Political Opinion Mining Through Twitter Data

Mackenzie Stueve

March 20, 2021

Abstract

Politics are a largely debated and widespread topic in today's day and age. Currently, in the United States there are two major political parties: Democrats and Republicans. As with most divided systems there exists a division in opinions and feelings towards one another. Throughout this paper the sentiment analysis methods of VADER, HuggingFace Transformer Pipelines, and TextBlob will be used to analyze twitter data relating to politics. 2,658 tweets have been examined, of these there is 28.4% total agreement between the three analysis methods.

Contents

List of Figures

List of Tables

1 Introduction

Politics have found a large place in the mainstream media within the recent past. Previously, politics were seen as a larger complex topic and were not always seen as socially acceptable to discuss publicly. This view on the topic of politics has shifted drastically. As news stories have become constantly accessible through technology including: laptops, smart phones, and television people are more knowledgeable about the inner workings of the government and politics as a whole. Everyday citizens are now being presented with enough information to create strong opinions on a variety of political topics and the parties. With this increased awareness, politics have grown into a frequent discussion topic and have captured the interest of a large range of people.

This is an important shift in terms of examining politics because as there is an increase in awareness and discussion, there also comes an increase in varying opinions. Frequently with the discussion of politics, a persons party affiliation plays a role in determining their beliefs. However as with many topics, there exists a political spectrum. A person can have a wide variety of experiences that can shift their political opinion on a topic, regardless of the person's party affiliation. It is important to know how lay people feel about political topics in order to determine whether policies will be passed when put up for vote. Another important outcome of tracking the public's feeling on the political parties is predicting how candidates will do in upcoming elections.

Social media has also been increasing in popularity in modern times. Different social media platforms have become known for specific purposes and types of sharing. Twitter is no different, it has become known as one of main social media options in order to share opinions. Users share these opinions in the form of "tweets" consisting of 280 text characters. The culmination of these tweets has resulted in an enormous data set on a wide variety of topics, one of these being politics. One specificity of Twitter is the use of hashtags in order to connect tweets of a similar topic.

Hashtags are words or phrases that begin with the pound symbol. They are used in the text of a particular tweet in order to link and connect the tweet to those that also contain the hashtag. On the Twitter website and app users are able to click on the particular hashtag, or search for them. When these actions are taken all tweets that have included that hashtag will be displayed. These hashtags are often words or phrases related to the topic that tweet is discussing, by including this the tweet is connected to a larger ongoing conversation about that topic.

These hashtags will play a large roll in being able to find tweets directly relating to politics. Certain hashtags are used at a higher rate in order to discuss political beliefs and parties. Three specific hashtags have been utilized in order to find the set of tweets that will be analysed throughout this project. By using these hashtags we are able to examine the sentiment of twitter users on the topic of politics and the particular parties. Besides ensuring the tweets are about the topic of politics, these hashtags will allow for conclusions to be made about the sentiment towards the parties and politics. These results will prove useful to those trying to analyze party popularity and shape political movements.

2 Background

The concept of Big Data is a growing field within Computer Science. It is the concept of working to analyze data sets that are too large to use traditional analysis methods to sort through. The field of Big Data works in order to be able to process this data and extract information from it in order to gain some form of knowledge. When working to define a topic for my thesis, this is something that really sparked my interest. Twitter has one of the largest data sets publicly available, which is the reason that was the social media site examined. The specific topic of politics was chosen, based upon the opinionated nature of the topic. Even within a political party, opinions can widely meaning there is large variation with the sentiment.

2.1 Sentiment Analysis and Machine Learning

When humans read text it is natural to detect the tone and emotion behind the words written. Different sentences and word combinations naturally trigger the human brain to understand whether the sentence contains negative, positive, or neutral concepts. This skill is what the field of sentiment analysis hopes to replicate. As machine learning has grown in reputability and continued to evolve the possibilities for sentiment analysis have continued to grow. This task raises specific challenges due to the human nature. When speaking people will often utilize irony or sarcasm, therefore the meaning of their words is different than the actual definitions. In addition, the English language is extremely sensitive therefore the inclusion of a single negation word like ”not” can be used to change the entire sentiment of a statement.

