Sentiment Analysis Of U.S. Presidential Election 2020

Sumana Vemuri, Aslam Baig, Bhanu Chander Reddy Sama, Bhavika Bhupatrai Chavda, Saida Muktar, Chandu Somepalli, Venkata Guru Siva Sai Nagarapu

University of Maryland, Baltimore County

Abstract

Trump or Biden; anyone with internet access in this world must for sure know these two names. With being one of the most-watched and followed elections throughout the world, Trump and Biden definitely have given us a show these past few months. These two men are polar opposites that had many people already set their votes from the beginning. However, there were also many things that occurred this past year that could really influence voters' opinions. What were those factors, and how did people's opinions sway in response to them?

Sentiment analysis is an evaluation of the opinion of the reader, writer, speaker, or people who have some opinion towards a certain topic. In the 2020 US presidential election, Donald Trump, Joe Biden, and Bernie Sanders were the top candidates, ending with a battle between Trump and Biden. The opinion of the American people on these candidates will definitely impact who would become the next president of the United States. This paper gives an overview of some of the important events that occurred during the months between January and November, analyzing the responses of people's opinions towards these candidates based on these events. Twitter and major news outlets were used to acquire a large diverse data set which included articles published from the month of January to November 2020 and the tweets of the candidates for the same period of time. Another dataset representing the public opinions on Twitter was collected using election specific keywords for the month of October. The collected tweets and articles were analyzed using different approaches of sentiment analysis and topic modeling, to determine the sentiments of the public. In this paper, we determine the topics that were in discussion throughout the year, the polarity variations, and how the media covered the news related to the candidates, we also cover the social media activity and reach of both Trump and Biden. We also showed how the sentiment of the tweets coming from different states varied and who was popular in these states. By analyzing all these outcomes from social media and digital media in the form of news articles, we understand the public opinion order to help predict who America was favoring at different times.

Introduction

Sentiment analysis is in its simplest form analyzing a text, whether it's a sentence, comment, or an entire document and determining whether it is positive, negative, or neutral. A sentiment analysis system for text analysis combines natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes, and categories within a sentence or phrase.

Sentiment analysis is profoundly useful for quickly gaining insights using large volumes of data. This allows it to extract subjective information in the source material and helps businesses to understand the social sentiment of their service while monitoring online conversations. It can be an essential part of your market research and customer service approach. Not only can you see what people think of your own products or services, but you can also see what they think about your competitors too. This will allow businesses, improve on things that their customers seem to react to negatively, and keep those that they are reacting positively to. The overall customer experience of your users can be revealed quickly with sentiment analysis, but it can get far more granular too. Sentiment analysis is widely used in social media monitoring. It allows us to gain an overview of the public view on certain topics. Social media monitoring via sentiment analysis is beyond just seeing what people like and don't like, but actually can be used to predict a variety of things. For example, shifts in sentiment on social media have been shown to correlate with shifts in the stock market.

Sentiment analysis is also a key tool that has been used to predict and influence elections in a variety of different countries. For example, the Obama administration used sentiment analysis to gauge public opinion to policy announcements and campaign messages ahead of the 2012 presidential election. Being able to quickly see the sentiment behind everything from forum posts to news articles means being better able to strategize and plan for the future.

Forecasting presidential elections has attracted a lot of attention in both academia and the public. Since the 2010s, as the popularity of social media big data increases, a strand of research has started to predict elections based on sentiments expressed on social media platforms such as Twitter. The main goal is usually to calculate the sentiment scores from related social media posts as accurately as possible. If the sentiment is positive towards a candidate, they predict this

candidate will win in the election. Twitter and social media do not give the whole story, however, since many people base their opinions on news outlets they follow. Therefore, this project aims to get outputs on predicting who has the edge to win the 2020 election, by determining the major issues that would influence the election results, comparing the result with social media and news article data, and visualizing various factors influencing the election and showing how people are reacting to Democrats and Republicans.

The best method in order to do these would be sentiment analysis. The reason sentiment analysis was used in this project is because we are using data in predicting peoples' votes, which can be projected by how they feel on certain hot topics of the election. Previous studies have shown that it is possible to predict election outcomes based on Twitter sentiment analysis. This is not something that is mind-blowing as but more of common sense that it could be done. Let's take a moment to think about why this would work. Why would one vote for one candidate instead of another? Their views on a lot of issues probably align with each other, correct? For example, if there is a person let's call him Joe, and he is born and raised in the United States, in a mostly white-dominated area, as he grows older he sees immigrants coming in and "taking" the jobs he used to see his other friends or families doing. Joe is angry and probably has some kind of hatred towards them. Joe is then, for extremely harsh immigration laws. So, if Joe is active on social media and sees a lot of Twitter posts about Trump's immigration laws and keeps retweeting, don't you think he is most likely going to vote for Trump? There are millions of Joe(s) in the United States, and so the reason sentiment analysis was chosen for this study is to analyze the opinions on certain issues of people like Joe, on hot topics that would influence people's votes in order to determine who the majority of voters were swaying towards at different times during this election. This paper is organized as follows. It first reviews the literature on previous similar work that has been done previously. It then differentiates our work from these other works that were reviewed. Moving on to introducing the readers to our research questions and the hypotheses we aim to prove throughout the paper. Next, we dig into our project pipeline talking about the data collection, cleaning, preprocessing, modeling, and visualization. Finally, we talk about the findings from the project and why they are important before a short discussion of our conclusions.

Literature Review

Usually, Sentiment Analysis is done in two main phases. The first being data collection using different platforms and media; and the next phase being the sentiments analyzing the sentiments using the data.

"Tweets may be considered relevant if they contain words from a list of target keywords that are compiled either manually... or in a semi-automatic fashion … through the expansion of a seed set." (Ansari et al., 2020, p. 1822). In their research paper, Ansari et. al. used various approaches to extract the sentiment from the tweets. With lexical analysis, the sentiment was analyzed based on the ratio of occurrence of the words with respect to one another. This happens in unsupervised methodology wherein the number of times the words were used is also calculated.

Advanced models and supervised methods were used when the tweets were either manually classified or according to their emotional context (Wang H. et al., 2020, p. 115-120). When deciding the emotional context, hashtags such as #yay, #upset, :D, :), #excited, etc. were used (Wick & Chambers, 2012, p. 603-612).

Research studies by A. Jain and P. Dandannavar (2017) focused on combining a lexical based approach with a learning-based approach to form a hybrid approach to sentiment analysis (Jain & Dandannavar, 2017, p. 36-42). Words like 'good', 'bad', 'better', etc. are important. Most machine learning techniques are commonly used for document-level analysis (Pang B. & Lee L., 2008, 1-135). The sentiment analysis of this kind is mainly done on 4 main levels. First, being Text Preprocessing, next being Feature Selection Methods, then Feature weighting, and representation schemes, and the last being the ML algorithm (Agarwal & Mittal, 2016, p. 5-13).

"Pang et al. are the first to use unigrams, bigrams, position-based features, POS-based features, adjectives, adverbs, and their combination as features for supervised document-level sentiment analysis." (Agarwal & Mittal, 2016).

From their paper, we can notice that unigram features work best as compared to all the other features.

Another approach was seen in a paper by Matsumoto et. al. wherein syntactic relations between words and word sequences were used. They were utilized as features to perform the sentiment analysis. They also made use of mining algorithms to understand the relation between sentiments and analyze the data. This data was that of movie reviews (Matsumoto S. et al., 2005).

"A proposed lexicon enhanced method for sentiment classification combines machine learning and semantic-orientation approaches into one framework that significantly improves sentiment classification performance." (Dang et al., 2010, p. 1). In their paper, Dang et. al. have utilized a combination of two. They have combined a semantic-orientation-based approach with machine learning algorithms or techniques. It can be noticed that this combination further enhances the sentiment analysis by a significant increase in precision and speed. A combination of unigrams and bigrams were used by them. "Their experimental results showed that rarely used sentiment features enhance the performance of sentiment analysis" (Agarwal & Mittal, 2016, p. 7).

Let us look at something called the appraisal groups or appraisal words. We know that to praise is to say something good. But to add an adverb to it and its combination makes it an appraisal group. "Appraisal groups are phrases that contain word groups like 'very beautiful', 'extremely bad,' etc., unlike individual words. These word groups are intensifiers or modifiers (i.e., very, extremely, etc.) to opinion words (i.e., beautiful, bad, etc.). These appraisal groups are used as features with bag-of-words features for sentiment analysis, resulting in improved performance." (Agarwal & Mittal, 2016, p. 7)

Similarly, sentiment analysis is done in our paper using two main platforms, one being the data collected from Twitter and the other using major News Channels such as CBS, CNN, Fox News, etc. that are published online. The research done from the social media platforms and the news channels brings forward large amounts of data that portrays the perspective of individuals from diverse backgrounds. The data collected is used to analyze behavioral patterns and interactions and changes within them as events passed from January 2020, until the result of the presidential election 2020, i.e. Joe Biden wins the elections.

Most previous works that have been done with sentiment analysis and election predictions have been done with social media platforms such as Twitter. However, we know that these forms of social media might not be accurate since a lot of posts and repost or likes may be purely on a sensitive intensive reaction based on their emotions in that moment. Since the data might purely be based on emotions, it could not give an accurate representation of where people actually stand on topics. For that reason ,we do not completely rely on only Twitter data to make our predictions, we have also used fact-based news articles.

Hypothesis/ Research question

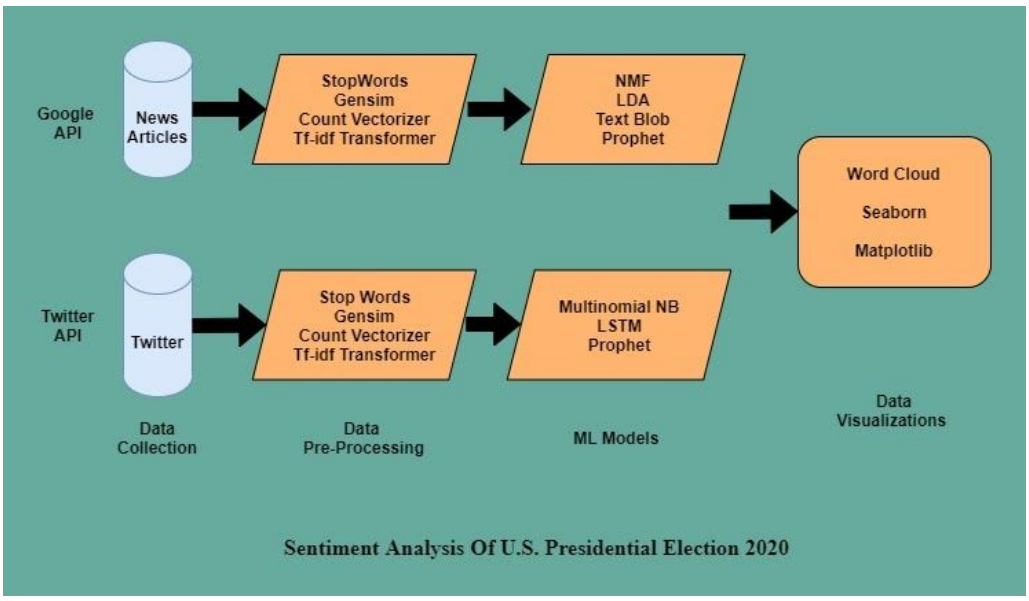
The 2020 election was one of the most-watched elections in the world. We have all seen how close the election was, which ended with Biden winning both the electoral and popular vote while breaking the record of the most popular votes. With an election this close, we have been analyzing what factors were swaying the decisions of voters. 2020 has been a long year with so many events that occurred, starting with Trump's impeachment trials, to COVID, and many more. This paper tries to answer the research question:

1. What are the factors that the American people were most concerned with when it came to choosing their next president?

There are a couple of hypotheses that we are trying to test. At first, we assume that both the candidates have the same chance of winning the election. This is a hypothesis that we will be testing mostly throughout the first few months in order to check where voters stand at the beginning. As previously stated, there is a lot that happened this year that could have swayed voters' opinion in either party's direction, but what were the main events and the second hypothesis is that we are assuming that all the major news outlets would be giving equal coverage to both the candidates and the average sentiment would be same. Our third hypothesis assumes that the most important and heated topics throughout the year that could affect the election are COVID-19, Health Care. Russia, Immigration Policy, Climate Change and China. One event that always makes a huge play in elections is immigration policies, the United States has so many complex views on immigration policies. President Trump won his 2016 election with a huge emphasis on building a wall to stop illegal immigrants, and now has thousands of

illegal immigrants displaced. With all that and more going on, we hypothesize that immigration policy is likely to play a huge part in the election. These past four years, President Trump has downplayed climate change claiming it does not exist, with protests breaking out throughout the country. Our last hypothesis is that both the candidates would be having the same twitter reach and would be equally active on twitter.

Project Pipeline:-



1. **Data Collection:-**

In simple words, data collection refers to the process of collecting data from different sources over a period in order to establish conclusions based on the patterns observed within the data. The basis of this paper is to find sentiment patterns based on the political events of 2020 that led to Joe Biden being elected the president of USA 2020. Herein, we collected data from two main sources i.e. Twitter using Twitter API; and News Channels articles from CBS, FOX, CNN, and NBC, using Google API. The data was collected in an organic fashion (from scratch).

Twitter allows one to interact with its posts or tweets, as it is commonly called, and collect its data using its specific API. It can provide information about specific users when the

user handles (usernames) are used to collect the data. The Twitter API enabled us to retrieve data from tweets related to Joe Biden and Donald Trump, the final candidates for the presidency, across different states of the United States of America. We recorded and collected month-wise tweets of the final candidates i.e. Joe Biden and Donald Trump, from the starting of the year 2020. Along with this we also collected data from general tweets based on election keywords for the month of October. Some of the keywords were 'Joe Biden', 'Donald Trump', 'excellent', 'COVID', 'election', 'fascism', 'America', etc. The entire process of data collection was done organically.

Google makes it possible for users to use its features and platforms by providing its services. These services or platforms are commonly available through what is commonly known as the Google API. The Google API makes it possible to use its features easily through the google maps API, google feeds API, google search API, and google friends API. In our research, we collected videos and news articles from various news channels like CBS, CNN, FOX, and NBC from the past 12 months. The data was collected organically using google API which was then loaded into a data frame which was then converted to a CSV file; which was in turn sent for the data preprocessing.

2. Data Pre-Processing:-

In our Project, since we are dealing with textual data it is important to preprocess the data. As you can see in our data frame, the first steps in the text mining process were to collect unstructured and semi-structured data from multiple data sources like Twitter and news articles. Next, the data was cleaned, converted into a structured format to analyze the patterns (visible and missing). Extracted information can be stored in a database, for example, to assist the decision-making process of an organization. Corpus preparation and cleaning were done using a series of packages running on top of Python such as the Natural Language Toolkit (NLTK) that provides stop-word removal, stemming, lemmatizing, tokenization, identifying n-gram procedures, and other data cleanings like lowercase transformation and punctuation removal. The preprocessing steps are supported in NLTK Library and contain the following patterns:

• Stop-word elimination: Removal of the most common words in a language that are not helpful and in general unusable in text mining like prepositions, numbers, and words that do not contain applicable information for the study. In fact, in NLP, there is no particular general list of stop words used by all developers who choose their list based on their goal to improve the recommendation system performance.

