

Contextual Text Mining on Social Media of Political Leaders Using Machine Learning Algorithms

Aleemullakhan Pathan¹, R. Sundar²

Department of Computer Science & Engineering, Madanapalle Institute of Technology & Science,
Madanapalle, AP, India

Email: ¹pathan.aleemullakhan75@gmail.com, ²drsundarr@mits.ac.in

Abstract

Now a day's most of the political leaders use social media to easily communicate with the people, such as sharing their ideas, promoting their policies etc. Contextual text mining is used to acknowledge the opinions of the political leaders as well as attitude on different subjects such as opinions, discussions and microblogs. Natural language processing (NLP) is included for performing the contextual text mining in order to provide the communication between the human and the machine with natural language. A new model has been implemented to compare the dataset of the political leaders which is extracted from facebook and twitter with different machine learning algorithms like Support Vector Machine (SVM), Naive Bayes Classifier (NBC) and Ensemble Learning Methods (ELM) to provide better and accurate results than other machine learning algorithms.

Keywords: Contextual Text Mining, Social Media, Natural language Processing (NLP), Machine Learning Algorithms.

1. Introduction

The process of analyzing the user reviews, opinions, emotions, sentiments along with various entities is known as contextual text mining. Anyone can share their opinions and post their views through the web. Contextual text mining is also an approach [1] of identifying and aggregating opinions of different issues. The analysis of text can be done by using the mechanism known as polarity. Polarity

makes the text analysis very easy based on the views posted online. Polarity is especially useful to indicate whether the posted online views are positive or negative. Contents with respect to the text unit are simply extracted, identified and characterized by using NLP or statistics or machine learning methods. Facts and opinions are various types of textual information. Facts are used to illustrate about entities, events, and their properties. Opinions are especially used to illustrate about sentiments, feelings or emotions. Majority of the works carried out on previous models using contextual text mining are only at the document level. The problem associated with that approach is that phrase – level opinion mining which represents about the views of the opinions directly about the online reviews that are used to retrieve the items and their predictions. The major goal of the proposed work is to overcome the drawbacks of previous models by analyzing massive amount of data using the entire dataset with the help of machine learning model and thus achieving the high performance and accuracy rate. The accuracy rate can be improved by applying the ensembling mechanism where various classifiers are combined.

1.1. Machine Learning

Machine learning is a subset of artificial intelligence (AI) that provides the ability to fundamentally gain information and strengthen the systems. Machine learning focuses on the development of system programs that can retrieve information. The main goal of machine learning is [2] to allow the systems to learn automatically without manual involvement. Machine learning makes most of the tasks very faster and easier when compared to human involved tasks. The major difference between machines and humans while performing their tasks is intelligence. Different types of information's is collected by the five senses such as vision, smell, taste, hearing, and tactility. The collected information is sent to the human brain through the neural system to perform perception task. Where as in machine learning, machines cannot handle the gathered information in an intelligent way. Because machines do not have the ability to analyse the information for classification, and machines do not learn from the experiences. NLP is a part of data science which includes systematic processes for analysing, understanding, and deriving the information from the text in an efficient and systematic way. NLP [3] is

responsible for organizing the huge amount of text data. and performing various types of tasks which are automated such as sentiment analysis, speech recognition, topic segmentation, automatic summarization, machine translation. It includes two steps such as: a. Text preprocessing b. Text feature space.

Text Pre-Processing: Among all the available data the most unstructured format of data is text. Text pre-processing is the combination of both standardization and cleaning of text and provides the noise free and analysis ready environment. It consists of three steps such as removal of noise, normalization of lexicon, standardization of object.

Text Feature Space: The main job of feature vectors is to analyse the pre-processed data. Some important jobs of NLP are: Classification of text, matching the text, resolution of coreference.

2. Related Work

A research paper by Priyanka Tyagi and Dr.R.C.Tripathi [4] used python language in a particular analysis to implement the classification algorithm based on collected data and characteristics are obtained using the probabilistic language model. Emotions are classified including positive, negative and neutral using a supervised machine learning algorithm named as K-nearest neighbor. Vishal Gupta and Gurpreet S.Lehal [5] used information extraction which is the disclosure by system of current, earlier unspecified information, by consequently obtaining information from several written resources as well as a survey of information extraction methods. A.Jeyapriya and C.S.Kanimozhi Selvi [6] illustrated the system depending on expression level to analyze user evaluations. It is applied to obtain the most crucial elements of an item and to forecast the direction of each element from the item assessments. It recognizes the emotional aspect of each element by supervised machine learning algorithms. Haji Binali, Vidyasagar Potdar, and Chen Wu [7] applied opinion mining structure and exhibits present domain of analysis. Well formed direction of particular words in records set up their provisional connection through opinion mining. Total element emotion can be communicated depending on its emotion terms by explicitly recognizing its characteristics and the views. Abdullah Alsaeedi and Mohammad Zubair Khan [8] used examination of

