Research Track
Thesis Report On

***"Prediction of USA Elections 2020 using Sentiment Analysis"***


***Submitted*** **By**

**Bhanu Kumar (2K19/SWE/04)**


**MASTER OF TECHNOLOGY**
**IN**
**SOFTWARE ENGINEERING**

Under the guidance
Of
**Mr. Rahul**
Assistant Professor, DTU


Department of Software Engineering
Delhi Technological University, Delhi


**July-2021**


**DELHI TECHNOLOGICAL UNIVERSITY**
**SHAHBAD DAULATPUR, DELHI-110042**

# DECLARATION

I hereby declare that the work presented in this report entitled "Prediction of USA Elections 2020 using Sentiment Analysis", in fulfillment of the requirement for the award of the MASTER OF TECHNOLOGY degree in Software Engineering submitted in Software Engineering Department at DELHI TECHNOLOGICAL UNIVERSITY, New Delhi, is an authentic record of my own work carried out during my degree under the guidance of Mr. Rahul.

The work reported in this has not been submitted by me for the award of any other degree or diploma.

Date: 31st July, 2021
Place: Delhi
**Bhanu Kumar (2K19/SWE/04)**

# CERTIFICATE

This is to certify that Bhanu Kumar (2K19/SWE/04) has completed the research titled "Prediction of USA Elections 2020 using Sentiment Analysis" under my supervision in fulfillment of the MASTER OF TECHNOLOGY degree in Software Engineering at DELHI TECHNOLOGICAL UNIVERSITY.

Rahul Digitally signed by Rahul
Date: 2021.08.09 21:03:50 +05'30'

**Mr. Rahul**
**(Supervisor)**
Assistant Professor, DTU

# ABSTRACT


Machine Learning (ML) is expanding its applications in our life as the amount of data stored on servers is increasing daily. With ample of provided applications to ease our workload and make us more efficient.

Exit Polls for elections although quite accurate cannot be completely relied upon. This can arise due to pressure from political leaders, peers; people who don't want to share their views etc.

There have been instances in past where the results of elections were completely contradictory to predictions based on exit polls. With social media, people have become more vocal about their views and perspectives with the privacy and security over internet.

With more people using social media to express their views, we can create several detailed and structured datasets according to our needs. This decreases time as compared to interviewing one person at a time, we can get data of millions promptly. This data can be classified on the basis of region, age, gender, etc.

Using ML Algorithms on these datasets we can predict the sentiment of these people and can get an accurate prediction for the elections.

We'll be performing Sentiment Analysis on one such dataset which consists of tweets extracted from Twitter.

This report will include using seven algorithms: Dictionary Based, Naïve Bayes, Support Vector Machine, Linear Regression, Logistic Regression and Bayesian Network and compare the results of these Algorithms along with their accuracies.

# ACKNOWLEDGEMENT

I am very thankful to **Mr. Rahul** (Assistant Professor, Department of Software Engineering) and all the faculty members of the Department of Software Engineering at DTU. They all provided me with immense support and guidance for the project.

I would also like to express my gratitude to the University for providing me with the laboratories, infrastructure, testing facilities, and environment which allowed us to work without any obstructions.

I would also like to appreciate the support provided to us by our lab assistants and our peer group who aided us with all the knowledge they had regarding various topics.

Bhanu Kumar
Roll No. 2K19/SWE/04

# INDEX

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **ML** | Machine Learning |
| **NB** | Naïve Bayes |
| **SVM** | Support Vector Machine |
| **NLP** | Natural Language Processing |
| **PSI** | Post Sentiment Information |
| **TMP** | Text Mining Process |

# CHAPTER 1

# INTRODUCTION

SA is an application of ML which involves predicting the emotions of the text which is provided to the algorithm. The output depends on the positive, negative or 0 zero value output of the algorithm where positive integer determines the positive behavior of text, negative value determines negative behavior of text and 0 for neutral text.

Different algorithms will generate different outputs. The difference in the reading is not uniform and the results will vary with each dataset provided to algorithms. Some algorithms will get more accurate with respect to time but for earlier stages they might be the worst choice.

We'll be using SA to predict the winner of USA Elections 2020 and verify the same with official result.

A dataset is created by extracting tweets from twitter using WebCrawler and Twitter API. The tweets were extracted in the months of October, 2020 and November 2020 with "Joe Biden" and "Donald Trump" as primary attributes for tweet extraction. A total of 712,238 tweets were extracted in the above two months which had "Joe Biden" or "Donald Trump" as characters in them. Languages in which tweets were extracted are English and Spanish (Most spoken languages in USA.)

Twitter has become a place for people to express their views and beliefs. As per [1] there are more than 340 million users across globe with around 500 million tweets uploaded every day. These tweets when extracted and evaluated can describe the demographics of its users.

In our case, consider a user wrote "Biden is the best suited for the post.", with word like "best" in the tweet, this will be considered as a positive opinion in favor of Biden. In other scenario where a tweet is "Joe Biden can't compete with Donald Trump" is negatively in favor of Joe Biden and positively in favor of Donald Trump. But to remove ambiguous data as it is a single tweet, it will only be considered negative opinion for Joe Biden.

