Authorship Analysis on Twitter during the Bundestag Election 2017:

Targeting Tweets from AfD and Left Politicians.

Computational Forensic Linguistics

Prof Gerold Schneider, Dr Manfred Klenner, Ahmet Uluslu

Nathalie Guibert

18-796-516

Table of Contents

Introduction

With nearly five billion people, dispersed on mainly fifteen Social Media platforms today, Twitter poses as the 12th largest platform worldwide (Kepios 2023), with nowadays more than 540 million compared to roughly a bit more than half of this number in 2017 (World of Statistics 2023). Serving as a micro-blog, this website attracts anyone who would like to share their daily life in less than 140 (since November 2017: 280) characters (Character Counter).

This short format does not only support reactive and fast-changing discussions but also attracts politicians to voice their slogans and opinions in a condensed and effective manner, as the prominent example of Trump in the 2016 US Presidential Election has shown (Buccoliero et al. 2020: 109ff.). This is reinforced by algorithms, which reinforce specific political content on Twitter (Huszár et al. 2021). Already in 2009, the platform played a role in the German National Election. While its prominence was still mainly reserved for "[…] bloggers and other technologically literate early adapters […]" (Jürgens & Jungherr 2015: 485), during the German Federal Election 2017 a majority of German politicians used the platform (Schmidt 2017).

However, fake accounts are an increasing problem for Social Media: They might share fake news, and may alter public opinion. This was for example the case during the 2016 US Presidential Election, during which Russian fake profiles "[…] spread misinformation and politically biased information" (Badawy et al. 2018: 258) or fake accounts pretending to be politicians, as happened when Elon Musk, Twitter's owner, provided the chance of users getting verified, just by paying eight dollars a month (Moran 2022).

Figure 1: Fake accounts pretend to be politicians spreading fake statements .[1]

[1] https://images.app.goo.gl/RL7tEvivV95ujkgT9

In order to detect such pseudo users, scholars have used methods that are known as Stylometry in authorship analysis (Cavalcante et al. 2014: 399): Support Vector Machines, neural networks (Khaled et al. 2018) and Naïve Bayes (Erşahin 2017) for instance. As a Tweet even requires more refined methods, due to its short length and thus, limited linguistic features, different approaches have been tested (Rocha et al. 2017).

In this paper, I research how and if Beatrix von Storch, Frauke Petry, and Sahra Wagenknecht, candidates for the German Bundestagswahl 2017, differ in their stylometric features making their styles distinct.

The paper is structured as follows: First, I summarise key literature concerning the use of authorship analysis on Social Media, predominantly on micro platforms, such as Twitter and also compare the usage and efficiency of different classifiers, as well as posing my research question. Secondly, I proceed to describe the Twitter data that I have used, as well as explain the features and classifiers in the methods section. Finally, I display the results and discuss them, give a conclusion and draw lessons for future research.

Literature Review

Based on the influence of fake accounts and the wrongful news they often spread, such as the example of the Russian influence in the 2016 US Presidential Election and the concrete fake Tweets by two accounts pretending to be Tony Blair and George W. Bush, the importance of authorship analysis remains high. In this literature review, I first discuss examples of Authorship Analysis on Twitter using various features, then I investigate what kinds of features and classifiers have been used to research authorship on Social Media – with a focus on Twitter. Lastly, I showcase an example of linguistic research on Tweets during the German Election 2017, however, distinguishing party leaning instead of authorship, which is where I see my contribution to the literature.

Authorship Analysis on Twitter

Petrik and Chuda analyse Twitter Feeds with celebrity accounts using a random forest classifier and the feature term frequency–inverse document frequency (tf-idf). They tried establishing the "[…] author's occupation, birthyear, fame and gender" (Petrik & Chuda 2019). After preprocessing, which included removing handles, substituting links, changing emojis to their meaning, and standard text preprocessing, they use the "Synth CSOB etic Minority Oversampling Technique (SMOTE)" and "Tomek links" to balance the dataset, as the feeds are of unequal length (Petrik & Chuda 2019). After extracting tf-idfs and reducing them, they used a random forest classifier to predict the characteristics of the famous Tweeters, ensuing an unsuccessful trial with a "convolutional recurrent neural network". Petrik and Chuda's analysis resulted in mixed f1 scores: This was true for the kind of occupation and also for the socioeconomic variables. None of the data reached good results, however (Petrik & Chuda 2019).

