Transformer-Based Analysis of Sentiment Towards German Political Parties on Twitter During the 2021 Election Year

Nils Constantin Hellwig
Media Informatics Group
University of Regensburg
Regensburg, Germany
nils-constantin.hellwig@student.ur.de

Markus Bink
Media Informatics Group
University of Regensburg
Regensburg, Germany
markus.bink@student.ur.de

Thomas Schmidt
Media Informatics Group
University of Regensburg
Regensburg, Germany
thomas.schmidt@ur.de

Jakob Fehle
Media Informatics Group
University of Regensburg
Regensburg, Germany
jakob.fehle@ur.de

Christian Wolff
Media Informatics Group
University of Regensburg
Regensburg, Germany
christian.wolff@ur.de

Abstract

Twitter has become an important platform for political discussions among both politicians and the public and was extensively used during the 2021 federal election in Germany. Previous research examined the sentiment of the major political actors during that election on Twitter, but it remains unclear how the German public responded to them on Twitter in terms of sentiment. We analyzed a corpus of 713,742 tweets mentioning the Twitter handle of 89 of the most important party and politician accounts. We annotated a subset of 2,000 of these tweets regarding their sentiment and used this and other annotated corpora to implement and evaluate sentiment analysis algorithms based on single-label classification (positive, negative and neutral). We achieved best results with the German BERT model *gbert-large* using a combination of our annotated corpus and a previously annotated corpus from the same context as training material. This model achieves an average accuracy of 81.8% in a 5x5 cross-validation setting. Applying sentiment analysis on the overall corpus revealed that the majority of the tweets expressed negative sentiments. We investigated sentiment developments per party and show that sentiment was driven by significant events such as the implementation of stricter COVID-19 regulations.

1 Introduction

In 2021, the 20th German federal election took place, with the reigning chancellor Angela Merkel not running again after 16 years in office. After the election, Angela Merkel's party, the Christian Democratic Union (CDU), was no longer part of the government and a coalition was formed consisting of the Social Democratic Party (SPD), the Green Party (BÜNDNIS 90/DIE GRÜNEN), and the Free Democratic Party (Liberals, FDP). According to the opinion polling institute *Infratest Dimap*, a strong change in the political mood in the form of voting intention could be observed among voters during the election year.[1] Due to ongoing restrictions in the wake of the pandemic, campaigning by the respective parties on social media platforms like Twitter[2] played a special role in this election. Twitter is one of the most popular social media platforms and a micro-blogging platform where users can send out short posts ("tweets") which can then be viewed by other users. Tweets are limited to 280 characters (as of January 2023) and may also contain images, videos, links or hashtags, i.e. keywords marked with a "#"-sign. It is possible to mention other users in tweets by using their Twitter handle (e.g. @OlafScholz for the current German chancellor's account).

Twitter has become a popular platform for all sorts of analysis in Natural Language Processing (NLP) and Computational Social Science (CSS) including sentiment analysis. Sentiment analysis, also known as opinion mining, is the computational method to predict the sentiment, attitude, or opinion of media, predominantly text (Liu, 2020) and has major application areas in the analysis of social media (Schmidt et al., 2020), online reviews (Fehle

[1] https://www.infratest-dimap.de/umfragen-analysen/bundesweit/sonntagsfrage

[2] As of July 2023, Twitter has been rebranded as X. However, we will use the name "Twitter" in this paper since the data was acquired before the rebranding and "Twitter" is still a common reference for the platform.

et al., 2023), healthcare NLP (Moßburger et al., 2020) or narrative texts (Schmidt and Burghardt, 2018). The method has also been used extensively in the political context on Twitter to quantify both public sentiment towards political parties and actors (Agarwal et al., 2018; Yaqub et al., 2020), predict election results (Ibrahim et al., 2015; Ramteke et al., 2016), and to describe and relate sentiment of parties with one another (Tumasjan et al., 2011; Caetano et al., 2018). Previous research analyzed the tweets of major political accounts during the 2021 federal election in Germany and identified, among other things, a tendency towards negativity by opposition parties and significant sentiment changes before and after election day (Schmidt et al., 2022).

We build on this research but shift the focus from the political actors to the general public. In this paper, we perform sentiment analysis to analyze how the most important German political parties and their politicians were perceived on Twitter during the 2021 federal election campaign year. We have built a corpus of 713,742 tweets that were posted throughout the election year 2021 and that mention a selection of 89 political party accounts and politicians from the major German parties using their Twitter handle. Our research questions are as follows:

- How does the sentiment of the tweets differ comparing the major parties and comparing opposition and government parties?
- How does the sentiment expressed in tweets change over the course of the election year?
- How does the sentiment of tweets from political parties differ compared to tweets from users mentioning accounts of those parties?

Our main contributions are as follows:

- Acquisition and preparation of a corpus consisting of 713,742 tweets mentioning (using @-sign) 89 Twitter accounts by the major political German parties.
- Annotation of sentiment for a sub-corpus of 2,000 tweets.
- A fine-tuned and optimized German BERT model using annotations as training material.
- The analysis of classification results on the entire corpus focused on the proposed research questions.

Although sentiment analysis of the tweets of political actors during the 2021 German Federal Election has been already explored (Schmidt et al., 2022), to the best of our knowledge, no prior work has investigated citizens' sentiment during this election.