twitter particulars that should be given enormous observation over the many years and includes deconstructing “tweets” and the capacity of these statements, analyzes several emotion assessments applied to twitter information and their results. Kassinda Francisco Martins Panguila and Chandra .J [9] described the extraction of several emotion actions to make a planned assurance and also support to classify emotion and fondness of people as transparent, conflicting or impartial. The famous classification techniques were used to obtain the emotion. The information was categorized through multilayer perceptron, convolutional neural networks, verified against the support vector network, random decision forest, classification tree, bayes theorem , etc. Priyavrat Chauhan, Nonita Sharma, and Geeta Sikka [10] applied the assessment about contextual text mining methods and the participation of the analysts to forecast polling outcome done with public network ideas, also allows analysis of inquiries that attempted to indicate the constitutional stand of connected users using public networks such as facebook and twitter, related with forecasting polling outcomes and impartial problems associated with contextual text mining. Shirin Hijaz Matwankar and Shubhash K. Shinde [11] defined a procedure that evaluates the constitutional strength that grabs actions of connected public, supporters and their connected actions. Constitutional outcomes are assessed by taking into account not just what customers are reporting, but also which connected society they are following and what their supporters are posting.. The procedure also evaluates constitutional results based on tweets gathered from the twitter. Brinda Hegde, Nagashree H S, and Madhura Prakash [12] described the method to obtain and evaluate tweets, categorize the tweets as positive or negative through machine learning methods with procedures, as well as performance assessment methods. The depreciation data file was acquired from Twitter via the Twitter API, pretreating was carried out using Natural Language Toolkit and Scikit-Learn, and as a result, it was subjected to analytical implementations including simple Bayes, logistic regression, and support vector networks. Risul Islam Rasel, Nasrin Sultana, Sharna Akhter, and Phayung Meesad [13] illustrated a method for obtaining opinions from public nodes as well as examine to see if they contain any explicit or abusive language.. Opinions are categorized among three different groups such as embarrassing, disgust presentation and neither. Form similarity inquiries are done to

recognize the interconnection among the forms. A precise text pretreating inquiry is done to generate an improved term carrier to train the prototype. Laszlo Nemes and Attila Kiss [14] used emotions as well as presentation related to customers about twitter public network, depending on the major tendency including NLP as well as emotion classification applying recurrent neural network(RNN). The trained prototype performs much more precisely, with a small mistakes, in choosing sentimental conformity in current time usually with problematic tweets. Ratna Patil and Sonia Arora [15] used public and subject environments which are merged by the laplacian matrix of the graph built by these environments and laplacian standardization is appended into the weblog contextual text mining prototype. Hypothetical outcomes on two real twitter data files exhibits that the proposed prototype can conquer standard techniques continually and consequently. In order to gather the tweets on pollings and analyse the opinions of the tweets, Kambhampati Kalyana Kameswari, J Raghaveni, R. Shiva Shankar, and Ch. Someswara Rao [16] utilised a machine learning-based classifier. The processed tweets were categorized using a supervised machine learning classification algorithm into three categories: positive, negative, and neutral about a particular politician. An amateurish classifier is used to categorise tweets as good, bad, or neutral. Gazi Imtiyaz Ahmad and Jimmy Singla [17] used fundamental ideas of contextual text mining luxurious dialect assets for unsupervised learning methods and exceptional calculated mechanisms for the supervised learning methods, in the area of natural language processing. Amruta U. Tarlekar, Manohar K.Kodmelwar [18] used examples taken from tweets or websites about constitutional rulers to recognise a variety of emotions that are negative as well as positive sides of communities. These sources included community twitter messages and a number of websites where the community shares its opinions about constitutional rulers. Depending on these outcomes, they also executed a combined design classification procedure to classify tweets or websites.

3. Proposed Work

The proposed work has several text processing and machine learning classification algorithms for determining the emotions of tweets. In order to depict the feature space, this mechanism has set targets such as negative (0), positive (1),

and neutral in combination with Term Frequency (TF) - Inverse Document Frequency (IDF). A number of machine learning techniques are evaluated to obtain the amazing results. These algorithms include Support Vector Machine (SVM), Naïve Bayes Classification (NBC), Random - Decision Forest (RDF), Decision Tree (DT), and Stacking Classifier (SC). The dataset of contextual text mining on social media of political leaders has been collected from the “Kaggle” dataset that can be crawled and labeled as either positive or negative. Emotions, usernames, and hashtags that are embedded in the data, this must be processed and converted into a standard format. Additionally, it is necessary to extract the text's useful properties, like unigrams and bigrams, that are used to express Tweets.

There are various algorithms to compare the datasets of political leaders which usually perform well in classification tasks, such as follows:

A. Naive Bayes Classification Algorithm: The Bayes rule is the foundation of the Naive Bayes classification method, which applies probability to examine a class based on a certain set of attributes.

B. Support Vector Machine Algorithm: the algorithms is applied to plot a graph by taking each unit of information such as an area in n-measurable place, where n is the number of attributes.

C. Random Decision Forest Algorithm: Random decision forest algorithm is applied for integrating various algorithms to create improved results for classification, statistical regression and other jobs.

D. Decision Tree Algorithm: Decision tree algorithm is applied for classification as well as statistical regression problems.

E. Stacking Classifiers: Stacking Classifiers integrate various machine learning procedures within one compatible design intending to reduce variance, bias, or enhance previsions.

The proposed model is depicted in figure 1.

