

Hochschule Rhein-Waal  
Fakultät Kommunikation und Umwelt  
Prof. Dr. Margarita Spirova  
Prof. Dr. Frank Zimmer

**Analyse der aktuellen und vergangenen politischen Landschaft Deutschlands im  
Hinblick auf die aktuelle Bundestagswahl mit Fokus auf die ansteigende Bedeu-  
tung von Social-Media im Wahlkampf**

**Bachelorarbeit**

vorgelegt von  
Jonathan Scheil

## Abstract

Since the last federal election in Germany, the CDU is no longer the strongest force and suffered large losses of votes. The SPD, Bündnis 90/Die Grünen and the FDP now make up the German parliament. Parts of the election campaign are taking place via social media. But how do the individual parties use modern media? Do the parties differ in the way they communicate their messages and, as a result, do they generate votes in different ways? I will examine these questions in more detail in this bachelor's thesis, starting with the reunification of Germany and including the last federal elections. The social media platform I am investigating for political communication is Twitter.

I presented the past and current parliamentary election results and differences in eastern and western Germany. To do this, I used sentiment analysis to examine tweets extracted from Twitter at the federal and local/state political level, for their content and the accompanying reactions. I created a detailed representation of the different social media appearances of politicians and determined whether appropriate conclusions can be drawn based on my results and data.

My analysis shows that at the parliamentary level, Annalena Baerbock of Bündnis 90/Die Grünen, followed by Dr. Alice Weidel of the AfD and Christian Lindner of the FDP, have the most likes and retweets. Olaf Scholz, with 826 tweets, is the most active on Twitter in 2021. At federal and local/state level, the topics of all parties are often similar. There are differences with regard to the core issues of the respective parties. I have illustrated this in each case.

Based on the politicians' use of Twitter, I could not draw any concrete conclusions about voter behaviour. For this it would be important to evaluate other media, e.g. Facebook.

## Inhaltsverzeichnis

1	Einleitung .....	1
2	Hauptteil.....	3
2.1	Die politische Lage Deutschlands, vergangene und aktuelle Situation .....	3
2.2	Thema Digitalisierung und Social-Media .....	7
2.3	Twitter .....	10
2.4	Data Mining.....	11
2.5	Text Mining .....	12
2.6	Opinion Mining .....	15
2.7	Machine learning Methoden .....	16
2.8	Überwachte Klassifikationsmethoden.....	17
2.8.1	Probabilistische Klassifikationen.....	17
2.8.2	Naive Bayes .....	17
2.8.3	Nicht-probabilistische Klassifikationen .....	18
2.8.4	Support Vektormaschinen (SVM) .....	18
2.9	Teilüberwachte Klassifikationsmethoden.....	18
2.9.1	Selbsttrainierte Ansätze .....	19
2.10	Nichtüberwachte Klassifikationsmethoden (Clustering) .....	19
2.10.1	Hierarchisches Clustering.....	20
2.10.2	Partitionelles Clustering.....	20
2.10.3	K-means.....	20
2.11	Sentiment Analysen .....	21
2.12	Ansätze von Sentiment Analysemethoden .....	21
2.13	Anwendungsebenen von Sentiment Analysen .....	23
2.13.1	Dokumentenebene .....	23
2.13.2	Satzebene .....	24
2.13.3	Aspektebene .....	24
2.13.4	Konzeptebene .....	25
2.14	Technische Umsetzung.....	25
2.14.1	Pandas.....	25
2.14.2	Numpy.....	25
2.14.3	Tweepy.....	26
2.14.4	TextBlob .....	26
2.14.5	NLTK.....	27
2.14.6	Matplotlib.....	27
2.14.7	Microsoft Excel .....	28
2.14.8	Power BI.....	28
3	Analysen der Politiker mit eingangs genannter Technik .....	28

3.1	Bundestagebene .....	32
3.1.1	Olaf Scholz, SPD.....	33
3.1.2	Armin Laschet, CDU.....	35
3.1.3	Annalena Baerbock, Bündnis 90/Die Grünen .....	37
3.1.4	Christian Lindner, FDP .....	39
3.1.5	Dr. Alice Weidel, AfD.....	41
3.1.6	Janine Wissler, Die Linke .....	43
3.2	Repräsentative Bundesländer Ost- und Westdeutschland.....	44
3.2.1	Kommunal-/Landesebene Westdeutschland .....	45
3.2.2	Kommunale-/Landesebene Ostdeutschland:.....	48
4	Zusammenfassung der Ergebnisse .....	58
4.1	Technische Limitationen der durchgeführten Arbeit.....	58
4.2	Ergebnisse der Auswertungen .....	59
4.2.1	Bundesebene.....	59
4.2.2	Kommunal-/Landesebene .....	60
4.3	Interpretation der Ergebnisse .....	63
5	Fazit.....	64
6	Literaturverzeichnis .....	66
7	Anhang .....	71
7.1	Anhangsverzeichnis.....	71
7.2	Visualisierungen Bundestagebene .....	72

## Abbildungsverzeichnis

Abbildung 1: Anzahl der Sitze der Parteien im Deutschen Bundestag von 1949 bis 2021 (Bundeswahlleiter, zitiert nach Statista 2021).....	5
Abbildung 2: Ergebnis der Bundestagswahl in Westdeutschland am 26. September 2021 (Zweitstimmen), (Infratest dimap, zitiert nach Statista 2021).....	5
Abbildung 3: Ergebnis der Bundestagswahl in Ostdeutschland am 26. September 2021 (Zweitstimmen), (Infratest dimap, zitiert nach Statista 2021).....	6
Abbildung 4: Wahlverhalten bei der Bundestagswahl am 26. September 2021 nach Alter (Stimmenanteile der Parteien), (Infratest dimap, zitiert nach Statista 2021).....	7
Abbildung 5: Durchschnittliche tägliche Nutzungsdauer ausgewählter Medien in Deutschland in den Jahren 2014 bis 2021 (in Minuten), (Seven.One Media GmbH, zitiert nach Statista 2021).....	9
Abbildung 6: Mediennutzung durch Jugendliche in der Freizeit in Deutschland 2021, eigene Abbildung, basierend auf Medienpädagogischer Forschungsverbund Südwest 2021.....	10
Abbildung 7: Data-Mining-Prozesse, eigene Abbildung, basierend auf Cleve und Lämmel, 2020, S. 5.....	12
Abbildung 8: Text-Mining-Prozess, eigene Abbildung, basierend auf Hippner und Rentzmann, 2006, S. 288.....	12
Abbildung 9: Machine Learning Methoden, eigene Abbildung, basierend auf Hemmatian & Sohrabi, 2017, S. 1507.....	16
Abbildung 10: Opinion Mining Ebenen, eigene Abbildung, basierend auf Hemmatian & Sohrabi, 2017, S. 1502.....	23
Abbildung 11: Anzahl aller im Jahre 2021 veröffentlichten Tweets, je Partei und Bundespolitiker gruppiert.....	32
Abbildung 12: Sentiment Verteilung und Anzahl Tweets, Olaf Scholz, SPD.....	33
Abbildung 13: Verteilung Polarität und Subjektivität von Tweets, Olaf Scholz, SPD.....	33
Abbildung 14: Wordcloud, Olaf Scholz, SPD.....	34
Abbildung 15: Mittelwert Likes je Sentiment, Olaf Scholz, SPD.....	35
Abbildung 16: Mittelwert Retweets je Sentiment, Olaf Scholz, SPD.....	35
Abbildung 17: Wordcloud, Armin Laschet, CDU.....	36
Abbildung 18: Mittelwert Likes je Sentiment, Armin Laschet, CDU.....	36
Abbildung 19: Mittelwert Retweets je Sentiment, Armin Laschet, CDU.....	37
Abbildung 20: Wordcloud, Annalena Baerbock, Bündnis 90/Die Grünen.....	38
Abbildung 21: Mittelwert Likes je Sentiment, Annalena Baerbock, Bündnis 90/Die Grünen.....	38
Abbildung 22: Mittelwert Retweets je Sentiment, Annalena Baerbock, Bündnis 90/Die Grünen.....	39
Abbildung 23: Wordcloud, Christian Lindner, FDP.....	39

Abbildung 24: Mittelwert Likes je Sentiment, Christian Lindner, FDP.....	40
Abbildung 25: Mittelwert Retweets je Sentiment, Christian Lindner, FDP.....	40
Abbildung 26: Wordcloud, Dr. Alice Weidel, AfD.....	41
Abbildung 27: Mittelwert Likes je Sentiment, Dr. Alice Weidel, AfD.....	42
Abbildung 28: Mittelwert Retweets je Sentiment, Dr. Alice Weidel, AfD.....	42
Abbildung 29: Wordcloud, Janine Wissler, Die Linke.....	43
Abbildung 30: Mittelwert Likes je Sentiment, Janine Wissler, Die Linke.....	44
Abbildung 31: Mittelwert Retweets je Sentiment, Janine Wissler, Die Linke.....	44
Abbildung 32: Anzahl aller Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland..	45
Abbildung 33: Anzahl aller negativen Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland.....	46
Abbildung 34: Anzahl aller positiven Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland.....	47
Abbildung 35: Anzahl aller neutralen Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland.....	47
Abbildung 36: Anzahl aller Tweets nach Parteien, Sachsen, Ostdeutschland.....	48
Abbildung 37: Anzahl aller negativen Tweets nach Parteien, Sachsen, Ostdeutschland.....	49
Abbildung 38: Anzahl aller positiven Tweets nach Parteien, Sachsen, Ostdeutschland.	50
Abbildung 39: Anzahl aller neutralen Tweets nach Parteien, Sachsen, Ostdeutschland.....	50
Abbildung 40: Anzahl der Politiker nach Social-Media-Kanal und Partei, Rheinland-Pfalz, Westdeutschland.....	52
Abbildung 41: Prozentuale Nutzung von Social-Media je Partei Rheinland-Pfalz Westdeutschland.....	52
Abbildung 42: Anzahl der Politiker nach Social-Media-Kanal und Partei, Sachsen, Ostdeutschland.....	53
Abbildung 43: Prozentuale Nutzung von Social-Media-Kanälen je Partei, Sachsen, Ostdeutschland.....	54
Abbildung 44: Wordcloud, Andreas Bleck, AfD Rheinland-Pfalz, Westdeutschland.....	55
Abbildung 45: Wordcloud, Edgar Naujok, AfD Sachsen, Ostdeutschland.....	55
Abbildung 46: Wordcloud, Thomas Hitschler, SPD Rheinland-Pfalz, Westdeutschland.	55
Abbildung 47: Wordcloud, Holger Mann, SPD Sachsen, Ostdeutschland.....	55
Abbildung 48: Wordcloud, Dr. Katja Leikert, CDU Hessen, Westdeutschland.....	56
Abbildung 49: Wordcloud, Dr. Otilie Klein, CDU Berlin, Ostdeutschland.....	56
Abbildung 50: Wordcloud, Helge Limburg, Bündnis 90/Die Grünen Niedersachsen, Westdeutschland.....	56

Abbildung 51: Wordcloud, Bernhard, Herrmann, Bündnis 90/Die Grünen Sachsen, Ostdeutschland.....	56
Abbildung 52: Wordcloud, Katja Adler, FDP Hessen, Westdeutschland.....	57
Abbildung 53: Wordcloud, Philipp Hartewig, FDP Sachsen, Ostdeutschland.....	57
Abbildung 54: Wordcloud, Heidi Rechinnek, Die Linke Niedersachsen, Westdeutschland.....	57
Abbildung 55: Wordcloud, Clara Bünger, Die Linke Sachsen, Ostdeutschland.....	57

## 1 Einleitung

Die Bundestagswahl 2021 hat aufgezeigt, dass sich die politische Landschaft in Deutschland stark verändert. Die CDU stellt erstmals seit 16 Jahren keinen Bundeskanzler/in mehr. Angela Merkels Ära geht zu Ende. Die Ampel-Koalition hat sich nach Wochen der Verhandlungen durchgesetzt (SPD, GRÜNE, FDP). Olaf Scholz, ehemaliger Finanzminister, Stellvertreter der Bundeskanzlerin ab März 2018 und vorhergehender Bürgermeister von Hamburg (2001- 2018), ist zum neuen Bundeskanzler ernannt worden. Im Vergleich zur Bundestagswahl 2017, musste die CDU bei den Erststimmen deutliche Verluste hinnehmen (-7,7%). Im Gegensatz dazu konnten die Bündnis 90/Die Grünen deutliche Stimmenanteile der Bevölkerung dazugewinnen (+5,9%). Einen leichten Zuwachs an Wählerstimmen konnte die SPD verzeichnen (+1,8%) sowie die FDP (+1,7%). Die AfD (-1,3%) und die Linke (-3,6%) verloren Wählerstimmen (Bundeswahlleiter, 2022).

Die Politik befindet sich mit der neuen Ampelkoalition und neuen schwierigen Aufgaben, wie z.B. der Bekämpfung der Coronapandemie, in einem großen Umbruch, der auch die nächsten Jahre in Deutschland stark prägen wird.

Aufgrund dieses großen Umbruchs möchte ich mich in dieser Bachelorarbeit auch der vergangenen und aktuellen politischen Landschaft in Deutschland widmen. Konkret heißt dies, dass ich mich dem Zeitraum zwischen dem Mauerfall und der Wiedervereinigung Deutschlands (9.11.1990) und der Bundestagswahl des Jahres 2021 (26.9.2021) annehme. Dabei gehe ich auf die allgemeinen politischen Wahlergebnisse ein. Außerdem werde ich mögliche regionale Unterschiede in Ost- und Westdeutschland aufzeigen und dementsprechend, vergangene und aktuelle Ergebnisse auf Bundestagebene analysieren.

Da Social-Media in den letzten Jahren einen immer größeren Stellenwert im Leben der Bürger Deutschlands eingenommen hat und auch der Wahlkampf in der Politik digitaler wird, habe ich mich außerdem dazu entschieden, eine Sentiment Analyse für Twitter durchzuführen. Ich analysiere dazu die Tweets des Jahres 2021 auf Bundestags- und (ausgewählter) Kommunal-/Landesebene der unterschiedlichen Parteien in Ost- und Westdeutschland. Diese visualisiere ich anschließend und ziehe mögliche Rückschlüsse über das Social-Media Verhalten der einzelnen Politiker der Parteien.

Ein wichtig zu erwähnendes Projekt, welches sich auch mit Sentiment Analysen beschäftigt, ist das SPARTA-Projekt (SPARTA, 2021). Es wurde innerhalb einer Zusammenarbeit der Universität der Bundeswehr München und der IABG mbH, einer in München ansässigen Firma, die sich mit Analysen und Testdienstleistungen beschäftigt,



gegründet. Ergebnis des Projektes ist eine Darstellung der Twitter-Aktivitäten der Bundestagskandidaten, die dabei jedoch Kommunalpolitiker außen vorlässt.

Das Ziel meiner Bachelorarbeit ist es, eine grundlegende Übersicht über die vergangene politische Landschaft und deren Veränderung ab der Wiedervereinigung darzustellen. Außerdem möchte ich die aktuelle politische Landschaft aufzeigen, mögliche regionale Einflussfaktoren näher beleuchten und auf die immer größer werdende Rolle von Social-Media eingehen. Anhand von aktuellen Twitter-Daten von Politikern Deutschlands soll eine möglicherweise unterschiedliche Nutzung dieses Mediums aufgezeigt werden.

Die vorliegende Arbeit ist in drei grundlegende Bereiche unterteilt. Zu Beginn wird auf die vergangenen Ergebnisse der Bundestagswahlen in Ost- und Westdeutschland eingegangen. Ich analysiere dazu die Daten der entsprechenden Wahlergebnisse und zeige mögliche regionale Unterschiede auf.

Der zweite Bereich wird sich mit dem Thema Digitalisierung, Social-Media und theoretischen Hintergründen zu Textklassifizierungsverfahren, wie der Sentiment Analyse, beschäftigen.

Der dritte Bereich thematisiert die von mir erhobenen Daten auf Bundestags- sowie Kommunal-/ und Landesebene und erläutert diese näher. Es werden die derzeitigen technischen Limitationen der Analyse aufgezeigt und mögliche Rückschlüsse, bezogen auf die Ergebnisse, ermittelt. Ich werde ein grundlegendes Fazit formulieren und abschließend einen Ausblick auf die weitere Forschung in diesem Themenbereich geben.

## 2 Hauptteil

### 2.1 Die politische Lage Deutschlands, vergangene und aktuelle Situation

Für die Betrachtung der näheren vergangenen politischen Lage, wird der Zeitraum ab dem Mauerfall und der Wiedervereinigung von Ost- und Westdeutschland am 09.11.1990 betrachtet. Dieser Tag markierte eine große Wende in der politischen Landschaft Deutschlands. Ausgelöst wurde dies durch den Reformkurs Michail Gorbatschows in Moskau, welcher in Massenprotesten und Ausreisebewegungen in Deutschland mündete. Die kommunistisch geprägte und von der SED-geleitete DDR (Deutsche Demokratische Republik) wurde durch neue Einflüsse der Rechtsicherheit, der Marktwirtschaft und dem Pluralismus (Gesellschaft in der freie Individuen unterschiedliche Meinungen verfolgen dürfen und diese respektiert werden) in Folge stark verändert. Die Integration der damals 16 Millionen Menschen aus der wirtschaftlich schwachen DDR in die westliche BRD (Bundesrepublik Deutschland), stellte dabei eine große zu bewältigende Aufgabe der damaligen Zeit dar (Tom Thieme, 2016, S. 7). Dabei wurde die wirtschaftliche Ungleichheit durch die Treuhand-Privatisierung, infolgedessen 85% der ostdeutschen Betriebe in westdeutsches Eigentum umgewandelt wurden und 10% zu ausländischen Besitzern wechselten, weiter vorangetrieben. Laut einer Untersuchung der Eigentümerstruktur wurde des Weiteren festgestellt, dass Mitte der 1990er Jahre über 70% der Immobilien in westdeutschen Besitz waren. Durch die Privatisierung staatlicher ostdeutscher Immobilien und die schwache Zahlungsfähigkeit der ostdeutschen Bürger, wurde die Eigentümerstruktur weiter westlich dominiert (Tom Thieme, 2016, S. 87). Auffallend war 1991 der hohe Anteil an Gewerkschaftsmitgliedern, dieser lag bei 35%. Im Vergleich dazu gehörten in Westdeutschland nur 15% einer Gewerkschaft an. Durch die Deindustrialisierung der ostdeutschen Wirtschaft in den folgenden Jahren kam es jedoch zu vielzähligen Austritten aus den Gewerkschaften. Bis zur Jahrtausendwende ging die Zahl bis auf 12% zurück und lag damit unter der Anzahl in Westdeutschland. Ein weiteres Merkmal war die hohe Wahlbeteiligung, dies zeigt z.B. die Volkskammerwahl im Jahr 1990 bei der 93,4% der Wahlberechtigten eine Stimme abgaben. Die Ostdeutschen besitzen dabei eher ein linksorientiertes, soziales Demokratieverständnis, im Gegensatz zum bürgerlich-liberalen Demokratieverständnis in Westdeutschland (Tom Thieme, 2016, S. 93).

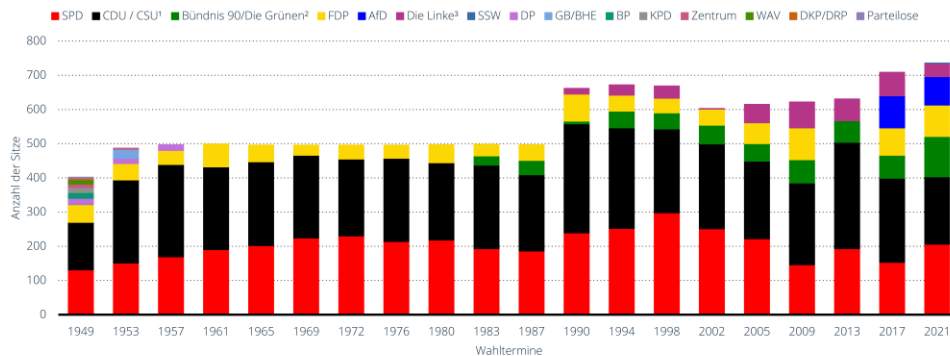
Bis in die heutige Zeit gibt es einen großen Dissens darüber, wie sehr sich die beiden Hälften des Landes angeglichen haben (Tom Thieme, 2016, S. 7). Nach Schätzungen des Deutschen Instituts für Wirtschaftsforschung (DIW) liegt der Medianwert für das Jahr 2012, wenn die Bevölkerung von Ost- und Westdeutschland in gleich große Teile

aufgeteilt wird, beim Nettovermögen der Ostdeutschen bei 8100 Euro und in Westdeutschland bei 21.200 Euro. Dabei besitzen ca. 39,9% der Ostdeutschen einen eigenen Wohnsitz, in Westdeutschland sind es 52,4%. Der Anteil der Ostdeutschen unter den reichsten 30% der Bundesbevölkerung beträgt nur 3 % bei einem Bevölkerungsanteil von 19% (Tom Thieme, 2016, S. 86). Auch der Bruttostundenlohn lag im Jahr 2012 in Ostdeutschland mit durchschnittlich 14,99 Euro deutlich unter den durchschnittlichen 20.10 Euro in Westdeutschland (Tom Thieme, 2016, S. 88). Durch weitere Rückschläge für die ostdeutsche, wirtschaftliche Lage, wie die Rentenreform in der Sozialpolitik von 2001, die das Rentenniveau senkte und somit besonders weniger gutverdienende Bürger in Ostdeutschland traf, sank ab 2004 die politische Identifikation zur Bundesrepublik (Tom Thieme, 2016, S. 95). Dies alles hat zur Folge, dass die politische Partizipation in Ostdeutschland, z.B. Parteienmitgliedschaft und Wahlbeteiligung geringer ausfällt (Rudzio, 2015, S.526 f.) und das Vertrauen in politische Institutionen schwächer ausgeprägt ist als im Rest von Deutschland (Rainer et al. 2018, S 109 f.). Analysen belegen des Weiteren, dass der Anteil von in Ostdeutschland lebenden Personen mit rechtspopulistischer und rechtsextremer Gesinnung höher ist als im Rest der Bundesrepublik (Decker et al. 2018, S 71 f.; Zick et al. 2016, S. 129 f.). Die wirtschaftlich negativen Transformationsprozesse nach der Wiedervereinigung haben, laut einschlägigen Publikationen, zur Wahrnehmung einer kollektiven Diskriminierung geführt (Best, 2016, S. 120), welche das Wahlverhalten und die politische Einstellung weiter beeinflusst hat (Becker & Naumann, 2020, S. 23).

Betrachtet man den Verlauf der Anzahl der Sitze im Bundestag, welcher in Abbildung 1 zu sehen ist, so erkennt man, dass ein Großteil der Wählerstimmen auf die CDU und SPD entfallen. Außerdem ist zu erkennen, dass die SPD nur im Jahr 1998 und 2021 mehr Abgeordnetensitze im Bundestag, als die CDU aufweisen konnte. Die restlichen Sitze teilten sich dabei in den Jahren 1990 bis 2013 die Linke, die FDP und Bündnis 90/Die Grünen. Seit 2017 ist auch die AfD (Alternative für Deutschland) im Bundestag vertreten (Bundeswahlleiter, zitiert nach Statista 2022).

## Anzahl der Sitze der Parteien im Deutschen Bundestag von 1949 bis 2021

Anzahl der Sitze der Parteien im Deutschen Bundestag bis 2021



**Hinweise:** Deutschland  
Weitere Angaben zu dieser Statistik, sowie Erläuterungen zu Fußnoten, sind auf [Seite 121](#) zu finden.  
**Quelle(n):** Bundeswahlleiter; ID:1222386

Abgeordneter Sitze **statista**

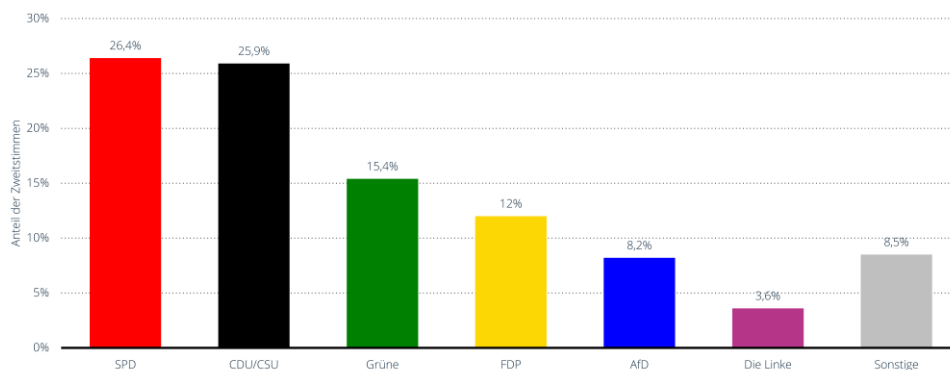
Abbildung 1: Anzahl der Sitze der Parteien im Deutschen Bundestag von 1949 bis 2021 (Bundeswahlleiter, zitiert nach Statista 2021)

Nachfolgend werden die aktuellen Wahlergebnisse der Zweitstimmen der Bundestagswahl am 26. September in Ost- und Westdeutschland kurz erläutert.

Abbildung 2 zeigt dabei, dass die stärkste Partei, nach Zweitstimmen gemessen, die SPD mit 26,4 % ist. Knapp dahinter befindet sich die CDU/CSU mit 25,9 %, auf dem dritten Platz Bündnis 90/Die Grünen mit 15,4%. Weitere Stimmen entfallen mit 12% auf die FDP, 8,2% entfallen auf die AfD, 3,6% auf Die Linke und 8,5% Sonstige Parteien (Infratest dimap, zitiert nach Statista 2022).