• Stemming: Convert words into their root, using stemming algorithms like Snowball Stemmer.

• Lemmatizing: Enhances system accuracy by returning the base or dictionary form of a word.

• Tokenizing: Divide the text input to tokens like phrases, words, or other meaningful elements resulting in a sequence of tokens.

• Identify n-gram procedures such as bigram (phrases containing two words) and trigram (phrases containing three words) words and consider them as one word.

After the preprocessing step, we have used a concept called gensim which consists of a document(text), corpus(collection of documents), Vector(Mathematical representation of documents), and model(transforming from one to another vector representation). First, we started with a corpus of documents. Next, we transformed these documents into a vector space representation. After that, we have created a model that transforms our original vector representation to TF-IDF. Finally, we used our model to calculate the similarity between some query documents and all the documents in the corpus. And also we applied a commonly used term-weighting method called TF-IDF, which is a pre-filtering stage with all the included TM methods. TF-IDF is a numerical statistical measure used to score the importance of a word (term) in any content from a collection of documents based on the occurrences of each word, and it checks how relevant the keyword is in the corpus. Also, it not only considers the frequency but also induces discriminative information for each term. Term frequency represents how many times a word appears in a document, divided by the total number of words in that document, while inverse document frequency calculates how many documents the term appears in and divides it by the number of documents in the corpus. Furthermore, calculating the TF-IDF weight of a term in a particular document requires calculating term frequency [TF(t, d)], which is the number of times that the word t occurred in document d; document frequency [DF(t)], which is the number of documents in which term t occurs at least once; and inverse document frequency

(IDF), which can be calculated from DF using the following formula. The IDF of a word is considered high if it occurred in a few documents and low if it occurred in many documents. The TF-IDF model is defined in the below Equations:

TF = num of occurrences of the word in documents/num of words in all documents

IDF = log(num of documents / num of documents with word occurs) (Frontiers in Artificial Intelligence, 2020)

3. **Models Used:-**

Topic Modelling:

Topic modeling is a machine learning technique that automatically analyzes text data to determine cluster words for a set of documents. This is known as 'unsupervised' machine learning because it does not require a predefined list of tags or training data that has been previously classified by humans. Topic modeling involves counting words and grouping similar word patterns to infer topics within unstructured data. The methods used are:

1. Latent Dirichlet Allocation(LDA):

Latent Dirichlet Allocation (LDA*)* is based on the same underlying assumptions: the distributional hypothesis, (i.e. similar topics make use of similar words) and the statistical mixture hypothesis (i.e. documents talk about several topics) for which a statistical distribution can be determined. The purpose of LDA is to map each document in our corpus to a set of topics that covers a good deal of the words in the document. LDA assumes that the distribution of topics in a document and the distribution of words in topics are Dirichlet distributions. Two hyperparameters control document and topic similarity, known as alpha and beta, respectively. A low value of alpha will assign fewer topics to each document whereas a high value of alpha will have the opposite effect. A low value of beta will use fewer words to model a topic whereas a high value will use more words, thus making topics more similar between them. A third hyperparameter has to be set when implementing LDA, namely, the number of topics the

algorithm will detect since LDA cannot decide on the number of topics by itself. (Introduction to Topic Modeling, 2019)

In our project, we have used the Gensim library for LDA. In Gensim, a document is an object of the text sequence type (string). A document could be anything from a short 140 character tweet, a single paragraph (i.e., journal article abstract), a news article, or a book. First, we obtain an id-2-word dictionary. For each headline, we will use the dictionary to obtain a mapping of the word id to their word counts. The LDA model uses both of these mappings. Next by generating LDA topics We will iterate over the number of topics, get the top words in each cluster and add them to a data frame.

2. Non-negative Matrix Factorization:

Non-negative Matrix Factorization (NMF) is a Family of linear algebra algorithms for identifying the latent structure in data represented as a non-negative matrix (Lee & Seung, 1999). NMF can be applied for topic modeling, where the input is a document-term matrix, typically TF-IDF normalized. Input Matrix (documents x terms) Input is of Document-term matrix A and Number of topics k. The output is Two k-dimensional factors W and H approximating A. With the rise of complex models like deep learning, we often forget simpler, yet powerful machine learning methods that can be equally powerful. NMF (Nonnegative Matrix Factorization) is an effective machine learning technique. NMF has a wide range of uses, from topic modeling to signal processing (Nonnegative Matrix Factorization); and is a matrix factorization method where we constrain the matrices to be nonnegative. To understand NMF, we should clarify the underlying intuition between matrix factorizations. (*Topic Modelling with Scikit-learn*, 2017, #5)

For the project, we use NMF to obtain a design matrix. To obtain results we are going to apply TF-IDF transformation to the counts. TF-IDF is the Common approach for weighting the score for a term in a document. Where Term Frequency (TF) is the Number of times a given term appears in a single document and Inverse Document Frequency (IDF) is the Function of a total number of distinct documents containing a term. The effect is to penalize common terms

that appear in almost every document. A similar vectorization approach can be used in Scikit-learn to produce a TF-IDF normalized document-term matrix. Again, we should perform additional preprocessing steps by passing the appropriate parameter values to TfidfVectorizer. The TF-IDF model transforms vectors from the bag-of-words representation to a vector space where the frequency counts are weighted according to the relative rarity of each word in the corpus.

3. Multinomial Naive Bayes Theorem:

With an ever-growing amount of textual information stored in electronic form such as legal documents, policies, company strategies, etc., automatic text classification is becoming increasingly important. This requires a supervised learning technique that classifies every new document by assigning one or more class labels from a fixed or predefined class. It uses the bag of words approach, where the individual words in the document constitute its features, and the order of the words is ignored. This technique is different from the way we communicate with each other. It treats the language like it’s just a bag full of words and each message is a random handful of them. Large documents have a lot of words that are generally characterized by very high dimensionality feature space with thousands of features. Hence, the learning algorithm requires to tackle high dimensional problems, both in terms of classification performance and computational speed. NAÏVE BAYES which is computationally very efficient and easy to implement is a learning algorithm frequently used in text classification problems. Two event models are commonly used namely Multivariate Bernoulli Event Model and Multivariate Event Model. The Multivariate Event model is referred to as Multinomial Naive Bayes. (Multinomial Naive Bayes Explained, 2020)

In our project, firstly we'll convert the raw messages (sequence of characters) into vectors (sequences of numbers) by using “split_into_tokens(message)”.The mapping is not 1-to-1. we'll use the bag of words approach, where each unique word in a text will be represented by one number.

BAG OF WORDS MODEL: It is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag(multiset) of its words, disregarding grammar and even word order but keeping multiplicity. The bag-of-words model has also been used for computer vision. The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each word is used as a feature for training a classifier.

Secondly, we used split_into_lemmas(a message which is the algorithmic process of determining the lemma of a word based on its intended meaning). Unlike stemming, lemmatization depends on correctly identifying the intended part of speech and meaning of a word in a sentence, as well as within the larger context surrounding that sentence, such as neighboring sentences or even an entire document.

Now we'll convert each message, represented as a list of tokens (lemmas) above, into a vector that machine learning models can understand.

Doing that requires essentially three steps, in the bag-of-words model:

1. counting how many times does a word occur in each message (term frequency)

2. weighting the counts, so that frequent tokens get lower weight (inverse document frequency)

3. normalizing the vectors to unit length, to abstract from the original text length

And finally, after the counting, the term weighting and normalization can be done with TF-IDF, using scikit-learn TfidfTransformer. There are a multitude of ways in which data can be preprocessed and vectorized. These two steps, also called "feature engineering", are typically the most time-consuming parts of building a predictive pipeline, but they are very important and require some experience. The trick is to evaluate constantly: analyze a model for the errors it makes, improve data cleaning & preprocessing, brainstorm for new features, evaluate. With messages represented as vectors, we can finally train our spam/ham classifier. This part is

straightforward, and there are many libraries that realize the training algorithms. We'll be using Naïve Bayes classifier to determine the accuracy of the model. There are many cases where polarity is zero because there is some data which either doesn't contain any text or simply have links or hashtags only. By using the Multinomial Naive Bayes theorem, our model will perform text classification for a given set of text (*Practical Data Science in Python*, 2014).

4. Text Blob:

Text Blob is a python library for Natural Language Processing (NLP). Text Blob actively used Natural Language ToolKit (NLTK) to achieve its tasks. NLTK is a library that gives easy access to a lot of lexical resources and allows users to work with categorization, classification, and many other tasks. TextBlob is a simple library that supports complex analysis and operations on textual data. For lexicon-based approaches, a sentiment is defined by its semantic orientation and the intensity of each word in the sentence. This requires a pre-defined dictionary classifying negative and positive words. Generally, a text message will be represented by a bag of words. After assigning individual scores to all the words, the final sentiment is calculated by some pooling operation like taking an average of all the sentiments. Typically, we quantify this sentiment with a positive or negative value, called polarity. The overall sentiment is often inferred as positive, neutral or negative from the sign of the polarity score. TextBlob returns the polarity and subjectivity of a sentence. Polarity lies between [-1,1], -1 defines a negative sentiment and 1 defines a positive sentiment. Negation words reverse the polarity. (Towards Data Science, 2020). For the project, we generated the polarity of tweets made from Joe Biden and Donald Trump, and also news articles published by different media. The sentiment score is generated based on a comparison of tweet words with positive and negative words lexicon. There are multiple ways to calculate sentiment scores. The approach followed here is to count the positive, neutral, and negative words in each tweet and assign a sentiment score. This way, the tweets can be ascertained how positive/ negative it is. The output gives us various graphical representations for an easy and better understanding of the sentiment on Trump and Biden.

5. Markovify:

Markovify is a Python package billed as a "simple, extensible Markov chain generator" used to build Markov models from a large corpus of text. Markovify works by reading a raw text as a string, splitting the input into sentences, and generating sentences based on parameters.

In the project, we generated Biden tweets made on Trump and Trump's tweets made on Biden using Markovify by giving our overall dataset as the text input.

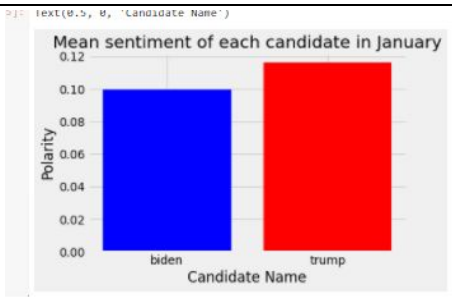
6. Long Short Term Memory(LSTM):

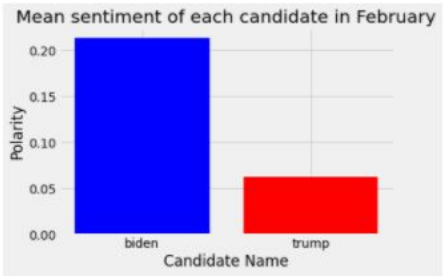
Long Short Term Memory networks, usually called "LSTMs", were introduced by Hoch Reiter and Schmidhuber. These have widely been used for speech recognition, language modeling, sentiment analysis, and text prediction. Before going deep into LSTM, we should first understand the need for LSTM which can be explained by the drawback of practical use of Recurrent Neural Network (RNN). So, let's start with RNN.

A drawback of RNN: During the training of RNN, the information goes in a loop again and again which results in very large updates to neural network model weights. This is due to the accumulation of error gradients during an update and hence, results in an unstable network. At an extreme, the values of weights can become so large as to overflow and result in NaN values. The explosion occurs through exponential growth by repeatedly multiplying gradients through the network layers that have values larger than 1 or vanishing occurs if the values are less than 1. Due to the above drawback of RNN, the scientists developed and invented a new variant of the RNN model, called the Long Short Term Memory. LSTM can solve this problem because it uses gates to control the memorizing process. (Towards Data Science, 2018)

4. Research Findings:-

MONTH WISE TOPIC MODELLING OF THE ARTICLES:-

<table>
<tr><th>Month</th><th>LDA</th><th>NMF</th><th>Sentiment</th></tr>
<tr>
<td>January</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>Trump</td><td>Biden</td><td>Democratic</td><td>Trump</td><td>Trump</td></tr>
<tr><td>1</td><td>said</td><td>Trump</td><td>CNN</td><td>Democratic</td><td>election</td></tr>
<tr><td>2</td><td>CBS</td><td>election</td><td>Trump</td><td>said</td><td>said</td></tr>
<tr><td>3</td><td>vote</td><td>Joe</td><td>said</td><td>President</td><td>Democratic</td></tr>
<tr><td>4</td><td>Biden</td><td>Democratic</td><td>voters</td><td>Republican</td><td>President</td></tr>
<tr><td>5</td><td>voters</td><td>Senate</td><td>political</td><td>Biden</td><td>Biden</td></tr>
<tr><td>6</td><td>House</td><td>President</td><td>main</td><td>Democrats</td><td>News</td></tr>
<tr><td>7</td><td>News</td><td>Mayor</td><td>President</td><td>campaign</td><td>Republican</td></tr>
<tr><td>8</td><td>President</td><td>would</td><td>impeachment</td><td>CNN</td><td>political</td></tr>
<tr><td>9</td><td>first</td><td>said</td><td>–</td><td>—</td><td>percent</td></tr>
<tr><td>10</td><td>told</td><td>Sen.</td><td>Election</td><td>state</td><td>Republicans</td></tr>
<tr><td>11</td><td>percent</td><td>campaign</td><td>election</td><td>–</td><td>news</td></tr>
<tr><td>12</td><td>—</td><td>Republican</td><td>percent</td><td>would</td><td>vote</td></tr>
<tr><td>13</td><td>CNN</td><td>vote</td><td>named</td><td>New</td><td>Democrats</td></tr>
<tr><td>14</td><td>National</td><td>win</td><td>sources</td><td>Iowa</td><td>Joe</td></tr>
<tr><td>15</td><td>Tuesday</td><td>Election</td><td>U.S.</td><td>voters</td><td>state</td></tr>
<tr><td>16</td><td>Republican</td><td>seat</td><td>Biden</td><td>presidential</td><td>Donald</td></tr>
<tr><td>17</td><td>would</td><td>Donald</td><td>trial</td><td>Senate</td><td>NBC</td></tr>
<tr><td>18</td><td>ballots</td><td>Republicans</td><td>town</td><td>debate</td><td>presidential</td></tr>
<tr><td>19</td><td>according</td><td>political</td><td>(CNN)</td><td>(CNN)</td><td>trust</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>started</td><td>main</td><td>biden</td><td>fox</td><td>cnn</td></tr>
<tr><td>1</td><td>iphone</td><td>sources</td><td>percent</td><td>trust</td><td>debate</td></tr>
<tr><td>2</td><td>magsafe</td><td>political</td><td>trump</td><td>trial</td><td>town</td></tr>
<tr><td>3</td><td>feature</td><td>adults</td><td>president</td><td>republicans</td><td>network</td></tr>
<tr><td>4</td><td>best</td><td>named</td><td>voters</td><td>hannity</td><td>democratic</td></tr>
<tr><td>5</td><td>getting</td><td>news</td><td>election</td><td>news</td><td>halls</td></tr>
<tr><td>6</td><td>12</td><td>source</td><td>vote</td><td>lineup</td><td>sanders</td></tr>
<tr><td>7</td><td>new</td><td>americans</td><td>votes</td><td>impeachment</td><td>warren</td></tr>
<tr><td>8</td><td>bronx</td><td>election</td><td>said</td><td>ran</td><td>february</td></tr>
<tr><td>9</td><td>torres</td><td>for</td><td>state</td><td>primetime</td><td>hampshire</td></tr>
<tr><td>10</td><td>san</td><td>oct</td><td>college</td><td>host</td><td>presidential</td></tr>
<tr><td>11</td><td>york</td><td>emerges</td><td>states</td><td>media</td><td>candidates</td></tr>
<tr><td>12</td><td>south</td><td>outlets</td><td>joe</td><td>democrats</td><td>qualify</td></tr>
<tr><td>13</td><td>breed</td><td>makeup</td><td>hunter</td><td>distrust</td><td>participate</td></tr>
<tr><td>14</td><td>francisco</td><td>commonly</td><td>electoral</td><td>trusted</td><td>iowa</td></tr>
<tr><td>15</td><td>ritchie</td><td>alignment</td><td>county</td><td>moderate</td><td>android</td></tr>
<tr><td>16</td><td>15th</td><td>versus</td><td>in</td><td>conservative</td><td>apps</td></tr>
<tr><td>17</td><td>priebus</td><td>tail</td><td>ballots</td><td>less</td><td>air</td></tr>
<tr><td>18</td><td>results</td><td>remaining</td><td>the</td><td>interviews</td><td>watch</td></tr>
<tr><td>19</td><td>city</td><td>resulted</td><td>wisconsin</td><td>flagship</td><td>national</td></tr>
</table>
</td>
<td>


