktplatz ausgetragen wurden. Heute bezeichnet man mit Forensik alle Arbeitsgebiete, die strafrechtlich und zivilrechtlich relevante Handlungen identifizieren, ausschließen, analysieren und rekonstruieren. Während Forensik in den USA eine eigene Fachrichtung darstellt, ist es

in Deutschland in das Gebiet der Kriminalwissenschaften eingebettet und stellt eine Verbindung von Kriminologie und Kriminalistik dar (siehe Abbildung 5). Typische Beispiele für forensische Fachgebiete sind die Gerichtsmedizin, Kriminalpsychologie oder die forensische Psychiatrie.²⁹

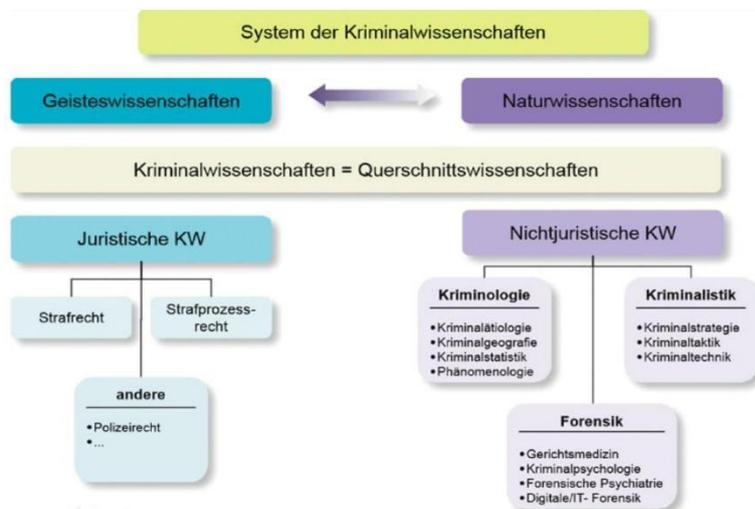


Abbildung 5: System der Kriminalwissenschaften nach Barthel³⁰

²⁸ Johnson und Kaye 2015, S. 545

²⁹ Labudde und Spranger 2017, S. 5–6

³⁰ Labudde und Spranger 2017, S. 6

Dazu zählt aber auch die noch junge Disziplin der digitalen Forensik, welche die Aufgabe hat, digitale Spuren zu sichern und diese für die Verwendung bei Gerichtsprozessen aufzubereiten und auszuwerten. Weitere synonyme Bezeichnungen für digitale Forensik sind Computer- oder IT-Forensik.

Durch die Verwendung wissenschaftlich akzeptierter Methoden wird ein größtmögliches Maß an Objektivität bei forensischen Untersuchungen garantiert. Dies ist notwendig, da Fehler bei der Durchführung oder das Ziehen von falschen Schlüssen zu erheblichen Konsequenzen für die Existenz, die Freiheit oder das Leben von Menschen haben können.³¹

³¹ Meier, S. 14–15

3 Ansätze der Sentimentanalyse

Grundsätzlich kann man im Bereich der Sentimentanalyse zwischen lernbasierten Ansätzen, lexikalisch-basierten Ansätzen und hybriden Ansätzen unterscheiden. Lernbasierte Ansätze verwenden dabei gelabelte Daten, lexikalisch-basierte Ansätze vertrauen auf Lexika bzw. zusätzlich noch ungelabelte Daten und hybride Ansätze stellen eine Mischung aus beidem dar. In der vorliegenden Arbeit wurde sich im Wesentlichen auf lernbasierte und lexikalische Ansätze beschränkt.

Allen diesen Ansätzen ist jedoch gemein, dass im ersten Schritt eine Vorverarbeitung der Daten unerlässlich ist, bevor es zu der eigentlichen Analyse kommt. Dies ist notwendig, um irrelevante Teile der Daten zu entfernen und den Text in eine für die Analyse geeignete Form zu transformieren. Irrelevante Teile der Daten könnten zum Beispiel HTML Tags, Skripte oder Werbung sein, welche die Performance und Genauigkeit der späteren Analyse beeinträchtigen. Zudem sollte in der Vorverarbeitungsphase eine Normalisierung der Daten durchgeführt werden, da die zu untersuchenden Texte, besonders in den sozialen Netzwerken, oftmals nicht den formalen und grammatikalischen Anforderungen gewöhnlicher Sprache entsprechen und damit auch schwer zu analysieren sind. Um dies zu realisieren werden oftmals spezielle mehrsprachige Lexika herangezogen, die gezielt auf die Vorverarbeitung von Texten aus sozialen Netzwerken ausgerichtet sind. Den wichtigsten Teil der Vorverarbeitung der Daten bilden aber Techniken aus der natürlichen Sprachverarbeitung wie Tokenization, Splitting, Entfernung von sogenannten Stop-Words oder Stemming. Tokenization beispielsweise teilt den Text in Wörter und Symbole auf, wohingegen das Splitting Satzgrenzen erkennt. Stop-words sind für gewöhnlich Wörter, die keine Meinung transportieren und somit für die Sentimentanalyse keinen Mehrwert darstellen und unter Stemming versteht man das Bilden von Stammformen der Wörter, um keine unnötigen Redundanzen entstehen zu lassen.³² Auf die genaue Umsetzung der Vorverarbeitung und die damit verbundenen spezifischen Eigenheiten der sozialen Netzwerke wird in Kapitel 4.1 noch einmal genauer eingegangen.

³² Dashtipour et al. 2016, S. 758–759

3.1 Lernbasierter Ansatz

Zunächst soll der lernbasierte Ansatz vorgestellt werden. Dieser nutzt zur Klassifikation Methoden des Information Retrieval, des Data Mining und kann in den Bereich maschinellen Lernens eingeordnet werden. Maschinelles Lernen beruht dabei im Wesentlichen auf dem Prinzip, dass Entscheidungen aufgrund gesammelter Erfahrungen getroffen werden aus denen Muster oder Regeln abgeleitet werden, mit dem Ziel Daten automatisiert in Zielgruppen einordnen zu können. Die Erfahrungen, die gesammelt werden, beziehen sich dabei auf die Merkmale der Daten bzw. Objekte, welche qualitativer oder quantitativer Natur sein können. Quantitativ bedeutet in dem Zusammenhang, dass ein Merkmal üblicherweise jeden Wert innerhalb eines Intervalls annehmen kann, wie z.B. Körpergröße oder Temperatur. Ein qualitatives Merkmal wiederum zeichnet aus, dass es nur bestimmte Werte oder Zustände annehmen kann, wie beispielsweise bei Monats- oder Geschlechtsangaben.

Um die Klassifikation mithilfe eines lernbasierten Ansatzes durchzuführen, ist es üblich eine Datenmatrix zu erstellen, wo alle Objekte mit ihren Ausprägungen eingetragen werden. Die Merkmale werden dabei oftmals mit Zahlen kodiert (z.B. blau = 1, gelb = 2 usw.), um die Objekte besser miteinander vergleichen zu können. Ausgehend von dieser Matrix wird nun versucht die Ähnlichkeit bzw. Unähnlichkeit (Distanz) einzelner Objekte zu berechnen, um diese daraufhin einer Klasse zuordnen zu können. Für die Berechnung der Ähnlichkeit gibt es verschiedene Methoden, welche als Ähnlichkeits- bzw. Distanzmaße bezeichnet werden. Dabei wird untersucht, wie nahe die Eigenschaften der Objekte beieinanderliegen und man geht davon aus, dass ähnliche Objekte den gleichen und unähnliche Objekte verschiedenen Gruppen angehören. Oftmals wird das Distanzmaß noch auf das Intervall $[0;1]$ skaliert, wobei 1 die maximale Ähnlichkeit darstellt und 0 die maximale Unähnlichkeit. Da die Sentimentanalyse im Wesentlichen auf binären Ähnlichkeitsmodellen aufgebaut ist, das heißt eine Eigenschaft ist entweder vorhanden oder nicht, können auch nur die Zustände null oder eins auftreten. Die verschiedenen Maße untersuchen damit die Eigenschaften nur darauf, ob sie gleich oder ungleich sind und unterscheiden sich nur in der Art und Weise der Messung. Ähnlichkeitsmaße vergleichen Wertekombination mit gleichen Werten wie 1/1 oder 0/0 und Distanzmaße bewerten hingegen die Wertepaare 0/1 und 1/0.³³

³³ Valerius 2012, S. 42–43

Das am häufigsten verwendete Distanzmaß ist die euklidische Distanz (*Formel siehe Abbildung 6*). Dabei entspricht der Abstand zweier Punkte der direkten Linie zwischen ihnen in einem Vektorraum (*siehe Abbildung 7*). Die Distanz lässt sich mit folgender Formel berechnen.

$$\text{Euklidische Distanz : } d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Abbildung 6: Formel der Euklidischen Distanz³⁴

Überträgt man dies auf binäre Merkmale, müssen nur noch ungleiche Wertepaare verglichen werden, da bei gleichen Merkmalen eine Distanz von Null entsteht.³⁵

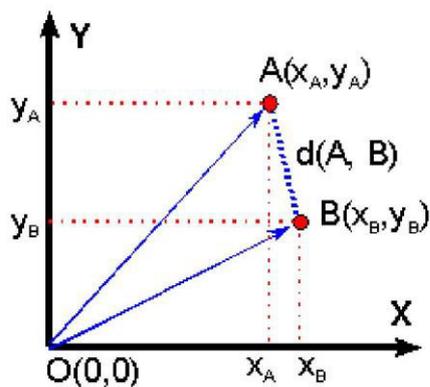


Abbildung 7: Modell der Euklidischen Distanz³⁶

Maschinelles Lernen spielt auch bei der Textklassifikation eine Rolle. Hierbei geht man davon aus, dass ein Dokument (bzw. ein Satz oder ein Wort) zu einem höher dimensional Raum von Dokumenten (bzw. Sätzen oder Wörtern) gehört, welcher verschiedene Klassen beinhaltet. Aufgabe des maschinellen Lernens ist es nun, die Dokumente mit Hilfe von Beispieldaten den Klassen zuzuordnen und dies wird dann später an neuen Dokumenten getestet. ³⁷

³⁴ Valerius 2012, S. 44

³⁵ Valerius 2012, S. 44–45

³⁶ Valerius 2012, S. 44

³⁷ Sprejz, S. 21–22

Bei der Klassifikation mittels lernbasierten Algorithmen geht es nun, wie bereits erwähnt, darum Daten aufgrund ihrer Eigenschaften in bestimmte Gruppen einzuordnen. Im Falle der Sentimentanalyse sind dies positive, negative und gegebenenfalls neutrale Dokumente bzw. Sätze. Für eine solche Klassifikation bedarf es zunächst Trainingsdaten, also bereits annotierte Dokumente. Diese können entweder selbst erstellt werden oder man greift auf bereits vorhandene Lösungen zurück. Diese werden auch als Korpus bezeichnet und man spricht bei diesen manuell erstellten Trainingsdaten auch von "supervised learning". Eine weitere Möglichkeit Trainingsdaten zu erzeugen, ist die Berücksichtigung der im Text verwendeten Emoticons (z.B. :-) , :-[) oder Hashtags (z.B. #fail), welche ein Dokument oder einen Satz relativ sicher als positiv oder negativ klassifizieren (siehe auch Kapitel 4.3.2). Man spricht in diesem Fall von semiautomatischer Klassifikation.

Die Aufgabe des Klassifikators ist es nun, anhand der Trainingsdaten Entscheidungsregeln abzuleiten, mit der sich unbekannte Daten in die Zielgruppen einteilen lassen. Dies wird als Generalisierungsfähigkeit bezeichnet und zeichnet einen guten Algorithmus aus. Damit dies geschieht soll er nicht zu spezifisch auf die Trainingsdaten ausgerichtet werden, denn sonst entsteht eine Überspezialisierung, das sogenannte Overfitting. Damit ist gemeint, dass der Algorithmus zwar in der Lage ist eine Zielgruppe sehr genau vorauszusagen, aber auch nur noch auf eine geringe Zahl von Datensätzen anwendbar ist.³⁸

Techniken des maschinellen Lernens greifen noch auf weitere Kennzahlen zurück, welche besonders bei der Vorverarbeitung von Bedeutung sind. Zunächst spielt hier der Begriff des N-grams eine wesentliche Rolle. Damit sind aufeinanderfolgende Sequenzen der Anzahl n im Text gemeint, wobei die Anzahl eins, unigram, die Anzahl zwei bigrams und die Anzahl drei trigrams genannt wird. Ein weiterer wichtiger Begriff ist die Dokumentenfrequenz, die die Gesamtanzahl der Dokumente bezeichnet, die ein bestimmtes Wort beinhalten. Für die Dokumentenfrequenz wird zusätzlich ein Grenzwert bestimmt, der entscheidet ob während der Vorverarbeitung ein Wort als Stop-Word entfernt wird oder nicht. Darüber hinaus stellt auch die Termfrequenz (TF) eine bedeutsame Kennzahl dar. Diese bezeichnet die Anzahl des Erscheinens eines bestimmten Wortes innerhalb eines Dokuments. Diese wird oftmals in Zusammenhang mit der inversen Dokumentenfrequenz verwendet, um häufig auftretende Wörter entsprechend zu gewichten (TF-IDF Weighting). Weitere Maße, die für maschinelles Lernen eingesetzt werden, sind die sogenannte Mutual Information (MI), die die Abhängigkeit von Variablen

³⁸ Valerius 2012, S. 45–46

zueinander bestimmt und das Information Gain (IG), dass die Menge von Informationen misst, die mit der Anwesen- oder Abwesenheit eines Wortes verbunden sind.³⁹

Ein bekanntes Werkzeug, welches zur Sentimentanalyse mithilfe lernbasierter Ansätze genutzt werden kann, ist WEKA (Waikato Environment for Knowledge Analysis). Dies ist ein frei verfügbares Tool, welches in Java geschrieben wurde und eine große Anzahl an Algorithmen des maschinellen Lernens und des Data Mining bereithält. Ein weiteres Werkzeug ist LIBSVM, eine Bibliothek, welche es dem Nutzer ermöglicht relativ einfach den Support Vector Maschine (SVM) Algorithmus zu implementieren.⁴⁰ Auf einige der wichtigsten Klassifikatoren wird in den folgenden Abschnitten genauer eingegangen.

3.1.1 Naive Bayes

Naive Bayes ist ein Klassifikator, der die Zuordnung von Datensätzen anhand bedingter Wahrscheinlichkeiten vornimmt. Diese Methode wurde vom englischen Mathematiker Thomas Bayes entwickelt und die Bezeichnung Naive beruht auf der Annahme, dass alle Eigenschaften eines Datensatzes voneinander unabhängig sind. Dies ist in der Praxis zwar unrealistisch, dennoch konnte der Algorithmus gute Ergebnisse erzielen.

Naive Bayes ist ein sehr beliebter Algorithmus, da er einfach und effizient ist und sich damit auf verschiedene reale Probleme anwenden lässt. Naive Bayes gehört zur Gruppe der überwachten und statistischen Lernmethoden und arbeitet mit einem Wahrscheinlichkeitsmodell, dass das Auftreten unsicherer Variablen voraussagt (*siehe Formel in Abbildung 8*).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Abbildung 8: Formel des Naive Bayes Algorithmus⁴¹

³⁹ Dashtipour et al. 2016, S. 760

⁴⁰ Dashtipour et al. 2016, S. 760

⁴¹ Gull et al. 2016, S. 1567

Im Fall der Sentimentanalyse ist dies die Wahrscheinlichkeit, dass ein Wort und in der Summe ein ganzes Dokument als positiv bzw. negativ eingestuft wird. Der Vorteil in der Verwendung dieser Methode liegt darin, dass er sich mit relativ wenig Trainingsdaten sehr gut trainieren lässt.⁴²

Bei der Berechnung betrachtet man eine Menge von Trainingsdaten mit d Datensätzen, die jeweils n Attribute besitzen, eingeteilt in k Klassen. Für jedes n -te Attribut wird nun eine Wahrscheinlichkeit durch die Häufigkeit seines Auftretens im Verhältnis zum Auftreten einer Klasse in der Gesamtmenge der Dokumente berechnet. Dies geschieht mit folgender Formel.

$L[n] = \text{Menge an Dokumenten in denen } n \text{ und } k \text{ vorkommen} / \text{Menge aller Dokumente in denen } k \text{ vorkommt.}$

Dieser Wahrscheinlichkeitswert wird nun für alle Attribute berechnet. Die finale Wahrscheinlichkeit der Zugehörigkeit eines Dokumentes zu einer Klasse berechnet sich schließlich aus dem Produkt aller Attribut Wahrscheinlichkeiten, multipliziert mit der Häufigkeit des Vorkommens einer Klasse in allen Dokumenten.⁴³

Vorteile von Naive Bayes sind wie bereits erwähnt seine einfache Interpretation und seine effiziente Arbeitsweise. Nachteilig ist die Annahme, dass alle Attribute voneinander unabhängig sind, was oftmals nicht der Realität entspricht.⁴⁴

3.1.2 Support Vector Machine (SVM)

Eine weitere Methode sind Support Vector Machines (SVM), welche in der Lage sind 2 Bereiche möglichst exakt voneinander zu trennen. Aufgrund seiner Eigenschaften gehört SVM in die Gruppe überwachter Lernverfahren.⁴⁵

Dieser Klassifikator bestimmt die Klasse von Objekten anhand ihrer Position in einem n -dimensionalen Vektorraum, wobei n die Anzahl der Objekte darstellt. Als Ausgangspunkt wird

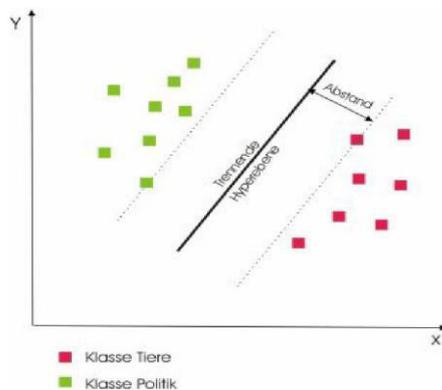
⁴² Patel et al. 2015, S. 25

⁴³ Valerius 2012, S. 47–48

⁴⁴ Singh und Shahid Husain 2014, S. 14

⁴⁵ Patel et al. 2015, S. 25–26

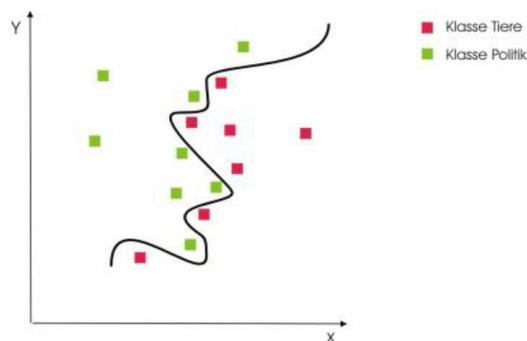
dabei wie bei den meisten lernenden Algorithmen eine Menge von Trainingsdaten genutzt, deren Klassenzugehörigkeit bereits bekannt ist. Ziel ist es nun in dem Vektorraum eine Linie



oder Trennfläche zu finden, welche die einzelnen Klassen möglichst sauber voneinander trennt und deren Abstand zueinander maximal ist. Eine solche Fläche nennt man Hyperebene (*siehe Abbildung 9*).