Many techniques have been developed in order to try and perform this judgement taking these challenges into account. As examined in Gonçalves et al. [2] the main thing that separates these different methods is what factors from the text are accounted for and what dictionary these words are compared to. Some methods rely upon the analysis of textual words, while others factor in things such as emoticons and punctuation. In addition, different systems have different baselines for words that it can account for, meaning some systems are better built to handle slang words and social media while others may be best for formal reports. The study covered in Gonçalves et al. [2] shows how metrics can be calculated in order to prove a clear value of how accurately sentiment can be predicted in order to compare the different methods.

2.2 Twitter API

Application Programming Interfaces, or APIs, are a means of allowing two software programs interact with one another. An API can allow for one software program to make calls to another and make requests for information or actions. Twitter grants access to their API to users who obtain a Twitter Developer account. Once granted this access, users are given unique log in information that will allow them to make a call to the Twitter API and be returned the requested information.

Every tweet sent is recorded within the Twitter database. All information attached to a specific tweet is placed alongside one another and accessible through calls to the Twitter API. This information includes, the unique tweet id, the user's handle, the text of the tweet, and various other important aspects. Developers can request specific tweets to be returned from the Twitter database by specifying parameters in the call to the Twitter API. We will explore these parameters farther within Section 3 and see how they can be used in order to get a specific data set.

3 Data Collection

At the start of this project the topic was originally climate change. I utilized six hashtags in pertains to climate change in order to collect around 900 tweets. Once I began running the different sentiment analysis methods outlined in Section 4.1 on these recorded tweets, the results showed the vast majority as neutral. In order to check if this was due to the analysis methods, or the tweet content, I ran analysis on the last 100 tweets in English, which returned a diverse range of sentiment. In addition after manually assessing around 100 of the recorded tweets, I concluded that the analysis was correct and the tweets were indeed neutral. Before switching topics, I examined the cause thoroughly and found this was caused by most tweets regarding climate change being informational instead of based in opinion. Due to this finding, the topic was changed to politics, which proved to be a more polarized and opinion based data set.

This research relies upon a data set of 2,658 tweets regarding to the topic of politics. I was granted a basic level Developer account in order to conduct this research. With this level, tweets that have been published within the past 7 days can be seen and pulled from the Twitter database. Due to this limitation all of the tweets recorded were collected between February 16, 2021 and February 23, 2021.

In order to extract this data a tool called Tweepy [7] was utilized. Tweepy is a Python tool created in order to help streamline the process of scraping the Twitter data in order to assist in collecting the necessary tweets. The Tweepy feature Cursor() was the method used in order to access the politics tweets examined throughout this research. Collecting Twitter data with Tweepy previously replied upon loops in order

Hashtags Examined
#Politics
#Democrat
#Republican

Table 1: A selected list of hashtags used on Twitter to discuss politics.

to iterate through pages of tweets and return the requested values. The Cursor function has eliminated this need and handles looking through multiple pages in order to minimize the run time of Twitter data extraction.

Within the Tweepy Cursor function the necessary parameters for the API call can be implemented in order to receive the specific data set that is requested. There exists a multitude of parameters that can be placed to limit the data set returned, some more generally relate to the content of the tweet while other instances are unique to Twitter. For this research specifically, the first condition was that the tweet contained one of the hashtags listed in Table 1. The second condition placed within this call is that the tweet be published in English, in order for the sentiment analysis methods listed in Section 4.1 to correctly perform the analysis.

The remaining conditions are specific to Twitter and play an important role with shaping the data analyzed in this research. Retweets and Quote Tweets are methods in which Twitter users can re-share the a tweet verbatim from another user. These have been excluded from the data set, in order to not alter the counts of each sentiment by having the same message recorded multiple times. Similarly reply tweets are when a user is replying directly to a posted tweet from another user. These have also been excluded in the parameters due to the unknown factor of the users relation to each other and any impact that could have on the reply message. These factors in summary have limited the tweets to original messages posted by the user, not in response to or taken from any other users. In addition, tweets that contain media, such as photos or videos, or links are excluded due to the lack in the ability to analyze any sentiment these attachments may contain or any changes they could make to the sentiment of the text.