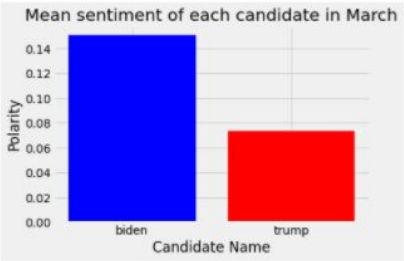
Some of the important topics were about the Impeachment motion against Trump and the democrats nominations. For the month of Jan, we can see that the mean sentiment for trump was greater than that of Biden.
</td>
</tr>
<tr>
<td>February</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>Trump</td><td>Biden</td><td>Democratic</td><td>Trump</td><td>Trump</td></tr>
<tr><td>1</td><td>said</td><td>Trump</td><td>CNN</td><td>Democratic</td><td>election</td></tr>
<tr><td>2</td><td>CBS</td><td>election</td><td>Trump</td><td>said</td><td>said</td></tr>
<tr><td>3</td><td>vote</td><td>Joe</td><td>said</td><td>President</td><td>Democratic</td></tr>
<tr><td>4</td><td>Biden</td><td>Democratic</td><td>voters</td><td>Republican</td><td>President</td></tr>
<tr><td>5</td><td>voters</td><td>Senate</td><td>political</td><td>Biden</td><td>Biden</td></tr>
<tr><td>6</td><td>House</td><td>President</td><td>main</td><td>Democrats</td><td>News</td></tr>
<tr><td>7</td><td>News</td><td>Mayor</td><td>President</td><td>campaign</td><td>Republican</td></tr>
<tr><td>8</td><td>President</td><td>would</td><td>impeachment</td><td>CNN</td><td>political</td></tr>
<tr><td>9</td><td>first</td><td>said</td><td>–</td><td>—</td><td>percent</td></tr>
<tr><td>10</td><td>told</td><td>Sen.</td><td>Election</td><td>state</td><td>Republicans</td></tr>
<tr><td>11</td><td>percent</td><td>campaign</td><td>election</td><td>–</td><td>news</td></tr>
<tr><td>12</td><td>—</td><td>Republican</td><td>percent</td><td>would</td><td>vote</td></tr>
<tr><td>13</td><td>CNN</td><td>vote</td><td>named</td><td>New</td><td>Democrats</td></tr>
<tr><td>14</td><td>National</td><td>win</td><td>sources</td><td>Iowa</td><td>Joe</td></tr>
<tr><td>15</td><td>Tuesday</td><td>Election</td><td>U.S.</td><td>voters</td><td>state</td></tr>
<tr><td>16</td><td>Republican</td><td>seat</td><td>Biden</td><td>presidential</td><td>Donald</td></tr>
<tr><td>17</td><td>would</td><td>Donald</td><td>trial</td><td>Senate</td><td>NBC</td></tr>
<tr><td>18</td><td>ballots</td><td>Republicans</td><td>town</td><td>debate</td><td>presidential</td></tr>
<tr><td>19</td><td>according</td><td>political</td><td>(CNN)</td><td>(CNN)</td><td>trust</td></tr>
</table>
</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>started</td><td>main</td><td>biden</td><td>fox</td><td>cnn</td></tr>
<tr><td>1</td><td>iphone</td><td>sources</td><td>percent</td><td>trust</td><td>debate</td></tr>
<tr><td>2</td><td>magsafe</td><td>political</td><td>trump</td><td>trial</td><td>town</td></tr>
<tr><td>3</td><td>feature</td><td>adults</td><td>president</td><td>republicans</td><td>network</td></tr>
<tr><td>4</td><td>best</td><td>named</td><td>voters</td><td>hannity</td><td>democratic</td></tr>
<tr><td>5</td><td>getting</td><td>news</td><td>election</td><td>news</td><td>halls</td></tr>
<tr><td>6</td><td>12</td><td>source</td><td>vote</td><td>lineup</td><td>sanders</td></tr>
<tr><td>7</td><td>new</td><td>americans</td><td>votes</td><td>impeachment</td><td>warren</td></tr>
<tr><td>8</td><td>bronx</td><td>election</td><td>said</td><td>ran</td><td>february</td></tr>
<tr><td>9</td><td>torres</td><td>for</td><td>state</td><td>primetime</td><td>hampshire</td></tr>
<tr><td>10</td><td>san</td><td>oct</td><td>college</td><td>host</td><td>presidential</td></tr>
<tr><td>11</td><td>york</td><td>emerges</td><td>states</td><td>media</td><td>candidates</td></tr>
<tr><td>12</td><td>south</td><td>outlets</td><td>joe</td><td>democrats</td><td>qualify</td></tr>
<tr><td>13</td><td>breed</td><td>makeup</td><td>hunter</td><td>distrust</td><td>participate</td></tr>
<tr><td>14</td><td>francisco</td><td>commonly</td><td>electoral</td><td>trusted</td><td>iowa</td></tr>
<tr><td>15</td><td>ritchie</td><td>alignment</td><td>county</td><td>moderate</td><td>android</td></tr>
<tr><td>16</td><td>15th</td><td>versus</td><td>in</td><td>conservative</td><td>apps</td></tr>
<tr><td>17</td><td>priebus</td><td>tail</td><td>ballots</td><td>less</td><td>air</td></tr>
<tr><td>18</td><td>results</td><td>remaining</td><td>the</td><td>interviews</td><td>watch</td></tr>
<tr><td>19</td><td>city</td><td>resulted</td><td>wisconsin</td><td>flagship</td><td>national</td></tr>
</table>
</td>
<td>


Some of the important topics were about the Impeachment motion against Trump and the democrat's nominations, the democrat's campaign, the Senate debate, and presidential candidates. For the month of February, we can see that the mean sentiment for Biden was greater than that of trump.
</td>
</tr>
</table>

March

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Biden	Broncos	Republican	Senate	coronavirus
1	coronavirus	coronavirus	said	said	said
2	said	Biden	told	Trump	Democratic
3	Democratic	vote	Trump	coronavirus	President
4	voting	game	--	President	media
5	House	said	News	(CNN)	Biden
6	election	economic	coronavirus	US	economic
7	President	would	CBS	House	think
8	Joe	video	officials	Pelosi	viewers
9	state	Democrats	Biden	Democratic	national
10	Trump	Senate	House	emergency	pay
11	health	mail	President	Republicans	Trump
12	Sanders	voting	Democrats	public	Sanders
13	Democrats	U.S.	announced	saying	primary
14	would	million	National	Democrats	CNN
15	Republicans	last	Thursday	McConnell	they've
16	primary	people	(CNN)	would	--
17	Tuesday	--	mail	still	China
18	presidential	Joe	three	aid	Broncos
19	(CNN)	Denver	DeWine	package	Super

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	iphone	senate	biden	same	really
1	started	republicans	tuesday	viewers	walden
2	magsafe	stimulus	delegates	wall	treasure
3	feature	rolled	primary	cable	smart
4	12	economic	sanders	coronavirus	he
5	best	aid	super	cnn	understand
6	getting	intended	joe	news	think
7	new	critical	democratic	said	coming
8	race	deliver	race	66	re
9	baltimore	trillion	vote	message	we
10	mayoral	businesses	polls	watchers	national
11	maryland	plan	state	msnbc	abates
12	hogan	thursday	voting	constant	greg
13	media	us	dewine	accurately	attempt
14	westmoreland	public	health	covering	congressman
15	trump	american	baltimore	52	said
16	york	cnn	mayoral	tv	confined
17	mount	democrats	states	network	currently
18	rushmore	payments	president	vaccine	response
19	county	proposal	election	37	mitigate

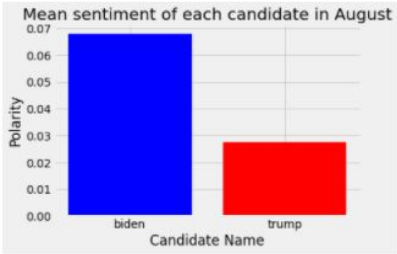


Some of the important topics were **Coronavirus, National emergency, China, Economy.** For the month of March, we can see that the mean sentiment for Biden was greater than that of trump.

April

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Democrats	Trump	Trump	Democratic	said
1	Trump	said	CBS	said	Democratic
2	health	coronavirus	House	White	President
3	public	say	coronavirus	Trump	Trump
4	black	protests	White	coronavirus	absentee
5	driven	President	said	people	coronavirus
6	said	state	state	state	state
7	would	Democratic	campaign	campaign	told
8	protests	(CNN)	Democratic	Biden	York
9	House	according	administration	states	United
10	coronavirus	statewide	political	calling	Senate
11	CBS	campaign	antibody	House	New
12	experts	governors	pandemic	United	ballot
13	say	told	Party	health	--
14	Biden	House	York	--	CBS
15	state	response	states	LaSalle	campaign
16	campaign	White	President	funding	sent
17	Senate	China	News	County	States
18	--	Chinese	went	government	CNN
19	people	racism	New	could	Pennsylvania

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	permission	73	trump	driven	protests
1	edit	hospitalized	democrats	states	june
2	article	place	democratic	stay	society
3	city	across	campaign	health	racism
4	york	restrictions	said	home	say
5	top	covid	biden	public	overall
6	there	19	white	dakota	now
7	new	000	president	orders	poll
8	car	new	black	decisions	cnn
9	feb	cuomo	senate	governors	24
10	indication	stahl	house	experts	despite
11	spend	minutes	state	at	ssrs
12	owned	60	coronavirus	facts	opinion
13	asian	10	the	cases	compared
14	field	rural	voters	saw	negative
15	unless	york	majority	south	steady
16	edited	cbs	news	science	largely
17	zero	sunday	presidential	statewide	justified
18	declares	news	cnn	notion	27
19	adds	people	administration	the	conducted

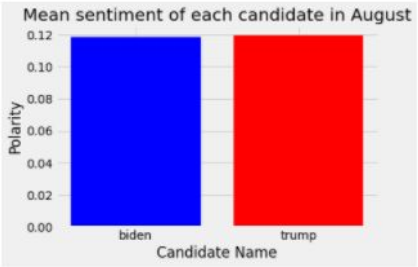


Some of the important topics were about **coronavirus, racism, Black lives matter, protest, statewide campaign, pandemic, united health, Biden campaign**. For the month of April, we can see that the mean sentiment for Biden was greater than that of trump.

May

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Democratic	Trump	Trump	Biden	Trump
1	said	coronavirus	President	Trump	President
2	Trump	national	coronavirus	would	said
3	House	news	Democratic	said	Republican
4	President	House	new	campaign	mask
5	Democrats	White	House	coronavirus	Yoho
6	state	New	women	state	get
7	--	President	Republican	President	(CNN)
8	Republican	said	voters	House	Carolina
9	last	TV	Donald	told	wearing
10	think	state	North	health	first
11	Senate	would	would	new	House
12	voters	voting	record	made	--
13	Biden	say	campaign	Sen.	abuse
14	told	told	called	Democrats	North
15	(CNN)	ballots	Senate	adviser	coronavirus
16	presidential	states	Vice	case	child
17	debate	Mike	Biden	--	week
18	New	election	voted	New	CNN
19	said.	Election	Republicans	(CNN)	state

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	bash	magsafe	trump	carolina	campaign
1	dana	feature	said	north	ellis
2	but	iphone	state	convention	sex
3	time	12	president	four	trump
4	unemployment	started	house	away	legalized
5	hassett	best	biden	gov	marriage
6	rate	getting	democrats	thing	pedophilia
7	double	new	republican	carried	bestiality
8	may	hampshire	national	tell	same
9	state	sanders	mask	played	2019
10	claims	iowa	democratic	emerging	cnn
11	debate	voters	coronavirus	purple	epstein
12	june	candidates	in	percentage	jeffrey
13	ballots	minds	voting	laboratory	made
14	could	none	the	fascinating	reviewed
15	democrats	slipped	think	gauging	obergefell
16	senior	convinced	new	cooper	gay
17	level	gains	cnn	ago	predicted
18	economic	steadiness	told	months	obscurity
19	tv	characterized	mail	decisions	hodges

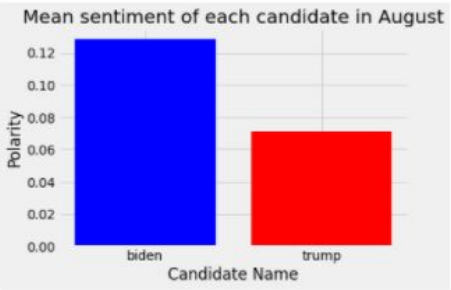


Some of the important topics were about **coronavirus, mask, election campaign, unemployment, economic-related**. For the month of May, we can see that the mean sentiment for Biden and Trump is almost the same.