Abbildung 9: Prinzip des SVM Verfahrens⁴⁶

SVM berechnet eine Hyperebene, die so im Raum liegt, dass der Abstand zu allen Klassen maximal ist. Oftmals ist es aber der Fall, dass die einzelnen Vektoren so zueinander im Raum liegen, dass sie nicht so einfach voneinander zu trennen sind. In der Abbildung beispielsweise



wäre dies nur durch eine schlangenförmige Linie möglich, deren Funktion jedoch schwer zu bestimmen ist (*siehe Abbildung 10*).

Abbildung 10: Problem bei der Klassifikation von Dokumenten⁴⁷

Überführt man das Problem jedoch in einen höher dimensional Raum reicht zur Trennung eine einfache gerade Ebene aus, welche mithilfe des Stützvektors (Support Vector) berechnet

⁴⁶ Zelch und Engel 2005, S. 4

⁴⁷ Zelch und Engel 2005, S. 4

werden kann (siehe Abbildung 11). Dies ist der Vektor, der senkrecht auf der Hyperebene steht und auf das am nächsten gelegene Dokument (bzw. Wort oder Satz) zeigt.

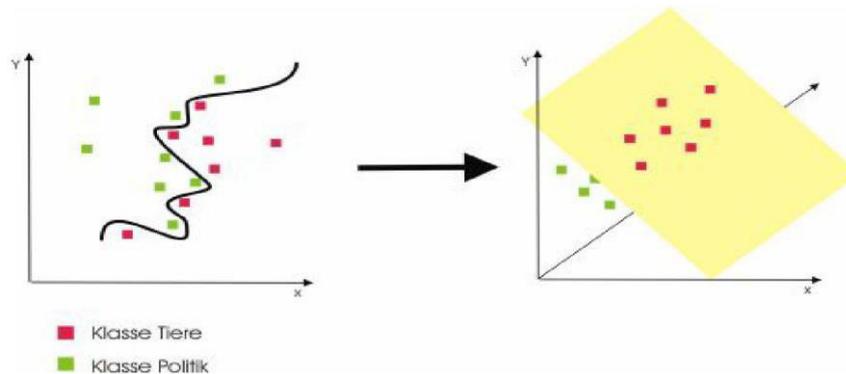


Abbildung 11: Überführen des Problems von SVM in einen höherdimensionalen Raum⁴⁸

Eine einzelne Dimension kann dabei als Eigenschaft eines Objektes verstanden werden. Im Falle von Texten sind dies die Wörter, wobei jedes Wort eine eigene Dimension darstellt. Da man jedoch das Vorhandensein jedes Wortes einzeln prüfen muss, liegen stets lediglich binäre Dimensionen vor.⁴⁹

Später werden die Hyperebenen wieder rücktransformiert und es entstehen gegebenenfalls nicht lineare Trennflächen, welche die Klassen jedoch sauber voneinander trennen.

Für ein neues Dokument muss dann nur noch sein Vektor berechnet und der Abstand zu dem Stützvektor bestimmt werden, um das Dokument einzuordnen. Der Vorteil dieses Verfahrens ist, dass damit Overfitting vermieden wird, da die Komplexität der Klassen schon im Algorithmus berücksichtigt wird.⁵⁰ Zudem zeichnet sich SVM durch seine gute Performance bei experimentellen Ergebnissen und seine geringe Abhängigkeit vom Testdatensatz aus. Nachteilig ist, dass bei fehlenden oder unregelmäßigen Daten eine Vorverarbeitung notwendig ist und das resultierende Modell oftmals schwer zu interpretieren ist.⁵¹

⁴⁸ Zelch und Engel 2005, S. 5

⁴⁹ Valerius 2012, S. 46

⁵⁰ Zelch und Engel 2005, S. 4–5

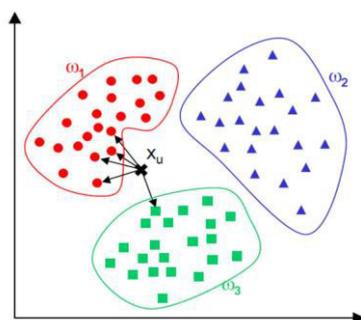
⁵¹ Singh und Shahid Husain 2014, S. 16

3.1.3 Maximale Entropie

Ein weiterer Klassifikator, der häufig zur Sentimentanalyse herangezogen wird, ist die maximale Entropie. Mit Entropie wird in diesem Fall das Maß an Unsicherheit bezeichnet, sodass eine geringe Entropie einen hohen Grad an Sicherheit darstellt und umgekehrt. Im Bereich der Textklassifikation wird die Entropie auf Wahrscheinlichkeitsverteilungen angewendet, welche ähnlich wie bei Naive Bayes aus diversen Variablen, wie Wörtern, Sätzen und deren Auftretenswahrscheinlichkeit bestehen. Ziel ist es eine Verteilung von Objekten in Klassen zu finden, die die maximale Entropie besitzt, da dann davon auszugehen ist, dass auch alle Fakten berücksichtigt und die Wahrscheinlichkeiten möglichst gleichmäßig verteilt sind.⁵²

3.1.4 Cluster

Cluster Klassifikatoren wie beispielsweise der Algorithmus K-Nearest Neighbor (KNN) sind instanzbasiertes Lernverfahren, welche die Zuordnung von Objekten zu Klassen anhand ihrer



Nähe zu anderen Objekten ermitteln. Ähnlich wie bei den Support Vektor Maschinen werden auch hier die Objekte als Vektoren in einem höher dimensionalen Vektorraum betrachtet, deren Distanz zueinander bestimmt wird (*siehe Abbildung 12*). Dabei wird bevorzugt auf das euklidische Distanzmaß zurückgegriffen (Formel siehe Kapitel 3.1).⁵³

Abbildung 12: KNN Klassifikation⁵⁴

Der hauptsächliche Vorteil bei der Verwendung von Cluster Algorithmen ist, dass er in der Lage ist Klassen oder Gruppen nahezu optimal voneinander zu trennen. Nachteilig ist jedoch, dass das Verfahren nicht so gut lernfähig ist und die Anzahl der zu bestimmenden Klassen für gewöhnlich unbekannt ist.⁵⁵

⁵² Valerius 2012, S. 49

⁵³ Valerius 2012, S. 47

⁵⁴ Valerius 2012, S. 47

⁵⁵ Singh und Shahid Husain 2014, S. 19

3.2 Lexikalischer Ansatz

Der lexikalische Ansatz verfolgt die grundsätzliche Idee alle meinungstragenden Worte des Textes mit einem semantischen Lexikon zu vergleichen, wo deren Stimmungsrichtung (Polarität) verzeichnet ist. Bei dieser Vorgehensweise wird auf das Bag of Words Modell zurückgegriffen, das bedeutet, dass alle Wörter im Text als eine ungeordnete Menge betrachtet und die Reihenfolge dabei vernachlässigt wird.

Für die Erstellung eines solchen semantischen oder auch Opinion Lexikon gibt es verschiedene Möglichkeiten. Ein wichtiger gemeinsamer Grundaspekt ist jedoch, dass stets der Anwendungskontext berücksichtigt werden muss, da dieser entscheidenden Einfluss auf die verwendeten Wörter hat. Zudem können Wörter in diversen Kontexten unterschiedliche Bedeutung und damit eine andere Polarität haben. Beispielsweise kann der Begriff "gering" bei dem Thema Kraftstoffverbrauch positiv besetzt sein und bei der Einschätzung von Leistungen eher negativ. Eine Möglichkeit ist es die entsprechenden Wörter manuell zu suchen und einzutragen. Diese manuelle Variante stellt jedoch einen enormen Zeit- und Arbeitsaufwand dar und ist zu einem gewissen Grad subjektiver Natur. Ganz davon abgesehen, dass ein Themenbereich nie ganz abschließend bearbeitet werden kann, was aber ein generelles Problem bei der Verwendung von Wörterbüchern darstellt.

Eine weitere Möglichkeit ist es ein semantisches Lexikon semi-automatisch zu erstellen. Dabei wird zunächst manuell eine Gruppe eindeutig positiv oder negativ besetzter Wörter bestimmt, die sogenannten Seeds, und diese sollen dann selbstständig, ähnlich einem Samen, selbstständig anwachsen. Diese Wörter müssen jedoch gut überlegt sein, da von ihnen die Qualität des späteren Lexikons abhängt, deshalb sollten möglichst eindeutige und kontextunabhängige Begriffe gewählt werden. Um diese Samenkörner zu vermehren, wird daraufhin automatisiert in Lexika nach Synonymen der Wörter gesucht und diese werden dem Wörterbuch hinzugefügt. Ein Nachteil dieser Methode ist jedoch, dass die Ähnlichkeit eines Wortes mit der Distanz zum ursprünglichen Samenkorn abnimmt, laut Studien sogar exponentiell, was dazu führt, dass nicht mehr gewährleistet werden kann, dass alle Wörter entlang eines solchen Pfades die gleiche Polarität besitzen. Godbole et al. konnten in einer Studie nachweisen, dass die meisten Wörter auf einem Pfad, ähnlich einer Normalverteilung, nicht eindeutig einer Meinungsrichtung zuzuordnen sind und daher nicht für ein Opinion Lexikon verwendet werden sollten. Besser sind nach Meinung der Autoren die Wörter, die sich am Rand der Normalverteilung befinden und sich deshalb eindeutig einer Polarität zuordnen lassen. Zusätzlich zu dieser Methode, können auch noch die Stärke der Pfade bei der Erstellung eines semantischen

Wörterbuchs berücksichtigt werden, was mithilfe von Gewichtungen, ähnlich dem Page Rank Algorithmus von Google, geschehen kann.⁵⁶

Die meist verwendete Möglichkeit zur Erstellung eines Opinion Lexikons greift auf bereits vorhandene Quellen zurück, wie das so genannte WordNet, welches von der Princeton Universität entwickelt wurde. Dies ist eine große Datenbank in englischer Sprache, welche Wörter nach ihrer Wortart gruppiert und Begriffe ähnlicher Bedeutung oder lexikalischer bzw. semantischer Verbindung verlinkt. Für die Sentimentanalyse existiert inzwischen eine eigene aufbereitete Form, das SentiWordNet, wo allen Wörtern bereits ihre Polarität (positiv, negativ oder neutral) zugewiesen wurde. Die Zuordnung zu einer Stimmung greift dabei auf die Verlinkungen zurück und weist allen Synonymen eines Wortes die gleiche Polarität zu. Dies geschieht in Form eines numerischen Wertes, der zwischen null und eins liegt, wobei auch die Summe aller Begriffe des SentiWordNet insgesamt eins ergibt. Möchte man nun die gesamte Polarität eines Satzes oder eines Dokuments bestimmen, werden einfach alle positiven und negativen Werte aufsummiert. Studien von Ohana et al. haben ergeben, dass das SentiWordNet mit einem positiven Vorhersagewert (Precision) von 65,85% ähnliche Werte erreicht, wie manuelle Opinion Wörterbücher, deren Genauigkeit die Autoren auf 69 % taxieren. Das SentiWordNet bildet somit die Grundlage für die meisten späteren semantischen Wörterbücher in den diversen Sprachen. Speziell auf die Lexika in deutscher Sprache soll in den nächsten Unterabschnitten noch einmal genauer eingegangen werden. Im Vergleich zu anderen semantischen Wörterbüchern, lässt sich feststellen, dass die Bestimmung der Polarität konsistent ist und sich nur quantitativ unterscheidet. Einen Vorteil ihrer Methode sehen die Autoren in der Verwendung von Phrasen, welche die Behandlung von Verneinungen und Verstärkungen vereinfacht.

Neben diesem einfachen Ansatz, wo Wörter und deren Polarität aufsummiert werden, existieren noch weitere semantische Methoden, wie die von Casey Whitelaw vorgeschlagenen Appraisal Groups. Dieser Begriff stammt aus der Emotionstheorie und bezeichnet eine Gruppe von Wörtern, die kontextunabhängig Meinungen verkörpert, wie beispielsweise "nicht sehr gut" oder "absolut hervorragend". Zusätzlich berücksichtigen Whitelaw et al. noch sogenannte Appraisal Modifier, also meinungsverändernde Wörter, wie "nicht", "sehr" oder "weniger", stellten jedoch fest, dass diese keinen signifikanten Einfluss auf die Genauigkeit der Klassifikation haben, da sie nur eine sehr geringe Menge aller Wörter ausmachten.⁵⁷

⁵⁶ Godbole et al. 2007, S. 1

⁵⁷ Whitelaw et al. 2005, S. 1

Ding et al. haben ähnlich wie Whitelaw versucht kontextunabhängige Wörter zu finden, mit dem Ziel ein universales Sentiment-Lexikon zu erstellen. Als Grundlage nutzten sie dabei das bereits vorgestellte "feature-based" Opinion Mining, also den Zusammenhang zwischen meinungstragenden Wörtern und den zugeordneten Features von Objekten. Die Polarität eines Satzes oder Dokuments wird dann anhand der Sentiment-Wörter in der Nähe eines Features bestimmt und verglichen, ob positive oder negative Stimmungsrichtungen überwiegen. Da diese Bestimmung anhand des Zusammenhangs von Sentiment-Wort und Feature teilweise nicht eindeutig ist, schlagen die Autoren die Einbeziehung von externen Informationen und Texten vor. Zusätzlich wurde in dieser Untersuchung eine Unterscheidung zwischen impliziten und expliziten Meinungsäußerungen getroffen, also die Art und Weise wie offensichtlich eine Meinung in einem Satz zum Ausdruck kommt.⁵⁸

3.2.1 SentiWS

Remus et al. veröffentlichten 2010 ein Wörterbuch zur deutschsprachigen Sentimentanalyse, genannt Sentiment Wortschatz oder kurz SentiWS. Dieses beinhaltet sowohl positiv, als auch

		Positive	Negative
Adjectives	Baseforms	784	698
	Inflections	11,782	10,604
Adverbs	Baseforms	6	4
	Inflections	0 ³	0 ³
Nouns	Baseforms	584	686
	Inflections	521	806
Verbs	Baseforms	312	430
	Inflections	2,453	3,100
All	Baseforms	1,650	1,818
	Inflections	14,756	14,510
	Total	16,406	16,328

negativ besetzte Wörter inklusive ihren Beugungsformen, welche in einem Intervall von -1 bis 1 gewichtet werden. Insgesamt stehen 16406 positive und 16328 negative Wortformen zur Verfügung (Stand Version 1.8b). Dazu zählen nicht nur Adjektive und Adverben, sondern auch Substantive und Verben (*siehe Abbildung 13*).⁵⁹

Abbildung 13: Überblick über den Inhalt von SentiWS⁶⁰

Ein Eintrag im Wörterbuch beinhaltet zu dem jeweiligen Wort seinen Part of Speech Tag, also seine Wortgruppe, welche nach dem Stuttgart-Tübingen-Tagset (STTS) bezeichnet wurde und sich stets auf die entsprechende Grundform bezieht. Darüber hinaus enthält ein Eintrag noch

⁵⁸ Valerius 2012, S. 38–41

⁵⁹ Remus et al. 2010, S. 1168

⁶⁰ Remus et al. 2010, S. 1168

Word	POS Tag	Weight	Inflections
harmonisch	ADJX ²	+0.5243	harmonische, harmonischst
Krise	NN	-0.3631	Krisen

die gewichtete Polarität und gegebenenfalls gebeugte Formen des Wortes (*siehe Abbildung 14*).⁶¹

Abbildung 14: Schema eines Eintrags in SentiWS⁶²

Um die semantische Orientierung von Wörter in SentiWS zu bestimmen wurde im Wesentlichen auf drei Quellen zurückgegriffen. Zum einen ist dies das General Inquirer Lexikon (GI), welches von Stone et al. 1966 veröffentlicht wurde. Dessen Wörter aus den Kategorien positiv bzw. negativ wurden mithilfe des Google Übersetzers ins Deutsche übertragen und manuell auf ihre Übereinstimmung der Polarität geprüft. Darüber hinaus wurden in bestimmten Bereichen (z.B. Finanzwesen) noch Wörter hinzugefügt. Die zweite Quelle waren Ergebnisse aus einer Studie zu Wörtern, die oft mit bestimmten Meinungen einhergehen. Dabei wurden die 200 häufigsten Wörter in positiven bzw. negativen Meinungen identifiziert und gegebenenfalls dem Wörterbuch hinzugefügt. Als dritte Quelle wurde das German Collocation Dictionary herangezogen, welches unter anderem Wörter gruppiert, die häufig mit bestimmten Substantiven auftreten. Daraus wurden die Wörter herausgesucht, die mit einer Meinung einhergehen, was auf ca. 27 % aller Gruppen (6932 Gruppen) zutraf.⁶³

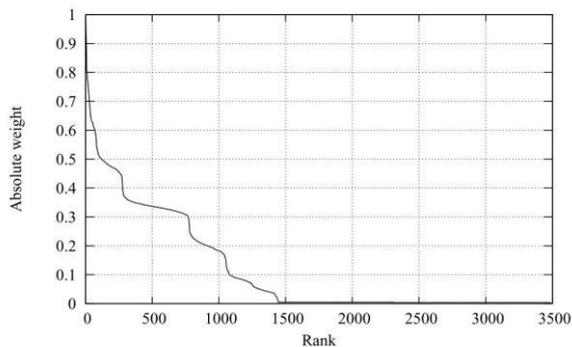
Zur Bestimmung der gewichteten Polarität wurde die so genannte punktweise Mutual Information (PMI) zurate gezogen. Diese bestimmt die semantische Orientierung (positiv bzw. negativ) eines Wortes aus seiner semantischen Beziehung zu anderen negativ bzw. positiv besetzten Worten. Der dadurch berechnete absolute Wert kann dann als Stärke der semantischen Orientierung verstanden werden. Die Werte wurden für jedes Wort aus dem Wörterbuch berechnet, zudem auf das Intervall [-1;1] skaliert und auf maximal 4 Dezimalstellen gerundet. Ein Wert von +1,0 bedeutet demnach absolut positiv und - 1,0 absolut negativ. Die Verteilung der

⁶¹ Remus et al. 2010, S. 1168–1169

⁶² Remus et al. 2010, S. 1168

⁶³ Remus et al. 2010, S. 1169

absoluten gewichteten Polarität der Wörter folgt dabei dem Gesetz von Zipf, d.h. ein paar wenige Wörter haben ein sehr hohes Gewicht, einige haben einen mittleren Wert und die größte