## Ergebnis der Bundestagswahl in Westdeutschland am 26. September 2021 (Zweitstimmen)

Ergebnis der Bundestagswahl in Westdeutschland 2021



**Hinweise:** Deutschland; 26. September 2021  
Weitere Angaben zu dieser Statistik, sowie Erläuterungen zu Fußnoten, sind auf [Seite 27](#) zu finden.  
**Quelle(n):** Infratest dimap; ID:1266643

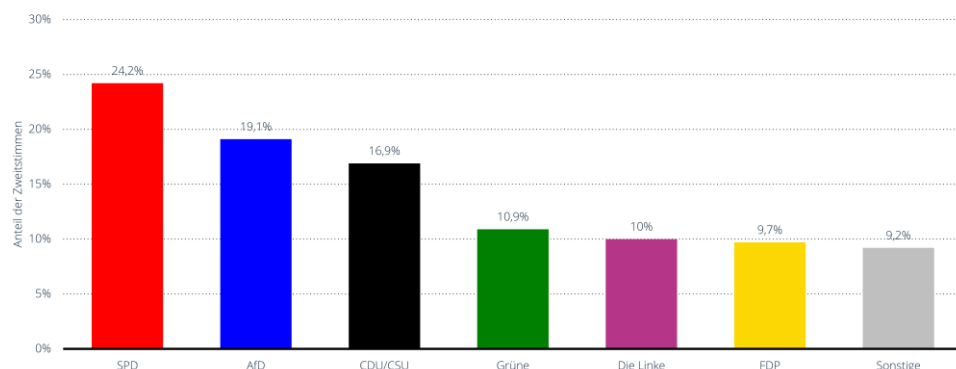
Wahlergebnisse **statista**

Abbildung 2: Ergebnis der Bundestagswahl in Westdeutschland am 26. September 2021 (Zweitstimmen), (Infratest dimap, zitiert nach Statista 2021)

Vergleicht man diese Ergebnisse mit den Bundesländern in Ostdeutschland (Abbildung 3), so zeichnet sich eine andere Verteilung ab. Auch in Ostdeutschland ist die stärkste Partei, nach Zweitstimmen gemessen, die SPD mit 24,2%. Dies stellt einen leichten Rückgang im Vergleich zu Westdeutschland dar. Wird nun jedoch die zweitstärkste Partei betrachtet, so erkennt man deutliche Unterschiede. Die AfD hat, mit 19,1% aller Zweitstimmen, noch vor der CDU/CSU mit 16,9%, den zweiten Platz für sich beanspruchen können. Damit hat die AfD in Ostdeutschland ein Plus von 10,9% gegenüber Westdeutschland verzeichnen können. Die CDU/CSU hingegen musste im direkten Vergleich ein Minus von 9% hinnehmen. Darauf folgt Bündnis 90/Die Grünen mit 10,9% (-4,5% im Vgl. zu Westdeutschland), Die Linken mit 10% (+6,4 im Vgl.), die FDP mit 9,7% (-2,3% im Vgl.) und sonstigen Parteien mit 9,2% (+0,7%). (Infratest dimap, zitiert nach Statista 2022).

#### Ergebnis der Bundestagswahl in Ostdeutschland am 26. September 2021 (Zweitstimmen)

Ergebnis der Bundestagswahl in Ostdeutschland 2021



Hinweis(e): Deutschland; 26. September 2021

Weitere Angaben zu dieser Statistik, sowie Erläuterungen zu Fußnoten, sind auf Seite 78 zu finden.

Quelle(n): Infratest dimap; ID 1266647

Wahlergebnisse **statista**

Abbildung 3: Ergebnis der Bundestagswahl in Ostdeutschland am 26. September 2021 (Zweitstimmen), (Infratest dimap, zitiert nach Statista 2021)

Es zeigt sich somit, dass es bis heute eine deutlich feststellbare Differenz im Wählerverhalten in Ost- und Westdeutschland gibt.

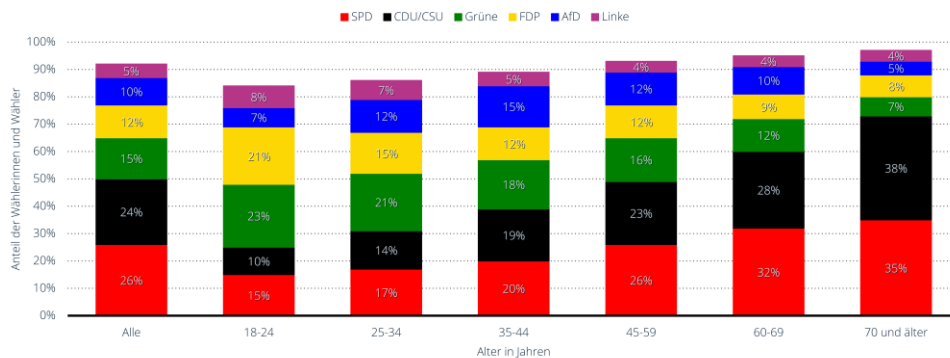
Blickt man auf das Wählerverhalten nach Altersgruppen in Abbildung 4, so kann eine weitere Wählerdynamik festgestellt werden. In den Altersgruppen von 18 – 24 Jahren ist eine deutliche Abnahme der alteingesessenen Parteien, wie der CDU/CSU und der SPD, zu erkennen. Diese können nur 10% respektive 15% der Wählerstimmen, für sich beanspruchen. Deutlich mehr Zuspruch bekommt Bündnis 90/Die Grünen (23%) und die FDP (21%). Die AfD liegt mit 7% leicht unter ihrem Endergebnis von ca. 10%. Die Linke liegt bei 8%. Dieser Trend ist besonders bei der jüngeren Wählerschaft zu beobachten und

schwächt sich mit ansteigendem Alter immer weiter ab. Die AfD kann, im Gegensatz zu Bündnis 90/Die Grünen und der FDP, jedoch bis einschließlich der Altersgruppe der 45-59-jährigen, ihre Stimmenanzahl, im Vergleich zu den 18-24-jährigen weiter ausbauen. Sie liegt zwischen 12 und 15%.

Betrachtet man die Wählerstimmen bei der ältesten Altersgruppe (70+), so dominieren die Parteien SPD mit 35% und CDU/CSU mit 38%. Bündnis 90/Die Grünen kommt nur auf 7%, Die FDP auf 8%, die AfD liegt bei 5% und die Linke bei 4% (Infratest dimap, zitiert nach Statista 2022).

#### Wahlverhalten bei der Bundestagswahl am 26. September 2021 nach Alter (Stimmenanteile der Parteien)

Wahlverhalten bei der Bundestagswahl 2021 nach Alter



**15** Hinweis(e): Deutschland; 26. September 2021; ab 18 Jahre; Wählerinnen und Wähler in Deutschland  
Weitere Angaben zu dieser Statistik, sowie Erläuterungen zu Fußnoten, sind auf Seite 85 zu finden.  
Quelle(n): Infratest dimap; ID: 1252992

Wahlverhalten **statista**

Abbildung 4: Wahlverhalten bei der Bundestagswahl am 26. September 2021 nach Alter (Stimmenanteile der Parteien), (Infratest dimap, zitiert nach Statista 2021)

## 2.2 Thema Digitalisierung und Social-Media

Die Digitalisierung schreitet unaufhörlich voran, sei es im beruflichen Umfeld, in der Kommunikation im privaten Raum oder in der Industrie. Waren in Deutschland in den 1990er Jahren Computer und die dazugehörigen Netzwerke noch einigen, wenigen Technikern und IT-Spezialisten vorenthalten, ist die Nutzung dieser für uns mittlerweile zum Alltag geworden (Kollmann & Schmidt, 2016, S. 4). In jedem Umfeld ist es kaum noch möglich, ohne die Digitalisierung auszukommen. Die berufliche Kommunikation läuft überwiegend über E-Mails und seit der Coronapandemie werden verstärkt Online-Meetings durchgeführt. In der Industrie werden tagtäglich riesige Datenmengen innerhalb und zwischen Firmen ausgetauscht. Auch im Privatleben wird immer mehr über Onlinedienste kommuniziert. Die Digitalisierung hat zu einem grundlegenden Strukturwandel im gesellschaftlichen und wirtschaftlichen Bereich weltweit geführt.

Jedes Jahr nutzen mehr Personen das World Wide Web. Im Zeitraum vom Januar 2020 bis Januar 2021 sind 316 Millionen neue Nutzer (+7,3%) weltweit hinzugekommen (Digital 2021: Germany, 2021, S. 8). Deutschland hat im Vergleich dazu im gleichen Zeitraum 1 Millionen (+1,3%) neue Nutzer dazu gewonnen (Digital 2021: Germany, 2021, S. 18). Die Gesamtanzahl der Internetnutzer in Deutschland liegt dabei bei 78,81 Millionen (94% der Bürger Deutschlands) (Digital 2021: Germany, 2021, S. 25).

Da Social-Media in dieser Arbeit eine grundlegende Rolle für die späteren Auswertungen einnimmt, erläutere ich diesen Begriff näher.

Social-Media wird allgemein unter dem Oberbegriff Web 2.0, oder auch Social Web zusammengefasst und beschreibt eine verbesserte Version des World Wide Web, in dem der Konsument zum Produzenten von Inhalten werden kann. Die Begriffsbezeichnung wurde dabei Mitte der 2000er-Jahre vom amerikanischen Verleger Tim O'Reilly weiterverbreitet (Schmidt, 2018, S. 16).

Social-Media als Begriffsbeschreibung steht grundlegend für alle Anwendungen, in denen ein Nutzer Daten, wie z.B. eigene Texte verfassen, editieren und veröffentlichen kann (Kaplan & Haenlein, 2010). Im täglichen Sprachgebrauch werden mit Social-Media im privaten Kontext Onlinekollaborationsplattformen, wie z.B. Facebook, Twitter oder Instagram und im Geschäftsumfeld Plattformen, wie z.B. XING oder LinkedIn, bezeichnet. Je nach Anwendungsfall und Plattform, können so Meinungen von Einzelpersonen miteinander ausgetauscht und ein offener Diskurs über die Themen des täglichen Lebens geführt werden. Auch ist es möglich, sich im Geschäftsumfeld mit anderen Arbeitnehmer/innen und Arbeitgeber/innen zu vernetzen und so die eigene digitale Reichweite zu erhöhen.

Wichtig beim Umgang mit Social-Media und der Berichterstattung über unterschiedlichste Plattformen hinweg, ist jedoch für den Nutzer den gewählten Kontext von Nachrichten zu erfassen und sich darüber im Klaren zu sein, dass es sich bei geteilten Beiträgen nicht immer um die Meinung einer breiten Masse handelt, sondern möglicherweise um Einzelpersonen, die mit ihren eventuell polarisierenden Beiträgen in die Öffentlichkeit gerückt werden können (Schweidel & Moe, 2014). Auch ist es möglich, dass sich Unwahrheiten sehr schnell über eine breite Masse von Nutzern hinweg verbreiten, da jeder Nutzer Beiträge verfassen kann, diese aber schwierig auf ihren Inhalt hin zu überprüfen sind.

Parallel zur eingangs erwähnten steigenden Internetnutzung, erhöht sich damit auch die Social-Media-Nutzung in Deutschland. Ca. 66 Millionen (78,7%) der Einwohner Deutschlands nutzen Social-Media in Ihrem Alltag (Datareportal, 2021, S. 17). Im vergangenen Jahr, von Januar 2020 bis Januar 2021, sind dabei 7,7 Millionen (+13,2%) Menschen dazugekommen (Datareportal, 2021, S. 18). Dabei befindet sich der Großteil der Nutzer von 82,2 % im Altersbereich von 18 – 54 Jahren (Datareportal, 2021, S. 45) und verbringt im Durchschnitt 1 Stunde und 24 Minuten am Tag mit Social-Media (Datareportal, 2021, S. 22). 24,8% der Nutzer benötigen Social-Media dabei auch für berufsbegleitende Tätigkeiten (Datareportal, 2021, S. 46). Social-Media hat sich, wie das World Wide Web, fest in den Haushalten Deutschlands etabliert.

Die klassischen Massenmedien, wie z.B. das Fernsehen, sind aktuell als zentraler Ort für die Informationsgewinnung noch am weitesten verbreitet, wie in Abbildung 5 zu sehen ist. Die tägliche Nutzungsdauer liegt dabei bei über 200 Minuten. Erkennbar ist allerdings, dass die Internetnutzung (in Minuten) über die letzten 7 Jahre deutlich zugenommen hat (Seven.One Media GmbH, zitiert nach Statista 2022).

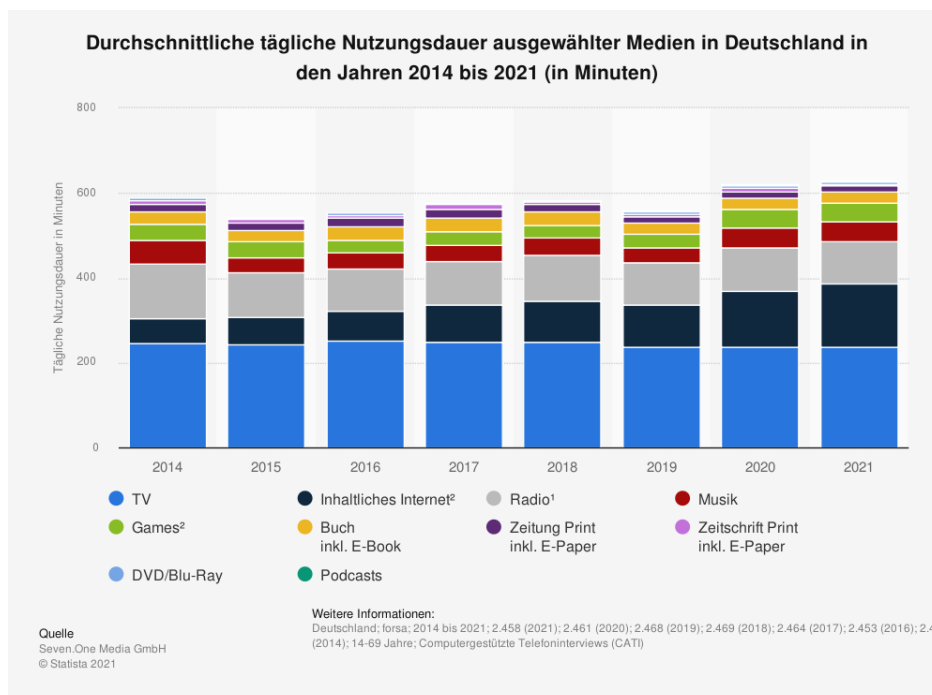


Abbildung 5: Durchschnittliche tägliche Nutzungsdauer ausgewählter Medien in Deutschland in den Jahren 2014 bis 2021 (in Minuten), (Seven.One Media GmbH, zitiert nach Statista 2021)



Die klassischen Medien, wie das Fernsehen, verlieren in den jüngeren Generationen (12-19 Jahre) immer weiter an Reichweite. Dabei verschiebt sich der Konsum von klassischen Massenmedien (Fernsehen, Zeitung) hin zu rein digitalen. Wie in Abbildung 6 erkennbar ist, nutzen im Jahr 2021 92% der Jugendlichen (12 – 19 Jahre) täglich ihr Smartphone und 88% das Internet. Dies ist für Kurznachrichtendienste, wie z.B. Twitter, vorteilhaft. Nur 48% nutzen das Fernsehen täglich. Betrachtet man die Mediennutzung in der Woche, so lesen nur 8% aller Befragten gedruckte Tageszeitungen (Siehe Abbildung 6).

Mediennutzung durch Jugendliche in der Freizeit in Deutschland 2021

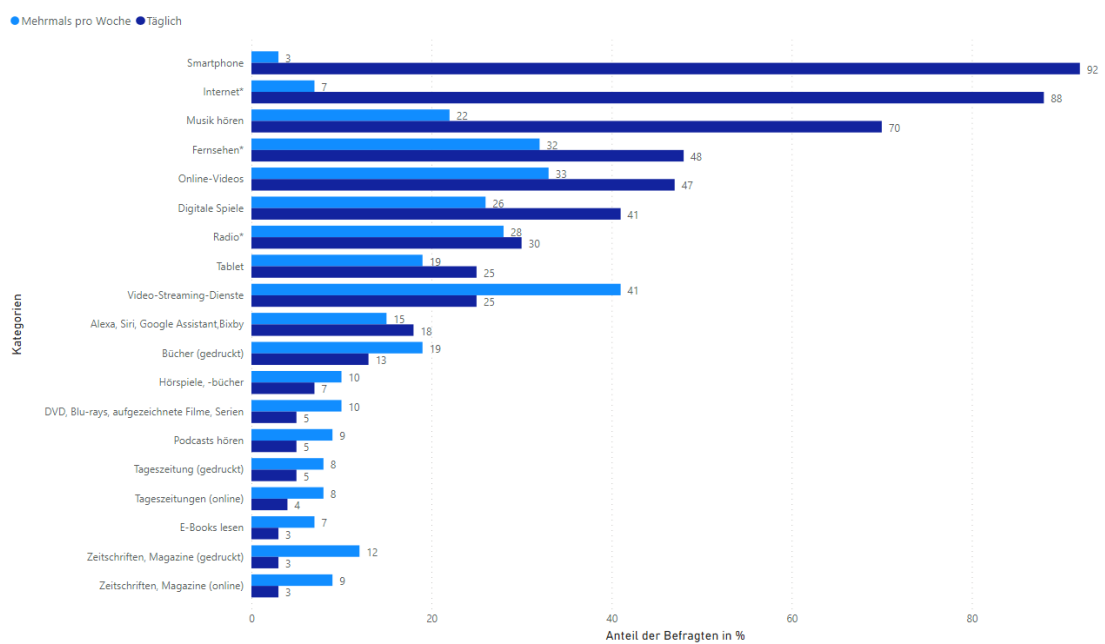


Abbildung 6: Mediennutzung durch Jugendliche in der Freizeit in Deutschland 2021, eigene Abbildung, basierend auf Medienpädagogischer Forschungsverbund Südwest 2021

## 2.3 Twitter

Für die in dieser Arbeit durchgeführten Analysen wird die Social-Media-Plattform Twitter genutzt. Nachfolgend wird daher das Thema Twitter in seiner Funktionsweise und Historie näher erläutert.

Twitter, damals noch „Twtr“ genannt, wurde als Social-Networking-Service am 8. März 2006 in Kalifornien, USA, von Jack Dorsey, Evan Williams, Biz Stone und Noah Glass gegründet. Ursprungsidee war es, auf der Technik der SMS-basierend (Short Message Service), innerhalb von 140 Zeichen kurze Nachrichten, so genannte „Tweets“ zu verschicken, wobei 160 Zeichen das damalige Limit für SMS-Nachrichten waren und die restlichen 20 Zeichen für einen Nutzernamen reserviert sein sollten (First Versions, 2015). Seit November 2017 sind bis zu 280 Zeichen verfügbar (Heise, 2017).

Twtr wurde im Oktober desselben Jahres vom neu gegründeten Mutterkonzern Obvious akquiriert. Im April des nächsten Jahres 2007 wurde Twitter Inc. gegründet und der Social-Media-Dienst von Obvious ausgegliedert (First Versions, 2015). In den darauffolgenden Jahren verzeichnete Twitter einen großen Zuwachs, so dass es weltweit bis zum ersten Quartal 2019 ca. 330 Millionen aktive Nutzer besaß (Twitter, 2019, zitiert nach Statista).

Twitter erreicht dabei potenziell 5,8 Millionen Menschen in Deutschland. Vergleicht man die letzten beiden Quartale sind ca. 350 Tausend Menschen und damit +6,4% hinzugekommen (Kemp, 2021).

Was Twitter und ähnliche Socia-Media-Plattformen, wie z.B. Facebook, jedoch von anderen abgrenzt, ist die zentrale Struktur, wie mit den Inhalten interagiert werden kann, die auf den jeweiligen Plattformen existieren. Gab es bei Facebook die Möglichkeit „Likes“ zu vergeben, so konnte man in Twitter „Follower“ erhalten. Später wurde die Einheit des „Retweets“ und des „Favs“ (ab November 2015 „Like“ genannt) hinzugefügt, um zusätzliche Funktionalitäten in der Interaktion zu bieten (Paßmann, 2018).

Retweets bieten die Möglichkeit, den Tweet eines anderen Bloggers mit den eigenen Followern zu teilen. Dabei wird der ursprüngliche Urheber des Tweets jedoch immer über eine eingebaute Retweet-Funktion mit angegeben (Wayback Machine, 2013).

Favs oder auch Likes (auf Deutsch auch bekannt als „Gefällt mir“), stellen eine Funktionalität dar, mit der Tweets mit einem Herzsymbol versehen werden können. Hauptgrund zum Liken eines Tweets ist es zu zeigen, dass man den Tweet gut findet, man diesem zustimmt oder auch unterstützt (Twitter, 2021).

## 2.4 Data Mining

Data Mining beschreibt die Wissenschaft, meist strukturierte Daten aus unterschiedlichen Datenquellen, wie z.B. relationalen Datenbanken, Datenbanken die Daten in Form von Tabellen abspeichern und relational miteinander verbunden sind (Steiner, 2021, S. 10), zu analysieren und auszuwerten (Cleve und Lämmel, 2020, S.38-39). Data Mining ist eine Sammlung von Techniken, Methoden und Algorithmen für die Analyse von Daten, die die Grundlage von komplexeren Ansätzen, wie z.B. der Business Intelligence oder dem Thema Big Data darstellt (Cleve und Lämmel, 2020, S.3).

Der Ablauf eines Data-Mining-Prozesses kann dabei in folgende 5 Schritte (Siehe Abbildung 7) unterteilt werden:

1. Selektion: Die zu analysierenden Daten werden ausgewählt und in einer Datenbank abgespeichert.

2. Datenvorverarbeitung: In dieser Phase wird die Datenbereinigung durchgeführt, fehlende oder fehlerhafte Werte müssen so weit korrigiert werden, dass die Daten korrekt weiterverarbeitet werden können.
3. Datentransformation: Oft müssen die zugrundeliegenden Daten vor dem Verfahren der Analyse in das richtige Datenformat gebracht werden. Dies ist besonders bei internationalen Datensätzen wichtig, da z.B. Datumsangaben in unterschiedlichen Ländern anders gehandhabt werden.
4. Data Mining: In dieser Phase läuft das eigentliche Data Mining ab, es wird ein Modell entwickelt, welches die Daten anschließend analysiert und in einer geeigneten Form wiedergibt.
5. Evaluation und Interpretation: Die ausgewerteten und analysierten Daten müssen geprüft und evaluiert werden, um festzustellen ob die gewünschten Ergebnisse und Schlussfolgerungen von den Daten abzuleiten sind.

(Cleve und Lämmel, 2020, S. 5)

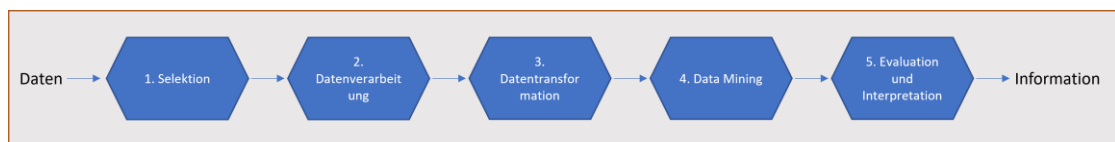


Abbildung 7: Data-Mining-Prozesse, eigene Abbildung, basierend auf Cleve und Lämmel, 2020, S. 5

## 2.5 Text Mining

Das Themenfeld des Text- und Data Mining wird oft als gleichwertig angesehen. Hauptunterschied zwischen den beiden Analyseformen ist jedoch, dass beim Text Mining, anders als beim Data Mining, meist unstrukturierte Textdokumente ausgewertet werden. Wobei Textdokumente an sich jedoch eine implizite Struktur in Form einer Grammatik beinhalten (Hippner und Rentzmann, 2006, S 287).

Text Mining ist schlussfolgernd ein Teilgebiet des Data Minings. Anders als beim Data Mining, müssen jedoch aufgrund der erhöhten Komplexität der Analyse von meist unstrukturierten Daten, diese vor der eigentlichen Datenanalyse und Auswertung in eine maschinell auswertbare Form gebracht werden. Das heißt speziell, dass eine zusätzliche linguistische Datenaufbereitung von Nöten ist, um fehlende Datenstrukturen zu rekonstruieren (Rajman und Veseley, 2004, S. 7). Die einzelnen Schritte des Text Minings (Siehe Abbildung 8)

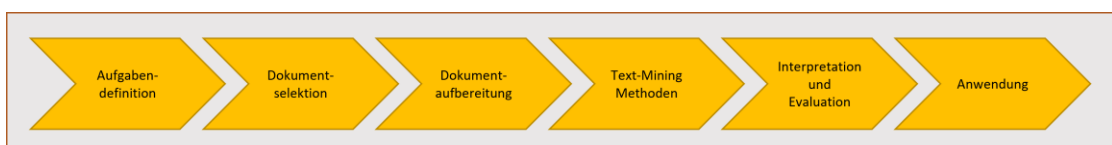


Abbildung 8: Text-Mining-Prozess, eigene Abbildung, basierend auf Hippner und Rentzmann, 2006, S. 288

ähneln teilweise dem des klassischen Data-Minings und sind wie folgt:

1. Aufgabendefinition: Die Problemstellung und das Ziel der Text-Mining-Analyse werden festgelegt.
2. Dokumentselektion: Im Hinblick auf das Ziel der Analyse werden möglicherweise relevante Artikel und Dokumente herausgefiltert und zusammengeführt (Sirmakessis, 2004, S 10 ff.).
3. Dokumentaufbereitung: Anders als beim Data-Mining müssen die ausgewählten Daten noch in für die Analyse abgestimmte Struktur gebracht werden. Dabei wird eine Merkmalsextraktion der vorhandenen Daten durchgeführt, bei der Terme aus den vorliegenden Texten extrahiert werden und als Repräsentation der einzelnen Dokumente dienen. Wichtig dabei ist, dass ein Term nur aus einem Wort, auch zusammengesetzten Wörtern, die eine eindeutige Bedeutung haben, oder einem Wortstamm bestehen darf (Hippner und Rentzmann, 2006, S. 288). Um diese Terme extrahieren zu können, gibt es unterschiedliche Techniken des Natural Language Processing:
  - Die morphologische Analyse: Bei der morphologischen Analyse wird der Unterschied sinntragender Wortbestandteile und Wortformen thematisiert. Dies soll dazu führen, dass die Komplexität der einzelnen Text-Mining-Analysen abnimmt (Sullivan, 2001, S. 32). Durch unterschiedliche Flexionsformen eines Wortes (Änderung eines Wortes im Hinblick auf grammatikalische Merkmale, die eine neue Wortform desselben Wortes kreiert), Beispiel dafür ist *lieft*, *laufen*, *läufst*, wird es möglich sie als zusammengehörig zu erkennen (Hippner und Rentzmann, 2006, S. 288).
    - Stemming: Das Stemming oder auch Stammformreduktion genannt, beschreibt ein Verfahren, bei dem verschiedene morphologische Varianten von Wörtern auf ihren Textstamm zurückgeführt werden (z.B. *gesuchte* - *suchen*) (Ferber, 2003, 41).
  - Die syntaktische Analyse: Die syntaktische Analyse fokussiert sich auf die Annotation einzelner Satzbausteine (Ule & Hinrichs, 2004, S.221). Als Annotation wird dabei eine Textauszeichnung genannt. Grundlage dafür ist eine Markierung von einzelnen Textabschnitten (Lobin, 2004, S. 51).
    - Kernstück der syntaktischen Analyse ist dabei das Part-of-Speech-(POS)Tagging, bei diesem werden alle Wörter in einem Satz auf ihre Wortarten annotiert (Adjektiv, Verb etc.). Um dies zu erreichen, kann beim POS-Tagging auf unterschiedliche Informationsquellen, wie z.B. Lexika

oder syntaktische Informationen zurückgegriffen werden (Manning & Schütze, 2003, S. 344).