June

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Trump	Trump	Senate	Trump	House
1	Democratic	said	said	said	Democratic
2	President	Democratic	President	Biden	White
3	Senate	Senate	Trump	Republican	Biden
4	said	primary	state	Democratic	Senate
5	primary	campaign	percent	voting	said
6	women	Democrats	Democratic	presidential	told
7	state	House	--	Democrats	Trump
8	Biden	Orange	McGrath	told	primary
9	presidential	County	House	delegates	state
10	Republicans	—	voter	like	presidential
11	GOP	President	state's	federal	would
12	would	New	Court	support	people
13	House	migrant	going	(CNN)	make
14	could	facing	Democrats	President	nomination
15	vote	political	Booker	people	—
16	violence	American	like	police	wear
17	Democrats	people	NRA	Sen.	President
18	—	calls	Republicans	National	campaign
19	Joe	Wayne	states	administration	(CNN)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	trump	delegates	cuomo	breakfast	white
1	primary	charlotte	andrew	enemies	house
2	biden	pageantry	reopen	faith	judd
3	democratic	meetings	and	allies	spokesman
4	president	24	homey	pelosi	deere
5	state	aug	matilda	know	cnn
6	said	assemble	status	going	services
7	republicans	before	deaths	arthur	manchurian
8	000	300	gives	inappropriate	brig
9	senate	speeches	undated	soon	updated
10	party	convention	mario	arrival	stands
11	donald	jacksonville	governor	join	distinguished
12	2016	moving	hospitalizations	remark	piece
13	votes	start	briefing	sometimes	servant
14	voting	business	parents	justification	eight
15	presidential	weekend	desire	prayer	comment
16	election	hold	stage	mormon	armed
17	in	three	photo	convict	tata
18	republican	days	cannot	romney	really
19	women	gop	covid	class	the

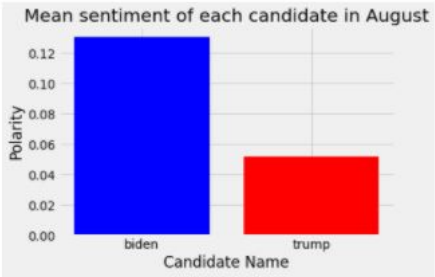


Some of the important topics were **migrants, presidential election, women vice president-elect, hospitalizations, national administrations**. For the month of June, we can see that the mean sentiment for Biden is more than 50% of Trump's sentiment

July

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Trump	President	Trump	said	Trump
1	said	Trump	President	Trump	campaign
2	campaign	coronavirus	say	Republican	Democratic
3	Republican	said	ballots	House	House
4	President	Trump's	supporters	Senate	President
5	presidential	rate	Hong	Trump's	Republican
6	would	House	told	President	two
7	Democratic	case	said	Sen.	–
8	House	US	—	election	presidential
9	White	fatality	election	week	said
10	Democrats	new	Facebook	Democrats	people
11	:	Tuesday	think	federal	told
12	political	protesters	mail	going	Department
13	Republicans	people	Republican	Republicans	White
14	people	(CNN)	federal	(CNN)	last
15	–	Americans	according	Supreme	Republicans
16	Donald	Chinese	CBS	orders	State
17	News	two	state	Election	would
18	Senate	Democrats	freedom	White	Party
19	Trump's	Jewish	campaign	could	political

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	iphone	senate	mcenany	trump	vice
1	started	executive	comments	election	pence
2	feature	democrats	trump	voting	president
3	magsafe	orders	added	mail	lacrosse
4	getting	house	think	campaign	wisconsin
5	12	white	donald	republican	mike
6	best	republicans	mainstream	ballots	speech
7	new	said	sense	tillis	didn
8	debate	meadows	equally	in	dana
9	video	people	sharpton	2020	mishandled
10	spin	expected	showman	app	record
11	above	staff	outright	voters	cooperative
12	coverage	to	consistent	absentee	lutheran
13	player	filibuster	hateful	presidential	lunchtime
14	coach	week	only	said	gave
15	room	chief	october	north	dairyland
16	cbsn	eliminate	hln	carolina	kozlov
17	600	50	needed	president	pre
18	carried	cnn	apologize	universal	commitment
19	february	so	mexicans	the	swung

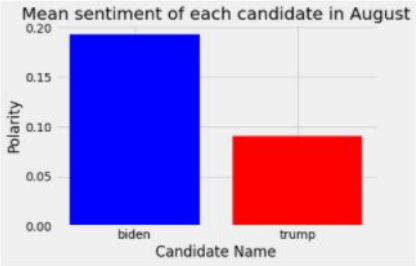


Some of the important topics were about **coronavirus, supreme orders, white democrats, mail voting, campaign mishandled the pandemic situation**. For the month of June, we can see that the mean sentiment for Biden is more than that of Trump's sentiment.

August

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Trump	Trump	Trump	Trump	said
1	million	Democratic	Democratic	Biden	Postal
2	–	said	White	Democratic	Trump
3	A18-49:	National	–	Republican	House
4	Social	President	say	Donald	would
5	Democratic	campaign	said	President	say
6	(R)Viewers:	political	President	political	Service
7	campaign	Convention	like	Joe	Americans
8	2	would	Americans	National	voters
9	Biden	absentee	political	–	coronavirus
10	Security	election	Democrats	first	—
11	said	Republican	House	said	Democratic
12	Democrats	Biden	voters	night	National
13	CBS	state	generic	Party	election
14	National	Democrats	new	convention	Biden
15	Republican	August	coronavirus	A18-49:	Donald
16	–	House	govern	Convention	Democrats
17	President	(CNN)	Republican	asked	President
18	House	Washington	Party	voters	United
19	told	–	majority	CNN	Republican

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	iphone	trump	history	hospitalized	results
1	magsafe	biden	stelter	73	46th
2	started	convention	competition	restrictions	projects
3	best	said	never	across	resultsjoe
4	feature	democratic	fox	place	statepresidential
5	12	mail	seen	covid	become
6	getting	president	like	19	presidential
7	new	in	news	000	us
8	virtual	democrats	govern	new	biden
9	dean	campaign	majority	vaccine	cnn
10	evening	republican	start	pledge	president
11	hundreds	national	white	leading	precincts
12	joint	ballot	party	confidence	hill
13	source	election	democrats	nine	iowa
14	ballots	the	kente	pharmaceutical	fiona
15	convention	cnn	asked	coronavirus	caucuses
16	in	postal	levels	unprecedented	university
17	according	party	backbone	signed	moines
18	good	republicans	aggressively	public	polling
19	live	vote	mcintosh	number	putin

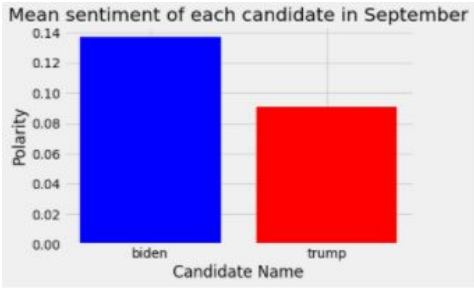


Some of the important topics were about **coronavirus, National president campaign, absentee election, Biden campaign, pharmaceutical related to covid, mail-in ballots**. For the month of August, we can see that the mean sentiment for Biden is more than that of Trump's sentiment.

September

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Supreme	Trump	Trump	Trump	said
1	Court	Supreme	said	Democratic	Senate
2	Senate	Court	Supreme	said	Court
3	said	Biden	Democrats	President	New
4	House	Democratic	Biden	Supreme	Supreme
5	Trump	said	political	would	Trump
6	told	President	—	Democrats	political
7	political	Democrats	would	Court	York
8	Republicans	would	Democratic	Biden	presidential
9	people	election	President	—	Republican
10	Democratic	Republicans	House	Donald	nominee
11	President	Joe	say	Joe	would
12	vote	coronavirus	Court	mail-in	election
13	Biden	Senate	White	told	says
14	presidential	Trump's	US	House	November
15	election	presidential	Republicans	Senate	—
16	would	told	presidential	nominee	Republicans
17	(CNN)	new	election	Justice	voters
18	2020	campaign	according	presidential	people
19	nominee	—	Joe	vote	Justice

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	greenspan	court	mail	margins	ads
1	never	supreme	absentee	senate	political
2	anything	trump	voting	presidential	facebook
3	crisis	justice	in	90	let
4	seen	president	postal	differences	users
5	ve	election	service	biden	see
6	covid	republicans	guilfoyle	lineup	people
7	like	nominee	universal	races	social
8	vaccine	said	officials	incorporate	issue
9	polls	senate	vote	leading	outright
10	trust	ginsburg	election	maybe	entirely
11	said	nomination	voters	exception	aim
12	19	would	ballots	91	misinformation
13	ll	democrats	must	presidency	library
14	politics	biden	attempt	chances	targeting
15	interfere	new	claim	average	features
16	what	the	distinction	single	neither
17	health	ruth	forms	takes	limiting
18	cnn	bader	indemocratic	indeed	main
19	coronavirus	confirmation	confuse	becomes	blog

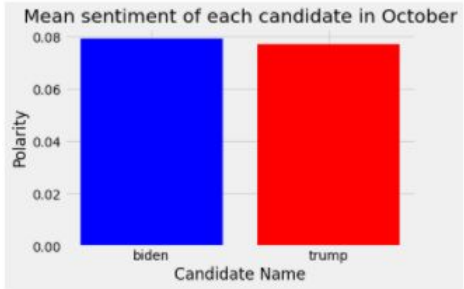


For the month of September, we can see that the mean sentiment for Biden was higher than that of trump and major topics were **Supreme Court, Coronavirus, Senate race, and facebook's political ads.**

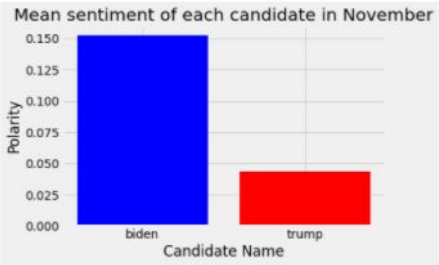
October

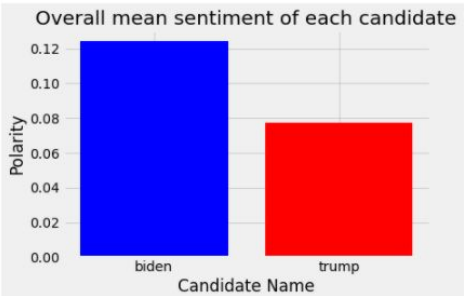
	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	Trump	Trump	Trump	said	Trump
1	said	House	Biden	Trump	said
2	Biden	President	President	election	Senate
3	Democrats	said	said	Biden	President
4	House	Biden	Joe	Democrats	Biden
5	campaign	White	voting	Republican	Democrats
6	President	told	debate	would	House
7	states	Republican	--	political	campaign
8	Republicans	campaign	Donald	Supreme	Republican
9	--	--	first	presidential	Republicans
10	Democratic	Democratic	Republicans	ballots	presidential
11	Republican	Joe	Democrats	House	state
12	swing	would	presidential	voters	election
13	voters	president	Democratic	President	told
14	would	state	campaign	campaign	White
15	White	Democrats	voters	Senate	--
16	officials	Election	House	state	News
17	Joe	Donald	—	—	voters
18	Sen.	tested	president	told	Democratic
19	Donald	former	election	White	(CNN)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05
0	nan	degrees	iphone	hospitalized	trump
1	zuckerberg	transport	magsafe	restrictions	biden
2	exceed	pfizer	12	73	president
3	exactly	94	feature	000	election
4	exaggerated	vaccine	best	place	cnn
5	exaggerating	keep	started	across	said
6	examination	bright	getting	19	house
7	examined	thanksgiving	new	covid	campaign
8	example	fda	hampshire	new	joe
9	exceedingly	guidelines	arizona	positive	democrats
10	executive	whistleblower	oct	hampshire	the
11	exceeds	development	supreme	tested	state
12	exception	house	voting	county	white
13	exchange	drives	requirement	pandemic	voters
14	excited	oklahomans	bush	thanksgiving	donald
15	excitement	happening	court	oct	court
16	exclusive	kitchens	delaware	first	debate
17	exact	tulsasome	york	lady	presidential
18	exacerbated	basket	university	many	republicans
19	evidence	baskets	poll	white	senate



Interestingly right before the elections, we can see that the mean sentiment for both Trump and Biden was the same for the month of October. Some of the important topics were about **Covid, Supreme Court, Pfizer Vaccine, and ballots.**

