# Understanding the Political Polarization in Twitter Amidst the COVID19 Pandemic

Lakmal Meegahapola, Farnaz Forooghifar, Negin Safaei, Haeeun Kim
Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland

## ABSTRACT

Diseases may be apolitical by nature, but its influence on society has political implications. This applies to the current outbreak of Coronavirus (COVID19) as well. Its influence varies from their different nationalities, geographies, financial status, social roles, and political affinities. In particular, during crisis situations, political factors become even more crucial because of the way politically motivated information are shared in social media, and the way influential figures and accounts could create politically biased perceptions on the same topic. In this study, we analyzed 144K tweets from 252 twitter accounts from United States that are either politicians, celebrities, or media organizations that have revealed their political affinities. Through our descriptive analysis and statistical analysis, we showed that tweet related features such as hashtags, number of URLs, and whether the tweet was retweeted can be used to discern political affinities as reflected by tweet content. Finally, by using features generated from the text and metadata of tweets, we inferred the political affinity (Democratic or Republican) of the writer of a tweet with an accuracy of 82.73% in a two-class inference task.

## KEYWORDS

covid19, twitter, democrats, republicans, bias, politics, topic modelling, disaster

## 1 INTRODUCTION

The COVID19 pandemic [27] has opened many research opportunities in different fields of research including medicine [28], context-aware mobile app development [31], and data analysis [5, 22]. Besides the medical data that are abundantly made available during this time period, another valuable data that is being produced in large scale is content from social media about COVID19 pandemic. One such promising source of data is Twitter, where many different types of users share their opinions and feelings every day [25]. Among the main challenges that have surfaced in Twitter during this time period is political polarization and spread of misinformation [2].

One effective diversity attribute that shapes the trajectory of social media websites such as Twitter is the *political orientation* of people who tweet [7]. Analyzing data, while taking into account the political affinities, might help to understand if the way that the country is handling the situation is acceptable by its people or not. It would also make it possible to understand the most pressing concerns of people during this time period, before it can result in public anger or protest [26]. Another interesting information is the change in the trend of topics that people are concerned about as the situation evolves. The ability to restrict the search in Twitter to selected users, locations, and topics makes it a powerful source to do data analysis for specific group of users or according to a specific topic.

In this study, specifically, we aim to analyze tweets by a set of twitter accounts from the United States (U.S.) during the time of covid19 pandemic. Our goal is to find the correlation of the level of users' involvement in politics and also their political affinity with their tweets. We have decided to restrict our analysis to three categories of twitter profiles, namely: politicians, celebrities, and media. We used different natural language processing (NLP) techniques to extract novel features from the tweets, and we analyzed the topics that are discussed, the trend of activities of the users, and their sentiment toward this pandemic from the perspective of political affinities. Primary contributions of this study are:

**Contribution 1:** We collect a dataset of 144K tweets during April-May 2020, from 252 accounts of politicians, celebrities, and media organizations in the U.S. which have shown Democratic or Republican political affinities in the past. Then, we conducted a descriptive analysis and a statistical analysis to understand underlying differences between the two political affinities for factors such as hashtags, word choice, favorite counts, retweet counts, sentiment analysis, and topic modelling using techniques such as Latent Dirichlet Allocation (LDA).
**Contribution 2:** We perform a novel task by inferring the political affinity of tweets with an accuracy of 82.83% using text and metadata, in a two-class classification task using Random Forest Classifiers, demonstrating the possibility of measuring political polarization in the U.S. during the COVID19 pandemic.

## 2 RELATED WORK

Related work can be divided into relevant literature on political polarization of the U.S. and disaster information diffusion in social media.

**Political Polarization in the U.S. Social Media.** A number of studies have addressed the political polarization in various social medias. The behavioral difference is visible in how political elites use social media. William et al. [4] found that political elites from the Democratic and Republican parties diffuse their ideology through different tactics of moral contagion. In their study, they stated that during the U.S. presidential election in 2016, the Republican political ideology was more impactfully delivered by using moral-emotional language that is related to religion and patriotism, compared to that of Democratic political ideology. During the presidential campaign, where Twitter played an unusually prominent role, the tweets made by both candidates were used to reach out to supporters, instead of interacting with the opponent [11]. The messages created by political elites are transmitted to and produced public reaction in different ways. In Youtube, during the U.S. presidential campaign in 2008 and 2012, Democratic campaign videos were shared faster and elicited more positive collective emotions,

**Table 1: Dataset Summary**

| Party | Type | # of Accounts | # of Tweets |
|---|---|---|---|
| Democrat | Politician | 50 | 17011 |
| Democrat | Celebrity | 49 | 13413 |
| Democrat | Media | 32 | 47974 |
| Republican | Politician | 49 | 17420 |
| Republican | Celebrity | 45 | 23610 |
| Republican | Media | 27 | 24450 |
| All | All | 252 | 143878 |

while it was the Republican campaign videos that were remembered longer inside the community [12]. During the presidential election in 2016, Democratic voters tweeted more about gay rights than Republicans whereas Republican voters tweeted more about abortion than their liberal counterparts. Furthermore, the overwhelmingly negative sentiment of conversations revealed more "crosstalk" from Democratic leaning users towards Republican candidates, and less vice-versa[20]. In this study, we focus on the political polarization during a pandemic, rather than a regular political event where polarization becomes more apparent than usual.