Menge hat eine niedrige Polarität (siehe Abbildung 15).⁶⁴

Abbildung 15: Verteilung der gewichteten Polarität und SentiWS⁶⁵

3.2.2 German Polarity Clues (GPC)

Für die Erstellung des German Polarity Clues wurde eine halb-automatische Übersetzung von bereits vorhandenen englisch sprachigen Sentiment Quellen genutzt. Da die existierenden Quellen sich sehr stark in der Anzahl der beinhalteten Begriffe und Features unterscheiden, wurde eine neue Ressource in drei Schritten entwickelt. Zunächst wurden die meist verwendeten englischsprachigen Quellen (z.B. SentiWordNet oder SentiStrength) in einem Experiment evaluiert, d.h. ihre Performance wurde in Abhängigkeit der Quantität ihrer Features getestet. Als nächstes wurden die umfangreichsten existierenden Wörterbücher ins Deutsche mithilfe einer Sprachsoftware übersetzt. Bei Wörtern mit mehreren Übersetzungsmöglichkeiten wurde sich auf eine Maximalanzahl von drei beschränkt. Darüber hinaus wurden alle Übersetzungen eines Wortes auch mit der Polarität des Originals belegt (positiv, negativ oder neutral). Dies hat zur Folge, dass sich die Gesamtgröße des übersetzten Wörterbuchs vom jeweiligen Original unterscheidet, aber auch das es gegebenenfalls zu Mehrdeutigkeiten einzelner Wörter kommt. Deshalb wurden in einem dritten Schritt nochmal eine manuelle Beurteilung aller Wörter des entstandenen GPC durchgeführt. Zusätzlich wurden 290 negierende Phrasen (z.B. nicht-schlecht) und noch nicht vorhandene positive oder negative Synonyme eingefügt,

⁶⁴ Remus et al. 2010, S. 1169–1170

⁶⁵ Remus et al. 2010, S. 1170

Overall Features:	10,141
No. Positive Features:	3,220
No. Negative Features:	5,848
No. Neutral Features:	1,073
No. Negation Features:	290
No. Noun Features:	4,408
No. Verb Features:	2,728
No. Adj/Adv Features:	2,604

wodurch insgesamt eine Gesamtanzahl von 10141 Begriffen entstanden ist (siehe Abbildung 16).⁶⁶

Abbildung 16: Verteilung der gewichteten Polarität und SentiWS⁶⁷

3.2.3 Sonstige

Rill et al. stellten ein weiteres deutschsprachiges Sentiment Lexikon zusammen. Die Liste enthält über 1700 meinungstragende Wörter wie Adjektive und Substantive, sowie 1400 daraus bestehende Phrasen und deren Polarität als Wert zwischen -1 und +1. Als Datengrundlage wurden Rezensionen und auch numerische Bewertungen von Kunden der deutschsprachigen Seite des Versandhandels Amazon genutzt.⁶⁸ Im Vergleich zu anderen semantischen Wörterbüchern, lässt sich feststellen, dass die Bestimmung der Polarität konsistent ist und sich nur quantitativ unterscheidet. Einen Vorteil ihrer Methode sehen die Autoren in der Verwendung von Phrasen, welche die Behandlung von Verneinungen und Verstärkungen vereinfacht.⁶⁹

Neben den speziell für die deutsche Sprache entwickelten Wörterbüchern existieren aber inzwischen auch deutsche Versionen der bekanntesten englisch sprachigen Wörterbücher, wie SentiWordNet oder SentiStrength, welche für verschiedene Plattformen und Programmiersprachen online heruntergeladen werden können.

⁶⁶ Waltinger 2010, S. 1639

⁶⁷ Waltinger 2010, S. 1639

⁶⁸ Rill et al. 2012, S. 305

⁶⁹ Rill et al. 2012, S. 311–312

4 Implementierung

4.1 Datenbeschaffung und Vorverarbeitung

Bevor man die Sentimentanalyse implementieren kann, müssen zunächst Rohdaten für die spätere Analyse gewonnen und vorverarbeitet werden, wenn man nicht auf bereits annotierte Texte zurückgreift. Dies sind Aufgaben, die eigentlich in das Tätigkeitsfeld des Opinion Mining fallen, jedoch für die Sentimentanalyse unerlässlich sind und daher im Folgenden kurz beschrieben werden sollen.

Generell geschieht die Datenbeschaffung typischerweise mithilfe eines Web-Crawlers oder einer speziell dafür entwickelten API (siehe folgende Kapitel), welche Nutzerdaten nach bestimmten Kriterien durchsucht und diese üblicherweise in einer Datenbank speichert.

Danach ist es notwendig den Rohtext zunächst in semantische Einheiten (in der Praxis z.B. Produktfeatures) zu zerlegen, welche dann mithilfe eines Algorithmus analysiert und bewertet werden können. Dazu teilt man zunächst jede der Nutzerdaten in Sätze auf (Tokenization) und weist jedem Wort darin seine semantische Bedeutung zu, also die Einteilung in Subjektive, Adjektive, Adverbien usw. Dies geschieht mithilfe des sogenannten Part-of-Speech (POS) Tagging.

Für die weitere Analyse werden jetzt nur die Sätze benötigt, welche auch getaggte Einheiten besitzen, alle anderen können verworfen werden. Zusätzlich ist es sinnvoll, die am häufigsten auftretenden Wörter zu identifizieren und gegebenenfalls mit ihren Synonymen zu versehen.⁷⁰

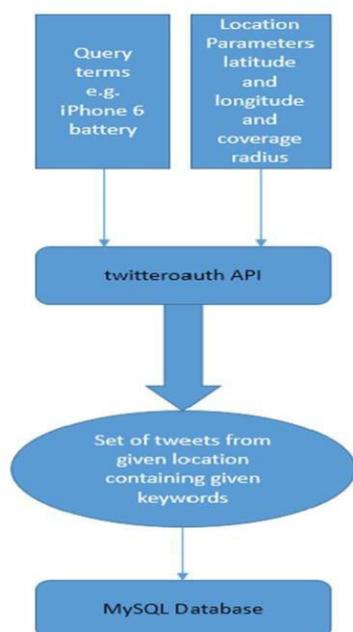
Da sich die Verfahren je nach verwendetem Datenmaterial etwas unterscheiden, sollen diese für die beiden bekanntesten sozialen Netzwerke Facebook und Twitter einzeln vorgestellt werden.

⁷⁰ Alvares et al. 2016, S. 89

4.1.1 Twitter

Anwar-Hridoy et al. beschreiben in ihrer Arbeit ein generelles Verfahren zum Opinion Mining von Twitter Daten. Demnach ist es notwendig, um die Meinung von Nachrichten auf Twitter zu analysieren, diese erstmal zu extrahieren. Dafür gibt es eine öffentlich zugängliche Twitter API, die dies automatisiert ermöglicht. Nachfolgend müssen die gesammelten Daten gefiltert werden, um Redundanzen oder unnütze Informationen zu entfernen. Zusätzlich ist es im Fall von Twitter sinnvoll, das Geschlecht der Autoren zu bestimmen, dass dies nicht von vornherein erfasst wird. Dafür wurden bereits spezielle Tools entwickelt, die auf der Basis von künstlicher Intelligent funktionieren, wie beispielsweise NamSor.⁷¹

Zur Extraktion von Tweets als Datengrundlage kann, wie bereits erwähnt, die eigens dafür



entwickelte Twitter API verwendet werden. Anwar Hridoy et al. verwendeten beispielsweise in ihrer Studie die Version "twitteroauth" von Williams aus dem Jahr 2012. Diese Version ist in PHP implementiert und kann direkt lokal oder in Web Servern verwendet werden. Für eine gezieltere Auswahl der Tweets können bei der Anfrage verschiedene Parameter verwendet werden und die Ergebnisse werden direkt in einer MySQL Datenbank gespeichert (siehe Abbildung 17).

Abbildung 17: Ablauf einer Extraktion von Tweets⁷²

Jeder Eintrag in der Datenbank entspricht dann einem einzelnen Tweet und enthält verschiedene Informationen wie Nutzernamen, Tweet ID, Text oder ähnliches. Anfangs beinhaltete die

⁷¹ Anwar Hridoy et al. 2015, S. 2

⁷² Anwar Hridoy et al. 2015, S. 6

Twitter API auch Information zur Lokalisierung des Tweets in Form von Längen- und Breitengraden. Dies wurde aufgrund von Sicherheitsbedenken und Beschwerden von Nutzern aber 2012 eingeschränkt, sodass die geografische Lage seitdem nicht direkt mit dem Tweet mitgeliefert wird. Jedoch besteht weiterhin die Möglichkeit während der Hauptabfrage in der Twitter API einen Ort als Filter Parameter einzugeben.⁷³

Nach der Extraktion durch die Twitter API müssen die Daten noch vorverarbeitet werden, da sie immernoch willkürliche Buchstaben oder andere unnütze Informationen enthalten können, welche die weiterführende Sentimentanalyse behindern würden.

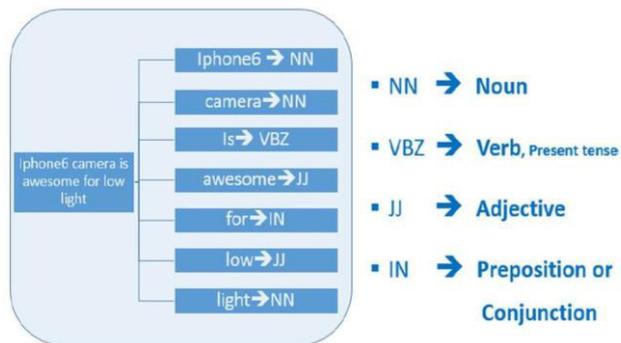
Zur Vorverarbeitung von Twitterdaten schlagen Zhang et. al verschiedene Schritte vor, wie das Entfernen von Retweets, das Überführen von Abkürzungen in ihre ungekürzte Form oder das Löschen von Links bzw. Benutzernamen. Beverungen et al. schlagen darüber hinaus noch das Entfernen fremdsprachiger Tweets vor oder solcher, die sich als SPAM herausstellen. Zusätzlich ist es auch sinnvoll Verneinungen direkt an die Wörter anzuhängen, auf die sie sich beziehen, bspw. durch einen Unterstrich. Dadurch vermeidet man, dass die Wörter positiv interpretiert werden und man muss in der späteren Analyse nur einzelne Wörter und nicht ihre Beziehung zueinander untersuchen. Zhang et al. schlagen auch noch vor, interrogative, also Fragesätze, im Vorfeld bereits auszusortieren, da diese in der Regel keine eigene Meinungsäußerung enthalten. Eine Satzsegmentierung, welche normalerweise einen wichtigen Vorverarbeitungsschritt darstellt, ist bei der Untersuchung von Twitterdaten normalerweise nicht notwendig, da diese aufgrund des begrenzten Zeichensatzes oftmals aus nur einem Satz bestehen.

Das nun vorliegende Rohmaterial kann nun weiteren Analysemethoden aus der natürlichen Sprachverarbeitung (NLP) unterzogen werden. Obwohl während der Vorverarbeitung versucht wurde, den Eigenheiten von Twitterdaten, wie ein begrenzter Zeichensatz, Umgangssprache, Benutzernamen und Retweets Rechnung zu tragen, kann es möglich sein, dass diese weiteren Methoden nicht oder nur begrenzt anwendbar sind. Dies muss jeweils vorher geprüft werden. Weitere Schritte wären jetzt die Zerlegung des Datenmaterials in Wörter (Tokenisierung), die Bestimmung der Wortarten (Part of speech Tagging) und das Überführen der Wörter in ihre Stammform (stemming).⁷⁴

⁷³ Anwar Hridoy et al. 2015, S. 5–6

⁷⁴ Valerius 2012, S. 38

Für die Englische Sprache kann man für das Part-of-Speech Tagging beispielsweise das Stanford Language Processing Tool (SNLP) nutzen, welches grammatikalische Abhängigkeiten von Wörtern innerhalb eines Satzes aufzeigt. Insbesondere die Beziehung von Subjektiven



und Adjektiven oder Objekten ist hierbei von Interesse. Wendet man dieses Tool nun auf die extrahierten Tweets an, bleiben nur diejenigen übrig, die mindestens eine dieser Abhängigkeiten beinhalten (siehe Abbildung 18).⁷⁵

Abbildung 18: POS Tagging an einem Beispiel⁷⁶

Inzwischen existiert dieses Tool auch für diverse weitere Sprachen, unter anderem für die Deutsche.

Danach ist die Vorverarbeitung abgeschlossen und es kann mit Hilfe von Algorithmen, wie Naive Bayes oder SVM oder einem semantischen Wörterbuch die Polarität der Meinungen ermittelt werden und die Ergebnisse können mithilfe der Metadaten von Twitter nach bestimmten Kriterien sortiert oder gefiltert werden, wie Geschlecht, Nationalität, Stadt usw.⁷⁷

4.1.2 Facebook

Um Facebook Daten zu sammeln, ist es notwendig, zunächst das Einverständnis des Nutzers einzuholen, da die Richtlinien bei Facebook darauf bedacht sind, die Privatsphäre des Nutzers zu schützen. Zu diesen Genehmigungen, die der Nutzer erteilen muss, gehören der Offline-Zugriff auf seine Daten, der Lesezugriff auf die Wand und der Zugriff auf Basisdaten, wie das Alter oder das Geschlecht. Danach kann über die Facebook API, einen vom Urheber eigens für solche Zwecke entwickelten HTTP Service, auf die Nutzerdaten zugegriffen werden. Dabei ist zu beachten, dass diese API auch gewissen Beschränkungen unterliegt, wie eine Begrenzung des Downloads von Nutzerdaten auf maximal 50 Nachrichten, die zudem nicht älter als

⁷⁵ Anwar Hridoy et al. 2015, S. 6–8

⁷⁶ Anwar Hridoy et al. 2015, S. 8

⁷⁷ Anwar Hridoy et al. 2015, S. 2

30 Tage sein dürfen. Jedoch können diese Beschränkungen durch geeignete Parameter übergangen werden, sodass es möglich ist die gesamte Nachrichtenhistorie eines Nutzers wiederherzustellen. Im Ergebnis erhält man so genannte JavaScript Object Notation (JSON) Streams, die wiederum aus einzelnen Schlüssel-Wert Paaren bestehen, welche Informationen über den Autor, den Empfänger, das Erstellungsdatum, den Text selbst oder die Anzahl an Likes einer Nachricht enthalten. Diese Datensätze können dann in einer Datenbank gespeichert und später analysiert werden.⁷⁸

Die darauffolgenden Vorverarbeitungsschritte ähneln dabei im Wesentlichen denen bei Twitter, weshalb nicht noch einmal gesondert darauf eingegangen wird. Unterschiede bestehen lediglich darin, dass Facebook Nachrichten aus mehr als einem Satz bestehen können und damit zusätzlich oftmals eine Zerlegung in einzelne Sätze (Tokenization) notwendig ist (*siehe*

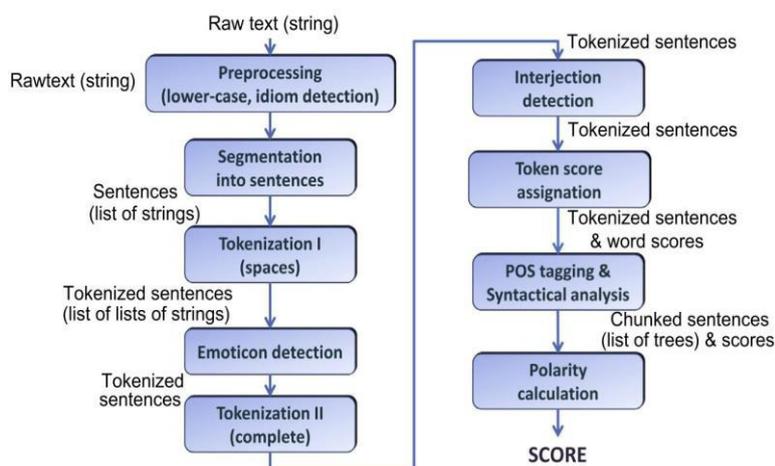


Abbildung 19). Dafür kann aber auf das Entfernen von Twitter typischen Daten wie Retweets, Hashtags oder Benutzernamen verzichtet werden.

Abbildung 19: Vorverarbeitungsschritte von Facebook Nachrichten⁷⁹

⁷⁸ Ortigosa et al. 2014, S. 533

⁷⁹ Ortigosa et al. 2014, S. 531

4.2 Polaritätsbestimmung

Nach der Datenbeschaffung und Vorverarbeitung kann nun die eigentliche Sentimentanalyse durchgeführt und damit die Polarität der Wörter bestimmt werden, die in irgendeiner Form eine Meinung repräsentieren.

Wie bereits in Kapitel 3 beschrieben wurde, kann man dafür auf lernbasierte oder lexikalische Ansätze und Verfahren zurückgreifen. Die lernbasierten Ansätze zeichnen sich besonders dadurch aus, dass sie neben explizit vorkommenden Begriffen, welche direkt mit einem Sentiment assoziiert werden, auch implizite Informationen klassifizieren können. Dafür benötigen sie jedoch eine größere Anzahl an Trainingsdaten und eine saubere Rechtschreibung und Grammatik. Somit eignen sich besonders gut für die Auswertung von Produktbewertungen und können in den unterschiedlichen Domänen präzisere Ergebnisse als Wörterbücher erzielen.

Betrachtet man jedoch soziale Netzwerke, welche oftmals nicht explizit einer Domäne zugeordnet werden können, sind lexikalische Ansätze besser geeignet, da diese keine oder nur wenige Testdaten benötigen und diese auch nicht erst für jede Domäne einzeln trainiert werden müssen. Zudem ist es in sozialen Netzwerken nicht sinnvoll, indirekt vorhandene Sentimente zu bestimmen, da diese die eigentliche Polarität des Satzes beeinflussen können.⁸⁰

Als Dimension der Sentimentanalyse empfiehlt sich die Satzebene, da Meinungsäußerungen aus sozialen Netzwerken oftmals knapp gehalten sind oder innerhalb eines Textes mehrere Entitäten thematisch behandelt werden, sodass diese segmentiert werden sollten.

Ein Problem bei der Implementierung ist, dass viele Forscher zwar die Methoden und Ergebnisse ihrer Forschung in Form von Papern veröffentlichen, den Quellcode ihres Algorithmus jedoch geheim halten, was die Weiterentwicklung und den gemeinsamen Erkenntnisgewinn erschwert.

Ein einfacher Algorithmus für das soziale Netzwerk Twitter, welcher frei zugänglich ist, wurde von Hu und Lin veröffentlicht. Dabei wird zunächst die semantische Orientierung der sogenannten Opinion Words mithilfe eines semantischen Wörterbuchs (seed list), ermittelt, welches als externe Quelle in den Algorithmus eingebunden wird. Für den Fall, dass einem Opinion Word eine Negation vorausgeht, wird die Polarität umgekehrt. (siehe Abbildung 20).