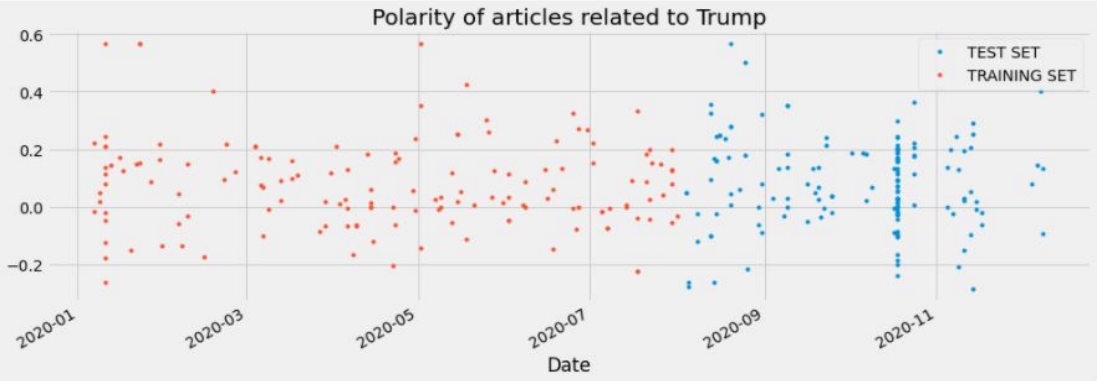
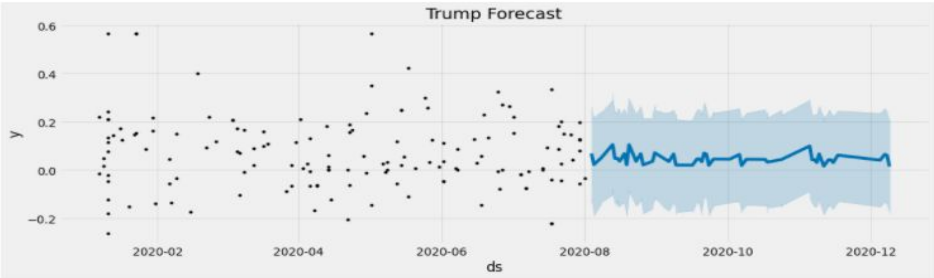
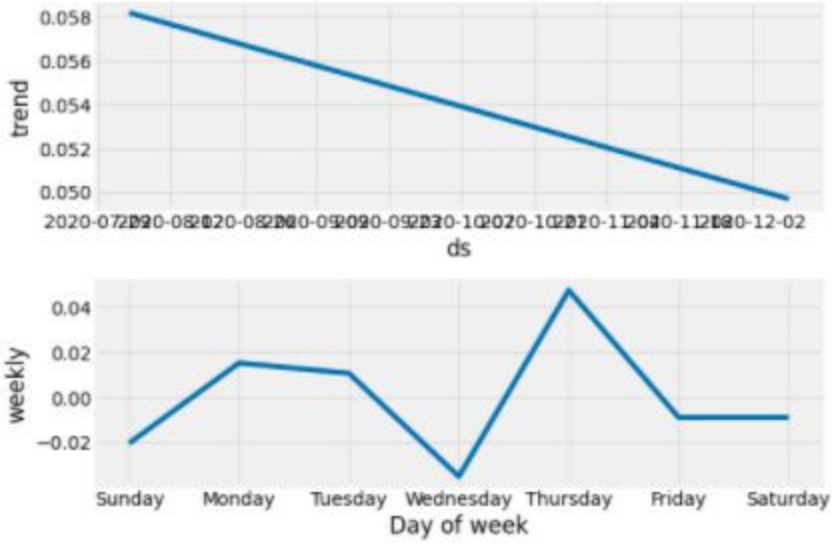
<table>
<tr>
<td>November</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>Trump</td><td>Trump</td><td>Biden</td><td>Biden</td><td>Biden</td></tr>
<tr><td>1</td><td>Biden</td><td>Biden</td><td>Trump</td><td>campaign</td><td>Trump</td></tr>
<tr><td>2</td><td>—</td><td>said</td><td>election</td><td>Trump</td><td>House</td></tr>
<tr><td>3</td><td>President</td><td>President</td><td>—</td><td>Joe</td><td>White</td></tr>
<tr><td>4</td><td>Joe</td><td>Senate</td><td>Republican</td><td>—</td><td>President</td></tr>
<tr><td>5</td><td>said</td><td>Georgia</td><td>percent</td><td>House</td><td>campaign</td></tr>
<tr><td>6</td><td>Donald</td><td>election</td><td>said</td><td>Republican</td><td>election</td></tr>
<tr><td>7</td><td>House</td><td>Joe</td><td>campaign</td><td>Democratic</td><td>vote</td></tr>
<tr><td>8</td><td>Republican</td><td>would</td><td>House</td><td>President</td><td>said</td></tr>
<tr><td>9</td><td>transition</td><td>Donald</td><td>President</td><td>said</td><td>Democratic</td></tr>
<tr><td>10</td><td>campaign</td><td>Republican</td><td>Biden's</td><td>Democrats</td><td>Obama</td></tr>
<tr><td>11</td><td>would</td><td>Democrats</td><td>White</td><td>White</td><td>Republican</td></tr>
<tr><td>12</td><td>state</td><td>state</td><td>Joe</td><td>election</td><td>Joe</td></tr>
<tr><td>13</td><td>White</td><td>campaign</td><td>Trump's</td><td>transition</td><td>Georgia</td></tr>
<tr><td>14</td><td>votes</td><td>voters</td><td>officials</td><td>told</td><td>Democrats</td></tr>
<tr><td>15</td><td>officials</td><td>lawsuit</td><td>votes</td><td>voters</td><td>state</td></tr>
<tr><td>16</td><td>told</td><td>House</td><td>Republicans</td><td>first</td><td>first</td></tr>
<tr><td>17</td><td>first</td><td>(CNN)</td><td>say</td><td>presidential</td><td>votes</td></tr>
<tr><td>18</td><td>official</td><td>—</td><td>state</td><td>would</td><td>News</td></tr>
<tr><td>19</td><td>presidential</td><td>court</td><td>confident"</td><td>Georgia</td><td>Biden's</td></tr>
</table>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>iphone</td><td>biden</td><td>kaitlan</td><td>votes</td><td>grant</td></tr>
<tr><td>1</td><td>getting</td><td>trump</td><td>ask</td><td>percent</td><td>that</td></tr>
<tr><td>2</td><td>magsafe</td><td>president</td><td>jiang</td><td>electoral</td><td>nih</td></tr>
<tr><td>3</td><td>feature</td><td>campaign</td><td>press</td><td>biden</td><td>wuhan</td></tr>
<tr><td>4</td><td>started</td><td>said</td><td>weijia</td><td>49</td><td>lab</td></tr>
<tr><td>5</td><td>best</td><td>house</td><td>answer</td><td>trump</td><td>work</td></tr>
<tr><td>6</td><td>12</td><td>transition</td><td>reporter</td><td>college</td><td>year</td></tr>
<tr><td>7</td><td>new</td><td>election</td><td>question</td><td>georgia</td><td>true</td></tr>
<tr><td>8</td><td>journalists</td><td>the</td><td>china</td><td>currently</td><td>scott</td></tr>
<tr><td>9</td><td>data</td><td>white</td><td>questions</td><td>vote</td><td>lightly</td></tr>
<tr><td>10</td><td>things</td><td>state</td><td>refused</td><td>winner</td><td>spending</td></tr>
<tr><td>11</td><td>title</td><td>joe</td><td>collins</td><td>pennsylvania</td><td>terminate</td></tr>
<tr><td>12</td><td>ix</td><td>senate</td><td>conference</td><td>project</td><td>reauthorized</td></tr>
<tr><td>13</td><td>transition</td><td>elect</td><td>response</td><td>win</td><td>daszak</td></tr>
<tr><td>14</td><td>schools</td><td>department</td><td>cbs</td><td>projected</td><td>cause</td></tr>
<tr><td>15</td><td>win</td><td>obama</td><td>white</td><td>states</td><td>collaborating</td></tr>
<tr><td>16</td><td>months</td><td>republican</td><td>told</td><td>arizona</td><td>usually</td></tr>
<tr><td>17</td><td>that</td><td>democrats</td><td>house</td><td>counted</td><td>pelley</td></tr>
<tr><td>18</td><td>promised</td><td>presidential</td><td>cnn</td><td>nevada</td><td>taken</td></tr>
<tr><td>19</td><td>biden</td><td>convention</td><td>trump</td><td>yet</td><td>when</td></tr>
</table>


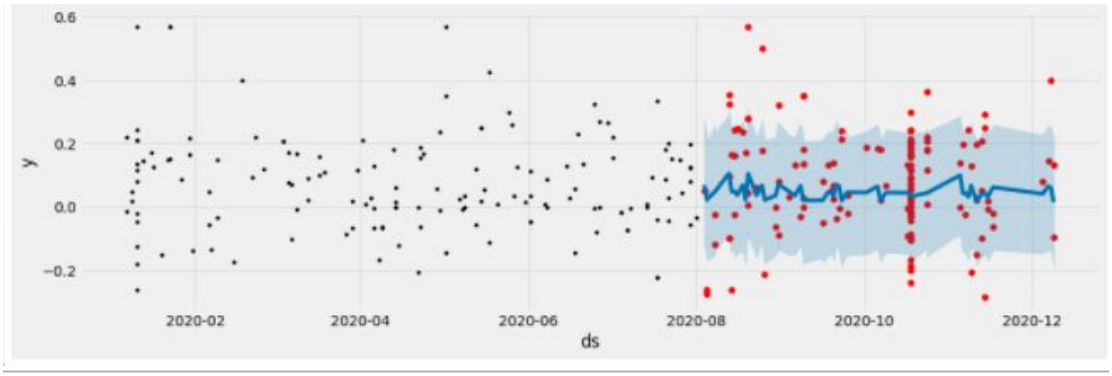
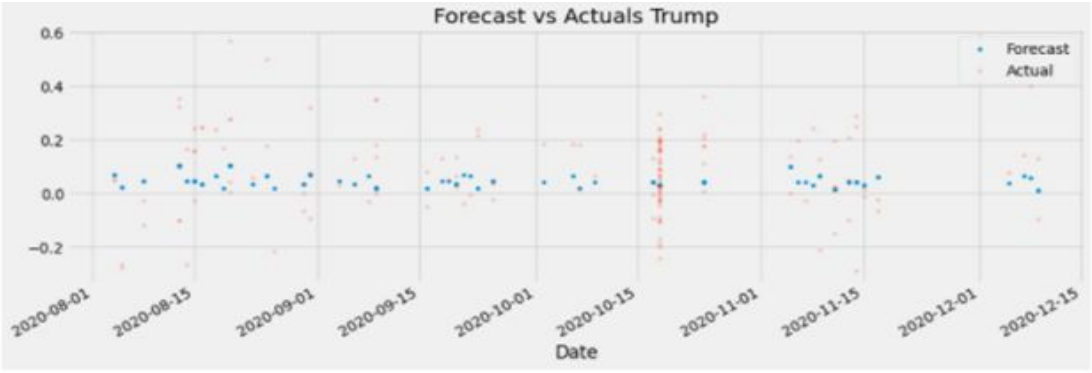
In the month of November, the articles mainly focused on topics such as the **transition process** as the initial trends showed that joe Biden was leading the race and **several lawsuits** that were filed in many battleground states by republicans. The mean sentiment for Biden was almost twice that of trump for the month of November.
</td>
</tr>
<tr>
<td>Overall Jan-November</td>
<td>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>said</td><td>Trump</td><td>Trump</td><td>Trump</td><td>Trump</td></tr>
<tr><td>1</td><td>Trump</td><td>President</td><td>said</td><td>said</td><td>President</td></tr>
<tr><td>2</td><td>Democratic</td><td>House</td><td>Democratic</td><td>Biden</td><td>said</td></tr>
<tr><td>3</td><td>Biden</td><td>campaign</td><td>Biden</td><td>Democrats</td><td>House</td></tr>
<tr><td>4</td><td>President</td><td>Democrats</td><td>election</td><td>Democratic</td><td>White</td></tr>
<tr><td>5</td><td>Republican</td><td>Biden</td><td>CNN</td><td>election</td><td>Democratic</td></tr>
<tr><td>6</td><td>House</td><td>said</td><td>--</td><td>Republican</td><td>Senate</td></tr>
<tr><td>7</td><td>Democrats</td><td>would</td><td>Republican</td><td>state</td><td>would</td></tr>
<tr><td>8</td><td>Republicans</td><td>political</td><td>would</td><td>--</td><td>coronavirus</td></tr>
<tr><td>9</td><td>—</td><td>Democratic</td><td>House</td><td>voting</td><td>told</td></tr>
<tr><td>10</td><td>New</td><td>coronavirus</td><td>presidential</td><td>Senate</td><td>voters</td></tr>
<tr><td>11</td><td>Donald</td><td>Joe</td><td>President</td><td>would</td><td>Supreme</td></tr>
<tr><td>12</td><td>Senate</td><td>Republican</td><td>White</td><td>Republicans</td><td>Court</td></tr>
<tr><td>13</td><td>people</td><td>told</td><td>campaign</td><td>President</td><td>say</td></tr>
<tr><td>14</td><td>presidential</td><td>(CNN)</td><td>coronavirus</td><td>Joe</td><td>vote</td></tr>
<tr><td>15</td><td>Joe</td><td>Donald</td><td>told</td><td>campaign</td><td>campaign</td></tr>
<tr><td>16</td><td>state</td><td>Trump's</td><td>political</td><td>House</td><td>(CNN)</td></tr>
<tr><td>17</td><td>campaign</td><td>White</td><td>News</td><td>new</td><td>Donald</td></tr>
<tr><td>18</td><td>two</td><td>presidential</td><td>voters</td><td>—</td><td>CNN</td></tr>
<tr><td>19</td><td>Sen.</td><td>--</td><td>say</td><td>political</td><td>News</td></tr>
</table>
<table>
<tr><th></th><th>Topic # 01</th><th>Topic # 02</th><th>Topic # 03</th><th>Topic # 04</th><th>Topic # 05</th></tr>
<tr><td>0</td><td>iphone</td><td>trump</td><td>pfizer</td><td>permission</td><td>73</td></tr>
<tr><td>1</td><td>magsafe</td><td>biden</td><td>94</td><td>article</td><td>hospitalized</td></tr>
<tr><td>2</td><td>feature</td><td>president</td><td>transport</td><td>app</td><td>restrictions</td></tr>
<tr><td>3</td><td>started</td><td>said</td><td>degrees</td><td>campaign</td><td>covid</td></tr>
<tr><td>4</td><td>12</td><td>election</td><td>vaccine</td><td>fauci</td><td>across</td></tr>
<tr><td>5</td><td>best</td><td>cnn</td><td>keep</td><td>trump</td><td>place</td></tr>
<tr><td>6</td><td>getting</td><td>campaign</td><td>bourla</td><td>wear</td><td>000</td></tr>
<tr><td>7</td><td>new</td><td>democratic</td><td>coronavirus</td><td>don</td><td>19</td></tr>
<tr><td>8</td><td>york</td><td>the</td><td>pharmaceutical</td><td>gunter</td><td>new</td></tr>
<tr><td>9</td><td>hampshire</td><td>democrats</td><td>pledge</td><td>surveillance</td><td>crisis</td></tr>
<tr><td>10</td><td>city</td><td>house</td><td>fda</td><td>egregious</td><td>anything</td></tr>
<tr><td>11</td><td>sanders</td><td>state</td><td>chance</td><td>contacts</td><td>never</td></tr>
<tr><td>12</td><td>bush</td><td>senate</td><td>companies</td><td>pair</td><td>seen</td></tr>
<tr><td>13</td><td>iowa</td><td>republican</td><td>nine</td><td>tool</td><td>ve</td></tr>
<tr><td>14</td><td>main</td><td>voters</td><td>developing</td><td>technology</td><td>like</td></tr>
<tr><td>15</td><td>voters</td><td>republicans</td><td>confidence</td><td>developing</td><td>coronavirus</td></tr>
<tr><td>16</td><td>cuomo</td><td>in</td><td>works</td><td>mask</td><td>cuomo</td></tr>
<tr><td>17</td><td>mayoral</td><td>joe</td><td>unprecedented</td><td>extraordinary</td><td>vaccine</td></tr>
<tr><td>18</td><td>candidates</td><td>presidential</td><td>approval</td><td>requests</td><td>york</td></tr>
<tr><td>19</td><td>race</td><td>vote</td><td>signed</td><td>power</td><td>positive</td></tr>
</table>


The above graph shows the overall mean sentiment of each candidate. We can see that the overall Biden polarity is more as compared to Trump. From topic modelling we found the major topics affecting this trend is coronavirus, presidential campaign, vaccine, pharmaceutical, travel restrictions, employment.
</td>
</tr>
</table>

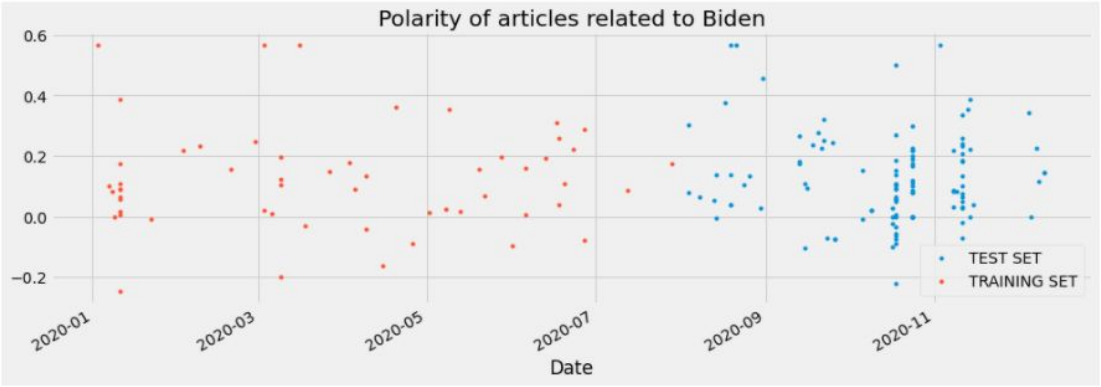
Trump Polarity Forecast using Prophet Model