**Disaster Information Diffusion** Scale-free and small-world property in Twitter has been studied for more than a decade [17]. Rongsheng et al. [9] found that the spread of disaster information diffusion shares the scale-free and small-world property with a looser organizational structure, in the case of Sina-Weibo and two earthquakes in China. In addition, several assessments have been done regarding the disaster response and recovery through sentimental analysis on Twitter data during specific disasters [16, 29]. We extend these works on social media disaster reaction by focusing on the current disaster case [8]. Since pandemic is inherently related to public health, we based our work on the previous studies on health information diffusion via microblogging media, which addressed social media as an important channel to assess health information [6, 13].

## 3 DATASET

### 3.1 Data Collection

In this study, our goal is to examine tweets from Twitter accounts that have previously shown political affinities. To have a broader understanding regarding tweeting patterns, we selected accounts that belong to three categories by considering links to both Democrats and Republicans using information from a plethora of sources. The categories are: (a) politicians, (b) celebrities, and (c) media. Hence, we extracted tweets from 252 Twitter accounts starting April 19, 2020, retrieving 200 tweets per profile per day until the end of May 2020. The limit on tweets per day was chosen to make sure that all the tweets that each account posts was captured in our dataset with the reasonable assumption that each account would not post more than 200 tweets per day. After omitting the duplicates, we obtained around 143878 tweets during this time period. The summary of the dataset is mentioned in Table 1. Moreover, all the selected accounts are from the U.S, and all of the gathered tweets are in English. Furthermore, to find suitable accounts for each category, we primarily relied on two factors: (a) their political affinity according to their posts and categorizations done by several websites [1] and (b) To make sure that the selected accounts are popular among Twitter users, we set the threshold of minimum number of followers to 25K, and we manually checked the number of recent tweets to confirm that the accounts are currently active.

[1]link to a file containing a list of sources

### 3.2 Feature Extraction

**Latent Dirichlet Allocation (LDA) Features.** Topic modeling is one of the most popular and powerful methods in text mining that is used to analyze the relationship between data, its context, and latent data discovery. Topic modeling has been applied in different fields such as linguistic science [23], media analysis [32], and political science [10]. Latent Dirichlet Allocation (LDA) [3] is one of the most popular topic modelling techniques. We use LDA in our study, and for different experiments, we use different number of topics as explained in Sections 5 and 6. Hence, if the number of topics used in the experiment is $\alpha$, we trained a topic model using $\alpha$ topics, and for each row in the dataset, we assigned a probability score for each topic. As an example, if $\alpha$=5, we would create five topics, and add five columns to the dataset corresponding to each topic, hence creating five new features for each data row that has values estimating the probability of a tweet belonging to specific topic.

**Sentiment Features.** We used two sentiment analysis techniques to generate features for our dataset using the text content: (1) VADER [15] - This is a sentiment analysis technique specifically used for social media content. For each tweet, this technique provided four sentiment scores, namely: positive, negative, neutral, and compound. We used these four features in the dataset, and (2) NLTK Sentiment Analysis [1] - NLTK sentiment analysis toolkit provided a sentiment score between -1 and +1 for each sentence in a tweet. We calculated the mean score for each tweet, and included it as a feature in the dataset.

**Hashtag Features.** Hashtag is a keyword which is used to describe the information on a tweet, and it helps users to discover tweets for specific topics. Because of the huge number of tweets per day, using hashtags is a way to organized information and discussion around specific topics or events on Twitter. Hashtags started being used in political issues in events like the 2009 Iran presidential election. At that time, #iranelection was the hottest topic on twitter [24]. As Twitter allows on-the-ground reporting of breaking news and democratic activism, it is suggested that Twitter is a democratic media [33]. In our analysis, we grouped tweets into six categories based on political affinity and type. For each of these groups, we obtained the 10 most common hashtags. We combined this list, and removed duplicates to obtain a list of hashtags for the analysis. This was done to make sure that we provide equal representation to all the different political groups and profile types in the analysis.

**Word Features.** Similar to the technique used in hashtag feature extraction, we extracted 10 words for each of the six groups, and obtained a set of words that have been highly used among democrats and republicans for the analysis.

### 3.3 Dataset Preparation

In the final dataset, we had attributes belonging to several categories, and we call them as feature groups throughout this report. (1) META - this feature group corresponds to tweet specific features

