Emotion analysis on tweets during the 2020 US elections

Kozłowska Marta[a], Lai Jaimy[a], Loureiro Simões Jessica[a] and Stachyra Piotr[a]

[a] *Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands*

Abstract

Emotion analysis is the use of natural language processing, computational linguistics and text analysis to study, identify or quantify subjective information or emotional states and moods. In this research, the goal is to identify emotional states positive and negatively between different political states within the US during the 2020 elections on Twitter and to include a method to understand which topics are mentioned. Conducting such research contributes to a better understanding and overview of the political climate on a social media platform as Twitter. The model used within this research is a toolkit called Pysentimiento that is based on the BERT and GPT model. Furthermore, an analysis on distinctive words has been conducted. The results of this study revealed that the dissimilarity between the distributions of emotion intensities and their frequencies occurs at a different level than the one initially assumed. These are similar for Democrat and swing states groups, which differ from Republican states. The most distinctive words analysis allowed for identifying the objects of political debate in the context of a particular emotion for each group of states. It was found that most of the tweets with dominantly negative emotions focused on the former US President, counting mailed ballots and changing states 'flipping' their political preference. The contents related to the issues related to the political debate and the state of democracy were also identified as distinctive for sad and fearful tweets. The most distinctive words of joyful, surprised and neutral tweets did not allow for extracting meaningful insights.

1. Introduction

Sentiment analysis (SA) is an ongoing field of research that focuses on the text mining field [1]. It does so by computationally analysing opinions, sentiment, and subjectivity of text. It is therefore, the computational study of people's opinions, attitudes, and emotions towards an entity that can be represented as individuals, events or topics [1]. Sentiment analysis is considered a classification problem. This problem comes with it's own challenges. Certain challenges that need to be taken into account are term presence and frequency, Parts-of-Speech (POS), opinion words and phrases to expressions as 'good or bad' or 'cost me an arm and a leg', negations and different languages [1]. Most of these problems arise from computers that do not understand semantics the same way humans do. It is therefore important to be aware of these possible complications when training a model. While sentiment analysis aims to get a grasp of a general feeling of the data, emotion analysis looks at it more in-depth. Sentiment analysis focuses on neutral, negative and positive feelings, whereas emotion analysis relies on a more complex and sophisticated system and makes distinctions between negative and positive emotions such as 'happy', 'angry' or 'disgust'. The benefit of emotion analysis is that it gives more valuable insights and understanding when comparing emotions whereas sentiment analysis could lead to oversimplification of emotions. However, this is highly dependent on the task at hand. One could argue that emotion analysis is more subjective. In a study on

CEUR Workshop Proceedings (CEUR-WS.org)

using hashtags as labels for a model to predict emotion in Twitter messages, Hasan et al. [2] indicated that labels from a general public are not reliable and inconsistent which shows that even for humans emotion classification is difficult. Twitter is a social media platform that allows users to interact with one another and share opinions, thoughts and experiences. As opposed to other social platforms, Twitter is interesting in that a lot of the dialogue is public and therefore open to analyze in order to get a view of users and their interactions [3]. Therefore, Twitter makes a great corpus for sentiment analysis and emotion analysis as people can express themselves freely.

Research by Vo and Collier [4] targets earthquake situations in Japan as an event for emotion analysis using Twitter. The analysis shows that Twitter users display fear and anxiety as emotions after an earthquakes occurrence. Using Twitter data, Imtiaz et al. [5] were able to analyze the Israel-Palestine conflict as perceived by Twitterers globally. Using pre-trained sentiment analysis models, the research was able to classify opinions into pro-Palestina, pro-Israel or neutral-based. According to Buccoliero et al. [6], Twitter has become the most important communication channel for political candidates. Politicians tweet their opinions and those tweets are spread by others through mentions and retweets. Some studies even suggest that social media activities have a connection with electoral outcomes. Politicians have picked up on this, as well as voters. Voters are interested in knowing more about a candidates' personality and in order to evaluate trustworthiness they will visit the profiles of candidates. Furthermore, politically interested and opinionated people also make a presence on Twitter. For some, Twitter is even a source of news or community where they are open to discuss and share opinions on their beliefs.

Various papers have tried to predict the elections using tweets. A research by Gaurav et al. [7] tried to predict the elections based on the number of times a candidate is mentioned prior to elections on Twitter. Ramteke et al. [8] conducted a similar research, proposing a machine learning model that can predict the elections using a two-stage framework and using sentiment analysis. Moreover, Gayo-Avello [9] notes that the predictive power of Twitter and elections is often exaggerated, difficult, and not reproducible and insists that using Twitter as a predictive measure is an over-generalization. Nevertheless, Murthy [10] states that while you cannot use Tweets to predict elections, they be used as a reactive measure, an echo of the media that shows what news users are focusing on and a general feeling towards.

One recent election that has woken the interest of people worldwide were the 2020 US elections with the two main candidates Donald Trump and Joe Biden. One important aspect in the US elections is the division of support in the different parties between the different states. They are generally divided into blue (democratic dominated), red (republican dominated) and swing (no obvious dominating party) states, which have numerous differences in dialogue and political attitudes. While many analyses have been done on elections over the years [6, 7, 8], this most recent one remains to be further explored.