⁸⁰ Schmidt 2012, S. 9–10

```

Algorithm 2: wordOrientation ()
1. Procedure wordOrientation(word, sentence)
2. begin
3. orientation = orientation of word in seed_list;
4. If (there is NEGATION_WORD appears closely arot
word in sentence)
5. orientation = Opposite(orientation);
6. end

```

Abbildung 20: Algorithmus zur Bestimmung der semantischen Wortorientierung⁸¹

Wurde die semantische Orientierung aller Opinion Words eines Satzes bestimmt, kann die Gesamtpolarität des Satzes durch Summierung von positiven und negativen Wörtern bestimmt werden (*siehe Abbildung 21*). ⁸²

```

Algorithm 1: SentenceOrientation ()
1. Procedure SentenceOrientation()
2. begin
3. for each opinion sentence si
4. begin
5. orientation = 0;
6. for each opinion word op in si
7. orientation += wordOrientation(op, si);
8. /*Positive = 1, Negative = -1, Neutral = 0*/
9. if (orientation > 0) si's orientation = Positive;
10. else if (orientation < 0) si's orientation = Negative;
11. else {
12. for each feature f in si
13. orientation +=
14. wordOrientation(f's effective opinion, si);
15. if (orientation > 0)
16. si's orientation = Positive;
17. else if (orientation < 0)
18. si's orientation = Negative;
19. else si's orientation = si-1's orientation;
20. }

```

Abbildung 21: Algorithmus zur Polarität eines Satzes⁸³

⁸¹ Alvares et al. 2016, S. 90

⁸² Alvares et al. 2016, S. 89–90

⁸³ Alvares et al. 2016, S. 89

4.3 Ergänzende Faktoren

Die Schwierigkeit bei der Sentimentanalyse von Nachrichten aus sozialen Netzwerken besteht darin, dass diese zu kurz oder zu unstrukturiert sind, um sie effektiv zu untersuchen. Um dieses Problem zu lösen, werden verstärkt weitere sogenannte nicht-textliche (non-textual) Features mit in die Analyse einbezogen, wie beispielsweise Bilder, Videos, Links oder Emojis. Des Weiteren spielen auch die Erkennung von Ironie bzw. Sarkasmus und die Glaubwürdigkeit der zugrunde liegenden Daten eine wichtige Rolle, um Sentimentanalyse noch effizienter einsetzen zu können. Neuere Forschungen betrachten zudem noch die soziale Interaktion zwischen den Nutzern als Faktor zur Bestimmung der Stimmungsrichtung. ⁸⁴

4.3.1 Ironie / Sarkasmus Erkennung

Sarkasmus wird in der Sentimentanalyse als ein Faktor angesehen, der die Polarität einer Nachricht ins Gegenteil verkehrt. Im Gegensatz zu einer einfachen Negation, wird bei Sarkasmus jedoch eine negative Meinung durch ausschließlich positive Worte vermittelt oder sogar durch übertrieben positive Begriffe noch verstärkt. Ebenso gibt es auch den Fall, dass mittels Sarkasmus eine negative Meinung ins positive gewandelt wird, was jedoch seltener vorkommt. Der Vorteil bei der Erkennung von Sarkasmus in Twitter Nachrichten ist, dass diese oftmals mittels hashtags wie #sarcasm oder #not als solche identifizierbar sind. ⁸⁵

Die Forscher Kunneman et al. untersuchten die Kennzeichen von Sarkasmus anhand niederländischer Twitter Nachrichten. Dazu trainierten sie einen Klassifikator mit Datensätzen, die explizite hashtags wie #sarcasm, #irony, #cynicism und #not enthielten und wendeten diesen Algorithmus dann auf einen Testdatensatz an. Er erreichte jedoch nur eine durchschnittliche Korrekturklassifikationsrate (Accuracy) von 35%, weshalb die Autoren zu dem Ergebnis kamen, dass es sehr schwierig ist, sarkastische von wörtlich gemeinten Tweets zu unterscheiden. Dennoch stellten sie die Hypothese auf, dass die Hashtags, welche Sarkasmus andeuten, den non-verbalen Ausdrücken in einer realen Konversation entsprechen und je mehr dieser Marker in einer Äußerung verwendet werden, desto eindeutiger ist diese verständlich. ⁸⁶

⁸⁴ Kuo et al. 2016, S. 399–400

⁸⁵ Kunneman et al. 2015, S. 500–501

⁸⁶ Kunneman et al. 2015, S. 507–508

Justo et al. untersuchten Merkmale von Sarkasmus in sozialen Netzwerken und entdeckten 1815 Stichwörter, die auf Sarkasmus hindeuten. Sie kann aber auch zu dem Schluss, dass zusätzliche Informationen, wie der Kontext oder die Länge einer Nachricht die Genauigkeit bei der Erkennung noch verbessern, da es auch Formen von Sarkasmus gibt, die nicht durch typische Stichwörter zu erkennen sind.⁸⁷

Klinger et al. untersuchten in einer weiteren Arbeit ein Modell zur Unterscheidung von Ironie und Sarkasmus bei Twitter Nachrichten. Sie stellten fest, dass dies nochmal deutlich schwerer ist, als eine reine Unterscheidung von sprichwörtlichen und sarkastischen bzw. ironischen Tweets. Dennoch arbeiten sie ein paar Faktoren heraus, die für die Erkennung von Ironie und Sarkasmus relevant sind. Dies ist zum einen die Länge der Nachricht, da ihren Ergebnissen nach, reguläre Tweets im Durchschnitt weniger Sätze als ironische und diese weniger als sarkastische Nachrichten enthalten. Jedoch beinhalten ironische Tweets mehr Markierungen, wie sarkastische. Die Autoren erklären dies damit, dass Ironie mehr Hinweise benötigt, um die Ironie zu vermitteln. Ein weiteres Merkmal ist, dass sarkastische Tweets mehr positive Wörter enthalten als ironische, wohingegen diese mehr negative Worte beinhalten. Dies geht mit dem allgemeinen Verständnis einher, das Ironie die wörtliche Bedeutung einer Aussage ins Gegenteil verkehrt und Sarkasmus die negative Form davon ist. Zudem werden in sarkastischen Tweets mehr Usernames verwendet, als in ironischen, da diese oftmals direkt an ein Gegenüber gerichtet sind, wohingegen Ironie eher genereller Natur ist.

Die Ergebnisse unterstützen die bereits vorhandenen Theorien, dass bestimmte wort-basierte Features und damit spezifisches Hintergrundwissen die Erkennung von Ironie und Sarkasmus verbessern.⁸⁸

4.3.2 Berücksichtigung von Emotionsausdrücken (Emojis)

Um während einer direkten Kommunikation Hinweise auf versteckte Intentionen zu geben, wird für gewöhnlich Körpersprache verwendet. Jedoch können Gesichtsausdrücke oder Gesten während einer schriftlichen Kommunikation, wie beispielsweise in einem sozialen Netzwerk, nicht zum Ausdruck gebracht werden. Dies wird heutzutage jedoch durch den Einsatz von Emotionsausdrücken, den sogenannten Emojis kompensiert, welche so effektiv wie möglich eingesetzt werden, um die eigene Meinung auszudrücken. Aktuelle Studien haben sogar

⁸⁷ Justo et al. 2014, S. 131–132

⁸⁸ Klinger und Ling 2016, S. 9–10

herausgefunden, dass das menschliche Gehirn Emojis nicht länger als Zeichen, sondern als wirkliche physische Verhaltensweisen wahrnimmt. Für die Sentimentanalyse ist diese Erkenntnis sehr hilfreich, da es hilft, nicht eindeutige Meinungsäußerungen zu klassifizieren. Beispielsweise lässt sich der Satz "Das Wetter ist wirklich schön heute" mit einem anschließenden weinenden Emoji eindeutig als negativ einstufen, wohingegen er ohne diese Markierung, höchstwahrscheinlich als positiv klassifiziert worden wäre.⁸⁹

In einigen aktuellen Studien und Wörterbüchern zur Sentimentanalyse (z. B. dem SentiStrength) werden Emojis auch schon berücksichtigt und ihnen wird, ähnlich wie einem meinungs-

Positive emoticons	:) :D :] :* ;) ;] 8) <3 ;P =P =) B) :-> :-D XD =3 ;> ;-) :') =>
Negative emoticons	:(:'(:-(=[:-@ :-< :-[:O :-/ :-\ :/ :\ :"(:-O O.o o.O :[:<

tragenden Wort, eine Polarität zugewiesen. Kadge et al. erstellten beispielsweise dafür eine eigene Tabelle, in der sie die einzelnen Emojis als positiv oder negativ klassifizierten (*siehe Abbildung 22*).⁹⁰

Abbildung 22: Einordnung von Emojis⁹¹

Zhang et al. belegten in einer Studie sogar, dass die Korrekturklassifikationsrate (Accuracy) linear zu der prozentualen Menge an Nachrichten, welche Emojis enthalten, ansteigt.⁹²

4.3.3 Glaubwürdigkeit

Etwa drei Viertel aller Erwachsenen, die über einen Online Zugang verfügen, nutzen, unter anderem, soziale Netzwerke als Informationsquelle, obwohl diese im Gegensatz zu politischen Blogs oder Nachrichten Webseiten als nicht sehr glaubwürdig eingestuft werden. Die Gründe warum Facebook als nicht sehr glaubwürdig eingestuft wird, liegen zunächst darin, dass Seiten, welche primär auf das Knüpfen sozialer Beziehungen ausgerichtet sind, generell als weniger vertrauenswürdig eingeschätzt werden. Ein weiterer Grund ist das häufige Fehlen von

⁸⁹ Huang et al. 2017, S. 145–146

⁹⁰ Kadge et al. 2016, S. 34

⁹¹ Kadge et al. 2016, S. 34

⁹² Huang et al. 2017, S. 150–151

Quellenangaben oder zumindest eine schlechte Nachprüfbarkeit aufgrund fehlerhafter Verlinkungen. Darüber hinaus fehlt es Facebook an einer Kontrollinstanz, wie etwa Lektoren oder Bearbeitern, die den Wahrheitsgehalt von Nachrichten prüfen.⁹³

Die Glaubwürdigkeit von Twitter wird generell als geringer angesehen als die von Blogs oder Facebook. Ähnlich wie Facebook verfügt Twitter nicht über ein Kontrollgremium, welches die Nachrichten auf ihre Authentizität prüft. Dies wird besonders bei Nachrichten zu aktuellen Ereignissen, Krisen oder Wahlen zu einem Problem, da in diesen Fällen viele schnell getippte Tweets in kurzer Zeit verschickt werden, was häufig zu Fehlinformationen oder Gerüchten führt.⁹⁴

Zu Beginn der Verbreitung von Twitter wurde der Informationsgehalt der Tweets von vielen sogar als sehr niedrig eingestuft. Dies änderte sich jedoch mit der zunehmenden Zahl bedeutender Personen aus Politik und Wissenschaft, die Twitter als Kommunikationsmedium nutzten. Castillo et al. untersuchten Twitter daraufhin nach Indikatoren in den Nachrichten, mit denen sich die Glaubwürdigkeit messen lässt. Sie ermittelten dabei mithilfe einer Testgruppe, dass der Anteil an relevanten Tweets bei 28,5 % lag. Aus dieser Menge wurden dann zufällig 383 Themen ausgewählt, deren Glaubwürdigkeit in drei Kategorien eingestuft wurde. Dabei wurden 41% aller Tweets als "almost certainly true", 31,8% als "likely to be false" und 8,6% als "almost certainly false" eingestuft. Diese manuellen Einschätzungen wurden dann als Trainingsdatensatz für einen Algorithmus genutzt der nur noch zwischen "almost certainly true" und "ambiguous" unterscheiden sollte. Dieser erreichte eine Korrektklassifikationsrate (Accuracy) von 86%. Im Ergebnis der Studie konnte festgestellt werden, dass Tweets, die keine URL enthalten generell als unglaubwürdiger gelten, ähnliches gilt für Tweets mit vielen positiv besetzten Begriffen. Nachrichten mit vermehrt negativen meinungstragenden Wörtern werden hingegen als glaubwürdig eingestuft, genau wie User, die viele Follower besitzen oder bereits viele Tweets verfasst haben. Ein weiteres positives Indiz für Glaubwürdigkeit sein nach Meinung der Autoren, die Anzahl der Retweets, die eine Nachricht bekommen hat. Die Ergebnisse kann man dahingehend interpretieren, dass Tweets, die Links enthalten generell glaubwürdiger sind, da diese Links als Quelle für die Aussagen betrachtet werden können. Gleiches

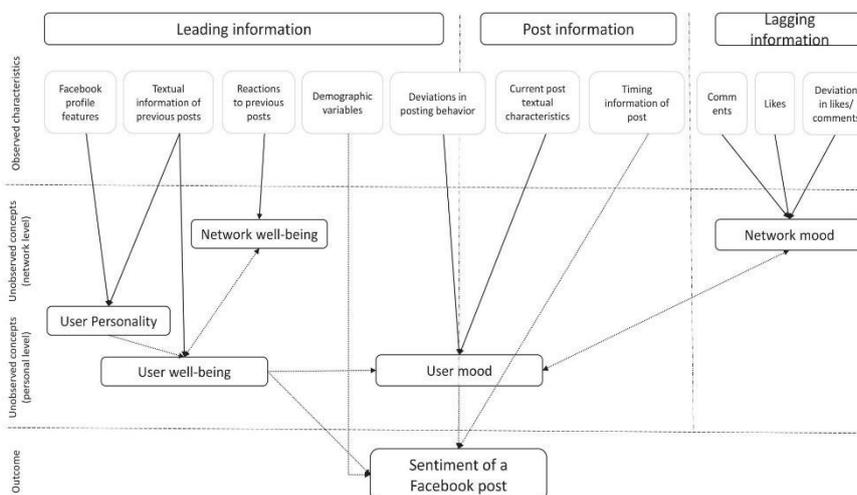
⁹³ Johnson und Kaye 2015, S. 545–546

⁹⁴ Johnson und Kaye 2015, S. 546

gilt für Personen, die viele Follower haben, da davon auszugehen ist, dass diese Autoren ihre Leser nicht durch Falschinformationen verlieren wollen.⁹⁵

4.3.4 Interaktion zwischen Nutzern

Meire et al. untersuchten in einer Studie den Einfluss von zusätzlichen Daten auf die Meinungen von Nutzern bei der Erstellung von Facebook Posts. Mit zusätzlichen Daten sind anleitende (leading) oder zurückbleibende (lagging) Informationen einer Nachricht gemeint, welche einen zusätzlichen Wert bei der Sentimentanalyse haben können. Zu den anleitenden Informationen zählen alle Informationen, die der Nutzer bereits im Vorfeld bei Facebook gesammelt hat und die seine Meinung beim Verfassen einer Nachricht beeinflussen können, wie zum Beispiel frühere Posts. Die Autoren der Studie bezeichnen dies als subjektives Wohlbefinden. Dieses sei auch eng verknüpft mit dem Phänomen der emotionalen Ansteckung, bei dem eine Person, die Meinung anderer oder den emotionalen Zustand nachahmt. Zu den zurückbleibenden Informationen zählen die Likes und die Kommentare auf einen Post oder abweichendes Verhalten im Bezug auf die Bewertung früherer Nachrichten. Beispielsweise wurde gezeigt, dass negative Posts sehr viele Kommentare hervorrufen und die Nutzer stärker emotional darauf reagieren, auf der anderen Seite erzeugen viele Likes positive Meinungen. All diese Faktoren beeinflussen, nach Meinung der Autoren, die Stimmung des Einzelnen (individual



mood) und die Stimmung des Netzwerks (network mood), welche wiederum in wechselseitiger Beziehung stehen (siehe Abbildung 23).⁹⁶

Abbildung 23: Faktoren zur Beeinflussung von Meinungen bei Facebook⁹⁷

⁹⁵ Valerius 2012, S. 28–30

⁹⁶ Meire et al. 2016, S. 100–101

⁹⁷ Meire et al. 2016, S. 100

5 Evaluation

5.1 Testdatensätze / Goldstandards

Neben der Entwicklung von lernbasierten Algorithmen im Bereich Sentimentanalyse ist es auch für die Evaluation notwendig, dass es vorhandene Goldstandards bzw. bereits bewertete Testdatensätze gibt. Jedoch nimmt das manuelle Labeln von Daten viel Zeit in Anspruch und ist kostenintensiv. Folglich gibt es im Bereich der Sentimentanalyse bisher recht wenige dieser Datensätze und diese repräsentieren oftmals nur eine geringe Bandbreite an Themen (z.B. Filme, Produkttypen) oder sind nur für englische Texte vorhanden. Boland et al. entwickelten daraufhin anhand von Amazon Produktkritiken einen recht umfangreichen Korpus von kommentierten Trainings- und Testdatensätzen in deutscher Sprache, welcher unterschiedliche Kategorien und Produkttypen beinhaltet. Dieser kann zukünftig dafür genutzt werden, um den Einfluss verschiedenartiger Datensätze auf Algorithmen zur Sentimentanalyse zu zeigen und diese mit Hilfe maschinellen Lernens zu verbessern.⁹⁸

Trainingsdaten sollten sich dadurch auszeichnen, dass sie in sich konsistent sind und wenn möglich einen endlichen Wertebereich besitzen, wobei jeder Klasse auch mindestens ein Beispiel zugeordnet wurde. Dadurch lässt sich ausschließen, dass neue Objekte nicht korrekt klassifiziert werden können. Wichtig ist zudem eine ausreichende Größe des Trainingsdatensatzes, damit auch eine repräsentative Häufigkeit erreicht wird und es damit den Algorithmen auch leichter fällt, Muster und Regeln zu erkennen. Dies verhindert auch Phänomene der Überanpassung (Overfitting), bei dem zu spezielle Regeln erlernt werden und neue Beispiele nicht mehr richtig klassifiziert werden können, da sie in ihren Eigenschaften nicht mehr den Trainingsdaten entsprechen.⁹⁹

Im Folgenden sollen ein paar dieser Testdatensätze bzw. Goldstandards, welche für die deutsche Sprache erschienen sind genauer vorgestellt werden.

⁹⁸ Boland et al. 2013, S. 5

⁹⁹ Sprejz, S. 22

5.1.1 Korpus nach Boland

In dem kommentierten Korpus nach Boland et al. konzentrierte man sich auf die satzweise Sentimentanalyse und klassifizierte diese in positiv, negativ, mixed und neutral. Die Einteilung wurde dabei manuell durch verschiedene wissenschaftliche Mitarbeiter vorgenommen. Idealerweise sollte der Kontext bei der Bewertung eines Satzes außer Acht gelassen werden, was jedoch nicht in jedem Fall möglich ist, wie der Einschätzung von Produktdetails ohne das nötige Expertenwissen. Diese Sätze, die ohne entsprechendes Hintergrundwissen nicht bewertet oder nicht eindeutig einer Kategorie zugeordnet werden konnten, wurden speziell markiert (Kategorie: kontrovers).