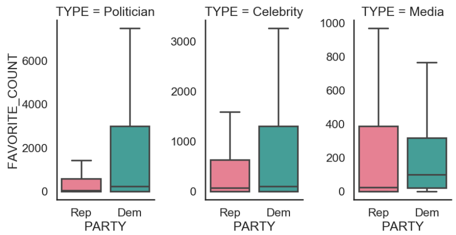


**Figure 1: Box plots for favorite count on tweets between different political parties and account types.**

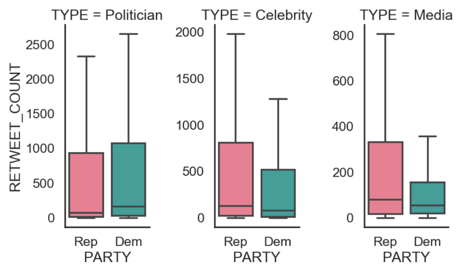


**Figure 2: Box plots for retweet counts between different political parties and account types.**

captured using the API (e.g. number of URLs, whether it is a retweet or not, favorite count, retweet count); (2) HASH - this feature group was based on common hashtags used in the dataset. We used 10 hashtags using the methodology described in Section 3.2. Hence, in the final dataset, for each tweet, we indicated a binary value (0 or 1) based on whether the tweet contained the hashtag or not; (3) WORD - using the methodology described in Section 3.2, we used words that come from accounts with democratic and republican affinities. Similar to hashtags, for each tweet, these features indicated whether the tweet contained the specific word or not; (4) SENTI - for each tweet, we extracted five sentiment scores using two techniques described in Section 3.2; and (5) LDA - we extracted scores for each tweet using $\alpha$ number of topics. We used $\alpha$ ranging from 3 to 150 in different experiments, hence changing the number of features in the dataset during each experiment.

## 4 DESCRIPTIVE DATA ANALYSIS

In this section, we analyze the dataset to understand underlying tweeting patterns from accounts with democratic and republican affinities. We sampled 130K tweets from the original dataset, equally split between the two political affinities. Figure 1 shows a boxplot of favorite counts for different user types and political affinities. Among politicians and celebrities, democrats had more favorites to their tweets in terms of the median, and also the spread. However, when considering media, even though the median was higher for democrats, republican tweets had a higher variation. Moreover, the number of retweets for different groups in Figure 2 show that democratic politicians receive a higher number of retweets on average compared to their republican counterparts. However, for celebrities and media, republican tweets had higher number of retweets comparatively. This suggests that republican media accounts have a wider network with many followers who retweet often.

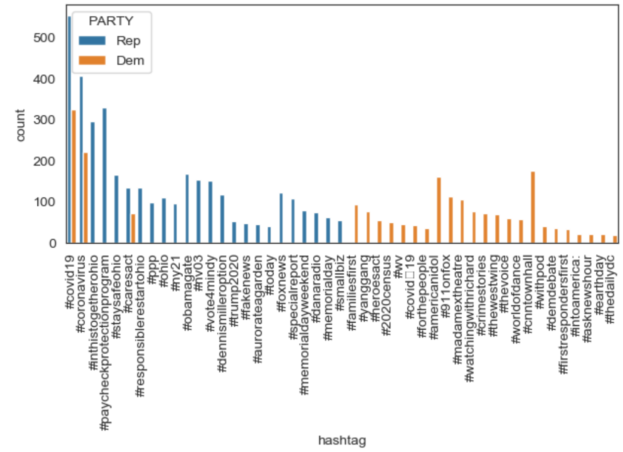


**Figure 3: Barplot for Hashtags from Different Political Affinity and Type Groups.**

To analyze the hashtags and the most common words in tweets, we followed the approach suggested in Section 3.2. We obtained a list of 46 hashtags and 29 words for the analysis. The barplot of hashtags is shown in Figure 3. We observe that apart from three hashtags used by both parties such as #covid19, #coronavirus, and #caresact, other hashtags in the list clearly showed significant differences in terms of usage among tweets from the two political affinities. Republicans used hashtags such as #inthistogetherohio, #paycheckprotectionprogram, #trump2020, #fakenews and #memorialday more often whereas accounts with democratic affinity used hashtags such as #familiesfirst, #heroesact, #madamextheatre, #cnntownhall as shown in Figure 3. We show an analysis of words in Figure 4, and compared to the hashtag usage, there is more overlap in the usage of words in tweets among the list we created. It is also interesting that the word 'coronavirus' is used almost four times more by Democrats as compared to Republicans. Some of the common words among Democrats were 'pandemic', 'health', 'worker', 'crisis', 'love' and 'thank', and on the contrary, Republicans often used words such as 'business', 'biden', and 'realdonaldtrump'. While Republicans tweeted more about economy and major political figures, they were reluctant to mention the urgency of the situation. Compared to the Republicans, the Democrats showed more concern towards the well-being of people.

## 5 STATISTICAL ANALYSIS

In this section, we consider features from four feature groups to look into different statistics related to individual features between the two political affinities *Democrats* and *Republicans*. However, we did not include features from topic modelling because of space limitations. Table 2 shows statistics such as t-statistic [18], p-value [14], and Cohen's-d (effect size) with 95% confidence interval (CI) [19] for all the features in the dataset for the two political affinities. Hence, the objective is to identify dataset features that would allow to discriminate between the two affinities. Moreover, since p-values are not sufficiently informative [21, 34], we additionally calculated cohen's-d [30] to help understand the statistical significance of the features. To interpret Cohen's-d, we used a commonly used rule-of-thumb: small effect size = 0.2, medium effect size = 0.5, and large