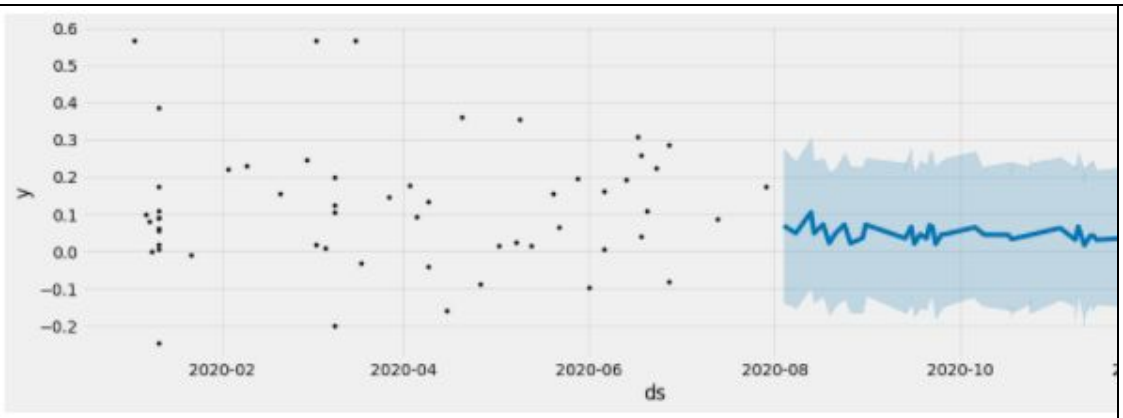
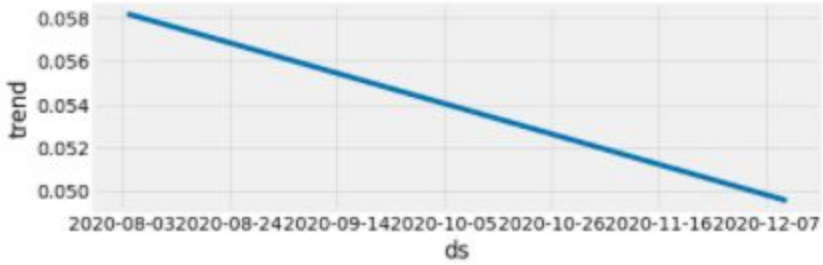
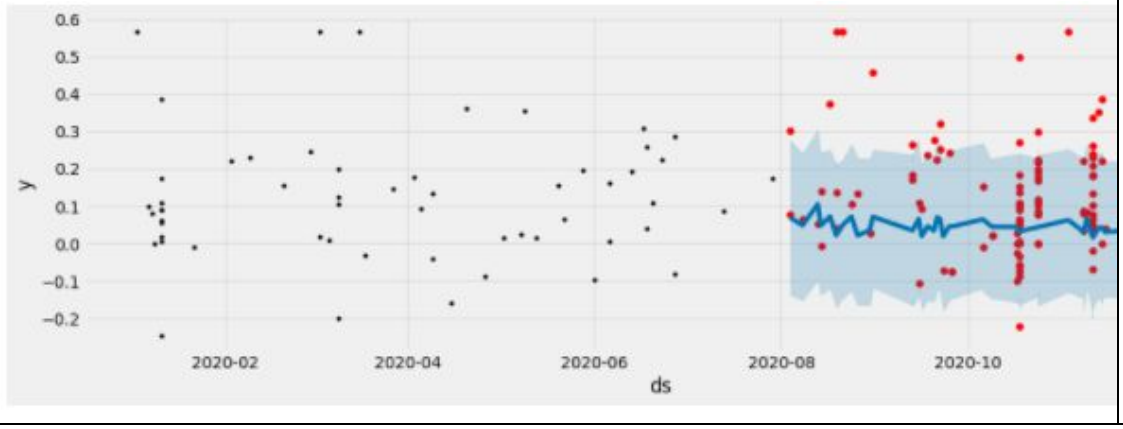
TRUMP	CONCLUSION

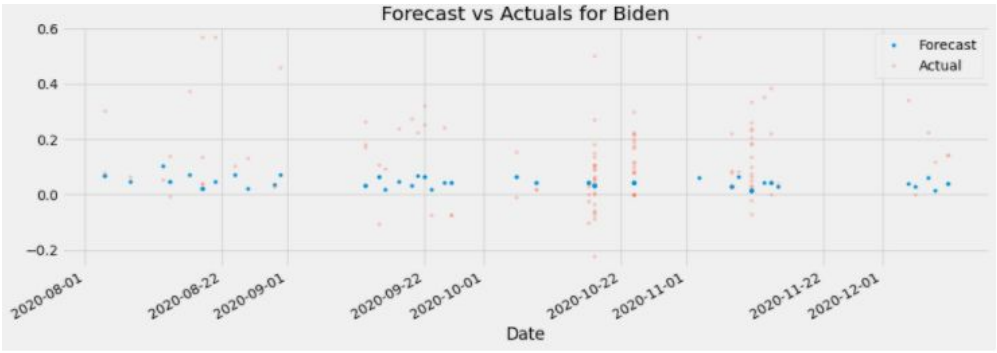
	We have split the data into training and test datasets and the data points are well distributed.
Plotting forecast 	Using the prophet model we have forecasted how the polarity would vary for the articles that are going to be related to trump.
	By using the prophet model we have seen how the trend is varying for Trump and we can also see how this has been varying throughout the week.

Plot the forecast with the actuals 	Our model has performed significantly well as we have obtained a Mean Squared Error 0.0249 and Mean Absolute Error of 0.126.
	A more detailed look at the forecast vs Actuals for trump related articles.

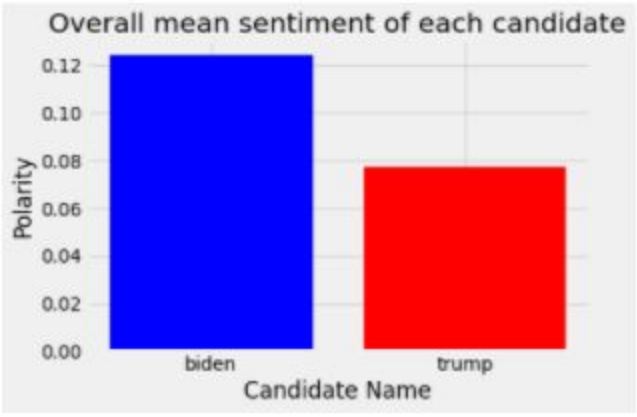
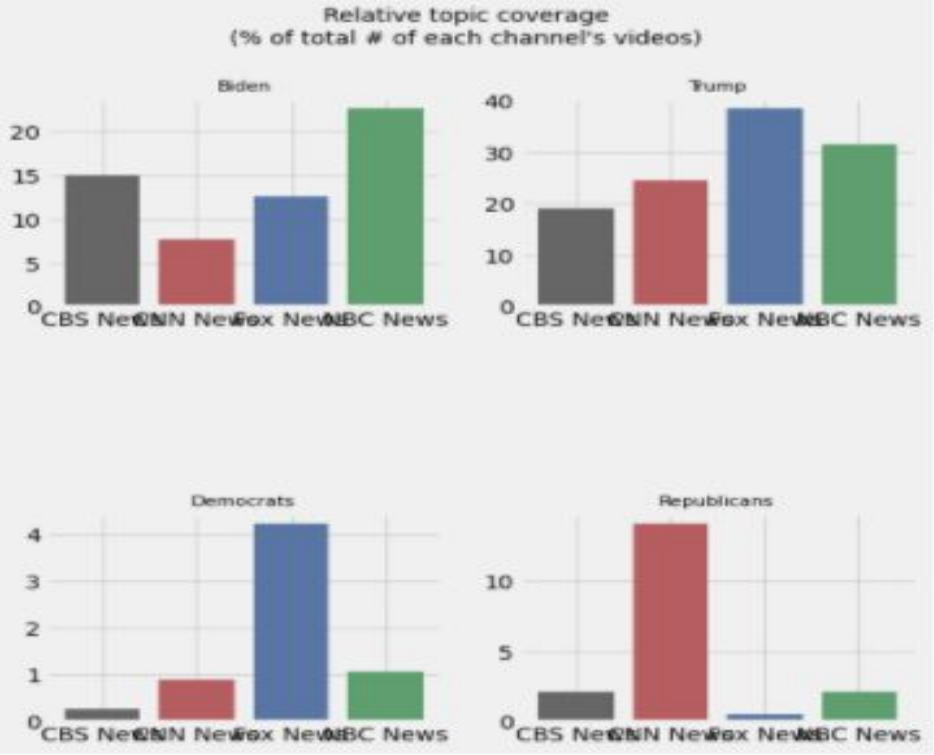
Biden- prophet forecast model

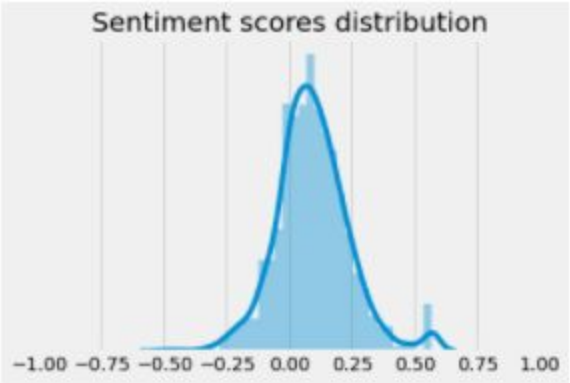
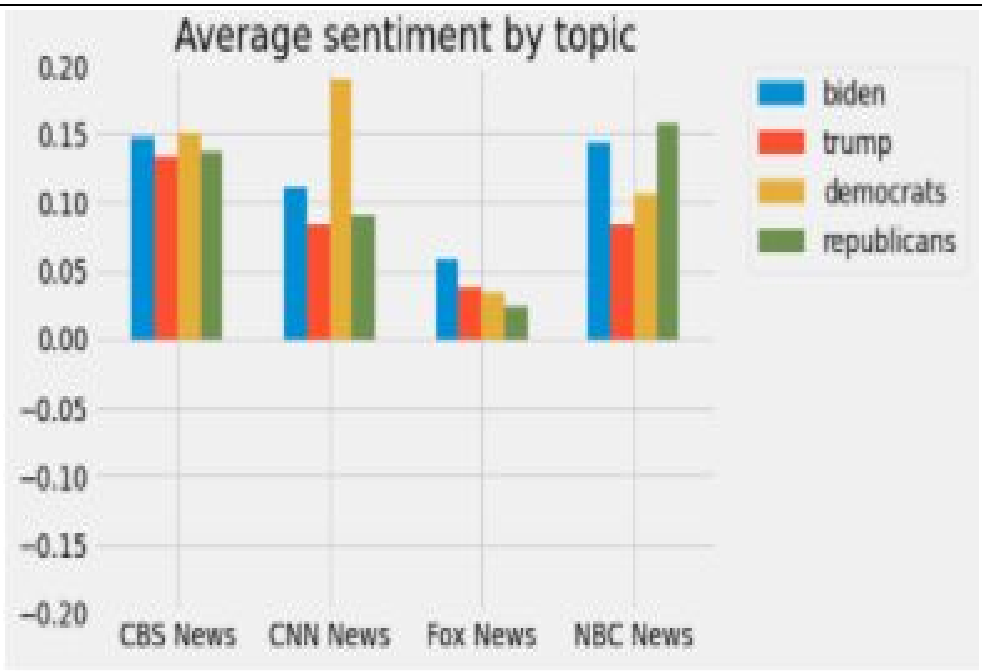
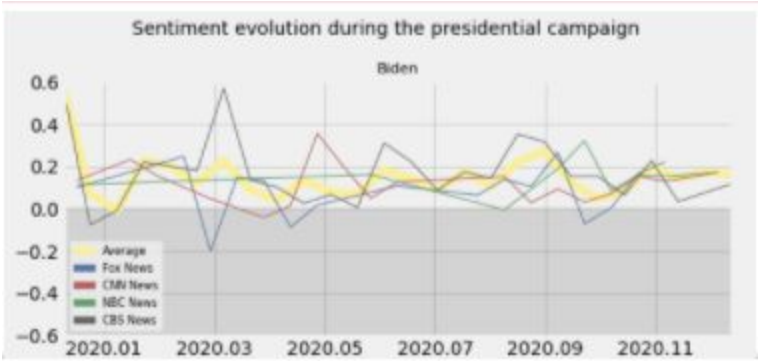
Biden	conclusion
	We have split the data into training and test datasets and the data points are well distributed.
Plot the forecast	

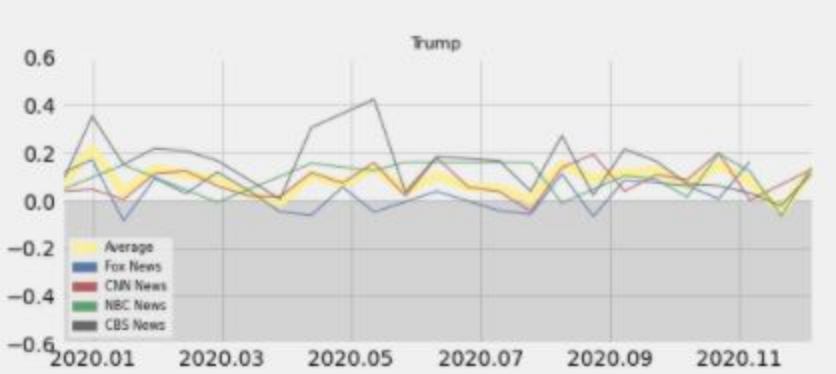
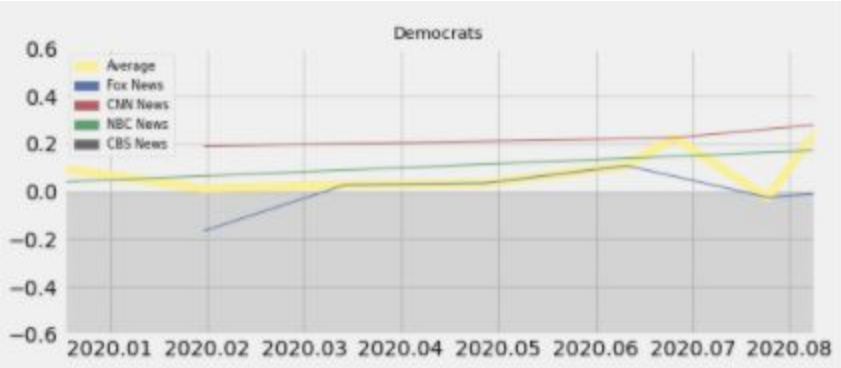
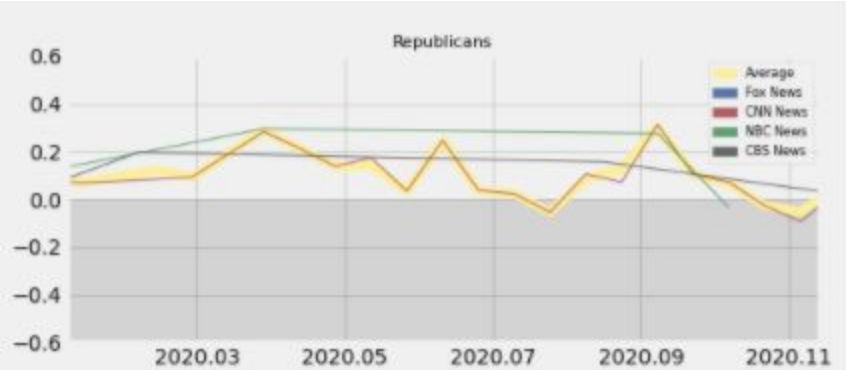
	Using the prophet model we have forecasted how the polarity would vary for the articles that are going to be related to Biden.
 	By using the prophet model we have seen how the trend is varying for Biden and we can also see how this has been varying throughout the week. We can observe that the trend is the same for both the candidates.
Plot forecast with actuals 	Our model has performed significantly well as we have obtained a Mean Squared Error 0.0297 and Mean Absolute Error of 0.1262.