The research question is therefore: *Based on most-liked political tweets during the US elections in 2020, are the emotions in the debate more positive or negative in swing states than safe states?* In the following section, the Dataset will be introduced and a descriptive analysis will be given. This is followed by an extensive explanation of the research method, which encompasses the pre-processing steps, the emotion analysis and the most distinctive words analysis. Then, the results of the study will be given, followed by an evaluation of the models performance. A conclusion and discussion on the results will then be given.

2. Data

In this section, the dataset and its structure will be described and explained.

2.1. Dataset

The dataset "US Election 2020 Tweets" [11] was retrieved from www.kaggle.com and it consists of data on tweets from 15th October 2020 to 8th November 2020. The data was collected by retrieving tweets that mention the two main candidates of the US 2020 elections with the keywords #DonaldTrump, #Trump, #JoeBiden and #Biden. The complete list of attributes can be found under appendix A.1. and the ones used for this study are as follows:

- tweet - Full tweet text
- likes - Number of likes on the tweet
- retweet_count - Number of retweets of the tweet
- user_followers_count - Followers count on tweet creator
- user_location - Location given on tweet creator's profile
- country - Country parsed from user_location
- state - State parsed from user_location

The dataset consists of around 1.5 million tweets in total, and 280.000 tweets in the U.S. alone. This large deduction is possibly due to the fact that not all users of the platform indicated their location and were therefore left out of the sample. The tweets that were used were only original tweets and not retweets or replies from other users, making the information on location allocated to the tweet less ambiguous.

2.2. Descriptive Statistics

After cleaning the data by extracting only tweets from users located in the U.S., the states were grouped by three different state categories: swing states, states dominated by republican voters and states dominated by democratic voters. This led to an imbalance in the data with people from democratic states having posted almost 6 times more tweets than people from republican states and people from swing states around 4 times more (see figure 1).

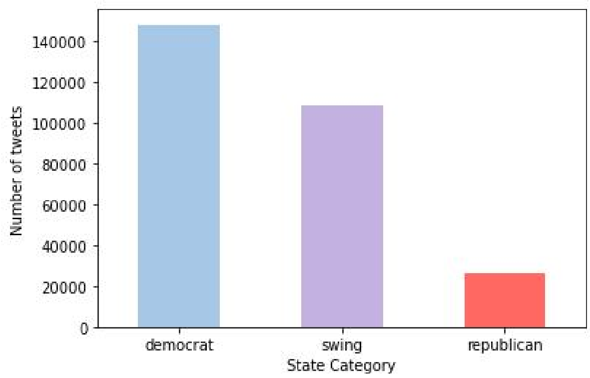

Figure 1: Number of tweets per state category.

To tackle this problem, a sample was taken of the 10.000 most liked tweets per state category, leading to an overall sample of 30.000 tweets in total. The argument for this method is that the goal of the study is to measure the *overall* emotion in the different state categories and thus a tweet that has received a like represents an agreement on its message and is therefore more relevant for this study. Consequently, this method will better reflect the voters' overall attitude towards the electives than other sampling methods for removing imbalanced data.

It can be argued that the likes on the selected tweets may not have originated from users in the same state category as the authors and that the fact that users from democratic states receive more likes in total than users from other states (see table 1) would lead to biased results. However, since the likes are not explicitly used in the analysis, but mainly taken as an indicator of message impact and agreement for sampling, this argument should not be relevant in this study's case.

dominant_party	**likes**
democrat	1.821.332
republican	101.473
swing	444.283

Table 1: Total likes per state category.

As can be seen in table 2, likes on tweets ranged between 1 and 165.702 with a mean of 79 likes per tweet. The retweet counts are generally lower. As can be seen in the percentiles, at least 25% of the tweets did not get any retweets and the mean retweet count is 18. Finally, the follower counts ranged from 0 to 5.75 million with an average of 53.388.

	likes	**retweet_count**	**user_followers_count**
count	30.000	30.000	30.000
mean	79	18	53.388
std	1291	197	265.827
min	1	0	0
25%	4	0	792
50%	9	2	3467
75%	23	6	16.544
max	165.702	20.615	5.750.841

Table 2: Descriptive Statistics.

3. Methods

The following section explains the process of analysing emotions and the objects they refer to. The emotion analysis is performed using a pre-trained model offered by Pysentimiento Python library and for the purpose of analysing the objects the emotions refer to, most distinctive words method was used.

3.1. Pre-processing

To pre-process the data, first and foremost, URLs and HTML tags were removed as these do not contain any information that could give an indication on a Twitterer's emotional state. From there on, the data was pre-processed two different ways in order to fit the task of finding the most distinctive words and to fit the emotion analysis model Pysentimiento. The data was not split, the 30000 tweets were the test sample.

For the task of finding the most distinctive words (MDW), the data was tokenized, stop words were removed using Spacy's default stop words list, as due to their commonness, they could distort the results of the MDW analysis. Moreover, the emphasis was put on unique, significant objects, so only nouns (objects), verbs (what is happening to objects) and adjectives (what the objects are like) were analysed. To avoid duplicated words differing just in their form, lemmatization was performed on the extracted tokens. Otherwise, the most output could contain the same word in different forms included multiple times, thus making the results harder to parse.

As for pre-processing for the task of emotion analysis, Pysentimento requires full strings as it has an internal tokenizer. Therefore, only URLs, HTML tags and mentions were removed from the strings to feed the model.

