Sentiment Analysis on Twitter for the Major German Parties during the 2021 German Federal Election

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`

Maximilian Weissenbacher
Media Informatics Group
University of Regensburg
Regensburg, Germany
`maximilian.`
`weissenbacher@ur.de`

Jonathan Richter
Media Informatics Group
University of Regensburg
Regensburg, Germany
`jonathan.richter`
`@ur.de`

Philipp Gottschalk
Media Informatics Group
University of Regensburg
Regensburg, Germany
`philipp.gottschalk`
`@ur.de`

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

Abstract

We present the results of a project performing sentiment analysis on tweets from German politicians and party accounts for the 2021 German federal election. We collected over 58,000 tweets from the Twitter accounts of the seven parties represented in the German Bundestag, of which a selection of 2,000 tweets were annotated by three annotators. Based on the annotated data, we implemented multiple sentiment analysis approaches and evaluated the sentiment classification performance. We found that transformer-based models like bidirectional encoder from transformers (BERT) performed better than traditional machine learning models such as Naive Bayes and lexicon-based models like GerVADER. The best performing BERT model achieved an accuracy of 93.3% and macro f1 score of 93.4%. Applying sentiment analysis on the overall corpus via this method showed that overall, negative sentiment was most frequent and that there were multiple major shifts in sentiment a few months before and after the election. Furthermore, we found that tweets from opposition parties had on average more negative sentiment than those from governing parties.

1 Introduction

The 2021 federal election in Germany led to a dramatic change in power of the leading parties. Angela Merkel's chancellorship and the reign of the CDU (Christlich Demokratische Union) came to an end after 16 years and a new coalition now forms the government (see table A.1 in the appendix for election results). Whereas former election campaigns only took place in the real world through posters and election events, ever since the rise of social media, campaigns additionally focus on gaining support on the internet (Freelon, 2017). During elections, politicians of all parties are strategic actors focused on gaining voters' support (Druckman et al., 2010). Besides online advertisements, political discussions via social media have gained more and more importance. This is a worldwide phenomenon but can especially be seen in the United States (Tumasjan et al., 2010) where former president Donald Trump used Twitter almost on a daily basis to share his opinion on a wide variety of topics. Twitter is a micro-blogging platform and one of the most popular social-media channels for online communication. Sharing content takes place in form of a short text, limited to 280 characters, which is called a tweet. Twitter has become an important platform for research in computational social science and a source for research conducting sentiment analysis (Drus and Khalid, 2019). Sentiment analysis is the computational method to predict the sentiment, attitude or opinion of media, predominantly text (Liu, 2015). It is often regarded as a classification task with the categories positive, neutral and negative (Wagh et al., 2018). Sentiment analysis can also be differentiated into three different description levels: document level, sentence level, feature level (Liu, 2015). In this study, we are focusing on complete tweets as level of analysis (i.e. sentiment analysis on document level). There are a variety of methods to perform sentiment analysis ranging from rule-based approaches to the application of transformer-based language

models (Drus and Khalid, 2019; Guhr et al., 2020).

In this study, we analyze the social media behaviour of German politicians and parties during the federal election of 2021 by applying sentiment analysis on the tweets of the entire election year for a selection of party accounts (58,864 tweets). The goal of this work is to gain insights about political parties' sentiment during the election year 2021. Our research questions are as follows:

- What is the best performing sentiment analysis technique in this use case of political tweets in regards to common methods and state-of-the-art recommendations?

- How does the sentiment of parties expressed in tweets differ from each other in general and with respect to government/opposition and election winner/loser relations?

- How does the sentiment of parties expressed in tweets change across the election year?

Our main contributions to the research area are as follows:

- The acquisition and preparation of all tweets of 89 Twitter accounts for the year 2021 of the most important German political parties (58,864 tweets)

- The annotation of a subset of 2,000 tweets with sentiment information for evaluation and machine learning (ML) purposes

- The implementation and evaluation of a lexicon-based approach, sentiment analysis based on traditional machine learning and the application of a large German BERT model on our annotated data set and a larger additional corpus

- The investigation of the above research questions applying the best performing sentiment model on our overall corpus

We release our annotated data sets and best performing model as well as additional data and visualizations via GitHub[1] to support further research. We apply the best performing model on our overall corpus to investigate the proposed research questions.

2 Related Work

Ever since the rise of social media, sentiment analysis on social media platforms is a very active research area (Wang et al., 2012; Elbagir and Yang, 2019). Sentiment analysis is used, for example, to explore sentiment in *Reddit* forums (Schmidt et al., 2020c; Moßburger et al., 2020), on Twitter (Elbagir and Yang, 2019) or social media artefacts like memes (Schmidt et al., 2020b). In the following chapters, we summarize important research in the context of political analysis on Twitter and offer an overview of current sentiment analysis methodology.

2.1 Sentiment Analysis on Twitter for Political Research

Research in political sentiment analysis on Twitter differs between the analysis of accounts of political actors and the analysis of public sentiment towards political events or actors. As examples for the latter, Bermingham and Smeaton (2011) investigated whether it is possible to predict the election results for the Irish general election 2011. The results showed that the analysis of sentiment indeed offers predictive qualities. Furthermore, there was a big sentiment-shift two days before the election day which already gave indications on the election results. In similar research for India, Sharma and Moh (2016) showed that parties which were mentioned in tweets with a positive sentiment are more likely to win election votes than parties with a negative sentiment.

