International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-12, MAR 2021

Analysis of Dubious Conversations on Online Forums

[1]Namita Mhatre, [2]Manav Mishra, [3]Kaustubh Lawale, [4]Prof. Govind Wakure

Department of Information Technology, MCT's Rajiv Gandhi Insititute of Technology, Andheri (W), Mumbai, India

mhatrenamita23@gmail.com, manav.mishra26@gmail.com, kaustubhlawale4@gmail.com, govind.wakure@gmai.com

Abstract - With the revolution of social media, it facilitated a lot in reshaping and redefining the way people communicate and express themselves. More and more people regardless of their age and gender are signing up for profiles on social networking platforms. Facebook, Twitter, Reddit are among some of the most used social networking forums. Data is being produced at an exponential rate by internet users on various platforms, it becomes mandatory to scrutinize and make use of the data for the Government organizations and understand the sentiment of the users. The rapidly progressing technology has resulted in the upsurge to many cybercrime activities which majorly include cyber bullying, harassing, making threats, etc. Sentiment analysis also known as opinion mining helps in digging into the words and context of the social media text data to understand their true nature in terms of polarity i.e. positive, negative, or neutral. In this paper we have considered various techniques of performing sentiment analysis in particular Machine learning, Deep learning and Lexicon based approach. This research paper elaborates the methodologies used to analyze the negative and positive relationship between the users by using comments and tweets etc.

Keywords — Deep Learning, Machine learning, , Natural Language Processing, Polarity, Sentiment analysis.

I. INTRODUCTION

In less than a generation, social media has altered the way people communicate with each other and express their emotions. Social media is rich with raw and unstructured data and with the development of technology, specifically in machine learning and artificial intelligence, allows the data to be processed and transforms it into convenient data that can aid many government and business organizations. Facebook is the largest social media platform in the world with nearly 2.79 billion users. Other popular platforms like Twitter and Reddit have around 400 million users monthly. In this study, we have chosen to work with online forums like Twitter, Reddit, and Facebook since we feel it is a better approximation of public sentiment as opposed to conventional internet articles and web blogs. The reason is that the amount of relevant data is much larger for these as compared to traditional blogging sites. Moreover, the response here is more prompt and also more general since the number of users who are active on the above-mentioned platforms is substantially more than those who write web blogs on a daily basis. Escalating criminal activities on digital mediums alerts the legal bodies to constantly monitor online activities. Social media bullying and trolling and have become serious concern these days and a system that is able to determine such texts would surely be of great use in creating a safer and bully-free online environment.

Several survey and facts have proved that it is difficult to manage information which constantly keeps changing on internet thus data mining is the optimal choice to analyze and gather data. Initially, researchers extensively focused on predicting positive sentiments. Later, the concept of negative sentiments slowly emerged, and gained popularity. What makes negative sentiment prediction more challenging than just a positive sentiment prediction is the fact that negative interactions are never explicitly available on any social networking website. For example, Facebook and Twitter have 'like', Reddit has 'up vote', but the distrust or negative relationship between two users is never explicitly available. But messages and comment can more effectively determine if the interaction is positive or negative. So that is where we design different methods to extract the hidden negative interactions out of the available information-centric interactions between users. Using varied data mining techniques, unprocessed data is extracted from a large text corpus and this data is transformed into structured data in pre-processing. Sentiment analysis is a natural language processing technique used to interpret and classify emotions in textual data. Current approaches to sentiment polarity analysis are either lexicon-based, machine learning, or deep learning based. In this paper we have tried to incorporate all three in analyzing the sentiment of data obtained from social networking platforms.

Natural Language Processing is a subdivision in the field of Artificial Intelligence. This field pursues to provide a link for interaction with humans through natural language. The development process of NLP includes several steps. The development cycle of NLP begins with collecting text. Text collected in a set is known as a corpus. Then, all the textual

 DOI : 10.35291/2454-9150.2021.0116

data is processed because not all text is structured. Further, the process continues to the initial pre-processing phase. Text processing is done using several NLP techniques. Pre-processing the data is the process of cleaning and preparing the text for classification Then, change to the feature engineering phase. This phase attempts to get attributes from unprocessed data. Due to existence of features that have been formed, the text becomes structured and ready for calculation to produce sentiment.