3.2. Analysing Emotions

To answer the question whether the emotions in the political debate are more positive or negative in swing states than in safe states, the analysis was based on the classification of emotions embedded in individual tweets. The most dominant emotions and the overall emotions intensities could be studied in terms of their distributions. This allowed for a transition from the lowest level of detail (individual tweets) to a more general level. This way, the results can be analyzed for each state and the desired groups of states - red (Republican), blue (Democrat) which constitute *safe* states and for the swing states. The assumption is that the areal extent in which the user posting a tweet was registered, corresponds to the political preference of the local population. Linking a tweet with the location is possible due to the fact that the US Election 2020 Tweets dataset contains features indicating the geographic location of the account which posted a tweet. To extract the positiveness or negativeness simply using sentiment analysis may be insufficient. Usually, the available pre-trained models allow for classifying the texts in terms of their sentiment as positive or negative, outputting polarity scores, yielding probabilities of the text being positive or negative and providing other functionality offered by NLP modules such as TextBlob [12]. Apart from the binary classification (positive or negative) and continuous outputs (polarity score), other models such as VADER produce a simple ordinal output, classifying the text stream as positive, neutral or negative and producing the intensity scores [13]. These NLP modules are well known and the VADER model's performance metrics can be better than the ones for a human being, achieving F1 score of 0.96 [13]. TextBlob uses the NLTK module functionality embedded in its API. The utilized NLTK classifier was trained on movies review corpus [12]. VADER is a rule-based model trained on a lexicon inspired by LIWC, GI, ANEW lexicons which was constructed using the wisdom-of-the-crowd approach [13]. Despite the recognition of these models, the outputs produced by them would not be informative enough in the studied context. First of all, these models do not allow for analyzing the emotions specifically, they can estimate the polarity or intensity, but emotions are more complex than just these two parameters. The expected output must contain details of the basic emotions such as anger, joy, sadness, fear, surprise, disgust [14]. Only then, it is possible to retrieve more informative characteristics of tweets. After obtaining these classification outputs, the intensity of these emotions can be measured in terms of their polarity, which is similar to the functionality provided by the discussed models.

Currently, there are few modules offering pre-trained models for classifying emotions with out-of-the-box functionality. Some of them predict emotions based on lexicons such as NRCLex [15]. Others were trained on social media and more specifically Twitter - for instance the Pysentimiento library. It was trained using BERTweet model on 5477 English tweets in 2021, so it is able to capture current specifics of language used on Twitter as well as the context of contemporary events, such as the recent political situation, economic situation, pandemics-related issues and other important topics of the debate on social media [16]. In general, it can be expected to produce the output based on up-to-date knowledge on the natural language characteristics. Apart from the fact that it was trained on the corpus originating from the same source as the one this study is concerned with, the recentness of the module is an important factor, as the study is related to the 2020 political events in the US. The model classifies emotions using labels which are quite coherent with Ekman's basic emotions, it is able to recognize: anger, disgust, fear, joy, sadness, surprise and additionally *others* which are referred to as neutral emotion [16]. Pysentimiento's functionality allows for retrieving the output indicating not only the emotions embedded in a tweet, but also the their probabilities. The only potential problem can be the fact that it was trained on a relatively small sample size (about 5500 tweets) and that the model's performance measured as a F1 score was quite low - around 0.6. Nonetheless, assigning labels to short text streams may be prone to subjectivity and even the everyday communication between people can be distorted by emotions misinterpretations.

To conclude, using Pysentimiento library, a more detailed, comprehensive spectrum of emotions present in the political discussion can be captured and further methods of analysis can be applied. The emotions provide multiple nominal classes opposed to previously mentioned outputs produced by some of NLP modules. The model performs multiclass classification and for each tweet a probability is returned for each available emotion label. These can be used to identify the most dominant emotion for a tweet. Such an approach allows for retrieving more informative results than those provided by binary outputs (positive or negative sentiment), simple ordinal outputs of a positive, neutral or negative sentiment or its strength. Based on the familiarity with the yielded classes, a formula for computing emotions intensity can be created in order to project the obtained classification outputs onto a continuous scale. The following simplified formula was used:

$$\begin{gathered} joy \times w_1 + sadness \times w_2 + anger \times w_3 + disgust \times w_4 \\ + \; surprise \times w_5 + fear \times w_6 + others \times w_7 \end{gathered}$$

where w_i is the weight assigned to each emotion. It can be assumed that the intensities express the sense of an emotion - its position on a continuous scale, where the middle is neutral, the negative and positive tails correspond to respectively: the strongest negative and the strongest positive emotions. Having the emotions' probabilities for each tweet, being able to identify tweet's most dominant emotion and to measure emotions sense (positive/negative/zero) as well as their strength (intensity), it is possible to perform quite a detailed emotions analysis required in the context of this study.

3.3. Analysing Objects

The second part of the study process was based on the idea that every emotion refers to some phenomenon or object. Considering the aim of this study, identifying them might give more information about the political debate. To complete this task, most distinctive words analysis was conducted. This particular method was chosen, because it allows for doing a multiple comparison between specific groups, e.g. tweets with dominant emotion of anger from blue states versus red and swing state and returns tokens that have significantly higher frequency in the target corpus, compared to a reference corpus [17]. Based on this tokens analysis of objects will be executed.