Considering the analysis of political actors, Tumasjan et al. (2010) analyzed the sentiment during the German federal election in 2009. For politicians and parties they discovered that the politicians' sentiment profiles reflected different nuances of the election campaign. Furthermore, polarizing politicians from the opposition showed inversed sentiment. Budiharto and Meiliana (2018) focused on the Indonesian presidential candidates and were able to predict popularity with various Twitter metrics including results of sentiment analysis. Recently Costa et al. (2021) analyzed the communication of parties and their sentiments in Portugal in one year. When comparing the results, the authors found a great variability between the parties. They revealed that the party being at the opposition had the most positive sentiment profile and the right wing generally expressed more positive sentiment than the left wing.

Overall, research shows that sentiment analysis of political actors and the public on Twitter can serve as source of analysis and predictor of popularity. Similar to previous research, we will focus on the identification of sentiment shifts and differences among parties.

2.2 Methods for Sentiment Analysis

Large transformer-based language models like BERT and ELECTRA are currently considered state-of-the-art for sentiment analysis tasks (Qiu et al., 2020; Chouikhi et al., 2021; Chan et al., 2020) and outperform traditional ML approaches using Naive Bayes or Support Vector Machines (SVM) (Geetha and Renuka, 2021). The large German language model *gbert* by *deepset* outperforms other models on a variety of tasks including sentiment analysis (Chan et al., 2020).

Nevertheless, another type of regularly used sentiment analysis approaches are lexicon-based methods. Lexicon-based sentiment analysis is a rule-based method using a dictionary, in which words with positive and negative connotations are stored.The basic idea is that the majority of the occurring words (or their values) of a class decides about the classification of a text unit, e.g. if predominantly positive connoted words occur in the text, it will be classified as positive (Jurek et al., 2015; Schmidt et al., 2021a). This branch of rule-based methods, while being outperformed by ML approaches in most settings, is still popular and common for German language research (Fehle et al., 2021). Lexicon-based methods are often applied in settings that lack annotated corpora and ML possibilities in German like literary texts (Schmidt and Burghardt, 2018; Schmidt et al., 2020a) or in human-computer interaction (Ortloff et al., 2019; Schmidt et al., 2020d). Thus we included this method in our evaluation. Indeed, in the context of the U.S. presidential elections 2016, the well-known lexicon-based sentiment analysis module VADER (Hutto and Gilbert, 2014) was used for the analysis of tweets (Elbagir and Yang, 2019). Besides lexicon-based methods, traditional ML approaches have also been used in research of political tweet analysis (Bermingham and Smeaton, 2011; Sharma and Moh, 2016). Traditional ML approaches follow a two-step process, which first extracts manually annotated features from the tweet to subsequently feed them into a classifier, e.g. SVM, which in turn makes predictions on novel (or un-

seen) data (Minaee et al., 2021). While transformer-based models have shown to outperform the aforementioned methods, we also implemented examples of lexicon-based methods and traditional ML to serve as baselines.

3 Methods

3.1 Data Acquisition

We gathered tweets from the seven parties currently represented in the German Bundestag for an entire year. For each party, we selected the ten most relevant politicians (according to their Twitter follower count) as well as the three largest official party-accounts (as of January 2022), which are mostly the national or regional party accounts (see Fig. 7 and Fig. 8 in the appendix for the full list of accounts). This results to tweets by 89 Twitter accounts (the party-accounts for the parties CDU and CSU were summarized to 4 accounts). In the following we do however report results for 6 parties by combining the tweets by CDU and CSU since both parties are in political proximity and the CSU is basically the Bavarian representative of the party. We used the *Scweet* (Jeddi and Bengadi, 2022) package for the acquisition of tweets, which downloads tweets from specific accounts and stores them in a CSV-File. For the data collection, we set the time frame to January 2021 to December 2021 to cover a large period before the election on September 26th as well as several months after the election. Tweet replies or retweets were not taken into account to obtain only those tweets that were written by the respective user themselves and thus contain the user's own wording and sentiment. The final tweet corpus contains of 58,864 distinct tweets. Table 1 summarizes general corpus statistics and further party information. The corpus consists of over 3 million tokens. A tweet consists on average of 53 tokens and the number of tweets per party differ with the AfD having the most tweets and the FDP the fewest.

3.2 Data Annotation

We selected a subset of 2,000 tweets using stratified (in respect to the proportion of tweets per party) random sampling to create an annotated sub corpus to use for evaluation and machine learning purposes. Each tweet was annotated by three annotators independently from each other. The annotators were three native-speaking students or research assistants respectively. We created an annotation

Partei	political orientation	pre-election	post-election	# tweets	%	# tokens	avg. tweet length
AfD	far right	opposition	opposition	11,625	20	592,828	51.00
CDU/CSU	center right	government	opposition	10,072	17	512,803	50.91
Die Linke	far left	opposition	opposition	9,628	16	522,322	54.25
FDP	liberal	opposition	government	6,610	11	356,789	53.98
Die Grünen	left, ecological	opposition	government	9,576	16	537,408	56.12
SPD	center left	government	government	11,353	19	623,572	54.93
Absolute	-	-	-	58,864	100	3,145,722	53.44

Table 1: General corpus statistics of the overall tweet corpus.