Cavalcante et al. on the other hand researched improvements that would increase the velocity of algorithms of authorship analysis and reason their study with the increasing need to investigate "cybercrimes" (2014: 399). In their research, they analyse Twitter data, for which they use the novel feature "character n-grams" and word unigrams" (Cavalcante et al. 2014: 400). The authors propose a technique that would fasten up "[…] efficiency one hundred to a thousand times faster than state-of-the-art counterparts" (Cavalcante et al. 2014: 399), therefore using the "Power Mean SVM" classifier, which previously achieved good performance and accuracy for large datasets (Cavalcante et al 2014: 400) on ten runs using "10-fold cross-validation" (Cavalcante 2014: 403). After keeping only English Tweets that have a length of n > 2, removing Retweets and similar basic preprocessing as Petrik and Chuda, they end up with 10 million Tweets by 10,000 authors they use for analysis. For the task of "Search-Space Reduction", in order to minimise the group of potential authors of a Tweet, Cavalcante et al, make use of a Cumulative Matching Curve. This approach leads to a significant reduction of the search space, and, therefore, helps to rank potential authors (Cavalcante et al. 2014: 405).

Comparison of Feature Extraction Methods on Social Media

In comparing different classifiers, namely Naïve Bayes and Logistic Regression, Aborisade and Anwar try to classify nearly 50,000 Tweets in a dozen known and a dozen unknown authors, depending on whether they are "[…] prominent people in the field of politics and media journalism" (Aborisade & Anwar 2018: 271). They follow a similar preprocessing routine as Cavalcante et al. by removing Retweets, as well as "[…] hyper-links, stop words, and outliers that do not add value to the classification of texts" (Aborisade & Anwar 2018: 271). As features, they make use of bags of words and tf-idf and end up achieving an accuracy of 89.8% for Naïve Bayes and 91.1% for the Logistic Regression with Conditional Maximum Likelihood (Aborisade & Anwar 2018).

Li et al. investigated over 9200 Facebook posts and their 30 authors based on their authenticity. Also here, the posts are not long, having a size of 20.6 words on average, which poses a common challenge to authorship analysis, as the previous authors have noted. In comparison, they list previous research which achieved 40.5% accuracy with an SVM and a bag of words approach, as well as 70% with a SCAP approach relying on character n-grams for "authorship identification" on Twitter (Li et al. 2017: 3). In their own analysis, Li et al. use the classifier SVM Light with a "leave-one-out (LOO) cross-validation method […]" – yielding over 650,000 tests in total – as a base (2017: 4), with which they able to get "[…] an average accuracy rate of 79.6% among 30 users" (2017: 1), and also compare their results with other classifiers, such as "[…] k-nearest neighbor, naive Bayes, decision tree, and neural network […]" (2017: 4). They extracted 233 classifiers, which can be subdivided into two subcategories: stylometric and Social Media features, which again, can be aggregated into categories. Tests have shown that Social Media features, such as "emoticons […], abbreviations[…]", capitalisation, punctuation, and lack of personal pronouns" (Li et al. 2017: 4) do only contribute 0.7 per cent in accuracy in the baseline model. All features yield the best accuracy and the lowest standard deviation. Employing the other aforementioned classifiers, they attained the best accuracy with a decision tree (77.4) and with a neural network using three layers the worst (66.8). To determine the superior performance between the two primary classifier candidates, an additional neural network was used, however, it did not outperform both of the others significantly (Li et al 2017).

Furthermore, Rocha et al. review authorship analysis on general long text, short text and Social Media posts and also explain the differences in features and their effectiveness. Generally, it can be said that the classifier SVM beat other methods of classification, such as SCAP, Naïve Bayes etc. The authors then explain different features, ranging from character-level n-grams to specific features that also take into consideration punctuation and emojis. They also employ their own study, in which they first analyse ten million Tweets of 10,000 authors in 2014, for which the authors are known and use mainly PMSVM and Random Forest classifiers. Rocha et al. find out that word 4-grams, as well as character-level 4-grams features, perform the best – particularly in connection with word- and POS-tags n-grams. Concerning classifiers, it was