The first step was to compile corpora. As it was stated before, tweets were grouped by two factors - dominant party (Republican, Democratic or swing) and the dominant emotion that was detected in a tweet (anger, disgust, fear, joy, others, sadness, surprise). This creates 21 groups in total (e.g. tweets with dominant emotion of anger from blue states), so 21 corpora to compare with each other. When doing a comparison one needs a target corpus and a comparative corpus [17], but in this study corpora will be used both as target and comparative corpora, depending on which group is analyzed. For each comparison (so different set of target and comparative corpus) most distinctive tokens will be different, because they are based on contrast between two corpora.

The second step is to count the occurrences of words and assign them weighted frequencies of words. This was done by calculating the log-likelihood measure; a bag-of-words model, which is said to give better results than tf-idf comparison [18]. This step required specific pre-processing, which was discussed in a previous section.

The third step is to get the distinctive words, which are chosen based on their position in the top n tokens as ranked by the log-likelihood. The log-likelihood measure was calculated as follows[17]:

$$LL = 2\left(a \times \ln(\frac{a}{E_1}) + b \times \ln(\frac{b}{E_2})\right)$$
$$E_1 = C \times \frac{a+b}{C+D}$$
$$E_2 = D \times \frac{a+b}{C+D}$$

where a and b is the frequency of a given word in corpus 1 and corpus 2, and C and D are the total counts of words in corpus 1 and corpus 2 [19].

The core of the case study is to compare emotions in safe (Democrat and Republican) versus swing states, however, determining the objects towards these emotions are directed may output significant insights. Therefore, the comparison of discussions was carried out for the same dominant emotion among three groups of states - red (Republican), blue (Democrat) and swing. This was done by creating 21 corpora, and then comparing one with the given dominant emotion and political party (e.g. angry tweets from Democrat states) with the same emotion and two other political affiliations (e.g. angry tweets from swing and Republican states). This comparison was done 21 times, each time for different emotions and political affiliations of a state. Such analysis of corpora and extraction of the most distinctive words might provide information about which topics are discussed within certain emotions and political groups on Twitter.

4. Results

The following section presents the results of the study explaining the outputs of emotions analysis, most distinctive words analysis as well as the methods for measuring accuracy of the emotion classification model.

4.1. Emotions analysis results

The emotions analysis resulted in producing interesting insights related not only to the emotions distribution among individual states, but also to their distribution in each of groups of states - in republican, democrat (safe) states and in swing states. The former can be considered an analysis at a microscale and the latter at a macroscale - the scope of this study. The emotions intensity was studied as it provided a general overview on the sense and strength of the emotions. A simplified formula for computing emotions intensities which was discussed in the previous section was used for the obtained predictions. The following plot shows the distribution of these intensities plotted on a continuous scale.

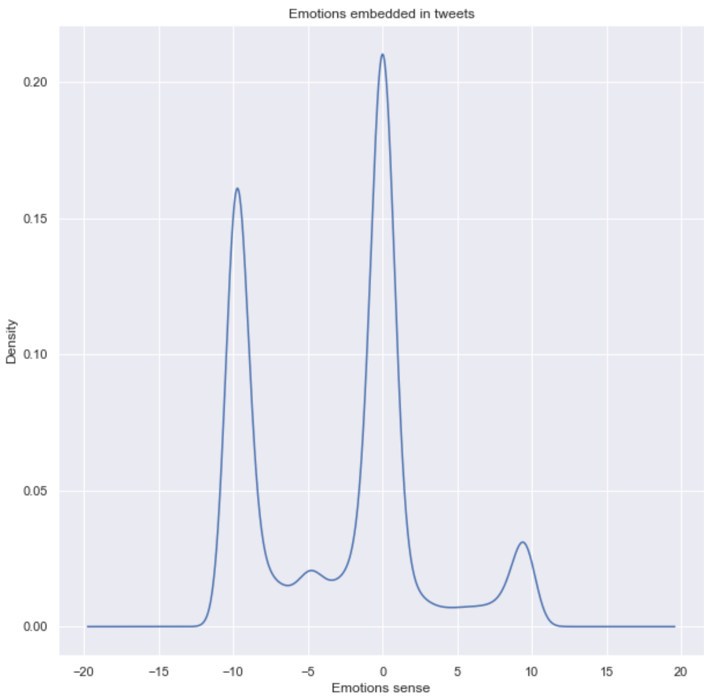

Figure 2: Emotion intensities distribution.

In this case, the emotions such as anger and disgust have the lowest weights of -10. Joy was assigned a weight of 10, sadness -1, fear and surprise -5 and others (neutral) assigned 0. In such a way, it is possible to capture the evidently negative emotions (anger, disgust) at the left tail of the scale, evidently positive emotions (joy) at the right tail of the scale, other (neutral) emotions just in the middle of the scale (0 or close to 0). Fear and surprise compose a bulge line near -5. The weaker the intensity of sad emotions and stronger of fear and surprise, the more evident it becomes. In the other case, sadness lifts the line around -1 making the shift from neutral to fear and surprise more gradual. The following plot shows the emotions intensities for each type of state.