manual with examples and instructions for the annotation of a tweet to ensure consistent annotation. Annotators were instructed to annotate the sentiment the tweet expresses. The annotation-classes were as follow:

1. **Positive.** Tweets with a predominantly positive sentiment

2. **Negative.** Tweets with a predominantly negative sentiment

3. **Neutral.** Tweets expressing no sentiment or neutral

4. **Mixed.** Tweets with a mix of positive and negative sentiment

Table 2 shows annotation examples. We used Fleiss' κ and Krippendorff's α as metrics to measure the inter-rater agreement between annotators. This was implemented with the *Statsmodels* (Seabold and Perktold, 2010) and *Krippendorff*[2] Python packages. The results of Fleiss' κ and Krippendorff's α with a value of 0.53 show moderate agreement according to the interpretation of Landis and Koch (1977). Indeed, agreement metrics for sentiment annotation on tweets do differ between very high and rather low depending on the number of classes and overall setting and our results are slightly below the average in similar settings (cf. Salminen et al., 2018). Studies in the context of German literary texts (Schmidt et al., 2019b,a) or movie subtitles (Schmidt et al., 2020a) do report similar or lower levels of agreement. In our case, the mediocre agreement shows the challenges of the annotation and that the tweets were often open to interpretation.

To deal with the mediocre agreement, the final annotation of a tweet was determined according to the majority of individual decisions. If no majority could be determined or the tweets were classified as mixed by the majority, these tweets were not considered in the further process. Table 3 shows the distribution of the annotated tweets. In total, this majority decision leads to an annotated corpus of 1,785 tweets.

3.3 Sentiment Analysis

We regard the sentiment analysis as single-label classification task with the classes positive, neutral and negative. We implemented and evaluated the following approaches:

3.3.1 Lexicon-Based Approaches

We used *GerVADER* (Tymann et al., 2019) which is a German adaption of the English tool *VADER* (Hutto and Gilbert, 2014) and showed positive results in the context of German social media content (Tymann et al., 2019). In *GerVADER* the German sentiment dictionary *SentiWS* (Remus et al., 2010) is used for the sentiment calculation. The lexicon consists of 1,650 positive and 1,818 negative words and their inflections resulting in over 32,000 different word forms. The words' sentiment is scaled between the values -1 and 1.

3.3.2 Traditional Machine Learning Approaches

We compared Multinomial Naive Bayes and Support Vector Machines. To train and test these models, a bag-of-words approach with 5-fold cross-validation was carried out. Since preprocessing of texts is recommended for these approaches (Krouska et al., 2016), we performed the following preprocessing steps: filtering punctuation, stop words and unique words, normalization via lower

Annotation	Tweet	Account
positive	@MikeJosef FFM ist ein engagierter SPD-Kandidat mit viel Einsatz und Ideen für seine Stadt Frankfurt am Main. Am 14.3. könnt Ihr ihn wählen, liebe Frankfurter*innen! Für eine lebenswerte, moderne und soziale Metropole im Herzen von Europa.	@OlafScholz
negative	Die CDU ist die Partei der sozialen Kälte. #Triell	@Ricarda_Lang
neutral	Es ist nicht die Zeit für Einen zu sagen: Ich mache alles. Wir müssen uns jetzt breit aufstellen. #CDUVorsitz #jetztabervoran	@n_roettgen
mixed	Medien berichten über Neuformierung der Parteispitzen von @spdde @Die_Gruenen + @CDU Vergleich hinkt, weil @CDU Weg aus tiefer Orientierungs- +Personalkrise sucht, während @spdde + @Die_Gruenen Personalwechsel eher herausfordernde Begleiterscheinungen politischen Erfolges sind	@Ralf_Stegner
no major-ity	Wir wollen nicht zurückfallen in ein Spiel der nationalen Mächte, in eine Zeit, in der man im permanenten, destruktiven Wettstreit war - sondern Dinge gemeinsam hinkriegen und an die Entspannungspolitik von Willy Brandt und Helmut Schmidt anknüpfen. #Progressives4Europe	@OlafScholz

Table 2: Annotation examples. First three examples annotators agree upon. Last example is annotated as negative, neutral and mixed.

Sentiment	Count	Percentage
Neutral	**763**	38,15%
Negative	**536**	26,80%
Positive	**486**	24,3%
No Majority	120	6,00%
Mixed	95	4,75%

Table 3: Sentiment class distribution of the annotated subset.

casing and lemmatization. The aforementioned steps were implemented in python using the libraries *NLTK*[3], *sklearn* (Pedregosa et al., 2011) and *spaCy* (Honnibal and Montani, 2017).

3.3.3 Transformer-Based Approaches

We also evaluated the, to our knowledge, one of the largest publicly available German transformer-based language model *gbert-base* by *deepset* (Chan et al., 2020). The model was acquired via the *Hugging Face* platform (Wolf et al., 2020) and was implemented with the library *Simple Transformers* (Rajapakse, 2019), an adaption of *Hugging Face*'s library *Transformers*.