- Das Lexika liefert Wörter und Wortarten, die beim POS-Tagging angenommen werden können.
- Die syntaktischen Informationen geben dabei Sequenzen von Wortarten an, durch die die wahrscheinlichste Wortart für das POS-Tagging ermittelt werden kann (Hippner und Rentzmann, 2006, S. 288).
- Aufbauend auf dem POS-Tagging kann anschließend eine zusätzliche Annotation mittels des Parsing erfolgen (Damascelli & Martelli, 2003, S.20). Das Parsing analysiert den vorhandenen Satzbau dabei nach den Wortstellungen im Satz und etikettiert diese nach ihrer Wortart. Dies kann vollautomatisch durch Annotationsmethoden geschehen. Die besten Verfahren wurden dabei für die englische Sprache entwickelt und erreichen eine Güte von 90% (Ule & Hinrichs, 2004, S. 235).
- Für die deutsche Sprache hat Wauschkuhn, mittels eines Werkzeugs zur partiellen syntaktischen Analyse, eine Güte von 85,7% erreicht (Wauschkuhn, 1996, S. 357).
- Das Parsing kann in zwei verschiedenen Formen eingeteilt werden: das partielle Parsing und das vollständige Parsing:
  - Das vollständige Parsing bezieht sich dabei auf jedes Wort und die Bestimmung der jeweiligen Wortform in einem Satz.
  - Beim partiellen Parsing ist es möglich, einzelne Wortformen unverbunden zu lassen und flache Analysen, wie z.B. Shallow Parsing, durchzuführen. Bei diesen wird sich auf die Bestimmung von Abfolgen von Wörtern konzentriert.

(Ule & Hinrichs, 2004, S. 236 ff.).

- Der große Vorteil der syntaktischen Analyse liegt darin, dass gezielt Daten aus syntaktischen Einheiten extrahiert werden können (Ule & Hinrichs, 2004, S. 236 ff.).

- Semantische Analyse: diese ist dazu da, um Kontexte in Texten zu erkennen und in bedeutungsunabhängige Einheiten zu zerlegen. So kann überprüft werden, in welchem Kontext ein Wort genutzt wird (Hippner und Retzmann, 2006, S. 289).

## 2.6 Opinion Mining

Aufbauend auf der vorangegangenen Erläuterung der Funktionsweise des Data- und Textminings, wird nun das Themenfeld des Opinion Minings näher erläutert. Opinion Mining wird teilweise auch als Synonym für die Sentiment Analyse verwendet, unterscheidet sich jedoch darin, dass beim Opinion Mining eine Meinung zu einem Thema zum Ausdruck gebracht wird und bei der Sentiment Analyse das vermittelte Gefühl des jeweiligen Autors im Vordergrund steht (Petz, 2018, S. 15 – 16). Aufgrund dieser engen sinngemäßen Verwandtschaft beider Begriffe, wird das Opinion Mining im Vorfeld zur Sentiment Analyse nun näher erläutert.

Das Opinion Mining kann als ein Themenfeld des Information Retrieval (Informationsbeschaffung), des Text Mining, des Web-Crawling, linguistischer Algorithmen und semantischer Technologien angesehen werden (Liu, 2012, S. 8). Das Ziel des Opinion Mining ist es, die Extraktion von Meinungen von Nutzern in unstrukturierten Texten zu automatisieren. Um dies zu erreichen, umfasst das Opinion Mining mehrere grundlegende Aufgaben, in denen auch die Sentiment Analyse als Teilaufgabe vorkommt:

- Sentiment Analyse: Der Zweck der Sentiment Analyse ist die Erkennung von Sentiments und der Meinungsuntersuchung, welche zum Forschungsgebiet des Text Mining gehört (Hemmatian und Sohrabi, 2017, S. 1498).
- Meinungsextraktion: Der Vorgang der Extraktion von Nutzermeinungen aus dem Internet wird Meinungsextraktion genannt. Der Hauptzweck der Meinungsextraktion ist es, herauszufinden, wie die Nutzer im Internet denken (Hemmatian und Sohrabi, 2017, S. 1498).
- Sentiment Mining: Das Sentiment Mining hat zwei Hauptziele. Das erste Ziel ist es zu bestimmen, ob der vorliegende Text objektive oder subjektive Sätze beinhaltet. Ein Satz wird dabei objektiv oder sachlich genannt, wenn dieser Fakten über ein Produkt enthält. Ein subjektiver Satz steht dabei für die individuellen Emotionen, die einem Produkt entgegengebracht werden. Das zweite Ziel ist es, die extrahierten Meinungen zu klassifizieren und diese in drei Kategorien einzuteilen: positiv, negativ und neutral (Farra et al. 2010).
- Subjektanalyse: Die Subjektanalyse schafft die Möglichkeit, subjektive Sätze zu identifizieren, zu klassifizieren und zu sammeln.
- Emotionale Analyse: Viele Wörter in einem Text sind entweder emotional positiv oder negativ behaftet. Die Emotionale Analyse spezifiziert, welche Aspekte im

Text Emotionen ausdrücken, mittels der Natural Language Processing Techniken (Grefenstette et al, 2004).

- Review Mining: Das Review Mining beschreibt eine Unterkategorie der Text Sentiment Analyse und fokussiert sich auf die Extraktion der Sentiments des Autors des zu analysierenden Textes. Das Endergebnis des Review Mining ist eine Zusammenfassung der analysierten Texte (Zhuang et al, 2006).

Neben dem Text Mining, welches sich mit der Aspektextraktion befasst und dem Opinion Mining, welches Meinungen aus Texten zu analysieren versucht, gibt es noch zusätzliche Verfahren, die eingesetzt werden, um Meinungen aus extrahierten Aspekten zu klassifizieren. Die nachfolgenden vorgestellten Methoden können dabei zwei Bereichen zugeordnet werden: den machine learning und lexikonbasierten Bereichen (Hemmatian & Sohrabi S. 1506).

## 2.7 Machine learning Methoden

Das Oberthema der Machine Learning Methoden bildet dabei nur die oberste Abstraktionsschicht und beinhaltet drei untergeordnete Techniken, die in diesem Bereich angewandt werden können (Siehe Abbildung 9). Die drei Techniken zur Klassifizierung von Meinungen zu Sentiments sind dabei unterteilt in supervised (überwachten), semi-supervised (teils-überwachten) und unsupervised (nicht überwachten). Im folgenden Abschnitt werden diese Techniken näher erläutert. Aufgrund des großen inhaltlichen Umfangs der einzelnen Techniken und den dazugehörigen Methoden werden zwar mehrere genannt, jedoch nur jeweils eine exemplarisch näher erläutert (Hemmatian & Sohrabi, 2017, S. 1507).

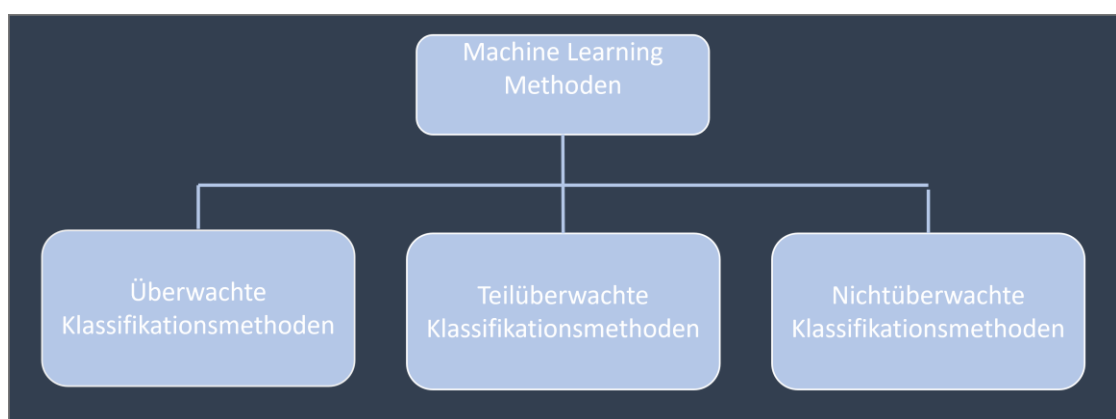


Abbildung 9: Machine Learning Methoden, eigene Abbildung, basierend auf Hemmatian & Sohrabi, 2017, S. 1507

## 2.8 Überwachte Klassifikationsmethoden

Bei den überwachten Lern-/Klassifizierungsmethoden wird ein Trainingsset an Daten verwendet, bei dem im Vorhinein schon spezifiziert wurde, welche Aspekte welchem Gefühl zugeordnet werden müssen. Die Machine Learning-Funktion versucht dabei, ein Muster beim Mappen des Input und Output (Output bedeutet hier dem Gefühl) zu erkennen und zu entwickeln. Auch die Sentiments wurden schon vorher in bestimmte Kategorien eingeteilt und spezifiziert. Die Machine Learning-Funktion/der Algorithmus versucht daraufhin, anhand von beispielhaft zugewiesenen Gefühlen zu unterschiedlichen Sentiments, die restlichen Gefühle korrekt einzuordnen (Hemmatian & Sohrabi S. 1506).

Die überwachten Lernmethoden können in zwei unterschiedliche Kategorien aufgeteilt werden: die probabilistischen Klassifikationen und die nicht-probabilistischen Klassifikationen (Hemmatian & Sohrabi, 2017, S. 1506 -1507).

### 2.8.1 Probabilistische Klassifikationen

Probabilistische Methoden, die auf Wahrscheinlichkeiten beruhen, zur Klassifizierung von Gefühlen zu Sentiments, sind ein beliebter Klassifikationsansatz im Forschungsfeld des maschinellen Lernens. Die Methoden leiten sich von probabilistischen Techniken ab, welche einen systematischen Weg hin zu statistischen Klassifikationen von komplexen Themengebieten, wie der natürlichen Sprache, bieten. Bekannte Beispiele für diese Methoden sind z.B. Naive Bayes, das Bayesian network oder auch die maximum entropy (Hemmatian & Sohrabi S. 1507).

### 2.8.2 Naive Bayes

Naive Bayes ist ein simpler und beliebter Ansatz im Bereich der Textklassifizierung. Der naive Bayes-Ansatz geht davon aus, dass in diesem Beispiel, jedes Wort „naiv“ unabhängig von allen anderen Wörtern in einem Satz vorkommt. Mittels der bedingten Wahrscheinlichkeitsrechnung, bei dem jedes Wort mit seinen Vorkommnissen im jeweiligen Datensatz mit der Anzahl an Wörtern verglichen wird und der Grundannahme einer Wahrscheinlichkeit für die Zugehörigkeit zu einem bestimmten Datensatz, kann so ein Wert für eine Wahrscheinlichkeit zur jeweiligen Klassifizierung bestimmt werden (Webb, 2017).



### 2.8.3 Nicht-probabilistische Klassifikationen

In manchen Fällen kann es vorkommen, dass aufgrund des gegebenen Problems, bzw. des zu klassifizierenden Themas, probabilistische Methoden nicht effektiv und akkurat sein können. In diesen Fällen kann die Klassifizierung stattdessen mit nicht-probabilistischen (ohne bekannte Wahrscheinlichkeit) Klassifikatoren durchgeführt werden. Besonders neurale Netzwerke, Support Vektormaschinen (SVM), nearest neighbor, decision trees und rule-based Methoden werden dabei häufig für Sentiment Analysen genutzt (Hemmatian & Sohrabi S. 1509).

### 2.8.4 Support Vektormaschinen (SVM)

SVM ist eine der beliebtesten überwachten Klassifikationsmethoden. SVM hat eine fundierte theoretische Basis und ist, laut dem Report von (Liu, 2007), wahrscheinlich eine der präzisesten Textklassifikationsmethoden. SVM stellen hierbei eine mathematische Methode dar, Objekte statistisch in unterschiedlichen Klassen einzuteilen. SVM kann für lineare und nichtlineare Klassifizierungen benutzt werden. Bei nichtlinearen Klassifizierungen wird hierbei ein Kernel, z.B. der polynomiale Kernel, genutzt, um Objektbereiche und deren Trennflächen in höhere Dimensionen zu überführen, um eine möglichst korrekte Klassifizierung zu ermöglichen. Die von der SVM gebildeten Klassen besitzen bei diesem Verfahren einen möglichst großen Bereich bis zur jeweiligen Klassengrenze. SVM wird daher auch als Large Margin Classifier bezeichnet. In der traditionellen Textkategorisierung erzielt es bessere Ergebnisse als naive Bayes (Joachims, 1998, S. 137 - 142).

## 2.9 Teilüberwachte Klassifikationsmethoden

Ein Nachteil der klassischen Textklassifikationsmethoden ist, dass diese nur vorher zugeordnete (labeled) Datenpaare zum Training der einzelnen Klassifikatoren verwenden. Da zugeordnete Datensätze jedoch oft schwierig, teuer oder zeitaufwendig zu erstellen sind und da sie Fachwissen von demjenigen abverlangen, der die Zuordnungen tätigt, liegt es nahe für diesen Bereich alternative Methoden in Betracht zu ziehen. Nicht zugeordnete Daten liefern allgemein schlechte Ergebnisse bei der Nutzung in Klassifikatoren. Diesem Problem widmen sich jedoch die teilüberwachten Klassifikationsmethoden. Bei diesen wird versucht, durch eine Kombination von großen Mengen an nichtzugeordneten und kleinen Mengen von zugeordneten Datenpaaren bessere Klassifikatoren zu bilden. Teilüberwachte Klassifikationsmethoden benötigen weniger menschliche Eingriffe und liefern eine höhere Präzision bei der Textklassifikation. Beispiele für unterschiedliche Ansätze von teilüberwachten Klassifikationsmethoden sind selbsttrainierte Ansätze, co-

trainierte, das multi-view learning, graphenbasierte Methoden und generative Ansätze (Hemmatian & Sohrabi, 2017, S. 1512 - 1517).

### 2.9.1 Selbsttrainierte Ansätze

Dieser Ansatz wird als eine der bekanntesten und beliebtesten Methoden, innerhalb der teilüberwachten Klassifikationsmethoden, angesehen und oft genutzt (Zimmermann et al., 2014; Gao et al., 2014; He und Zhou, 2011). Bei diesem Ansatz werden Textklassifikatoren zunächst mit einem kleinen Datensatz von zugeordneten Datenpaaren trainiert. Der trainierte Klassifikator wird im nächsten Schritt mit nicht zugeordneten Testdaten gespeist. Die Datenpaaren, die daraufhin mit der höchsten Zuverlässigkeit zugeordnet wurden, werden im darauffolgenden Schritt mit in den Trainingsdatensatz aufgenommen. Dieser Prozess wird so oft wiederholt, bis eine große Menge an zuverlässig zugeordneten Datenpaaren entstanden ist. Dieser Ansatz nutzt also einen kleinen Satz von zugeordneten Datenpaaren, um daraufhin nicht zugeordnete Daten korrekt zuweisen zu können. Da dieser Ansatz jedoch stark darauf beruht, dass möglichst zuverlässig zugeordnete Datenpaare die Vorhersage im nächsten Durchlauf verbessern, kann mit schlecht zugeordneten Datenpaaren die Textklassifikation auch inkorrekte Ergebnisse liefern (Hemmatian & Sohrabi, 2017, S. 1517).

## 2.10 Nichtüberwachte Klassifikationsmethoden (Clustering)

Bei nichtüberwachten Klassifikationsmethoden ist nur der Inputwert der Trainingsdatensätze bekannt. Die auf Clustering-basierenden Ansätze ermöglichen es, ohne manuelles eingreifen und ohne linguistische Daten oder Trainingszeit, mittelmäßig gute Klassifikationsergebnisse zu liefern (Hemmatian & Sohrabi S. 1519).

Das Ziel beim Clustering ist es, Strukturen in Texten zu finden, die vorher unentdeckt blieben. Das Clustering teilt Daten, aufgrund von strukturellen Gemeinsamkeiten und Unterschieden, in unterschiedliche Gruppen (Cluster) ein. Daraus ergeben sich Gruppen, in denen Daten, aus einer bestimmten Perspektive gesehen, alle Ähnlichkeiten miteinander besitzen und maximal unterschiedlich zu anderen Gruppen von Daten sind (Hemmatian & Sohrabi S. 1519).

Das nichtüberwachte Lernen des Clusterings wird in zwei grundlegende Kategorien eingeteilt: das hierarchische Clustering und das partitionelle Clustering (Hemmatian & Sohrabi S. 1519).

### 2.10.1 Hierarchisches Clustering

Das hierarchische Clustering besteht aus einem Set von verschachtelten Clustern, die als ein Baum angeordnet sind (Li & Liu, 2014). Bei dieser Methode ist es Clustern erlaubt, Untergruppen zu besitzen. Anhand der Struktur des hierarchischen Clusterings, können auch hier zwei Kategorien unterschieden werden (Hemmatian & Sohrabi S. 1524):

1. Trennende Methoden (divisive): In trennenden oder auch top-down Methoden werden zuerst alle Daten als ein Cluster angesehen. Im nächsten Schritt werden alle Daten die wenig Gemeinsamkeiten zu anderen besitzen in separate Cluster aufgeteilt. Dies wird so lange wiederholt, bis in den einzelnen Clustern nur ein Datensatz vorhanden ist.
2. Bottom-up oder agglomerativen Methoden: Bei diesen Methoden werden zuerst alle Datensätze als separate Cluster angesehen und anschließend, je nach Gemeinsamkeit, Datensätze in Clustern zusammengefasst. Diese iterative Methode wird dabei so oft durchlaufen, bis nur noch ein Cluster oder eine vorher festgelegte Menge an Clustern existieren.

(Hemmatian & Sohrabi, 2017, S. 1524 - 1525).

### 2.10.2 Partitionelles Clustering

Das partitionelle Clustering zeichnet sich dadurch aus, dass es mit Sammlungen von Datensätzen arbeitet, die keine Überschneidungen besitzen und jeweils nur zu einem Cluster zugewiesen werden. Das Ziel dabei ist es, die Datensätze so zu Clustern, dass diese mit anderen, mit denen sie die meisten Gemeinsamkeiten besitzen, zusammengeführt werden und zu anderen Clustern die größten Unterschiede aufweisen (Li & Liu, 2014). Das Kriterium, um die Gemeinsamkeiten festzustellen, ist dabei die euklidische Distanz. Beispiele für diese Methode sind k-means oder fuzzy c-means (Hemmatian & Sohrabi, 2017, S. 1523).

### 2.10.3 K-means

Das k-means Clustering basiert darauf, dass am Anfang eines jeden Clusterings eine gewählte Zahl, hierbei k, an zufälligen Punkten in einem, z.B. zwei oder dreidimensionalen, Datenraum gewählt werden. Nachdem dieser erste Schritt durchgeführt wurde, werden die euklidischen Distanzen der einzelnen Datenpunkte und den neuen zufällig gewählten Punkten verglichen. Die Datenpunkte, die dabei am nächsten an einem der neuen Punkte liegen, werden zum jeweiligen Cluster des neuen Punktes hinzugefügt.

Im dritten Schritt wird die euklidische Distanz des Mean-Wertes von den neu gebildeten Clustern errechnet und anschließend wieder neue Cluster gebildet. Dies geschieht so lange, bis die Datenpunkte nach einem Durchgang zu keinem anderen Cluster zugeordnet werden konnten. Abschließend wird die Varianz zwischen den Datenpunkten gemessen und ein neuer Clusterdurchlauf durchgeführt. Die Anzahl der Durchläufe kann dabei anfangs angegeben werden. Ist das Clustering abgeschlossen wird die Cluster-Variante gewählt, in der die geringste Varianz zwischen den Datenpunkten bestand. Werden die Varianzergebnisse gemessen, so wird ersichtlich, dass mit jedem neuen zufälligen Wert (k) die Varianz geringer wird, jedoch immer ab einem bestimmten Punkt in einem Liniendiagramm abknickt und die Varianz langsamer sinkt. Dies ist der Wert, der für das k im k-means Clustering gewählt werden sollte (Wiedenbeck & Züll, 2001, S. 14-15).

## 2.11 Sentiment Analysen

Das Themenfeld der Sentiment Analyse beschäftigt sich damit, veröffentlichten Medien eine Emotion zuzuordnen und diese anschließend entsprechend auszuwerten. Die zwei grundlegenden Aufgaben dabei sind es, Wörter die Gefühle symbolisieren zu extrahieren und diese anschließend auszuwerten. Die Zuordnung erfolgt dabei in drei verschiedenen Sentiments (Gefühlen): positiv, negativ oder neutral. Für diese Zuordnung können dabei unterschiedlichste Algorithmen benutzt werden. (Böck et al, 2017, S 13. und Cambria et al, 2017, S.4).

## 2.12 Ansätze von Sentiment Analysemethoden

Bereits existierende Ansätze von Sentiment Analysemethoden können in drei Hauptkategorien eingeteilt werden: wissensbasierte Methoden, statistische und hybride Ansätze.

Wissensbasierte Ansätze sind dabei sehr beliebt, aufgrund ihrer einfachen Zugänglichkeit, geringen Investitionskosten und ihrer geringen Komplexität. Bei diesen Ansätzen erfolgt die Zuordnung über lexikalische Methoden, dabei werden einzelne oder auch mehrere zusammenhängende Wörter mit Sentiments, wie z.B. „gut“ oder „schlecht“ aus Lexika mit den analysierten Wörtern verglichen und daraus abgeleitet, in welches Sentiment das jeweilige Wort gehört. Die Anzahl der Vorkommen dieser Wörter wird anschließend ausgewertet, um einen Score für den jeweiligen Text zu erstellen (Böck et al, 2017, S 13. und Cambria et al, 2017, S 5). Dieser Score gibt das allgemeine Sentiment für den analysierten Text an, -1 bis >0 negativ, 0 neutral und 0 bis 1 positiv.

Der Hauptschwachpunkt von wissensbasierten Ansätzen für die Sentiment Analyse ist dabei jedoch die schlechte Auswertung von Sentiments von Wörtern, wenn linguistische Regeln hinzugezogen werden (Cambria et al., 2017, S. 6). Im Speziellen sind damit

Situationen gemeint, in denen Ironie, Sarkasmus oder auch doppelte Verneinungen vorkommen. In solchen Fällen ist es wahrscheinlich, dass eine simple wissensbasierte Implementation einer Sentiment Analyse schlechte und wenig aussagekräftige Ergebnisse liefert.

Um diesem Problem entgegenzuwirken, wurden komplexere und ressourcenintensivere Ansätze entwickelt, die auf eine größere Wissensbasis zugreifen können, um Wörter und Bedeutungen akkurater zu erfassen. Implementationen, die nicht auf umfangreiche Wissensbasen zugreifen können, besitzen jedoch weiterhin das grundlegende Problem, die Semantik der natürlichen Sprache nicht gut erfassen zu können (Cambria et al., 2017, S. 6). Die Validität der wissensbasierten Ansätze beruht demnach stark auf der Größe und Qualität der genutzten Trainingsdatensätze.

Ein weiterer Nachteil der wissensbasierten Ansätze liegt in der starren Einteilung von Wörtern in verschiedene Sentiments. Diese strikte Umsetzung verhindert eine Flexibilität, die gebraucht wird, um mit unterschiedlichen Konzepten in Texten akkurat umzugehen (Cambria et al., 2017, S. 6).

Statistische Ansätze der Sentiment Analyse, z.B. SVM (Support Vector Machine) und deep learning, sind beliebt, um Affekte in Texten zu klassifizieren. Bei dieser Methodik können Texte durch einen Machine-Learning-Algorithmus nach ihren Sentiments ausgewertet werden. Um gute Ergebnisse zu erzielen, muss der Machine-Learning-Algorithmus allerdings im Vorfeld mit großen Trainingsätzen versorgt werden, in denen die jeweiligen Wörter, die für die Sentimenteinteilung ausschlaggebend sind, bereits korrekt annotiert wurden. Durch dieses Verfahren ist es ML-Algorithmen möglich, die Wertigkeit von Schlüsselwörtern der Analyse zu erlernen sowie beliebige Nicht-Schlüsselwörter und wiederkehrende Wörter in unterschiedlichen Kontexten richtig auszuwerten (Cambria et al., 2017, S. 6).

Allgemein sind statistische Ansätze jedoch nicht gut geeignet, um die Semantik in Texten korrekt auszuwerten, da lexikalische oder auch wiederkehrende Wörter individuell gesehen keinen großen Wert haben, um korrekte Sentimentklassifizierungen durchzuführen. Von diesem Problem ausgehend ergibt sich dabei die Tatsache, dass statistische Ansätze nur akzeptable Ergebnisse liefern, wenn der Textinput für die ML-Algorithmen groß ist. Dieser Ansatz eignet sich damit zwar für Sentiment Analysen auf Dokument- oder Paragraphenebene, nicht jedoch für kürzere Texte, wie z.B. einzelne Sätze (Cambria et al., 2017, S. 6).

Hybride Ansätze vereinen die Vorteile von wissensbasierten und statistischen Ansätzen, um Sentiment Analysen durchzuführen und qualitativ hochwertige Auswertungen zu erhalten (Cambria et al., 2017, S. 6).

Sentiment Analysen besitzen ein großes Potenzial als Hilfskomponenten für andere Technologien und Systeme zu fungieren. Sentiment Analysen können die Möglichkeiten von Customer-Relationship-Management-Systemen (CRM-Systemen) und Empfehlungssystemen erweitern, in dem sie es ermöglichen herauszufinden, welche Funktionalitäten von Kunden geschätzt werden und welche Produkte aufgrund von schlechten Bewertungen nicht mehr empfohlen werden sollen. Weitere Anwendungsgebiete können die Analyse von erfolgreichen Lehrmaterialien oder erfolgreichen Entertainment-Methoden sein sowie für die Spam-Erkennung und Filterung in der Kommunikation über Social-Media. Unternehmen investieren immer größere Beträge in Marketing-Strategien und die Erkennung und Vorhersage der Meinung und Reaktion von Kunden auf Ihre Produkte und Brands. Die Entwicklung von Tools und Software, die es ermöglichen Sentiments im Web automatisch und in Echtzeit zu analysieren und einzugruppieren, repräsentiert dabei einen der aktivsten Forschungs- und Entwicklungsbereiche (Cambria et al, 2017, S. 2).

### 2.13 Anwendungsebenen von Sentiment Analysen

Wie im vorherigen Abschnitt erläutert, können Sentiment Analysen über unterschiedliche Ansätze durchgeführt werden. Je nach Ansatz sind diese für unterschiedliche Anwendungsfälle besser oder schlechter geeignet. Neben der Auswahl des Ansatzes der Analyseverfahren, ist ein weiterer wichtiger Aspekt bei der Durchführung von Sentiment Analysen, die Ebene, auf der diese durchgeführt wird. Grundlegend gibt es vier unterschiedliche Analyseebenen zwischen denen unterschieden werden kann (Siehe Abbildung 10).