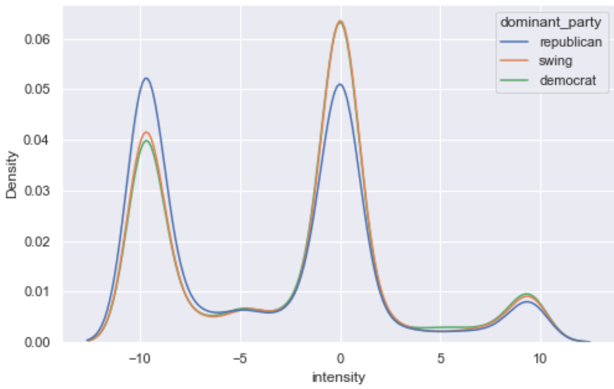

Figure 3: Emotion intensities for each group of states.

The difference between the distribution in democrat and swing states is less visible in this plot, but plotting individual graphs for these groups revealed that their distributions are in fact close to identical. Below, blue plot line represents emotions intensities of tweets originating from democrat states group, green from swing states and red from democrat states group.

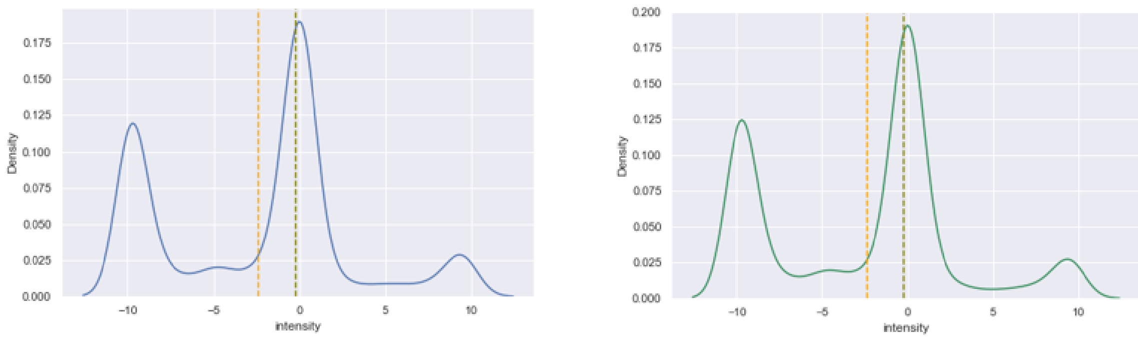

Figure 4: Emotion intensities for the Democrat (blue) and swing (green) states groups.

The dashed lines represent mean (orange) and median (olive). These two distributions differ from the distribution of emotions intensities for tweets posted by users registered in today's republican states shown in the plot below.

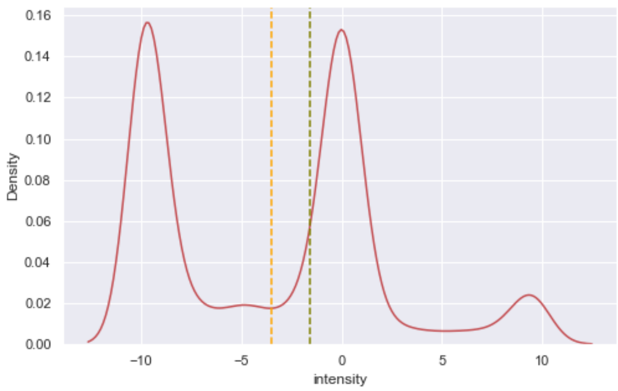

Figure 5: Emotion intensities for the Republican states tweets group.

All of the distributions share quite a similar shape - hill on the left for negative emotions, a hill in the middle for neutral emotions and a slight hill on the right - for the positive emotions intensities. The difference occurs at another level of detail. In case of Republican states, there are more negative emotions in tweets compared to Democrat or swing states. The intensities close to sadness, fear and surprise emotions are distributed more evenly in swing and democrat states, as the plot line height decreases between 0 and -5 less evidently than in the Republican states group. The W-like shape between 0 and -10 values in swing and democrat states is more asymmetrical, compared to quite a symmetrical W-like shape in the plot line for the Republican states group, in which case the spread for these emotions is oscillating closer to the value -5. In general, it indicates a weaker influence of sadness intensity and sharper impact of fear or surprise. In swing and Democrat states there were more tweets containing neutral emotions, seen as the middle hill.

Therefore, the Democrat states group has a similar distribution of emotions intensities to the one for swing states group's tweets. This is not the case for Republican states in which case it seems that the growth in negative intensities occurred at the expense of neutral emotions. The number of joyful tweets is just slightly smaller than in two other groups of states. As mentioned before, Democrat and swing states have a broader spread for fearful, surprised and sad emotions intensities in tweets than the Republican states group.