We used gbert-base[4] and fine-tuned it to the downstream task of sentiment classification differing between three different data sets for the training: (1) the 1,785 annotated tweets of our own data set,

---

[3] https://www.nltk.org/
[4] https://huggingface.co/deepset/gbert-base

(2) the freely available *GermEval 2017* data set (Wojatzki et al., 2017), consisting of around 28,000 annotated German posts from various social media sources, representing one of the largest data sets of German sentiment-annotated posts, and (3) the combination of data sets (1) and (2). Each model is trained and evaluated in 5x5 stratified setting containing only the annotated data set. For methods (2) and (3) the *GermEval* data set is added to the training set while the test sets remain the same (consisting only of the annotated data). In the following, we refer to these approaches as BERT-1, BERT-2 and BERT-3 respectively. Each model is fine-tuned according to the default recommendations of BERT (Devlin et al., 2018) and trained for 4 epochs, with a train and evaluation batch size of 32, learning rate of 4e-5 and Adam optimizer for stochastic gradient descent. As GPU, a Tesla K80 was used.

4 Results

4.1 Evaluation of the Different Approaches

To evaluate the different approaches we used well established ML evaluation metrics including accuracy, macro (ignoring class distribution) and weighted (including class distribution in the calculation) f1 score.

Table 4 shows the results of the different approaches. For the traditional ML and transformer-based approaches we report averages over all 5 runs.

	SVM	NB	GerVADER	BERT-1	BERT-2	BERT-3
Accuracy	57.6	65.0	52.0	85.8	81.5	**93.3**
F1 Macro	54.5	65.3	52.0	82.1	73.8	**93.4**
F1 Weighted	55.9	65.1	54.0	85.9	81.5	**93.3**

Table 4: Results of the evaluation of the different sentiment analysis approaches. Best results per metric are marked in bold.
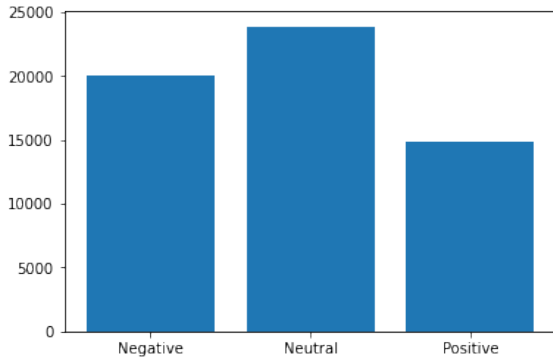

Figure 1: Overall sentiment distribution with 25% positive, 34% negative and 41% neutral tweets.

Figure 2: Percentage distribution of sentiment classes for all parties.

The best overall performance was achieved with BERT-3, followed by BERT-1 as the second best approach. The BERT-3 model reached an accuracy of 93.3%, a macro and weighted f1 score of 93.4 % and 93.3%. Thus, the best run of this model was used to predict the sentiment of the whole corpus of 58,864 tweets. In terms of traditional ML approaches, the Naive Bayes classifier performed best with an accuracy of 65.0% and macro and weighted f1 scores of 65.1% and 65.3% respectively. SVM performed considerably worse with an accuracy of 57.6% and macro as well as weighted f1 scores of 54.5% and 55.9%. GerVADER obtained the worst accuracy score with 52.0% and worst macro and weighted f1 scores with 52.0% and 54.0%.

4.2 Data Analysis

We classified each of the tweets of our overall corpus with the best run of BERT-3 and analyze the results in the following chapter. We focus on party-based and diachronic analysis.

Figure 1 shows the distribution of neutral, positive and negative sentiment predictions for all tweets. Figure 2 gives a more detailed view on the sentiment distribution per party. Overall, most of the tweets were predicted as neutral which is in line with the distribution of the annotated data set. Additionally, there are more negative than positive tweets. Regarding specific parties, the AfD
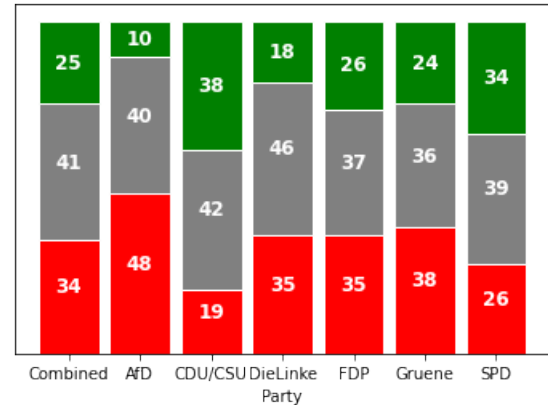
(Alternative für Deutschland) is the party with the highest percentage of negative tweets. Die Grünen has the second most percentage of negative tweets. Additionally, AfD got the lowest count of positive tweets. Parties which were part of the opposition before the election such as AfD, FDP (Freie Demokratische Partei), Die Grünen and Die Linke express more negative sentiment than the two government parties SPD (Sozialdemokratische Partei Deutschlands) and CSU/CDU who indeed have the highest percentage of tweets classified as positive.

For semantic analysis, we looked at word clouds for the different sentiment classes after stop words removal. The word clouds for the overall corpus - Figure 3 and Figure 4 - as well as further term frequency analysis that can be found in our github repository, show that topics like "Corona", "lockdown", "Afghanistan" or "Klimawandel" (German for "climate change") are often mentioned in negative tweets. Positive tweets, however, frequently treat acceptance speeches with words like "Danke" (German for "Thanks"). Additionally, they often include mentions of the own party that the specific account represents. In negative tweets, there are regularly mentions of competing parties.