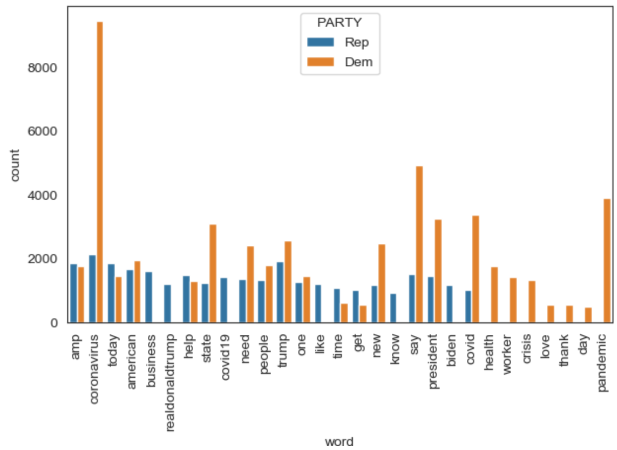


**Figure 4: Barplot for Most Common Words from Different Political Affinity and Type Groups.**

effect size = 0.8. Moreover, we calculated 95% confidence interval for Cohen's-d. The higher the cohen's-d value, the stronger the possibility of discriminating the two groups using the considered feature. A confidence interval does not include zero increases the confidence in the calculated effect size.

Table 2 show that features such as 'number of URLs' and 'retweeted' have high t-statistics and above small effect sizes, indicating signs of discriminating the two groups using those features. Moreover, vader features from the feature group SENTI too had high t-statistics, even though the cohen's-d score were small. However, all the features included in the table had positive confidence intervals for cohen's-d showing high reliability on those scores. Features such as 'coronavirus','realdonaldtrump', and 'pandemic' from WORD had closer to small effect sizes with positive confidence intervals. From the group HASH, #paycheckprotectionprogram and #ppp that indicate the same semantic meaning were among the features with top five t-statistic and Cohen's-d values in that feature group.

Moreover, we analyzed and compared the topic model features across classes of democrats and republicans for different number of topics as shown in Table 3. Here, we generated features using different number of topics, and for each combination, we calculated the maximum and mean of both t-statistic and cohen's-d. Results show that the all four attribute values decreased with increasing number of topics. This suggest that, when the number of topics is increased, informativeness of each single topic in discriminating between the two classes of interest diminished.

## 6 INFERENCE

The goal of the two-class inference task was to use different subsets of features in the training set, and calculate the accuracy. The target variables were "democrats" and "republicans" that indicate political affinity of the account that tweeted. Moreover, we used random forest classifiers (RF), changing number of trees between 200 and 500. Furthermore, when preparing the dataset, we made sure that the classes are balanced by up-sampling the minority classes to get balanced dataset. The baseline for experiments is 50%, since the classes were balanced in the inference task.

For inference tasks we used 10 hashtags with highest t-statistics: #inthistogether, #paycheckprotectionprogram, #coronavirus, #ohio, #ppp, #obamagate, #vote4mindy, #foxnews, #911onfox, and #staysafeohio. Moreover, we used ten words with highest t-statistics: coronavirus, say, realdonaldtrump, pandemic, health, worker, people, business, president, and crisis. Hence, for the inference task, we used four features from META, five feature SENTI, ten features each from WORD and HASH. In addition to these 29 features, for different tasks, we used different $\alpha$ to generate topic modelling features.

The objective of this task was to use different feature group combinations in the inference and examine how the accuracy varies. We set the number of topics as 30 for this specific task, and the results are summarized in Table 4. As individual feature groups, SENTI achieved an accuracy of 70.19% in inferring the political affinity. On the other hand, feature group with least accuracy was HASH with an accuracy of 64.36% which is only a slight increase over the baseline. The feature group META+HASH+WORD shows an accuracy of 76.23% for political affinity inference. This feature group combination only used tweet metadata and content to generate features, hence showing that even by using the tweet content, it is possible to infer the affinity with a high accuracy. This accuracy can be increased by over 6% to 82.73% when using features from LDA and SENTI, showing the usefulness of using additional features derived using topic modelling and sentiment analysis techniques. Moreover, the last row of the table shows the feature group combination when the variable *Type* (media, politician, and celebrity) is added to the features. The accuracy bumped to 83.78% with an increase of 1% from the prior feature group combination.

Moreover, in order to examine the sensitivity of the inference to different number of topics derived from topic modelling ($\alpha$), we changed $\alpha$ from 3 to 150 and calculated the overall accuracy of the inference model. Results indicated that (figure 6 in appendix A) the accuracy would increase for $\alpha$ between 3 and 35, where as the curve is more or less flattened after that point. However, it should be highlighted that for all $\alpha$, accuracy values varied only in a small range between 81.8% and 83.1%.

As a summary, these results showed that it is indeed possible to infer binary political affinities using content metadata of tweets. In such an urgent disaster situation, where the diffusion of misinformation and biased arguments may have a critical effect to society, intelligent systems with automatic political affinity inference capabilities could benefit users to have access to more balanced news and correct information.