For individual emotions occurrences, it is also possible to spot a certain pattern. For the purpose of this part of emotions analysis, the most dominant emotion was assigned to a tweet based on its probability score. After examining five most common emotions for each group of states, it is quite obvious that the emotions distributions for Democrat and swing states groups are indeed very similar - not only for the emotions intensities, but also for emotions frequencies.

democrat: [('others', 5394), ('disgust', 2637), ('joy', 985), ('anger', 665), ('fear', 115)]
republican: [('others', 4604), ('disgust', 3391), ('anger', 927), ('joy', 811), ('sadness', 104)]
swing: [('others', 5424), ('disgust', 2701), ('joy', 903), ('anger', 689), ('fear', 105)]

The five most common emotions for Democrat and swing states groups are exactly the same: others (neutral), disgust, joy, anger and fear. Their frequencies are also quite similar. For the republican states group, the first two most common emotions are the same as for democrat and swing states, however, their frequencies are significantly different. There are approximately 800 fewer neutral tweets in the republican states group than in democrat or swing states group and there are almost 700 more disgusted tweets in republican states group. Additionally, the third common emotion is not joy, but anger, joy is in the fourth place and then instead of fear there is sadness. In the democrat and swing states group, surprise and sadness are least significant in terms of their counts and in republican states group these are fear and surprise. The last two emotions frequencies were not shown above to improve the readability of the results, however, in case of Republican states group that was fear (94) and surprise (69); in case of Democrat states: sadness (115) and surprise (89); in case of swing states: surprise (90) and sadness (88). The model's secondary prediction (for each tweet, the emotion with the second highest probability) was also analyzed. Apart from the difference in joy and others (neutral) in which case joy was in the first place and others in the second place for Democrat and swing states groups and a reverse order for them for Republican states group, the rest of the emotions followed the same order. Anger was more frequent in Republican states group and fear in Democrat states group.

4.2. Object analysis results

Most distinctive words analysis showed intriguing insight into the objects that are related to a particular emotion and occur in specific case of states' political affiliation. In this section results from a few analyses are presented. The focus in the analysis was placed on anger. The top 5 most distinctive features for tweets in Democratic states, that had anger as the dominant emotion can be seen in Table 3.

count	word	llr	p-value
1	goprussia	31.020941	2.552595e-08
2	workload	26.329566	2.878483e-07
3	@morningmika	26.329566	2.878483e-07
4	able	24.146064	8.929849e-07
5	reduce	21.743114	3.117066e-06

Table 3: Top 5 most distinctive features for Democratic states in tweets that had anger as the dominant emotion.

The first distinctive feature for Democratic states in tweets that had anger as the dominant emotion is 'goprussia', which suggest that The Republican Party (for which GOP is a shortcut) has some tight relationship with Russia. The second one is 'workload', which was contained in tweets which discussed the amount of work that is put into counting mailed ballots during the election. These also addresses the issue on how Donald Trump wanted to stop counting them. The third is '@morningmika', which is the Twitter username of Mika Brzezinski - an American journalist, talk show host, liberal political commentator, and author who currently co-hosts weekday morning broadcast show Morning Joe, where she criticized Donald Trump. She is mostly mentioned in tweets criticizing Donald Trump, and Fox News. The fourth is 'able', which is mostly used in tweets about the election, but both favouring Donald Trump and Joe Biden. The fifth is 'reduce', which was again in Tweets about the election, but mostly in favour of mail, and dropbox and criticizing Trump.

In general in Democratic states the tweets criticized The Republican Party and Donald Trump, and there was a lot of discussion about the election itself. This discussion was mainly about controversy around the election, and weather there was some frauds involved. People on Twitter focused on the mailing of votes that was more popular due to the epidemiological situation. This mailing system was criticized by Donald Trump and his supporters, who mentioned that it is used to ’steal the election’, but praised by Joe Biden as being safer.

The top 5 most distinctive features for tweets in Republican states, that had anger as the dominant emotion are shown in Table 4. Here the first one is ’horse’, which are tweets criticizing Donald Trump, and his actions ’being like a horse ass’, but there are also tweets comparing Joe Biden to a Trojan horse. The second one is ’toddler’ and tweets with this word mostly compare Trump to a toddler. The third is ’quintessence’, which again criticises Trump saying that he is ’quintessence of mendacity’.

count	word	llr	p-value
1	horse	27.229001	1.807274e-07
2	toddler	26.905560	2.136434e-07
3	quintessence	26.905560	2.136434e-07
4	limit	26.905560	2.136434e-07
5	apprehension	26.905560	2.136434e-07

Table 4: Top 5 most distinctive features for Republican states in tweets that had anger as the dominant emotion.

Fourth is ’limit’, which is again about criticizing Donald Trump, and writing that he is ’limited’. Fifth is ’apprehension’, which is again about criticizing Donald Trump.

In general those most distinctive words in Republican states criticize Donald Trump and are more about his personality, than his action.

The top 5 most distinctive features for tweets in swing states, characterized by anger can be seen in Table 5.

count	word	llr	p-value
1	flip	12.618366	0.000382
2	ONEV1	11.879205	0.000568
3	tired	10.434768	0.001237
4	yrs	9.843913	0.001704
5	blue	8.269728	0.004031

Table 5: Top 5 most distinctive features for swing states in tweets that had anger as the dominant emotion.

The first most distinctive word is ’flip’, which is mostly used in the context of describing a state that changes its political affiliation. The second one is ’ONEV1’, which is a Twitter Community whose goal is to engage with voters of all ages and build political power for the Democrat’s Candidate. The third is ’tired’, which is mostly used in tweets saying that people are tired of Donald Trump. The fourth is ’yrs’, which probably means years, and is used mostly in tweets in support of Donald Trump. The fifth is ’blue’, which probably means the Democratic Party, and there are tweets both supporting and criticizing it.