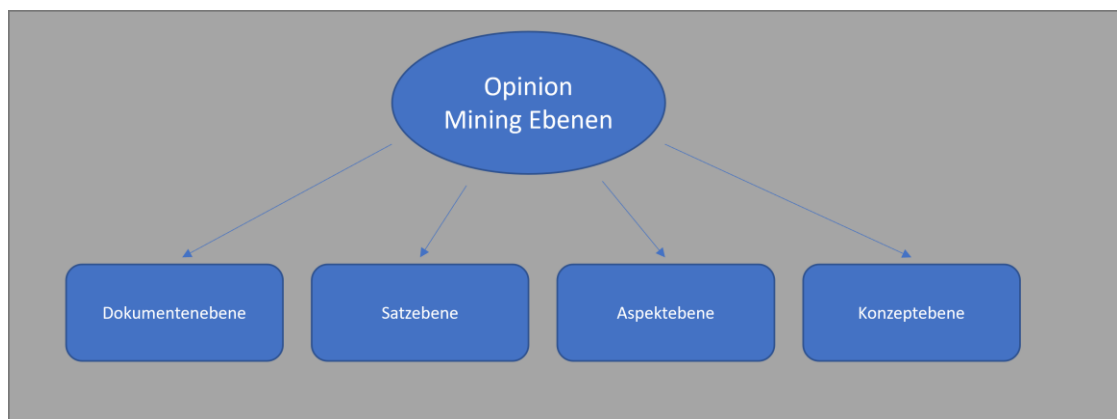


Abbildung 10: Opinion Mining Ebenen, eigene Abbildung, basierend auf Hemmatian & Sohrabi, 2017, S. 1502

#### 2.13.1 Dokumentenebene

Die Dokumentenebene stellt die allgemeinste Ebene dar, auf der Sentiment Analysen durchgeführt werden können. Auf der Dokumentenebene werden alle analysierten Sentiments im Text zusammengefasst und das gesamte Dokument entweder als positiv oder

negativ eingestuft (Pang et al, 2002). Analysen auf Dokumentenebene eignen sich dann, wenn die zu analysierenden Texte auf einem einzigen Thema basieren, entweder positiv oder negativ ausfallen und automatisiert ausgelesen werden sollen (Moraes et al., 2013). Ein Vorteil der Sentiment Analyse auf Dokumentenebene, ist die hohe Analysegeschwindigkeit aufgrund der fehlenden Detailinformationen, die in anderen Ebenen beachtet werden müssen. Analysen auf Dokumentenebene stellen, historisch gesehen, die zu Beginn des Forschungszweiges am häufigsten vertretene Analyseart da (Hemmatian und Sohrabi, 2017, S. 8).

Da auf dieser Analyseebene das gesamte Dokument als ein Objekt mit einem Sentiment angesehen wird, eignet sich diese Ebene jedoch nicht für präzisere Analysen und Vergleiche.

### 2.13.2 Satzebene

Durch die fehlenden Detailinformationen der Sentiment Analyse auf Dokumentenebene waren die Ergebnisse für eine genauere Einschätzung zu allgemein. Um dieses Problem zu lösen, widmete sich die weitere Forschung der Sentiment Analyse auf Satzebene (Wilson et al., 2005; Marcheggiani et al., 2014, S. 273 - 285; Yang & Cardie, 2014, S. 325 - 335; Appel et al., 2016). Vorteil der Analyse auf Satzebene ist es, dass jeder Satz auf sein Sentiment hin untersucht werden kann und damit detailliertere Informationen in auszuwertenden Texten liefert. Die Analyse besteht dabei aus zwei grundlegenden Schritten (Liu und Zhang, 2012):

1. Analyse der Subjektivität und Objektivität des Satzes und anschließende Einteilung in jeweiligen Bereich
2. Auswertung der Polarität des Satzes. Ist er positiv, negativ oder neutral.

Nachteil der Analyse auf Satzebene ist die steigende Komplexität und Dauer der Durchführung im Vergleich zur allgemeineren Dokumentenebene.

### 2.13.3 Aspektebene

Da Sentiment Analysen auf Dokument- und Satzebene nicht immer alle nötigen Informationen liefern, wie z.B. welche Aspekte eines Textes positiv oder negativ aufgefasst werden, wurde für diese Fälle eine zusätzliche Analyseebene geschaffen. Die Aspektebene analysiert die eigentliche Meinung, die im Text geäußert wird direkt, anstatt Strukturen wie Dokumente oder Sätze zu fokussieren (Liu, 2012). Das Ziel auf dieser Analyseebene ist es, die einzelnen Aspekte im Text zu extrahieren und diesen eine Polarität zuzuordnen. Dadurch können Sentiments über verschiedene Aspekte in einem Text gebildet und ausgewertet werden.

#### 2.13.4 Konzeptebene

Die Konzeptebene stellt aufgrund der Tiefe der Sentiment Analyse, unter den zuvor vorgestellten Analyseebenen eine Ausnahme dar. Die Konzeptebene basiert auf der Forschung konzeptioneller Informationen über Gefühle und Emotionen in der natürlichen Sprache. Konzeptuelle Ansätze fokussieren sich dabei auf die Analyse von Semantik in Texten, in denen Emotionen nicht explizit, sondern nur implizit vermittelt werden (Poria et al, 2014).

### 2.14 Technische Umsetzung

Die technische Umsetzung der nun nachfolgenden Analysen wurden mittels der Programmiersprache Python und den Programmbibliotheken Textblob, Tweepy, Numpy, Pandas, NLTK und Matplotlib erstellt. Außerdem wurden Visualisierungen mittels Excel und Power BI angelegt.

#### 2.14.1 Pandas

Pandas ist eine open-source Programmbibliothek, um mittels der Programmiersprache Python einfach, schnell und flexibel Open-Source-Daten zu extrahieren, analysieren und manipulieren (Pandas, 2021).

Hauptfunktionalitäten von Pandas:

- Ein schneller und effizienter DataFrame zur Objektmanipulation mit eingebauter Indexierung.
- Werkzeuge, um Daten zu lesen und in „in-memory-Datenstrukturen in den Formaten CSV, Microsoft-Excel, SQL Database und HDF5-Format zu schreiben.
- Intelligente Datenanordnung und integriertes Handling von fehlenden Daten.

#### 2.14.2 Numpy

Numpy ist eine open-source Programmbibliothek, die sich dem Themenfeld der Array-Programmierung widmet und in diesem Themenfeld mit Python am häufigsten vertreten ist. Numpy ermöglicht es, Daten in Vektoren, Matrizen und höherdimensionierten Arrays anzusprechen, diese zu manipulieren und Operationen mittels dieser auszuführen (Harris, 2020, 30.12.2021).



Hauptfunktionalitäten von NumPy:

- Das NumPy-Array ist die Kernfunktionalität von NumPy und ist dazu da, um multidimensionale Arrays effizient anzusprechen und abzuspeichern. Es ermöglicht eine große Palette an wissenschaftlichen Berechnungen. Es besteht aus einem Pointer, der auf einen Datenspeicherort zeigt und zusätzlichen Metadaten, um den Datenspeicherort so zu interpretieren, dass das Programm weiß welche Daten dort gespeichert sind. Diese Metadaten bestehen aus dem Datentyp, der Datenform und Datenschritten zwischen den dazugehörigen Datenblöcken (Harris, 2020).

### 2.14.3 Tweepy

Tweepy ist eine open-source Programmbibliothek, um mittels Python auf die Twitter API zugreifen zu können (Tweepy, 2021).

### 2.14.4 TextBlob

TextBlob ist eine open-source Programmbibliothek, um Textdaten zu verarbeiten. TextBlob bietet eine einfache API um natural language processing (NLP) Aufgaben durchzuführen.

Hauptfunktionalitäten von TextBlob im NLP-Bereich sind:

- Substantiv(noun)-Phrasen Extraktion
  - Part-of-speech tagging (POS-Tagging)
  - Sentiment Analysen
  - Klassifikationen (Naive Bayes, Decision Tree)
  - Tokenization (aufsplitten von Text in Wörter und Sätze)
  - Wörter und Satz Frequenzen (?)
  - Parsing
  - N-Gramme
  - Wordbeugung (Gruppierung von Wortbeugungen zu einem Element)
  - Rechtschreibkorrektur
  - Funktionalität, um zusätzliche NLP-Modelle mittels Erweiterungen hinzuzufügen
  - WordNet integration
- (TextBlob, 2021)

### 2.14.5 NLTK

NLTK oder auch Natural Language Toolkit genannt, ist eine Sammlung von open-source-Programmbibliotheken, um mittels Python mit Daten der menschlichen Sprache zu arbeiten. Durch NLTK ist es möglich, auf über 50 Trainingsressourcen, wie z.B. WordNet zuzugreifen. Zusätzlich dazu besteht die Möglichkeit auf diverse Libraries im Bereich Natural Language Processing zuzugreifen.

Hauptfunktionalitäten von NLTK im Bereich Natural Language Processing sind:

- Klassifizierung
- Tokenisierung
- Stemming
- Tagging
- Parsen
- Semantisches Reasoning

(NLTK, 2021)

### 2.14.6 Matplotlib

Matplotlib ist eine open-source Programmbibliothek um statische, animierte und interaktive Visualisierungen mit Python zu erstellen.

Hauptfunktionalitäten von Matplotlib:

- Möglichkeit aus einer großen Auswahl von unterschiedlichen qualitativ hochwertigen Visualisierungen auszuwählen
- Erstellung von interaktiven Visualisierungen mit der Möglichkeit, heranzuzoomen und zu aktualisieren
- Anpassung von visuellen Stilen und dem Layout der Visualisierungen
- Visualisierungsexport in unterschiedliche Dateiformate
- Integration in JupyterLab und Möglichkeit Matplotlib direkt in grafische Benutzeroberflächen einzubauen
- Funktionalität third-party Programmbibliotheken mit Matplotlib zu nutzen

(Matplotlib, 2021)

### 2.14.7 Microsoft Excel

Microsoft Excel ist das verbreitetste Tabellenkalkulationsprogramm der Firma Microsoft.

Hauptfunktionalitäten von Excel:

- Tabellenkalkulationsfunktionen
- Makro-Funktionen
- Erstellung von Visualisierungen
- Analysefunktionen (z.B. um den Durchschnitt zu berechnen)
- Bietet viele Schnittstellen zu anderen Microsoft-Softwareprodukten und darüber hinaus

(Microsoft, 2022b)

### 2.14.8 Power BI

Power BI ist ein Geschäftsanalyse Dienst von Microsoft. Mit diesem ist es möglich, Interaktive Datenvisualisierungen zu erstellen und Geschäftsanalyse-Funktionen zu nutzen. Der Nutzer kann mit der Software eigene Reports und Dashboards erzeugen.

Hauptfunktionalitäten von Power BI:

- Relationale Datenbankfunktionalitäten für importierte Datensätze
- Echtzeitanalysen
- Erstellen professioneller interaktiver Berichte
- Automatisierung von Workflows durch Makros
- Aufbereitungsfunktionalitäten für Daten
- Viele Schnittstellen zu anderen Softwareprodukten

(Microsoft, 2022a)

## 3 Analysen der Politiker mit eingangs genannter Technik

Da, wie in der Einleitung erwähnt, bereits Analysen auf Bundestageebene in der Forschungsgemeinde bestehen, wenn auch nur in geringer Anzahl, werden sich die nachfolgenden Analysen in dieser Arbeit, nicht nur auf diese politische Ebene beschränken. Zusätzlich zur Bundestageebene, wird auch die Kommunalpolitiker/Landespolitikerebene näher beleuchtet, um mögliche Unterschiede aufzuzeigen.

Aufgrund der bereits im vorhergehenden Abschnitt zur politischen Historie in Ost- und Westdeutschland festgestellten Unterschiede, werden die zu analysierenden Politiker, außerdem repräsentativ aus den beiden genannten Regionen ausgewählt. So soll es möglich werden, eventuell bestehende Unterschiede in der politischen Kommunikation sichtbar zu machen. Da die Ergebnisse in Bezug zur letzten Bundestagswahl im Jahr 2021 gesetzt werden sollen, werden Tweets seit Jahresbeginn 2021, genauer dem 01.01.2021 bis zum Ende des Jahres, dem 31.12.21, ausgewertet. Die auszuwertenden Parteien werden nach dem aktuellen 20. Bundestag ausgewählt und sind daher folgend:

- SPD (Sozialdemokratische Partei Deutschlands)
- CDU/CSU (Christlich Demokratische Union Deutschlands/Christlich Soziale Union in Bayern)
- Bündnis 90/Die Grünen
- FDP (Freie Demokratische Partei)
- AFD (Alternative Für Deutschland)
- Die Linke

(Bundeswahlleiter, 2022)

Fraktionslose Bundestagsabgeordnete werden in dieser Analyse, aufgrund der geringen Anzahl und der damit einhergehenden schlecht vergleichbaren statistischen Ergebnisse, ausgelassen. Aus den genannten Parteien ergeben sich repräsentativ folgende Spitzenkandidaten für die Sentimentsanalysen auf Bundestagebene, die entweder direkte Kanzlerkandidaten darstellen oder hohe Ämter in ihren jeweiligen Parteien beziehen und aktiv im Wahlkampf teilnahmen:

Auf Bundestagebene ausgewählte Spitzenkandidaten:

SPD: Olaf Scholz

CDU: Armin Laschet

Bündnis 90/Die Grünen: Annalena Baerbock

FDP: Christian Lindner

AFD: Dr. Alice Weidel

Die Linke: Janine Wissler

#### **Kommunal-/Landesebene:**

Rahmenbedingung bei der Analyse von Kommunal-/Landespolitikern sind die entsprechenden Bundesländer in Ost-, bzw. Westdeutschland. Des Weiteren mussten alle ausgewählten Kommunal-/Landespolitiker einen Twitter-Account besitzen und diesen aktiv

betreiben. Eine aktive Nutzung mit möglichst vielen Tweets ist Voraussetzung, um möglichst aufschlussreiche Ergebnisse zu erzielen. Als Kennzahl für eine aktive Benutzung wurde der Wert von mindestens 100 Tweets innerhalb des Jahres 2021 festgelegt. Diese Kennzahl repräsentiert den größten gemeinsamen Nenner unter den ausgewählten Politikern.

Die ursprüngliche Idee war es, nur Kommunal- bzw. Landespolitiker zu analysieren, die direkt in den Bundestag gewählt wurden, um nur die Politiker auszuwählen, die einen eindeutigen Erfolg durch ihre Kommunikation zu ihren Wählern erzielen konnten. Dieses Vorhaben musste jedoch, aufgrund von fehlenden Repräsentanten der jeweiligen Parteien in den entsprechenden Bundesländern, verworfen werden. Schlussfolgernd wurden in die Analysen Kommunal-/Landespolitiker miteinbezogen, die direkt in ihr Bundestagsmandat gewählt wurden sowie über entsprechende Listenplätze ihrer Partei. Die restlichen Kommunal-/Landespolitiker wurden nach dem Zufallsprinzip ausgewählt. Die Analyse der Politiker auf Bundes- und Kommunal-/Landesebene wird nachfolgend auf unterschiedliche Art und Weise dargestellt. Aufgrund der großen Anzahl an Kommunal-/Landespolitikern und der damit einhergehenden großen Anzahl an Daten, werden diese teilweise je nach Ihrem jeweiligen Bundesland gruppiert dargestellt und direkt miteinander verglichen. Auf Bundestagebene, aufgrund der höheren Priorität und der geringeren Anzahl an Politikern, werden detailliertere Analyseergebnisse je Politiker dargestellt. Die Ergebnisse auf Bundestagebene werden in 5 Visualisierungen dargestellt.

Auf Bundestagebene werden folgende Daten visualisiert:

#### **Sentiment-Verteilung mit Darstellung der Polarität und Subjektivität:**

Diese Visualisierung gibt eine Übersicht über die Verteilung der Subjektivität und Polarität der analysierten Tweets an. Die Polarität befindet sich hierbei auf der x-Achse des Koordinatensystems und kann von -1 (negativ) über 0 (neutral) bis zu 1 (positiv) die Verteilung des Sentiments der Tweets angeben. Auf der y-Achse ist die Ausprägung der Tweets zu erkennen, von 0 (objektive Texte) bis hin zu 1 (subjektive Texte).

#### **Sentiment-Verteilung mit Darstellung der Anzahl aller Tweets und ihren dazugehörigen Sentiments:**

Durch diese Visualisierung ist es möglich, einen schnellen Überblick über die Verteilung der Tweets der analysierten Person und den jeweiligen Sentiments zu erkennen. Auf der x-Achse des Koordinatensystems befindet sich die Anzahl der Tweets und komplementär dazu, auf der y-Achse, die jeweiligen Sentiments der Tweets. Die Daten werden dabei, zur einfachen Darstellung, in einem vertikalen Balkendiagramm dargestellt.

Um einen detaillierteren Einblick auf die Resonanz der veröffentlichten Tweets zu erhalten, wurden zwei Grafiken zur Darstellung der Mittelwerte von Retweets und Likes je nach Sentiment entworfen.

### **Vergleich aller veröffentlichten Tweets der Bundespolitiker im Jahr 2021**

Diese Visualisierung zeigt die Anzahl, aller veröffentlichten Tweets des Jahres 2021, der Bundespolitiker je nach Partei gruppiert an. Dadurch soll ein Überblick über die Aktivität der einzelnen Bundespolitiker ermöglicht werden. Auf der x-Achse befinden sich die Bundespolitiker je nach ihrer Partei gruppiert. Auf der y-Achse ist die Anzahl der Tweets erkennbar.

### **Mittelwert von Retweets je Sentiment**

In dieser Visualisierung wird das arithmetische Mittel aller Likes, verteilt auf die jeweiligen Sentiments der dazugehörigen Tweets, ermittelt. Auf der x-Achse befindet sich der Mittelwert der Anzahl der Retweets und auf der y-Achse die jeweiligen Sentiments als vertikales Balkendiagramm visualisiert. Sinn und Zweck dieser Darstellung ist es, auf einen Blick zu erkennen, wie oft Tweets der unterschiedlichen Sentiments retweetet (geteilt) werden und sich so mehr verbreiten als andere.

### **Mittelwert von Likes je Sentiment**

Diese Visualisierung kann komplementär zur Visualisierung des Mittelwerts der Retweets je Sentiment angesehen werden. Unterschied bei dieser Darstellung ist, dass auf der x-Achse anstatt der Retweets, der Mittelwert der Likes dargestellt wird. Vorteil dieser Darstellung ist es, die direkte Resonanz auf die jeweils veröffentlichten Inhalte mit einem Blick erfassen zu können. So können Rückschlüsse für die Interaktion zwischen Politiker und Twitter-Nutzern gezogen werden.

### **Wordcloud:**

Für jeden Politiker wird eine Wordcloud erstellt, die auf den zuvor für die Analyse vorbereiteten Texten der Tweets des Jahres 2021 basiert. Die Wordcloud erstellt dabei eine Visualisierung der meistvertretenen Themen bzw. Wörtern, die in den analysierten Texten gefunden wurden. Je größer das entsprechende Wort dargestellt wird, desto häufiger wurde es genannt. Um bessere Ergebnisse zu erzielen, wurden deutsche Stoppwörter mittels eines Text-Corpus von Stoppwörtern aus der NLTK-Programmbibliothek und einer manuell erstellten Erweiterung dieser herausgefiltert. Stoppwörter sind in diesem Kontext Wörter, die sehr häufig vorkommen und keine Relevanz für die Erfassung der Themenschwerpunkte der analysierten Person darstellen und daher vernachlässigt werden können.

### 3.1 Bundestagebene

Da alle ausgewählten Politiker der Bundestagebene eine ähnliche proportionale Verteilung der Tweets je Sentiment aufweisen, wird nachfolgend auf das Ergebnis der Visualisierung von Olaf Scholz exemplarisch eingegangen und die restlichen Visualisierungen der Sentimentsverteilung für die anderen Politiker im Anhang aufgeführt. Ebenso stimmen die Darstellungen zur Anzeige der Polarität und Subjektivität in großen Teilen überein. Daher wird auch hier exemplarisch beim Politiker Olaf Scholz darauf eingegangen, der Rest wird im Anhang aufgeführt.

#### Vergleich aller veröffentlichten Tweets auf Bundesebene nach ausgewählten Politikern, 2021

Abbildung 11, stellt die Aktivität auf Twitter, durch die Anzahl aller veröffentlichten Tweets, des Jahres 2021 dar. Mit 826 Tweets hat Olaf Scholz der SPD, mit einem großen Abstand, die meisten Tweets verfasst. Mit 241 Tweets weniger, also 585, befindet sich Christian Lindner der FDP auf dem zweiten Platz. Die drittaktivste Politikerin ist Janine Wissler der Linken mit 478 veröffentlichten Tweets. Dr. Alice Weidel der AfD setzte 309 Tweets ab. Annalena Baerbock von Bündnis 90/Die Grünen verschickte 281 Tweets und auf die geringste Tweet-Anzahl kommt Armin Laschet der CDU, mit 186 Tweets.

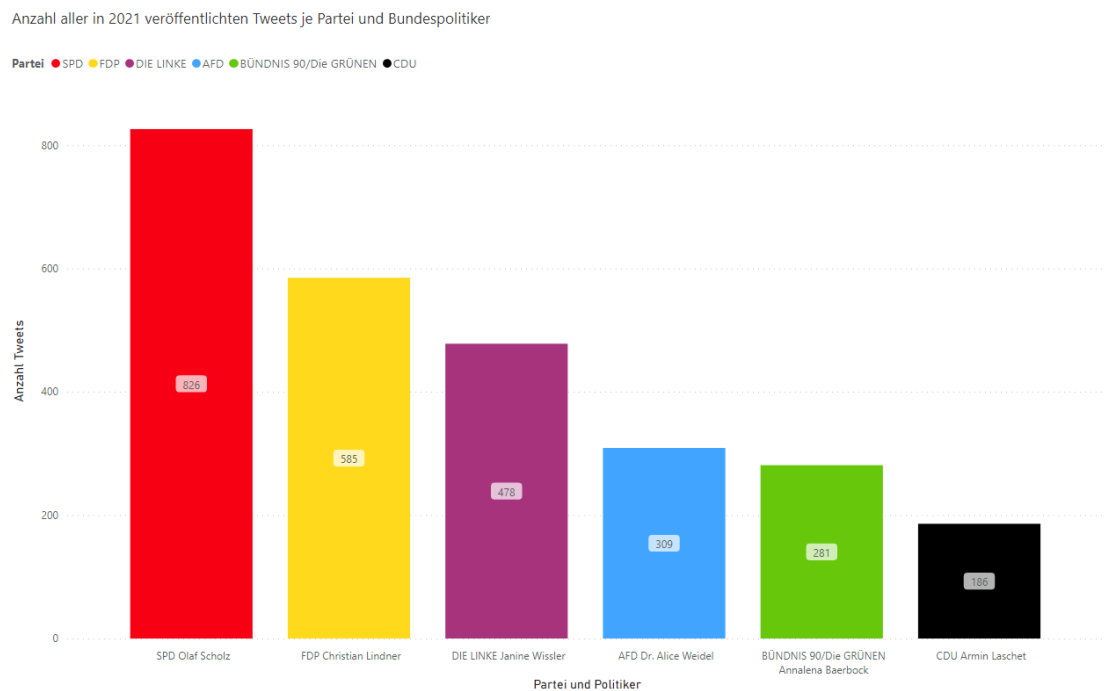


Abbildung 11: Anzahl aller im Jahre 2021 veröffentlichten Tweets, je Partei und Bundespolitiker gruppiert

### 3.1.1 Olaf Scholz, SPD

Olaf Scholz Twitter-Profil hat im Jahre 2021 insgesamt 826 Tweets veröffentlicht, davon ist der Großteil, 766 Stück (92,7%) neutral, gefolgt von 51 positiven Tweets (6,2%) und 9 negativen (1,1%). Olaf Scholz hat damit im Durchschnitt 2,2 Tweets pro Tag veröffentlicht. Eine Besonderheit ist, dass seine Twitter-Aktivität mit seinen 826 Tweets, unter den ausgewählten Spitzenkandidaten des Bundestages, am höchsten ist (Siehe Abbildung 12).

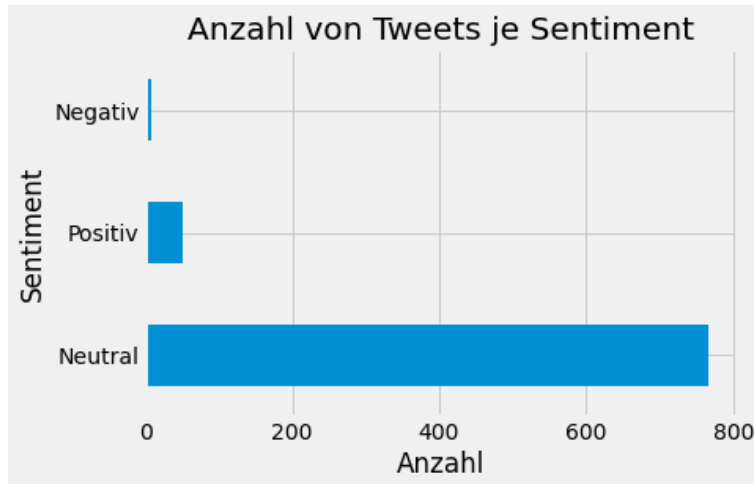


Abbildung 12: Sentiment Verteilung und Anzahl Tweets, Olaf Scholz, SPD

Olaf Scholz Tweets sind überwiegend neutral und damit bei einer Polarität von 0. Auffallend ist, dass seine neutral eingeordneten Tweets einen deutlichen Ausschlag in die Subjektivität beinhalten (Siehe Abbildung 13). Diese Dynamik kann entstehen, wenn Tweets oft aus der eigenen Position heraus „Ich finde, Ich denke etc.“ geschrieben werden, dabei aber eher allgemeine Aussagen betreffen.

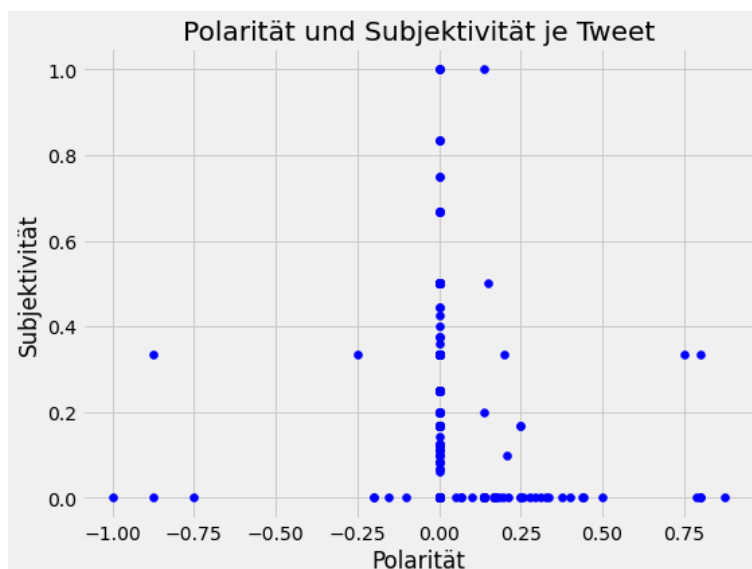


Abbildung 13: Verteilung Polarität und Subjektivität von Tweets, Olaf Scholz, SPD





Vergleicht man die Mittelwerte der Likes je Sentiment (Abbildung 15) erkennt man, dass die Anzahl der Likes und damit die positive Resonanz bei positiven und neutralen Tweets am höchsten ist. Mit ca. 60% der Anzahl der Likes, ist die Resonanz auf negative Tweets deutlich von den anderen beiden Sentiments abgeschlagen.