For diachronic analysis, we calculated a mean sentiment value by assigning -1 to negative, +1 to positive and 0 to neutral tweets. We then summed

Figure 3: Word cloud of negative tweets with all parties combined.

Figure 4: Word cloud of positive tweets with all parties combined.

the values for all tweets of a month per party and calculated the average. The lower the number the more negative, the higher the more positive. Figure 5 shows the mean sentiment per month of the different parties in 2021, with the dashed line symbolizing the election month. First, the figure shows that each party has nearly the same tops and valleys. It can be seen that there is a decrease in sentiment from June to August over all parties. This sentiment decrease turns around before the election in September, where all parties increased their mean sentiment. Surprising winners like FDP got a strong increase also after the election, whereas election losers like AfD or Die Linke got a sentiment decrease after the election.

To present more detailed results shortly before and after the election, Figure 6 shows the average sentiment value of each party's tweets over a 6-week period before and after the election on Sept. 26, 2021. For the average sentiment of all parties, there is a noticeable drop for mid to late August. However, the average sentiment of all parties increased significantly one week before the election. For the parties CDU/CSU, SPD, Die Grünen, this trend remains until one week after the election be-

fore the average sentiment drops again. For the parties FDP and AfD, sentiment remains roughly the same in the week after the election, while the average sentiment of the party Die Linke drops immediately after the election. A rise in sentiment can be seen again towards the end of October and the beginning of November.

5   Discussion

Considering the performance of the sentiment analysis approaches, results of the current state-of-the-art are confirmed with transformer-based models outperforming other approaches and the best model achieving an accuracy of 93% in a three class setting. However, in regards to the traditional machine learning approaches, please note that we did not include the "GermEval 2017" data set for training as we did in the BERT setting. The lexicon-based approach performs worst which is due to the fact of very bad recall values for the neutral class. Investigating the results of the different BERT approaches, we see that a combination of the "Germeval 2017" data and our data set for training achieves the best results (BERT-3) which proves that more data of the same domain for training is beneficial for overall performance.

Considering the analysis of the tweet classification on the overall corpus, we identified a predominance of neutral sentiment followed by negative sentiment for the overall distributions of the entire year. The higher frequency of negative sentiment compared to positive may be due to the period in which we collected the tweets. In 2021 the Covid pandemic posed major challenges to everyday life and was present all over the media. As the decisions of the government in dealing with the virus were often much disputed by the parties, this may explain the overall negative sentiment. This can be seen by inspecting the word clouds of negative tweets from the different parties. The overall word cloud for the negative tweets (see Fig. 3) indeed contains the word "Corona" in contrast to the positive word cloud for which this word is often missing.

Our results regarding differences between reigning parties and the opposition are contrary to research by Costa et al. (2021). They noticed that parties at the opposition had the most positive sentiment profile. We observed a more negative overall sentiment by the opposition parties AfD, Die Linke, Die Grünen and the FDP in comparison to the reign-
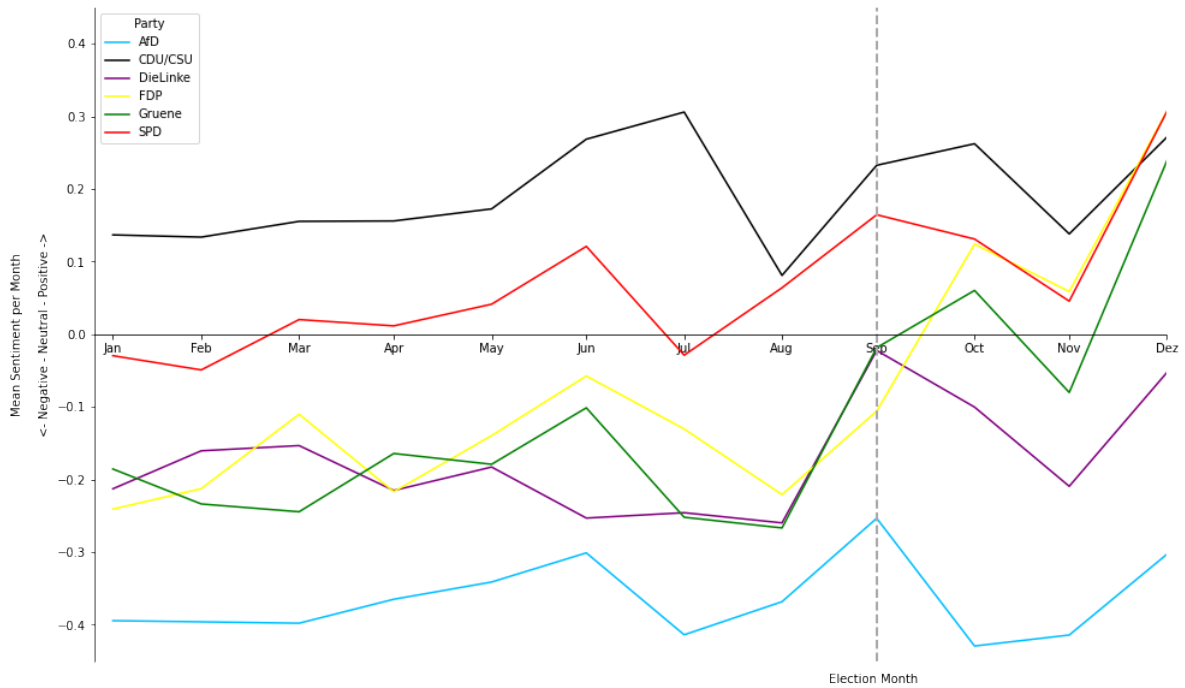
Figure 5: Mean sentiment per month for the political parties over the whole election year.
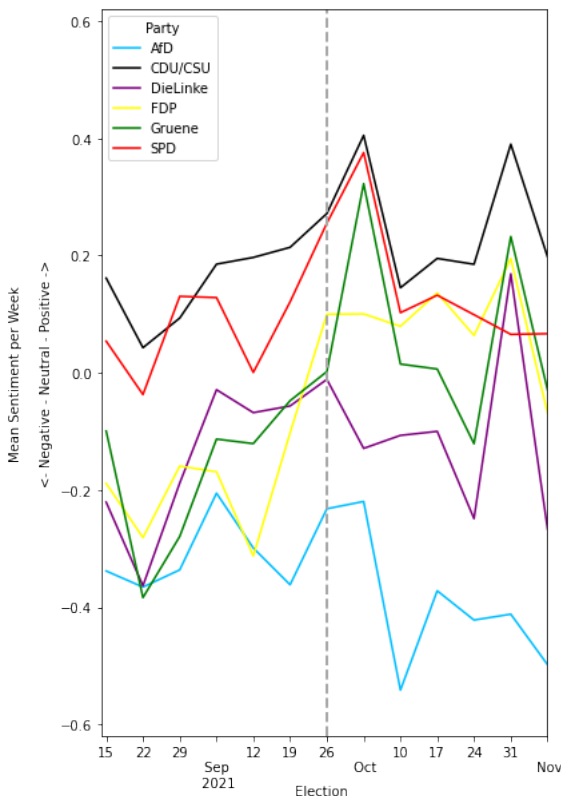