In the swing states the most distinctive words are more associated with the Democratic Party. The ’ONEV1’ is a phenomenon that is very specific for Twitter, and ’flip’ might suggest that in swing states people on Twitter talk more about themselves, that is the swing states. The other emotions also provided interesting insights. For instance in case of Republican states' tweets for the disgust emotion, keywords such as 'trumplie' or 'republicansforbiden' were found. For Democrat states 'debt' would be distinctive for the disgusted tweets corpora of these states, and for swing states, this emotion was related to the topics regarding the US Senate. In the context of dominantly sad tweets, the keywords suggest discussions on the pandemics in case of tweets from Democrat states, the political debate in case of Republican and hope for better future in case of swing states' tweets. The MDW in the corpus related to fear indicated impeachment topics for swing, issues related to democracy for Democrat and to Donald Trump as well as the pandemics for Republican states group. The analysis in the context of joy, neutral and surprise emotions provided less meaningful results.

4.3. Evaluating the accuracy of emotion classification procedure

As it was mentioned before, emotion classification is a relatively difficult task to be performed accurately, as it may be prone to subjectivity, especially when more emotions labels are considered. Due to this fact the evaluation of the output is somehow complex to conduct.

In this study, the accuracy of the model was measured in terms of the emotions which were assigned the biggest probability for each tweet. A qualitative analysis was performed for 50 tweets and then by merging the evaluated tweets with the tweets labeled by the model by tweets index, it was possible to compare the labels assigned by human with two labels of biggest probabilities given by the model. If the primary label of the model was identical with the label given by human, the model was awarded a point. The same was applied even if the primary label did not correspond to the manually assigned label, but the secondary label did. Only if neither primary, nor secondary label was the same as manually assigned labeled, it was counted as an error. The model achieved accuracy of 86% of correctly classified emotions. The following confusion matrix shows the determined model accuracy:

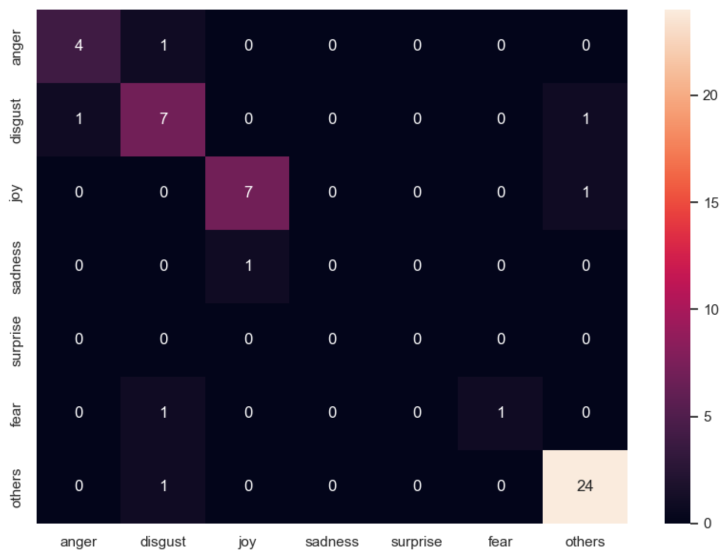

Figure 6: Confusion matrix for model's classification task.

5. Conclusion and Discussion

The political discussion on Twitter in the context of emotions shared by users is a complicated phenomenon which must be studied carefully. The analysis performed in this study covered the problem from different perspectives. The emotions were analyzed in terms of their intensities and their frequency distributions. Apart from that, the most distinctive words analysis was conducted to identify objects of these emotions. The study proved that the distributions of emotion intensities for swing and democrat states were similar to each other and different from the distribution of republican states. The emotions present in the Democrat states group's tweets are quite similar to those of the swing states group in the context of emotions intensities and their distribution, but also in the context of the dominant emotions frequencies. At the same time these remain different from the characteristics of the emotions embedded in the content posted by users registered in the Republican states. In their case, more negativity was found - in terms of the dominant emotions frequencies and of emotion intensities. Therefore, it cannot be stated that there is a difference between *safe* and *swing* states in terms of the political discussion emotions expressed on Twitter, however, the difference occurs at an initially unexpected level. The emotions intensities and frequencies distribution of the tweets posted in Republican states group differs from those of the Democrat and swing states groups. In terms of the most distinctive words analysis, some of the keywords were hashtags or account names. In such cases, it is easier to determine the political affiliation, but it provides less information about the object of the expressed emotions. In other words, alone they were often not enough to understand the context and additional, qualitative assessment of a tweet was required to determine meaning. It was also found that the Democratic states criticized Trump and Republicans, but the tweets from Republican states did that as well, most probably it is due to the fact that the Democratic minority in the Republican states generates most content and it is the reason why these anti-republican posts were found. Another reason might be that Republicans criticized Donald Trump as well, supporting other Republican fractions. Some of those most distinctive words from Democratic states related to this criticism were in the same group of tweets referencing a very similar text and specific keywords. This group was so different than the rest, that it produced few distinctive words, which provided less information about other objects. It is also interesting that many tweets talked about the former US President - Donald Trump in a negative manner in all states groups, regardless of their political affiliation, perhaps this is an indication that the Democratic supporters are much more active on Twitter, however, this may be a topic of a further study on online political debate in the USA. If this is true, then, despite balancing the data using 10000 most-liked tweets for each of the states group category, more tweets from Democrat states may have been selected, as the likes are not constrained to states borders or political view. Apart from the mentioned 'hidden' sample imbalance o, there are other limitations of this study. First of all, the results cannot be extrapolated to the general population of U.S. voters. Since Twitter is in itself a sample of the general population, it is certain that various demographic groups were either over- or under-represented in this research. Therefore, while we did find specific emotions in the different state categories, those emotions reflect only the attitude of a specific portion of the voters population.