The previous parameters have been chosen in order to narrow down the tweets returned, in order to select the data set. The last parameter is used in order to ensure the tweets can be correctly analyzed. It specifies the tweets should be returned in ”extended” mode, which means that the entire text of the tweet will be returned instead of a ”truncated”, or shortened, version of the tweet. This ensures that the entire message of the tweet will be able to be analyzed for sentiment and no words will be excluded from the analysis. With all of these parameters in place a call to the Twitter API is made through the Tweepy Cursor function. The tweets are returned from this call in the form of a model class instance.

4 Experiment Design

The first step of implementing this project is to collect a data set using the methods outlined in Section 3. After these steps have been executed the data is returned as a model class instance. There is one more necessary condition to check before the tweet can be included within the data set, that it is not already included. When calling to the Twitter API there is a parameter that can be enacted called ”since”, that takes a date and time and retrieves tweets exclusively after the date. This created a problem because the ”since” parameter only takes dates in the format of ”yyyy:mm:dd hh:mm” while the Twitter API returns dates in the format of ”yyyy:mm:dd hh:mm:ss”. Due to this, it is possible to either miss tweets within the remainder of a minute or have duplicates of tweets from the minute remainder. In order to accommodate this, I recorded the time the call to the Twitter API was completed. The ”yyyy:mm:dd hh:mm” of this time was used as the since parameter for the next API call. Another statement was used in order to check that only tweets produced after the ”ss” portion of the return time were recorded.

After checking for this, the remaining tweets were processed in order to save the necessary data. For each tweet the user’s handle, the hashtags from Table 1 that it contains, the date and time of the tweet, and the actual text of the tweet is accessed and stored. There are various other characteristics of a tweet that are available from the API; however, only those necessary for future analysis have been saved for this project. All of these values are stored within a data frame using the package Pandas [6].

Once all of the information has been saved within the data frame, sentiment analysis can now be performed on the text of each tweet. In order to do this the text element of each tweet was passed into the three methods of sentiment analysis. The specifics of these methods can be seen in Section 4.1. The numeric values and the classifications returned from these methods are stored in a new data frame. This data frame is then full-joined with the data frame previously mentioned and finally exported to a comma separated value, or CSV, file.

With all of the necessary information running experiments in order to draw important information from these tweets can now be performed. These experiments included seeing how frequently the methods of analysis return the same value, or different values and the reasoning. Exploring which methods were most successful and how the hashtag impacts sentiment. The results of these experiments will be examined within Section 5.

VADER Compound Score Meaning		
Negative	Neutral	Positive
$CompoundScore \leq -0.05$	$-0.05 < CompoundScore > 0.05$	$0.05 \leq CompoundScore$

Table 2: The formulas used in order to determine sentiment from VADER Compound Score.

4.1 Methods

4.1.1 VADER

Valance Aware Dictionary for sEntiment Reasoning, or VADER for short, is a well proven method for performing sentiment analysis. VADER does not require training data and is particularly good at performing sentiment analysis on social media data sets. As discussed in Hutto and Gilbert [4] it was designed for microblog-like contexts and was evaluated over multiple data types of this form including: social media text, editorials, product reviews and movie reviews. VADER takes into account several different factors of the text in order to assess the sentiment score.

These factors include: punctuation, capitalization, degree modifiers, and conjugations. Punctuation is taken into account because marks such as exclamation points, can be used in order to increase the sentiment felt whether positive or negative. Capitalization is taken into account similarly, due to the fact capitalized words often hold a stronger feeling associated with that word. Degree modifiers such as: very, extremely, and really, are also used in order to amplify the sentiment felt with the word being modified. Conjugations are used in a different means for performing the sentiment analysis. Frequently, conjugations can bring forth a contrast in opinion, by VADER taking these into account it can allow for there to be note of a change in sentiment throughout a tweet.