## 7 DISCUSSION

We found different tweeting patterns between Democrats and Republicans during the COVID19 crisis. We observed the clearly separable use of hashtags and different lists of frequently used words between two political affinities. Our findings extend the previous studies on political polarization in social media during the presidential elections[11, 12], an event that is inherently 'political'. Even during the pandemic outbreak, a disaster that is usually considered neutral, political polarization was observed. Considering the importance of having access to correct, up-to-date information, especially in the amidst of disaster, such an polarization may deter

**Table 2: Comparative statistics of features across classes "democrats" and "republicans": t-statistic, p-value, and cohen's-d with 95% confidence intervals. Features are grouped by feature group and sorted based on the decreasing order of cohen's-d. Only five features each from HASH and WORD feature groups are presented here due to space limitations.**

| | | Democrats | | Republicans | | Statistics | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Feature** | **Feature Group** | **mean** | **std** | **mean** | **std** | **abs. mean diff.** | **t-statistic** | **p-value** | **cohen's-d [95% CI]** |
| number of urls | META | 0.6670 | 0.6670 | 0.4843 | 0.5252 | 0.1872 | 67.72 | $< 10^{-5}$ | 0.3576, [0.3472, 0.3681] |
| retweeted | META | 0.0768 | 0.2664 | 0.1447 | 0.3518 | 0.0678 | 41.57 | $< 10^{-5}$ | 0.2174, [0.2069, 0.2278] |
| retweet count | META | 1236.43 | 8556.12 | 1590.86 | 5910.69 | 354.43 | 8.96 | $< 10^{-5}$ | 0.0482, [0.0378, 0.0585] |
| favorite count | META | 3473.60 | 31800.30 | 2350.83 | 13161.60 | 1122.77 | 8.45 | $< 10^{-5}$ | 0.0461, [0.0357, 0.0565] |
| vader compound | SENTI | 0.0578 | 0.4878 | 0.1030 | 0.4738 | 0.0451 | 17.69 | $< 10^{-5}$ | 0.0938, [0.0834, 0.1042] |
| vader positive | SENTI | 0.0962 | 0.1189 | 0.1062 | 0.1342 | 0.0100 | 14.99 | $< 10^{-5}$ | 0.0789, [0.0685, 0.0893] |
| vader negative | SENTI | 0.0738 | 0.1007 | 0.0665 | 0.1049 | 0.0073 | 13.42 | $< 10^{-5}$ | 0.0709, [0.0605, 0.0813] |
| vader neutral | SENTI | 0.8298 | 0.1391 | 0.8271 | 0.1535 | 0.0027 | 3.51 | $< 10^{-3}$ | 0.0185, [0.0081, 0.0288] |
| nltk sentiment | SENTI | 0.0638 | 0.1812 | 0.0672 | 0.2010 | 0.0033 | 3.35 | $< 10^{-3}$ | 0.0176, [0.0073, 0.0280] |
| coronavirus | WORD | 0.1506 | 0.3577 | 0.0678 | 0.2515 | 0.0827 | 51.91 | $< 10^{-5}$ | 0.2677, [0.2577, 0.2777] |
| realdonaldtrump | WORD | 0.0058 | 0.0763 | 0.0401 | 0.1962 | 0.0342 | 46.53 | $< 10^{-5}$ | 0.2304, [0.2201, 0.2401] |
| pandemic | WORD | 0.0709 | 0.2567 | 0.0258 | 0.1587 | 0.0450 | 40.79 | $< 10^{-5}$ | 0.2112, [0.2012, 0.2212] |
| say | WORD | 0.0820 | 0.2744 | 0.0390 | 0.1936 | 0.0430 | 35.11 | $< 10^{-5}$ | 0.1810, [0.1711, 0.1910] |
| health | WORD | 0.0559 | 0.2297 | 0.0309 | 0.1731 | 0.0249 | 23.84 | $< 10^{-5}$ | 0.1226, [0.1156, 0.1297] |
| #paycheckprotectionprogram | HASH | 0.0001 | 0.0084 | 0.0096 | 0.0972 | 0.0095 | 28.05 | $< 10^{-5}$ | 0.1373, [0.1276, 0.1478] |
| #inthistogetherohio | HASH | 0.0000 | 0.0034 | 0.0042 | 0.0647 | 0.0042 | 18.71 | $< 10^{-5}$ | 0.0916, [0.0819, 0.1015] |
| #ppp | HASH | 0.0001 | 0.0103 | 0.0036 | 0.0601 | 0.0035 | 16.63 | $< 10^{-5}$ | 0.0815, [0.0716, 0.0918] |
| #coronavirus | HASH | 0.0098 | 0.0988 | 0.0188 | 0.1361 | 0.0090 | 15.06 | $< 10^{-5}$ | 0.0757, [0.0658, 0.0889] |
| #ohio | HASH | 0.0000 | 0.0048 | 0.0028 | 0.0523 | 0.0027 | 14.96 | $< 10^{-5}$ | 0.0732, [0.0631, 0.0839] |