	A more detailed look at the forecast vs Actuals for trump related articles.

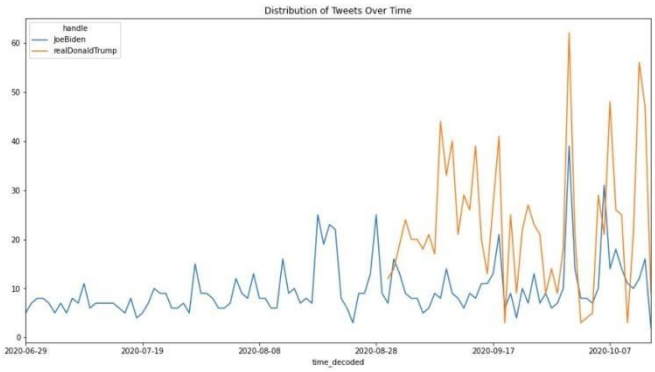
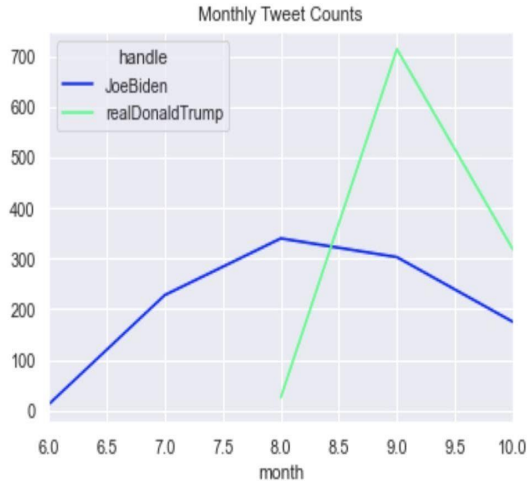
Overall graphs

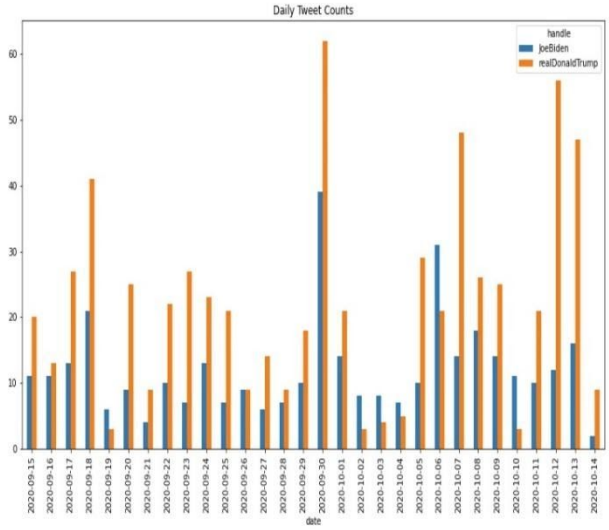
Graph	conclusion
	If we consider all the articles from the month of Jan to November, we can see that the mean polarity for Biden is more as compared to Trump.
	Since the data is collected for various news channels, we are showing how each news channel is reacting for Biden, Trump, Democrats, and Republicans.

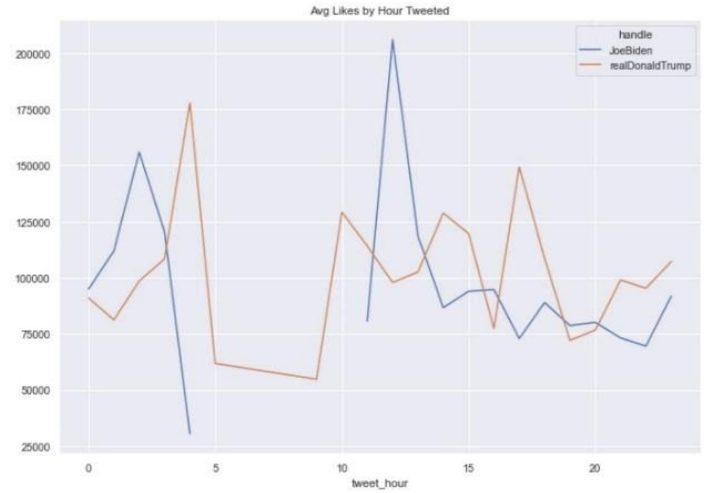
	The graph shows sentiment scores distribution for all the articles related to Biden, Trump, Democrats, and Republicans.
	We can see that CBS has maintained a far neutral approach while covering election topics and CNN has covered quite extensively about democrats. Interestingly even the fox news average sentiment for Biden is more than that of trump and NBC has a slightly higher average sentiment for republicans than for democrats.
	This is a time-series graph for sentiment evolution during the presidential campaign for Biden. Here, we see how different news channels are distributed for different months and the average sentiment for all news channels for the articles related to Biden

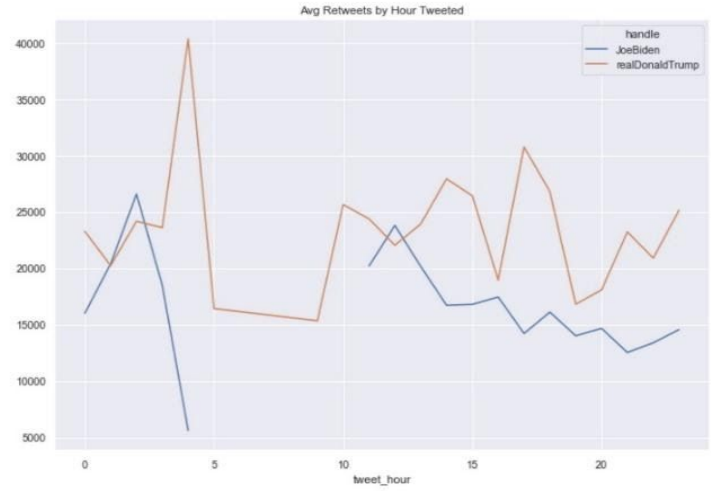
	This is a time-series graph for sentiment evolution during the presidential campaign for Trump. Here, we see how different news channels are distributed for different months and the average sentiment for all news channels for the articles related to Trump
	According to the Average sentiment evolution of the presidential campaign, we can say that during January and March, the sentiment was neutral, but during July it was quite positive. CNN sentiment for all the month is positive whereas Fox news sentiment was negative during January but later in the month of June the sentiment was positive.
	From this graph of the average sentiment evolution of the republicans, we can see that the average sentiment trend has been decreasing in the last couple of months before the elections

Candidates Twitter Handle Analysis:-

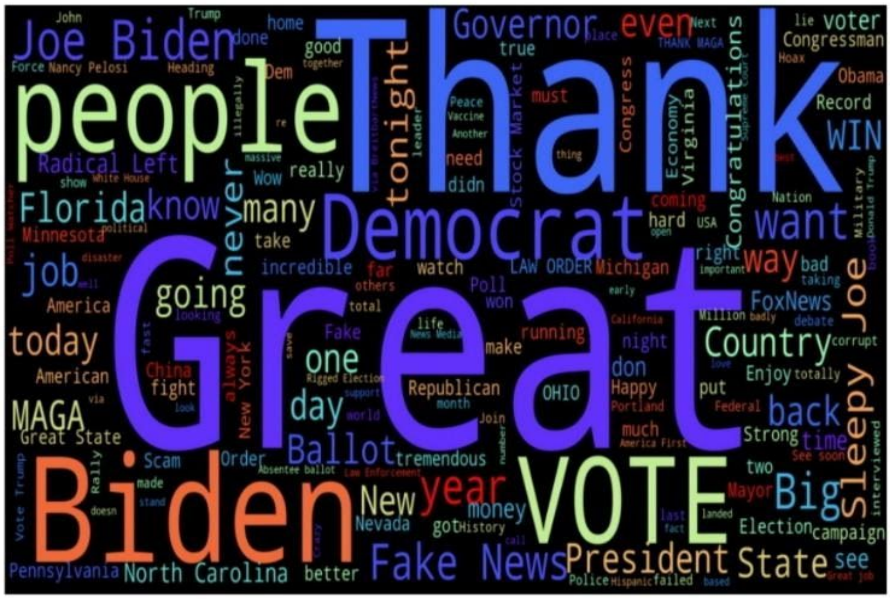
 	From the visualizations, we can conclude that Biden was less active compared to Trump on twitter, as Trump tweeted twice as many tweets as Biden from the month of August to October. But Joe Biden's distribution was well distributed over time as he maintained a constant rate of tweets for a longer period of time.

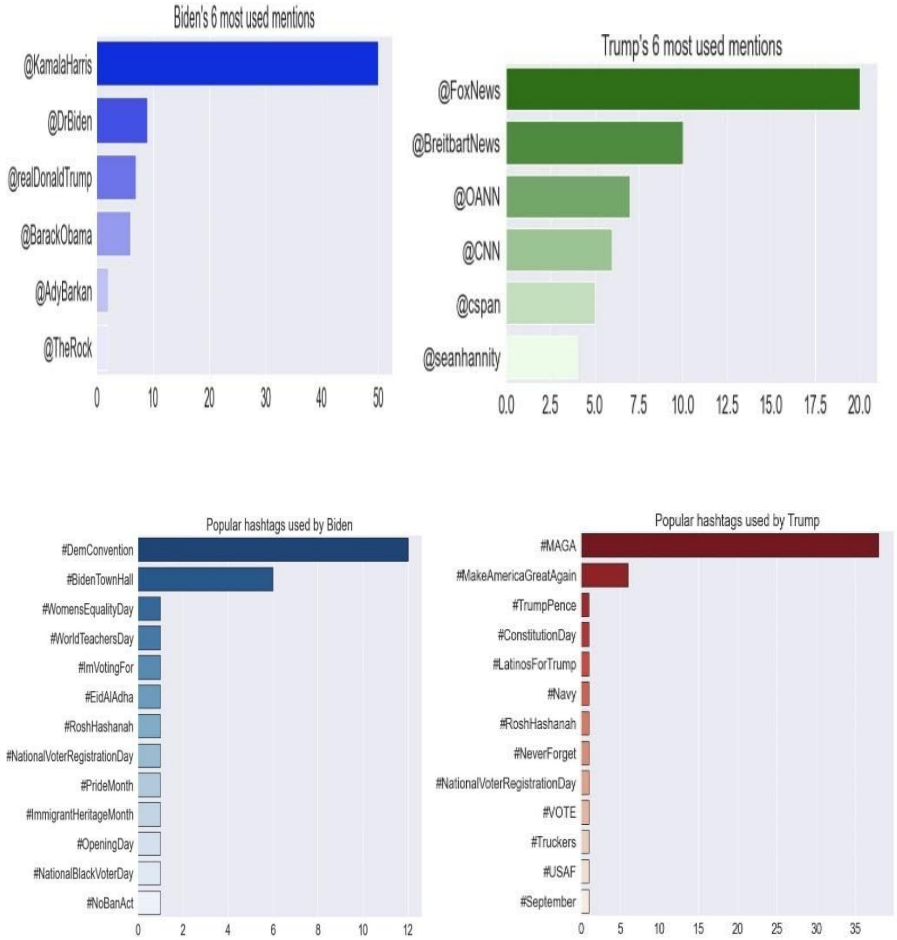






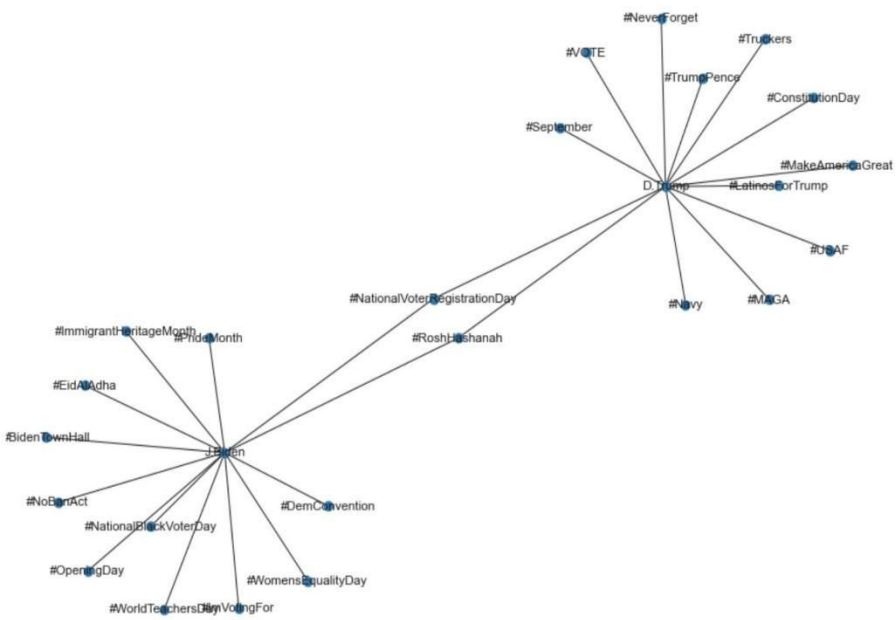
From the above graphs we can see that Trump was having better reach of retweets, but Biden managed to have more likes at certain hours of the day. A general conclusion can be drawn that Trump was more active on social media and he had a fair amount of reach than Mr. Biden.

Bidens word cloud	This visualization is word cloud for what Biden has tweeted and we can see that he has mentioned Trump's name quite a few times along with covid, building back the nation, climate change, future, economy.
Trump's word cloud 	This visualization is a word cloud for what Trump has tweeted and we can see that Trump has used a lot of words like great, incredible, fake news, Jobs, China and has also targeted Biden using some phrases like 'sleepy' and 'radical left'.

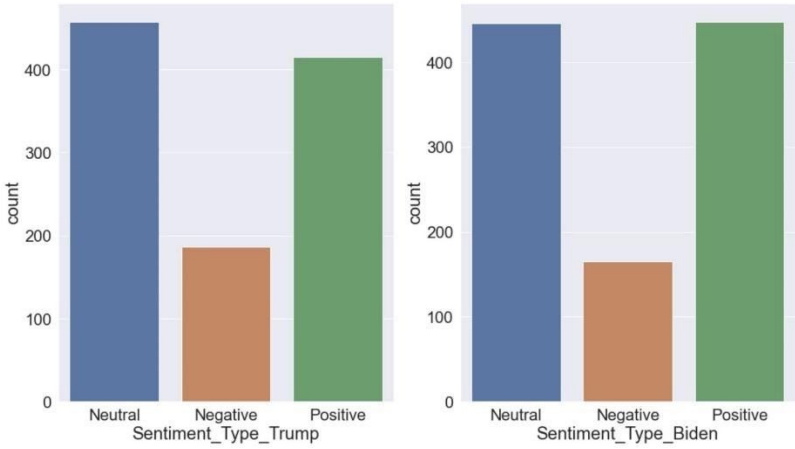


From the visualizations we can see that Biden mentioned Kamala Haris, Dr.Biden, Barack Obama, and Donald Trump many times. Trumps majorly mentioned news channels.