Utilizing these factors VADER returns multiple values that can be used in order to analyze the sentiment of the provided text. These values are a compound score, a positive score, a negative score, and a neutral score. The positive, negative, and neutral score are all positive numbers that when added together equal a value of 1. The compound score is an average of each individual lexicon score from the text input that ranges between -1 and 1. This compound score is the value used in order to determine the sentiment and the equation used to determined this can be seen in Table 2.

By accounting for these specific textual features VADER has been able to create an extremely successful sentiment analysis tool. In the study discussed by Hutto and Gilbert [4] it found VADER has a remarkable 96% accuracy rating. This is a higher rate than for humans reading and analysing the sentiment of tweets, which was found to be around 84%.

4.1.2 Hugging Face Transformer Pipeline

The Hugging Face Transformer Pipeline is the next method implemented for this project. It is designed for Natural Language Processing, or NLP for short researchers and educators wanting to work with large-scale transformer models. Another audience that this method has been created for is users looking to download a pre-trained model for NLP projects. Hugging Face's Pipelines are based upon a pre-trained Transformer model. For sentiment analysis, the method "distilber-base-uncased-finetuned-sst-2-english" is the default style for analysis within the transformer pipeline [8].

It outputs a word response of negative or positive, there is no option of neutral with the Hugging Face Pipeline. In addition it returns a probability score. This is a value that is between 0 and 1 and represents the probability of that sentiment being correct. Higher scores represent more confidence in that sentiment, while lower scores represent that the feeling may be more neutral. The pipeline function has been analyzed at length and as recorded in Wolf et al. [8] has been found to carry an impressive 91% accuracy rating.

4.1.3 TextBlob

TextBlob is a tool used for text analysis and NLP. It has many functionalities including tokenizing words, spelling correction, noun phrase extraction, and multiple others. The premise of TextBlob being created was to make a simple API for common NLP problems[5]. In order to utilize the TextBlob sentiment analysis future, the text being analyzed is passed into the TextBlob function. An object is then returned and from this the elements needed for sentiment analysis can be accessed.

TextBlob returns the sentiment evaluation as two numeric scores, polarity and subjectivity. The polarity is a value between -1 and 1 that determines the sentiment of the text. The equations needed in order to determine the sentiment classification of "negative", "positive", or "neutral" can be found located in Table 3. The other score, subjectivity, is a value between 0 and 1, inclusive. A value being closer to 0 marks it as more objective, while on the opposite being closer to 1 means the text is more subjective.

In order to determine these scores, TextBlob is pre-trained and relies upon PatternAnalyzer[5]. The values are decided upon through a series of decisions and forms of analysis. If TextBlob does not have a word of phrase within its library, where all of the known words and their sentiment are scored, it will simply ignore the word and not factor it into the polarity score. With the words contained in the library, TextBlob will assign polarity to both words and phrases and at the end of the text input it will average all of these scores[1]. Using this system TextBlob has been proven to be decently effective, with a 76% effectiveness as recorded in Hasan et al. [3].

TextBlob Compound Score Meaning		
Negative	Neutral	Positive
$PolarityScore \leq -0.01$	$-0.01 < PolarityScore > 0.01$	$0.01 \leq PolarityScore$

Table 3: The formulas used in order to determine sentiment from TextBlob Score.

5 Results and Analysis

5.1 Analysis of Methods

There was a total of 2,658 tweets collected within this research. All three forms of analysis outlined previously in Section 4.1 were performed on every tweet. Figures 1 and 2 shows a count of how many tweets were labeled negative, positive, and neutral by each analysis method. As can be seen in the graph, some analysis methods have a tendency towards a specific sentiment. The HuggingFace Transformer Pipeline returns negative more frequently than any other analysis method. In contrast, TextBlob returns positive more than either of the two methods. VADER returns positive and negative at an equal rate, with a difference of less than 30 tweets.