To perform sentiment analysis on the tweets,

We need to extricate them from Twitter using Twitter API, and then noise removal is to be performed to remove unwanted data. To extract information from the tweets several extraction techniques can be applied like Aggregation and Weighing Scheme, Weighted Score Computing, Classification Methods like Neutral, Polar or Irrelevant. These extraction techniques can categorize tweets into positive, negative or neutral.

Here, we'll perform Dictionary Based Classifiers, Support Vector Machine (SVM) Classifiers, Naïve Bayes Classifiers, Linear Regression Classifiers, Logistic Regression Classifiers, Maximum Entropy Classifiers and Bayesian Network Classifiers to find the which of these can most accurately predict the outcome of the 2020 elections.

# CHAPTER 2

# RELATED WORK

Plenty of research is done regarding SA and on various algorithms of ML. As ML is a pristine topic, its applications on prediction of election results are very limited. So, research available for this trend is vague.

[2] Classifies NLP into Text Mining and Opinion Mining. It segregates emotions and sentiments from normal text blocks. Sentiment Analysis is used by various organizations for market surveys which includes analyzing sentiments of customers, their likes and dislikes and predicting the impact of their product in those demographics.

[3] segregates the meaning of sentiment into 3 layers which include: Opinion Layer, Emotion Layer and Opinions on the basis of Emotion Layer.

[5] suggests that tweets ending with emoticons such as “:)”, “:(” or “:|” can also be used to determine the sentiment of a person. Here we’ll also be considering emoticons in our analysis.

To have a better dataset for our analysis on, we extracted tweets for English and Spanish languages as they are the most spoken languages in the USA. This have also improved our demographics of the tweets we mined.

Spanish tweets were first translated to English and analyzed alongside native English tweets.

[6] used SWAB (Social Web Analysis Buddy) together with R language to perform a basic Sentiment Analysis of tweets by extracting tweets from Twitter and data filtration followed by Naïve Bayes classifiers, performing it in four phases and taking input in hashtags and number of tweets restricted between 5 and 1000.

[7] performs sentiment analysis of tweets using python language after authenticating user and display an error message if authentication fails.

[8] describes uses of sentiment analysis and opinion mining for social networking sites along with challenges that can arise in the process.

[9] trace the use of sentiment analysis in healthcare based on text-based information.

[10] proposes a method to predict movement in stock price by mining twitter data.

[11] used sentiment analysis to predict winner of 2013 Pakistan Elections using Naïve Bayes, SVM and Chi-square Automatic Interaction Detector (CHAID) algorithms.

# THEORY

## SENTIMENT ANALYSIS PROCESS AND TECHNIQUES

### i. Process to perform Sentiment analysis

Here we'll discuss the process undertaken to perform SA.

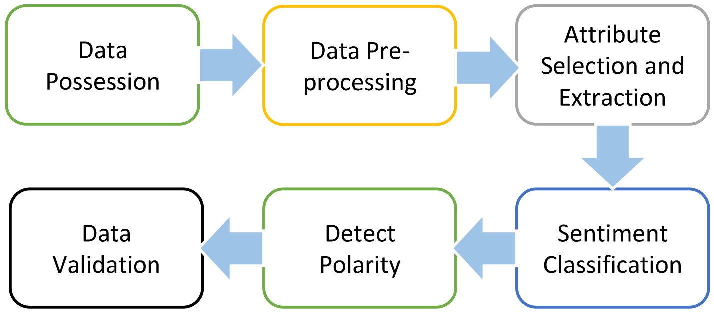


Figure 1: Process for performing Sentiment Analysis

### 1. Data Possession

Required data regarding USA Elections was extracted from Twitter using crawling on tweets based on hashtags such as #Trump, #Biden, #USelections, #US2020elections, etc. Also, region was made limited to USA to improve accuracy. The tweets were stored in tabular form with date, time, tweet and a unique identifier for each tweet to remove duplicate tweets from same person.

### 2. Data Pre-processing

Unwanted data such as noise, ambiguous data, URLs are removed from the tweets to make operations on them more efficient and reduce size of dataset.

Spanish tweets were translated to English.

This was done mainly because Spanish dictionary was limited on TextBlob and it wouldn't have been as efficient as compared to English Dictionary available.

### 3. Attribute Selection and Extraction

From each tweet features are extracted using extraction techniques for precise extraction using techniques Aggregation and Weighing Scheme, Weighted Score Computing etc.

### 4. Sentiment Classification

After previous steps, sentiment analysis techniques are applied. We used Naïve Bayes, Support Vector Machine (SVM) and Dictionary Based algorithms.

5. **Detect Polarity**

Polarity of statements is detected whether it is positive, neutral or negative based Sentiment Classification.

6. **Data Validation**

The result of previous step is validated, based on which the result is evaluated to demonstrate the fidelity of the result.

### ii. SENTIMENT ANALYSIS TECHNIQUES

Figure 2 displays various techniques/algorithms that can be used for SA.