PMSVM, SCAP and PPM-5 methods that outperformed the classic SVM classifier. With a Cumulative Match Curve, they were also able to define the threshold of minimum authors to get decent accuracy, which is reached with 200 tweets per user. Lastly, they also researched how to pair features to get the best results: Overall, every feature had a significant contribution, however POS N-grams turned out to be the best overall feature set. Also using every feature resulted in a competitive performance time- and space-wise. Then the authors analyse Tweets where the authors are partially not known, thus, splitting 50/50 in known and unknown Tweets for each of the 50 known authors. They have chosen W-SVM, Logistic Regression and RBF SVM as classifiers, and increased the number of unknown authors by 50 each time. Increasing the degree of anonymity leads to a decrease in accuracy. In total, W-SVM Advantage still slightly performed the best, whereas Logistic Regression was the worst out of the three classifiers. In the end, they give an overview of approaches that have been used for authorship analysis, mostly using long texts so far, for example, English prose. For Social Media, most of the authors have used SVM and Naïve Bayes (Rocha et al. 2017).

Previous Studies on Political Authorship in German Elections

Cohrs and Petersen, on the other hand, investigated the party affiliation of politically active Twitter users during the 2017 election of the Bundestag in Germany with nine parties. Making the differentiation between using entire accounts to determine partisanship and unique Tweets, they apply different compositions of features, such as n-grams, PoS tagging and weighting by word frequency. Using a probabilistic classifier, they were able to attain an accuracy of 72% for entire accounts with higher-order Markov chains and 36% for individual Tweets with unigrams. In comparison, a "[…] random guessing baseline system" had 11% accuracy for both modi (Cohrs & Petersen 2019: 2).

Given this overview of the existing literature, one may see that a lot of approaches have been tested. Whereas SVM seems to be by far the most used and the most successful, other, simpler approaches, such as Naïve Bayes have been tried to handle Social Media data. It appears that authors have a problem with short texts due to restrictions of Social Media platforms, which they tried to handle with extensive preprocessing and specific features. Often, there is also the question of how to reduce complexity, which, however, shall not be the issue of this paper. Only a little research has been done on Social Media posts in the context of politics, and more specifically, elections in non-English speaking countries, which is why I pose the following research question: *Which classifiers and features are the most effective in distinguishing German politicians?*

Data

Overview Data

Contrary to most previous approaches, which did not collect Social Media data in a specific context, I analyse Tweets of three German politicians – two of them belonged to the "AfD" at the time (@Beatrix_vStorch, @FraukePetry) and one during the time frame to "the Linke" (SWagenknecht). The posts have been collected manually, due to the restrictions of the Twitter API, during the election period from 1.4.2017 until 1.10.2017 via Twitter's advanced search (Example of a search term: (from:FraukePetry) until:2017-10-01 since:2017-04-01). My choice of the analysed users is content-based reasoned – as Wagenknecht of "the Linke" had been compared to spreading the same rhetoric as party AfD (Busemann 2016). However, in this analysis, I focus on the Stylometry of the authors in order to find out, if similar content may have also contained related Stylometry.

Table 1: Number of Tweets before preprocessing

User	Beatrix_vStorch	FraukePetry	SWagenknecht
Number of Tweets	1426	362	185

Preprocessing Data

According to previous literature, preprocessing is a necessary task for Social Media posts and particularly, to remove bias. In the Twitter corpus that I have collected, links, emojis, user handles and hashtags can be found, which may lead to considerable biases as they provide inferences as to who the author is. For instance, if a politician adds a hashtag of their own party (e.g. #AfD), this might help the classifier to identify the author more easily, without it indicating anything really about the linguistic features of the author (Rocha et al. 2017: 20). This why I consequently replace hashtags by "meta tags" "HT", URLs with "URL", e-mails with EMAIL, and references to users as indicated by @ by “REF”, as done by the other authors. This is easily doable by the search and replace function in Excel. More generally, I remove every Emoji by hand, as too many different kinds have been used – e.g. also Emoticons using special characters. Furthermore, I remove Tweets that are shorter than 4 words (or HT and REF), posts marked as Retweets and Tweets that are largely not in German, for consistency and further processing reasons – for example, PoS Tags. This is compliant with previous research. After preprocessing, I am left with the same amount of Tweets for Sahra Wagenknecht, 355 for Frauke Petry, and 1262 for Beatrix von Storch.