Des Weiteren haben wurde in dem Korpus der Tatsache Rechnung getragen, dass die Sentimentanalyse Domain, Themen und Zeit abhängig ist und es somit notwendig ist, dass fertig bewertete Datensätze einen großen Bereich davon abdecken, um lernende Algorithmen weiterzuentwickeln. Boland et al. haben das so gelöst, indem sie ihren Korpus in weitere Sub-Korpusse unterteilt haben, welcher verschiedene Produkte und Produkttypen umfasst (*siehe Abbildung 23*).¹⁰⁰

Subcorpus	#positive sentences	#negative sentences	#mixed sentences	#neutral sentences	#controversial sentences
Washing Machine	1272	1455	314	1190	106
Camera	3765	4525	915	4425	561
Book	1085	903	196	1278	69
Tablet	2594	3337	827	2318	21
Mobile	1296	1282	334	994	779
Smartphone	7841	10379	2193	6793	20
Total	17853	21881	4779	16998	1556

Abbildung 23: Verteilung von Produktkategorien im Korpus¹⁰¹

Im Ergebnis der Arbeit von Boland et al. entstand ein Korpus aus 6 untergeordneten Produkttypen (Waschmaschine, Kamera, Bücher, Tablet, Mobil und Smartphone), deren bewertete Sätze in 5 Kategorien (positiv, negativ, gemischt, neutral und kontrovers) eingeteilt werden

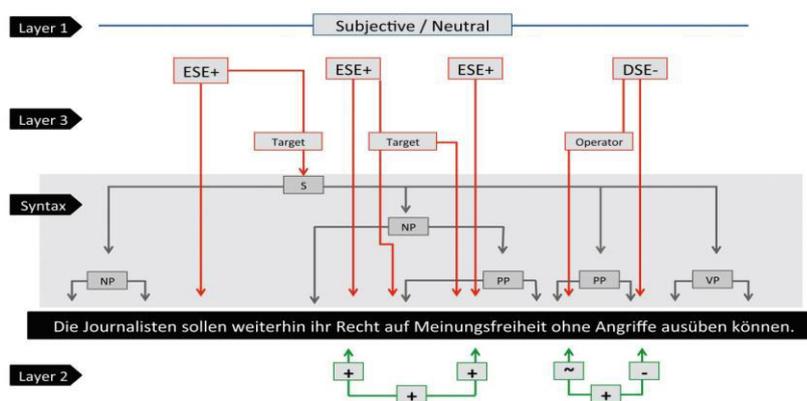
¹⁰⁰ Boland et al. 2013, S. 6

¹⁰¹ Boland et al. 2013, S. 12

können. Wurden Sätze von mehreren Mitarbeitern bewertet, wurde das mehrheitliche Votum als Grundlage für die finale Bestimmung der Kategorie genutzt. Ist keine mehrheitliche Meinung erzielt worden, wurde der betreffende Satz in die Kategorie kontrovers eingestuft. Zudem war eine Zielsetzung, dass positive und negative Meinungen etwa ausbalanciert sind, dies konnte in der Praxis jedoch nicht ganz erreicht werden, da bei der Bewertung von Produkttypen oftmals die negativen Meinungen vorherrschen,¹⁰²

5.1.2 MLSA

Eine weiterer Korpus für die Sentimentanalyse in deutscher Sprache ist der 2012 vorgestellte multi-layered reference corpus for German-language sentiment analysis (MLSA). Dieses Projekt zeichnet sich besonders dadurch aus, dass die 270 bewerteten Sätze in drei verschiedenen Granularitätsstufen betrachtet werden können. Schicht 1 betrachtet den Satz als Ganzes und damit seine gesamte Polarität, Schicht 2 ist Wort bzw. Wortgruppen basierend und Schicht 3 betrachtet einzelne Ausdrücke, wie beispielsweise eine direkte Rede (siehe Abbildung 24). Um die Verlässlichkeit des Korpus zu gewährleisten, wurde jeder Satz von mehreren Juroren



bewertet und eine durchschnittliche paarweise Übereinstimmung gebildet.¹⁰³

Abbildung 24: Ausschnitt aus dem Mehrschichtensystem des MLSA¹⁰⁴

Schicht 1 im Model MLSA umfasst die Bewertung auf Satzebene und ist somit der grobkörnigste Anmerkungsbereich. Hier wird ein Satz als Ganzes betrachtet und im ersten Teil in subjektiv oder objektiv unterschieden. Subjektiv heißt in diesem Zusammenhang, dass eine

¹⁰² Boland et al. 2013, S. 12

¹⁰³ Gindl et al. 2012, S. 3551

¹⁰⁴ Gindl et al. 2012, S. 3552

Meinung oder Einstellung mit einer Äußerung einhergeht und objektiv bedeutet, dass ein Satz lediglich eine Information transportiert, aber keine explizite Einstellung dazu (*Ergebnisse dazu in Abbildung 25*).

Tag	# of tags in consensus
subjective	147
objective	71
no consensus	52

Abbildung 25: Verteilung der subjektiv bzw. objektiv Bewertungen in Schicht 1¹⁰⁵

Als zweite Dimension der ersten Schicht wird die Polarität einer Äußerung bestimmt, dabei ist eine negative Polarität synonym zu einer negativen Äußerung und das selbe gilt umgekehrt

Tag	# of tags in consensus
positive	55
negative	78
neutral	75
no consensus	62

für positive Polarität. Ist ein Satz mehrdeutig oder enthält überhaupt keine explizite Meinung, wird er mit einer neutralen Polarität gekennzeichnet (*siehe Abbildung 26*).¹⁰⁶

Abbildung 26: Verteilung der Polarität in Schicht 1¹⁰⁷

Schicht 2 des MLSA Modells konzentriert sich auf die Polarität von Wörtern bzw. Wortphrasen, besonders Nominal- und Präpositionalphrasen (NP und PP). Dabei wurden sowohl subjektive als auch objektive Teile bewertet, basierend auf ihrer syntaktischen Struktur. Ziel ist es, mithilfe dieser Schicht, Algorithmen anzulernen, wie sie die Polarität komplexer Phrasen, basierend auf seinen einzelnen Teilphrasen, bestimmen können. Auf der Ebene der Phrasen wurde die Symbole + für positiv, - für negativ, 0 für neutral und # für mehrdeutig verwendet. Darüber hinaus wurden die Phrasen in eckige Klammern gesetzt und das jeweilige Symbol angefügt. Auf der Wortebene wurden die Symbole % für vermindernd, ^ für verstärkend und ~ für inver-

Tag	Marker	#Words	Examples	#Top Phrases	#All Phrases
positive	+	335	hope	158	275
negative	-	362	doubt	180	300
intensifier	^	63	heavy	n.a.	n.a.
diminisher	%	9	low	n.a.	n.a.
shifter	~	51	against	n.a.	n.a.
bipolar	#	n.a.	n.a.	21	54
neutral	0	n.a.	n.a.	10	12

tierend verwendet.

Die Tabelle zeigt die Statistiken zu den Ergebnissen aus Schicht 2.

Abbildung 27: Verteilung der Polarität in Schicht 2¹⁰⁸

¹⁰⁵ Gindl et al. 2012, S. 3553

¹⁰⁶ Gindl et al. 2012, S. 3552–3553

¹⁰⁷ Gindl et al. 2012, S. 3553

¹⁰⁸ Gindl et al. 2012, S. 3554

Die Spalte Top Phrases bezeichnet dabei diejenigen Phrasen, die direkt unterhalb von Schicht 1 stehen, also unmittelbar zu einem Satz zusammengesetzt werden können. Die Spalte "All Phrases" bezeichnet alle Phrasen, die unterhalb von Schicht 1 stehen und mit einer Polarität versehen wurden. Diese schließen auch die Top-Phrasen ein. Die Ergebnisse zeigen, dass es eine Tendenz zu Negativität gibt, sowohl auf Wort-, als auch auf Phrasenebene, während Neutralität eher selten vorkommt. Zudem kommen verstärkende oder invertierende Wörter deutlich häufiger vor als vermindernde. Letztendlich wird noch deutlich, dass Sätze auch eine neutrale Polarität haben können, obwohl sie Wörter oder Phrasen beinhalten, die als positiv oder negativ eingestuft wurden.¹⁰⁹

Schicht 3 des MLSA Models konzentriert sich auf die Beurteilung von Ausdrücken, welche sich auf eine bestimmte Entität, zum Beispiel eine Person oder ein Produkt beziehen. Dabei werden lexikalische Einheiten mit Anmerkungen versehen, wie private Zustände, die nach außen noch nicht offensichtlich oder verifizierbar sind. Unterschieden werden diese in die drei Typen, objektive Sprachereignisse (OSE), direkte Sprachereignisse (DSE) und explizit subjektive Äußerungen (ESE). Letzteres werden genutzt, um bestimmte Gefühle auszudrücken, ohne diese direkt zu äußern (*siehe Abbildung 28*).

	Merged	Annotator 1	Annotator 2
DSE	656	642	638
ESE	734	692	713
OSE	7	7	6

Abbildung 28: Verteilung der Typen in Schicht 3¹¹⁰

Zusätzlich werden die lexikalischen Einheiten in Elemente unterteilt, wie Quelle (source), Ziel (target), Modulationen (modulation) oder Operatoren (operator). Ein weiteres Element ist die Polarität (polarity), welche die Stimmung bezüglich des Ziels einer Äußerung widerspiegelt

	Merged	Annotator 1	Annotator 2
Source	261	254	249
Target	1124	1053	1074
Operator	60	54	58
Modulation	160	147	155
Polarity	23	23	18
Support	130	126	127

(*siehe Abbildung 29*) Beispiele dafür sind Präposition wie für bzw. gegen, welche als Marker fungieren, ob die Meinung positiv oder negativ gegenüber einem Ziel eingestellt ist.¹¹¹

Abbildung 29: Verteilung der Kategorien in Schicht 3¹¹²

¹⁰⁹ Gindl et al. 2012, S. 3553–3554

¹¹⁰ Gindl et al. 2012, S. 3554

¹¹¹ Gindl et al. 2012, S. 3554–3555

¹¹² Gindl et al. 2012, S. 3554

Wie aus Abbildung 28 ersichtlich wird, gibt es nur wenige objektive Sprachelemente, da diese oftmals keine Meinung bzw. eine Haltung des Autors gegenüber einem Ziel beinhalten. In Abbildung 29 sieht man zudem, dass weniger Quellen wie Ziele identifiziert wurden. Dafür gibt es zwei Gründe. Zum einen bezieht sich die Quelle oftmals implizit auf den Verfasser des Satzes und wurde deshalb nicht explizit vermerkt. Zum anderen gibt es relativ viele explizit subjektive Ausrücke (ESE), wo kein direkter Zusammenhang der Meinung zu einer Quelle hergestellt werden kann.¹¹³

5.1.3 Sonstige

Sänger et al. veröffentlichten einen Korpus zu Nutzerrezension von mobilen Applikationen in deutscher Sprache. Sie bezeichneten ihn als Sentiment Corpus of App Reviews (SCARE) und er besteht aus 1760 manuell annotierten Rezensionen zu Applikationen aus dem Google Play Store, welche 2487 Features und 3959 subjektive Phrasen beinhalten.¹¹⁴

Cieliebak et al. veröffentlichten SB10k, einen Korpus mit etwa 10.000 deutschsprachigen Twitter Nachrichten.¹¹⁵

Klinger et al. veröffentlichten den Bielefeld University Sentiment Analysis Corpus for German and English (USAGE), welcher annotierten Phrasen und damit verbundene Features aus Amazon Rezensionen beinhaltet. Der Datensatz besteht aus 622 Englischen und 611 Deutschen Rezensionen, um zu verdeutlichen, wie man Methoden der Sentimentanalyse auf verschiedene Sprachen übertragen kann.¹¹⁶

Sidarenka veröffentlichte den Potsdam Twitter Sentiment Corpus (PotTS), einen Korpus mit 7992 deutschsprachigen Tweets, welche manuell annotiert wurden. Dieser beinhaltet Markierungen zu Quelle und Ziel einer Aussage, sowie zu den meinungstragenden Wörtern und mögliche Verneinungen oder Modifikationen¹¹⁷

¹¹³ Gindl et al. 2012, S. 3555

¹¹⁴ Mario Saenger et al., S. 1114

¹¹⁵ Mark Cieliebak et al. 2017, S. 45

¹¹⁶ Roman Klinger und Philipp Cimiano, S. 1

¹¹⁷ Sidarenka 2015, S. 1133

5.2 Evaluationsmaße

5.2.1 Konfusionsmatrix

Zur Bestimmung der Güte eines Algorithmus wird häufig die sogenannte Konfusionsmatrix herangezogen. Diese vergleicht die Vorhersage des Klassifikators mit einem Vergleichsdatensatz, der als korrekt klassifiziert betrachtet wird. Dabei können vier Fälle unterschieden werden: wahr-positiv, wahr-negativ, falsch-positiv und falsch-negativ (*siehe Abbildung 30*). Als wahr-positiv und wahr-negativ werden die Zustände angesehen, bei denen der Klassifikator die Objekte korrekt als positiv bzw. negativ bewertet hat. Die Fälle falsch-positiv und falsch-negativ bezeichnen die Fälle bei denen der Klassifikator die Objekte falsch eingeordnet hat, also entweder wahre Werte als falsch oder falsche Werte als wahr. Die Güte eines Algorithmus bestimmt sich somit aus dem Verhältnis von richtig zu falsch eingeordneten Objekten.¹¹⁸

	Relevant	Irrelevant
Detected Opinions	True Positive (tp)	False Positive
Undetected Opinions	False Negative (fn)	True Negative

Abbildung 30: Konfusionsmatrix¹¹⁹

5.2.2 Positiver Vorhersagewert (Precision)

Der positive Vorhersagewert bzw. die Precision gibt den Anteil der korrekt als positiv klassifizierten Ergebnisse an der Gesamtheit der als positiv klassifizierten Ergebnisse an (erste Zeile der Konfusionsmatrix).

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad \text{Abbildung 31: Formel für Precision}^{120}$$

¹¹⁸ Valerius 2012, S. 56

¹¹⁹ Singh und Shahid Husain 2014, S. 13

¹²⁰ Singh und Shahid Husain 2014, S. 13

5.2.3 Sensitivität (Recall)

Das Maß Sensitivität bzw. Recall misst den Anteil der richtig klassifizierten positiven Objekte an der Gesamtheit aller tatsächlich positiven Objekte.

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad \text{Abbildung 32: Formel für Recall}^{121}$$

5.2.4 Korrektklassifikationsrate (Accuracy)

Die Korrektklassifikationsrate bzw. Accuracy gibt den Anteil aller Objekte an, die korrekt klassifiziert wurden.

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \quad \text{Abbildung 33: Accuracy als Formel}^{122}$$

5.2.5 F-Maß

Zusätzlich kann auch noch das F-Maß herangezogen werden. Dieser Wert vereint die Maße Precision und Recall zu einer Variablen (*siehe Abbildung 33*). Dies ist hilfreich, da diese jeweils einzeln nicht sehr aussagekräftig sind. Beispielsweise liegt der Recall bei 100%, wenn ein System alle untersuchenden Dokumente wieder als Ergebnisse zurückliefert. Genauso liegt die Precision bei 100%, wenn nur ein Dokument klassifiziert wird, dieses jedoch korrekt. Um diese Zustände zu vermeiden verbindet das F-Maß die beiden Werte durch das harmonische Mittel oder mit Gewichtung in Form eines Parameters.¹²³

$$F = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Abbildung 34: Formel für F-Maß}^{124}$$

¹²¹ Singh und Shahid Husain 2014, S. 13

¹²² Singh und Shahid Husain 2014, S. 13

¹²³ Valerius 2012, S. 56

¹²⁴ Singh und Shahid Husain 2014, S. 13

5.3 Evaluation vorgestellter Verfahren

Im folgenden Kapitel sollen einige der bereits vorgestellten Ansätze und Verfahren noch einmal hinsichtlich ihrer Evaluationsergebnisse genauer betrachtet werden.

Buetow et al. zeigten in einer Studie zur Sentimentanalyse von deutschsprachigen News Artikeln, dass nicht alle Techniken zur Datenvorverarbeitung aus dem Information Retrieval auch für die Sentimentanalyse sinnvoll sind. Beispielsweise haben die Autoren festgestellt, dass ein einfaches Hinzufügen der Termfrequenz keine Verbesserung der Ergebnisse brachte (*siehe*

Abbildungen 35 und 36).

	Negative	Neutral	Positive
Precision	67.7 %	76.3 %	74.1 %
Recall	51.5 %	90.3 %	43.2 %
Accuracy	74.8 %		

Abbildung 35: Evaluation der Resultate ohne Vorverarbeitung¹²⁵

	Negative	Neutral	Positive
Precision	67.5 %	76.5 %	72.6 %
Recall	51.9 %	90.0 %	43.5 %
Accuracy	74.7 %		

Abbildung 36: Resultate mit Hinzufügen der Termfrequenz¹²⁶

Dies hat ihrer Meinung nach damit zu tun, dass häufig auftretende Wörter wie beispielsweise Artikel keinen Einfluss auf die Polarität eines Satzes haben. Ähnliches gilt für eine Kombination aus Termfrequenz und inverser Dokumentenfrequenz (*siehe* *Abbildung 37*).

	Negative	Neutral	Positive
Precision	60.4 %	82.5 %	59.3 %
Recall	67.3 %	78.3 %	62.3 %
Accuracy	73.5 %		

Abbildung 37: Resultate mit Hinzufügen von TF-IDF¹²⁷

Diese soll eigentlich mehr Gewicht auf selten auftretende Wörter legen, was bei traditionellen Anwendungen des Information Retrieval wie Suchmaschinen auch sinnvoll ist. Jedoch haben bei der Sentimentanalyse bestimmte häufig auftretende Wörter wie "gut" oder "schlecht" eine hohe Relevanz und sollten daher nicht abgewertet werden. Sinnvoll ist hingegen die Berücksichtigung von n-grams, da ansonsten einzelne Wörter wie "sehr" keine Bedeutung haben, da

¹²⁵ Buetow et al. 2013, S. 6

¹²⁶ Buetow et al. 2013, S. 6

¹²⁷ Buetow et al. 2013, S. 6

sie diese erst in Kombination, wie beispielsweise "sehr gut", oder "sehr schlecht" erhalten (siehe Abbildung 38).

	Negative	Neutral	Positive
Precision	69.3 %	71.3 %	72.3 %
Recall	29.8 %	96.1 %	26.3 %
Accuracy	71.2 %		

Abbildung 38: Ergebnisse ohne Berücksichtigung von 2-grams¹²⁸

Auch die Entfernung von Stop-Wörtern unterliegt bei der Sentimentanalyse gewissen Einschränkungen, da dieses Verfahren oftmals mithilfe von Listen durchgeführt wird, welche aber wiederum Wörter wie "sehr" enthalten können, die für die Bestimmung der Polarität relevant sind. Somit ist es auch hier notwendig im Vorfeld alle meinungstragenden Begriffe aus der Liste der Stop-Wörter zu entfernen bzw. muss erst einmal untersucht werden, ob gewisse Wörter in bestimmten Fällen, zum Beispiel bei der Verwendung von Ironie, in irgendeiner Form

	Negative	Neutral	Positive
Precision	65.7 %	75.6 %	75.0 %
Recall	48.9 %	90.7 %	40.6 %
Accuracy	74.0 %		

eine Meinung verkörpern (*Ergebnisse siehe Abbildung 39*).