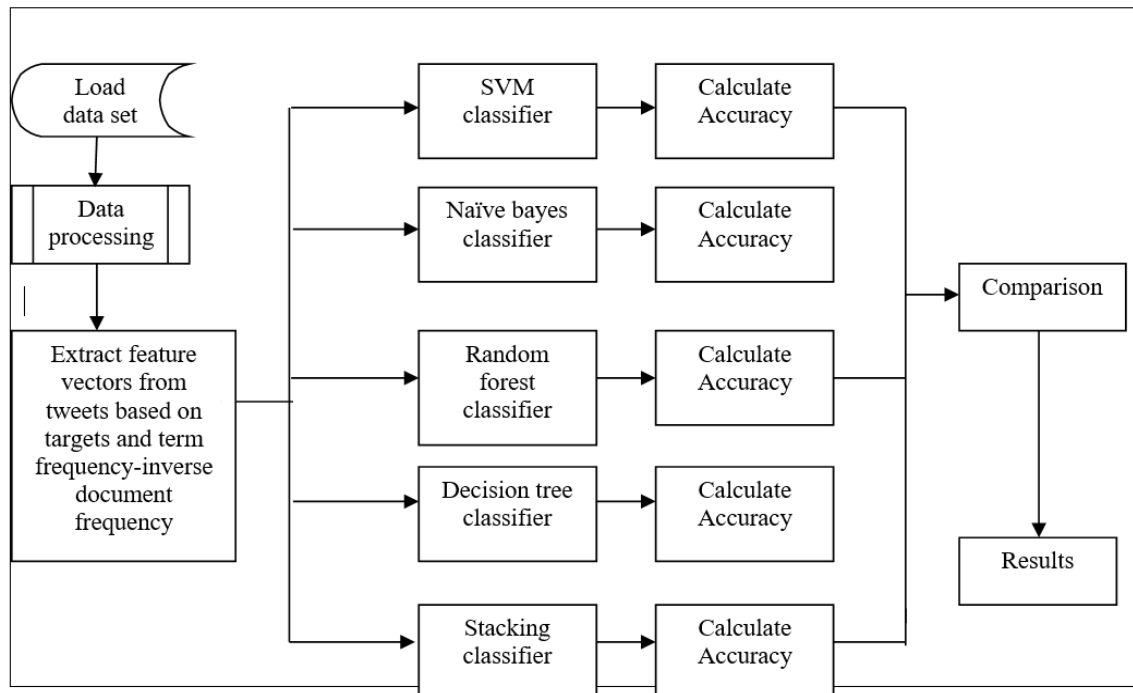


Figure 1. Proposed Architecture

Performance Prediction: Performance Prediction of contextual text mining on social media of political leaders can be measured by collecting the dataset from the twitter. The first step is to register for a twitter account and then extract the tweets from it. Then, using the Twitter website, the developers constructed a Twitter development application that gives them access to specific keys. Additionally, the tweets are extracted from the database using these keys and saved as a.csv (comma-separated value) file. The resulting.csv (comma-separated value) file is utilised as an input to produce the desired outcome from the sentiment analysis. By using the keys the extraction of the tweets are started. Further the access tokens are generated via these keys. After that tokenization is used to split the words from the tweets and then every word is referred with a dictionary and its polarity is measured. Term frequency-inverse document frequency (TF-IDF) and machine learning techniques like Support Vector Machines (SVM), Naive Bayes Classifier (NBC), and Random Decision Forest (RDF) are used to analyze tweets for sentiment. Therefore, two output target types have been created named as negative tweets and positive tweets.

- Negative tweets are tweets with sentiment 0
- Positive tweets are tweets with sentiment 1

The working mechanism of contextual text mining is described with two phases such as Building phase and Operational phase. The building phase consists of six steps, such as:

1. Extracting Dataset from the Twitter: In this method dataset have been retrieved using twitter application programming interface.

2. Loading Dataset: In this method dataset reviews have been loaded using panda's library.

3. Preprocessing Data: In this method pre-processed text-based reviews are represented earlier to produce characteristics for machine learning. Preprocessing steps are depicted in figure 2.

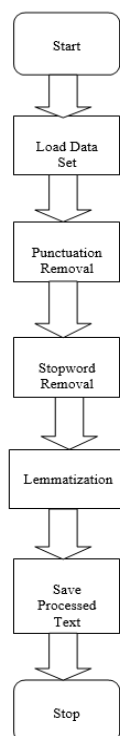


Figure 2. Preprocessing Steps

4. Feature Extraction: In this method natural language processing techniques have been used to convert text-based reviews to feature vectors.

5. Split Dataset into Train and Test Phase: In this method dataset can be splitted into two parts, namely train dataset and test dataset.

6. Training Classifier with Python Scikit Learn: In this method python scikit learn have been used along with classifier to perform training.

The operational phase is described with two methods such as:

1. Perform Predictions: In this step the predictions for predicting the dataset which are related to the new document for the future usage is performed . This predicted dataset is used to build a model.

2. Performance Calculation: the classification algorithms is used to fit the predicted data; this would run the collected data over the current data and give us the accuracy of the classification.

Datasets Used: The collection contains references to people as well as words, emoticons, symbols, and Uniform Resource Locators (URL). Uniform Resource Locators (URLs) and references have no relationship to sentiment prediction; however, words and emotions do. Therefore, Uniform Resource Locators (URL) and references cannot be considered. The proposed model used two type of datasets namely preprocessed train dataset, preprocessed test dataset. The data is offered as a comma-separated value (CSV) file that includes tweets and the thoughts that go with them. The training dataset consists of a comma-separated value (CSV) file of the following types: tweet_id, sentiment, and tweet, where tweet_id is a single integer that uniquely identifies a tweet and sentiment is either 1 (positive) or 0 (negative). Similarly, the test dataset is a comma separated value (CSV) file of type tweet_id, tweet. The Statistics of preprocessed train data set contains the information about Tweets, User Mentions, Unigrams, Bigrams with respect to the total, unique, average, max, positive, and negative values. These details are tabulated in Table 1.