Methods

Now that I have the data, methodological approaches have to be chosen to apply to the linguistic corpus. Based on previous literature, characteristics, such as “character-level n-grams, word-level n-grams, part-of-speech n-grams, and diverse lexical and syntactic statistics as features have been used (Rocha et al. 2017: 14).

Letting me be guided by previous literature, I investigate their used features and classifiers and also combine them in new ways. For instance, in Rocha et al.’s first experiment, they discovered that word 4-grams and character 4-grams are features that have been fruitful in their authorship classification. To explain both features, I would briefly try to make an example: A character 4-gram of the misspelt word “mispell” would be the following strings: “misp”, “ispe”, “spel”, “pell” that still have enough overlap with the correctly written word “misspell”. With this approach, the effect of irregularities, such as spelling mistakes – that do not happen as a pattern – can be mitigated. The word 4-grams work similarly to the characters, thus, with entire words – strings limited by white spaces or punctuation. Next, I investigate the usage of PoS tags as bigrams. This yields the following applied tags for an example sentence: “The(DET) house(NOUN) is(VERB) green(ADJ).”. As my posts are in German, I use spaCy’s German language model “de_core_news_sm”. Additionally, as a statistic feature, I make use of tf-idf, to determine the relevance of certain tokens, given a corpus. In my analysis, I use the aforementioned explained features uniquely and also all combined as a Feature Set.

Furthermore, I was able to establish that the Support Vector Classification (SVC) of the module Support Vector Machine, serves as the most popular classifier for Social Media posts. Authors, such as Aborisade and Anwar, have made use of Naïve Bayes, which often serves as a general baseline model, and Logistic Regression. However, the application of the latter classifier on Social Media data has not been widespread so far, and also in using it, has resulted in mixed results. These classifiers are all part of supervised learning techniques.

After splitting in train and test data on an 80/20 basis, I oversample the minority classes of @FraukePetry and @SWagenknecht to the size of the majority class @Beatrix_vStorch to address the class imbalance. Furthermore, I make use of the Stratified K Fold algorithm, using five folds, in order to retain the original proportion of samples.

Results

In this section, I briefly present the results from the classification. As mentioned before, I have used Naïve Bayes, Support Vector Machine and Logistic Regression as classifiers, as well as the following features: Word 4-grams, PoS bigrams, Character 4-grams, Tf-idf, and all features together.

Table 2: Naive Bayes Results

Accuracy	Word 4-grams	PoS bigrams	Character 4-grams	Tf-idf	All features
Mean (rounded)	0.858	0.768	0.956	0.909	0.962
Standard Deviation (rounded)	0.155	0.014	0.008	0.005	0.006

Table 3: SVM Results

Accuracy	Word 4-grams	PoS bigrams	Character 4-grams	Tf-idf	All features
Mean (rounded)	0.981	0.931	0.974	0.946	0.979

Standard Deviation (rounded)	0.004	0.011	0.005	0.011	0.006

Table 4: Logistic Regression Results

Accuracy	Word 4-grams	PoS bigrams	Character 4-grams	Tf-idf	All features
Mean (rounded)	0.980	0.904	0.975	0.929	0.978
Standard Deviation (rounded)	0.006	0.009	0.006	0.013	0.007

Based on previous literature, I have chosen accuracy as the main means of comparison. Overall, my study yields rather high results. I shall address this in the discussion. Furthermore, the predictions of the SVM are the most accurate. This confirms the findings of previous literature (see Rocha et al 2017). Contrary to Rocha et al., significantly worse results for Logistic Regression could not have been reproduced, and this classifier performs even better than Naive Bayes. Investigating the results of which features and feature sets produced the best results, the last model, including all the features performs the best with an accuracy of 0.973. However, this is not far from the second-best performer with 0.968 for the character 4-grams, which also have been successful in previous studies. The worst outcome was on average given by the PoS bigrams (0.868), where it achieved the lowest accuracy with the Naive Bayes classifier (0.768). Whereas the standard deviation was generally relatively small, it was particularly large for word 4-grams with Naive Bayes as classifier (0.155).

Discussion

Following the demonstration of the results, I would like to discuss them now and point out potential shortcomings.