Sentiment Analysis is a field of study which is a part of NLP and aims to interpret people's opinions, on certain topics, about any event, etc. There are different algorithms that can be implemented in sentiment analysis models, depending on how much data needs to be analyzed, and how accurate should the model be. Rule-based approach uses a set of human-crafted rules to help analyze the subjectivity, polarity, or the subject of an opinion. These rules may include various NLP techniques developed in computational linguistics, such as: Stemming, tokenization, and parsing, Lexicons. Contrary to rule-based systems, automatic approach doesn't rely on manually crafted rules, but on machine learning techniques. A sentiment analysis task is usually expressed as a classification problem, whereby a classifier is given a text and returns a category, e.g. positive, negative, or neutral. Deep learning which is considered as a further development of machine learning uses multiple algorithms in a progressive chain of events to solve complex problems and allows you to tackle accurately massive amounts of data.

II. LITERATURE REVIEW

To understand the necessity and process of sentiment analysis of social media data and to start in on providing our analysis, we first need to comprehend the existing research and the work done in this field. In this section, we illustrate an incisive review performed on the existing pieces of work. Accordingly, several different research papers have been explored to gather the related information regarding the project.

A research paper [4] conducted a systematic literature review which provides information on studies of sentiment analysis in social media data. The paper made the following three contributions. First, it introduced the various methods used for analyzing sentiments which includes Lexicon based method i.e. SentiWordnet and TF-IDF while for machine learning it is Naïve Bayes and SVM. Second, they identify what is the most common type of social media site to extract information for sentiment analysis that is Twitter. Third, they demonstrated the application of sentiment analysis in social media.

In a paper by [2], the twitter data is collected and passed through machine learning classifiers. After being classified by individual classifiers, a voted classification mechanism has been used to finally obtain the class of the "tweets" and the percentage confidence on it. Polarity method for classification has also been used to find the percentage of positive and negative tweets. Lastly, deep learning models have been implied to classify the tweets. Models like RNN, LSTM, CNN-RNN, have been utilized to classify the tweets. The paper concluded that deep learning models and their various combinations showed better performance compared to machine learning algorithms. This Survey paper presented an overview on the recent updates in SA algorithms and applications. Fifty-four of the recently published and cited articles were categorized and summarized. These articles give contributions to many SA related fields that use SA techniques for various real-world applications. After analyzing these articles, it is clear that the enhancements of SC and FS algorithms are still an open field for research. Naı¨ve Bayes and Support Vector Machines are the most frequently used ML algorithms for solving SC problem. They are considered a reference model where many proposed algorithms are compared to. Information from micro-blogs, blogs and forums as well as news source, is widely used in SA recently.[5]

This research design was an experimental research where the twitter API was used to get as many as 100 tweets. The initial step of using the system was done by inputting the topic on the system. The numbers of tweets produced were as many as 100 in which 65 tweets included in the positive classification. Furthermore, 23 tweets are included in the negative classification while 12 tweets are included in the neutral classification. In this study 250 English datasets were provided and 112 adjectives were used as lexicon. Then, a comparison was made of previous studies using the machine learning approach. The paper stated that sentiment analysis using the lexicon approach has lower accuracy than using machine learning approach.[3]

Sentiment analysis is one of the computational techniques of opinion, sentiments, and the variety of texts subjectivity. In this paper, the methodology of determining these public opinions was discussed. The development of a program for sentiment analysis was done to create a platform for social network analysis. This paper also discusses the sentiment analysis design, gathering data, training the data, and visualizing the data using the Python library. Finally, a platform is designed in order for other users to search the sentiment results of particular topics of interest. A total of 3000 Reddit data and 3000 Twitter data has been gathered, cleaned, analyzed, and visualized in this research.[1]

III. METHODOLOGY

The analysis of the data collected from varied social media was carried out using multiple techniques which included machine learning, deep learning as well as lexicon-based approach. These practices were done to make sure that our system should be able to provide the versatility and

 DOI : 10.35291/2454-9150.2021.0116

alternatives when it comes to detecting the sentiment of the data accurately.