Table 1. Statistics of Preprocessed Train Dataset

	Total	Unique	Average	Max	Positive	Negative
Tweets	200	-	-	-	400312	399688
User Mentions	200	-	0.4917	12	-	-
Unigrams	9823554	181232	12.279	40	-	-
Bigrams	9025707	1954953	11.28	-	-	-

The Statistics of preprocessed test data set contains the information about Tweets, User Mentions, Unigrams, Bigrams with respect to the total, unique, average, max, positive, and negative values. These details are tabulated in Table 2.

Table 2. Statistics of Preprocessed Test Dataset

	Total	Unique	Average	Max	Positive	Negative
Tweets	50	-	-	-	-	-
User Mentions	50	-	0.4894	11	-	-
Unigrams	24562	7828	12.286	36	-	-
Bigrams	225730	6865	11.29	-	-	-

Metrics Used: NLP are used to convert text-based reviews to feature vectors, NLP techniques that are used to convert text to vectors is known as term frequency-TF-IDF, TF metric is a process of measuring the frequently occurred term in a document. Most probably a term would appear much more in long documents than short documents. Thus, the term frequency (TF) is also represented as Number of times 't' appears in a document divided by the total number of terms in the document. Inverse Document Frequency (IDF) metric is used to measure the

importance of term. Thus, the inverse document frequency (IDF) is also denoted as $\log_e$ of total number of documents divided by the number of documents with term 't' in it. Therefore, two types of features are extracted from the dataset known as unigrams and bigrams. The well known and simple mechanism for text classification is named as unigram that consists of single words or tokens. A total of 181232 unique words are extracted from the dataset. To generate the vocabulary, only the top N words are used, where N is 15000 for sparse vector classification and 90000 for dense vector classification. Similar to Bigrams, word pairs in the dataset that appear consecutively in the corpus are likewise designated as such. Thus, 1954953 unique bigrams in total were retrieved from the dataset.

4. Results and Discussion

Table 3 illustrates about the comparison of accuracy of all ML techniques.

1. Support Vector Machine (SVM) has obtained an accuracy and performance rate of 75.70 percent.
2. Random Decision Forest (RDF) has obtained an accuracy and performance rate of 76.30 percent.
3. Naïve Bayes Classifier (NBC) has obtained an accuracy and performance rate of 73.10 percent.
4. Decision Tree (DT) has obtained an accuracy and performance rate of 68.40 percent.
5. The Proposed Model has obtained an accuracy and performance rate of 79.80 percent.

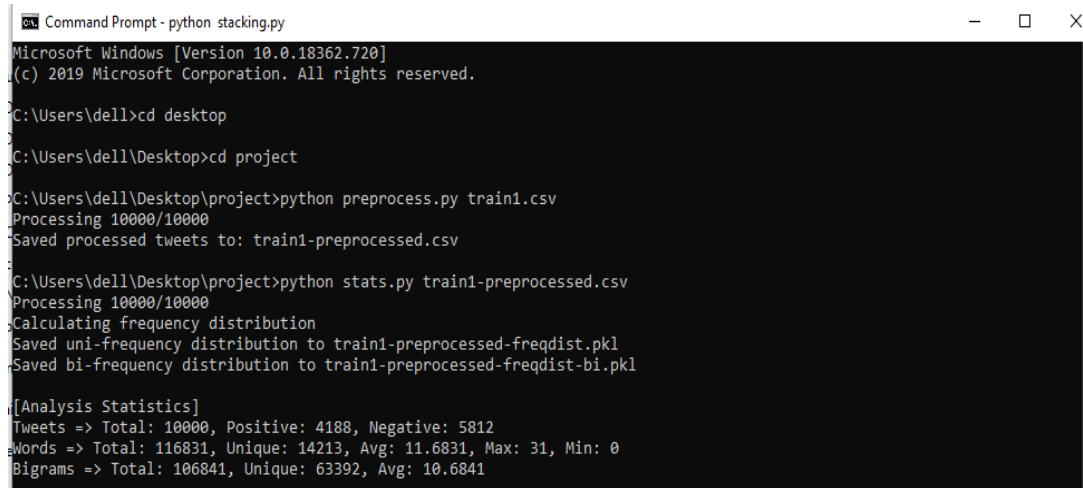
By comparing the performance and accuracy of various machine learning models it has been observed that the proposed model is having better and accuracy and performance rate of 79.80 percent.

Table 3. Comparison of Accuracy of all ML Techniques

S.no	Classification Techniques	Accuracy
1	SUPPORT VECTOR MACHINE (SVM)	75.70
2	RANDOM DECISION FOREST (RDF)	76.30
3	NAIVE BAYES CLASSIFIER (NBC)	73.10
4	DECISION TREE (DT)	68.40
5	PROPOSED MODEL	79.80

a. Data Pre-Processing and Statistics

The below diagram i.e figure 3 illustrates about data pre-processing and Statistics which has the information about total tweets of 10000, positive tweets of 4188, negative tweets of 5812, total words of 116831, unique words of 14213, average words of 11.6831, maximum words of 31, minimum words of 0, total bigrams of 106841, unique bigrams of 63392, average bigrams of 10.6841.



```

Command Prompt - python stacking.py
Microsoft Windows [Version 10.0.18362.720]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\dell>cd desktop
C:\Users\dell\Desktop>cd project
C:\Users\dell\Desktop\project>python preprocess.py train1.csv
Processing 10000/10000
Saved processed tweets to: train1-preprocessed.csv
C:\Users\dell\Desktop\project>python stats.py train1-preprocessed.csv
Processing 10000/10000
Calculating frequency distribution
Saved uni-frequency distribution to train1-preprocessed-freqdist.pkl
Saved bi-frequency distribution to train1-preprocessed-freqdist-bi.pkl

[Analysis Statistics]
Tweets => Total: 10000, Positive: 4188, Negative: 5812
Words => Total: 116831, Unique: 14213, Avg: 11.6831, Max: 31, Min: 0
Bigrams => Total: 106841, Unique: 63392, Avg: 10.6841

```

Figure 3. Data-Preprocessing and Statistics

b. Unigrams

The graph plotted in figure 4 illustrates about the unigrams information. Unigram has been obtained by plotting the graph between the total number of words and their corresponding frequencies. Total number of words are ranged from 1000 to 4000 and frequencies of top 20 unigrams.