First, what first strikes the eye when looking at the previously presented results are the general high values of accuracy. This might be due to two reasons: The author's linguistic features are very distinct or (and) there still might be some bias in the datasets. I argue that the latter might be the case, although having previously followed carefully the approaches of literature to avoid such tendencies in the first place, for example, by replacing handles, hashtags, links, emails and references to the own party. Perhaps, mentions of other parties, which are not hashtags, might need to be renamed, too. However, due to the specific period in the political agenda of Germany, the three Tweeters, Beatrix von Storch, Frauke Petry and Sahra Wagenknecht, have often used short and similarly structured attacks aimed at their closest opponents – the government or the parties they are tied with in the electoral battle – as part of their rhetoric and stylistic devices.

Concerning the availability of data, Rocha et al. have argued that one needs to have 200 Tweets per user, in order to get a "[…] decent accuracy" (2017: 22ff.). This is, however, not true for my data. Using oversampling for the minority classes, has yielded good predictions. However, the effects of oversampling on performance are still under-researched (García et al. 2020). In order to avoid too significant consequences, perhaps more genuine data could be generated for other analyses. This may include a broader authorship, thus, even more authors, to validate results on a larger level, as well as, a longer time period. For instance, if I aim to compare if similar content also implies similar linguistic features, then investigating posts for a longer period than just an election might be fruitful since I expect more profound, and, therefore, more lexically nuanced posts than just party election slogans for the most part. Whereas Twitter might not be as widespread in Germany as in other countries, it might make sense to set a primary focus on the US, in which politics are widely discussed on Twitter or to also combine data from different Social Media platforms.

In my code, I have also created confusion matrices for every classifier and feature, and summed up the predicted classified Tweets and put in a heat map with the actual values. The confusion matrix of the Naïve Bayes model with the feature word 4-grams is hereby particularly interesting. It shows that for this configuration of the model, Beatrix von Storch has also often been misclassified as Frauke Petry and Sahra Wagenknecht equally. This, however, is not often the case, so that misclassifications generally do not happen more often than in the table 5. Table 6 presents the most successful model, clearly indicating the beige pattern of correlation between each author with themselves and, thus, the rest in black, as only little misclassifications are to be found.

Table 5: Confusion Matrix NB - Word4grams

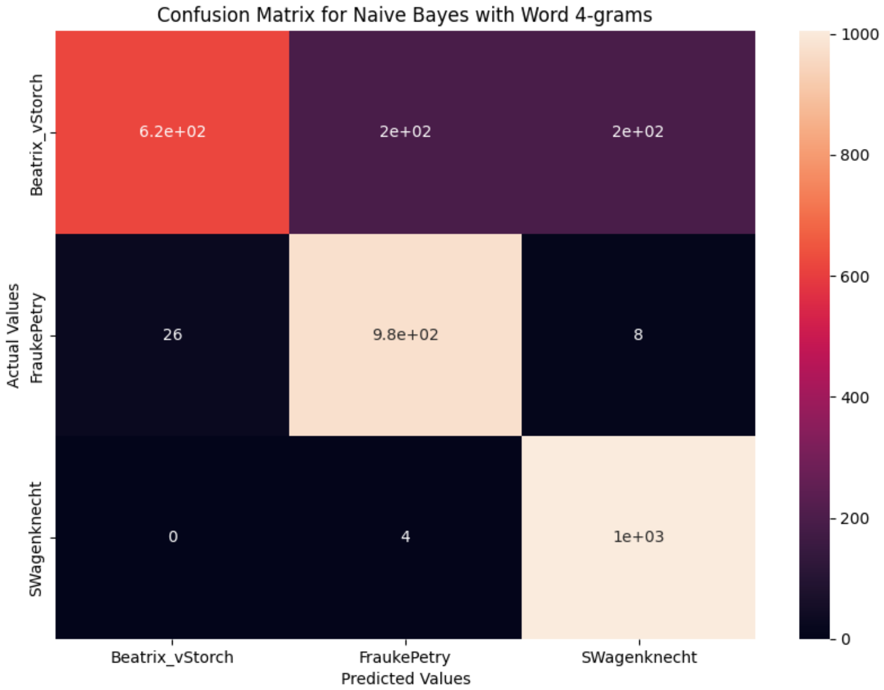
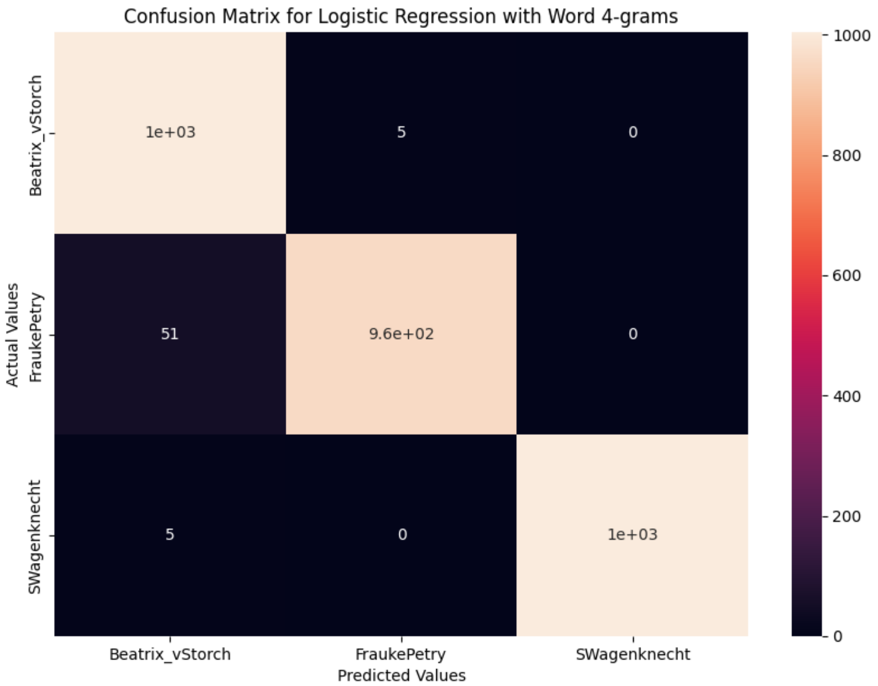