I. **Analysis of tweets collected from Twitter using Machine Learning Algorithms**:

1. Collecting Data: In this approach, the initial step was to extract a sample twitter dataset of 50,000 tweets in CSV format. This dataset consisted of 30,000 training set and further 20,000 tweets were used for testing the model. The training dataset included pre-labeled tweets, where label '1' denoted the tweet is racist/sexist (negative) and label '0' denoted the tweet is not racist/sexist (positive). Our objective was to predict the labels on the given test dataset. The Figure No 1 given below, includes the steps to be followed to achieve the sentiment analysis of the given dataset. Each of these steps are elaborated further in detail.

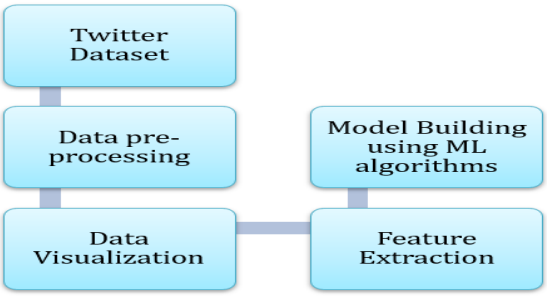


Figure No. 1 Proposed System of Twitter Sentiment Analysis

2. Data pre-processing: The preprocessing of the textual data is a crucial step as it makes the raw text data ready for mining. The intension of this step is to clean the noise in the data which is lacking relevance to find the sentiment of tweets such as punctuation, special characters, numbers, and terms which don't carry much meaning in context to the text data.

-Removing Twitter Handles (@user_handle) :The tweets consist of a number of twitter handles (@user_handle), that is how a user is recognized on Twitter. We removed all the user handles as they don't give much information.

-Removing Punctuations, Numbers, and Special Characters

-Removing Short Words : Short words use in tweets don't provide much information overall, rather removing them would help in better feature selction. Selecting the length of the words which we want to remove, we decided to remove all the words having length 3 or less. For example, terms like "his", "is", "oh" are not of much use.

-Tokenization: Tokenization is the process of dividing a sentence, paragraph, or an entire document into smaller units, such as individual words also called as terms. Each of these small scale units are called as tokens.

```
tokenized_tweet = combi['tidy_tweet'].apply(lambda x: x.split()) # tokenizing
tokenized_tweet.tail(10)
```

```
49149    [loving, life, #createyourfuture, #lifestyle, #holiday, hyatt, regency, long, beach]
49150    [black, professor, demonizes, proposes, nazi, style, confiscation, white, assets, like, germany, #breaking]
49151    [learn, think, positive, #positive, #instagram, #instagood]
49152    [love, pretty, happy, fresh, #teenilicious, #fixdermateen, #generationz, #pretty, #fresh]
49153    [damn, tuff, ruff, muff, techno, city, uktnx, #web, hardcore, gabba]
49154    [thought, factory, left, right, polarisation, #trump, #uselections, #leadership, #politics, #brexit, #blm]
49155    [feeling, like, mermaid, #hairflip, #neverready, #formal, #wedding, #gown, #dresses, #mermaid]
49156    [#hillary, #campaigned, today, #ohio, used, words, like, assets, liability, never, once, #clinton, thee, word, #radicalization]
49157    [happy, work, conference, right, mindset, leads, culture, development, organizations, #work, #mindset]
49158    [song, glad, free, download, #shoegaze, #newmusic, #newsong]
Name: tidy_tweet, dtype: object
```

Figure No. 2 Tokenized Tweets

-Stemming : Normalizing a text means reducing variations (specially style variation) in the most possible way. One of the ways to accomplish this is by stripping off word suffixes. It is a rule-based technique of removing the suffixes from a word. For example – "eat", "eating", "eats" are the different variations of the word – "eat". In this NLTK's Porter Stemmer function is used to Normalize the tweets. Figure No. 2 demonstrates the tokenized data, after the application and completion of pre-processing of raw data.