2 Related Work

2.1 Methods for Sentiment Analysis

Previous sentiment analysis research on Twitter has employed diverse approaches, ranging from lexicon-based methods (Elbagir and Yang, 2019; Hutto and Gilbert, 2014) to machine learning approaches like support vector machines (Awwalu et al., 2019; Xia et al., 2021), word embeddings (Lilleberg et al., 2015; Joulin et al., 2017) or neural networks (Zhang et al., 2018; Minaee et al., 2021; Xia et al., 2021). However, transformer-based models such as BERT (Devlin et al., 2019) and ELECTRA (Clark et al., 2020) which are trained on huge amounts of unlabeled textual data are currently considered state-of-the-art in a variety of NLP tasks including sentiment analysis (Dang et al., 2020; Qiu et al., 2020; Schmidt et al., 2021a). BERT-based models are available for many languages, and there are versions that have been fine-tuned on specific domains or languages. For example, for the German language, *deepset*[3] published large models that are trained on over 160 GB of German texts (Chan et al., 2020). In the context of Twitter, there are also some BERT-based models such as *BERTweet* (Nguyen et al., 2020) and *TwHIN-BERT* (Zhang et al., 2023) that have been fine-tuned on English tweets. In the field of political sentiment analysis, transformer-based models usually outperform lexicon-based methods and traditional machine learning methods (Chintalapudi et al., 2021; Fehle et al., 2021; Schmidt et al., 2022). Thus, we will focus on this approach for the implementation of our sentiment analysis.

2.2 Sentiment Analysis in the Context of Twitter for Political Research

Analyzing the sentiment of politicians' or political party tweets has been shown to accurately reflect the political orientation of these politicians or political parties. Tumasjan et al. (2011) found that the party sentiment profiles corresponded to how similar their political views between parties were.

[3] https://www.deepset.ai/

Additionally, politicians from opposing parties expressed opposing sentiments. Moreover, tweets during the 2016 American presidential election of both users and political actors have been used to identify homophily, i.e. “the tendency for individuals to interact with similar others”(Fu et al., 2012; Caetano et al., 2018). More recently, Schmidt et al. (2022) showed that in the 2021 German Federal election, the sentiment expressed in tweets of major parties was largely negative. Additionally, governing parties expressed more positive sentiments compared to those in the opposition.

Tweets have also been used to localize public opinion towards political actors in elections (Agarwal et al., 2018; Yaqub et al., 2020). Using both, geospatial data and sentiment analysis of tweets, Agarwal et al. (2018) have shown how political actors were perceived across the globe in the context of the EU Referendum regarding whether the UK should leave or stay in the EU. Likewise, Yaqub et al. (2020) evaluated the similarity between the sentiment of location-based tweets and on-ground public opinion and show that it corroborates with the election result. Similarly, Chaudhry et al. (2021) analyzed Twitter sentiment before, during, and after the 2020 US election on a state level. They find that the sentiment corresponded to a large degree with the final election results. Ali et al. (2022) investigated the sentiment expressed in tweets about Joe Biden and Donald Trump in the lead up to and aftermath of the 2020 US presidential election. Their findings indicate that following the election outcome, there was an increase in positive sentiment towards the winner Joe Biden. Using the public citizen sentiment of tweets regarding political candidates has also been shown to be useful in predicting the outcome of elections (Ibrahim et al., 2015; Ramteke et al., 2016).

3 Methodology

3.1 Data Acquisition

In order to capture the sentiment towards political actors on Twitter, we collected tweets that used the Twitter handles (using the @-sign) of a selection of accounts of parties represented in the Bundestag. The seven parties in the Bundestag were taken into account (whereby these were the same parties before and after the election). For each party, the ten politicians' accounts with the most followers were considered (as of January 2022). In addition, tweets were collected that mentioned the three official party accounts with the most followers of the seven parties. Since the parties CDU and CSU form a parliamentary group and the CSU represents the state of Bavaria only, these parties were considered as a single party in the following analysis. Thus, the party accounts of CDU and CSU were summarized to four accounts. In total, 89 accounts were included in the analysis, which are the same as those considered by Schmidt et al. (2022) to enable direct comparisons between the sentiment expressed by political actors and the sentiment of public citizens towards them in the discussion. A full list of the accounts can be found in tables 6 and 7 in the appendix.

For the collection of tweets, we used *Twint*[4], a Python library that allows downloading large amounts of tweets. For each account, tweets were collected for two random days in each month of 2021. For one day, all tweets that mentioned the account with an @-sign were scraped. We have chosen these selection criteria due to resource and API limitations we would encounter when working with all tweets mentioning the 89 accounts for this year ($\sim$ 11 million tweets). We argue that the acquired corpus is still appropriate in size and representative in the context of our research goals.

Subsequently, tweets that did not have a German language code were filtered, as well as tweets in which the account under consideration mentioned itself. After filtering, the final corpus consists of 707,241 tweets and over 22 million tokens in total (see table 1 for general statistics of the corpus). The accounts of the parties SPD and CDU/CSU, which formed the government until the federal election, were mentioned in far more tweets and the party DIE LINKE the least compared to the other parties.

3.2 Data Annotation

We annotated a subset of randomly selected 2,000 tweets in order to train a machine learning model. The proportion of tweets related to a party in the annotated subset corresponded to the proportion in the entire corpus. The tweets were annotated independently of each other by five native speakers who were students or research assistants. Annotators received an annotation manual and a guided instruction session. Each tweet of the annotation subset should be assigned to one of the following sentiment labels by the annotators:

[4] https://github.com/twintproject/twint, https://github.com/kevctae/twint

Mentioned Party	Political Orientation	Pre-Election	Post Election	# Tweets	%	# Tokens	avg. Tweet Length
SPD	center left	government	government	228,415	32.3	7,153,549	31.32
CDU/CSU	center right	government	opposition	227,683	32.2	7,097,145	31.17
DIE GRÜNEN	left, ecological	opposition	government	73,261	10.4	2,408,946	32.88
FDP	liberal	opposition	government	79,815	11.3	2,607,610	32.67
AfD	far right	opposition	opposition	57,572	8.1	1,636,144	28.42
DIE LINKE	far left	opposition	opposition	40,495	5.7	1,340,331	33.10
Total	-	-	-	707,241	100	22,243,725	31.45

Table 1: General corpus statistics.

1. **positive**: Tweet has a predominantly positive connotation.
2. **negative**: Tweet has a predominantly negative connotation.
3. **neutral**: Tweet has a neutral sentiment tone.
4. **mixed**: Tweet contains positive and negative elements, with no predominant tendency towards positive/negative connotation.