Figure 6: Average sentiment per week for the political parties in the 6-week period before and after the election.

ing parties rather consistently throughout the year with major shifts appearing after the election with the new reigning parties becoming more positive (see fig. 5).

Next to the general distributions, we also investigated sentiment progressions throughout the year. The first shift of sentiment in figure 5, which occurs for almost all parties in July could be explained with the flood disaster in west and middle Europe. It posed tremendous challenges to the country and a lot of people were hurt, lost their homes or died due to the catastrophe. In August, all parties except AfD and SPD had one of their lowest mean sentiment. One reason for this could be the the withdrawal of American troops from Afghanistan which has been heavily debated. One indicator of this is the vocabulary used in negative tweets by all parties in August. Tweets often refer to "Afghanistan," "Kabul," "Taliban" or "Ortskräfte" (German for "local forces"), which leads to the conclusion that topics related to troop withdrawals in Afghanistan were often criticized by the parties.

Looking at the period of a few weeks around the election, several sentiment changes are noticeable (see fig. 6). Towards the election week the sentiment of all parties increased again after the rather low average sentiment of July and August. If we compare the changes in sentiment in detail for the week after the election and in context with the

results of the election (see table A.1), we identify that for the clear winners and losers of the election, such as the SPD, Die Grünen (both winners) and Die Linke it is also reflected in their sentiment trend. For those parties for which the proportional change in votes tended to be small, no major changes in sentiment can be observed. Only the CDU/CSU contradicts this pattern: the party records the highest percentage loss of all parties, 8.8 %, but still shows a strong increase in sentiment. This may be due to the optimistic attitude of the CDU/CSU towards the emerging opportunities of once again belonging to the opposition rather than the government-forming parties after a long period of time.

After the average sentiment of the parties went back to previous levels in mid-October, the next burst of positive sentiment towards the end of October and the beginning of November of some parties can be explained by the fact that the formation of a coalition of the governing parties was finalized. It has to be kept in mind that the new government constellation wasn't build directly after the election. The new government constellation with the SPD, FDP and Die Grünen are ruling just since November. In addition, the first session of the Bundestag of the new election period was held and a new president for the Bundestag was chosen. This is reaffirmed with the general vocabulary used between the last week of October and the first week of November. Examples are an increasing use of words and phrases like "Herzlichen Glückwunsch" (German for "congratulations"), "Bundestagspräsidentin" (German for "President of the Bundestag") and "Demokratie" (German for "democracy"). In autumn, it can be seen that the mean sentiment of most parties was on a lower level again, most likely caused by stronger Covid restrictions and more infections in Germany. However, the sentiment of all parties rose to the end of the year with events like Christmas and New year's Eve.

While our work provides in-depth insight on the sentiment of political parties before, during and after the German federal election, there are certain limitations we want to approach in future work. First, we only annotated a small subset of the overall corpus and achieved mediocre agreement among annotators. We currently plan further annotation studies with an extended annotation manual and guided training annotations to improve upon this

problem. Furthermore we intend to discuss examples with low agreement to investigate this problem and we will annotate on a more fine-grained level marking words and word sequences to get a better understanding of the sentiment expression in the tweets and explore other prediction approaches. More annotation are beneficial for more precise evaluations and can improve the performance of our models.