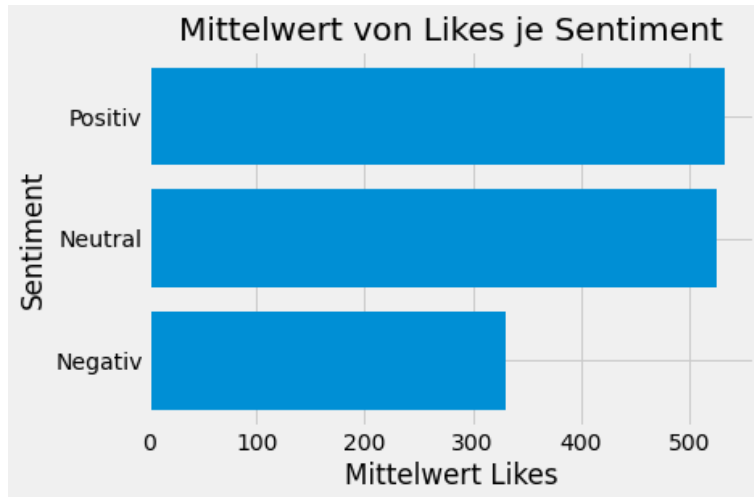


Abbildung 15: Mittelwert Likes je Sentiment, Olaf Scholz, SPD

Komplementär dazu erkennt man beim Mittelwert der Retweets, der Anzahl an Teilungen der Tweets von Olaf Scholz von seinen Lesern (siehe Abbildung 16), dass diese positive Tweets deutlich häufiger teilen als neutrale und negative Tweets.

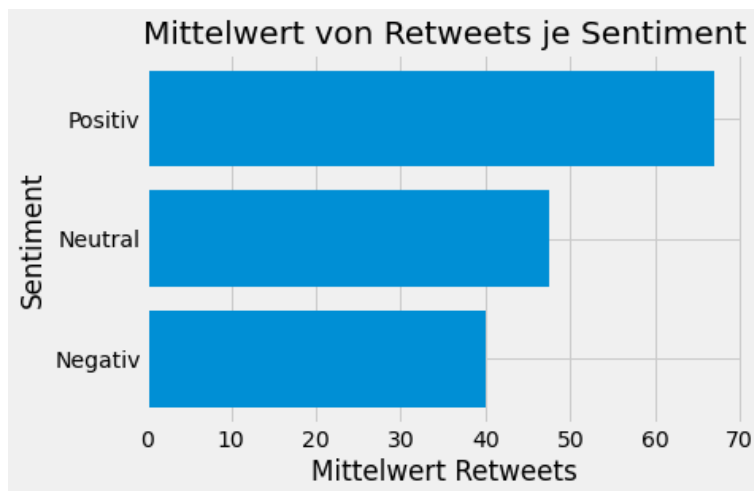


Abbildung 16: Mittelwert Retweets je Sentiment, Olaf Scholz, SPD

### 3.1.2 Armin Laschet, CDU

Armin Laschet legt laut der analysierten Tweets, sein Hauptaugenmerk auf die Themen Pandemie, Deutschland, das Bundesland in seinem Zuständigkeitsbereich Nordrhein-Westfalen und Schlüsselthemen wie die Zukunft, Gespräche und das Thema Kampf (Siehe Abbildung 17).



Wird dazu der Mittelwert der Retweets je Sentiment betrachtet, ergibt sich ein ähnliches Ergebnis. Neutrale Tweets wurden mit etwas über 65-mal retweetet, danach kommen negative Tweets mit einer Retweetanzahl zwischen 45-50 und am wenigsten retweetet werden positive Tweets, 35-40 (Siehe Abbildung 19).

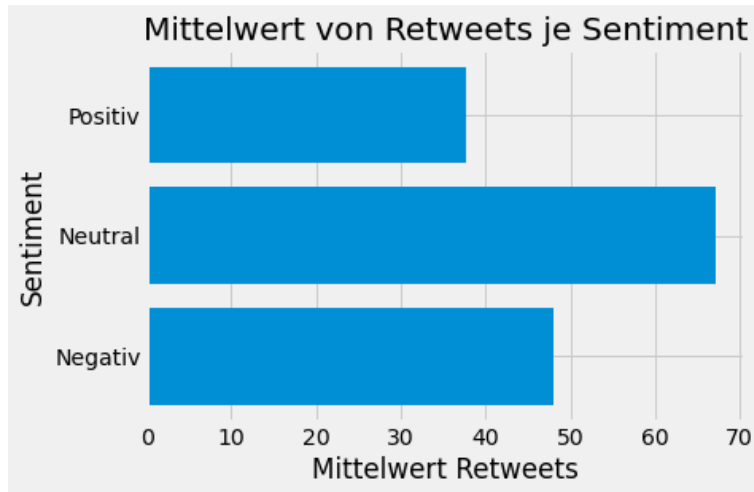


Abbildung 19: Mittelwert Retweets je Sentiment, Armin Laschet, CDU

### 3.1.3 Annalena Baerbock, Bündnis 90/Die Grünen

Laut den meistgenutzten Wörtern von Annalena Baerbock, liegt Ihr Fokus auf den Themen Klimaschutz, Menschen, der Bundestagswahl und Bundesregierung, Deutschland und Europa, der Pandemie, Freiheit und gemeinsam (Siehe Abbildung 20).

Wörter für Themenbereiche:

- **Klimaschutz:** Klimaschutz, Klimakrise
- **Menschen:** Menschen, Kinder
- **Bundestagswahl und Bundesregierung:** Bundesregierung, Demokratie, btw21
- **Deutschland und Europa:** Deutschland, Europa, EU
- **Pandemie:** Pandemie, Corona
- **Freiheit:** Freiheit
- **Gemeinsam:** gemeinsam





Auch die Mittelwerte der Likes je Sentiment bei Christian Lindner sind im Vergleich zu den Politikern Olaf Scholz und Armin Laschet erhöht. Auffallend ist der hohe Mittelwert von 1750 Likes bei positiv eingeordneten Tweets. Damit unterscheidet sich das Ergebnis grundlegend von den anderen Politikern, bei denen das positive Sentiment die niedrigsten Zahlen vorweisen konnte. Neutrale Tweets erhalten als Mittelwert etwas über 1000 Likes, negative 1000. Damit sind beide Sentiments, was die Likes angeht, sehr nah beieinander (Siehe Abbildung 24).

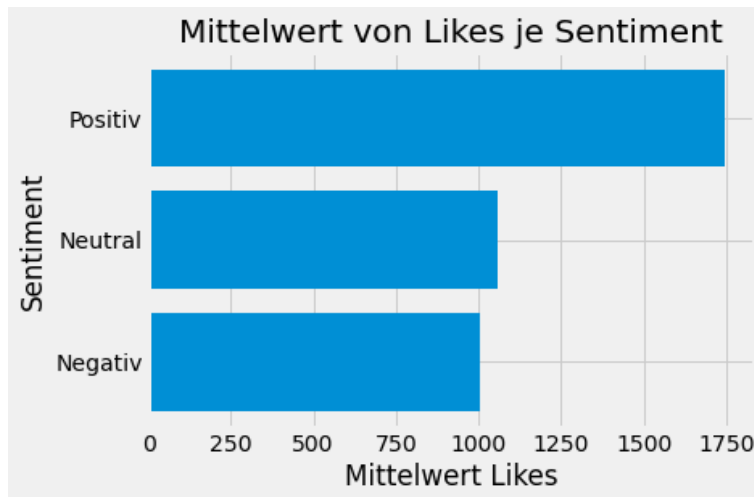


Abbildung 24: Mittelwert Likes je Sentiment, Christian Lindner, FDP

Auch bei der Retweetanzahl ergibt sich ein ähnliches Bild. Mit über doppelt so vielen Retweets pro Tweet liegt das positive Sentiment vorne. Neutrale und negative Tweets verzeichnen ca. die gleiche Anzahl von Retweets, 70 Stück (Siehe Abbildung 25).

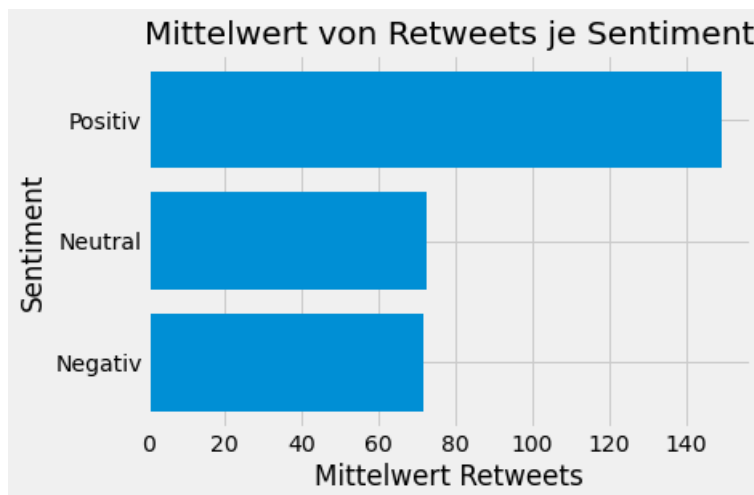


Abbildung 25: Mittelwert Retweets je Sentiment, Christian Lindner, FDP



### 3.1.5 Dr. Alice Weidel, AfD

In dieser Visualisierung sind die Themen, auf die sich fokussiert wurde, außer der eigenen Partei, Corona, SPD, CDU, die Bundesregierung, Bürger, Migration, Freiheit, Politik, die Ampel-Regierung und Europa (Siehe Abbildung 26).

Wörter für Themenbereiche:

- **Eigene Partei:** AfD, AfD Bundestag, AfD btw21
- **Corona:** Corona, Lockdown, Impflucht
- **SPD und CDU:** SPD, CDU
- **Regierung:** Regierung, Kanzlerin, Bundestag, Bundesregierung, Merkel
- **Bürger:** Bürger, Menschen
- **Migration:** Migration
- **Freiheit:** Freiheit
- **Politik:** Politik
- **Ampel-Regierung:** Ampel
- **Europa:** EU



Abbildung 26: Wordcloud, Dr. Alice Weidel, AfD

Auch bei den Mittelwerten der Likes je Sentiment bei Dr. Alice Weidel kann eine erhöhte Aktivität festgestellt werden. Die meisten Likes erhält sie bei positiven Tweets (ca. 2500), gefolgt von neutralen (ca. 1900) und danach negativen (ca. 1750). (Siehe Abbildung 27).



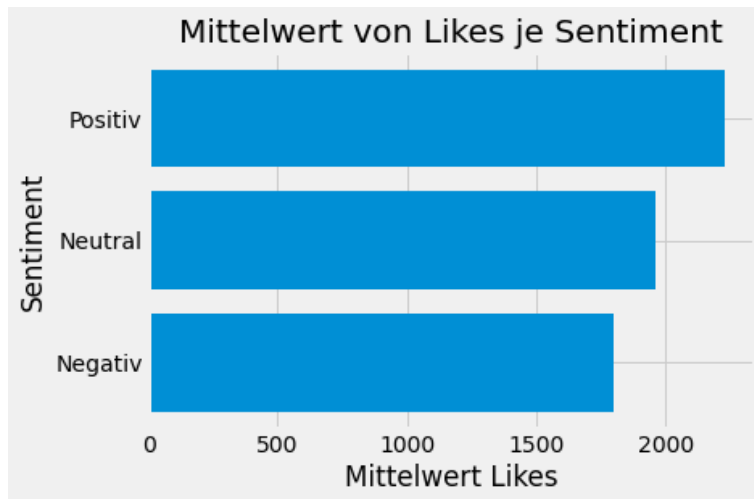


Abbildung 27: Mittelwert Likes je Sentiment, Dr. Alice Weidel, AfD

Bei den Mittelwerten der Retweets je Sentiment kann die gleiche Struktur wie bei den Likes beobachtet werden. Die meisten Retweets erhalten positive Tweets (über 500 Stück), gefolgt von neutralen (über 400) und negativen knapp dahinter (auch über 400 Retweets). Damit ist hier über alle Sentiments hinweg, ein Mittelwert von über 400 erkennbar, welcher der höchste von allen ausgewerteten Bundespolitikern ist (Siehe Abbildung 28).

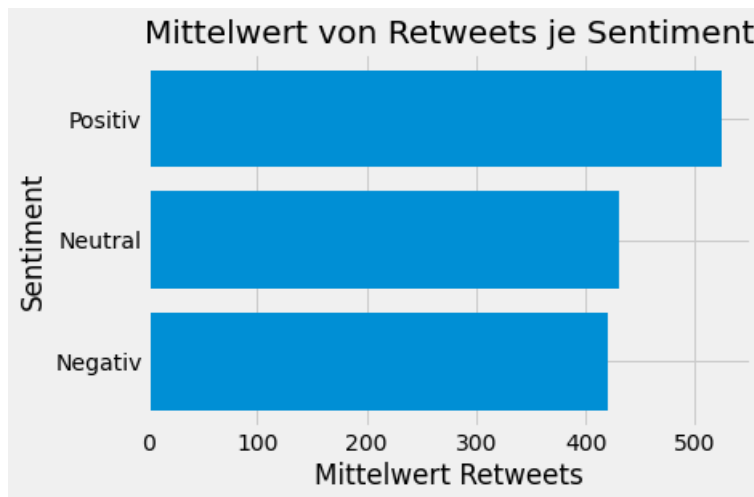


Abbildung 28: Mittelwert Retweets je Sentiment, Dr. Alice Weidel, AfD



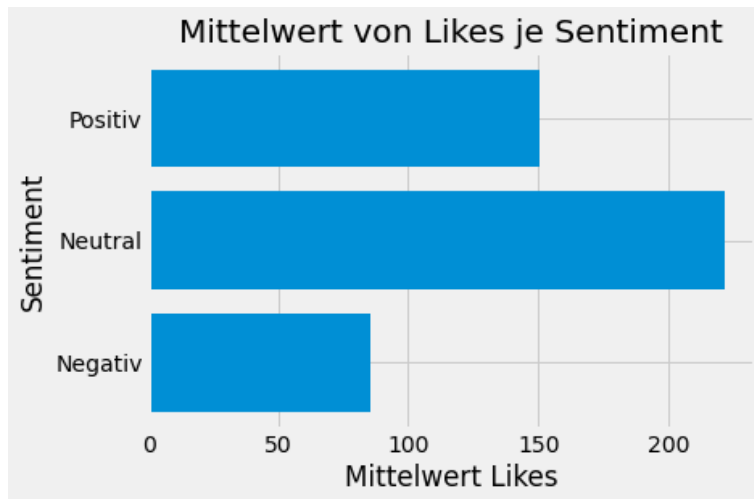


Abbildung 30: Mittelwert Likes je Sentiment, Janine Wissler, Die Linke

Auch bei den Mittelwerten der Retweets ergibt sich proportional das gleiche Bild. Neutral eingetragene Tweets werden im Mittel am meisten retweetet (ca. 35-mal), gefolgt von dem positiven Sentiment (im Bereich von 15 - 20 Tweets) und dem negativen (7 – 10 Tweets). Ähnlich wie bei der Like-Anzahl, ist auch der Mittelwert der Retweets bei Janine Wissler, von den ausgewählten Bundespolitikern, am niedrigsten (Siehe Abbildung 31).

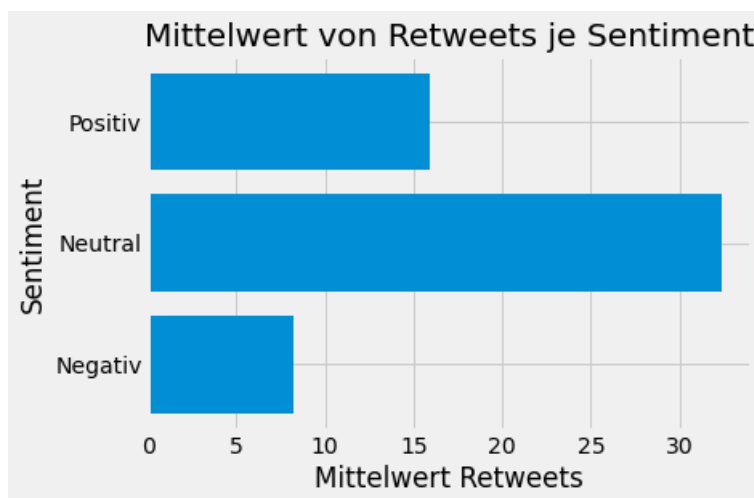


Abbildung 31: Mittelwert Retweets je Sentiment, Janine Wissler, Die Linke

### 3.2 Repräsentative Bundesländer Ost- und Westdeutschland

Für Westdeutschland wurde ursprünglich repräsentativ das Bundesland Rheinland-Pfalz ausgewählt. Für Ostdeutschland das Bundesland Sachsen. Hintergrund dieser Auswahl war, dass beide Bundesländer in wichtigen Eckpunkten, wie Population (Rheinland-Pfalz 4,098 Millionen, Sachsen 4,056 Millionen) und Fläche (Rheinland-Pfalz 19.858 km<sup>2</sup>, Sachsen 18.449 km<sup>2</sup>) (Statistisches Bundesamt, 2021) vergleichbar sind. Aufgrund eingangs erwähnter Limitationen der möglichen Kandidaten für die Analyse (Politiker muss

Twitter besitzen und im Jahr 2021 aktiv, also mindestens 100 Tweets veröffentlicht haben) mussten für Westdeutschland Politiker aus den umliegenden Bundesländern Hessen und Niedersachsen und für Ostdeutschland ein Politiker aus Berlin hinzugezogen werden. Aufgrund der Menge der analysierten Daten auf Kommunal-/Landespoliterebene, werden die nachfolgenden Ergebnisse, je nach Partei und Bundesland, gruppiert dargestellt.

### 3.2.1 Kommunal-/Landesebene Westdeutschland:

Repräsentativ für die einzelnen Parteien in Westdeutschland wurden die folgenden Politiker ausgewählt:

- Andreas Bleck für die AfD in Rheinland-Pfalz (über Landesliste gewählt)
- Thomas Hitschler für die SPD in Rheinland-Pfalz (direkt gewählt)
- Dr. Katja Leikert für die CDU in Hessen (über Landesliste gewählt)
- Katja Adler für die FDP in Hessen (über Landesliste gewählt)
- Heidi Reichinnek für die Linke in Niedersachsen (über Landesliste gewählt)
- Helge Limburg für Bündnis 90/Die Grünen in Niedersachsen (über Landesliste gewählt)

Betrachtet man die repräsentativ ausgewählten Kommunal- und Landespolitiker in den folgenden Visualisierungen, so erkennt man, dass die Vertreter der SPD mit 970 und die CDU, mit 836 Tweets im Jahr 2021, mit großem Abstand die meisten Tweets verfassten (Siehe Abbildung 32).

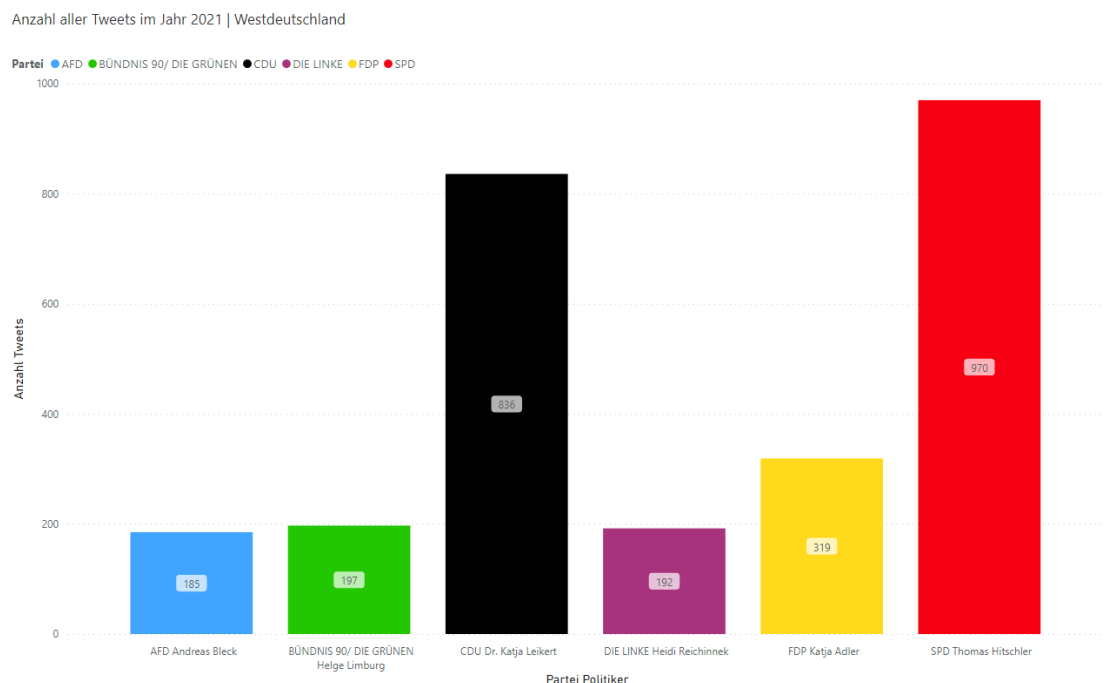


Abbildung 32: Anzahl aller Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland

Auffallend ist, dass der Repräsentant der AFD, der mit 185 Tweets die geringste Tweet-Anzahl vorweist, eine unverhältnismäßig hohe Anzahl an negativen Tweets hat (7 Tweets) (siehe Abbildung 33). Bei den Parteien Bündnis 90/Die Grünen, der Linken und der FDP, ist dies nicht zu beobachten. Die höchste Anzahl an negativen Tweets verzeichnen die Repräsentanten der SPD und CDU mit jeweils 12 Tweets.

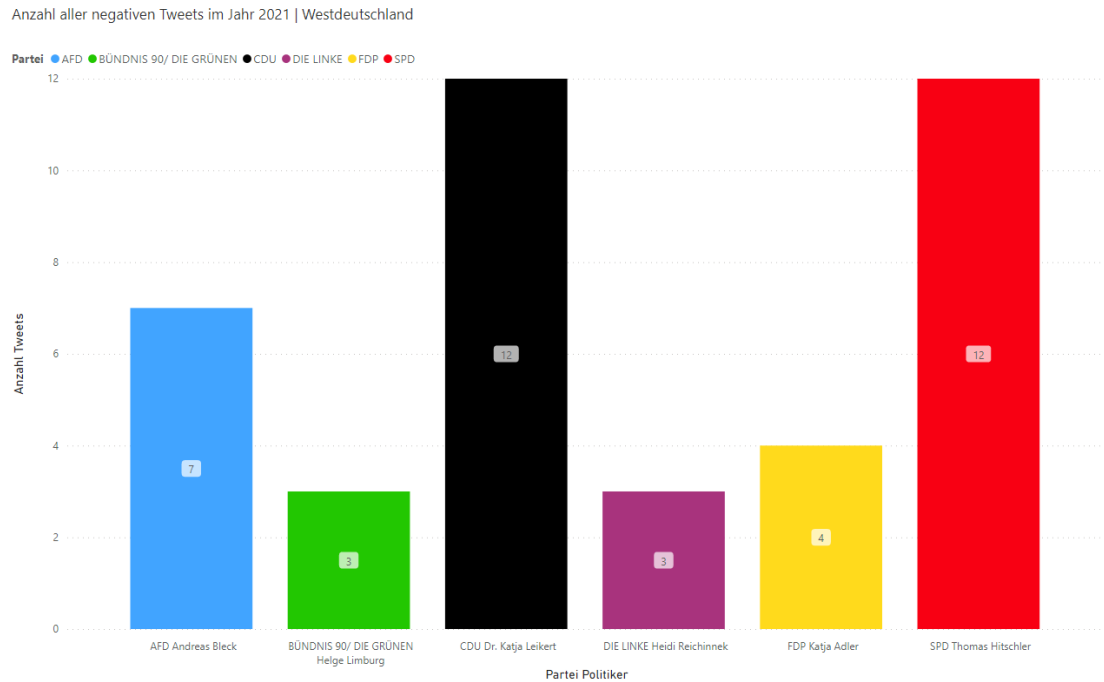


Abbildung 33: Anzahl aller negativen Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland

Betrachtet man in Abbildung 34 die Verteilung der positiven Tweets, so lässt sich feststellen, dass der Repräsentant der CDU deutlich mehr Tweets (+28 Stück) als die SPD veröffentlicht hat. Dies ist besonders auffällig, da der Repräsentant der SPD, allgemein mehr Tweets besitzt. Des Weiteren ist die proportional hohe Anzahl an positiven Tweets (24 Stück) bei der Repräsentantin der Linken auffällig. Ausserdem ist anzumerken, dass Bündnis 90/Die Grünen trotz ungefähr gleicher Menge an veröffentlichten Tweets, wie die AFD und die Linke, mit 2 positiven Tweets, die niedrigste Anzahl innehat.