Examples of annotations are shown in table 4 in the appendix. We acquired three annotations per tweet. Fleiss‘ κ and Krippendorff's α were calculated to measure the inter-rater agreement. Both Fleiss‘ κ and Krippendorff's α are 0.61; percentage-wise agreement is on average 66%. These values point towards substantial agreement according to Landis and Koch (1977).

Annotation	# Tweets	Proportion
positive	**120**	6,00%
negative	**976**	48,80%
neutral	**777**	38,85%
mixed	87	4,35%
no majority	40	2,00%

Table 2: Distribution of the sentiment classes of the annotated subset.

We assigned each tweet the majority annotation class and removed all tweets with no majority or mixed as majority annotation class since we perform sentiment analysis on a three class setting (neutral, positive, negative). The annotated corpus consists of 1,873 tweets after this filtering. The distribution of the majority labels for the annotated tweets is shown in table 2. The majority of tweets were annotated as negative (48.8%) while only few tweets are annotated as positive (6%).

3.3 Sentiment Analysis Model Training

Since large language models such as BERT are considered state-of-the-art in text classification, we decided to use *gbert-large* by *deepset*, a pre-trained model based on the BERT architecture (Chan et al., 2020) and one of the largest German language transformer-based models, as the base model. It also proved to be the best classification model in a similar setting (Schmidt et al., 2022). The model was loaded and implemented via *Hugging Face's*[5] model hub and fine-tuned for the downstream task of single-label classification on tweets with the classes: negative, positive and neutral. We used three different data sets for this fine-tuning process: (1) our 1,873 annotated tweets, (2) the 1,785 annotated tweets by Schmidt et al. (2022) which consists of tweets by politicians of the same election context and (3) the *GermEval 2017* dataset (Wojatzki et al., 2017). *GermEval 2017* consists of German sentiment-annotated posts from the field of customer feedback (Wojatzki et al., 2017) and is one of the most popular training corpus for sentiment analysis in German. We used the 26,209 annotated documents, referred to as the "main dataset" by Wojatzki et al. (2017). We evaluated a total of 4 different approaches with these three datasets in a 5x5 stratified cross-validation setting:

- **BERT-1:** Using 80% of dataset (1) for training and evaluating the model accordingly with 20% for all 5 cross-validation runs.
- **BERT-2:** As of BERT-1 + dataset (2) for training.
- **BERT-3:** As of BERT-1 + dataset (3) for training.

[5] https://huggingface.co/

- **BERT-4:** As of BERT-1 + dataset (2) and (3) for training.

All models were trained for five epochs, with a batch size of 16 for both training and evaluation. AdamW (Loshchilov and Hutter, 2019) was used as optimizer with a learning rate of 5e-6. This hyperparameter setting proved to achieve best results in our experiments. All models were trained on an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB VRAM. The evaluation was solely carried out on the respective subset of our annotated dataset. For the implementation of the models and the evaluation we used *Pytorch* (Paszke et al., 2019), *Transformers* (Wolf et al., 2020) and *scikit-learn* (Pedregosa et al., 2011).

4 Results

4.1 Evaluation of the BERT Models

The results for the evaluation are shown in table 3. For all metrics, the best performance was achieved with BERT-2. The average accuracy is 81.8%, with precision and recall being higher for the negative and neutral classes than for the positive class. The worst accuracy was achieved with BERT-3, although BERT-1 and BERT-4 are only slightly better.

4.2 Analysis of Classification Results

All 707,241 tweets in the corpus were then classified using BERT-2. Thus, the final fine-tuned model was trained with 3,658 tweets: 606 (17%) positive, 1,512 (41%) negative and 1,540 (42%) neutral. We first present distribution and word frequency results and follow up with the analysis of time-based sentiment progressions. Please refer to table 5 in the appendix for election results to support the analysis and interpretation of the data.

4.2.1 General Analysis

As figure 1 shows, the majority of tweets were classified as negative (54.4%). Looking at the parties individually, we can see that for each party over 50% of the tweets were classified as negative, which is about the same as for the annotation subset. Tweets mentioning AFD accounts have the largest share of tweets classified as negative, while the FDP has the smallest share. However, for each party, over 50% of the tweets were classified as negative. Furthermore, it can be observed that the sentiment of the tweets that mentioned the parties that formed a government after the election (SPD,

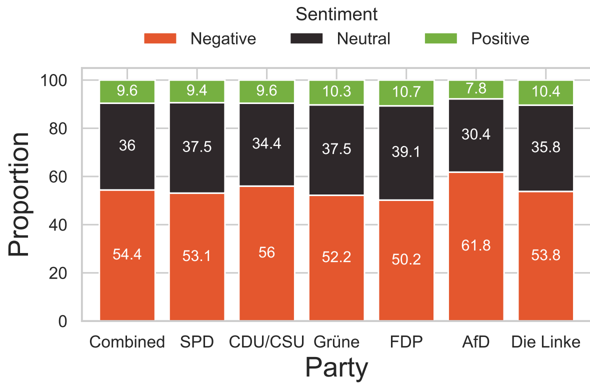

Figure 1: Distribution of the sentiment annotation for all parties.

DIE GRÜNEN, FDP) have the lowest proportion of negative tweets in comparison to the parties that would become the opposition (CDU/CSU, AfD, DIE LINKE). This might be due to an overall more positive representation in the public after the election for the winning parties.

For preliminary semantic analysis, we investigated the word frequencies of the three sentiment classes, looking for the most common positive or negative terms. In order to enhance the interpretability of the results, we removed stop words and all @-mentions from the tweets. Analyzing the word frequencies in negative tweets revealed frequent occurrences of terms such as “Corona”, “Merkel” or “Impfung” (German for vaccination), showing the importance of COVID in the political discourse of that year. Terms such as “Danke” (thank you), “gut” (good) or “Herzlichen Glückwunsch” (congratulations) are most common among the positive tweets, indicating that post-election celebrations were the major source for positive tweets. Word clouds illustrating the word frequencies of all negative and positive tweets are presented in figures 4 and 5 in the appendix.