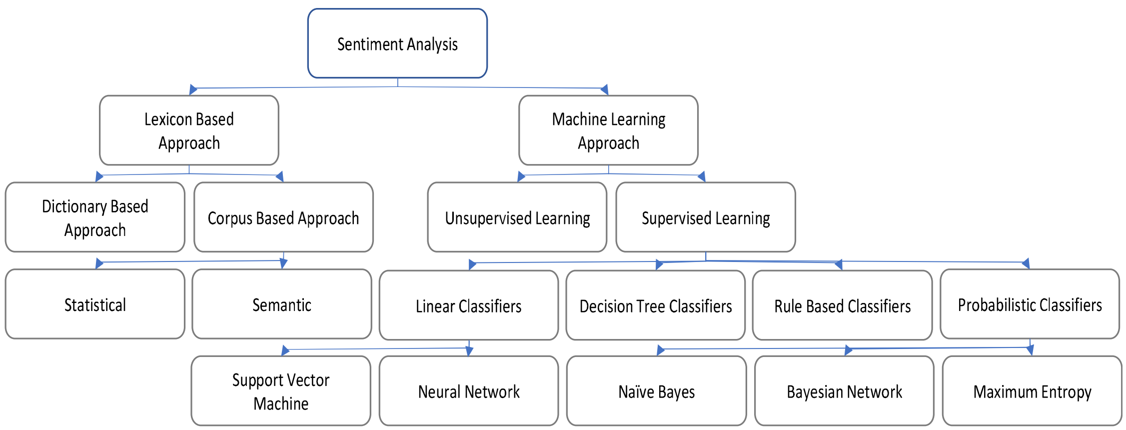


Figure 2: Sentiment Analysis Techniques

We'll be using and comparing Supervised Learning Approach.

Sentiment Analysis can be performed using several techniques on the basis of their approach are classified into two namely Lexical Based Approach and Machine Learning Based Approach.

Figure 2 shows classification of techniques used for Sentiment Analysis based on the approaches used.

Lexical Based approach on the basis of usage is bifurcated into Dictionary Based and Corpus Based approach. Lexical approach scores data based on existing vocabulary or lexicons. This is considered slightly inaccurate as compared to Machine Learning approach [12]. Corpus based is further classified into Statistical and Semantic based approach.

Machine Learning on the other hand is classified into Supervised Learning and Unsupervised Learning approaches.

Unsupervised Learning approach involves feeding preprocessed data into the model and the model is supposed to learn the structure based on input. This approach is useful when the data is not labeled or the structure of data is unknown.

Supervised Learning involves using two data sets as explained by [13] among which one is train dataset and is trained on an example input with humans labelling the output data. The model then learns from the input provided to improve its accuracy to expected output.

In Supervised Learning, there are four classifiers: Decision Tree Classifiers, Linear Classifiers, Probabilistic Classifiers and Rule Based Classifiers.

Linear Classifiers involve Support Vector Machine (SVM) and Neural Networks while Probabilistic Classifiers involve Naïve Bayes, Maximum Entropy and Bayesian Network Classifiers.

# <u>CHAPTER 3</u>

# PROPOSED WORK

We have extracted and downloaded the dataset from Twitter for 712,238 tweets using web-crawling and TextBlob. The resulting dataset was filled with ambiguous data and was partially inefficient due to various redundancies.

To overcome these obstacles, firstly ambiguous data was removed followed by normalizing dataset according to needs.

For different algorithms different techniques were used for best efficiency.

**a. Dictionary Based Sentiment Analysis**

Data in form of several words is added to a dictionary or model manually where positive, negative and neutral sentiments of the words are known. These words are known as seeds. The algorithm then increases and improves the size of dataset by searching in others dictionaries available for the synonyms and antonyms of the word. When no new word can be found, this iterative process halts.

Based on [14], we perform Sentiment Analysis using Text Mining Process as shown in figure 3.

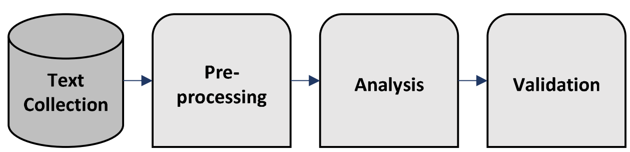


Figure 3: Text Mining Process

1. **Text Collection**

The process starts with Text Collection where the first step is to generate data. These databases can be Static (remains alike for the whole process) or Dynamic (can be updated at any instance.)

2. **Pre-processing**

This is a structured approach which involves

i) Tokenization: for identifying words in input text.

ii) Data Filtering: removes or update redundant data into more logical words. Example: “Best” and “Bestttttttt” will be considered same word and will be considered a single occurrence when arriving in a same sentence. Also “Bestttttttt” will be updated to “Best” when arriving exclusively in a sentence.

iii) Stop Word Removal: This removes frequently occurring words such as adverbs, articles, prepositions and conjunctions.

iv) Stemming: reduces word to its native/root form. Example: slowly, slower to slow.

### 3. Analysis

This is considered the fundamental step of Text Mining. Some significant knowledge can be extracted at this process.

### 4. Validation

Qualitative and quantitative measures are taken to validate the analysis of previous steps. Here, we can return to previous steps to improve our results by preforming modifications and alternate methods.