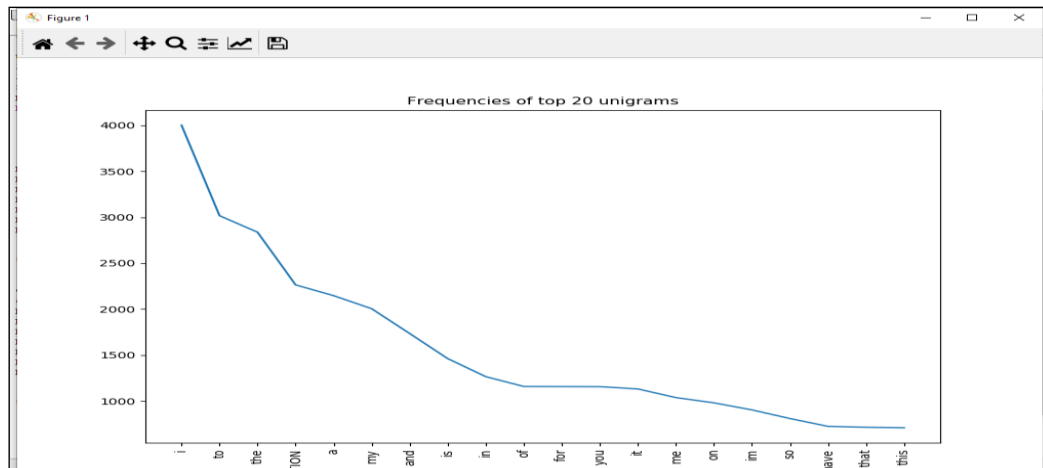


Figure 4. Unigrams

c. Unigrams Zipf Law

The graph plotted in figure 5 illustrates about the unigrams frequencies follow zipf's law. Where x-axis shows the log (rank) and y-axis shows the log (frequency). The log(frequency) is ranged from 5.0 to 9.0 and log (rank) is ranged from 0 to 4.

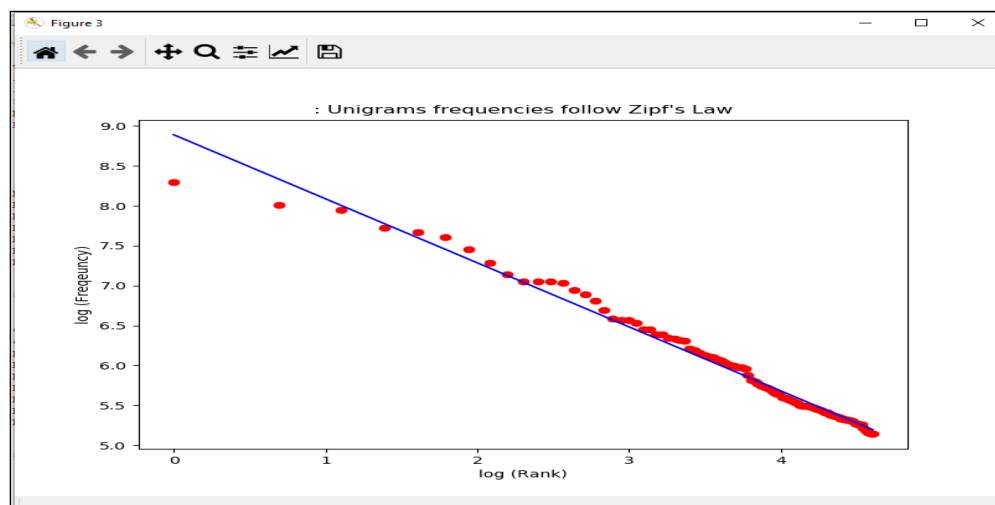
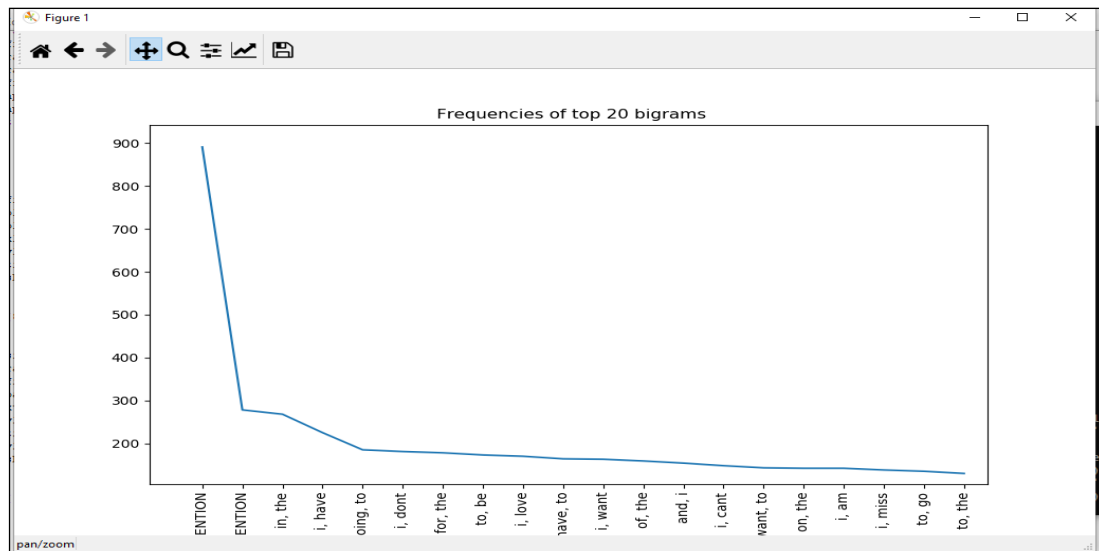
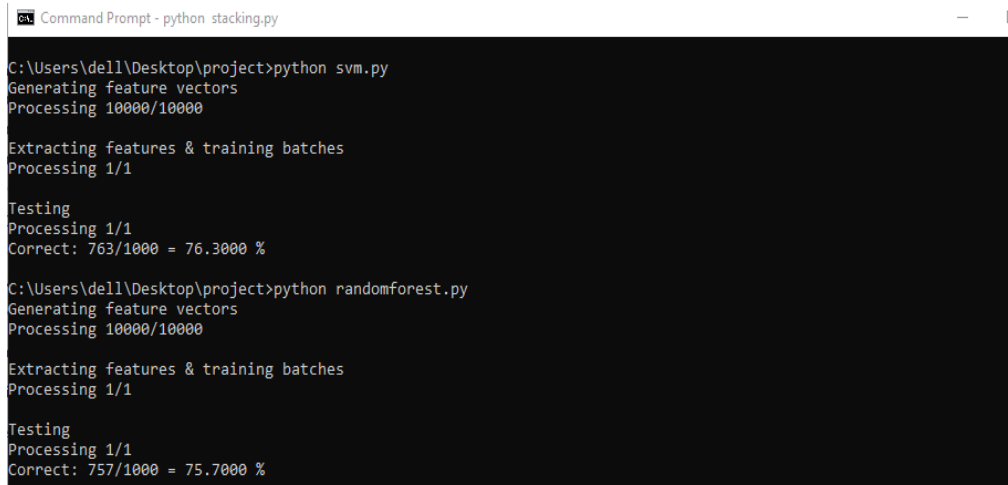