4.3 Diachronic Sentiment Analysis

We also carried out a diachronic sentiment analysis (similar to Schmidt et al., 2022). Tweets that were classified as positive were assigned +1, neutral tweets 0, and negative tweets -1. These values were then aggregated for tweets of each month and party for the election year 2021 and a mean sentiment score was calculated by averaging this value with the number of all tweets of that month and party (see figure 2). It is noticeable that the parties’ curves are often in sync with each other and are

	BERT-1 $_1$	BERT-2 $_{1+2}$	BERT-3 $_{1+3}$	BERT-4 $_{1+2+3}$
Accuracy	80.1	**81.8**	79.7	80.4
F1 Macro	74.3	**77.5**	73.9	75.2
F1 Micro	80.1	**81.8**	79.7	80.4
F1 Weighted	80.0	**81.7**	79.5	80.2
Precision $_{\text{positive}}$	71.0	**71.0**	69.8	69.3
Precision $_{\text{negative}}$	83.5	**84.8**	83.0	83.8
Precision $_{\text{neutral}}$	77.2	**79.8**	76.9	77.6
Recall $_{\text{positive}}$	55.0	**66.7**	55.8	60.0
Recall $_{\text{negative}}$	86.2	**86.6**	85.2	85.4
Recall $_{\text{neutral}}$	76.4	**78.1**	76.3	77.2

1 = Our Annotations, 2 = Annotations by Schmidt et al. (2022), 3 = GermEval 2017

Table 3: Results of the training of the different BERT models for the classification of sentiment.

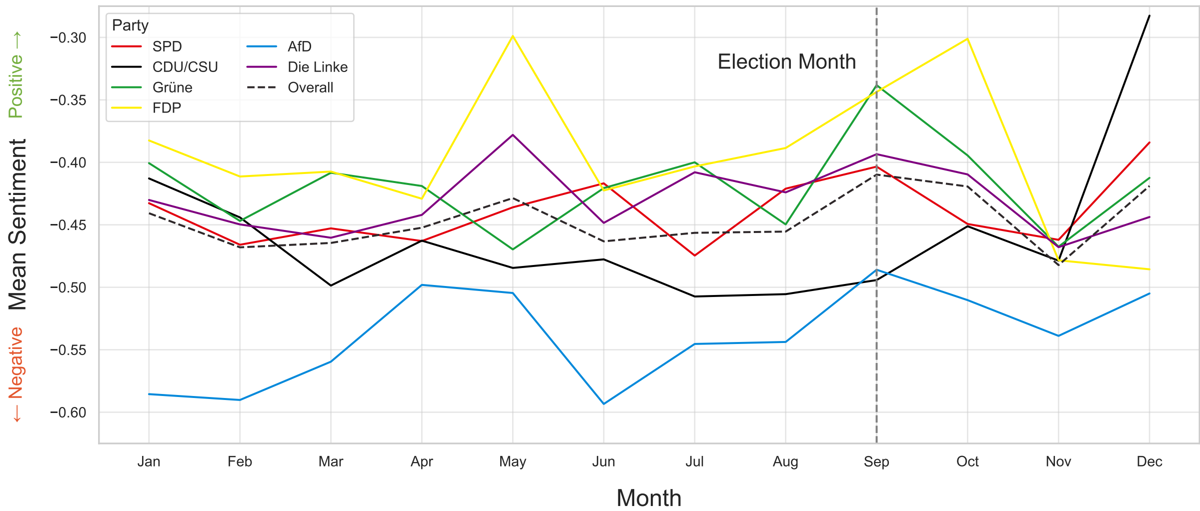

Figure 2: Mean sentiment of tweets mentioning the political accounts over the course of the election year.

constantly below -0.3 with only a few exceptions showing the dominant overall negativity in tweets that are mentioning political actors. It can be seen that the sentiment of the tweets deteriorated for all parties from January to February and October to November and improved from November to December. Tweets mentioning the AFD are the most negative for 11 months compared to the other parties, while for six months tweets mentioning the FDP are the most positive compared to the other parties.

To take a closer look at the period around election day, figure 3 shows the average sentiment of tweets mentioning the respective party for six weeks before and after the election day. Six weeks before the election day there are only small outliers but in general the sentiment remains approximately constant. Within the week starting on the election day, it is particularly noticeable that the sentiment of tweets mentioning DIE GRÜNEN was more positive compared to those mentioning DIE GRÜNEN in the previous week (about +0.4) and there is a clear outlier. The tweets mentioning the other two election winners (in terms of percentage gains) SPD and FDP were also more positive compared to the other three parties CDU, AFD and DIE LINKE, which recorded percentage losses in the election. Finally, within a week starting on 17 October, tweets that mentioned the SPD, DIE GRÜNEN and the FDP became more positive compared to the previous week, especially for the SPD (a major election winner) a clear change can be observed (about +0.3).