Abbildung 39: Ergebnisse mit der Entfernung von Stop-Wörtern¹²⁹

Das stemming, was eine weitere Methode der Vorverarbeitung darstellt, kann nach Meinung der Autoren, aber direkt auf die Sentimentanalyse angewendet werden, da hierbei leichte Modifikationen eines Wortes auf dieselbe Stammform reduziert werden.¹³⁰

Bei seiner Veröffentlichung war SentiWS das erste deutschsprachige Wörterbuch zur Sentimentanalyse und deshalb entsprechend schwer zu evaluieren. Um dennoch die Performance einschätzen zu können, wurde ein Testkorpus von 480 Sätzen aus Internetforen erstellt, deren Länge zwischen 4 und 40 Wörtern variierte. Diese wurden daraufhin von zwei Juroren manuell bewertet, welche jedem Wort eine Polarität zuordneten (positiv, negativ oder keine Polarität). Die Ergebnisse wurden dann noch nach Wortgruppen getagged und mit den Einträgen im SentiWS verglichen. Dabei zeigte sich, dass SentiWS seine Stärken bei der Erkennung negativer Wörter hat (F-Maß 0,86), aber auch insgesamt zu überzeugen wusste (Precision=0,96, Recall = 0,74 und F-Maß = 0,84 siehe Abbildung 40).

¹²⁸ Buetow et al. 2013, S. 7

¹²⁹ Buetow et al. 2013, S. 7

¹³⁰ Buetow et al. 2013, S. 6

Typische Fehler bei der Erkennung (Recall) waren fehlende Wörter und Wortformen, domäne-spezifische Wörter und Fremdwörter. Fehler, die zu einer Verringerung der Präzision führten, waren hingegen Ambiguitäten, also Wörter, die in einem Satz positiv und in einem anderen negativ besetzt waren.¹³¹

	Pos			Neg			
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>	
Rater 1	0.94	0.70	0.80	0.99	0.74	0.85	C
Rater 2	0.89	0.62	0.73	0.97	0.72	0.83	C
Consensus	0.94	0.72	0.82	0.99	0.76	0.86	C

Abbildung 40: Evaluation von SentiWS¹³²

Für die Evaluation der Ergebnisse von GPC wurde wie bereits erwähnt, zunächst die Performance von Englischsprachigen Ressourcen getestet. Dabei schnitt Subjectivity Clues im Verhältnis zu der Anzahl seiner Feature am besten ab. Zudem konnte in dem Experiment gezeigt werden, dass subjektive erstellte Wörterbücher in Kombination mit maschinellem Lernen bes-

Resource	Model	F1-Positiv
English Subjectivity Clues	SVM-Linear	.83
	SVM-RBF	.82
English SentiWordNet	SVM-Linear	.83
	SVM-RBF	.81
English SentiSpin	SVM-Linear	.83
	SVM-RBF	.81
English Polarity Enhancement	SVM-Linear	.84

essere Ergebnisse zeigen als grundlegende Algorithmen wie Naive Bayes oder maximale Entropie (siehe Abbildung 41).

Abbildung 41: Evaluation von englischsprachigen Features¹³³

Diese Resultate wurden nun als Referenz für Beurteilung deutschsprachiger Sentiment Wörterbücher genutzt. Hier zeigte sich, dass selbst eine einfache Übersetzung der englischsprachigen Ressourcen schon eine Verbesserung der Performance erbrachte. Insgesamt konnte aber festgestellt werden, dass die Erstellung des neuen GPC nochmal eine Steigerung zu bereits vorhandenen Wörterbüchern bringt (siehe Abbildung 42).

¹³¹ Remus et al. 2010, S. 1170

¹³² Remus et al. 2010, S. 1171

¹³³ Waltinger 2010, S. 1642

Resource	Model	F1-Positive	F1-Negative	F1-Average
German SentiSpin Star1+2 vs. Star4+5	SVM-Linear	.827	.828	.828
	SVM-RBF	.830	.830	.830
German SentiSpin Star1 vs. Star5	SVM-Linear	.857	.861	.859
	SVM-RBF	.855	.858	.857
German Subjectivity Star1+2 vs. Star4+5	SVM-Linear	.810	.813	.811
	SVM-RBF	.804	.803	.803
German Subjectivity Star1 vs. Star5	SVM-Linear	.841	.842	.841
	SVM-RBF	.834	.834	.834
GermanPolarityClues Star1+2 vs. Star4+5	SVM-Linear	.875	.730	.803
	SVM-RBF	.866	.661	.758
GermanPolarityClues Star1 vs. Star5	SVM-Linear	.875	.876	.876
	SVM-RBF	.855	.850	.853

Abbildung 42: Evaluation von deutschsprachigen Sentiment Wörterbüchern¹³⁴

Dies ist insofern erstaunlich, da während der Evaluation nur 2700 Features genutzt wurden und damit deutlich weniger als beispielsweise bei der deutschen Version von SentiSpin, einem weiteren englischsprachigen Sentiment Wörterbuch. Somit scheint nach den Ergebnissen eine manuelle Erstellung von Vokabeln in Verbindung mit den zugehörigen Verneinungen und Synonymen eine sehr effektive Möglichkeit, um Sentimentanalyse in deutschen Texten durchzuführen.¹³⁵

¹³⁴ Waltinger 2010, S. 1642

¹³⁵ Waltinger 2010, S. 1640

6 Forensische Anwendungsgebiete

Im folgenden Kapitel soll mögliche Anwendungsmöglichkeiten der Sentimentanalyse am Beispiel des Forschungsgebiets der Forensik vorgestellt und diskutiert werden. Dabei wird auch auf bereits vorhandene Forschung in diesem Bereich Bezug genommen.

Eine erste Anwendung, welche man dem Gebiet der IT-Forensik oder der Cybersicherheit zuordnen könnte, ist die Untersuchung und Beobachtung von Radikalismus, Terrorismus und Extremismus. Die Verbreitung derartigen Gedankenguts ist eng mit der Entwicklung der sozialen Medien verbunden und wurde durch die mobile Nutzung mithilfe von Smartphones noch verstärkt. Deutschsprachige Forschung existiert für diese Anwendungen bisher nur in sehr geringem Maße. Vereinzelt vergibt die bereits erwähnte Interest Group for German Sentiment Analysis (IGGSA) Projekte in diesem Bereich, jedoch lassen sich auch grundlegende Erkenntnisse aus englischsprachigen Studien übertragen.

Die Kenianischen Wissenschaftlern Wanjala und Kahonge haben beispielsweise versucht durch Opinion Mining und Sentimentanalyse einen forensischen Beweis für Hass-Reden in den sozialen Netzwerken zu finden.¹³⁶ Als Datenmaterial nutzten sie dafür 20 öffentliche und beliebte Webseiten für die Verbreitung von Radikalismus, soziale Netzwerke wie Facebook und Twitter, sowie Blogs aus dem Dark Web. Diese Daten wurden mithilfe verschiedener Crawler und APIs gesammelt und in einer mySQL Datenbank gespeichert. Danach wurden die Daten für die Sentimentanalyse durch Tokenisierung und Part-of-Speech tagging vorverarbeitet und normalisiert. Für die Sentimentanalyse wurde das englischsprachige SentiWordNet zur Bewertung der Testdaten ausgewählt und der Algorithmus dann mithilfe von Naive Bayes trainiert.¹³⁷

¹³⁶ Wafula Wanjala und Kahonge 2016, S. 39

¹³⁷ Wafula Wanjala und Kahonge 2016, S. 40–41

Im Ergebnis wurden dann eine einzelne URLs in Verbindung mit jeweils einem positiven und negativen Wert der Polarität ausgegeben (siehe Abbildung 43).

Title	Link	Positive Sentiment	Negative Sentiment
senior unep employees celebrate tom Cholmondel ey for killing Kenyans -"	http://www.kahawatungu.com/2016/09/02/senior-unep-employees-celebrate-tom-cholmondel-ey-killing-kenyans/\n'	Positive: 0.3	Negative: 0.4

Abbildung 43: Polarität einzelner URLs¹³⁸

Darüber hinaus wurde eine bestimmte Facebook Gruppe genauer untersucht und die diskutierten Themen statistisch erfasst, was den Beweis für eine politische Form des Radikalismus erbrachte (siehe Abbildung 44).

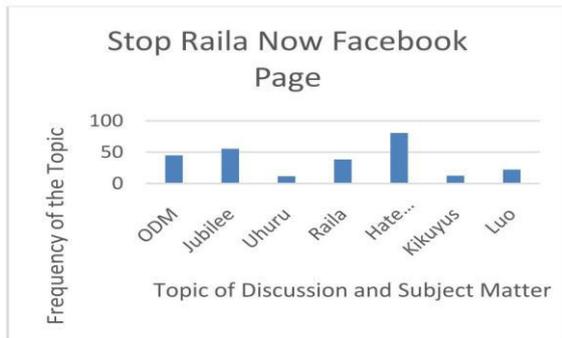


Abbildung 44: Verteilung der Themen einer beobachteten Facebook Gruppe¹³⁹

Ähnlich wurden auch die 20 Webseiten statistisch ausgewertet. Zusätzlich hat man aufgezeigt, welches Medium hauptsächlich für die Verbreitung von Hass-Reden genutzt wird (siehe Abbildung 45).¹⁴⁰

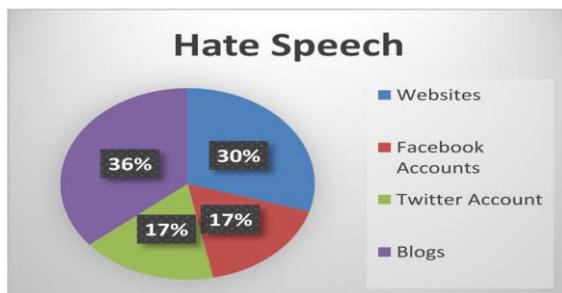


Abbildung 45: Verteilung der Medien zum Thema Hass-Reden¹⁴¹

¹³⁸ Wafula Wanjala und Kahonge 2016, S. 43

¹³⁹ Wafula Wanjala und Kahonge 2016, S. 44

¹⁴⁰ Wafula Wanjala und Kahonge 2016, S. 43–45

¹⁴¹ Wafula Wanjala und Kahonge 2016, S. 44

Schlussendlich kamen Wanjala et al. zu dem Fazit, dass Opinion Mining und Sentimentanalyse generell gut geeignet sind, um soziale Medien auf Hass-Reden zu untersuchen und deren Ergebnisse als Beweise zu sichern. Sie verzeichneten einen hohen Anteil des radikalen Gedankenguts in Blogs im Vergleich zu Webseiten oder Facebook Accounts und erklärten es sich damit, dass diese oftmals speziell anlässlich bestimmter politischer Ereignisse verfasst werden, wohingegen Facebook und Twitter Posts eher genereller emotionaler Natur sind. Zukünftiger Forschungsbedarf besteht nach Meinung der Autoren noch in der Verbesserung der Performance von Crawlern und Datenbanken, welche in der Lage sein müssen, das Netz noch tiefgründiger nach radikalem Datenmaterial zu durchsuchen und dieses aufzubereiten, wenn möglich sogar in Echtzeit. Dazu veröffentlichten sie eine API in der Programmiersprache PHP, die für die Analyse von sozialen Netzwerken eingesetzt und weiterentwickelt werden kann.¹⁴²

Für die deutsche Sprache entwickelten Hartung et al. einen lernbasierten Ansatz zur Identifikation von Rechtsextremismus anhand von Twitter Profilen. Der Ansatz basiert auf textlichen Hinweisen, sprachlichen Mustern und emotionalen Merkmalen, welche anhand bestimmter Features klassifiziert wurden (*siehe Abbildung 46*).

	Category N	Category R
words	heute, wir, beim, 's, the, uhr, geht, für_die, to, of, glückwunsch, Stellenangebot, feuerwehren, morgen, danke	d18, deutschland, #wehrdich, brd, deutsche, DTL, asylanten, Bürgerinitiative_wir, demonstration, asyl, herbstoffensive, asylbewerber, durchgeführt_!, volk, !!
network	nabu, ldkth, awo, mikemohring, bodoramelow, mdr_th, cdu_fraktion, naju, lmvth, der_AWO, lakoth16, DGB_Jugend, jef_de, dgb, bund_net	NPD, FN, suegida, identitäreaktion, weimarticker, sternde, spiegelonline, der_NPD, goldenerlöwe, npdde, Thüringer_NPD, agnordost, antifa, NPD-Landesverband, AB.ERFURT
local	erfurt, in_Erfurt	hildburghausen, Jena, nordhausen, Saalfeld, Erfurt_), suhl, Webetgasse, Bauvereinstraße, Hußstraße, Gera_Zschochernstraße
regional	Thüringen, thüringer, in_Thüringen, r2g, hochwasser, gebietsreform	kyffhäuserkreis, wir_lieben, lieben_Thüringen
national/global	TTIP, #mitredeneu, Bund, ews2014	#merkelmussweg, genozid
emotionality	— None —	EMO_TWEETS_PROP>0.3, MOST_FREQ_EMOTION=Disgust

Abbildung 46: Top Features zur Erkennung von Rechtsextremismus¹⁴³

¹⁴² Wafula Wanjala und Kahonge 2016, S. 45–46

¹⁴³ Hartung et al. 2017, S. 5

Ihrer Meinung nach ist es sinnvoller ganze Twitter Profile zu analysieren und nicht nur einzelne Tweets. Dazu teilten sie die Nutzerprofile in die Kategorien R (right-wing extremist) und N (non-extremist) ein. Zur Klassifikation wurde der SVM Algorithmus verwendet¹⁴⁴

Im Ergebnis erreichten sie eingeschränkte Precision von 25%, jedoch einen Recall von 95%, sodass der manuelle Arbeitsaufwand zur Erkennung von Rechtsextremismus durch diesen Ansatz stark reduziert wird.¹⁴⁵

Darüber hinaus entwickelten Al-Rowaily et al. ein zweisprachiges Lexikon zur Sentimentanalyse im Dark Web (BiSAL), was der Cyber-Sicherheit und Terrorismusbekämpfung dienen soll. Dies besteht aus einem Teil für Englisch Sprache (SentiLEN) und einem für arabische Wörter (SentiLAR). Dabei wurden für SentiLEN 279 Wörter identifiziert, die mit Cyber-Bedrohungen, Radikalismus o.ä. in Verbindung stehen (Auszug *siehe Abbildung 47*) und für SentiLAR 1019. Zum praktischen Einsatz von BiSAL wurde ein Web-Interface implementiert, was halb-automatisiert die Polarität von Meinungen ermittelt.¹⁴⁶

Words	Morphological variations	Sentiment polarity
ambush	ambush, ambushed, ambushes, ambushing, ambuscade	-0.86
arm	arm, armed, arms, arming	-0.54
assassin	assassination, assassinate, assassinating, assassinated, assassin, assassinator, assassigators	-0.95
assault	assault, assaulting, assaulted, assaulter, assaultive, assaulters	-0.84
attack	attack, attacked, attacking, attacks, attacker, attackers	-0.90
belief	belief, beliefs, believe, believer, believing, believed, believes, believable, believably, believing	+0.68
blast	blast, blasts, blasting, blaster, blasted	-0.61
blow	blow, blows, blew, blown, blower, blowing, blowy	-0.75
blood	blood, bloody, blooded, bleeding, bloods	+0.08
body	body, bodies, bodily, bodied, bodying	+0.08
bomb	bomb, bombs, bombing, bombings, bomber, bombers, bombed	-0.85
burn	burn, burning, burns, burned, burner, burnable	-0.69
business	business, businesses, businessman	+0.73
bust	bust, busts, buster, busty, busting, busted	-0.53
camp	camp, camps, camper, camping, campy, camped	-0.52
capture	capture, captured, captures, capturing, capturer, capturers	-0.35
casualty	casualty, casualties	-0.85
change	change, changes, changing, changed	+0.68
checkpoint	checkpoint, checkpoints	+0.08
chief	chief, chiefs	+0.73
support	support, supports, supporter, supporters, supporting, supported, supportive	+0.95

Abbildung 47: Liste von Einträgen des SentiLEN mit Polaritätswert¹⁴⁷

¹⁴⁴ Hartung et al. 2017, S. 2

¹⁴⁵ Hartung et al. 2017, S. 1

¹⁴⁶ Al-Rowaily et al. 2015, S. 53

¹⁴⁷ Al-Rowaily et al. 2015, S. 60

Die Wörter, welche für BiSAL verwendet wurden, stammen zum einen aus dem Department of Homeland Security (DHS) und einem Dark Web Portal. Die Wortliste aus dem DHS beinhaltet Begriffe, welche häufig in Zusammenhang mit Heimatschutz, Terrorismus oder Cyber-Sicherheit verwendet werden.

Diese wurden in diverse Kategorien eingeteilt, von denen aber lediglich sechs für das Thema der Studie interessant waren. Von diesen sechs Kategorien wurde mit Hilfe von Stammformenbildung (stemming) letztlich 99 Begriffe für das englischsprachige Wörterbuch SentiLEN ausgewählt (*siehe Abbildung 48*).

Category	No. of words	Exemplar words
Domestic security	53	assassination, attack, drill, exercise, recovery, shooting, evacuation, deaths, explosion, gangs, security, threat, bomb
HAZMAT and nuclear	34	hazmat, nuclear, toxic, plume, radiation, radioactive, chemical, biological, epidemic, hazardous, incident, infection
Terrorism	55	terrorism, terror, attack, target, jihad, extremism, radicals, plot, nationalist, fundamentalism
Infrastructure study	35	collapse, subway, power
South west border violence	63	violence, gang, drug, border, gunfight, trafficking, kidnap, bust
Cyber security	25	botnet, virus, trojan, hacker, worm, scammer
Total	265	

Abbildung 48: Kategorien des DHS mit Beispielwörtern zum Thema Cyberkriminalität¹⁴⁸

Die zweite Quelle, das Dark Web Portal ist eine Sammlung von Dark Web Foren, welche zu gewaltsamen Aktivitäten aufrufen und Informationen zu Cyber-Angriffen, Radikalismus oder ähnlichem vertreiben. Darunter fallen auch 29 Jihadisten Foren, welche überwiegend in englischer und arabischer Sprache geführt werden.

Words	Frequency
Weapons	918
Terrorists	738
Violence	1089
Brothers	1692
Militants	1680
Suicide	1148
Bomb	1317
Killing	1589
Fighting	1785
Death	1214
Injured	1008

Aus einem dieser Foren, Ansar1, wurden 205 Englische Begriffe als Seed Words für das Wörterbuch SentiLEN ausgewählt (Beispiele *siehe Abbildung 49*).