**Table 3: Comparative statistics of topic model features across classes "democrats" and "republicans" for different number of topics. Mean and Maximum of t-statistic and cohen's-d based results are included.**

| # of topics | maximum t-statistic | mean t-statistic | maximum cohen's-d | mean cohen's-d |
|---|---|---|---|---|
| 3 | 44.8921 | 29.8019 | 0.2390 | 0.1581 |
| 4 | 42.7796 | 21.2210 | 0.2283 | 0.1127 |
| 5 | 38.8133 | 18.8462 | 0.2074 | 0.1001 |
| 6 | 36.9744 | 20.8717 | 0.1977 | 0.1107 |
| 7 | 34.7234 | 19.1773 | 0.1856 | 0.0928 |
| 8 | 34.6744 | 17.5067 | 0.1857 | 0.0928 |
| 10 | 32.0082 | 15.6414 | 0.1715 | 0.0829 |
| 15 | 28.2354 | 13.6842 | 0.1493 | 0.0723 |
| 20 | 28.4012 | 12.2689 | 0.1484 | 0.0651 |
| 25 | 30.9418 | 11.5765 | 0.1668 | 0.0651 |
| 30 | 28.0685 | 10.8192 | 0.1500 | 0.0574 |

**Table 4: Inference Accuracies for Different Feature Group Combinations**

| Feature Group | Accuracy | Precision | Recall |
|---|---|---|---|
| Baseline | 50.00% | 50.00% | 50.00% |
| HASH | 64.36% | 65.41% | 61.59% |
| LDA | 68.43% | 68.83% | 68.24% |
| WORD | 69.41% | 67.31% | 69.32% |
| SENTI | 70.19% | 70.43% | 70.24% |
| META | 70.90% | 70.99% | 70.89% |
| META+HASH+WORD | 76.23% | 77.41% | 76.28% |
| META+HASH+WORD+SENTI | 79.37% | 79.38% | 79.49% |
| META+HASH+WORD+LDA | 80.19% | 80.21% | 80.19% |
| META+HASH+WORD+LDA+SENTI | 82.73% | 82.71% | 81.18% |
| META+HASH+WORD+LDA+SENTI+Type | 83.78% | 83.84% | 83.78% |

overcoming the situation, restoring resilience, and recovering from the consequences quickly, for the society as a whole. This leads to a question on whether it is appropriate for data scientists and computational social scientists to keep an apolitical attitude as possible in every question they encounter. In our study, rather than merely presenting the findings, we introduced a machine-mediated inference model that can be used to balance Twitter timelines and mitigate the polarization, which will eventually lead to a more resilient society.

## 8 CONCLUSION AND FUTURE WORK

In our study, we (1) observed the different tweeting patterns of U.S. twitter profiles from different political affinities; and (2) inferred the political affinity of a tweet using the tweet content and metadata with an accuracy of 82.73%. Our findings suggest that the political polarization persisted even during the pandemic crisis and could be identified by using machine-mediated inference models.

There are several limitations in our work. Firstly, we used only a limited number of Twitter profiles, assuming that the sample is adequately representative. This is due to the lack of well-established knowledge on influential twitter accounts that have revealed their political affiliation. For a more reliable and robust analysis, a higher number of profiles and a better system of labeling the users is suggested. Moreover, in this work, we modelled the political affinity to be a binary variable that represent democrats and republicans, whereas in the real-world it always lies on continuous spectrum.

This work can be extended towards several directions. Firstly, in-depth analysis on tweet content can be done. This can be achieved by applying other NLP techniques, such as N-gram, or by combining the word analysis and sentiment analysis. This approach could discover how common hashtags or words were expressed in different tones and contexts, or reversely, how different hashtags and keywords were used to deliver similar messages. Secondly, our study can be extended towards temporal analysis, in order to align the timeline with the major events and updates during the selected time period. Such an approach may enable the comparison of reaction speed in different political affinities and issues. Moreover, similar approach and methodology can be expanded towards other sectors of the U.S society. And finally, our approach can be use to represent the political landscape of other societies whose political structure and ideologies are different from those of the U.S.

## REFERENCES


[1] 2020. *NLTK Sentiment Analysis.* Retrieved June 09, 2020 from https://www.nltk.org/api/nltk.sentiment.html

[2] BBC. 2020. *Coronavirus: Twitter will label Covid-19 fake news.* BBC. Retrieved May 12, 2020 from https://www.bbc.com/news/technology-52632909

[3] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3, null (March 2003), 993–1022.

[4] William J Brady, Julian A Wills, Dominic Burkart, John T Jost, and Jay J Van Bavel. 2019. An ideological asymmetry in the diffusion of moralized content on social media among political leaders. *Journal of experimental psychology. General* 148, 10 (October 2019), 1802—1813. https://doi.org/10.1037/xge0000532

[5] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. COVID-19: The First Public Coronavirus Twitter Dataset. *arXiv preprint* (March 2020). https://doi.org/arXiv:2003.07372

[6] Xuewei Chen, Jennifer L. Hay, Erika A. Waters, Marc T. Kiviniemi, Caitlin Biddle, Elizabeth Schofield, Yuelin Li, Kimberly Kaphingst, and Heather Orom. 2018. Health Literacy and Use and Trust in Health Information. *Journal of Health Communication* 23, 8 (2018), 724–734. https://doi.org/10.1080/10810730.2018.1511658 arXiv:https://doi.org/10.1080/10810730.2018.1511658 PMID: 30160641.

[7] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing.* IEEE, 192–199.

[8] United States Department of Homeland Security. [n.d.]. *COVID-19 Disaster Declarations.* Retrieved June 6, 2020 from https://www.fema.gov/coronavirus/disaster-declarations

[9] Li Libing Zhang Qingpeng Dong, Rongsheng and Cai Guoyong. 2018. Information Diffusion on Social Media During Natural Disasters. *IEEE Transactions on Computational Social Systems* 5, 1 (2018), 265–276.