Figure 5. Unigrams Zipf Law**d. Bigrams**

The graph plotted in figure 6 illustrates about the bigrams information. Bigram has been obtained by plotting the graph between the total number of words and their corresponding emotions. Total numbers of words are ranged from 200 to 900 and frequencies of top 20 bigrams.

**Figure 6. Bigrams****e. Result of Support Vector Machine (SVM) & Random Decision Support (RDS) Algorithm:**

The below diagram i.e figure 7 illustrates about the results of Support Vector Machine (SVM) and Random Decision Support (RDS) Algorithm. By applying extracting features and training batches Support Vector Machine (SVM) has obtained an accuracy of 75.7 percent and Random Decision Support (RDS) has obtained an accuracy of 76.3 percent.



```

Command Prompt - python stacking.py

C:\Users\dell\Desktop\project>python svm.py
Generating feature vectors
Processing 10000/10000

Extracting features & training batches
Processing 1/1

Testing
Processing 1/1
Correct: 763/1000 = 76.3000 %

C:\Users\dell\Desktop\project>python randomforest.py
Generating feature vectors
Processing 10000/10000

Extracting features & training batches
Processing 1/1

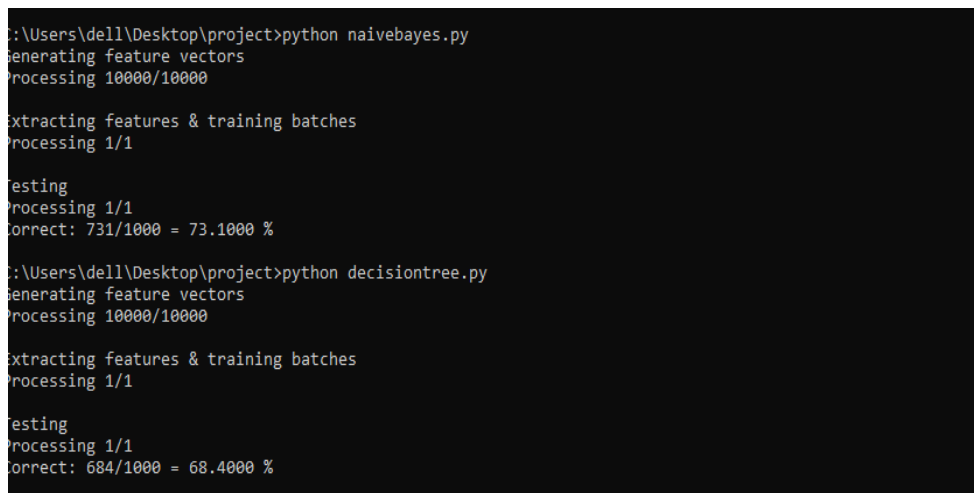
Testing
Processing 1/1
Correct: 757/1000 = 75.7000 %

```

Figure 7. Result of Support Vector Machine (SVM) & Random Decision Support (RDS) Algorithm

f. Result of Naïve Bayes Classifier (NBC) & Decision Tree (DT) Algorithm:

The below diagram i.e figure 8 illustrates about the results of Naive Bayes Classifier (NBC) and Decision Tree (DT) Algorithm. By applying extracting features and training batches Naive Bayes Classifier (NBC) has obtained an accuracy of 73.10 percent and Decision Tree (DT) Algorithm has obtained an accuracy of 68.40 percent.



```

C:\Users\dell\Desktop\project>python naivebayes.py
Generating feature vectors
Processing 10000/10000

Extracting features & training batches
Processing 1/1

Testing
Processing 1/1
Correct: 731/1000 = 73.1000 %

C:\Users\dell\Desktop\project>python decisiontree.py
Generating feature vectors
Processing 10000/10000

Extracting features & training batches
Processing 1/1

Testing
Processing 1/1
Correct: 684/1000 = 68.4000 %

```

Figure 8. Result of Naïve Bayes Classifier (NBC) & Decision Tree (DT) Algorithm

g. Result of Proposed Model Algorithm

The below diagram such as figure 9 illustrates about the result of Proposed Model Algorithm. By applying extracting features and training batches Proposed Model Algorithm has obtained an accuracy of 79.80 percent.

```

C:\Users\dell\Desktop\project>python stacking.py
Generating feature vectors
Processing 10000/10000

Extracting features & training batches
Processing 1/1

Testing
Processing 1/1
Correct: 798/1000 = 79.8000 %

C:\Users\dell\Desktop\project>

```

Figure 9. Result of Proposed Model Algorithm

h. Comparison Graph of Various Algorithms

The graph plotted in figure 10 shows the comparison graph of various machine learning algorithms. Where x-axis shows the various machine learning algorithms and y-axis shows the corresponding accuracy and performance rates of those machine learning algorithms.

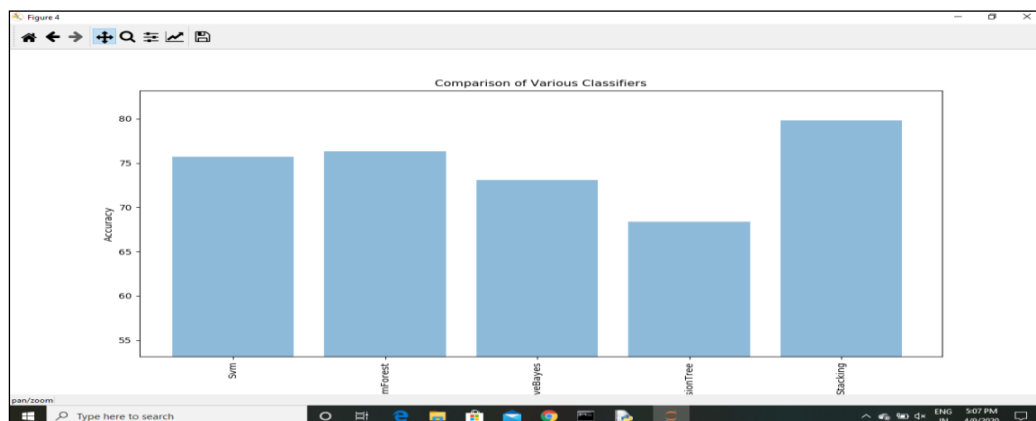


Figure 10. Comparison Graph of various Machine Learning Algorithms

Comparison of accuracy of Various Machine Learning Algorithms

The graph plotted in figure 11 shows the comparison of accuracy of various machine learning algorithms. Where x-axis shows the various machine learning algorithms and y-axis shows the corresponding accuracy and performance rates of those machine learning algorithms. The Proposed Algorithm has the highest accuracy compared to the other machine learning algorithms.

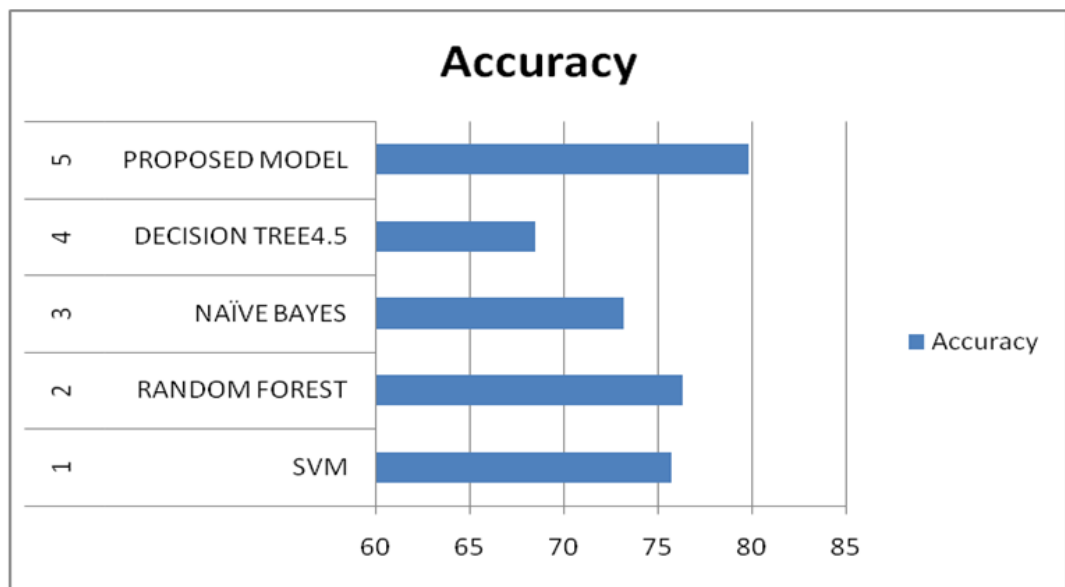


Figure 11. Comparison Results

5. Conclusion

The use of contextual text mining for tweets has a broad opportunity as it examines the emotions of the people in an area and thus all the methodologies and assessment can be depending on how the people of that area acknowledge several political mandates. A dataset has been extracted from twitter and used term frequency-inverse document frequency to create feature vectors from this dataset. The experimental results demonstrate that the methods used are very promising in executing their jobs. Moreover, the accuracy of several classifiers is compared in determining the emotions of tweets, the results showing that naive bayes gave best accuracy compared to remaining classifiers in determining the emotions of tweets.