Abbildung 49: Liste von Begriffen und deren Häufigkeit aus dem Ansar1 Datensatz¹⁴⁹

¹⁴⁸ Al-Rowaily et al. 2015, S. 55

¹⁴⁹ Al-Rowaily et al. 2015, S. 54

Für das arabisch sprachige Wörterbuch SentiLAR wurden, die am häufigsten benutzten Begriffe des arabischen Alokab Forums, welches wiederum im Dark Web Portal enthalten ist, ausgewählt.¹⁵⁰

Zur Bestimmung der Polarität des SentiLEN wurden in BiSAL die vier Sentiment Wörterbücher, AFFIN, SentiWordNet, General Inquirer und SentiStrength herangezogen und deren Punktwerte in einer einheitlichen Funktion zusammengefasst. Da jedes der Wörterbücher eine andere Skala zur Bestimmung der Polarität verwendet, wurden die Wörter dazu zunächst auf die Skala [-1;1] normalisiert und danach wurde ein Durchschnittswert über alle vier Lexika gebildet (siehe Abbildung 50).¹⁵¹

Word(w_i)	AFFIN(d_1)		SentiWordNet(d_2)		GeneralInquirer(d_3)		SentiStrength(d_4)		$\hat{\delta}(w_i)$
	$P_0(w_i, d_1)$	$P_n(w_i, d_1)$	$P_0(w_i, d_2)$	$P_n(w_i, d_2)$	$P_0(w_i, d_3)$	$P_n(w_i, d_3)$	$P_0(w_i, d_4)$	$P_n(w_i, d_4)$	
	[-4,+2]	[-1,+1]	[-0.88,+0.63]	[-1,+1]	[-4,+2]	[-1,+1]	[-4,+3]	[-1,+1]	
Attack	-1	0	-0.25	-0.17	-4	-1.0	-3	-0.71	-0.47
Harm	-2	-0.33	-0.42	-0.39	-4	-1.0	-3	-0.71	-0.61
Bomb	-1	0	-0.19	-0.08	-4	-1.0	-2	-0.43	-0.38
Secret	0	+0.33	+0.13	+0.33	-1	0	-1	-0.14	+0.13

Abbildung 50: normalisierte und durchschnittliche Punktwerte aus dem SentiLEN¹⁵²

Da es für die arabische Sprache bisher kein Sentiment Wörterbuch gab, wurde mithilfe von drei wissenschaftlichen Experten ein eigenes erstellt. Dazu wurden die 2000 Nachrichten aus dem Alokab Dark Web Forum zunächst verschiedenen NLP Techniken unterzogen, wie beispielsweise dem Entfernen von Stop- Wörtern. Danach konnten die verbliebenen Begriffe von den Experten in die Kategorien positiv und negativ jeweils auf der Skala von [0,1] klassifiziert werden. Am Ende wurden noch diejenigen Wörter aussortiert, die von jedem Experten einen Punktwert von null erhalten haben, da diese keine Meinung transportieren, sodass letztendlich 1019 Wörter übrig blieben. Ähnlich wie bei den englischen Begriffen wurde anhand einer einheitlichen Formel der Durchschnittswert über alle drei Scores gebildet und auf die Skala [-1;1] normalisiert.¹⁵³

Ein weiteres forensisches Anwendungsgebiet, wo Sentimentanalyse eine sinnvolle Ergänzung darstellt, ist die Bestimmung der Urheberschaft einer Nachricht. In diesem Bereich werden

¹⁵⁰ Al-Rowaily et al. 2015, S. 54–55

¹⁵¹ Al-Rowaily et al. 2015, S. 55–57

¹⁵² Al-Rowaily et al. 2015, S. 58

¹⁵³ Al-Rowaily et al. 2015, S. 58–60

oftmals die N-grams, also die meinungstragenden Wortgruppen, statistisch ausgewertet und zumeist mit Cluster Algorithmen, wie KNN, einzelnen Autoren Profilen zugeordnet. Dieses Forschungsgebiet beschränkt sich aktuell jedoch ebenso zumeist auf die englische Sprache.¹⁵⁴

Die Sentimentanalyse spielt jedoch nicht nur in der IT-Forensik zukünftig eine wichtige Rolle, sondern kann auch auf Aufgabengebiete der allgemeinen Forensik, wie die forensische Psychologie oder die forensische Psychiatrie angewendet werden.

Ein Beispiel, deren Forschung ursprünglich dem rein medizinischen Bereich zuzuordnen ist, wäre die Ermittlung der Stimmungslage von Patienten, die unter schwerwiegenden Krankheitsbildern leiden, um diese gezielter psychologisch betreuen oder behandeln zu können. Beispielsweise bei Krebspatienten kann die Behandlung oder der Tod von Mitpatienten zu seelischen und emotionalen Störungen, wie Depressionen führen, was dem eigenen Heilungsverlauf nicht förderlich ist. Deshalb ist hier eine dauerhafte Beobachtung des Stimmungsbildes ein wichtiger Teil der Behandlung. Viele dieser Patienten nutzen zudem ein soziales Netzwerk um sich mit anderen Menschen über ihre Erkrankung auszutauschen oder sich gegenseitig zu unterstützen. Die Brasilianischen Wissenschaftler Rodrigues et al. untersuchten diese Thematik und veröffentlichten ein Sentimentanalyse Tool, mit Namen SentiHealth-Cancer (SHC-pt), was dazu dienen soll, den emotionalen Zustand von Krebspatienten anhand ihrer Facebook Posts zu bestimmen. Dazu sammelten sie Daten aus zwei Facebook Gruppen zum Thema Krebserkrankungen, welche überwiegend portugiesisch-sprachig waren und bestimmten die Polarität mit den semantischen Wörterbüchern Semantria und SentiStrength. Zur Evaluation der Ergebnisse übersetzten sie die Texte anschließend ins Englische und werteten die Ergebnisse aus. Mit einer Korrektklassifikationsrate (Accuracy) von durchschnittlich 66% lagen ihre Ergebnisse deutlich über den bisherigen zu diesem Thema.¹⁵⁵

Eine forensische Übertragung dieser Ergebnisse wäre beispielsweise hinsichtlich des Suizidrisikos der Patienten sinnvoll. Britische Forscher untersuchten bereits in diesem Zusammenhang Aufzeichnungen aus den sozialen Netzwerken auf Hinweise nach Selbstmordversuchen und konnten mittels überwachten Lernverfahren eine Korrektklassifikationsrate von 86,6% erreichen.¹⁵⁶

¹⁵⁴ Peng et al. 2016, S. 171

¹⁵⁵ Rodrigues et al. 2016, S. 80

¹⁵⁶ Schoene und Dethlefs 2016, S. 128

Auf eine ähnliche Anwendung fokussierten sich Eichstaedt et al., indem sie den Einfluss von negativen sozialen Beziehungen und negativen Emotionen auf die Sterblichkeit an arteriellen Herzkrankungen (AHD) untersuchten. Dazu analysierten sie 10 Millionen Tweets, welche überall in den USA versendet wurden und verglichen die dort geäußerten Meinungen mit der tatsächlichen Sterblichkeitsrate, welche vom Center for Disease Control and Prevention (CDC) veröffentlicht wurde (siehe *Abbildung 51*).¹⁵⁷

Sie kamen zu dem Ergebnis, dass die verwendete Sprache auf Twitter ein plausibler Indikator für das Risiko einer arteriellen Herzkrankung sein kann. Dabei erhöhen negative Beziehungen und Stress das Risiko, wohingegen positive Beziehungen vorbeugend wirken. Dasselbe gilt für die besprochenen Themen.¹⁵⁸

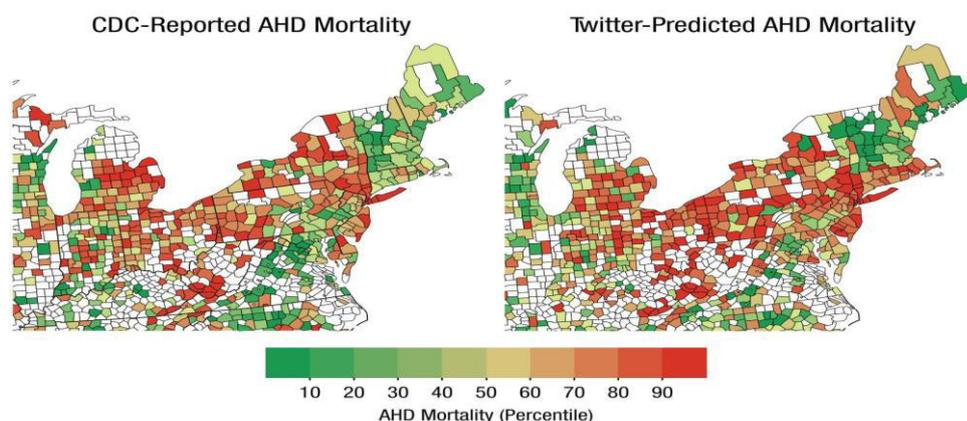


Abbildung 51: Landkarte mit Sterblichkeitsrisiko an AHD¹⁵⁹

Zudem kann Sentimentanalyse in sozialen Netzwerken zur Identifikation oder Vorhersage von Infektionen bzw. Seuchen genutzt werden. Liu et al. entwickelten in diesem Zusammenhang ein Tool, welches das physische Aktivitätslevel anhand von Twitterdaten in Echtzeit beobachtet.¹⁶⁰ Sie wiesen in ihrer Arbeit daraufhin, dass regelmäßige physische Aktivität ein wichtiger Faktor ist, um Krankheiten zu vermeiden und aktuell nur etwa 20 % aller erwachsenen Amerikaner das empfohlene Aktivitätslevel erfüllen (etwa 150 min. sportliche Betätigung von mittlerer Intensität pro Woche). Zwar existieren bereits Studien, welche darauf abzielen, Gruppen

¹⁵⁷ Eichstaedt et al. 2015, S. 1

¹⁵⁸ Eichstaedt et al. 2015, S. 6

¹⁵⁹ Eichstaedt et al. 2015

¹⁶⁰ Liu und Young 2016, S. 1

oder Populationen zu identifizieren, welche keiner regelmäßigen sportlichen Betätigung nachgehen, jedoch haben diese den Nachteil, dass die Erhebungen sehr zeit- und ressourcenintensiv sind und damit eine zeitliche Verschiebung von bis zu zwei Jahren vom Beobachten der Inaktivität bis zur Auswertung der Studie entsteht. Daher ist es nach Meinung der Autoren sinnvoll Daten aus sozialen Netzwerken auszuwerten, da diese in Echtzeit vorliegen und damit Hilfsorganisationen gezielter eingreifen können. Eine Möglichkeit diese Daten auszuwerten ist, neben der Modellierung des Themas oder einer Strukturanalyse des Netzwerks, die Sentimentanalyse.¹⁶¹

Des Weiteren untersuchten Denecke et al. die Anwendung von Sentimentanalyse auf klinische Dokumente, wie radiologischen Berichte, Entlassungspapiere oder medizinische Blogs und stellten fest, dass auch in diesen formal objektiven Dokumenten Meinungen zum Ausdruck gebracht werden, welche den Gesundheitszustand oder die Behandlung des Patienten betreffen. Im Gegensatz zu medizinischen Texten aus den sozialen Medien, enthalten klinische Dokumente zwar weniger meinungstragende Wörter, der generelle Ansatz lässt sich aber durchaus übertragen.¹⁶² Mögliche Klassen und Kriterien, nach welchen sich klinische Dokumente einstufen lassen sind in *Abbildung 52* zu sehen.

Entity	Possible sentiment values
Health status	Improve, worsen
Medical condition	Present, improve, worsen
Diagnosis	Certain, uncertain, preliminary
Effect of a medical event	Critical, non-critical
Medical procedure	Positive or negative outcome, successful or unsuccessful
Medication	Helpful, useless, serious adverse events

Abbildung 52: mögliche Sentiment Klassen im medizinischen Kontext¹⁶³

Zusammenfassend lässt sich sagen, dass Sentimentanalyse auch in forensischen Anwendungsgebieten durchaus eine sinnvolle Ergänzung darstellen kann. Dies gilt besonders für Aspekte der IT-Forensik oder der Cybersicherheit, wie Radikalismus oder Terrorismus. Dort ist es sinnvoll Meinungsäußerungen oder ganze Profile von verdächtigen Personen mit Hilfe der Sentimentanalyse zu klassifizieren. Aber auch Erkenntnisse der letztgenannten Studien, welche hauptsächlich aus medizinischen Gesichtspunkten veranlasst wurden, lassen sich auf

¹⁶¹ Liu und Young 2016, S. 1–2

¹⁶² Denecke und Deng 2015, S. 17

¹⁶³ Denecke und Deng 2015, S. 18

Bereiche der Forensik übertragen. Beispielsweise können die Ergebnisse von Eichstaedt et al. auch auf gerichtsmedizinische Untersuchungen, wie eine Herzerkrankung infolge von Cybermobbing erweitert werden. Genauso kann man den Ansatz von Denecke, dass klinische Dokumente, welche formal objektiv sind, auch Meinungen enthalten können, genauso gut auf forensische oder psychologische Gutachten anwenden, wenn dafür entsprechende Zielklassen formuliert werden. Beispielsweise könnte man anhand der Ergebnisse von körpereigenen Spuren, wie Knochenmaterial, DNA, Blut o.ä. die Todesursache einer Person klassifizieren. In diesem Zusammenhang untersuchten bereits Niederländische Wissenschaftler Protokolle von psychologischen Sitzungen nach Mustern zur Identifikation oder Vorhersage von Kindesmissbrauch und erreichten mithilfe von Naive Bayes und SVM Klassifikatoren sehr gute Ergebnisse.¹⁶⁴

Die meisten der bisherigen Ansätze zur Sentimentanalyse, welche sich auf die Forensik anwenden lassen, existieren zwar nur in englischer Sprache, jedoch lassen sie sich oftmals ohne größeren Aufwand auch auf die deutsche Sprache übertragen. Beispielsweise könnte man das englischsprachige Wörterbuch zur Terrorismus Erkennung SenitLEN einfach ins Deutsche übersetzen, man müsste bloß die Polaritätswerte der deutschen Entsprechung gegebenenfalls anpassen. Zudem existiert bisher auch noch kein einheitlicher Korpus an Trainingsdaten für die Forensik, weder in englischer noch in deutscher Sprache. Dies ist jedoch auch schwierig, da dieses Forschungsgebiet sehr breitgefächert ist, sodass man aufgrund der Spezifität der Algorithmen für jeden Teilbereich einen eigenen Korpus entwickeln müsste. Für einige der Teilgebiete der Forensik gibt es jedoch schon erste Ansätze, wie besonders die medizinischen Studien zeigen.

In den Anlagen werden einige der aktuellen Forschungen, welche direkt oder indirekt der Forensik zugeordnet werden können, noch einmal in Form einer Übersicht dargestellt.

¹⁶⁴ Amrit et al. 2017, S. 402

7 Zusammenfassung und Fazit

Die Auswertung diverser Studien hat gezeigt, dass soziale Netzwerke wie Facebook oder Twitter eine sehr gute Datenquelle für aktuelle Verfahren der Sentimentanalyse sind. Wie gut die Ergebnisse im Einzelnen sind, hängt jedoch stark von der Domäne und der damit verbundenen verfügbaren Datenmenge ab. Zudem ist es durch gewisse datenschutzrechtliche Einschränkungen schwierig Nutzer- oder Nutzergruppen genau zu filtern. Dies wird besonders im Anwendungsbereich der Forensik relevant. Beispielsweise ist bei Twitter weder das Alter noch die genaue geografische Position des Nutzers verfügbar, was eine Lokalisierung oder die genaue Identifizierung eines Verdächtigen erschwert. Dennoch ist die Forensik eine Domäne, wo Sentimentanalyse durchaus eine sinnvolle Ergänzung darstellen kann, sei es bei der Früherkennung von Terrorismus oder auch bei der Bestimmung des Suizidrisikos anhand von Daten aus sozialen Netzwerken.

Die Sentimentanalyse von sozialen Netzwerken kann aber auch zu Kontrolle, Manipulations- oder Missbrauchszwecken eingesetzt werden, beispielsweise indem totalitäre Regime die Einstellung der Bevölkerung zu Themen wie Politik oder Religion messen. Aber auch in einer Demokratie könnten Unternehmen dadurch die Einstellung ihrer Mitarbeiter zur Firma oder zu Vorgesetzten bestimmen. Dies muss einem beim Nutzen von öffentlichen Plattformen wie Twitter bewusst sein.¹⁶⁵

Betrachtet man die aktuellen Algorithmen zur Sentimentanalyse im Allgemeinen, verfügen diese schon über eine hohe Klassifikationsgüte, jedoch wird es wichtig sein diese noch weiter zu verfeinern. Ein wichtiger Schritt ist sicherlich die Übertragung dieser Verfahren auf die diversen Sprachen. Speziell für das Deutsche wurde durch die Entwicklung von semantischen Wörterbüchern wie dem SentiWS oder dem German Polarity Clues schon eine gute Basis gelegt. Auch wenn einem bewusst sein muss, dass die Feinheiten der menschlichen Sprache viel zu komplex sind, um mittels eines automatisierten Sprachverarbeitungssystems ein perfektes Ergebnis zu erreichen.

¹⁶⁵ Valerius 2012, S. 78–79

Im Hinblick auf die einzelnen Klassifikationsalgorithmen konnte sich keiner als vollkommen überlegen auszeichnen. Die lernbasierten Verfahren können zwar in Einzelfällen und in bestimmten Domänen sehr gute Ergebnisse liefern, jedoch sind sie wenig generalisierungsfähig, da die Trainingsdaten immer eine hohe Ähnlichkeit zu den Testdaten haben müssen. Möchte man lernbasierte Verfahren auf mehrere Sprachen anwenden, muss für jede Sprache eine eigene Trainingsdatenmenge erstellt werden, was einen erheblichen Aufwand bedeutet. Für eine Problematik oder Domäne, in der es bisher wenige bzw. keine annotierten Testdatensätze gibt, ist ein lexikalischer Ansatz somit besser geeignet. Dies würde auch auf das Forschungsgebiet der Forensik zutreffen.

Um wirklich gute Ergebnisse mit einem lexikalischen Ansatz zu erzielen, muss der Algorithmus jedoch nicht nur das reine Vorhandensein von Wörtern betrachten, sondern auch weitere Herausforderungen, die mit Nutzer generierten Inhalten einhergehen, berücksichtigen. Ein aktuelles Problem ist, dass die meisten Wörterbücher statischer Natur sind und damit nicht in Lage unbekannte Wörter zu klassifizieren. Gerade wenn es sich um Wörter mit großer Bedeutung für die Polarität eines Satzes handelt, bedeutet dies einen erheblichen Nachteil in der Genauigkeit der Klassifikation. Des Weiteren kann es Wörter geben, die je nach Kontext eines Satzes ironisch bzw. sarkastisch verwendet werden oder der Fall, dass in einem Satz verschiedene Meinungen verknüpft werden. Dies sind dann Sätze, die sehr schwer klassifiziert werden können. Die besondere Schwierigkeit besteht darin, dass gerade sprachliche Mittel wie Ironie oder Sarkasmus, nicht auf speziellen Begriffen oder Satzstellungen beruhen, sondern tiefgreifendes Wissen zum Kontext der Aussage erfordert. Spätestens wenn zur richtigen Interpretation einer Aussage Insiderwissen notwendig und dieses nicht öffentlich zugänglich ist, wird ein Algorithmus an dieser Aufgabe scheitern. Jedoch muss einschränkend gesagt werden, dass die fehlende Erkennung von Sarkasmus die Genauigkeit von Algorithmen nur unwesentlich verschlechtert, da diese Form nur einen sehr geringen Teil der Aussagen ausmachen.

Im Gegensatz zu der Erkennung von Ironie oder Sarkasmus, kann die Glaubwürdigkeit von Aussagen mithilfe der sozialen Netzwerkstruktur inzwischen schon recht gut bestimmt werden. Gerade zusätzliche Informationen, wie Links oder Retweets bei Twitter, geben dabei wichtige Hinweise.