Second of all, the majority of Twitter users are mere observers. Most twitter users do not tweet about politics, meaning that the political dialogue, and therefore the political attitude of the voters, is created by a minority of users. This further strengthens the argument of inability to extrapolate the obtained results and conclusions to the entire population. Moreover, such an extrapolation could strengthen the stereotypes and generalized assumptions regarding supporters of a certain political affiliation, which could have harmful effects. The results obtained in this study should rather indicate that in general, the most liked tweets related to political discussion lean from a mixture of emotions oscillating around neutrality towards negativity, which should be avoided.

Another limitation is the fact that replies to tweets were not included in the analysis. Unlike retweets, which are a re-post of the same text, replies to tweets contain new information and reactions to the original tweets. The fact that this information was discarded from the dataset may have led to an incomplete image of the overall emotions in the sample of voters on the social platform. Further research should be done using different kinds of samples (not only retrieved from Twitter) or by correcting for demographic biases in order to find results that are more representative of the general population. In addition, not only original tweets but also replies to tweets could be incorporated in the analysis. Moreover, a model such a BERTweet can be trained specifically for the objectives of this tasks to achieve better accuracy. An alternative measure techniques were used to evaluate Pysentimiento's model performance in case of this study and it is unclear if a similar solution was used to test the module's functionality upon its release. Apart from the fact that 0.02 of the entire sample of tweets were evaluated, the performance can be considered satisfactory, at the level of 86% correctly or almost correctly classified emotions. It can be considered much better compared to the accuracy of around 60% achieved during the module testing by its authors.

References

[1] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams engineering journal 5 (2014) 1093–1113.

[2] M. Hasan, E. Agu, E. Rundensteiner, Using hashtags as labels for supervised learning of emotions in twitter messages, in: ACM SIGKDD workshop on health informatics, New York, USA, 2014.

[3] S. Macskassy, On the study of social interactions in twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 6, 2012.

[4] B.-K. Vo, N. Collier, Twitter emotion analysis in earthquake situations, International Journal of Computational Linguistics and Applications (2013).

[5] A. Imtiaz, D. Khan, H. Lyu, J. Luo, Taking sides: Public opinion over the israel-palestine conflict in 2021 (2022).

[6] L. Buccoliero, E. Bellio, G. Crestini, A. Arkoudas, Twitter and politics: Evidence from the us presidential elections 2016, Journal of Marketing Communications 26 (2018) 1–27. doi:10.1080/13527266.2018.1504228.

[7] M. Gaurav, A. Srivastava, A. Kumar, S. Miller, Leveraging candidate popularity on twitter to predict election outcome, 2013. doi:10.1145/2501025.2501038.

[8] J. Ramteke, S. Shah, D. Godhia, A. Shaikh, Election result prediction using twitter sentiment analysis, in: 2016 International Conference on Inventive Computation Technologies (ICICT), volume 1, 2016, pp. 1–5. doi:10.1109/INVENTIVE.2016.7823280.

[9] D. Gayo-Avello, No, you cannot predict elections with twitter, IEEE Internet Computing 16 (2012) 91–94. doi:10.1109/MIC.2012.137.

[10] D. Murthy, Twitter and elections: Are tweets, predictive, reactive, or a form of buzz, Information 18 (2015). doi:10.1080/1369118X.2015.1006659.

[11] M. Hui, Us election 2020 tweets, 2020. URL: https://www.kaggle.com/manchunhui/us-election-2020-tweets, accessed Jan. 18, 2022.

[12] S. Loria, Textblob, 2020. URL: https://textblob.readthedocs.io/en/dev/api_reference.html#module-textblob.classifiers.

[13] C. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, 2015.

[14] P. EKMAN, Introduction, Annals of the New York Academy of Sciences 1000 (2006) 1–6. doi:10.1196/annals.1280.002.

[15] M. Bailey, Metalcorebear/nrclex: An affect generator based on textblob and the nrc affect lexicon. note that lexicon license is for research purposes only., ???? URL: https://github.com/metalcorebear/NRCLex.

[16] J. M. Pérez, J. C. Giudici, F. Luque, pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks, arXiv preprint arXiv:2106.09462 (2021).

[17] P. Pojanapunya, R. W. Todd, Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis, Corpus Linguistics and Linguistic Theory 14 (2018) 133–167. URL: https://doi.org/10.1515/cllt-2015-0030. doi:doi:10.1515/cllt-2015-0030.

[18] A. Gelbukh, G. Sidorov, E. Lavin-Villa, L. Chanona-Hernández, Automatic term extraction using log-likelihood based comparison with general reference corpus, volume 6177, 2010, pp. 248–255. doi:10.1007/978-3-642-13881-2_26.

[19] P. b. Rayson, Log-likelihood and effect size calculator., 2008. URL: https://ucrel.lancs.ac.uk/llwizard.html.