Based on equations

$$\text{Precision} = \frac{tp}{\mathrm{tp} + fp}$$

$$\text{Recall} = \frac{tp}{\mathrm{tp} + fn}$$

$$\text{Accuracy} = \frac{tp+tn}{\mathrm{tp} + fn+tn+fp}$$

$$\text{F-measures} = \frac{2*Precision*Recall}{Precision+Rec}$$

Table 1 shows segregation of tweets specific to both candidates based on Dictionary Based Approach.

| Name | Number of Tweets |
| --- | --- |
| Donald Trump | 278,194 |
| Joe Biden | 308,566 |
| Both Names | 125,478 |

Table 1

**b. Naïve Bayes based Sentiment Analysis**

Naïve Bayes uses probabilistic approach to train classifier. The basic principle of this approach involves a training classifier to assume the features of a specific class variable as independent. As per [13], naïve bayes is considered as one of the easiest and most commonly used algorithms. This model calculates probability in regards to total occurrences of words in the documents.

We used,

$$p(\text{Label/Feature}) = \frac{p(Label) * p(Feature/Label)}{p(Feature)}$$

to create Figure 5.

**c. Support Vector Machine based Sentiment Analysis**

SVM (Support Vector Machine) is applied to two-group classification complications because of their learning behavior as per [15]. SVM classifies given text into positive, neutral or negative based on sentiment involved. SVM has potential to handle large attributes. As per [16], SVM can handle various set of examples because of its robust nature.

SVM uses probability with its linear properties.

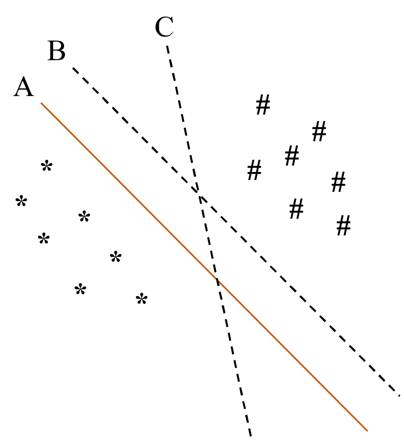


Figure 4: SVM on a Classification Problem

Consider a classification problem where Figure 6 shows two classes "*" and "#" with three hyperplanes A, B and C. Hyperplane A yields the most efficient separation between classes B and C because of its nominal distance to all the data points is maximum of both classes. Hyperplane A hereby constitutes the maximum separation margin.

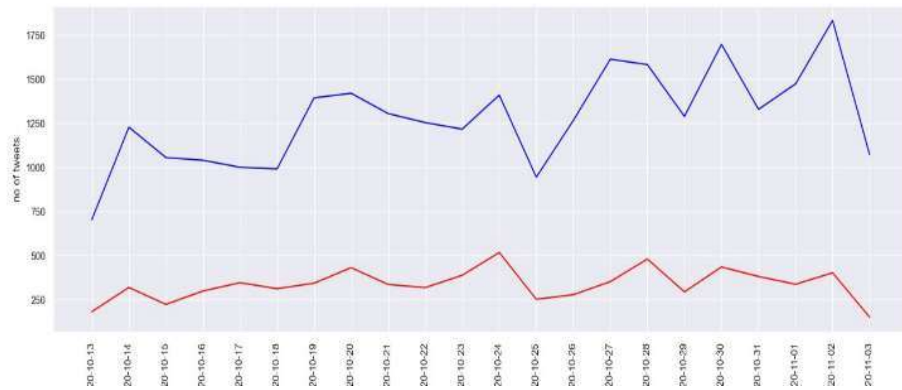

Figure 5: Positive and Negative tweets for Donald Trump for the month of October,2020.

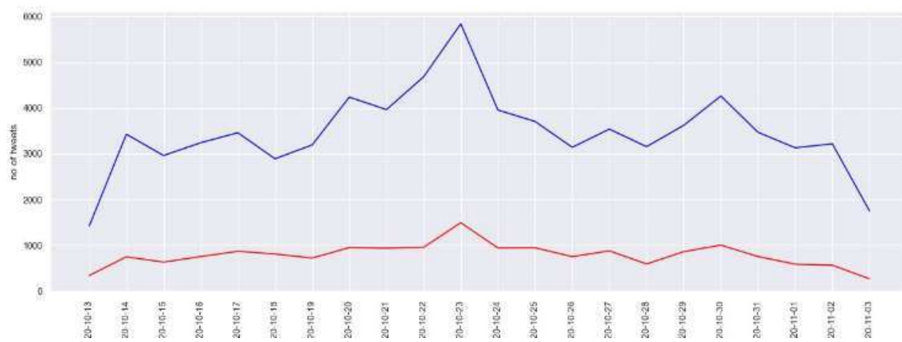

Figure 6: Positive and Negative tweets for Joe Biden for the month of October,2020.

In Figures 7 and 8, blue line represents positive tweets whereas red line represents negative tweets. These graphs were created with data collected each day for the given timeframe.