[10] Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. 2012. Mining contrastive opinions on political texts using cross-perspective topic model. In *Proceedings of the fifth ACM international conference on Web search and data mining.* 63–72.

[11] Jennifer Fromm, Stefanie Melzer, Björn Ross, and Stefan Stieglitz. 2018. *Trump Versus Clinton: Twitter Communication During the US Primaries.* Springer International Publishing, Cham, 201. https://doi.org/10.1007/978-3-319-90312-5_14

[12] David Garcia, Fernando Mendez, Uwe Serdült, and Frank Schweitzer. 2012. Political Polarization and Popularity in Online Participatory Media: An Integrated Approach. In *Proceedings of the First Edition Workshop on Politics, Elections and Data* (Maui, Hawaii, USA) *(PLEAD '12).* Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/2389661.2389665

[13] Liza Greenberg, Guy D'Andrea, and Dan Lorence. 2004. Setting the Public Agenda for Online Health Search: A White Paper and Action Agenda. *J Med Internet Res* 6, 2 (8 Jun 2004), 18. https://doi.org/10.2196/jmir.6.2.e18

[14] Greenland Sander, Senn Stephen J., Rothman Kenneth J., Carlin John B., Poole Charles, Goodman Steven N., and Altman Douglas G. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology* 31, 4 (2016), 337–350. https://doi.org/10.1007/s10654-016-0149-3

[15] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *ICWSM.*

[16] Mehdi Jamali, Ali Nejat, Souparno Ghosh, Fang Jin, and Guofeng Cao. 2019. Social media data and post-disaster recovery. *International Journal of Information Management* 44 (2019), 25 – 37. https://doi.org/10.1016/j.ijinfomgt.2018.09.005

[17] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. Why We Twitter: Understanding Microblogging Usage and Communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (San Jose, California) *(WebKDD/SNA-KDD '07).* Association for Computing Machinery, New York, NY, USA, 56–65. https://doi.org/10.1145/1348549.1348556

[18] Tae Kim. 2015. T test as a parametric statistic. *Korean Journal of Anesthesiology* 68 (11 2015), 540. https://doi.org/10.4097/kjae.2015.68.6.540

[19] Daniël Lakens. 2013. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 4: 863. *Frontiers in psychology* 4 (11 2013), 863. https://doi.org/10.3389/fpsyg.2013.00863

[20] Huyen T. Le, G. R. Boynton, Yelena Mejova, Zubair Shafiq, and Padmini Srinivasan. 2017. Revisiting The American Voter on Twitter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17).* Association for Computing Machinery, New York, NY, USA, 4507–4519. https://doi.org/10.1145/3025453.3025543

[21] Dong Kyu Lee. 2016. Alternatives to P value: confidence interval and effect size. In *Korean journal of anesthesiology.*

[22] Edward Livingston and Karen Bucher. 2020. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA Infographic* (March 2020). https://doi.org/10.1001/jama.2020.4344

[23] Daniel A McFarland, Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D Manning, and Daniel Jurafsky. 2013. Differentiating language usage through topic models. *Poetics* 41, 6 (2013), 607–625.

[24] Negar Mottahedeh. 2015. *# iranelection: Hashtag solidarity and the transformation of online life.* Stanford University Press.

[25] Dhiraj Murthy. 2018. *Twitter.* Polity Press Cambridge, UK.

[26] Immanuel Ness. 2015. *Encyclopedia of American social movements.* Routledge.

[27] World Health Organization. 2020. *Coronavirus disease (COVID-19) pandemic.* WHO. Retrieved June 1, 2020 from https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[28] World Health Organization. 2020. *"Solidarity" clinical trial for COVID-19 treatments.* WHO. Retrieved May 27, 2020 from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments

[29] J. Rexiline Ragini, P.M. Rubesh Anand, and Vidhyacharan Bhaskar. 2018. Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management* 42 (2018), 13 – 24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004

[30] Marnie E. Rice and Grant T. Harris. 2005. Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r. *Law and Human Behavior* 29, 5 (01 Oct 2005), 615–620. https://doi.org/10.1007/s10979-005-6832-7

[31] Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina, and Joe Hasell. 2020. *Coronavirus Pandemic (COVID-19).* Retrieved June 2, 2020 from https://ourworldindata.org/coronavirus

[32] Konstantinos N Vavliakis, Andreas L Symeonidis, and Pericles A Mitkas. 2013. Event identification in web social media through named entity recognition and topic modeling. *Data & Knowledge Engineering* 88 (2013), 1–24.

[33] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management.* 1031–1040.