Anzahl aller positiven Tweets im Jahr 2021 | Westdeutschland

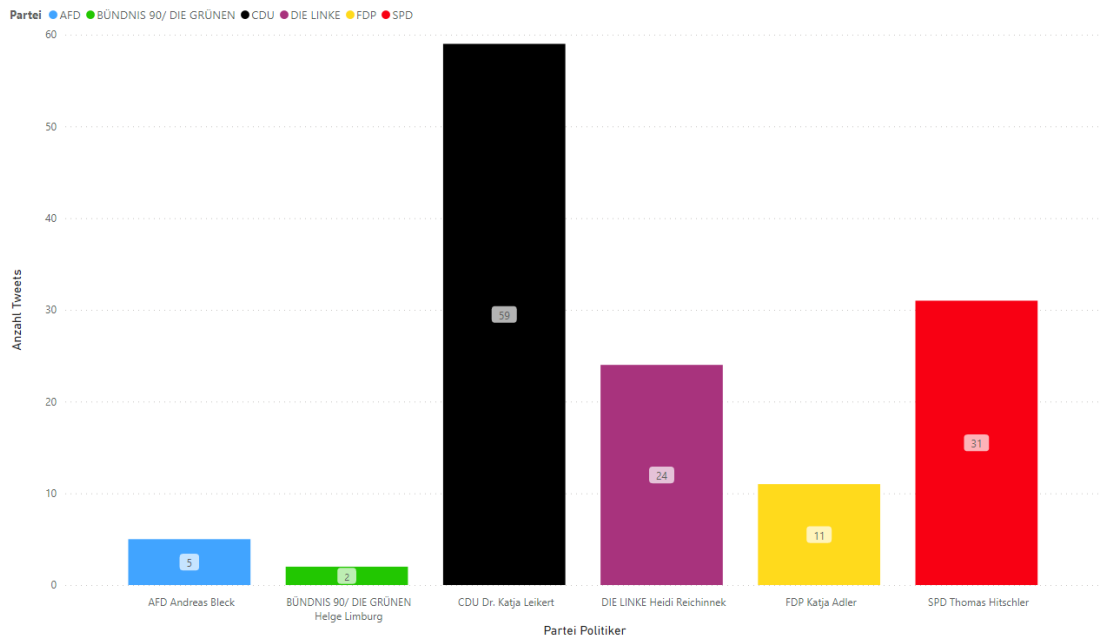


Abbildung 34: Anzahl aller positiven Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland

In Abbildung 35 ist zu erkennen, dass der Großteil aller veröffentlichten Tweets in das neutrale Sentiment eingeordnet wurden. Auffallend ist die sehr ähnliche Verteilung der Tweets in Vergleich zu Abbildung 32 in der die Gesamtanzahl aller Tweets dargestellt wurde.

Anzahl aller neutralen Tweets im Jahr 2021 | Westdeutschland

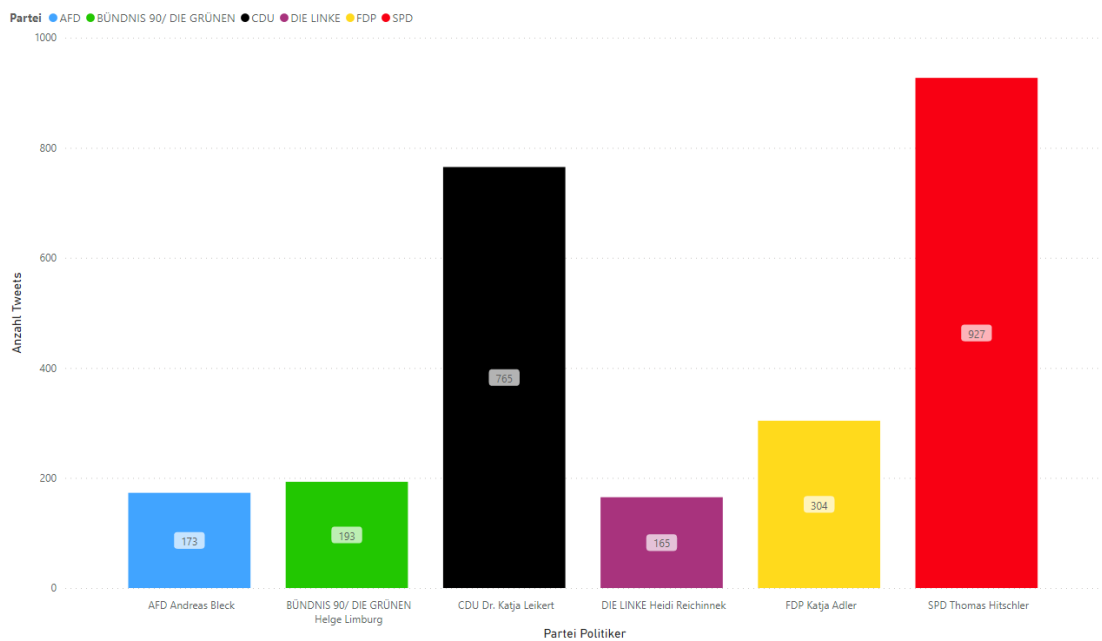


Abbildung 35: Anzahl aller neutralen Tweets nach Parteien, Rheinland-Pfalz, Westdeutschland

### 3.2.2 Kommunale-/Landesebene Ostdeutschland:

Repräsentativ für die einzelnen Parteien in Westdeutschland wurden die folgenden Politiker ausgewählt:

- Dr. Otilie Klein für die CDU in Berlin (über Landesliste gewählt)
- Philipp Hartewig für die FDP in Sachsen (über Landesliste gewählt)
- Clara Bünger für die Linke in Sachsen (über Landesliste gewählt)
- Bernhard Herrmann für Bündnis 90/Die Grünen in Sachsen (über Landesliste gewählt)
- Edgar Naujok für die AfD in Sachsen (direkt gewählt)
- Holger Mann für die SPD in Sachsen (über Landesliste gewählt)

#### Tweetverteilung der Repräsentanten von Ost- und Westdeutschland je Partei

In der folgenden Abbildung 36, welche die Anzahl aller Tweets der Repräsentanten ihrer jeweiligen Partei ist Ostdeutschland abbilden, ist zu erkennen, dass die Linke mit großem Abstand, die meisten Tweets verfasst hat (1311 Tweets). Mit etwas über der Hälfte der Tweets (754) reiht sich der Repräsentant des Bündnis 90/Die Grünen mit den zweitmeisten Tweets ein. Anzumerken ist, dass Bündnis 90 die Grünen, die SPD (696 Tweets) und die CDU (615 Tweets) alle eine ähnliche Quantität der Tweets vorweisen. Auf dem vorletzten Platz liegt die FDP (221 Tweets) und die geringste Tweet-Aktivität verzeichnet die AfD (108 Tweets).

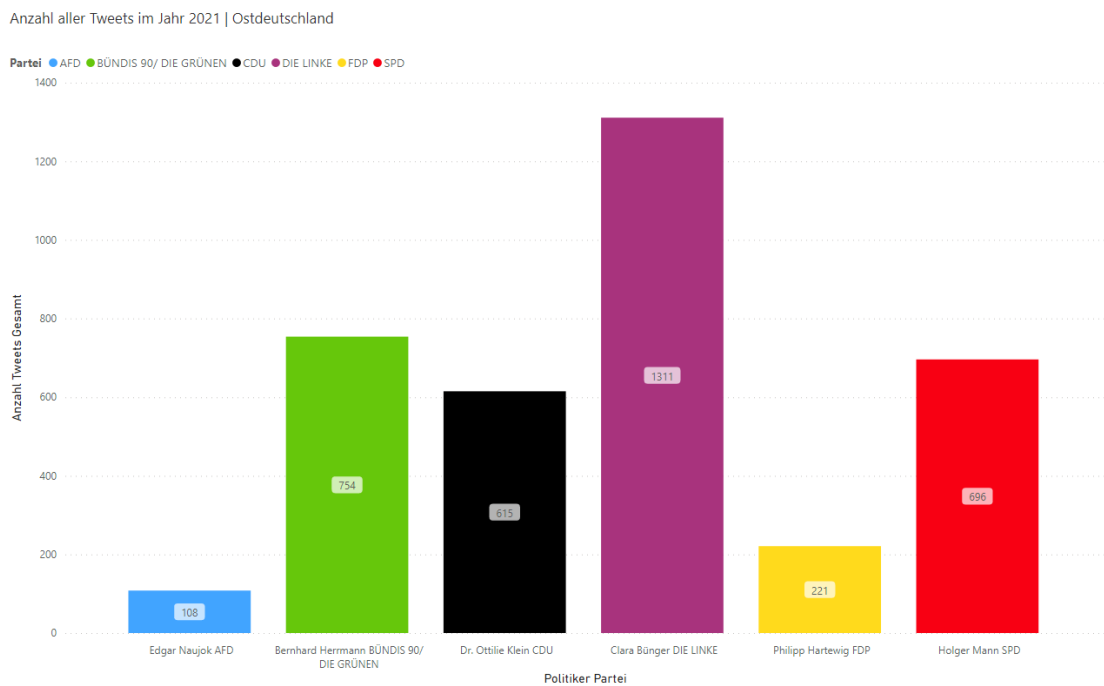


Abbildung 36: Anzahl aller Tweets nach Parteien, Sachsen, Ostdeutschland

In Abbildung 37 ist die Verteilung aller negativen Tweets über die Repräsentanten der Parteien hinweg zu erkennen. Auffällig hierbei ist, dass die Verteilung im Vergleich zur Abbildung 1 mit allen veröffentlichten Tweets, nahezu identisch ist. Der einzige nennbare Unterschied ist, dass nun die SPD am zweithäufigsten Tweets dieses Sentiment veröffentlicht hat (21 Tweets) und vor Bündnis 90/Die Grünen (17 Tweets) liegt.

Anzahl aller negativen Tweets im Jahr 2021 | Ostdeutschland

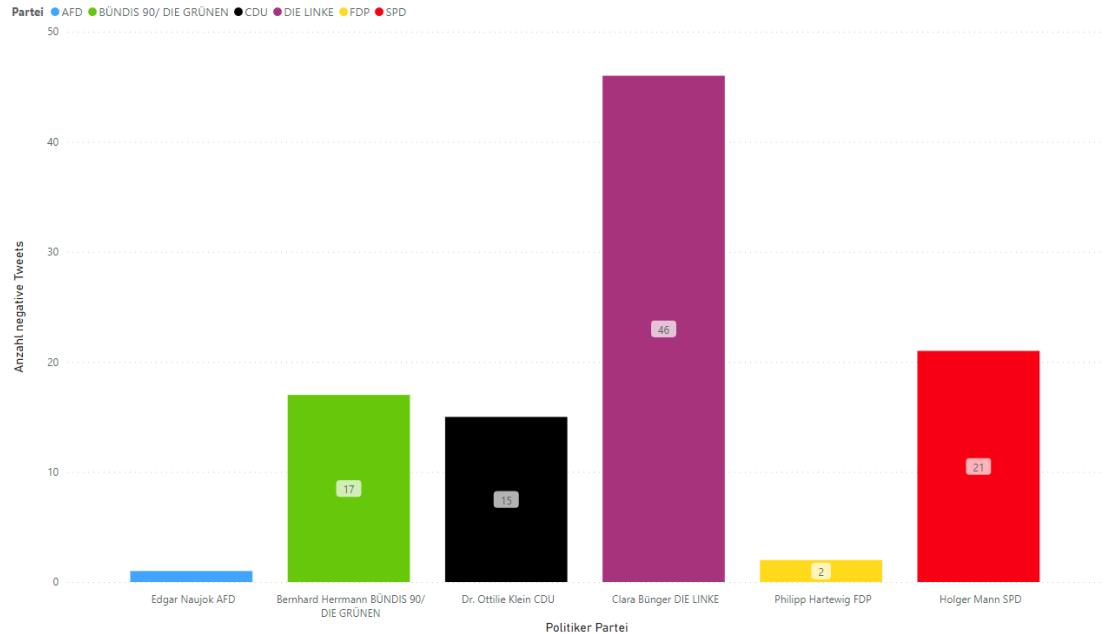


Abbildung 37: Anzahl aller negativen Tweets nach Parteien, Sachsen, Ostdeutschland

Auch Abbildung 38, in der alle positiven Tweets aufgelistet werden, ähnelt Abbildung 36 und 37 in großen Teilen. Die CDU hat eine proportional geringe Anzahl von positiven Tweets.



Anzahl aller positiven Tweets im Jahr 2021 | Ostdeutschland

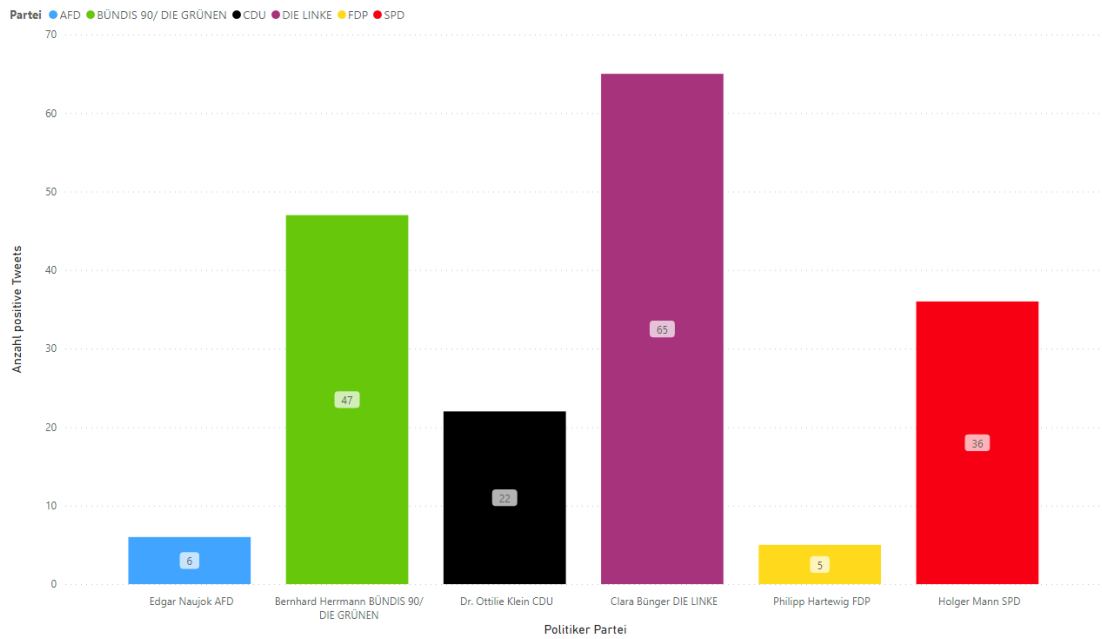


Abbildung 38: Anzahl aller positiven Tweets nach Parteien, Sachsen, Ostdeutschland

Abbildung 39 von Ostdeutschland zeigt, ähnlich wie Abbildung 35 von Westdeutschland, dass die meisten der ausgewerteten Tweets dem neutralen Sentiment zugeordnet wurden. Dabei ist die Verteilung der Tweets proportional vergleichbar mit Abbildung 36, welche alle Tweets abgebildet hat.

Anzahl aller neutralen Tweets im Jahr 2021 | Ostdeutschland



Abbildung 39: Anzahl aller neutralen Tweets nach Parteien, Sachsen, Ostdeutschland

## **Anzahl der Politiker in Werten und Prozent und deren Social-Media-Nutzung in Ost- und Westdeutschland**

In diesem Abschnitt geht es um die tatsächliche Nutzung von Social-Media innerhalb der Parteien in Ost- und Westdeutschland. Für dieses Vorhaben wurden alle neu gewählten Bundestagsabgeordneten, dies umfasst vorherige Kommunal-/Landes- und bereits bestehende und wiedergewählte Bundespolitiker, entsprechend ihrer Parteizugehörigkeit und Region der Tätigkeit, in Werten und Prozenten aufgelistet. Auf der x-Achse werden die jeweiligen Parteien gruppiert dargestellt, auf der y-Achse wird die Anzahl der Politiker dargestellt. In den jeweiligen gruppierten Parteien sind drei unterschiedlich eingefärbte Balken zu erkennen. Der orangefarbene Balken gibt die Gesamtzahl der Politiker der jeweiligen Partei an, der hellblau eingefärbte Balken, die Anzahl der Politiker der jeweiligen Partei, die mindestens einen Social-Media-Kanal nutzen und der dunkelblaue, die Anzahl der Politiker, die ein Twitter-Profil besitzen. In den nachfolgenden Abbildungen 40 und 42, welche die Anzahl der Abgeordneten in den Parteien darstellen, wird speziell auf die Gesamtanzahl der Politiker eingegangen, da die Verteilung der Social-Media-Nutzung prozentual näher in Abbildung 41 und 43 dargestellt wird. Die Informationen aus Abbildung 40 und 42 können so komplementär zu Abbildung 41 und 43 angesehen werden.

### **Westdeutschland**

Grundlegend wurden in Westdeutschland, hier Rheinland-Pfalz, 32 Abgeordnete gewählt, in Ostdeutschland, hier Sachsen, 38. Damit ist eine leicht höhere Abgeordnetenanzahl in Sachsen zu erkennen. In Abbildung 40 kann festgestellt werden, dass die meisten Abgeordneten in Rheinland-Pfalz der SPD zugehörig waren, mit insgesamt 12 Abgeordneten. Nachfolgend hat die CDU 9 Abgeordnete aufweisen können. Bündnis 90/Die Grünen und die FDP liegen mit jeweils 5 Abgeordneten beide auf dem dritten Platz. Die zweitniedrigste Anzahl an gewählten Abgeordneten gibt es bei der AFD mit 4 Abgeordneten und deutlich abgeschlagen, mit einem Abgeordneten, ist die Linke.

Anzahl Politiker je nach Social-Media und Twitter Nutzung aufgelistet | Westdeutschland

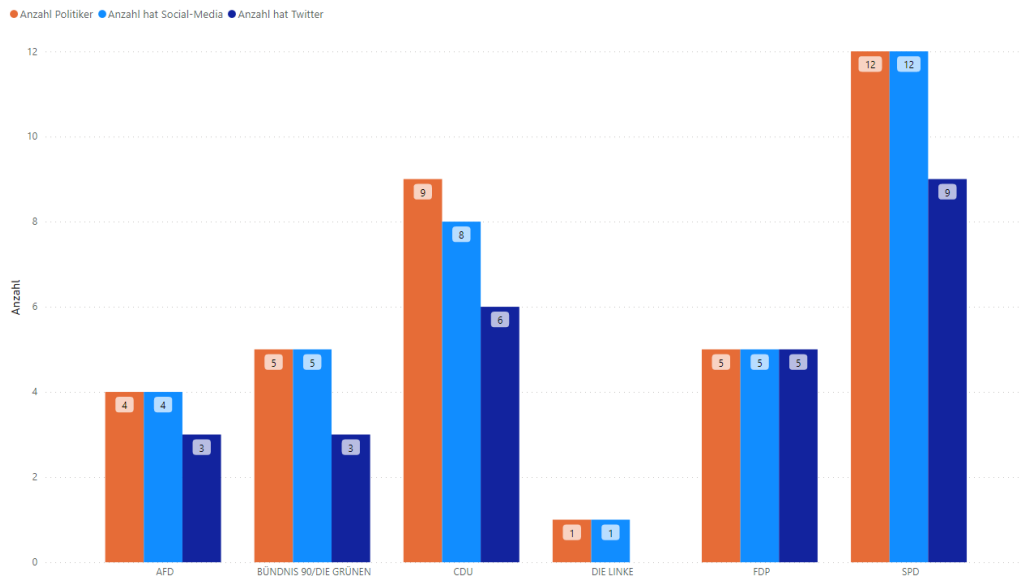


Abbildung 40: Anzahl der Politiker nach Social-Media-Kanal und Partei, Rheinland-Pfalz, Westdeutschland

In Abbildung 41 ist zu erkennen, dass außer bei der CDU, jeder Abgeordnete der anderen Parteien einen Social-Media-Kanal zur Kommunikation mit seinen Wählern betreibt. Betrachtet man die Verteilung der Twitter-Accounts, so besitzen nur in der FDP alle Abgeordneten auch diesen. Auffallend ist, dass die Politiker der restlichen Parteien zwar häufig Social-Media-Kanäle betreiben, dabei aber nur 75 % (SPD, AFD) bis hin zu 66,67 % (CDU) und 60% der Abgeordneten von Bündnis 90/Die Grünen einen Twitter-Account besitzen. Einzig die Linke hat keinen Twitter-Nutzer, allerdings wurde für diese Partei auch nur ein einzelner Abgeordneter gewählt.

Prozentuale Nutzung von Social Media bei Kommunalpolitikern in Westdeutschland

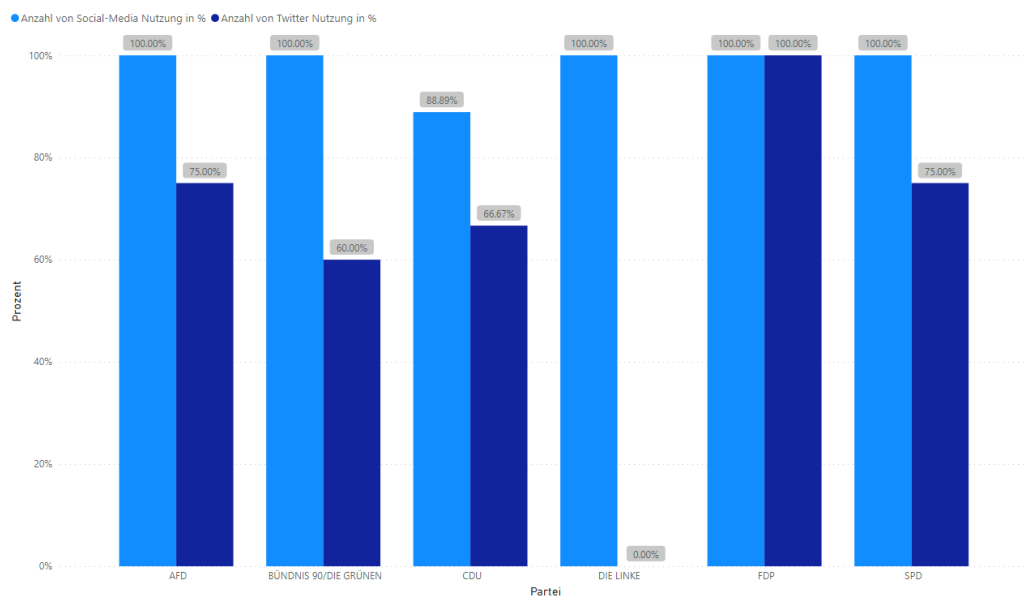


Abbildung 41: Prozentuale Nutzung von Social-Media je Partei Rheinland-Pfalz Westdeutschland

## Ostdeutschland

Betrachtet man nun die gleichen Visualisierungen für Sachsen in Ostdeutschland, so kann eine deutlich unterschiedliche Verteilung der Politiker in Abbildung 42 festgestellt werden. In Sachsen ist die Partei mit den meisten Abgeordneten die AFD mit 10 Personen, gefolgt wird diese von der SPD mit 8 Abgeordneten. Auffällig hierbei ist, dass die SPD, genau wie in Rheinland-Pfalz, Westdeutschland, auch hier mehr Abgeordnete hat als die CDU. Die CDU besitzt in Ostdeutschland 7 Abgeordnete. Die restlichen Parteien liegen in Sachsen sehr nah, in der Verteilung der Abgeordneten, beieinander. Die FDP liegt, gemessen an der Zahl der Abgeordneten, mit 5 Abgeordneten hinter der CDU. Die niedrigste Abgeordnetenanzahl haben Bündnis 90/Die Grünen und die Linke mit jeweils 4 Personen. Anmerken sollte man hierbei, dass die Linke deutlich mehr Abgeordnete besitzt als in Westdeutschland (im Vgl. dazu nur 1 Abgeordneter).

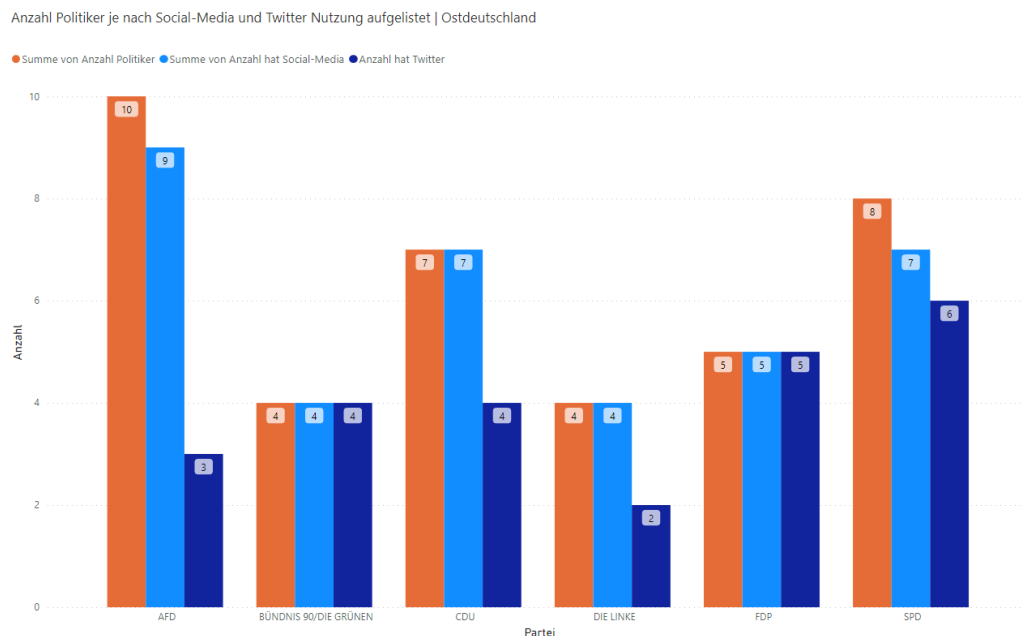


Abbildung 42: Anzahl der Politiker nach Social-Media-Kanal und Partei, Sachsen, Ostdeutschland

Die prozentualen Ergebnisse aus Abbildung 43 lassen erkennen, dass die Abgeordneten aller Parteien, bis auf die AFD und SPD, einen Social-Media-Kanal zur Kommunikation nach außen benutzen. Bei der AFD besitzen 90% der Abgeordneten einen Social-Media-Kanal, bei der SPD sind es nur 87,5%. Schaut man sich nun die Verteilung der Twitter-Kanäle an, so stellt man fest, dass es in Sachsen prozentual weniger Nutzer gibt als in Rheinland-Pfalz. Besonders hervor stechen hier Bündnis 90/Die Grünen und die FDP, bei denen jeder Abgeordnete einen Twitter-Account besitzt. Bei der SPD sind es 75% aller Abgeordneten, bei der CDU nur noch 57,14%, bei der

Linken nur noch die Hälfte, also 50% und die niedrigste Anzahl an Twitter-Accounts sind bei der AfD mit 30% aller Abgeordneten zu finden.