### d. Bayesian Network

Bayesian Network involves a set of variables represented by a probabilistic graph along with a acyclic graph directing their conditional dependencies.

[17] proposes a Gaussian Process Dynamic Bayesian Model to learn dependencies in relationships and map a time series for sentiments relating to that subject.

[18] proposed an approach based on Bayesian Network to create a model of observed sentiments.

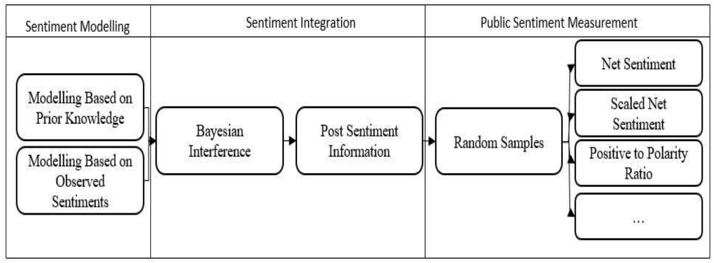


Figure 7: Bayesian Based Approach for Public Sentiment

## 1. Sentiment Modelling:

Here sentiments are modelled based on probabilities. Dirichlet Distribution along with Multinomial Distribution are used to design sentiment probabilities.

[18] uses distinct functions for underlining models

### 1.1 Modelling Based on Prior Knowledge

$$p(\boldsymbol{\theta}) = p(\theta_1, \theta_2, \theta_3) = Dir(\alpha_1, \alpha_2, \alpha_3) = \frac{1}{B(\alpha_1, \alpha_2, \alpha_3)} \prod_{i=1}^{3} \theta_i^{\alpha_i - 1}.$$

Here *B* is a multivariate Beta Function. α are shape parameters and θ are sentiment probabilities.

### 1.2 Modelling Based on Observed Sentiments

$$L(\boldsymbol{X}|\boldsymbol{\theta}) = L(x_1, x_2, x_3 | \theta_1, \theta_2, \theta_3) = \frac{\Gamma(\sum_i x_i + 1)}{\prod_i \Gamma(x_i + 1)} \prod_i \theta_i^{x_i}.$$

Here $x_1$ are negative sentiments, $x_2$ are neutral sentiments and $x_3$ are positive sentiments.

## 2. Sentiment Integration:

Sentiment Integration is used to integrate the system based on following parameters

### 2.1 Bayesian Interference

Bayesian Interference is used for statistical interference and Bayes' Theorem is used for updating the probability.

### 2.2 Post Sentiment Information

Shape parameters are introduced in interviewed dataset. This makes the process easier and efficient.

### 3. Public Sentiment Measurement

Public Sentiment Measurement involves summarizing sentiment information through aggregating sentiment probabilities. Histograms are generated based on Beta Distributions like (Negative + Neutral or Positive or Neutral).

Using Bayesian Network on our dataset, following graph shows the output after applying this approach.

### e. Maximum Entropy

Maximum Entropy approach involves probability distribution which is no less than other members of specified class. If the contents of distribution are unknown except the class to which it belongs, least-informative entropy is selected and made default based on largest entropy distribution.

[19] uses Maximum Entropy Model to perform Sentiment Analysis on Tibetan Sentences. They divide the stated model into 3 modules in form of a public model for training and test execution purposes.

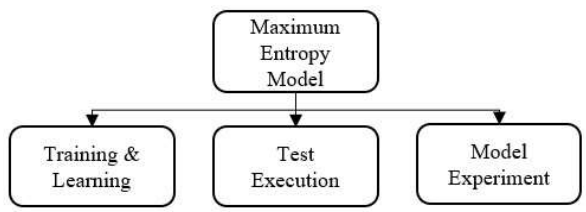


Figure 8: Maximum Entropy Model

[20] inscribe the exude of Word Identification in Sentiment Analysis on an opinionated sentence. They compare their supervised model to lexicon-based model and improvement in their approach compared to later.

[21] makes a comparison between Support Vector Machine and Maximum Entropy for performing Sentiment Analysis on Grab User Reviews.

### f. Linear Regression

Linear Regression performs regression tasks based on independent variables. It uses linear technique between input($x$) and output($y$).

Hypothesis function for Linear Regression:

$y = \theta_1 + \theta_2 . x$

Where $x$ is input training data, $y$ denotes labelled data, $\theta_1$ acts as intercept and $\theta_2$ is coefficient of $x$.

### g. Logistic Regression

Logistic Regression involves predicting the probability of a target variable. The dependent or target variable can have only two possible classes.

[22] uses Logistic Regression and Naïve Bayes for Sentiment Classification on Big Data of Twitter Reviews using Hadoop and Mahout and concludes with Logistic Regression being more accurate as compared to Naïve Bayes.

[23] uses Multinomial Logistic Regression on Sentiment Analysis using K-Fold Validation Technique to measure performance of this approach. They conclude by publishing that the number of folding doesn’t affect the performance.

# CHAPTER 4

# RESULTS

In this Chapter, we will see the results of the different algorithms used.

For ease of understanding graphs for each algorithm is designed which consists of percentage of positive, negative and neutral tweets for both candidates.