When using the three analysis methods as factors, comparisons can be drawn about them. As can be seen in Figure 3 VADER and TextBlob have the highest frequency of agreeing on the sentiment. This happens a total of 1,439 times, showing that VADER and TextBlob agree roughly 54% of the time. Conversely HuggingFace Transformers Pipeline and TextBlob agree on the least amount of tweets around 37.9% of the total set. HuggingFace Transformer Pipeline, VADER, and TextBlob have total agreeance, meaning they all all produced the same sentiment for 754 of the tweets within this data set, which accounts for around 28% of the tweets.

In order to select the most accurate sentiment analysis method, a random sampling test was performed. 50 tweets were chosen from the complete data set at random and were manually analyzed for sentiment. It was recorded how many times the manual sentiment aligned with the value output by the three methods. Based on this the manual sentiment detection agreed with HuggingFace Transformer Pipelines 66% of the time. Where TextBlob and VADER agreed with the human decision on 74% of the tweets. Showing that according to this study, TextBlob and VADER are the most effective means for sentiment analysis.

5.1.1 Outliers

For this research, outliers were defined as tweets where each analysis method returned a different sentiment. There exists 302 instances of this for the total data set. Through examining these tweets conclusions can be drawn about why they were so difficult to classify. By first manually looking at the outliers there seemed to be a significant portion of these outliers that contained questions. After running analysis, it turns

Recorded Sentiment By Analysis Method

	Vader Classify	Transformer Classify	TextBlob Classify
Negative	1174	2141	745
Positive	1148	517	1196
Neutral	336	0	717

Figure 1: Table showing how many instances of each sentiment was recorded by each analysis method.