Eine weitere Schwierigkeit besteht in der bereits genannten Abhängigkeit von Wörtern zu einer bestimmten Domäne, das bedeutet, dass sie je nach Themengebiet unterschiedlich bedeutsam für die Klassifikation einer Meinung sind. Außerdem ist es oftmals der Fall, dass Wörter unterschiedlich buchstabiert bzw. geschrieben werden, sodass es eine große Herausforderung darstellt, diese ihrer ursprünglichen Wortform und damit ihrer Polarität zuzuordnen. Zu-

dem stellt auch die häufige Vernachlässigung von Emotionsausdrücken (Emojis) bei den diversen Ansätzen ein Problem dar, da diese ein recht sicheres Indiz für die Stimmungsrichtung eines Satzes sind und damit für die Genauigkeit der Analyse einen deutlichen Mehrwert darstellen würden. ¹⁶⁶

Ein weiterer wichtiger Punkt, welcher zukünftig noch verbessert werden muss, ist die Erkennung von relevantem Datenmaterial. Beispielsweise haben Studien bei Twitter gezeigt, dass allein das Filtern anhand themenspezifischer Hashtags nicht ausreicht, um relevante Daten zu identifizieren. Um dem entgegenzuwirken, müsste man für jedes Themengebiet Listen mit Wörtern erstellen, welche einen hohen Bezug dazu haben bzw. dieses Thema ausschließen. Positiv hervorzuheben ist hierbei, dass immer mehr Anbieter von sozialen Netzwerken APIs zur Verfügung stellen, die es ermöglichen ausreichend Datenmaterial für die Analyse zu sammeln. ¹⁶⁷

Jedoch müssen einem, wie bereits am Anfang des Kapitels erwähnt, auch immer die rechtlichen Einschränkungen dieser Verfahren bewusst sein, beispielsweise beim Zugriff auf Nutzerdaten von Facebook oder Twitter. Dies spielt gerade in Anwendungsbereichen, wie der Forensik, welche in enger Zusammenarbeit mit dem Rechtssystem stehen, eine sehr wichtige Rolle. Zudem lassen sich oftmals nur gegenwärtige Daten gewinnen, da der Zugriff auf historisches Datenmaterial entweder beschränkt oder kostenintensiv ist, was im Budget der Forschung berücksichtigt werden sollte. Des Weiteren lassen sich die gewonnenen Erkenntnisse nicht immer verallgemeinern, da beispielsweise Twitter Nutzer generell jünger sind, als der allgemeine Internet Nutzer und man auch beim Download über die eigene API nicht sicher sein kann, dass man eine repräsentative Auswahl getroffen hat. Der letztgenannte Punkt ist dabei für die Anwendung der Sentimentanalyse in der Forensik vermutlich von untergeordneter Bedeutung, da dort oftmals Einzelfälle betrachtet werden, jedoch können auch dort repräsentative Gesamtaussagen, beispielsweise über eine beobachtete Facebook Gruppe von Bedeutung sein.

Weiterer Bedarf besteht auch in der Entwicklung von klaren Richtlinien oder standardisierten Protokollen zur Vorverarbeitung und Analyse von Daten aus sozialen Netzwerken, da bisher zwar eine Fülle von ähnlichen Herangehensweisen und Methoden existiert, aber kein eindeutig

¹⁶⁶ Kadge et al. 2016, S. 31

¹⁶⁷ Valerius 2012, S. 81–82

einheitliches Konzept.¹⁶⁸ Die ist auch dadurch bedingt, dass die meisten Sentimentanalyse-systeme von Forschern und kleinen Softwarefirmen entwickelt werden, welche ihre Erkenntnisse zwar in Form von Papern veröffentlichen, jedoch den Quellcode geheim halten. Bei Softwarefirmen kommt verstärkend noch der finanzielle Aspekt dazu. Dieser Sachverhalt erschwert aber das Implementieren und Weiterentwickeln von Frameworks zu diesem Thema und die einzelnen Entwickler können auch nicht voneinander profitieren. Es bleibt abzuwarten, wie die Entwicklung in diesem Bereich in Zukunft weitergeht, wobei davon auszugehen ist, dass die Nachfrage nach komplexen Sentimentanalyzesystemen zunehmen wird.¹⁶⁹

Zusammenfassend zur vorliegenden Arbeit, lässt sich festhalten, dass die Analyse von Daten aus sozialen Netzwerken ein enormes Potential bietet, welches sich auch auf das Gebiet der Forensik ausweiten lässt. Dabei kann neben der hier betrachteten Sentimentanalyse auch auf weitere Techniken, wie die Bestimmung des Themas einer Konversation (topic modeling) oder die Lokalisierung von Tweets zurückgegriffen werden. Dies alles kann genutzt werden, um Kriminalität in jeglicher Form zu beobachten oder vorherzusagen. Beispielsweise lassen sich damit die Strecken von Drogenhandel nachvollziehen oder es können gezielt Personen identifiziert werden, die Opfer von Gewalttaten wurden. Zur Realisierung eignen sich dafür besonders lexikalische Ansätze, da ihre Algorithmen robuster und domäne-unabhängiger sind als lernbasierte Ansätze. Zukünftig wird es jedoch sinnvoll sein, beide Ansätze in Form von hybriden Verfahren sinnvoll zu verknüpfen, um auch die Stärken manuellen Lernens, wie das Einbeziehen von Kontext oder die höhere Präzision nutzen zu können. Letztendlich wird man aber nie ganz auf eine subjektive und manuelle Reflexion der Ergebnisse verzichten können, da die menschliche Einschätzung auch durch moderne Algorithmen vermutlich nicht zu ersetzen ist.¹⁷⁰

¹⁶⁸ Liu und Young 2016, S. 3–4

¹⁶⁹ Valerius 2012, S. 80–81

¹⁷⁰ Schmidt 2012, S. 10–11

Literatur

Adarsh M J; Pushpa Ravikumar, PhD (2015): Survey: Twitter data Analysis using Opinion Mining. In: *International Journal of Computer Applications* 128 (5), S. 34–36. DOI: 10.1002/9781119197249.ch1.

Al-Rowaily, Khalid; Abulaish, Muhammad; Al-Hasan Haldar, Nur; Al-Rubaian, Majed (2015): BiSAL - A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. In: *Digital Investigation* 14, S. 53–62. DOI: 10.1016/j.diin.2015.07.006.

Alvanaki, Foteini (2014): Mining Interesting Events on Large and Dynamic Data. Dissertation. Universität des Saarlandes, Saarbrücken, zuletzt geprüft am 03.07.2017.

Alvares, Benito; Thakur, Nishant; Patil, Siddhi; Fernandes, Daniel; Jain, Kavita (2016): Sentiment Analysis using Opinion Mining. In: *International Journal of Engineering Research & Technology (IJERT)* (5), zuletzt geprüft am 03.07.2017.

Amrit, Chintan; Paauw, Tim; Aly, Robin; Lavric, Miha (2017): Identifying child abuse through text mining and machine learning. In: *Expert Systems with Applications* 88, S. 402–418. DOI: 10.1016/j.eswa.2017.06.035.

Anwar Hridoy, Syed Akib; Ekram, M. Tahmid; Islam, Mohammad Samiul; Ahmed, Faysal; Rahman, Rashedur M. (2015): Localized twitter opinion mining using sentiment analysis. In: *Decis. Anal.* 2 (1). DOI: 10.1186/s40165-015-0016-4.

Boland, Katarina; Wira-Alam, Andias; Messerschmidt, Reinhard (2013): Creating an Annotated Corpus for Sentiment Analysis of German Product Reviews. GESIS - Leibniz Institute for the Social Sciences, zuletzt geprüft am 17.07.2017.

Buetow, Florian; Schultze, Florian; Strauch, Leopold (2013): Semantic Search: Sentiment Analysis with Machine Learning Algorithms on German News Articles. Technische Universität Berlin. Berlin, zuletzt geprüft am 17.07.2017.

Dashtipour, Kia; Poria, Soujanya; Hussain, Amir; Cambria, Erik; Hawalah, Ahmad Y. A.; Gelbukh, Alexander; Zhou, Qiang (2016): Multilingual Sentiment Analysis. State of the Art and

Independent Comparison of Techniques. In: *Cognitive Computation* 8, S. 757–771. DOI: 10.1007/s12559-016-9415-7.

Denecke, Kerstin; Deng, Yihan (2015): Sentiment analysis in medical settings. New opportunities and challenges. In: *Artificial intelligence in medicine* 64 (1), S. 17–27. DOI: 10.1016/j.art-med.2015.03.006.

Eichstaedt, Johannes C.; Schwartz, Hansen Andrew; Kern, Margaret L.; Park, Gregory; Lubarthe, Darwin R.; Merchant, Raina M. et al. (2015): Psychological language on Twitter predicts county-level heart disease mortality. In: *Psychological science* 26 (2), S. 159–169. DOI: 10.1177/0956797614557867.

Gindl, Stefan; Remus, Robert; Clematide, Simon; Waltinger, Ulli; Petrakis, Stefanos; Klenner, Manfred et al. (2012): MLSA — A Multi-layered Reference Corpus for German Sentiment Analysis, zuletzt geprüft am 17.07.2017.

Godbole, Namrata; Srinivasaiah, Manjunath; Skiena, Steven (2007): Large-Scale Sentiment Analysis for News and Blogs. Hg. v. Google inc., zuletzt geprüft am 23.10.2017.

Gull, Ratab; Shoaib, Umar; Rasheed, Saba; Abid, Washma; Zahoor, Beenish (2016): Pre Processing of Twitter's Data for Opinion Mining in Political Context. In: *Procedia Computer Science* 96, S. 1560–1570. DOI: 10.1016/j.procs.2016.08.203.

Hartung, Matthias; Klinger, Roman; Schmidtke, Franziska; Vogel, Lars (2017): Identifying Right-Wing Extremism in German Twitter Profiles: a Classification Approach. Hg. v. Har (Lecture notes in computer science), zuletzt geprüft am 22.08.2017.

Huang, Faliang; Zhang, Shichao; Zhang, Jilian; Yu, Ge (2017): Multimodal learning for topic sentiment analysis in microblogging. In: *Neurocomputing* 253, S. 144–153. DOI: 10.1016/j.neucom.2016.10.086.

Johnson, Thomas J.; Kaye, Barbara K. (2015): Reasons to believe. Influence of credibility on motivations for using social networks. In: *Computers in Human Behavior* 50, S. 544–555. DOI: 10.1016/j.chb.2015.04.002.

Justo, Raquel; Corcoran, Thomas; Lukin, Stephanie M.; Walker, Marilyn; Torres, M. Inés (2014): Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. In: *Knowledge-Based Systems* 69, S. 124–133. DOI: 10.1016/j.knosys.2014.05.021.

Kadge, Sanam; Panchbhai, Saba; Padghawala, Sakina; Chhipa, Kausar (2016): Emoticon Analysis with Dynamic Text based Opinion Mining. In: *IJCA* 136 (9), S. 29–36. DOI: 10.5120/ijca2016908573.

Klinger, Roman; Ling, Jennifer (2016): An Empirical, Quantitative Analysis of the Differences between Sarcasm and Irony. Universität Stuttgart. Stuttgart (Lecture notes in computer science), zuletzt geprüft am 22.08.2017.

Kumar, K. NandhaG.; T, Christopher (2015): Opinion Mining: A Survey. In: *International Journal of Computer Applications* (113), zuletzt geprüft am 03.07.2017.

Kunneman, Florian; Liebrecht, Christine; van Mulken, Margot; van den Bosch, Antal (2015): Signaling sarcasm. From hyperbole to hashtag. In: *Information Processing & Management* 51 (4), S. 500–509. DOI: 10.1016/j.ipm.2014.07.006.

Kuo, Yau-Hwang; Fu, Meng-Hsuan; Tsai, Wen-Hao; Lee, Kuan-Rong; Chen, Ling-Yu (2016): Integrated microblog sentiment analysis from users? Social interaction patterns and textual opinions. In: *Appl Intell* 44 (2), S. 399–413. DOI: 10.1007/s10489-015-0700-z.

Labudde, Dirk; Spranger, Michael (2017): Forensik in der digitalen Welt. Berlin, Heidelberg: Springer Berlin Heidelberg, zuletzt geprüft am 03.10.2017.

Liu, Sam; Young, Sean D. (2016): A survey of social media data analysis for physical activity surveillance. In: *Journal of Forensic and Legal Medicine*. DOI: 10.1016/j.jflm.2016.10.019.

Mario Saenger; Ulf Leser; Peter Adolphs; Steffen Kemmerer; Roman Klinger: SCARE -- The Sentiment Corpus of App Reviews with Fine-grained Annotations in German, zuletzt geprüft am 22.08.2017.

Mark Cieliebak; Jan Milan Deriu; Dominic Egger; Fatih Uzdilli (2017): A Twitter Corpus and Benchmark Resources for German Sentiment Analysis. In: *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, S. 45–51, zuletzt geprüft am 17.07.2017.

Meier, Stefan: Digitale Forensik in Unternehmen, zuletzt geprüft am 03.10.2017.

Meire, Matthijs; Ballings, Michel; van den Poel, Dirk (2016): The added value of auxiliary data in sentiment analysis of Facebook posts. In: *Decision Support Systems* 89, S. 98–112. DOI: 10.1016/j.dss.2016.06.013.

- Ortigosa, Alvaro; Martín, José M.; Carro, Rosa M. (2014): Sentiment analysis in Facebook and its application to e-learning. In: *Computers in Human Behavior* 31, S. 527–541. DOI: 10.1016/j.chb.2013.05.024.
- Patel, Vishakha; Prabhu, Gayatri; Bhowmick, Kiran (2015): A Survey of Opinion Mining and Sentiment Analysis. In: *IJCA* 131 (1), S. 24–27. DOI: 10.5120/ijca2015907218.
- Peng, Jian; Choo, Kim-Kwang Raymond; Ashman, Helen (2016): Bit-level n -gram based forensic authorship analysis on social media. Identifying individuals from linguistic profiles. In: *Journal of Network and Computer Applications* 70, S. 171–182. DOI: 10.1016/j.jnca.2016.04.001.
- Piryani, R.; Madhavi, D.; Singh, V. K. (2017): Analytical mapping of opinion mining and sentiment analysis research during 2000? 2015. In: *Information Processing & Management* 53 (1), S. 122–150. DOI: 10.1016/j.ipm.2016.07.001.
- Remus, Robert; Quasthoff, Uwe; Heyer, Gerhard (2010): SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. Universität Leipzig. Leipzig, zuletzt geprüft am 17.07.2017.
- Rill, Sven; Adolph, Sven; Drescher, Johannes; Reinel, Dirk; Scheidt, Jörg; Schütz, Oliver et al. (2012): A phrase-based opinion list for the German language. University of Applied Sciences Hof. Hof, zuletzt geprüft am 17.07.2017.
- Rodrigues, Ramon Gouveia; das Dores, Rafael Marques; Camilo-Junior, Celso G.; Rosa, Thierson Couto (2016): SentiHealth-Cancer. A sentiment analysis tool to help detecting mood of patients in online social networks. In: *International journal of medical informatics* 85 (1), S. 80–95. DOI: 10.1016/j.ijmedinf.2015.09.007.
- Roman Klinger; Philipp Cimiano: The USAGE review corpus for fine-grained, multi-lingual opinion analysis, zuletzt geprüft am 21.08.2017.
- Schmidt, Irina (2012): SentiStrength. Universität Koblenz-Landau. Koblenz, zuletzt geprüft am 27.09.2017.
- Schoene, Annika Marie; Dethlefs, Nina (2016): Automatic Identification of Suicide Notes from Linguistic and Sentiment Features. In: Nils Reiter, Beatrice Alex und Kalliopi A. Zervanou (Hg.): Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural He-

Literatur

ritage, Social Sciences, and Humanities. Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. Berlin, Germany, 01.08.2016 - 01.08.2016. Stroudsburg, PA, USA: Association for Computational Linguistics, S. 128–133, zuletzt geprüft am 16.10.2017.

Anlagen

Publikationen zu forensischen Anwendungsgebieten der Sentimentanalyse

Autor	Jahr	Ziel	Sprache	Datenquelle	Methode	Evaluation
Al-Rowaily et al.	2015	Erstellung eines bilingualen Wörterbuchs zur Terroris- muserkennung	Englisch / Arabisch	Staatliche Be- hörde (DHS), Dark Web Portal	Lexikalisch, ei- genes Wörter- buch	-
Amrit et al.	2017	Mustererkennung bei medi- zinischen Daten zur Identifi- zierung von Kindesmiss- brauch	Englisch	Hilfsorganisation (GGD)	Lernbasiert, SVM, Naive Bayes	Accuracy = 0,822, Recall = 0,870
Brocardo	2015	Verifizierung von Autoren anhand unstrukturierter Da- ten	Englisch	Twitter, Email (En- ron Korpus)	Lernbasiert, SVM	Equal Error Rate (EER) = min. 9,98 %
Denecke et al.	2015	Aufzeigen von Anwen- dungsmöglichkeiten der Sentimentanalyse auf medi- zinische Dokumente	Englisch	Blogs (WebMD), Datenbank (MIMC II), DrugRa- tingz.com	Lexikalisch, SentiWordNet	-
Eichstaedt	2015	Vorhersage von Herzer- krankungen anhand der bei Twitter verwendeten Spra- che	Englisch	Twitter, öffentliche Behörde (CDC)	Lexikalisch, ei- gene Wörter- bücher	Korrelations- koeffizient nach Pear- son (r) = max. 0,45
Hartung	2017	Identifizierung von Rechts- extremismus anhand von Twitter Profilen	Deutsch	Twitter	Lexikalisch, eignes Wörter- buch	-
Liu	2016	Echtzeit-Überwachung des physischen Aktivitätslevels	Englisch	Twitter	Lexikalisch, SentiWordNet	-
Nirkhi	2016	Verifizierung des Autors von Online Nachrichten	Englisch	Enron Korpus	Lernbasiert, Cluster	-

Anlagen

Peng	2015	Bestimmung der Urheber- schaft in sozialen Netzwer- ken	Englisch		Lernbasiert, Cluster	Accuracy = max. 0,97
Rodrigues	2016	Bestimmung der Stim- mungslage von Krebspati- enten anhand ihrer Aktivität in sozialen Netzwerken	Portugie- sisch, Englisch	Facebook	Lexikalisch, Semantria, SentiStrength	Accuracy = min. 0,58
Schoene	2016	Identifizierung von Ab- schiedsbriefen aus unstruk- turierten Daten	Englisch	Genuine Suicide Notes (GSN)	Lernbasiert, Naive Bayes	Accuracy = 0,86
Schrading	2015	Analyse von häuslichem Missbrauch anhand von Da- ten aus sozialen Netzwer- ken	Englisch	Twitter, Reddit	Lernbasiert, SVM, neuro- nale Netze	Accuracy = 0,78
Wafula Wanjala	2016	Identifikation von Hass Re- den aus sozialen Netzwer- ken	Englisch	Facebook, Twit- ter, Blogs	Lernbasiert, Naive Bayes	-

Selbstständigkeitserklärung

Hiermit erkläre ich, dass die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Chemnitz, den 24.10.2017

Tobias Münzner