Table 6: Confusion Matrix LR Word4grams



Conclusion

This paper has investigated the accuracy of different features (Word 4-gram, PoS bigrams, Character 4-grams, Tf-idf, and as a combination) using three classifiers (Naïve Bayes, SVC, Logistic Regression) that have also been applied in previous analyses. Whereas I could confirm previous findings indicating that SVC generally outperforms the other classifiers, I cannot declare that this is by a large margin compared to the other classifiers. For instance, I could not reproduce the comparatively poor results for the Logistic Regression classifier, that Rocha et al.'s research yielded (2017). Generally, the model with all the features has proven to be the strongest, however, the accuracy of 4-character gram models is not far off the best configuration.

Given the arguments I made in the discussion concerning potential lingering biases and data size, I would only argue with some slight hesitation that the authors, therefore, are distinct writers – at least in their Stylometry. Therefore, regarding my hypothesis that similar content might indicate similar linguistic features, cannot be confirmed. However, future studies need to address the concerns which I have outlined before. It could be, for instance, a fruitful undertaking to research Stylometry for longer periods of time and – in the political sphere – with more Social Media-affine political actors. Furthermore, politicians with similar values and attitudes could be analysed transnationally, thus, potentially indicating specific stylistics for such actors.

Overall, this study has contributed to the still under-researched area of authorship analysis on Social Media, and, more specifically, in a political context. Given the rising need to counterfight the dangers of fake news and slanders spread by fake profiles, authorship analysis on Social Media platforms becomes increasingly important, particularly on platforms on which it has been easy to create such imitation accounts – as seen in the case of Tony Blair and George W. Bush on page one – or in the anonymous spreading of barely moderated content facilitated as in the case of Telegram, for example. Whereas some have argued that the popularity of Twitter has declined since the takeover by Elon Musk, important lessons might still be drawn from Tweets during specific situations, as the research by Badawy et al. on the interference of Russian fake accounts in the 2016 US election demonstrates (2018). Overall, authorship analysis will increasingly become important in the field, not only for forensics or Social Media analysis but also for AI detection in allegedly human-written texts.