3. Data Visualization: Here the processed data in explored further to gain more insights. Data visualization is the act of taking data and settling it into a visual environment like a graph and charts. It also makes the process of detecting patterns, trends, and outliers in groups of information easier.

-To recognize the most common words used : A WordCloud is used for this wherein it creates a visualization of the most frequently used words which appear in larger size and less frequent words appear in smaller size. To understand the frequently used words separate WordClouds were built for negative and positive words. Figure No. 3 gives a visualization of negative and positive words WordCloud. In the negative words WordCloud, "trump", "racist", "white" are some of the most used words where as "love", "life", "smile" are frequently used positive words.

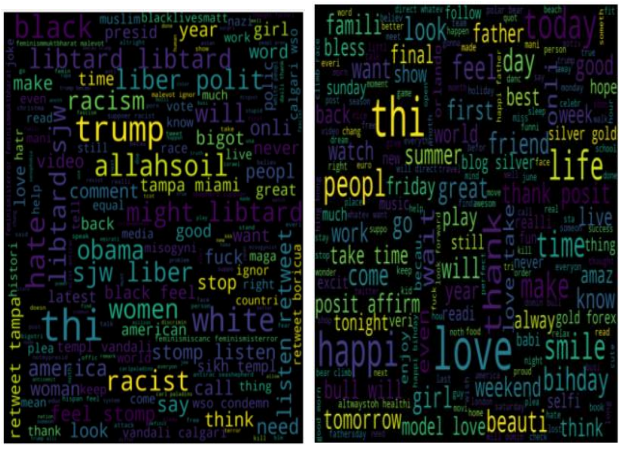


Figure No. 3 WordCloud of Negative and Positive words in Tweet

-To recognize the impact of hashtags: Hashtags in twitter are equivalent with the ongoing trends on twitter at any point in time. Here we tried to check if the hashtag adds

 DOI : 10.35291/2454-9150.2021.0116

value to our sentiment analysis problem. To understand this, the most used positive and negative hashtags were plotted in a bar graph as given in Figure No. 4.

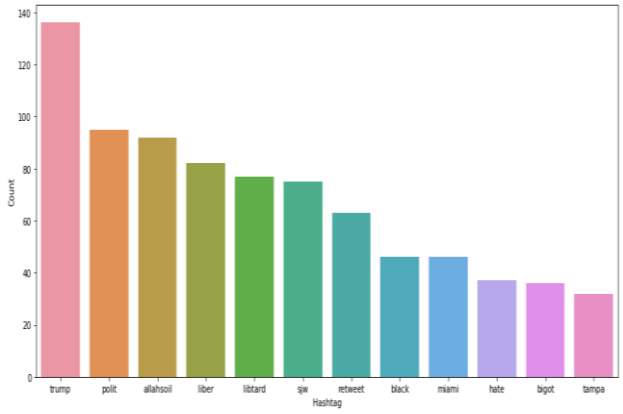


Figure No. 4 Bar graph of the Most used Negative hashtags

4. Feature Extraction: To study processed data, the data should be converted into features. Feature extraction is the technique that selects and combine variables into features and correctly describes the original data set as a whole., and minimizes the amount of data that needs be processed. Depending upon the usage, features are constructed using varied techniques**,** such as– Bag-of-Words, TF-IDF, and Word Embeddings.

-Bag-of-Words Features : Bag-of-Words is a technique to express text as numerical features. It is a description of how many times words occurs in a text document.

-TF-IDF Features: Term frequency-inverse document frequency TF-IDF is a method that signifies how important a word is to a document in a corpus. It is different from the bag-of-words approach such that it takes into account, not just the occurrence of a word in a single document or tweet but in the entire corpus.

-Word Embedding: Word embeddings are a sort of word representation which lets words with comparable meaning have same type of representation.