5 Discussion

In the following section, we discuss and interpret the overall results and highlight interesting findings. To discuss our third research question, we refer to

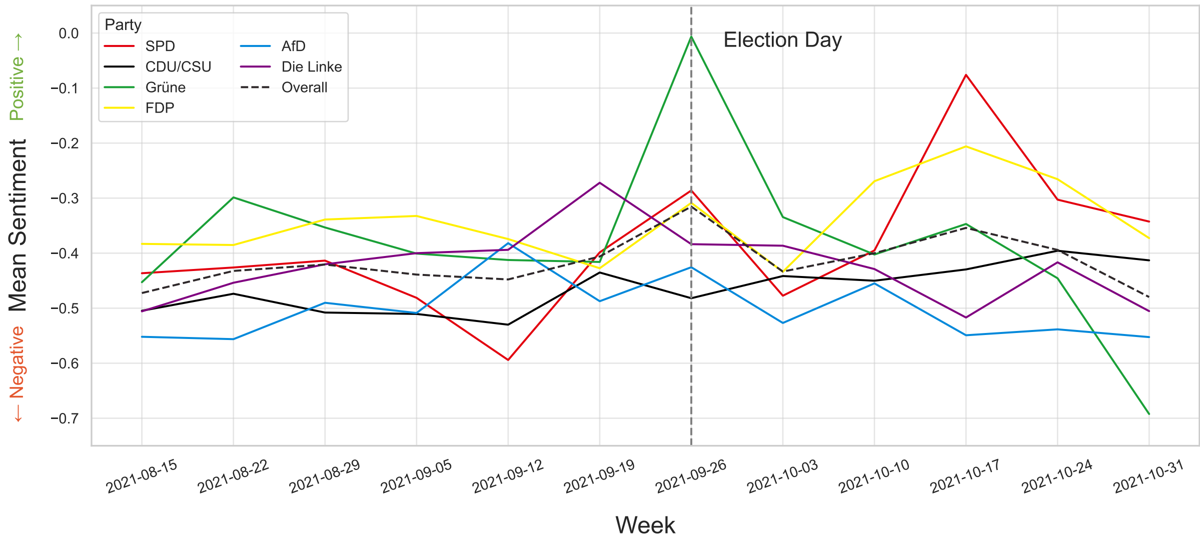

Figure 3: Mean sentiment of tweets mentioning parties over the course of 6 weeks before and after the election.

research by Schmidt et al. (2022) who did a similar study but analyzed the tweets of the 89 political actors themselves and not tweets mentioning them.

Considering the general corpus, it is noticeable that the parties in power until the election, CDU/CSU and SPD, were mentioned more often by tweets than the four opposition parties. Furthermore, it can be seen that the tweets are on average shorter than the tweets from the accounts of the political parties themselves. The tweets by the accounts of the political parties are 53.4 tokens long on average (Schmidt et al., 2022), whereas the tweets that mention the political parties, as shown in this paper, are only 31.45 tokens long. This may be because politicians, being in the public eye, are more cautious about the language and information they share. Conversely, public citizens may simply want to express their emotions towards others, and therefore, do not feel the same pressure to use more words and explain themselves in more detail.

We annotated a subset of 2,000 tweets from the corpus and achieved substantial agreement among the annotators. Sometimes, annotations showed disagreement, particularly in cases where tweets contained ironic and sarcastic language, or expressed mixed sentiments. This made it difficult for individual annotators to determine the overall sentiment of the tweet, resulting in varied interpretations. The annotated data set was then used to fine-tune a BERT model. The best of the evaluated models achieved an accuracy of about 81.8%. Methods of hyperparameter optimization and dealing with the class imbalance by assigning weights to labels for loss calculation during training showed no improvements. However, overall, the accuracies are in line with similar classification results in German (Chan et al., 2020).

Using the best model, we then classified the sentiment of the tweets of the entire corpus, with more than half of the tweets being classified as negative and less than 10% as positive. Compared to the tweets from the accounts of the political parties themselves (Schmidt et al., 2022), the sentiment is far less positive and more negative in average. This could possibly be attributed to politicians using positive and diplomatic language to gain support for their policies while avoiding offending anyone, whereas citizens tend to use negative language to express their frustration or dissatisfaction with political events or decisions. Regarding party-based classification results, we showed that tweets referring to the AFD were most often classified as negative compared to the other parties. Tweets about the election winner parties showed the most positive sentiment.

Subsequently, we analyzed how the sentiment in the tweets has evolved over the course of the election year. We identified several overall sentiment drops and peaks. The drop in sentiment observed in tweets from January to February could be explained by the ongoing discussions of state-level COVID-19 regulations during that period, as indicated by the corresponding term frequencies for these months. In these two months, terms such as “Lockdown”, “Pandemie” (pandemic) and “Corona” were frequently used in tweets. The drop in sentiment from October to November can probably also be explained by the fact that new COVID-19 restrictions were discussed after the summer, at which time similar terms were mentioned in the tweets. At the time of election day in September, the peak of DIE GRÜNEN is particularly noticeable and can be explained because they recorded the strongest percentage gain compared to the other parties. These findings are consistent with previous research (Ali et al., 2022) which has shown that there tends to be an increase in positive sentiment regarding the winning candidate around the time of the election and the announcement of the results. On the other hand, the CDU is the party where the tweets mentioning their accounts have the lowest sentiment in September and the week after the election compared to the other parties. Presumably, this can be explained since they lost the most votes in percentage compared to the last election. Finally, the increase in sentiment from November to December is likely due to the election of a new federal cabinet. This is supported by the most frequent words used in these months like "Glückwunsch", "Gratulation" (both congratulation) and "Erfolg" (success). Comparing the changes in sentiment over the year, there are differences and similarities comparing tweets from political party accounts (Schmidt et al., 2022) and tweets mentioning them, as studied in this paper. Our findings indicate that the parties do not have similar tops and lows and that the parties' courses of the sentiment are more asymmetric to each other. Nevertheless, the highs and lows in February, November and December are also recognizable and prove that major international and national events influence both in similar way politicians' tweets and tweets by the public about them. Our results also show that the sentiment of the AFD is more negative in most months compared to the other parties. But in comparison, the CDU is not the party whose tweets show the most positive sentiment in most months. Among the tweets we looked at, the CDU is the party with the most positive tweets only in December, most often it is the FDP.