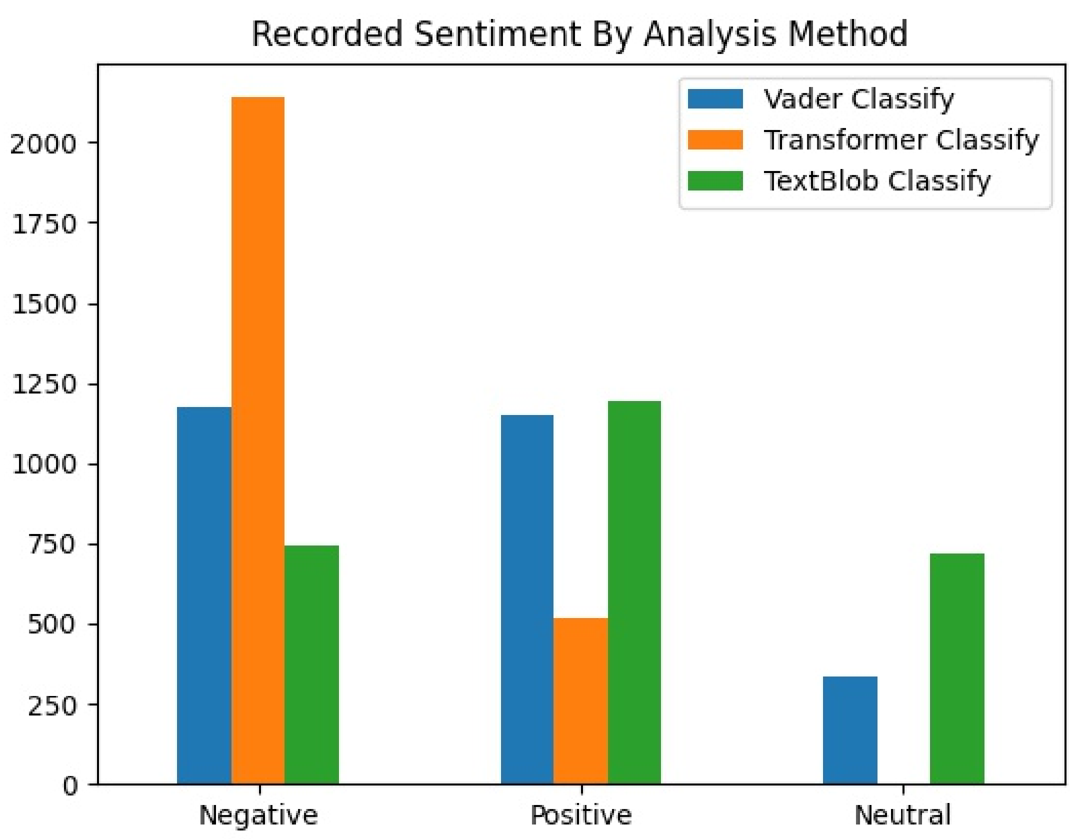

Figure 2: Graph showing how many instances of each sentiment was recorded by each analysis method.

Comparing Similarity Between Sentiment Analysis Methods

	Vader & Transformer	Vader & TextBlob	Transformer & TextBlob
Agree	1416	1439	1009
Disagree	1242	1219	1649

Figure 3: Table showing how frequently the analysis methods agree on sentiment.

out there was 84 outliers that contained questions. Some of these questions seemed to have a sarcastic connotation behind them, while others seemed to act as a general two option question in order to spark debate. These differences can be explained by the nature of questions. When asking a question, it can provide a challenge to detect the emotion of the person asking the question because they are asking in order to get the respondent's opinion and reaction.

The other division of tweets that the analysis methods seemed to have a hard time detecting, is tweets that one of the methods deemed neutral. Since HuggingFace Transformer Pipeline does not have the functionality to return a neutral sentiment, if VADER or TextBlob return nuetral there will be some disagreement. While only 11.4% of the overall tweets are outliers, the outlier tweets deemed neutral make up 28.7% of the overall neutral tweets. This shows how neutral sentiment can be difficult to detect, since it is simply the absence of the positive or negative sentiment.

5.2 Hashtag Analysis

The 2,658 tweets were chosen based upon the hashtags in Table 1. Of these tweets the sentiment analysis methods had total agreement on 754 tweets. These tweets are the set used in order to perform the analysis done below. By only using tweets with total agreement, it proves that the proper sentiment is being recorded for each hashtag.

The number of times each hashtag was recorded can be seen in Figure 4. Knowing which hashtags are most utilized can be important for understanding the size of the audience this tweet will most likely attract. With this knowledge the next most important factor is the sentiment of each hashtag. These values can be seen in Figure 5. All three of the hashtags have mostly negative tweets. While #politics and #republican have more recorded tweets, #democrats has a higher percentage of negative tweets. Of the tweets containing #politics or #republicans they have 55.2% and 54.7% of those tweets being negative, respectively. The #democrats records a higher rate of 66# of the tweets being negative. From looking at the contents of these tweets it appears Twitter users utilize these hashtags in order to talk about a party and not one they belong