5. Building a model:After data preparation and feature extraction, the next step was to create sentiment analysis model. This problem of sentiment analysis is modeled as a classification problem. For that we used different Machine learning classification algorithm namely:

-Logical Regression: Logistic regression which is a classification ML algorithm assigns observations to a discrete set of classes. It is a predictive analysis algorithm. As it is based on probability, it predicts the probability of occurrence of an event by placing the data in a logic function.

-Support Vector Machine: SVM is a supervised classification machine learning algorithm. The goal of this algorithm is to generate the best line or decision boundary that can partition n-dimensional space into classes so that the new data point can be put in the correct category in the future. This best boundary is called a hyperplane. SVM selects the extreme points or vectors that assist in forming the hyperplane. Support vectors are basically these extreme cases.

-Random Forest: Random Forest is a classification algorithm, it consist of several decision trees on different subsets of the given dataset. It takes the average of the decision trees to provide better the predictive output of that dataset. The "forest" that Random forest algorithm builds, is an ensemble of decision trees, mostly trained using the “bagging” method. The basic idea of the bagging method is that the combination of learning models improves the overall output. So rather than relying on one decision tree, this algorithm takes the predictive output from each tree and predicts the final output based on the majority votes of predictions.

-Extreme Gradient Boost : XGBoost designed for speed and performance is an implementation of gradient boosted decision trees. It fast learning is provided through parallel and distributed computing and efficient memory usage is also offered. It is an ensemble learning technique. Occasionally, it may not be enough to rely upon the outputs of only one ML model. Ensemble learning provides a systematic solution to unite the predictive power of multiple learners. The result of this is a single model which is able to the aggregated output from several models. In boosting, as the trees are built sequentially each subsequent tree focuses on reducing the errors of the previous one. Each tree learns from its former and updates the errors. Subsequently, the tree that extends next in the sequence will learn from an updated version of the residuals. In contradiction to bagging techniques like Random Forest, in involves trees are growing to their maximum extent, boosting makes use of trees with fewer splits. Small trees like this which are not very deep, are highly interpretable.

II. Analysis of comments collected from Facebook using Deep Learning:

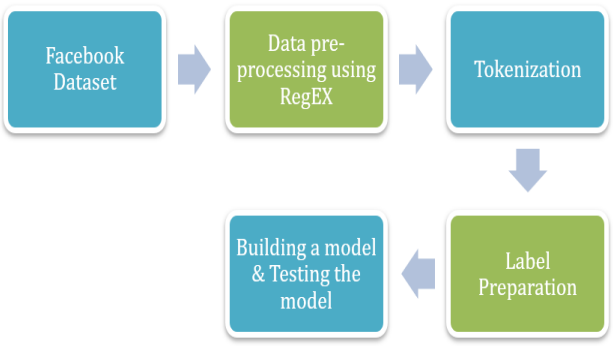


Figure No.5 Proposed System of Facebook Sentiment Analysis

 DOI : 10.35291/2454-9150.2021.0116

1. Data Collection and Preprocessing: In this approach a dataset of 1000 Facebook comments was downloaded. This dataset was divided into training 67% and testing 33% set. With this Neural network, our aim was to predict whether or not a comment is Positive (P) or Negative (N) also the comments with the sentiment labeled different (O) were of no use to us, therefore it's far from the dataset. The text data is lowercased, and the symbols are removed through RegEx. Regular Expression RegEx is an object that can be used for information extraction from any data primarily string whether it is number letter or punctuations. RegEx will allow you to check and match any character combination in strings. RegEx, sequence of characters defines a search pattern

2. Tokenization: Tokenization is a common task in language process (NLP). It's an elementary step in each ancient information science strategies like Count Vectorizer and Advanced Deep Learning-based architectures like Transformers.

3. Label preparation: Here the labels area is unit checked once the removal of the "other" sentimental comments. Additionally, some preparation to the rule, such as making the test ready, and training sets ready.