References

Badawy, Adam; Ferrara, Emilio; Lerman, Kristina (2018): Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) , Barcelona, Spain, pp. 258-265. DOI: 10.1109/ASONAM.2018.8508646

Busemann, Hans Edzard (2016): World: FEATURE-Braune Antifa-Torte gegen rote Wagenknecht. https://www.reuters.com/article/idUSKCN0YK0DV/ (last accessed 19.1.2024)

Buccoliero, Luca; Bellio, Elena; Crestini, Giulia; Arkoudas, Alessandra (2020): Twitter and politics: Evidence from the US presidential elections 2016. In: Journal of Marketing Communications, Vol. 26(1), pp. 88-114. https://doi.org/10.1080/13527266.2018.1504228

Cavalcante, Thiago; Rocha, Anderson; Carvalho, Ariadne (2014): Large-Scale Micro-Blog Authorship Attribution: Beyond Simple Feature Engineering. In: Bayro-Corrochano, Eduardo; Hancock, Edwin (eds): Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science, Vol. 8827, Cham: Springer. https://doi.org/10.1007/978-3-319-12568-8_49

Character Counter: Twitter Character Counter. https://charactercounter.com/twitter (last accessed 16.1.2024)

Cohrs, Enum; Petersen, Wiebke (2019): Guessing a tweet author's political party using weighted n-gram models. In: Jarmo Harri Jantunen, Sisko Brunni, Niina Kunnas, Santeri Palviainen, and Katja Västi (eds.): Proceedings of the Research Data and Humanities (RDHum) 2019 Conference: Data, Methods and Tools, number 17 in Studia Humaniora Ouluensia, pp. 101-113.

Erşahin, Buket; Aktaş, Özlem; Kılınç, Deniz; Akyol, Ceyhun (2017): Twitter fake account detection. In: 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, pp. 388-392, DOI: 10.1109/UBMK.2017.8093420

García, Vicente; Sánchez, Josep S.; Marqués, Ana I.; Florencia, Rogelio; Rivera, Gilberto (2020): Understanding the apparent superiority of over-sampling through an analysis of local information for class-imbalanced data. In: Expert Systems with Applications, Vol. 158, pp. 1-19. https://doi.org/10.1016/j.eswa.2019.113026

Huszár, Ferenc; Ktena, Sofia I.; O'Brien, Conor; Belli, Luca; Schlaikjer, Andrew; Hardt, Moritz (2021): Algorithmic amplification of politics on Twitter. In: PNAS, Vol. 119(1). https://doi.org/10.1073/pnas.2025334119

Jürgens, Pascal; Jungherr, Andreas (2015): The Use of Twitter during the 2009 German National Election. In: German Politics, Vol. 24(4), pp. 469-490. DOI: 10.1080/09644008.2015.1116522

Kepios (2023): Global Social Media Statistics. https://datareportal.com/social-media-users (last accessed 8.1.2024)

Khaled, Sarah; El-Tazi, Neamat; Mokhtar, Hoda M. O. (2018): Detecting Fake Accounts on Social Media. In: 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, pp. 3672-3681. DOI: 10.1109/BigData.2018.8621913.

Li, Jenny S.; Chen, Li-Chiou; Monaco, John V.; Singh, Pranjal; Tappert, Charles C. (2017): A comparison of classifiers and features for authorship authentication of social networking messages. In: Concurrency Computat.: Pract. Exper., Vol. 29(e3918). DOI: 10.1002/cpe.3918

Petrik, Juraj; Chuda, Daniela (2019): Twitter Feeds Profiling With TF-IDF. In: Notebook for PAN at CLEF 2019. https://ceur-ws.org/Vol-2380/paper_215.pdf

Rocha, Anderson; Scheirer, Walter J.; Forstall, Christopher W.; Cavalcante, Thiago; Theophilo, Antonio; Shen, Bingyu; Carvalho, Ariadne R. B.; Stamatatos, Efstathios (2017): Authorship Attribution for Social Media Forensics. In: IEEE Transactions on Information Forensics and Security, Vol. 12(1), pp. 5-33. DOI: 10.1109/TIFS.2016.2603960.

Schmidt, Jan-Hinrik (2017): Verbreitung, Aktivität und Informationsquelle: Twitter-Nutzung von Kandidierenden der Bundestagswahl 2017. In: Media Perspektiven, Heft 12, pp. 616-629. https://www.ard-media.de/fileadmin/user_upload/media-perspektiven/pdf/2017/1217_Schmidt.pdf (last accessed 16.1.2024)

World of Statistics (2023) [Twitter] 1 November. Available at: https://x.com/stats_feed/status/1719774296946385149?s=20 (last accessed 16.1.2024)