References

- [1] Bo Pang and Lillian Lee, “Opinion Mining and Sentiment Analysis”, Foundations and Trends In Information Retrieval, Vol. 2, No.1-2, 2008, pp. 1-135.
- [2] Benjamin M. Abdel-Karim, Nicolos Pfeuffer and Oliver Hinz, “Machine learning in Information systems - a bibliographic review and open research issues”, Electronic Markets, Pages 643-670, April 2021.
- [3] Bill Manaris, “Natural Language Processing: A Human-Computer Interaction Perspective”, Vol. 47, 1998, pp. 1-66.
- [4] Priyanka Tyagi and Dr. R.C. Tripathi, “A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data”, International Conference on Advanced Computing and Software Engineering, 2019.
- [5] Vishal Gupta and Gurpreet S. Lehal, “A Survey of Text Mining Techniques and Applications”, Journal of Emerging Technologies in Web Intelligence, Vol. 1, No. 1, August 2009.
- [6] A.Jeyapriya and C.S.Kanimozhi Selvi, “Extracting Aspects and Mining Opinions in Product Reviews Using Supervised Learning Algorithm”, International Conference on Electronics and Communication Systems, 2015.
- [7] Haji Binali, Vidyasagar Potdar, and Chen Wu, “A State of the art opinion mining and its application domains”, IEEE International Conference on Industrial Technology, pp. 1–6, February 2009.
- [8] Abdullah Alsaedi and Mohammad Zubair Khan, “A Study on Sentiment Analysis Techniques of Twitter Data”, International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019.
- [9] Kassinda Francisco Martins Panguila and Chandra .J, “Sentiment Analysis on Social Media Data Using Intelligent Techniques”, International Journal of Engineering Research and Technology, Volume 12, Number 3, 2019, pp. 440-445.

- [10] Priyavrat Chauhan, Nonita Sharma, and Geeta Sikka, “The emergence of social media data and sentiment analysis in election prediction”, *Journal of Ambient Intelligence and Humanized Computing*, February 2021.
- [11] Shirin Hijaz Matwankar and Shubhash K. Shinde, “Sentiment Analysis for Big Data using Data Mining Algorithms”, *International Journal of Engineering Research & Technology*, Vol. 4, Issue 09, September-2015.
- [12] Brinda Hegde, Nagashree H S, and Madhura Prakash, “Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets”, *International Research Journal of Engineering and Technology*, Volume. 05, Issue. 06, June-2018.
- [13] Risul Islam Rasel, Nasrin Sultana, Sharna Akhter, Phayung Meesad, “Detection of Cyber-Aggressive Comments on Social Media Networks: A Machine Learning and Text mining approach”, *International Conference on Natural Language Processing and Information Retrieval*, September 2018.
- [14] Laszlo Nemes and Attila Kiss, “Social media sentiment analysis based on COVID-19”, *Journal of Information and Telecommunication*, Volume 5, Issue 1, 2021.
- [15] Ratna Patil and Sonia Arora, “Sentiment Analysis of Social and Topic Context Using Machine Learning Techniques”, *Journal of Critical Reviews*, Volume 7, Issue 10, 2020.
- [16] Kambhampati Kalyana Kameswari, J Raghaveni, R. Shiva Shankar, and Ch. Someswara Rao, “Predicting Election Results using NLTK”, *International Journal of Innovative Technology and Exploring Engineering*, Volume-9, Issue-1, November 2019.
- [17] Gazi Imtiyaz Ahmad and Jimmy Singla, “Machine Learning Techniques for Sentiment Analysis of Indian Languages”, *International Journal of Recent Technology and Engineering*, Volume-8, Issue-2, September 2019.
- [18] Amruta U. Tarlekar, Manohar K.Kodmelwar, “A Survey of Mining Online Public Opinions from twitter on Political Domain Using Machine Learning Techniques”, *International Journal of Science and Research*, Volume 3 Issue 11, November 2014.

Author's Biography



Mr. Aleemullakhan Pathan received B.Tech and M.Tech from Jawaharlal Nehru Technological University, Kakinada, AP. He is currently associated with Madanapalle Institute of Technology and Science, Madanapalle-517325, India as Assistant Professor of Computer Science and Engineering of the Institute since the year 2022. His research areas include Data Mining, Machine Learning, Deep Learning, Artificial Intelligence, Cryptography and Network Security. He published four research papers in various reputed International Journals and Conferences. He is the member of IE and CSI.



Mr. R.Sundar received Ph.D (Tech., CSE) from Sathyabama University, Chennai, Tamil Nadu, India. He is currently associated with Madanapalle Institute of Technology and Science, Madanapalle-517325, India as Assistant Professor of Computer Science and Engineering of the Institute since the year 2021. His research areas include Mobile Adhoc Network, Machine Learning, Deep Learning. He has published four papers in International Journals and Conference proceedings. He has also attended more than five conferences at National and International levels.