On a methodological level, while an accuracy of 93% represents current state-of-the-art results in sentiment analysis in German (Chan et al., 2020), there is room for improvement. We see potential in further pretraining the language model with texts of political Twitter as recommended in the research area of domain adaptation of language models (Gururangan et al., 2020). Furthermore, the exploration of more sophisticated emotion categories instead of basic sentiment could lead to further more fine-grained insights. Indeed, recent experiments in the branch of emotion classification for German texts (Schmidt et al., 2021b,c) show the possibilities of the application of transformer-based models for multi-class emotion classification. We intend to integrate emotion annotation in our annotation process as well.

Please also note that we only investigated a subset of party representatives and that the selection as well as Twitter overall do not represent the entire party and its political dissemination, especially in lights of different parties pursuing different goals on Twitter or even having varying emphasize considering the usage of Twitter. It is also noteworthy that Twitter is not as popular in Germany as in other countries. According to current surveys only 10% of Germans use Twitter regularly[5] compared to 23% of U.S. adults.[6] Thus the implications and the importance of Twitter for political parties are limited. Nevertheless the importance of Twitter grows in Germany as well and we intend to build upon our research as described to further gain insights about the influence and development of sentiment of German political actors.

---

[5]https://de.statista.com/statistik/daten/studie/171006/umfrage/in-anspruch-genommene-angebote-aus-dem-internet/
[6]https://www.statista.com/statistics/232818/active-us-twitter-user-growth/

References

Adam Bermingham and Alan Smeaton. 2011. On using twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, pages 2–10.

Widodo Budiharto and Meiliana Meiliana. 2018. Prediction and analysis of indonesia presidential election from twitter using sentiment analysis. *Journal of Big data*, 5(1):1–10.

B Chan, S Schweter, and T Möller. 2020. German's next language model. arxiv. *arXiv preprint arXiv:2010.10906*.

Hasna Chouikhi, Hamza Chniter, and Fethi Jarray. 2021. Arabic sentiment analysis using bert model. In *International Conference on Computational Collective Intelligence*, pages 621–632. Springer.

Carlos Costa, Manuela Aparicio, and Joao Aparicio. 2021. Sentiment analysis of portuguese political parties communication. In *The 39th ACM International Conference on Design of Communication*, pages 63–69.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

James N Druckman, Cari Lynn Hennessy, Martin J Kifer, and Michael Parkin. 2010. Issue engagement on congressional candidate web sites, 2002—2006. *Social Science Computer Review*, 28(1):3–23.

Zulfadzli Drus and Haliyana Khalid. 2019. Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161:707–714.

Shihab Elbagir and Jing Yang. 2019. Twitter sentiment analysis using natural language toolkit and vader sentiment. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, volume 122, page 16.

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.

Deen Freelon. 2017. Campaigns in control: Analyzing controlled interactivity and message discipline on facebook. *Journal of Information Technology & Politics*, 14(2):168–181.

MP Geetha and D Karthika Renuka. 2021. Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model. *International Journal of Intelligent Networks*, 2:64–69.

Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans Joachim Böhme. 2020. Training a broad-coverage german sentiment classification model for dialog systems. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 1627–1632.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. *arXiv:2004.10964 [cs]*. ArXiv: 2004.10964.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

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

Yassine Ait Jeddi and Soufiane Bengadi. 2022. Scweet. a simple and unlimited twitter scraper with python. https://github.com/Altimis/Scweet. Accessed on 2022-06-07.

Anna Jurek, Maurice D Mulvenna, and Yaxin Bi. 2015. Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, 4(1):1–13.

Akrivi Krouska, Christos Troussas, and Maria Virvou. 2016. The effect of preprocessing techniques on twitter sentiment analysis. In *2016 7th international conference on information, intelligence, systems & applications (IISA)*, pages 1–5. IEEE.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Bing Liu. 2015. Sentiment analysis: mining opinions, sentiments, and emotions.

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 (CSUR)*, 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.

Anna-Marie Ortloff, Lydia Güntner, Maximiliane Windl, Thomas Schmidt, Martin Kocur, and Christian Wolff. 2019. Sentibooks: Enhancing audiobooks via affective computing and smart light bulbs. In *Proceedings of Mensch Und Computer 2019*, MuC'19, page

863–866, New York, NY, USA. Association for Computing Machinery.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. Pre-trained Models for Natural Language Processing: A Survey. *arXiv:2003.08271 [cs]*. ArXiv: 2003.08271.

T. C. Rajapakse. 2019. Simple transformers. https://github.com/ThilinaRajapakse/ simpletransformers.

Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. Sentiws-a publicly available german-language resource for sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Joni O Salminen, Hind A Al-Merekhi, Partha Dey, and Bernard J Jansen. 2018. Inter-rater agreement for social computing studies. In *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 80–87. IEEE.

Thomas Schmidt, Marlene Bauer, Florian Habler, Hannes Heuberger, Florian Pilsl, and Christian Wolff. 2020a. Der einsatz von distant reading auf einem korpus deutschsprachiger songtexte. In Christof Schöch, editor, *DHd 2020: Spielräume; Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts; Universität Paderborn, 02. bis 06. März 2020*, pages 296–300. Paderborn, Germany.

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, Manuel Burghardt, Katrin Dennerlein, and Christian Wolff. 2019a. Sentiment Annotation for Lessing's Plays: Towards a Language Resource for Sentiment Analysis on German Literary Texts. In Thierry Declerck and John P. McCrae, editors, *2nd Conference on Language, Data and Knowledge (LDK 2019)*, pages 45–50. Leipzig, Germany.