6 Limitations and Future Work

Our work provides insights on how the political parties were perceived on Twitter in the election year 2021 and we contribute resources to the research area of sentiment analysis in German. However, there are limitations of our work that we intend to address in future work: Due to the high number of tweets mentioning party accounts, we decided not to collect tweets for all days within the election year and instead acquired tweets for two random days of each month in 2021. This certainly limits the representativeness of our corpus, since critical events or fluctuations in public sentiment may have been overlooked, like the 2021 European Floods, killing 196 in Germany[6] which had a strong impact in Germany during the election campaign. Furthermore, our corpus contains tweets that mention multiple accounts, which can dilute the sentiment targeted at the primary party or politician of interest. Another limitation is the accuracy of the trained model. While it is in line with similar studies and evaluation results (Chan et al., 2020), we plan to improve accuracies by annotating more tweets and exploring more methods of hyperparameter optimization. We want to address the performance also with more sophisticated methods to deal with class imbalance (Buda et al., 2018). Moreover, we will investigate the addition of more complex classes similar to emotion classification (Schmidt et al., 2021b; Dennerlein et al., 2023), as the annotators also reported that nuanced emotions occurred often. Furthermore, Twitter also offers multimedia content that we intend to explore via computer vision based sentiment- and emotion analysis (Schmidt et al., 2021c; Schmidt and Wolff, 2021; El-Keilany et al., 2022). Lastly, we also want to highlight that Twitter is not as popular in Germany as in other countries and thus represents a limited subsection of public social media sentiment. According to surveys, 10% of Germans use Twitter regularly[7] compared to 23% of U.S. adults.[8] In addition to that, we intend to improve upon the semantic exploration of our data via more sophisticated methods like topic modeling and named entity recognition. On the annotation side, we plan to investigate possibilities of more fine-grained annotation to gain a better understanding of the annotation theory on this material. Parts of our research and more information about this project are publicly available to support further research in this area.[9]

[6] Cf. https://en.wikipedia.org/wiki/2021_European_floods.

[7] https://de.statista.com/statistik/daten/studie/171006/umfrage/in-anspruch-genommene-angebote-aus-dem-internet/

[8] https://www.statista.com/statistics/232818/active-us-twitter-user-growth/

[9] https://github.com/NilsHellwig/Twitter_German_Federal_Election_Perception_2021

References

Amit Agarwal, Ritu Singh, and Durga Toshniwal. 2018. Geospatial sentiment analysis using twitter data for uk-eu referendum. *Journal of Information and Optimization Sciences*, 39(1):303–317.

Rao Hamza Ali, Gabriela Pinto, Evelyn Lawrie, and Erik J Linstead. 2022. A large-scale sentiment analysis of tweets pertaining to the 2020 us presidential election. *Journal of big Data*, 9(1):1–12.

Jamilu Awwalu, Azuraliza Abu Bakar, and Mohd Ridzwan Yaakub. 2019. Hybrid n-gram model using naïve bayes for classification of political sentiments on twitter. *Neural Computing and Applications*, 31(12):9207–9220.

Mateusz Buda, Atsuto Maki, and Maciej A. Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106:249–259.

Josemar A Caetano, Hélder S Lima, Mateus F Santos, and Humberto T Marques-Neto. 2018. Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 american presidential election. *Journal of internet services and applications*, 9(1):1–15.

Branden Chan, Stefan Schweter, and Timo Möller. 2020. German's next language model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6788–6796, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Hassan Nazeer Chaudhry, Yasir Javed, Farzana Kulsoom, Zahid Mehmood, Zafar Iqbal Khan, Umar Shoaib, and Sadaf Hussain Janjua. 2021. Sentiment analysis of before and after elections: Twitter data of us election 2020. *Electronics*, 10(17):2082–2108.

Nalini Chintalapudi, Gopi Battineni, and Francesco Amenta. 2021. Sentimental analysis of covid-19 tweets using deep learning models. *Infectious Disease Reports*, 13(2):329–339.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. In *International Conference on Learning Representations*.

Nhan Cach Dang, María N Moreno-García, and Fernando De la Prieta. 2020. Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3):483–512.

Katrin Dennerlein, Thomas Schmidt, and Christian Wolff. 2023. Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century. *Digital Scholarship in the Humanities*, 38(4):1466–1481.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Alina El-Keilany, Thomas Schmidt, and Christian Wolff. 2022. Distant Viewing of the Harry Potter Movies via Computer Vision. In *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022).*, pages 33–49, Uppsala, Sweden.

Shihab Elbagir and Jing Yang. 2019. Twitter sentiment analysis using natural language toolkit and vader sentiment. In *Proceedings of the international multiconference of engineers and computer scientists 2019*.

Jakob Fehle, Leonie Münster, Thomas Schmidt, and Christian Wolff. 2023. Aspect-based sentiment analysis as a multi-label classification task on the domain of german hotel reviews. In *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*, Ingolstadt, Germany. KONVENS 2023 Organizers.

Jakob Fehle, Thomas Schmidt, and Christian Wolff. 2021. Lexicon-based sentiment analysis in German: Systematic evaluation of resources and preprocessing techniques. In *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pages 86–103, Düsseldorf, Germany. KONVENS 2021 Organizers.

Feng Fu, Martin A Nowak, Nicholas A Christakis, and James H Fowler. 2012. The evolution of homophily. *Scientific reports*, 2(1):845.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the Eigth International AAAI conference on web and social media*, volume 8, pages 216–225.

Mochamad Ibrahim, Omar Abdillah, Alfan F Wicaksono, and Mirna Adriani. 2015. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In *2015 IEEE international conference on data mining workshop (ICDMW)*, pages 1348–1353. IEEE.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.

John Richard Landis and Gary Grove Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Joseph Lilleberg, Yun Zhu, and Yanqing Zhang. 2015. Support vector machines and word2vec for text classification with semantic features. In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 136–140. IEEE.

Bing Liu. 2020. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. Machine Learning Repository. arXiv:1711.05101. Version 3.

Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. Deep learning–based text classification: A comprehensive review. *ACM Computing Surveys*, 54(3):1–40.

Luis Moßburger, Felix Wende, Kay Brinkmann, and Thomas Schmidt. 2020. Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum. In *Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task*, pages 70–81, Barcelona, Spain (Online). Association for Computational Linguistics.

Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 9–14, Online. Association for Computational Linguistics.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 8026–8037, Red Hook, NY, USA. Curran Associates Inc.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897.

Jyoti Ramteke, Samarth Shah, Darshan Godhia, and Aadil Shaikh. 2016. Election result prediction using twitter sentiment analysis. In *2016 international conference on inventive computation technologies (ICICT)*, volume 1, pages 1–5. IEEE.

Thomas Schmidt and Manuel Burghardt. 2018. An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 139–149, Santa Fe, New Mexico. Association for Computational Linguistics.

Thomas Schmidt, Katrin Dennerlein, and Christian Wolff. 2021a. Emotion classification in German plays with transformer-based language models pretrained on historical and contemporary language. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 67–79, Punta Cana, Dominican Republic (online). Association for Computational Linguistics.

Thomas Schmidt, Katrin Dennerlein, and Christian Wolff. 2021b. Using Deep Learning for Emotion Analysis of 18th and 19th Century German Plays. In Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke, Niels-Oliver Walkowski, Joëlle Weis, and Ulrike Wuttke, editors, *Fabrikation von Erkenntnis: Experimente in den Digital Humanities. Teilband 1*. Melusina Press, Esch-sur-Alzette, Luxembourg.

Thomas Schmidt, Alina El-Keilany, Johannes Eger, and Sarah Kurek. 2021c. Exploring Computer Vision for Film Analysis: A Case Study for Five Canonical Movies. In *2nd International Conference of the European Association for Digital Humanities (EADH 2021)*, Krasnoyarsk, Russia.

Thomas Schmidt, Jakob Fehle, Maximilian Weissenbacher, Jonathan Richter, Philipp Gottschalk, and Christian Wolff. 2022. Sentiment analysis on Twitter for the major German parties during the 2021 German federal election. In *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)*, pages 74–87, Potsdam, Germany. KONVENS 2022 Organizers.

Thomas Schmidt, Florian Kaindl, and Christian Wolff. 2020. Distant reading of religious online communities: A case study for three religious forums on reddit. In *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference (DHN 2020)*, pages 157–172, Riga, Latvia.

Thomas Schmidt and Christian Wolff. 2021. Exploring Multimodal Sentiment Analysis in Plays: A Case Study for a Theater Recording of Emilia Galotti. In *Proceedings of the Conference on Computational Humanities Research 2021 (CHR 2021)*, pages 392–404, Amsterdam, The Netherlands.

Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2011. Election forecasts with twitter: How 140 characters reflect the political landscape. *Social science computer review*, 29(4):402–418.

Michael Wojatzki, Eugen Ruppert, Sarah Holschneider, Torsten Zesch, and Chris Biemann. 2017. GermEval 2017: Shared Task on Aspect-based Sentiment in Social Media Customer Feedback. In *Proceedings of the GermEval 2017 – Shared Task on Aspect-based Sentiment in Social Media Customer Feedback*, pages 1–12, Berlin, Germany.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Ethan Xia, Han Yue, and Hongfu Liu. 2021. Tweet sentiment analysis of the 2020 u.s. presidential election. In *Companion Proceedings of the Web Conference 2021*, WWW '21, page 367–371, New York, NY, USA. Association for Computing Machinery.

Ussama Yaqub, Nitesh Sharma, Rachit Pabreja, Soon Ae Chun, Vijayalakshmi Atluri, and Jaideep Vaidya. 2020. Location-based sentiment analyses and visualization of twitter election data. *Digital Government: Research and Practice*, 1(2):1–19.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1253.

Xinyang Zhang, Yury Malkov, Omar Florez, Serim Park, Brian McWilliams, Jiawei Han, and Ahmed El-Kishky. 2023. Twhin-bert: A socially-enriched pre-trained language model for multilingual tweet representations at twitter. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 5597–5607, New York, NY, USA. Association for Computing Machinery.

A Appendix

A.1 Annotation Examples

Annotation	Tweet	Author	Mentioned Account (Party)
positive	(Offenbar) Unpopular opinion: Ich mag @ArminLaschet als Persönlichkeit und kann ihn mir als Kanzler durchaus vorstellen.	@fredschorn	@ArminLaschet (CDU)
negative	@derspiegel Scholz versagt irgendwie, @Karl_Lauterbach, tun Sie etwas.	@1worldvs1virus	@Karl_Lauterbach (SPD)
neutral	@rRockxter @europeika @CDU Was hat denn die CDU mit dem Christentum zu tun?	@123JulianN321	@CDU (CDU)
mixed	@GrueneBundestag @BriHasselmann Grüne Verbots Partei.. ahnungslos Glücklich	@Paellamixta	@GrueneBundestag (Die Grünen)
no majority	@gb_1960 @SWagenknecht Die Gehirnwäsche hat gewirkt. Du hättest herrlich in die DDR gepasst.	@MaierJrg1	@SWagenknecht (Die Linke)

Table 4: Annotation Examples: For the first four tweets, the annotators were unanimous, the last example was annotated as neutral, positive and mixed (no majority).

A.2 Results of German Federal Election 2021

Party	Full Name	2021	2017	Change
SPD	Social Democratic Party of Germany	25.7 %	20.5 %	+ 5.2 %
CDU/CSU	Christian Democratic Union/ Christian Social Union (Bavaria)	24.1 %	32.9 %	- 8.8 %
Die Grünen	The Greens	14.8 %	8.9 %	+ 5.9 %
FDP	Free Democratic Party	11.5 %	10.7 %	+ 0.8 %
AfD	Alternative for Germany	10.3 %	12.6 %	- 2.3 %
Die Linke	The Left	4.9 %	9.2 %	- 4.3 %

Table 5: Election results of the 2021 federal election and changes compared to the previous election in 2017.