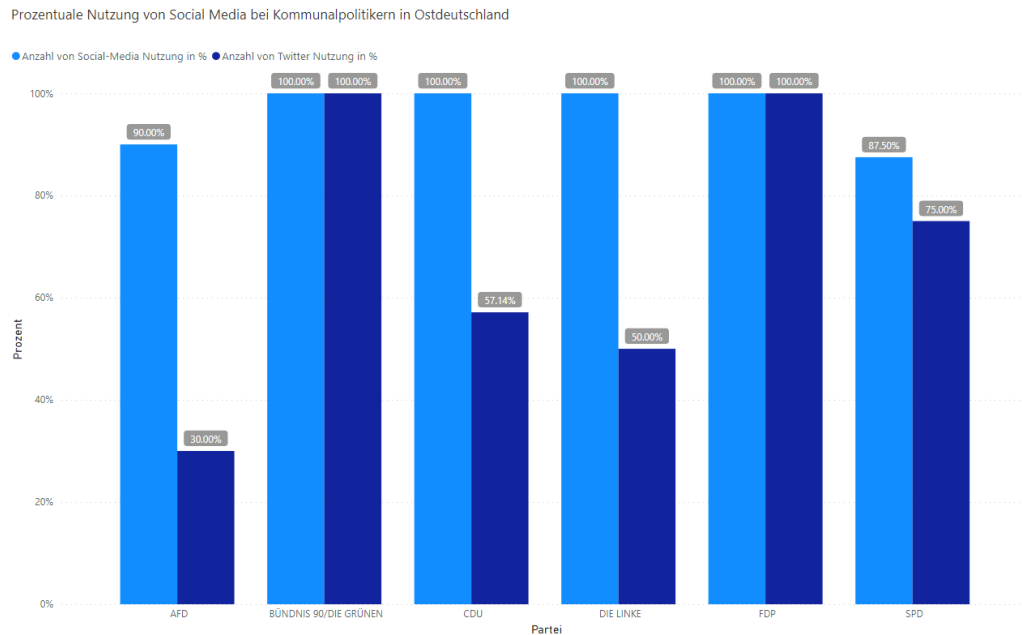


Abbildung 43: Prozentuale Nutzung von Social-Media-Kanälen je Partei, Sachsen, Ostdeutschland

Darstellung der Themenschwerpunkte der Repräsentanten der einzelnen Parteien in Ost- und Westdeutschland mittels Wordclouds

Im folgenden Abschnitt werden die Themenschwerpunkte der einzelnen Repräsentanten der unterschiedlichen Parteien, mittels einer Wordcloud, aufgeteilt in Ost- und Westdeutschland, zu Vergleichszwecken nebeneinander aufgeführt. Dabei werden die Themenschwerpunkte des jeweiligen Repräsentanten aufgeführt und anschließend die Wörter, zugehörig zum Thema, dargelegt. Die linke Seite repräsentiert die Repräsentanten aus Westdeutschland, die rechte Seite Ostdeutschland. Die Repräsentanten sind nach Parteien gruppiert.

### AfD

Vergleicht man die Themenschwerpunkte beider Repräsentanten, so fällt auf, dass beide unabhängig von ihrer Parteizugehörigkeit zur AfD, Gemeinsamkeiten im Bereich Bundesregierung (Wörter: Bundesregierung und Merkel, Bundestag) sowie Corona (Wörter: Corona, Geimpfte, infektiös) aufweisen. Unterschiedlich ist die beim Repräsentanten ausgeprägte Thematik der Migration (Wörter: Migrationspakt, Einbürgerung), China (Wort: China) und der Polizei (Wort: Polizei). Der Repräsentant der AfD in Westdeutschland thematisiert mehr die Themen Bündnis 90/Die Grünen (Wörter: Grünen,







Bundestag, Staatsregierung) häufiger erwähnt und das Thema Sport (Wort: Sport), Unterstützung (Wörter: Unterstützung) sowie Grundrechte, Respekt und Frieden (Wörter Grundrechte, Respekt, Frieden). Im Fokus der FDP in Westdeutschland steht die Freiheit (Wort: Freiheit), die Bildung (Wörter: Bildung, Schule, Kinder), die Frau (Wort: Frau), die eigene Partei (Wort: FDP) und Deutschland (Wort: Deutschland) (zu sehen in Abbildung 52 und 53).



Abbildung 52: Wordcloud, Katja Adler, FDP Hessen, Westdeutschland



Abbildung 53: Wordcloud, Philipp Hartewig, FDP Sachsen, Ostdeutschland

## Die Linke

Für die Repräsentanten der Linken in Ost- und Westdeutschland gibt es Überschneidungen bei den Themen CDU (Wort: CDU), Menschen (Wörter: Menschenrechte, Menschen) und Deutschland (Wort: Deutschland) und generell der Politik (Wort: Politik). Im Osten wird sich des Weiteren noch auf die Themen Europa (Wörter: EU, EU-Außengrenze), Rechte (Wörter: Recht, rechte, rechten) und Migration sowie Außenpolitik (Wörter: Frontex, Afghanistan, Griechenland, Schutzsuchenden, Gefahr) konzentriert. Im Westen sind häufiger die Themen Kinder (Wort: Kinder, Schulen) andere Parteien (Wort: SPD, FDP, Parteien), das Bundesland (Wort: Niedersachsen) und der Rat von Osnabrück, Abkürzung osrat (Wort: osrat) (zu sehen in Abbildung 54 und 55).

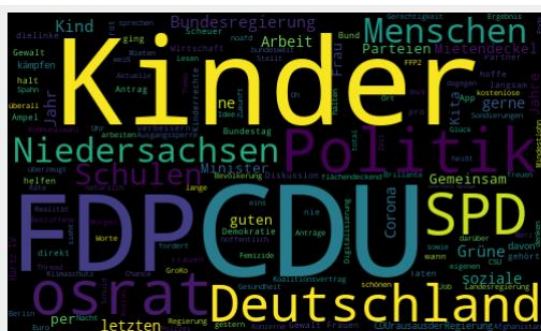


Abbildung 54: Wordcloud, Heidi Reichinnek, Die Linke Niedersachsen, Westdeutschland



Abbildung 55: Wordcloud, Clara Bünger, Die Linke Sachsen, Ostdeutschland



## 4 Zusammenfassung der Ergebnisse

### 4.1 Technische Limitationen der durchgeführten Arbeit

Bei den erarbeiteten Ergebnissen ist zu beachten, dass diese aufgrund von technischen Limitationen bei der Wahl des Textanalyseverfahrens keine 100% Abbildung der Realität gewährleisten können. Die Anzahl verfügbarer Textkorpora der deutschen Sprache für das Training von geeigneten Algorithmen im Social-Media-Bereich ist sehr limitiert. Eine weitere Einschränkung, ergebend aus der aktuellen Rechtslage, ist die Speicherung von Tweets nur durch entsprechende IDs, die auf die jeweiligen Tweets verweisen. Es dürfen keine Klartexte abgespeichert werden. Ein im Rahmen der Recherche untersuchter deutscher Textkorpora für Twitter ist SB10K, welcher im Jahre 2017 im Rahmen einer Kooperation der SpinningBytes AG (SpinningBytes, 2022) und der Zürcher Hochschule für Angewandte Wissenschaften (ZHAW, 2022) entstand und im Paper „A Twitter corpus and benchmark resources for german sentiment analysis“ (Cieliebak et al., 2017) eingeführt wurde. Dieser beinhaltet ca. 10.000 von drei Kuratoren ausgewertete Tweets mit jeweiligem Sentiment. Entsprechende Skripte zur Extraktion der genannten Tweets konnten diese jedoch nicht mehr auffinden. Dies kann geschehen, wenn die ursprünglichen Tweets oder Accounts nicht mehr existieren. Aufgrund eines fehlenden Trainings- und Testdatensatzes im geeigneten Umfang, wurden die entsprechenden Analysen in dieser Arbeit, mit einem bereits bestehenden lexikalischen Textklassifizierungsverfahren der deutschen Sprache durchgeführt. Dieses Verfahren kann, im Gegensatz zu einigen antrainierten Alternativen, nur einzelne Worte je nach ihrem Sentiment auswerten und erzielt schlechtere Ergebnisse, wenn Ironie/Sarkasmus oder doppelte Verneinungen in Aussagen analysiert werden müssen. Ein weiterer Einflussfaktor ist die geringe Datenmenge, von maximal 280 Zeichen pro Tweet, die der Algorithmus zur Verfügung hat, um das entsprechende Sentiment zu berechnen. Eine weitere technische Hürde, seitens der Twitter-API, bzw. seitens Tweepy, welches ich zur Kommunikation mit der Twitter-API aus Python heraus benutzt habe, war die unvollständige Extraktion von Tweets eines bestimmten Zeitraums. Dieses Problem trat auf, wenn die zu analysierende Person eine bisher unbekannte, größere Menge an aufeinanderfolgenden Retweets im eigenen Profil besaß, jedoch selbst nichts twitterte. Im Laufe der Erarbeitung und Recherche konnte dieses Problem durch eine Filterung aller Tweets behoben werden, sollte aber bei der Arbeit mit Tweepy beachtet werden.

## 4.2 Ergebnisse der Auswertungen

Da diese Arbeit sich auch zum Ziel gesetzt hat, mögliche Schlussfolgerungen zur Bundestagswahl des Jahres 2021 mittels der Analysen treffen zu können, wurde der Zeitraum vom 01.01.2021 bis 31.12.2021 für die Tweetsammlung ausgewählt.

### 4.2.1 Bundesebene

Die Auswertung der Politiker auf Bundestageebene mit Twitter hat ergeben, dass die Zustimmung bei Annalena Baerbock, Bündnis 90/Die Grünen, am höchsten ist. Sie verzeichneten im Mittel bei neutralen Tweets über 4000 Likes, bei negativen ca. 3000 und positiven ca. 1000. Annalena Baerbock hält die zweithöchste Retweetanzahl der Bundespolitiker. An zweiter Stelle lässt sich Dr. Alice Weidel der AfD einordnen. Sie hat im Mittel über alle Sentiments mehr als 1700 Likes je Tweet und besitzt die höchste gemittelte Retweetanzahl mit über 400. Auf dem dritten Platz befindet sich Christian Lindner der FDP mit über 1000 Likes im neutralen und negativen Sentiment und ca. 1750 im positiven. Auch wenn die Retweetanzahl verglichen wird befindet sich dieser auf dem dritten Platz. Auf dem gemeinsamen 4. Platz befinden sich Armin Laschet der CDU und Olaf Scholz der SPD. Beide verzeichnen, wenn alle Sentiments betrachtet werden, eine sehr ähnliche Like- (400-600) und Retweetanzahl (ca. 40 -70). Die geringste Zustimmung erhält Janine Wissler der Linken, mit ca. 70 – 250 Likes und ca. 7 – 35 Retweets.

Auffallend ist eine erhöhte Zustimmung der positiv eingeordneten Tweets bei den Politikern Olaf Scholz, Dr. Alice Weidel und Christian Lindner. Armin Laschet, Annalena Baerbock und Janine Wissler hingegen, erhalten bei Ihren neutralen Tweets mehr Resonanz.

Die ausgewerteten Themenbereiche der Bundespolitiker ergaben, dass sich die Wörter Corona/Pandemie, Deutschland und Menschen/Bürger generell nachweisen ließen.

Speziellere Hauptthemenbereiche, nach Politiker eingeordnet, sind:

#### **Dr. Alice Weidel (AfD)**

- AfD

#### **Janine Wissler (Die Linke)**

- Solidarität, die Linke

#### **Christian Lindner (FDP)**

- Bundesregierung, Freiheit, Alexej Nawalny

**Annalena Baerbock** (Bündnis 90/Die Grünen)

- Demokratie, Kinder, Klimaschutz

**Olaf Scholz** (SPD)

- Zukunft, Respekt, spd21 (SPD-Bundespartei 21)

**Armin Laschet** (CDU)

- Europa, Nordrhein-Westfalen

## 4.2.2 Kommunal-/Landesebene

Auf der Kommunal-/Landespolitikerebene wurde durch historisch bedingte Differenzen entschieden, Repräsentanten aus Ost- und Westdeutschland miteinander zu vergleichen. Für jede Partei wurde ein Repräsentant ausgewählt, welcher einen Twitter-Account besitzt und diesen aktiv betreibt (100+ Tweets im Jahr 2021), des Weiteren sollte der gewählte Politiker vorher noch nie in den Bundestag gewählt worden sein. Abseits dieser Limitationen wurden die Politiker per Zufallsprinzip ausgewählt.

**Repräsentanten Ostdeutschland**

In Ostdeutschland ist der Repräsentant der Linken mit über 1300 Tweets am aktivsten, gefolgt wird dieser von Bündnis 90/Die Grünen, der CDU und der FDP. Die geringste Twitteraktivität wurde beim Repräsentanten der AfD ausgemacht. Die Analyse hat des Weiteren ergeben, dass Bündnis 90/Die Grünen proportional zur Menge der abgesetzten Tweets, besonders viele positive veröffentlicht hat. Von der reinen Anzahl hat die Linke die meisten positiven Tweets (65), gefolgt von Bündnis 90/Die Grünen (47) und der SPD (36). Danach kommt die CDU (22), die AfD (6) und die FDP (5). Betrachtet man die Themenschwerpunkte, die mittels der Wordclouds ausgewertet wurden, so ist erkennbar, dass die Themen „Menschen, Sachsen, Corona/Pandemie/Impfung und Bundestagswahl 21“ die am meisten vertretenen Themen sind.

Spezielle Hauptthemenbereiche der Politiker in Sachsen und Berlin, Ostdeutschland, sind nach Politiker sortiert:

**Clara Büniger, die Linke** (Sachsen)

- EU-Außengrenze, Frontex, Afghanistan

**Philipp Hartewig, FDP** (Sachsen)

- Sport, Zukunft

**Bernhard Herrmann, Bündnis 90/Die Grünen (Sachsen)**

- Chemnitz, Klimaschutz, Grüne

**Edgar Naujok, AfD (Sachsen)**

- Migrationspakt, AfD, Zeit, Landkreis Leipzig, Polizei

**Dr. Otilie Klein, CDU (Berlin)**

- Gespräch, Mitte, mitteninMitte, Berlin (CDU-Kreisverband Berlin Mitte)

**Holger Mann, SPD (Sachsen)**

- Leipzig, SPD, Klimaschutz

**Sachsen, Ostdeutschland allgemein**

Analysiert man die Anzahl aller gewählten Abgeordneten in Sachsen, so stellt man fest, dass die AfD mit 10 Abgeordneten die meisten Bundestagszugänge aus Sachsen vermelden konnte. Am zweithäufigsten wurden Politiker der SPD (8) gewählt. Gefolgt von der CDU (7) Personen. Die FDP konnte 5 Abgeordnete dazugewinnen, Die Linke und Bündnis 90/Die Grünen nur 4. Betrachtet man die Social-Media-Nutzung der gewählten Abgeordneten, so ist zu erkennen, dass bis auf die SPD (87,5%) und die AfD (90%), in jeder anderen Partei jeder gewählte Politiker mindestens einen Social-Media-Kanal hat. In diesen Parteien besitzen jedoch nur alle Vertreter der FDP und Bündnis 90/Die Grünen auch einen Twitter-Account. Die restliche Verteilung gibt an, dass ca. 75% aller SPD-Abgeordneten einen Twitter-Account besitzen, gefolgt von ca. 57% bei der CDU, 50% bei den Linken und 30% bei der AfD.

**Repräsentanten Westdeutschland**

In Westdeutschland sind die Repräsentanten der SPD (970) und CDU (836) deutlich aktiver als die der anderen Parteien. Mit weniger als der Hälfte der Tweets der CDU, befindet sich die FDP mit 304 Tweets auf dem dritten Platz. Gefolgt von Bündnis 90/Die Grünen, die Linke und schließlich die AfD.

Bei den negativ eingeordneten Tweets befinden sich die CDU und SPD mit jeweils 12 Stück auf dem ersten Platz. Proportional viele hat jedoch auch die AfD mit 7, die zweithöchste Anzahl, obwohl sie bei allen abgesetzten Tweets mit 185 auf dem letzten Platz ist.

Bei der Verteilung der positiven Tweets ist die CDU mit 59 weit vor der zweithöchsten Anzahl 31, der SPD. Proportional zu allen verfassten Tweets gesehen, hat die Linke mit

24 Tweets besonders viele positive verfasst. Dahinter kommt die FDP mit 11, die AfD mit 5 und Bündnis 90/Die Grünen mit nur 2 Tweets.

Betrachtet man die neutralen Tweets, so ergibt sich die gleiche Reihenfolge und proportionale Verteilung wie bei allen Tweets. Themenschwerpunkte der Repräsentanten sind allgemein am häufigsten die Menschen, Corona/Pandemie, Arbeit und Kinder.

Spezielle Hauptthemenbereiche der Politiker in Rheinland-Pfalz, Hessen und Niedersachsen, Westdeutschland sind nach Politikern sortiert:

**Heidi Reichinnek, die Linke** (Niedersachsen)

- FDP, SPD, CDU, Schulen, osrtat (Rat der Stadt Osnabrück)

**Katja Adler, FDP** (Hessen)

- Freiheit, Bildung, Frau, FDP, Schulen

**Helge Limburg, Bündnis 90/Die Grünen** (Niedersachsen)

- Itnds (Landtag Niedersachsen), Gesellschaft, Familien, Solidarität

**Andreas Bleck, Afd** (Rheinland-Pfalz)

- AfD, Grünen, Baerbock, Laschet, Merkel, Fluthilfe

**Dr. Katja Leikert, CDU** (Hessen)

- Europa, EU, Hanau

**Thomas Hitschler, SPD** (Rheinland-Pfalz)

- Südpfalz, Bundestag, SPD

Betrachtet man die Anzahl aller gewählten Abgeordneten in Rheinland-Pfalz, so ist zu erkennen, dass die SPD mit 12 Abgeordneten, noch vor der CDU mit 9, am häufigsten gewählt wurde. Von der FDP und Bündnis 90/Die Grünen wurden jeweils 5 Politiker gewählt, kurz darauf kommt die AfD mit 4 Abgeordneten. Für die Linke wurde nur 1 Politiker in den Bundestag gewählt. Bei der Analyse der Social-Media-Kanäle wurde festgestellt, dass bis auf die CDU, alle Politiker der anderen Parteien mindestens einen besitzen. Auch in Westdeutschland haben alle gewählten Politiker der FDP einen Twitter-Account. Bei der SPD und AfD besitzen 75% der Abgeordneten einen Account, in der CDU ca. 67%, bei Bündnis 90/Die Grünen 60% und der Repräsentant der Linken Partei besitzt keinen Twitter-Account. Damit ist in Westdeutschland eine leicht höhere Social-Media/Twitter-Nutzung als in Ostdeutschland zu beobachten.

### 4.3 Interpretation der Ergebnisse

Schaut man sich die nun vorgestellten Ergebnisse an, so wird klar, dass Annalena Baerbock, Dr. Alice Weidel und Christian Lindner in ihrer Social-Media-Aktivität, zumindest auf Twitter, augenscheinlich erfolgreicher sind als Olaf Scholz und Armin Laschet. Janine Wissler bekommt unter den Bundestagsabgeordneten dabei die geringste Resonanz und Bestätigung durch ihre Leser. Im Falle von Olaf Scholz muss jedoch bei der Einordnung der Ergebnisse beachtet werden, dass dieser mit 826 Tweets, die mit Abstand meisten Tweets unter den analysierten Bundespolitikern veröffentlicht hat. Durch diesen Faktor wird seine reale Reichweite und Zuspruch höher liegen als die Mittelwerte vermuten lassen. Die Likes und Retweet-Anzahlen verteilten sich bei Olaf Scholz auf mehr Tweets, daher sind die einzelnen Werte niedriger. Des Weiteren kann nachgewiesen werden, dass die FDP, AfD und SPD eine sehr hohe Nennung der eigenen Parteien vorweisen können. Damit scheinen diese mehr auf die eigene Partei fokussiert zu sein, als sich mit anderen thematisch auseinanderzusetzen. Auffällig ist, dass die CDU besonders häufig Europa, bzw. die EU thematisiert. Dies könnte für eine offenere Ausrichtung für die Themen der EU stehen. Über die Parteien hinweg konnten einige Themen wie die aktuelle Coronapandemie, die häufige Nennung der Bürger und Menschen sowie der Arbeit festgestellt werden. Diese bilden damit die Hauptthemengebiete der analysierten Politiker ab. Auffällig war auch die geringe Anzahl an neu gewählten Abgeordneten der AfD in Sachsen, Ostdeutschland, die Social-Media-Kanäle besaßen und Twitter nutzten. Da die AfD in Sachsen und den anderen neuen Bundesländern in Ostdeutschland jedoch deutlich bessere Wahlergebnisse erhielt, könnte dies entweder durch das allgemeine Wahlprogramm der AfD (welches die Menschen in Ostdeutschland generell besser anspricht) erklärt werden oder durch einen verstärkten Wahlkampf vor Ort, im direkten Austausch mit den Wählern und dafür weniger über Social-Media-Kanäle.

## 5 Fazit

Möchte man nun wissen, wer von den untersuchten Politikern am effektivsten auf Social-Media ist, so ist es auf Bundestagebene Annalena Baerbock von Bündnis 90/Die Grünen, gefolgt von Dr. Alice Weidel von der AfD und Christian Lindner der FDP. Diese Bundespolitiker erhalten höhere Like- und Retweetzahlen pro veröffentlichtem Tweet. Betrachtet man die Hauptthemen der Politiker, so kann man Überschneidungen über die Parteien hinweg feststellen. Die Hauptthemen des Jahres 2021 sind zum einen die Coronapandemie und zum anderen der Bürger, welcher im Zentrum steht. Andere oft genannte Themen waren die jeweiligen Parteien der Politiker und die Bundestagswahl 2021. Die Parteien haben schlussfolgernd nach der Auswertung noch parteispezifisch unterschiedliche Themen. Der Unterschied in der Kommunikation nach außen ist aber nicht so markant, wie ursprünglich angenommen. Es gibt aber in Ost- und Westdeutschland bis heute Unterschiede im Wählerverhalten. Viele davon werden durch eine wirtschaftliche Schwäche im Osten weiter begünstigt und sind teilweise eine Auswirkung der Entscheidungen, welche unmittelbar nach der Wiedervereinigung getroffen wurden. Nach wie vor sind in Ostdeutschland Parteien, wie die Linke und die AfD, erfolgreicher als im Westen Deutschlands, in denen die SPD, die CDU und Bündnis 90/Die Grünen erfolgreicher sind.

Die Nutzung von Social-Media für Politiker und Parteien wird immer wichtiger. Die Nutzerzahlen der jeweiligen Social-Media-Plattformen in den letzten Jahren steigen, wie nachgewiesen, immer weiter an. Dies belegen auch die jeweiligen Auswertungen in den verschiedenen Kapiteln. Ob und wie groß der Einfluss von Social-Media mittels Twiternutzungsdaten auf die Bundestagswahl im Jahr 2021 ist, kann jedoch nicht sicher festgestellt werden, da die ausgewerteten, vorliegenden Daten zu gering sind, um daraus stichhaltige Schlussfolgerungen zu ziehen. Wähler nutzen unterschiedliche Informationskanäle und Wege/Erfahrungen, um zu einer Wählermeinung zu kommen. Da der Bereich Social-Media in den nächsten Jahren durch steigende Nutzerzahlen allerdings immer wichtiger wird, ist eine weitere Betrachtung und Auswertung der Art und Weise wie Politiker online interagieren und wie erfolgreich sie damit sind, auch über unterschiedliche Social-Media-Kanäle hinweg, ein spannendes und wichtiges Thema, welchem weiter nachgegangen werden sollte.

Meine Bachelorarbeit hat erste Erkenntnisse zum Bereich des Social-Media-Verhaltens von Politikern Deutschlands, im Zusammenhang der Bundestagswahlen des Jahres 2021 und speziell der Kommunal- und Landespolitikerebenen in Ost- und Westdeutschland, aufgezeigt. Aufgrund des gewählten lexikalischen Textanalyseverfahrens, mittels der Python-Programmiersprache Textblob und der geringen Anzahl und Verfügbarkeit

von deutschen Textkorpora im Bereich Social-Media, ist es sinnvoll, auf diesen Ergebnissen weiter aufzubauen und diese zu erweitern. Die Erstellung von zusätzlichen deutschen Textkorpora im Bereich-Social-Media ist somit ratsam und würde weitergehende und präzisere Textanalysen mit antrainierten Textanalyseverfahren ermöglichen. Bei den hier präsentierten Ergebnissen muss des Weiteren beachtet werden, dass diese nur einen der in Deutschland genutzten Social-Media-Kanäle abbilden und die Ergebnisse auf anderen Kanälen abweichen können. Die Analysen wurden mittels der Programmiersprache Python durchgeführt. Die Durchführung in anderen Programmiersprachen, z.B. R, sind möglich. Eine weitergehende Erforschung von anderen Social-Media-Kanälen wie, z.B. Facebook, welches am zweitmeisten in Deutschland genutzt wird (Datareportal, 2021, S. 47), könnte neue wichtige Erkenntnisse für die Forschungsgemeinde liefern.



## 6 Literaturverzeichnis

Appel O, Chicalana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. *Knowl Based Syst* S. 108–124

Böck, Matthias, Felix Köbler, Eva Anderl, Linda Le (2017) *Social-Media-Analyse – mehr als nur eine Wordcloud: HMD Best Paper Award 2016 (essentials)*, 1. Aufl. 2017, München, Deutschland: Springer Vieweg.

Bundeswahlleiter (2017) Ergebnisse Deutschland - Der Bundeswahlleiter, Der Bundeswahlleiter, Verfügbar unter: <https://www.bundeswahlleiter.de/bundestagswahlen/2017/ergebnisse/bund-99.html> [aufgerufen am 08.01.2022].

Bundeswahlleiter (2021) Ergebnisse Deutschland - Der Bundeswahlleiter. Zitiert nach Statista (2021b) Verfügbar unter: <https://de.statista.com/statistik/daten/studie/1222386/umfrage/sitze-der-parteien-im-deutschen-bundestag/> [aufgerufen am 08.02.2022].

Bundeswahlleiter (2022) Ergebnisse Deutschland, Verfügbar unter: <https://www.bundeswahlleiter.de/bundestagswahlen/2021/ergebnisse/bund-99.html> [aufgerufen am 3.02.2022]

Cambria, Erik, Dipankar Das, Sivaji Bandyopadhyay, Antonio Feraco (2018) *A Practical Guide to Sentiment Analysis (Socio-Affective Computing, Band 5)*, Softcover reprint of the original 1st ed. 2017, Singapur, Kolkata, Singapur, Indien: Springer.

Cieliebak, Mark, Jan Milan Deriu, Dominic Egger, Fatih Uzdilli, SpinningBytes AG (2017) A Twitter corpus and benchmark resources for german sentiment analysis | ZHAW digitalcollection, in: Publikationen School of Engineering, 14.12.2017, S. 45–51, Verfügbar unter: <https://digitalcollection.zhaw.ch/handle/11475/1856> [aufgerufen am 02.02.2022].

Cleve, Jürgen, Uwe Lämmel (2020) *Data Mining (De Gruyter Studium)*, 3. Auflage, Wismar, Deutschland: De Gruyter Oldenbourg.

Damascelli, A.T., Martelli, A. (2003) *Corpus linguistics and computational linguistics: an overview with special reference to English*. Turin: Celid

Dpa (2017) Twitter verdoppelt maximale Länge der Tweets auf 280 Zeichen, heise online, Verfügbar unter: <https://www.heise.de/newsticker/meldung/Twitter-verdoppelt-maximale-Laenge-der-Tweets-auf-280-Zeichen-3883047.html> [aufgerufen am 18.12.2021].