4. Building a Model: In this section, the algorithm is implemented with following features prepared. As we are working with text stream, sequential model is the most preferred one. A sequential model is suitable for a plain stack of layers where every layer has one input and output tensor.

5. Testing and Training the model: Here we train the model, plotted a histogram over the 7 epochs and plotting the accuracy and loss, Testing the model, and retrieving score and accuracy. Also, we validated for the model's accuracy in predicting either a positive, or a negative score. The histogram in Figure No.6 shows how the accuracy and loss of the model has increased and decreased respectively over a period of 7 epochs.

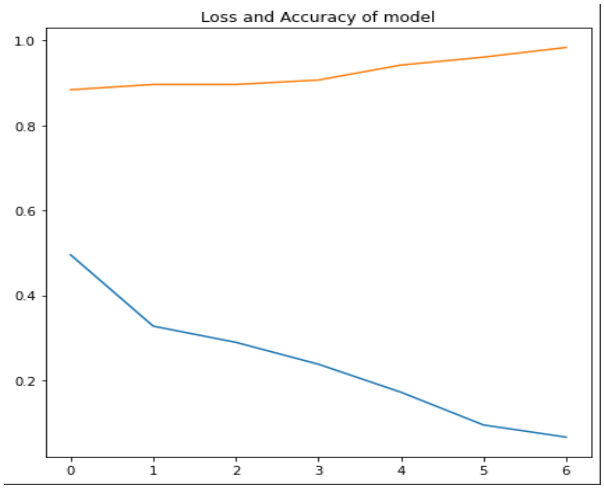


Figure No.6 Facebook Loss and Accuracy

III. Sentiment Analysis using Lexicon approach:

1. Reddit: PRAW which is stands for "Python Reddit API Wrapper", is a Python package which allows for simple access to Reddit's API. PRAW is simple to utilize and follows all of Reddit's API rules. PRAW was used to establish a connection to Reddit API. Consequently, we were able to get any particular subreddit information mentioned like posts and comments using its functions. The next step was obtaining data from subreddit: Post title, comments, and replies. Posts on subreddit square measure divided into 2 components, the title and the comment section. Some titles have a low description that represents a read of the post author. Most voted comment and its replies are visible on prime of the comment section. VADER that is Valence Aware dictionary for sentiment reasoning is a lexicon and rule-based sentiment analysis tool that's specifically attuned to sentiments expressed in social media. VADER was imported for sentiment analysis of Reddit comments and topics. We can access the VADER operating sia object. The sia.polarity takes a sentence as input and offers sentiment score as output. From nltk_sentiment() sentences are categorized into positive, negative, and neutral sentiment.

2. Twitter: A twitter developer account is created and a python library "Tweepy" is used to access and fetch tweets from Twitter API. Cursor object is used to access our own tweets from our own timeline, accessing user tweets from a specific user whose tweets we want to analyze. Pagination was done to make sure the content is divided across series of pages. For analyzing the tweets Pandas and Numpy library was used. Pandas will allow for the tweets to be stored into data frame.

After data frame creation of each tweet, we can visualize the data with a time series graph. We plotted the number of likes a tweet got on a given day, for which Matplotlib was used. TextBlob is a Python library that processes matter information. It provides associate API to process (NLP)tasks like part-of-speech tagging, phrase extraction, sentiment analysis, classification, translation, and more. Hence, TextBlob was used for sentiment analysis of the tweets, to understand the polarity of tweets, positive or negative. In Figure No.8, twitter data frame was created where each tweet had unique ids. This data frame consisted of different parameters namely: Length, data, source, sentiment, number of likes/retweets. The number of likes and retweets each day of all the fetched tweets are represented in Figure No.7 as a Time series graph.