A.3 Twitter accounts for data acquisition

A.3.1 Parties

SPD	CDU	CSU	Die Grünen	FDP	AfD	Die Linke
@spdde Follower: 417k Tweets: 22,138	@CDU Follower: 378k Tweets: 37,100	@CSU Follower: 229k Tweets: 9,072	@Die_Gruenen Follower: 649k Tweets: 30,560	@fdp Follower: 414k Tweets: 27,981	@AfD Follower: 173k Tweets: 8,330	@dieLinke Follower: 350k Tweets: 14,135
@spdbt Follower: 217k Tweets: 9,809	@cducsubt Follower: 166k Tweets: 13,250		@GrueneBundestag Follower: 186k Tweets: 6,399	@fdpbt Follower: 39k Tweets: 8,194	@AfDimBundestag Follower: 68k Tweets: 4,713	@Linksfraktion Follower: 108k Tweets: 2,994
@jusos Follower: 77k Tweets: 1,847	@Junge_Union Follower: 79k Tweets: 931		@gruene_jugend Follower: 76k Tweets: 1,290	@fdp_nrw Follower: 28k Tweets: 884	@AfDBerlin Follower: 19k Tweets: 364	@dielinkeberlin Follower: 19k Tweets: 1,228

Table 6: The 3 main accounts with the most followers for each party (as of January 2022).

A.3.2 Politicians

SPD	CDU	CSU
@Karl_Lauterbach Follower: 770k Tweets: 132,526	@jensspahn Follower: 279k Tweets: 35,571	@Markus_Soeder Follower: 341k Tweets: 30,495
@HeikoMaas Follower: 660k Tweets: 6,431	@ArminLaschet Follower: 188k Tweets: 36,161	@DoroBaer Follower: 103k Tweets: 2,560
@OlafScholz Follower: 324k Tweets: 27,414	@_FriedrichMerz Follower: 179k Tweets: 23,651	@andreasscheuer Follower: 63k Tweets: 2,431
@KuehniKev Follower: 323k Tweets: 5,192	@JuliaKloeckner Follower: 74k Tweets: 3,357	@ManfredWeber Follower: 54k Tweets: 527
@larsklingbeil Follower: 116k Tweets: 5,669	@n_roettgen Follower: 68k Tweets: 4,645	@DerLenzMdB Follower: 10k Tweets: 236
@hubertus_heil Follower: 108k Tweets: 2,406	@PaulZiemiak Follower: 58k Tweets: 12,723	@hahnflo Follower: 9k Tweets: 2,900
@EskenSaskia Follower: 101k Tweets: 7,180	@groehe Follower: 49k Tweets: 79	@smuellermdb Follower: 9k Tweets: 239
@Ralf_Stegner Follower: 64.9k Tweets: 7,061	@HBraun Follower: 39k Tweets: 3,212	@DaniLudwigMdB Follower: 8k Tweets: 3,821
@KarambaDiaby Follower: 55.6k Tweets: 392	@rbrinkhaus Follower: 30k Tweets: 4,280	@ANiebler Follower: 6k Tweets: 25
@MiRo_SPD Follower: 39k Tweets: 350	@tj_tweets Follower: 17k Tweets: 396	@MarkusFerber Follower: 5k Tweets: 21

Die Grünen	FDP	AfD	Die Linke
@cem_oezdemir Follower: 290k Tweets: 9,942	@c_lindner Follower: 552k Tweets: 19,942	@Alice_Weidel Follower: 138k Tweets: 9,367	@SWagenknecht Follower: 518k Tweets: 7,177
@GoeringEckardt Follower: 202k Tweets: 5,227	@MaStrackZi Follower: 46k Tweets: 2,453	@Joerg_Meuthen Follower: 76k Tweets: 4,813	@GregorGysi Follower: 439k Tweets: 1,722
@JTrittin Follower: 115k Tweets: 1,782	@MarcoBuschmann Follower: 46k Tweets: 10,062	@Beatrix_vStorch Follower: 68k Tweets: 3,962	@katjakipping Follower: 130k Tweets: 1,072
@KonstantinNotz Follower: 85k Tweets: 2,144	@KonstantinKuhle Follower: 44k Tweets: 2,710	@gottfriedcurio Follower: 37k Tweets: 275	@DietmarBartsch Follower: 82k Tweets: 3,409
@RenateKuenast Follower: 77k Tweets: 2,026	@johannesvogel Follower: 38k Tweets: 2,121	@MalteKaufmann Follower: 36k Tweets: 5149	@anked Follower: 43k Tweets: 935
@Ricarda_Lang Follower: 65k Tweets: 3,546	@Wissing Follower: 32k Tweets: 2,805	@JoanaCotar Follower: 30k Tweets: 4,330	@b_riexinger Follower: 41k Tweets: 1,399
@KathaSchulze Follower: 37k Tweets: 4,609	@Lambsdorff Follower: 27k Tweets: 884	@Tino_Chrupalla Follower: 21k Tweets: 2,875	@jankortemdb Follower: 34k Tweets: 743
@BriHasselmann Follower: 37k Tweets: 1,795	@ria_schroeder Follower: 23k Tweets: 359	@StBrandner Follower: 23k Tweets: 11,914	@Janine_Wissler Follower: 37k Tweets: 1,046
@nouripour Follower: 29k Tweets: 505	@LindaTeuteberg Follower: 23k Tweets: 328	@GtzFrmming Follower: 17k Tweets: 984	@SevimDagdelen Follower: 35k Tweets: 172
@MiKellner Follower: 28k Tweets: 3,436	@f_schaeffler Follower: 20k Tweets: 1,092	@PetrBystronAFD Follower: 17k Tweets: 496	@SusanneHennig Follower: 29k Tweets: 4,463

Table 7: The 10 accounts with the most followers for each party (as of January 2022).

A.4 Word clouds for positive and negative tweets

Figure 4: Word cloud created of negative tweets in the corpus for all parties.(These visualizations were generated by the Python package *wordcloud*.)

Figure 5: Word cloud created of positive tweets in the corpus for all parties.