### a. Dictionary Based Technique

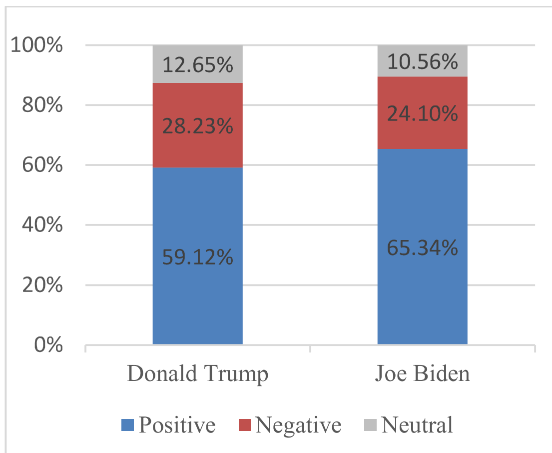


Figure 9: Dictionary Based Approach

Figure 4 shows that Joe Biden had 65.34% positive tweets as compared to 59.12% of Donald Trump stating possibility of Joe Biden winning the 2020 USA Elections.

### b. Naïve Bayes Technique

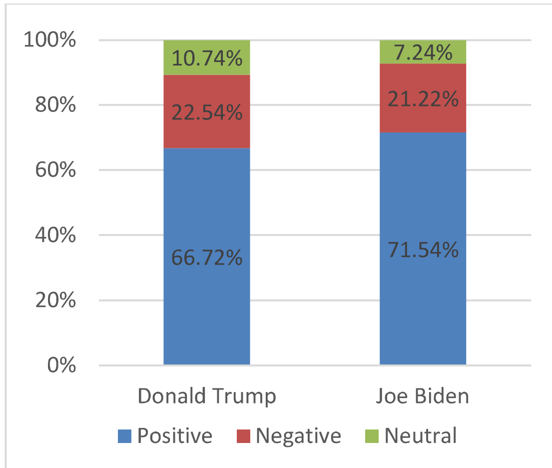


Figure 10: Naïve Bayes Approach

Figure 5 shows Joe Biden had 71.54% positive tweets as compared to Donald Trump with 66.72% positive tweets stating likelihood of Joe Biden winning the 2020 USA Elections.

### c. Support Vector Machine Technique

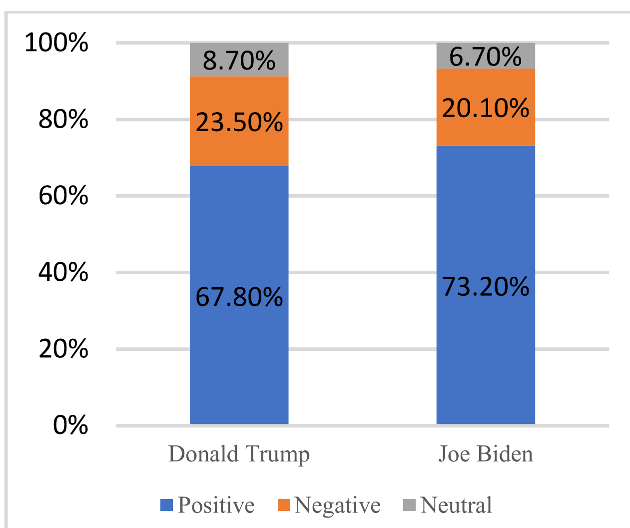


Figure 11: Support Vector Machine Approach

Figure 9 shows 73.20% positive tweets in favour of Joe Biden as compared to 67.80% positive tweets for Donald

Trump. As per calculated data, likelihood of Joe Biden winning elections is more as compared to Donald Trump.

### d. Bayesian Network

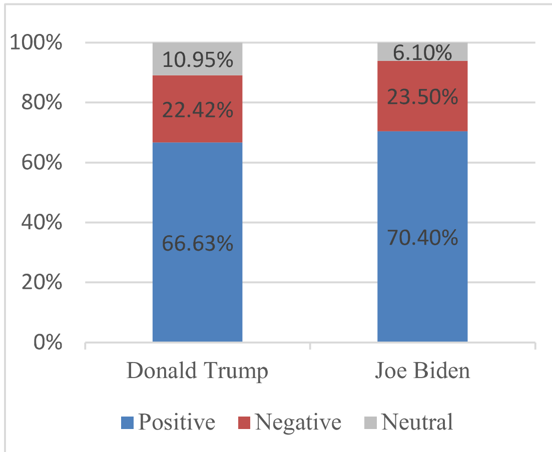


Figure 12: Bayesian Network Approach

Figure 11 shows Joe Biden in lead with 70.40% positive tweets as compared to 66.63% positive tweets in favour of Donald Trump.