[34] Koji Yatani. 2016. *Effect Sizes and Power Analysis in HCI.* Springer International Publishing, Cham, 87–110. https://doi.org/10.1007/978-3-319-26633-6_5

# A MORE RESULTS

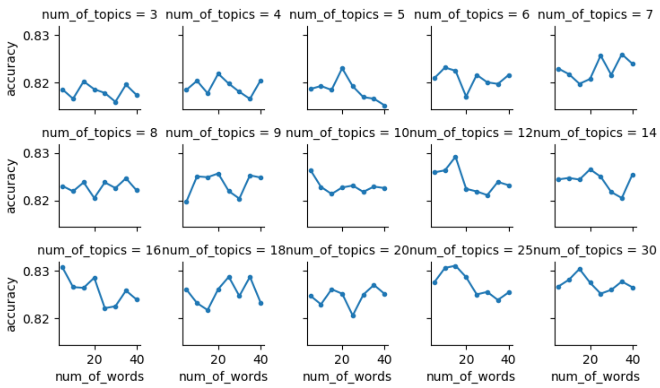


**Figure 5: Number of words vs. Accuracy plot for different number of topics.**

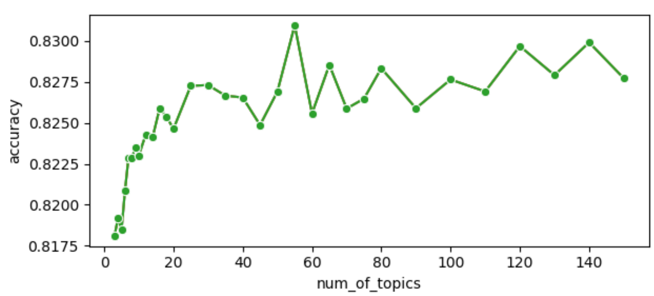


**Figure 6: Number of topics vs. Average Accuracy using Random Forest Classifiers.**

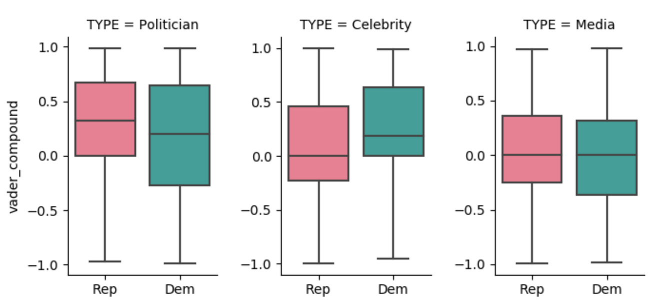


**Figure 7: Boxplot for the feature Vader compound**

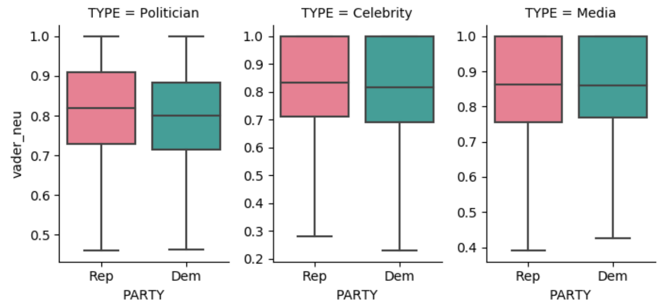


**Figure 8: Boxplot for the feature Vader Neutral**

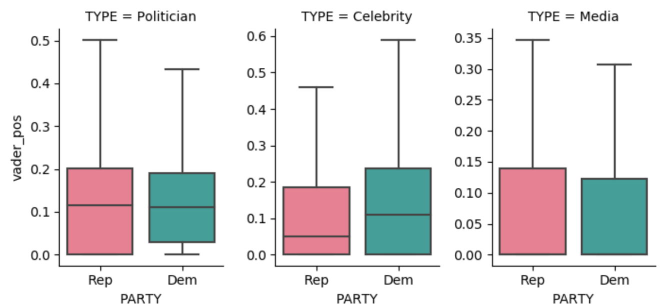


**Figure 9: Boxplot for the feature Vader Positive**

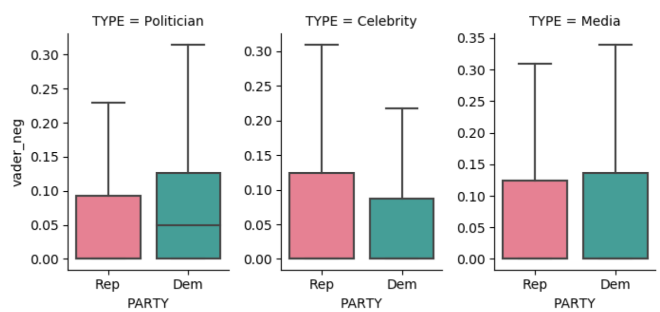


**Figure 10: Boxplot for the feature Vader Negative**

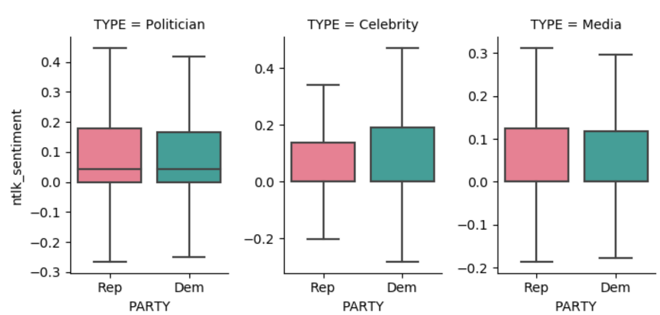


**Figure 11: Boxplot for the feature NLTK sentiment**