 DOI : 10.35291/2454-9150.2021.0116

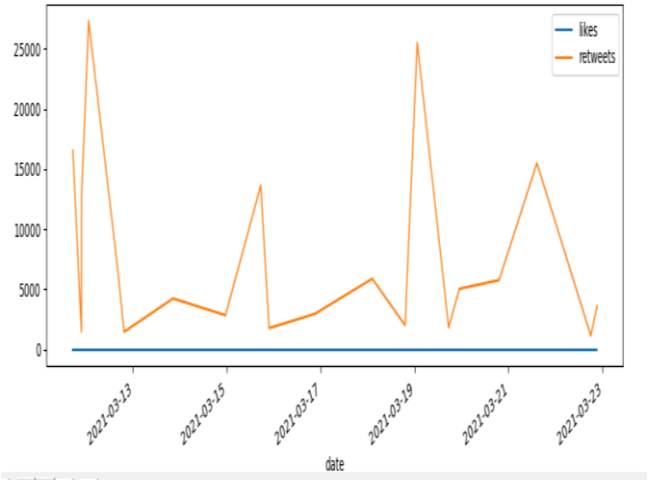
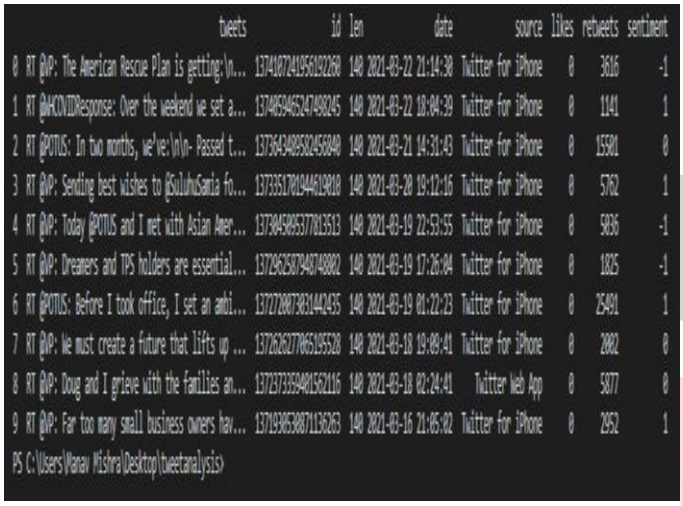


Figure No.7 Twitter Time series graph



	tweets	id	len	date	source	likes	retweets	sentiment
0	RT @VP: The American Rescue Plan is getting:\n...	1374107241956192268	140	2021-03-22 21:14:38	Twitter for iPhone	0	3616	-1
1	RT @WHCOVIDResponse: Over the weekend we set a...	1374059465247498245	140	2021-03-22 18:04:39	Twitter for iPhone	0	1141	1
2	RT @POTUS: In two months, we've:\n\n- Passed t...	1373643489850456840	140	2021-03-21 14:31:43	Twitter for iPhone	0	15981	0
3	RT @VP: Sending best wishes to @SoulmuSanta fo...	1373351701944619010	140	2021-03-20 19:12:16	Twitter for iPhone	0	5762	1
4	RT @VP: Today @POTUS and I met with Asian Amer...	1373045985377013513	140	2021-03-19 22:53:55	Twitter for iPhone	0	5036	-1
5	RT @VP: Dreamers and TPS holders are essential...	1372963587949748802	140	2021-03-19 17:26:04	Twitter for iPhone	0	1835	-1
6	RT @POTUS: Before I took office, I set an ambi...	1372728073831444435	140	2021-03-19 01:22:23	Twitter for iPhone	0	25491	1
7	RT @VP: We must create a future that lifts up ...	1372626277065195528	140	2021-03-18 19:09:41	Twitter for iPhone	0	2002	0
8	RT @VP: Doug and I grieve with the families an...	1372373359401562116	140	2021-03-18 02:24:41	Twitter Web App	0	5877	0
9	RT @VP: Far too many small business owners hav...	1371938638871126263	140	2021-03-16 21:05:02	Twitter for iPhone	0	2952	1

PS C:\Users\Manav Mishra\Desktop\tweetanalysis>

Figure No.8 Twitter Data frame with sentiment

IV. RESULTS

Online Platform	Approach	Final Output	Programming Language	Limitation
Twitter	Machine Learning XGBoost	F1 score 0.67	Python	F1 score can be improved with hyperparameter tuning
Facebook	Deep Learning LSTM	Accuracy 88%	Python	Deep learning requires a very large dataset to accurately train the model.
Reddit	Lexicon based VADER	Negative 40.2% Neutral 38.86% Positive 20.9%	Python	Limited to 3600 API calls/ hr.
Twitter	Lexicon based TextBlob	83%	Python	Limited to 100 API calls/ hr.