Some of the popular tags used by Biden are, Democratic convention, women's equality day, Immigration heritage month, National black day. Trump, tweeted about making America great again, constitution day, vote , national vote registration day.
which means that Biden and Trump are addressing these situations or problems and utilizing social media to share his thoughts on trending topics and which will impact the presidential election.



The graph depicts how Biden and Trump tweeted on trending topics and how they are related to both the candidates.



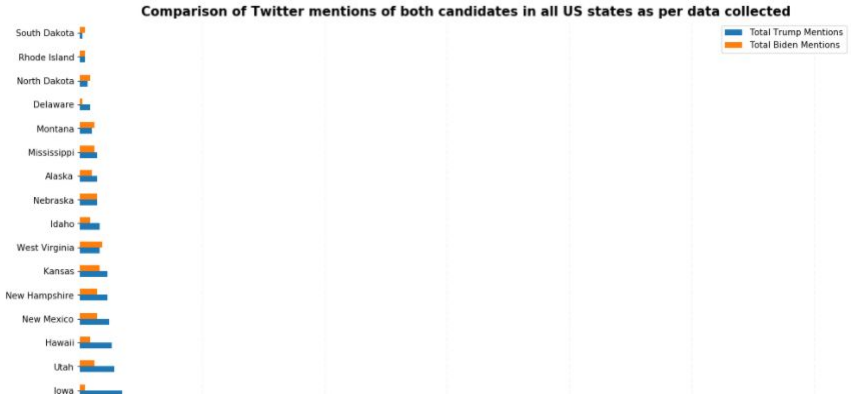

```
my_1st_tweet = 'MAKE AMERICA GREAT AGAIN. VOTE!!!'
my_2nd_tweet = 'Now more than ever, we need a president who will choose compassion over cruelty.'
print("Tweet #1:", "'",my_1st_tweet, "'", ' \n \n', "I'm about %.0f%%" % (100 * max(nb_detector.predict_proba([my_1st_tw
print("Tweet #2:", "'",my_2nd_tweet, "'", ' \n \n', "I'm about %.0f%%" % (100 * max(nb_detector.predict_proba([my_2nd_tw
```

```
Tweet #1: ' MAKE AMERICA GREAT AGAIN. VOTE!!! '

 I'm about 69% sure this was tweeted by realDonaldTrump
Tweet #2: ' Now more than ever, we need a president who will choose compassion over cruelty. '

 I'm about 86% sure this was tweeted by JoeBiden
```

From the visualization we can see that the neutral sentiment for Both Biden and Trump is the same.
We can also conclude that negative sentiment for Trump is slightly more for Trump as compared to Biden and the positive sentiment for Biden is slightly higher than that of Trump.

We are using the multinomial NB for text classification. Using this model we can predict the percentage of accuracy of tweet i.e., if it is Biden or Trump tweets.

KEYWORDS:

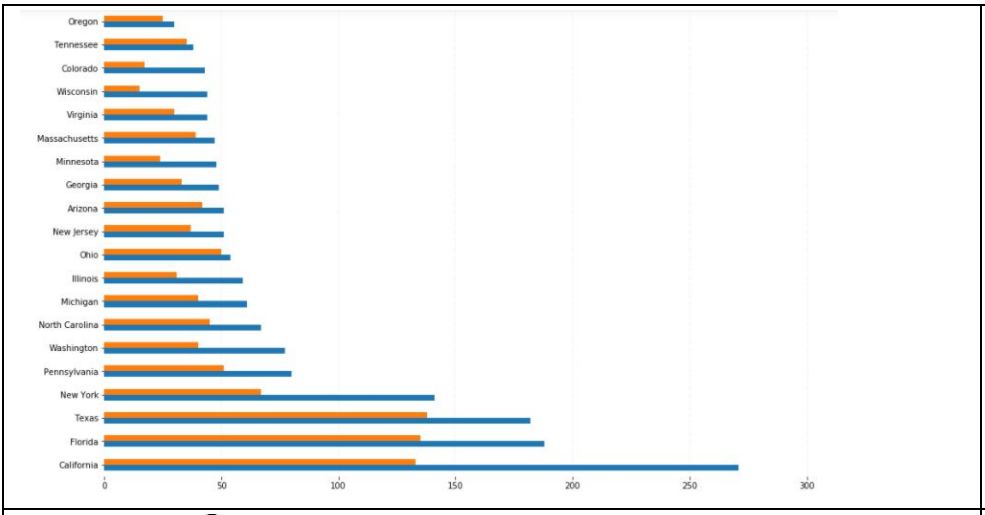
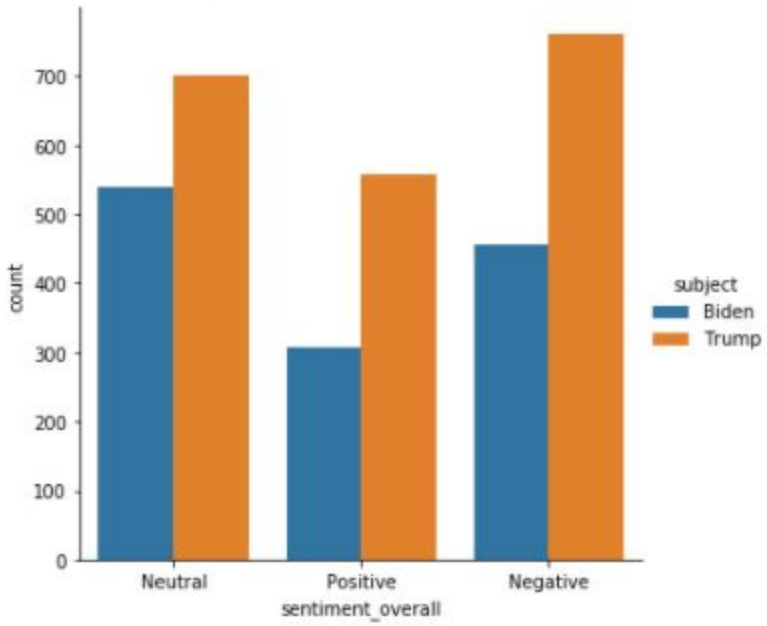


We have collected tweets related to election related keywords. The Keywords that we have used are

'#vote','#Election2020','PresidentialDebate','#Debates2020','#Election2020','#vpdebate','Trump','Biden', 'kamala', 'pence'.

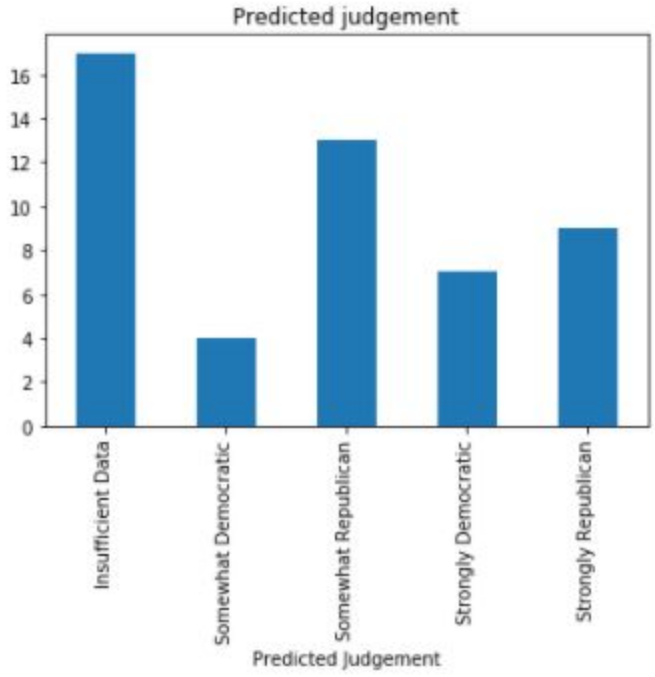
This data collected is only for the month of October.

In the visualization we have plotted graphs for how people from various regions/states are reacting for the presidential election. From this, we can see in which state the Democrats or republicans are leading.

	Note: This is not the final result, since we collected data only for the month of october. If we collect data for all months we can come to a conclusion, but we cannot Totally depend on social media tweets since these tweets are mostly related to emotions.
	From the keywords dataset we collected, we cleaned the data and applied a text blob to the tweets to find how people are reacting for the two candidates(Trump and Biden). From the graph we can see that the negative, positive and neutral count for Trump is more as compared to Biden.

State	Trump Positive	Trump Negative	Trump Neutral	Biden Positive	Biden Negative	Biden Neutral
Alabama	30	22	48	42	26	32
Alaska	29	0	71	40	20	40
Arizona	25	37	37	19	26	55
Arkansas	44	28	28	14	29	57
California	28	36	36	21	36	43
Colorado	19	35	47	24	41	35
Connecticut	18	50	32	38	12	50
Delaware	0	75	25	0	0	100
Florida	28	34	38	27	35	39
Georgia	31	31	39	36	30	33
Hawaii	23	38	38	0	25	75
Idaho	38	25	38	50	50	0
Illinois	24	47	29	19	42	39
Indiana	40	40	20	37	32	32
Iowa	41	41	18	50	50	0

From the keywords dataset, we can segregate the total number of positive, negative and neutral tweets for both Trump and Biden for different states.



The graph shows the predicted judgement on various keywords related to the dataset. We can say that most of the data is insufficient and we can see that somewhat republician and strongly republicans are higher when compared to somewhat democratic and strongly democratic.

Note: We cannot conclude from this graph, as this data is related to only October month.

Conclusion:-

Coming to their twitter handles, Trump was more active on social media and he had a fair amount comparatively. But Joe Biden's tweets were consistent over time as he maintained a constant rate of tweets for a longer period of time. Major keywords used by Biden were 'building back the nation, climate change, future, economy' while Trump used words like 'great, incredible, fake news, jobs, China, covid'. Biden promoted the Vice-Presidential Candidate Kamala Harris a lot through his twitter handle. Trump used Fox News and OANN. The overall neutral tweets were equal from both the twitter handles with Trump having a slight edge in negative tweets and Biden having the edge in positive tweets. We then used Multinomial Naive Bayes to perform text classification and used this model to classify some recent tweets of both the candidates and our model predicted it correctly. From the keywords dataset we collected from twitter using election specific keywords we found the trend for various states for both candidates and found that Trump was more popular on social media which again proved our last hypothesis of both candidates having equal reach wrong. We used the LSTM model to analyze the sentiment of tweets, the accuracy of the model is 88.46%. Since the election result is based on each state's electoral votes we tried to analyze the sentiment for each candidate in all states and based on the positive and negative counts of tweets we called the states for the respective parties. Based on our analysis there were 9 strongly republican states and 7 strongly democratic states along with 4 somewhat democratic and 7 somewhat republican states. We were not able to call the results in 17 states due to insufficient data.

Recommendations for Future Research:

Perform twitter sentiment analysis and topic modeling for all months to know the trend and various topics.

Using neural networks for topic modelling of news articles.

Data collection can be made better by collecting data from all sources like social media, youtube, etc.

Bibliography

Agarwal, B., & Mittal, N. (Eds.). (2016). *Prominent Feature Extraction for Sentiment Analysis* (Vol. 2). Springer. 10.1007/978-3-319-25343-5

Ansari, M. Z., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). *Analysis of Political Sentiment Orientations on Twitter* (Vol. 167). Procedia Computer Science. https://doi.org/10.1016/j.procs.2020.03.201.

Dang, Y., Zhang, Y., & Chen, H. (2010). A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews. *Univeristy of Arizona*. https://d1wqtxts1xzle7.cloudfront.net/35991839/15_A_Lexicon_Enhanced_Method_for_Sentiment_Classification.pdf?1418901523=&response-content-disposition=inline%3B+filename%3DA_Lexicon_Enhanced_Method_for_Sentiment.pdf&Expires=1606591421&Signature=CCUdREeJcHMG

Frontiers in Artificial Intelligence. (2020, july 14). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. *Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis*. https://www.frontiersin.org/articles/10.3389/frai.2020.00042/full#h4

Introduction to Topic Modeling. (2019, SEPTEMBER 26). Introduction to Topic Modeling. https://monkeylearn.com/blog/introduction-to-topic-modeling/

Jain, P., & Dandannavar, P. (2017). Text Analytics Framework Using Apache Spark and Combination of Lexical and Machine Learning Techniques. *International Journal of Business Analytics and Intelligence*, *5 (1)*, 36-42.

https://d1wqtxts1xzle7.cloudfront.net/58180399/Text_Analytics_Framework_using_Apache_Spark_and_Combination_of_Lexical_and_Machine_Learning_Techniques.pdf?1547462977=&response-content-disposition=inline%3B+filename%3DText_Analytics_Framework_using_Apache_Sp

Liu, R., Yao, X., Guo, C., & Wei, X. (n.d.). Can We Forecast Presidential Election Using Twitter Data? An Integrative Modelling Approach. *Annals of GIS*.

Matsumoto S., Takamura H., & Okumura M. (2005). Sentiment classification using word subsequences and dependency sub-trees. In *Proceedings of the 9th Pacific-Asia conference on advances in knowledge discovery and data mining (PAKDD)* (pp. 301-311).

Multinomial Naive Bayes Explained. (2020, SEPTEMBER 20). Multinomial Naive Bayes Explained. https://www.mygreatlearning.com/blog/multinomial-naive-bayes-explained/#:~:text=The%20Multivariate%20Event%20model%20is%20referred%20to%20as%20Multinomial%20Naive%20Bayes.&text=Naive%20Bayes%20is%20based%20on,occurrence%20of%20the%20other%20feature.

Pang B., & Lee L. (2008). *Opinion mining and sentiment analysis. Foundations and trends in information retrieval* (Vol. 2, 1-2). Now Publishers.

Practical Data Science in Python. (2014, DECEMBER 10). Practical Data Science in Python. https://radimrehurek.com/data_science_python/

Řehůřek., R. (2009, November 04). *Gensim (topic modelling for humans)*. Gensim. https://radimrehurek.com/gensim/auto_examples/core/run_core_concepts.html#:~:text=C

orpora%20serve%20two%20roles%20in%20Gensim%3A%20Input%20for,annotations%20or%20tagging%20documents%20by%20hand%2C%20is%20required

Topic Modelling with Scikit-learn. (2017). DEREK GREENE. http://derekgreene.com/slides/topic-modelling-with-scikitlearn.pdf

Towards Data Science. (2018, FEBRAUARY 19). *Long Short Term Memory*. Long Short Term Memory. https://towardsdatascience.com/understanding-lstm-and-its-quick-implementation-in-keras-for-sentiment-analysis-af410fd85b47

Towards Data Science. (2020, june 27). *Sentiment Analysis using TextBlob*. Sentiment Analysis using TextBlob. https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524

Wang H., Can D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2020). A system for real-time Twitter sentiment analysis of 2012 US Presidential election cycle. *ACL (System Demonstrations)*, 115-120. https://www.aclweb.org/anthology/P12-3020.pdf

Wick, M. M., & Chambers, N. (2012). Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter. 603-612. https://apps.dtic.mil/dtic/tr/fulltext/u2/a589957.pdf