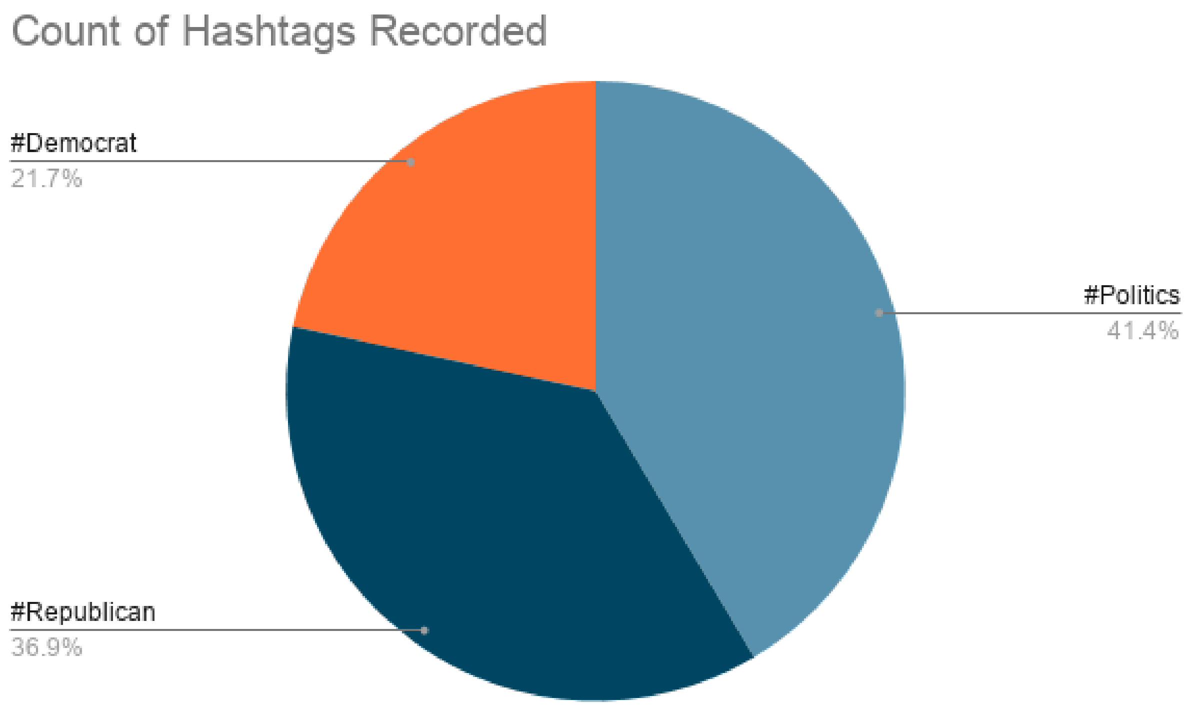

Figure 4: Number of total agreement Tweets with each hashtag.

to. This means that the sentiment of each hashtag aligns with the perception of that party to others. Analysis such as this above can be extremely useful in being able to understand the feelings of users towards different political parties and movements. Information such as this can provide insight to those working in political campaigns and lobbying, by allowing them to have some foresight into how these messages will be perceived.

6 Conclusion & Future Works

Natural Language Processing is a growing field in computer science due to its applications. Previously language processing tasks, like sentiment analysis, were restricted by time constraints. By training computer models to perform these tasks, previously reserved for humans, these limitations are removed and they can be applied on a larger scale. Having data that has been assigned the same sentiment through two or more separate analysis methods ensures that the true sentiment of a given text is being detected. As can be seen through this research, combining the methods of VADER and TextBlob can create a large data set of tweets with the correctly identified sentiment.

The Twitter API is an extremely large and well maintained data set. Having this data available provides