Thomas Schmidt, Johanna Dangel, and Christian Wolff. 2021a. Senttext: A tool for lexicon-based sentiment analysis in digital humanities. In Thomas Schmidt and Christian Wolff, editors, *Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021)*, volume 74, pages 156–172. Werner Hülsbusch, Glückstadt.

Thomas Schmidt, Katrin Dennerlein, and Christian Wolff. 2021b. 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. 2021c. 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*.

Thomas Schmidt, Philipp Hartl, Dominik Ramsauer, Thomas Fischer, Andreas Hilzenthaler, and Christian Wolff. 2020b. Acquisition and analysis of a meme corpus to investigate web culture. In *15th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2020, Conference Abstracts*, Ottawa, Canada.

Thomas Schmidt, Florian Kaindl, and Christian Wolff. 2020c. 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, Miriam Schlindwein, Katharina Lichtner, and Christian Wolff. 2020d. Investigating the relationship between emotion recognition software and usability metrics. *i-com*, 19(2):139–151.

Thomas Schmidt, Brigitte Winterl, Milena Maul, Alina Schark, Andrea Vlad, and Christian Wolff. 2019b. Inter-rater agreement and usability: A comparative evaluation of annotation tools for sentiment annotation. In *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft (Workshop-Beiträge)*, pages 121–133, Bonn. Gesellschaft für Informatik e.V.

Skipper Seabold and Josef Perktold. 2010. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

Parul Sharma and Teng-Sheng Moh. 2016. Prediction of indian election using sentiment analysis on hindi twitter. In *2016 IEEE international conference on big data (big data)*, pages 1966–1971. IEEE.

Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4.

Karsten Tymann, Matthias Lutz, Patrick Palsbröker, and Carsten Gips. 2019. Gervader-a german adaptation of the vader sentiment analysis tool for social media texts. In *LWDA*, pages 178–189.

Bhagyashri Wagh, JV Shinde, and PA Kale. 2018. A twitter sentiment analysis using nltk and machine learning techniques. *International Journal of Emerging Research in Management and Technology*, 6(12):37–44.

Hao Wang, Doğan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 system demonstrations*, pages 115–120.

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. *Proceedings of the GermEval*, pages 1–12.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 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.

A Appendix

A.1 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 per party for the election years
2017 and 2021.

A.2   Twitter Accounts from Data Acquisition

AFD	Links	SPD	Grüne	FDP	CDU	CSU
@Alice_Wei del 138k	@SWagenk necht 518k	@Karl_Laut erbach 770k	@cem_oez demir 290k	@c_lindner 552k	@jensspa hn 279k	@Markus_ Soeder 341k
@Joerg_Me uthen 76k	@GregorG ysi 439k	@HeikoMaa s 460k	@GoeringE ckardt 202k	@MaStrackZ i 46k	@ArminLa schet 188k	@DoroBaer 103k
@Beatrix _vStorch 68k	@katjakippi ng 130k	@OlafSchol z 324k	@JTrittin 115k	@MarcoBusc hmann 46k	@Friedrich Merz 179k	@andreass cheuer 63k
@Gofffried Curio 37k	@DietmarB artsch 82k	@KuehniKe v 323k	@Konstanti nNotz 85k	@Konstantin Kuhle 44k	@JuliaKlo eckner 74k	@Manfred Weber 54k
@MalteKau fmann 36k	@anked 43k	@larsklingb eil 116k	@RenateK uenast 77k	@johannesv ogel 38k	@n_roettg en 68k	@DerLenz MdB 10k
@JoanaCot ar 30k	@b_riexing er 41k	@hubertus_ heil 108k	@Ricarda_ Lang 65k	@Wissing 32k	@PaulZie miak 58k	@hahnflo 9k
@Tino_Chr upalla 21k	@jankorte mdb 34k	@EskenSas kia 101k	@KathaSc hulze 37k	@Lambsdorff 27k	@groehe 49k	@smueller mdb 9k
@StBrandn er 23k	@Janine_ Wissler 37k	@Ralf_Steg ner 64,9k	@BriHasse lmann 37k	@ria_schroe der 23k	@HBraun 39k	@DaniLudw igMdB 8k
@GtzFrmmi ng 17k	@SevimDa gdelen 35k	@Karamba Diaby 55,6k	@nouripour 29k	@LindaTeute nberg 23k	@rbrinkha us 30k	@ANiebler 6k
@PetrBystr onAFD 17k	@Susanne Hennig 29k	@MiRo_SP D 39k	@MiKellner 28k	@f_schaeffle r 20k	@tj_tweets 17k	@MarkusFe rber 5k

Figure 7: Ten biggest user user accounts of all parties used for the acquisition of tweets.

AFD	Links	SPD	Grüne	FDP	CDU	CSU
@AfD 173k	@dieLinke 350k	@spdde 417k	@Die_Gruen en 649k	@fdp 414k	@CDU 378k	@CSU 229k
@AfDimBund estag 68k	@Linksfraktio n 108k	@spdbt 217k	@GrueneBun destag 186k	@fdpbt 39k	@cducsubt 166k	
@AfDBerlin 19k	@dielinkeber lin 19k	@jusos 77k	@gruene_jug end 76k	@fdp_nrw 28k	@Junge_Union 79k	

Figure 8: Three biggest main accounts of all parties used for the acquisition of tweets.