Farra N, Challita E, Assi RA, Hajj H (2010) Sentence-level and document-level sentiment mining for arabic texts. In: *Proceedings of IEEE international conference on data mining workshops*, S. 1114–1119

Ferber, R. (2003) *Information Retrieval – Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web*. Heidelberg: dpunkt.verlag

First Versions (2015) Twitter, Verfügbar unter: <https://www.firstversions.com/2015/07/twitter.html> [aufgerufen am 28.11.2021].

Gao W, Li S, Xue Y, Wang M, Zhou G (2014) Semi-supervised sentiment classification with self-training on feature subspaces. In: Su X, He T (eds) *Chinese Lexical Semantics. CLSW 2014. Lecture notes in computer science*, vol 8922. Springer, Cham, S. 231–239

Grefenstette G, Qu Y, Shanahan JG, Evans DA (2004) Coupling niche browsers and affect analysis for an opinion mining application. In: *Proceedings of RIAO '04*

Coupling approaches, coupling media and coupling languages for information retrieval, S. 186–194

Harris, Charles (2020) Array programming with NumPy, Nature, Verfügbar unter: <https://www.nature.com/articles/s41586-020-2649-2> [aufgerufen am 30.12.2021].

He Y, Zhou D (2011) Self-training from labeled features for sentiment analysis. Inf Process Manag 47, S. 606–616

Hemmatian, Fatemeh, Mohammad Karim Sohrabi (2017) A survey on classification techniques for opinion mining and sentiment analysis, in: *Artificial Intelligence Review*, Bd. 52, Nr. 3, S. 1495–1545, Verfügbar unter: <https://link.springer.com/article/10.1007/s10462-017-9599-6>

Hippner, Hajo, René Rentzmann (2006) Text Mining, in: Informatik-Spektrum, Bd. 29, Nr. 4, S. 287–290, Verfügbar unter: doi:10.1007/s00287-006-0091-y.

Infratest dimap (2021a), [www.infratest-dimap.de](http://www.infratest-dimap.de). Zitiert nach Statista (2021c). Ergebnis der Bundestagswahl in Westdeutschland 2021, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/1266643/umfrage/ergebnis-der-bundestagswahl-in-westdeutschland/> [aufgerufen am 08.02.2022].

Infratest dimap (2021b), [www.infratest-dimap.de](http://www.infratest-dimap.de). Zitiert nach Statista (2021d). Ergebnis der Bundestagswahl in Ostdeutschland 2021, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/1266647/umfrage/ergebnis-der-bundestagswahl-in-ostdeutschland/> [aufgerufen am 08.02.2022].

Infratest dimap (2021c), [www.infratest-dimap.de](http://www.infratest-dimap.de). Zitiert nach Statista (2021e). Wahlverhalten bei der Bundestagswahl 2021 nach Alter, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/1257097/umfrage/wahlverhalten-bei-der-bundestagswahl-nach-alter/> [aufgerufen am 08.02.2022].

Joachims T. (1998) Text categorization with support vector machines: learning with many relevant features, Springer, Berlin

Kaplan, A. M., & Haenlein, M. (2010) Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. Verfügbar unter: <https://doi.org/10.1016/j.bushor.2009.09.003>

Datareportal (2021) Digital in Germany: All the Statistics You Need in 2021, DataReportal – Global Digital Insights, Verfügbar unter: <https://datareportal.com/reports/digital-2021-germany> [aufgerufen am 11.02.2022].

Kollmann, Tobias, Holger Schmidt (2016) *Deutschland 4.0: Wie die Digitale Transformation gelingt*, 1. Aufl. 2016, Essen, Deutschland: Springer Gabler.

Li G, Liu F (2014) Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. Appl Intell 40, S. 441–452

Liu B (2007) Web data mining: exploring hyperlinks, contents, and usage data. Springer, Berlin

Liu B (2012) Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. Morgan & Calypool Publishers, S. 1–167. Verfügbar unter: <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>

Liu B, Zhang L (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) Mining text data. Springer, Boston, MA, S. 415–463

Liu, Bing (2012) Sentiment analysis and opinion mining, Morgan & Claypool Publishers, San Rafael

- Lobin, H. (2004) Textauszeichnung und Dokumentgrammatiken. In: Lobin, H., Lemnitzer, L. (Hrsg.) *Texttechnologie – Perspektiven und Anwendungen*. Tübingen: Stauffenburg, S. 51–82
- Manning, C.D., Schütze, H. (2003) *Foundations of Statistical Natural Language Processing*. Cambridge (MA): Massachusetts Institute of Technology
- Marcheggiani D., Täckström O., Esuli A., Sebastiani F. (2014) Hierarchical multi-label conditional random fields for aspect-oriented opinion mining. In: de Rijke M et al (eds) *Advances in information retrieval. ECIR 2014. Lecture notes in computer science*, vol 8416. Springer, Cham, S. 273–285
- Matplotlib — Visualization with Python (2021) Matplotlib - Visualization with Python, Verfügbar unter: <https://matplotlib.org/> [aufgerufen am 30.12.2021].
- Medienpädagogischer Forschungsverbund Südwest (2021), JIM-Studie 2021, Verfügbar unter: [https://www.mpfs.de/fileadmin/files/Studien/JIM/2021/JIM-Studie\\_2021\\_barrierefrei.pdf](https://www.mpfs.de/fileadmin/files/Studien/JIM/2021/JIM-Studie_2021_barrierefrei.pdf) [aufgerufen am 08.02.2022].
- Microsoft (2022a), Datenvisualisierung | Microsoft Power BI Verfügbar unter: <https://powerbi.microsoft.com/de-de/> [aufgerufen am 10.01.2022].
- Microsoft (2022b), Excel-Hilfe und -Learning - Microsoft-Support, Microsoft, [Verfügbar unter: <https://support.microsoft.com/de-de/excel> [aufgerufen am 10.01.2022].
- Moraes R, Valiati JF, Neto WPG (2013) Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert Syst Appl* 40, S. 621–633
- NLTK (2021) NLTK: Natural Language Toolkit, Verfügbar unter: <https://www.nltk.org/> [aufgerufen am 01.01.2022].
- Pandas - Python Data Analysis Library (2021) pandas - Python Data Analysis Library, Verfügbar unter: <https://pandas.pydata.org/about/index.html> [aufgerufen am 30.12.2021].
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, vol 10, S. 79–86
- Paßmann, Johannes (2018) *Die soziale Logik des Likes: Eine Twitter-Ethnografie*, 1. Aufl., Siegen, Deutschland: Campus Verlag.
- Paßmann, Johannes, Campus Verlag (2018) *Die soziale Logik des Likes: Eine Twitter-Ethnografie*, Frankfurt am Main, Deutschland: Campus Verlag.
- Petz, Gerald (2018) *Opinion Mining im Web 2.0: Ansätze, Methoden, Vorgehensmodell*, 1. Aufl. 2019, Linz, Österreich: Springer Gabler.
- Poria S, Cambria E, Winterstein G, HuangGB (2014) Sentic patterns: dependency-based rules for concept-level sentiment analysis. *Knowl Based Syst* S. 69:45–63
- Prof. Tom Thieme (2016) *25 Jahre deutsche Einheit – Kontinuität und Wandel in Ost- und Westdeutschland*, Universitätsverlag Chemnitz
- Rajman, M., Vesely, (2004) From Text to Knowledge: Document Processing and Visualization: a Text Mining Approach. In: Sirmakessis, S. (Hrsg.) *Text Mining and its Applications – Results of the NEMIS Launch Conference*. Berlin u.a.: Springer
- Schmidt, Jan-Hinrik (2017) *Social Media (Medienwissen kompakt)*, 2. Aufl. 2018, aktualisierte und erw., Hamburg, Deutschland: Springer VS.

Schweidel, D. A., & Moe, W. W. (2014) Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research*, 51 (August), 387–402.doi:10.1509/jmr.12.0424.

Seven.One Media GmbH (2021), <https://www.sevenonemedia.de/>. Zitiert nach Statista (2021f). Tägliche Nutzungsdauer ausgewählter Medien in Deutschland bis 2021, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/992180/umfrage/taegliche-nutzungsdauer-von-medien-in-deutschland/> [aufgerufen am 08.02.2022].

Sirmakessis, S. (Hrsg.) (2004) Text Mining and its Applications – Results of the NEMIS Launch Conference. Berlin, Heidelberg

SPARTA (2021) Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr, Verfügbar unter: <https://dtecbw.de/sparta> [aufgerufen am 28.11.2021].

SPD – Olaf Scholz (2022) Über mich, Verfügbar unter: <https://olaf-scholz.spd.de/ueber-mich/> [aufgerufen am 08.12.2022].

SpinningBytes (2022) SpinningBytes – Text Analytics & Speech Processing, Verfügbar unter: <https://spinningbytes.com/> [aufgerufen am 02.02.2022].

Statista (2021a) Monatlich aktive Nutzer (MAU) von Twitter weltweit bis zum 1. Quartal 2019, Statista, Verfügbar unter: <https://de.statista.com/statistik/daten/studie/232401/umfrage/monatlich-aktive-nutzer-von-twitter-weltweit-zeitreihe/> [aufgerufen am 05.11.2021].

Statistisches Bundesamt (2021), Daten aus dem Gemeindeverzeichnis – Bundesländer mit Hauptstädten nach Fläche, Bevölkerung und Bevölkerungsdichte 31.12.2020, Verfügbar unter: [https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/02-bundeslaender.xlsx?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/02-bundeslaender.xlsx?__blob=publicationFile) [aufgerufen am 17.01.2022].

Steiner, René (2021) Grundkurs Relationale Datenbanken: Einführung in die Praxis der Datenbankentwicklung für Ausbildung, Studium und IT-Beruf, 10., akt. Aufl. 2021, Wiesbaden, Deutschland: Springer Vieweg.

Sullivan, D (2001) Document Warehousing and Text Mining. New York u.a.: John Wiley & Sons

TextBlob: Simplified Text Processing (2021) TextBlob 0.16.0 documentation, Verfügbar unter: <https://textblob.readthedocs.io/en/dev/> [aufgerufen am 30.12.2021].

Tweepy (2021) Tweepy, Verfügbar unter: <https://www.tweepy.org/> [aufgerufen am 30.12.2021].

Twitter (2019) Twitter Q1 2019 Letter to Shareholders. Zitiert nach Statista, Verfügbar unter: <https://de-statista-com.ezproxy2.hsrw.eu/statistik/daten/studie/232401/umfrage/monatlich-aktive-nutzer-von-twitter-weltweit-zeitreihe/> [aufgerufen am 19.12.2021].

Twitter (2021) Twitter: So markierst du einen Tweet mit „Gefällt mir“, Verfügbar unter: <https://help.twitter.com/de/using-twitter/liking-tweets-and-moments#> [aufgerufen am 18.12.2021].

Ule, T., Hinrichs, E. (2004) Linguistische Annotation., In: Lobin, H., Lemnitzer, L. (Hrsg.) Texttechnologie – Perspektiven und Anwendungen. Tübingen: Stauffenburg, S. 217–243

Wauschkuhn, O. (1996) Ein Werkzeug zur partiellen syntaktischen Analyse deutscher Textkorpora. In: Gibbon, D. (Hrsg.) Natural Language Processing and Speech

Technology – Results of the 3rd KONVENS Conference, Bielefeld. Berlin u.a.: Mouton de Gruyter, S. 357–368

Wayback Machine (2013) Retweeten eines Tweets, Verfügbar unter: <https://web.archive.org/web/20131230233624/https://support.twitter.com/groups/52-connect/topics/211-tweeting/articles/20170060-retweeten-eines-tweets> [aufgerufen am 18.12.2021].

Webb, Geoffrey I. (2017) Naïve Bayes, in: *Encyclopedia of Machine Learning and Data Mining*, S. 895–896, Verfügbar unter: doi:10.1007/978-1-4899-7687-1\_581.

Wiedenbeck, M., & Züll, C. (2001) Klassifikation mit Clusteranalyse: grundlegende Techniken hierarchischer und Kmeans-Verfahren. (GESIS-How-to, 10). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-.

Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase level sentiment analysis. In: Proceedings of HLT/EMNLP-05

Yang B, Cardie C (2014) Context-aware learning for sentence-level sentiment analysis with posterior regularization. In: ACL, no 1, S. 325–335

ZHAW Zurich University of Applied Sciences (2022) Welcome to the ZHAW, Verfügbar unter: <https://www.zhaw.ch/en/university/> [aufgerufen am 02.02.2022].

Zhuang L, Jing F, Zhu XY (2006) Movie review mining and summarization. In: Proceedings of the 15th ACM international conference on information and knowledge management, S. 43–50

Zimmermann M, Ntoutsis E, Spiliopoulou M (2014) A semi-supervised self-adaptive classifier over opinionated streams. In: 2014 IEEE international conference on data mining workshop (ICDMW). IEEE, S. 425–432

## 7 Anhang

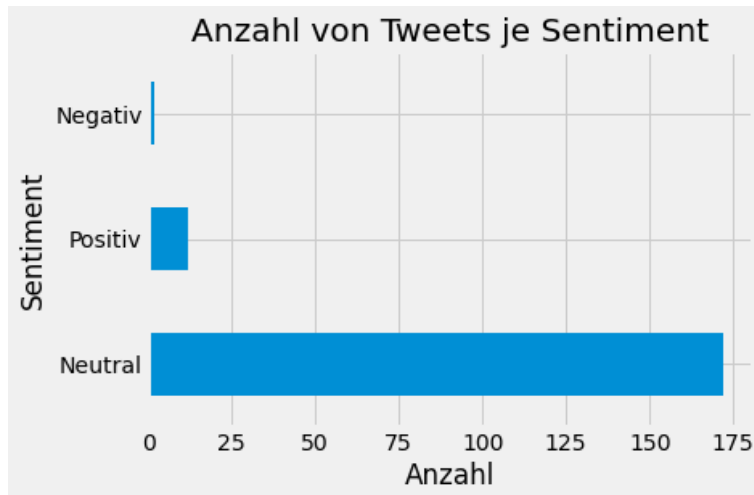
### 7.1 Anhangsverzeichnis

Anhang 1: Visualisierungen Bundestagebene .....	72
---	----

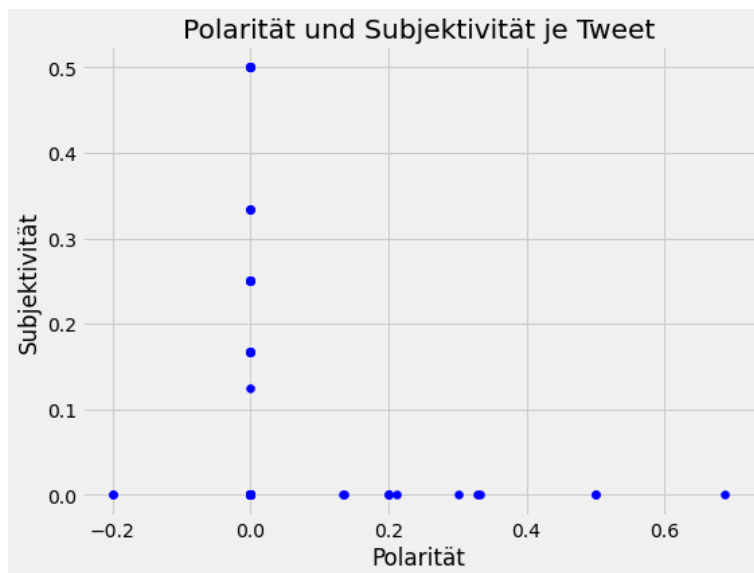
## 7.2 Visualisierungen Bundestagebene

Im Folgenden die Visualisierungen zu den einzelnen Bundestagsabgeordneten, die im Hauptteil zusammengefasst wurden.

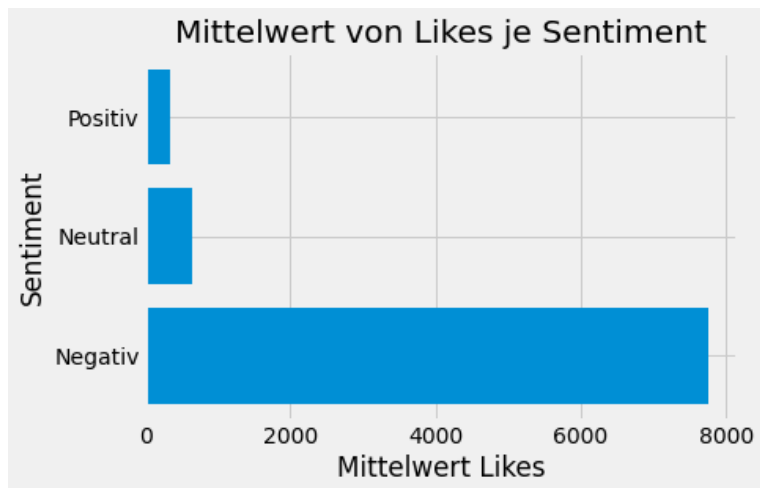
### CDU, Armin Laschet



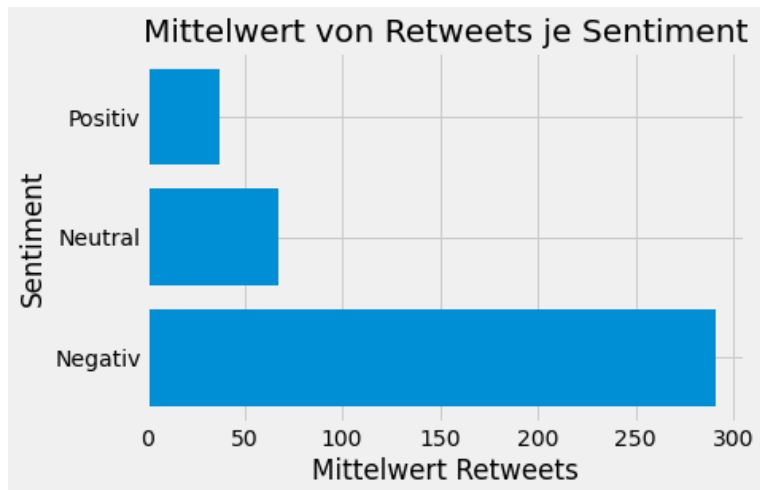
*Sentiment Verteilung und Anzahl Tweets, Armin Laschet, CDU*



*Verteilung Polarität und Subjektivität von Tweets, Armin Laschet, CDU*

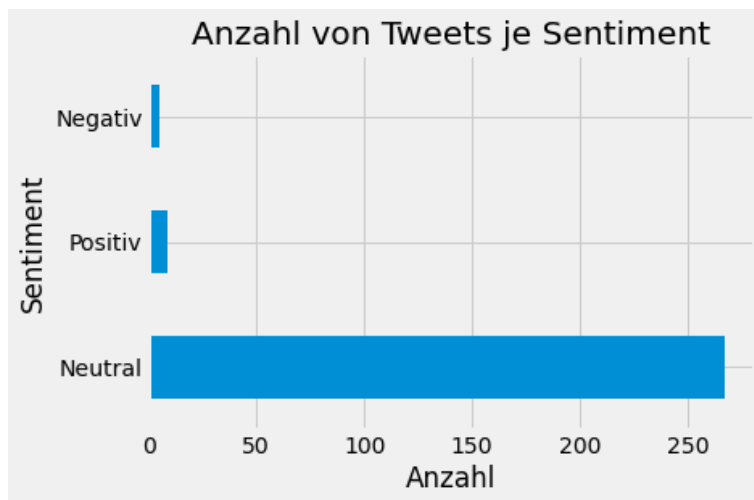
**CDU, Armin Laschet mit negativem Tweet (15.000+ Likes)**

Anzahl Tweets je Sentiment (mit negativem Tweet 15.000+ Likes), Armin Laschet, CDU

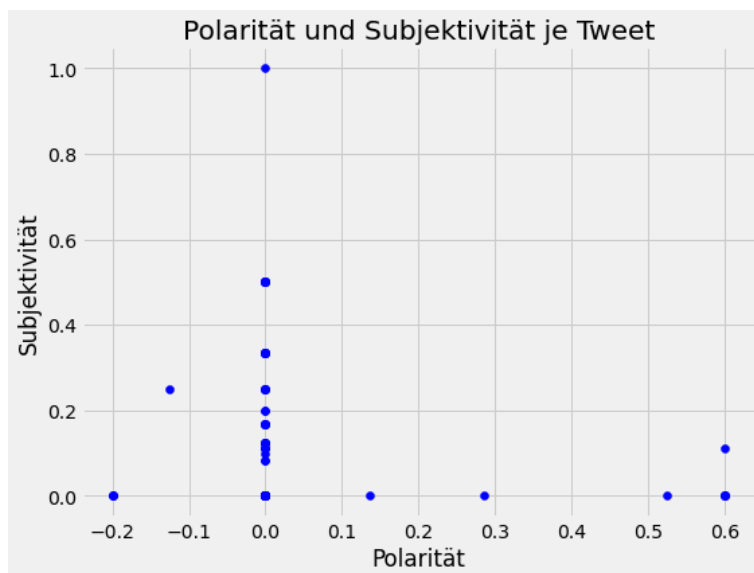


Mittelwert von Retweets je Sentiment (mit negativem Tweet 15.000+ Likes), Armin Laschet, CDU



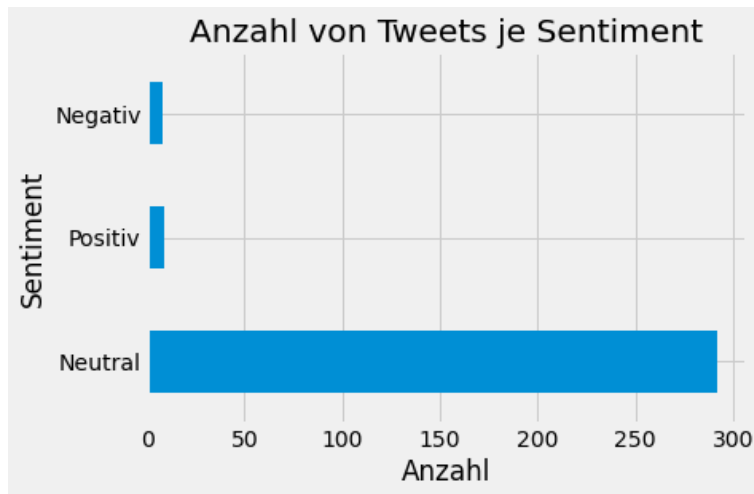
**Bündnis 90/ Die Grünen, Annalena Baerbock**

*Sentiment Verteilung und Anzahl Tweets, Annalena Baerbock, Bündnis 90/Die Grünen*

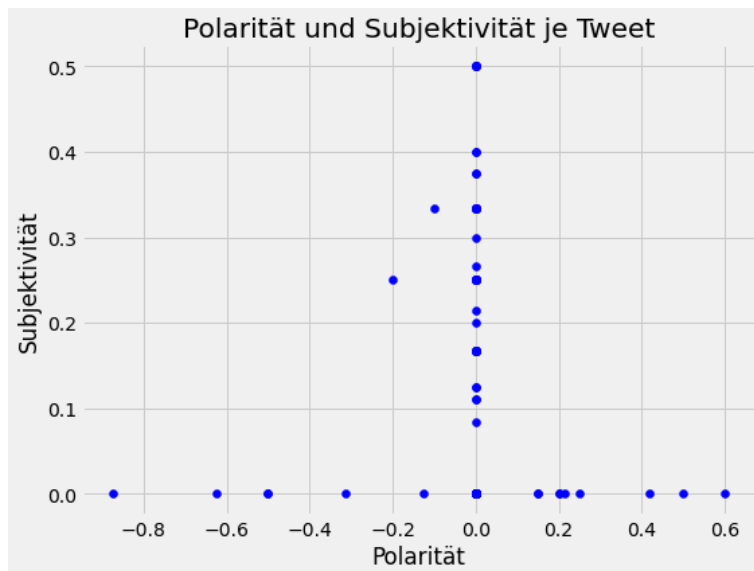


*Verteilung Polarität und Subjektivität von Tweets, Annalena Baerbock, Bündnis 90/Die Grünen*



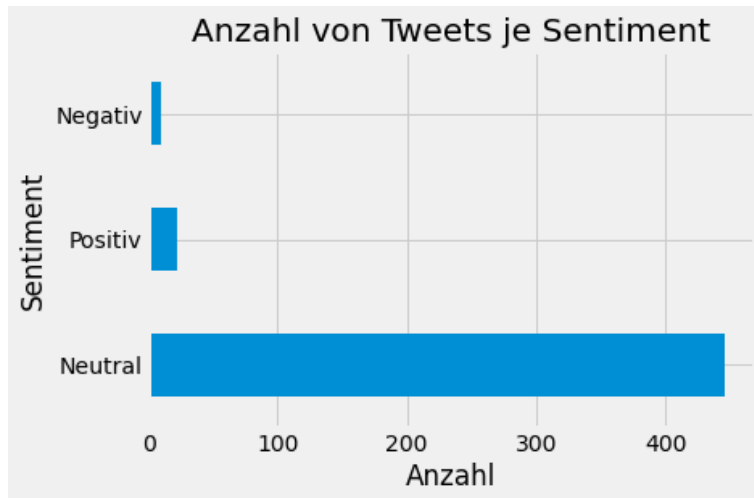
**AfD, Dr. Alice Weidel**

*Sentiment Verteilung und Anzahl Tweets, Dr. Alice Weidel, AfD*

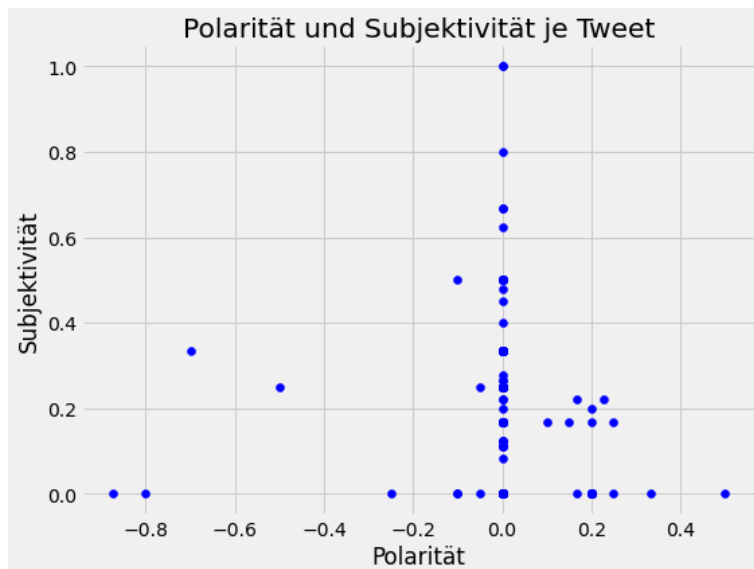


*Verteilung Polarität und Subjektivität von Tweets, Dr. Alice Weidel, AfD*

## Die Linke, Janine Wissler



Sentiment Verteilung und Anzahl Tweets, Janine Wissler, Die Linke



Verteilung Polarität und Subjektivität von Tweets, Janine Wissler, Die Linke