Figure No.9 Comparison Of Various Approaches Implemented

V. CONCLUSION & FUTURE SCOPE

Analysing data that's being generated through social media like Twitter, Reddit, Facebook, etc and understanding the behavior of individuals exploiting this opportunity is useful as it will generate informative insights for the business and government organization. The law implementing organizations are finding solutions to detect such illegal posts that are in the form of text for criminal investigation. The analysis carried out in the research can be very useful for this purpose. Twitter sentiment anslysis was conducted on sufficiently large dataset using Machine Learning and the same was done for Facebook dataset using Deep learning techniques. Also lexicon approach was used for Reddit and twitter respectively. Out of all the techniques Deep learning performed the best with 88% accuracy using LSTM but it needs a larger dataset of facebook comments for further training. For the Twitter dataset F1 score of 0.67 was obtained using Extreme gradient boosting algorithm for 20,000 tweets dataset. Machine Learning approach worked well for the twitter dataset whereas lexicon approach doesn't provide as accurate results as the other two also the data that can be fetched from the API is restricted to a smaller number at a time.

REFERENCES

[1] T. Surya Gunawan, N. Aleah Jehan Abdullah, M. Kartiwi and E. Ihsanto, "Social Network Analysis using Python Data Mining," 2020 8th International Conference on Cyber and IT Service Management (CITSM), Pangkal, Indonesia, 2020, pp. 1-6, doi: 10.1109/CITSM50537.2020.9268866.

[2] Y. Chandra and A. Jana, "Sentiment Analysis using Machine Learning and Deep Learning," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 1-4, doi: 10.23919/INDIACom49435.2020.9083703.

[3] Kusrini and M. Mashuri, "Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 2019, pp. 365-368, doi: 10.1109/ICAIIT.2019.8834477.

[4] Zulfadzli Drus, Haliyana Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review", Procedia Computer Science, Volume 161, 2019, Pages 707-714, ISSN 1877-0509

[5] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Volume 5, Issue 4, 2014, Pages 1093-1113, ISSN 2090-4479.

[6] D. Das and P. Sharma, "Algorithm for prediction of negative links using sentiment analysis in social networks," 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, Spain, 2017, pp. 1570-1575.

[7] S. Ahmed and F. Muhammad, "Using Boosting Approaches to Detect Spam Reviews," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, Dhaka, Bangladesh, 2019, pp. 1-6.

 DOI : 10.35291/2454-9150.2021.0116

[8] L. Cheng and S. Tsai, "Deep Learning for Automated Sentiment Analysis of Social Media," *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Vancouver, BC, Canada, 2019, pp. 1001-1004.

[9] Srivastava, Tanya & Mangalagowri, R & Dudala, Shailesh. (2018). MONITORING OF SUSPICIOUS DISCUSSIONS ON ONLINE FORUMS USING DATA MINING. 10.13140/RG.2.2.11235.91680.

[10] K. Glass and R. Colbaugh, "Estimating the sentiment of social media content for security informatics applications," Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, Beijing, China, 2011, pp. 65-70, doi: 10.1109/ISI.2011.5984052.

[11] Emma Haddi, Xiaohui Liu, Yong Shi,"The Role of Text Pre-processing in Sentiment Analysis", Procedia Computer Science, Volume 17, 2013, Pages 26-32, ISSN 1877-0509.

[12] 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, USA, 2016, pp. 568-571, doi: 10.1109/WI.2016.0097.

[13] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[14] Loria, S. (2018). Textblob Documentation. Release 0.15, 2.

 DOI : 10.35291/2454-9150.2021.0116