### e. Maximum Entropy

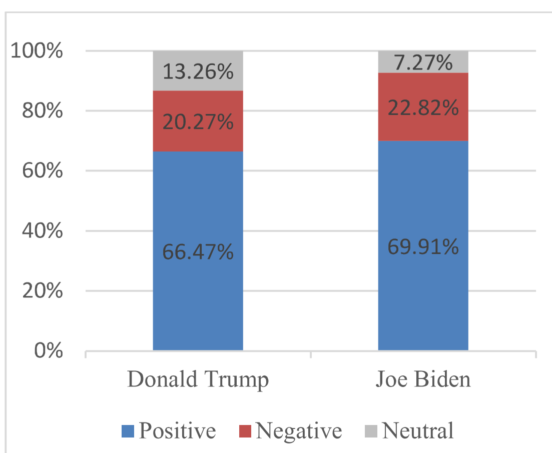


Figure 13: Maximum Entropy Approach

Here it is visible that Joe Biden had more positive tweets in his favour as compared to Donald Trump. But Joe Biden also had more negative tweets as compared to his counterpart. This graph represents that more people were having neutral views about Donald Trump.

### f. Linear Regression

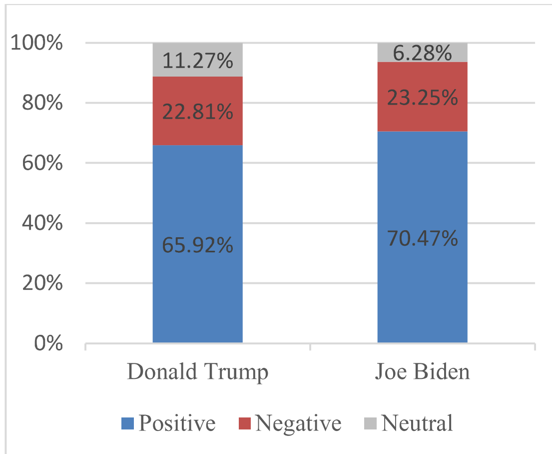


Figure 14: Linear Regression Approach

The graph shows a similar result like Maximum Entropy with Joe Biden having more Positive and Negative tweets in his favour.

### g. Logistic Regression

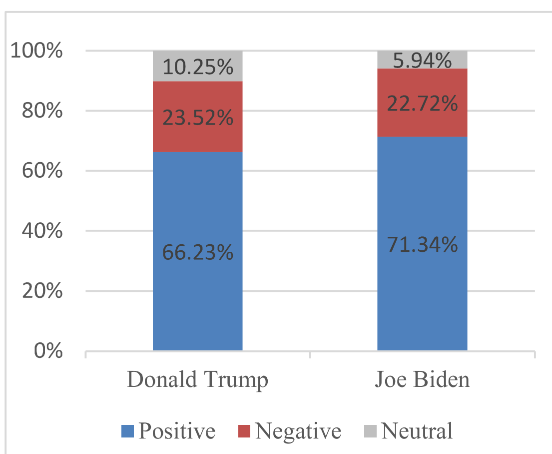


Figure 15: Logistic Regression Approach

This Graph concludes that Joe Biden should be in lead with the number of positive tweets. As a result, we can assume that according to this approach Joe Biden should be the potential winner.

To show collective result along with accuracy of these algorithms, Table 2 is created displaying all the required information.

| Name of Candidate | Dictionary Based | | | Naïve Bayes | | | Support Vector Machine | | |
|---|---|---|---|---|---|---|---|---|---|
| | Positive | Negative | Neutral | Positive | Negative | Neutral | Positive | Negative | Neutral |
| Donald Trump | 59.12% | 28.23% | 12.65% | 66.72% | 22.54% | 10.74% | 67.80% | 23.50% | 8.70% |
| Joe Biden | 65.34% | 24.10% | 10.56% | 71.54% | 21.22% | 7.24% | 73.20% | 20.10% | 6.70% |
| Accuracy | 71.2% | | | 86.6% | | | 88.7% | | |

| Name of Candidate | Bayesian Network | | | Maximum Entropy | | | Linear Regression | | |
|---|---|---|---|---|---|---|---|---|---|
| | Positive | Negative | Neutral | Positive | Negative | Neutral | Positive | Negative | Neutral |
| Donald Trump | 66.63% | 22.42% | 10.95% | 66.47% | 20.27% | 13.26% | 65.92% | 22.81% | 11.27% |
| Joe Biden | 70.40% | 23.50% | 6.10% | 69.91% | 22.82% | 7.27% | 70.47% | 23.25% | 6.28% |
| Accuracy | 85.2% | | | 84.6% | | | 84.3% | | |

| Name of Candidate | Logistic Regression | | |
|---|---|---|---|
| | Positive | Negative | Neutral |
| Donald Trump | 66.23% | 23.52% | 10.25% |
| Joe Biden | 71.34% | 22.72% | 5.94% |
| Accuracy | 85.9% | | |

TABLE 2

# CONCLUSION

Table 2 exhibits the output of each algorithm we used for both candidates along with accuracy.

Accuracy of Support Vector Machine (SVM) was highest at 88.7% followed by Naïve Bayes at 86.6% and least efficient among the seven was Dictionary Based Approach at 71.2%. All of the approaches predicted Joe Biden (Democratic Party) winning the elections with majority of electoral votes.