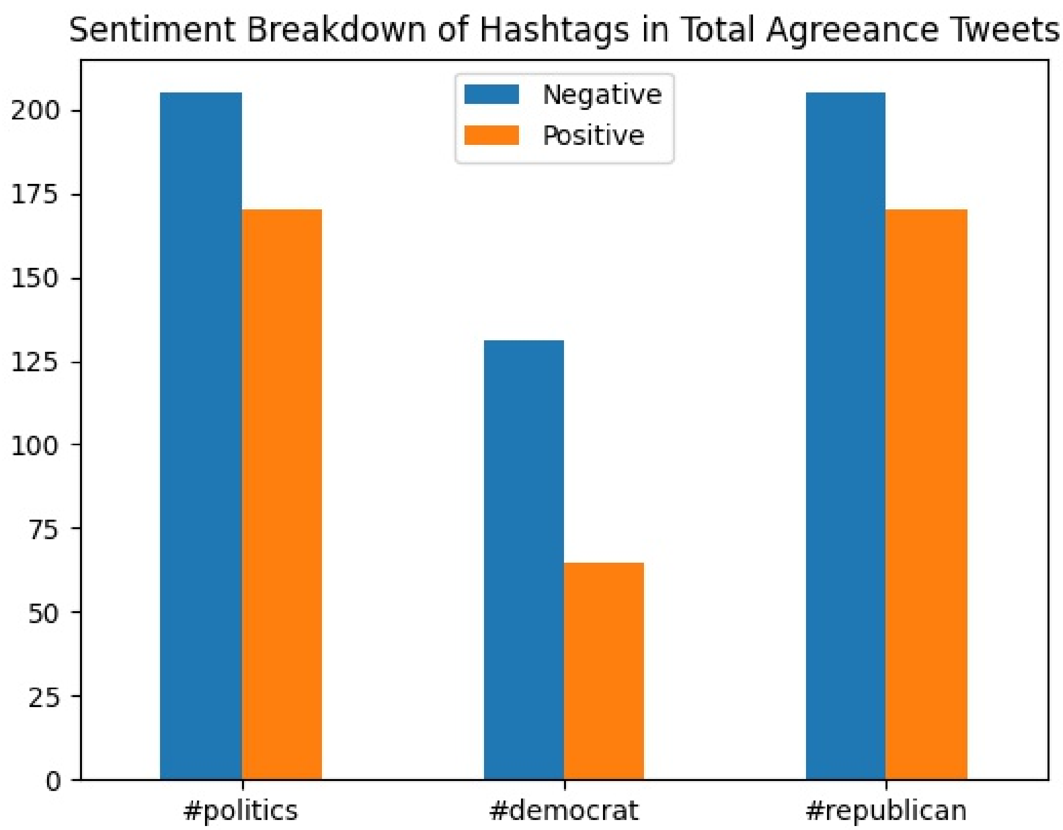

Figure 5: Sentiment of each hashtag recorded in tweets with total agreement.

the opportunity for analysis to be performed in order to draw larger understandings. Sentiment analysis like that examined within this project can provide insight to political figures and those working to pass legislation. By examining the sentiment of individual hashtags the general perception of the parties can be determined. Having this knowledge will allow those trying to sway political perception to appropriately direct their tweet at the audience and sentiment they intended.

In this project certain tweets were excluded from the data set, specifically those that contained media, were retweets or quote tweets, or were replies to other tweets. In the future, adding a method for photo analysis will allow for a wider range of tweets to be examined and create a more complete analysis of the available tweets. Another method that could be used to expand upon this project is working to incorporate the number of quote tweets and retweets. By taking a count of these actions and multiplying the tweet sentiment by the number of times retweeted the number of people who agree with that sentiment will be taken into account. This new value can be factored into the overall sentiment of a given hashtag again providing a more clear vision of the complete sentiment.

References

[1] Shahul ES. *Sentiment Analysis in Python: TextBlob vs Vader Sentiment vs Flair vs Building It From Scratch*. Feb. 2021. URL: `https://neptune.ai/blog/sentiment-analysis-python-textblob-vs-vader-vs-flair`.

[2] Pollyanna Gonçalves et al. "Comparing and combining sentiment analysis methods". In: *Proceedings of the first ACM conference on Online social networks*. 2013, pp. 27–38.

[3] Ali Hasan et al. "Machine Learning-Based Sentiment Analysis for Twitter Accounts". In: *Mathematical and Computational Applications* 23.1 (2018). ISSN: 2297-8747. DOI: `10.3390/mca23010011`. URL: `https://www.mdpi.com/2297-8747/23/1/11`.

[4] C.J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". In: Jan. 2015.

[5] Steven Loria. *textblob Documentation*. Oct. 2020. URL: `https://textblob.readthedocs.io/_/downloads/en/dev/pdf/`.

[6] Wes McKinney. *pandas: powerful Python data analysis toolkit Release 1.2.3*. Mar. 2021. URL: `https://pandas.pydata.org/docs/pandas.pdf`.

[7] Joshua Roesslein. *tweepy Documentation Release 3.7.0*. Nov. 2018. URL: `https://tweepy2.readthedocs.io/_/downloads/en/latest/pdf/`.

[8] Thomas Wolf et al. “Transformers: State-of-the-Art Natural Language Processing”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. URL: `https://www.aclweb.org/anthology/2020.emnlp-demos.6`.