The result of elections was declared indeed in favor of Joe Biden (Democratic Party) winning the 2020 USA Elections with majority of electoral votes. The competition was close in reality and exit polls of few News channels were contradicting the actual result. Our technique predicted the overall victory of “Joe Biden” but with more detailed dataset for each constituency, a wider and more detail picture can be painted where we can determine who is winning at specific location.

There is still scope to improve these approaches, but this research was bounded by current knowledge available regarding this application.

Dictionary for secondary languages can be improved in TextBlob so that they can be used directly instead of translating them to English as some words may lose their meaning in translation.

# REFERENCES


1 https://www.omnicoreagency.com/ttwitte-statistics/

2 P. Sharma and T. Moh, "Prediction of Indian election using sentiment analysis on Hindi Twitter," 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, 2016, pp. 1966-1971, doi: 10.1109/BigData.2016.7840818.

3 Q. Li, S. Shah, R. Fang, A. Nourbakhsh and X. Liu, "Tweet Sentiment Analysis by Incorporating Sentiment-Specific Word Embedding and Weighted Text Features," 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Omaha, NE, 2016, pp. 568-571, doi: 10.1109 /WI.2016.0097.

4 L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learninig Methods," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154183.

5 A. Pak, and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), pp. 1320–1326, May 2010.

6 V. Prakruthi, D. Sindhu and D. S. Anupama Kumar, "Real Time Sentiment Analysis of Twitter Posts," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 29-34, doi: 10.1109/CSITSS.2018.8768774.

7 S. Dhawan, K. Singh and P. Chauhan, "Sentiment Analysis of Twitter Data in Online Social Network," 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 2019, pp. 255-259, doi: 10.1109/ISPCC48220.2019.8988450.

8 Sudipta Roy, Sourish Dhar, Arnab Paul, Saprativa Bhattacharjee, Anirban Das, Deepjyoti Choudhury," Current Trends of Opinion Mining and Sentiment Analysis in Social Networks", International Journal of Research in Engineering and Technology, Volume 2, Special Issue 2, December 2013.

9 M. Taimoor Khan, Shehzad Khalid, "Sentiment Analysis for Health Care", International Journal of Privacy and Health Information Management, 2015.Li Bing, Chan, K.C.C., Ou, C.," Public Sentiment Analysis in Twitter Data for Prediction of A Companys Stock Price Movements",11th International Conference on e-Business Engineering (ICEBE), 2014 IEEE.

10 Mahmood, T.,Iqbal, T. ; Amin, F. ; Lohanna, W.; Mustafa, A.,Mining Twitter big data to predict 2013 Pakistan election winner,16th International Multi Topic Conference (INMIC),2013 IEEE.

11 Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA (2017) Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLOS ONE 12(2): e0171649.

12 https://journals.plos.org/plosone/article?id =10.1371/journal.pone.0171649

13 https://www.irjet.net/archives/V5/i4/IRJET-V5I4673.pdf

14 Reshma Bhonde, Binita Bhagwat, Sayali Ingulkar and Apeksha Pande, "Sentiment Analysis Based on Dictionary Approach", International Journal of Emerging Engineering Research and Technology, Volume 3, Issue 1, January 2015, PP 51-55 ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

15 C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning. 20, no. 3, pp. 273–297, 1995.

16 Zainuddin, Nurulhuda & Selamat, Ali. (2014). Sentiment analysis using Support Vector Machine. I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology, Proceedings. 333-337. 10.1109/I4CT.2014.6914200.

17 H. Liang, U. Ganeshbabu and T. Thorne, "A Dynamic Bayesian Network Approach for Analysing Topic-Sentiment Evolution," in IEEE Access, vol. 8, pp. 54164-54174, 2020, doi: 10.1109/ACCESS.2020.2979012.

18 Y. Chen, W. Ji and Q. Wang, "A Bayesian-Based Approach for Public Sentiment Modeling," 2019 Winter Simulation Conference (WSC), 2019, pp. 3053-3063, doi: 10.1109/WSC40007.2019.9004846.

19 X. Yan and T. Huang, "Tibetan Sentence Sentiment Analysis Based on the Maximum Entropy Model," 2015 10th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), 2015, pp. 594-597, doi: 10.1109/BWCCA.2015.32.

20 X. Fei, H. Wang and J. Zhu, "Sentiment word identification using the maximum entropy model," Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering(NLPKE-2010), 2010, pp. 1-4, doi: 10.1109/NLPKE.2010.5587811.

21 B. A. Dewanti Putri, A. Uswatun Khasanah and A. Azzam, "Sentiment Analysis on Grab User Reviews Using Support Vector Machine and Maximum Entropy Methods," 2019 International Conference on Information and Communications Technology (ICOIACT), 2019, pp. 468-473, doi: 10.1109/ICOIACT46704.2019.8938527.

22 A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," 2017 International Conference on Computer Communication and Informatics (ICCCI), 2017, pp. 1-5, doi: 10.1109/ICCCI.2017.8117734.

23 W. P. Ramadhan, S. T. M. T. Astri Novianty and S. T. M. T. Casi Setianingsih, "Sentiment analysis using multinomial logistic regression," 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC), 2017, pp. 46-49, doi: 10.1109/ICCEREC.2017.8226700.

# **END**