Sentiment Analysis for Company Recruitment Process Using Twitter Data

Prof. Vidya Raut*, Vanita Amdapure, Sneha Bhoyar, Sakshi Patil, Ashlesha Ghayde

Computer Engineering Department, Cummins College of Engineering for women RTMNU, Nagpur, Maharashtra, India

ABSTRACT

In this system we provide a novel technique to recruit a right person for company by obtaining a behavioral analysis of person from the Twitter stream . We tend to propose a strategy to associate flexible sentiment analysis approach that analyzes twitter posts and extracts opinion and responses of various user's on the various social topics in time period. The catchword is to assist firms apprehend their employees skills and general read beforehand therefore on build their method for hiring process easier. To make a platform for firms that they'll use to examine the generated report concerning the employees that applied for job in their company. The report is generated on basis of overall analysis of employee's interaction on twitter, the opinion and views shared by them, conjointly response to social topics. Victimization this analysis, a report is generated that predicts the temperament of the employee showing their positive and negative interactions and their war social topics within the graphical illustration.

Keywords : Sentiment analysis, Machine Learning, Natural Language Processing, Python , API, Naïve Bayes Classifier Machine Learning Approach, Lexicon based Approach.

I. INTRODUCTION

Twitter is a micro-blogging website, having over 100 million users generating and social networking service on which over 500 million tweets post every day. With such large audience, Twitter has consistently attracted users to convey their opinions and perspective about any issue, brand, company or any other topic of interest. Due to this reason, Twitter is used as an informative source by many organizations, institutions and companies.

On Twitter, users are allowed to share their opinions in the form of tweets, using only 280 characters. People tend to express their opinions about events in the real life and such opinions contain valuable information for market research, brand monitoring and political polls by using slang, abbreviations, emoticons, short forms etc.

In proposed system we used sentiment analysis to extract sentiment from tweets. The results from this is used in our proposed system to analyzing and monitoring changes of sentiment with an event, sentiments regarding a particular brand or release of a particular product, analyzing public view of government policies etc.

We study how to perform sentiment analysis on Twitter data using Python and applied machine learning models on data. This model extract sentiment words from unlabeled Twitter datasets, which are

preliminary filtered using the domain-independent lexicons.

It is difficult to achieve the high quality of sentiment analysis on such short informal texts as tweets because people use a lot of slang, vulgarity and out-of-vocabulary words to state their opinions about various objects and situations. Thus, special methods for processing social media data streams should be developed. We proposed and evaluated our approach for English language. To evaluate the quality of the created lexicons extrinsically, we conduct the experiments on the tweet subjectivity using machine learning techniques with polarity calculation tasks using various lexicons.

The key advantage of the proposed to make hiring process easier for company.

II. LITERATURE SURVEY

In this paper [1] Parinya Sanguansat proposed the sentiment analysis system in Thai language to monitor business brand image via social media. Identified four sentiments (positive, negative, neutral and need). In this paper [2] Nehal Mamgain performing sentiment analysis of people's opinions regarding top colleges in India. Bayes' theorem was used for spelling correction. Comparison between the results using machine learning algorithms: Naïve Bayes and Support Vector Machine and an Artificial Neural Network model. Rincy Jose proposed Sentiment analysis is the computational study of opinions, sentiments, evaluations, attitudes, views and emotions expressed in text. It refers to a classification problem where the main focus is to predict the polarity of words and then classify them into positive or negative sentiment. Sentiment analysis over Twitter offers people a fast and effective way to measure the public's feelings towards their party and politicians.

In this paper [3] use automatically classifies the sentiment of tweets by combining machine learning classifiers with lexicon based classifier. The new combination of classifiers are SentiWordNet classifier, naive bayes classifier and hidden markov model classifier. In this paper [4] Manju Venugopalan proposed work aims at developing a hybrid model for sentiment classification that explores the tweet specific features and uses domain independent and domain specific lexicons to offer a domain oriented approach and hence analyze and extract the consumer sentiment towards popular smart phone brands .

This Paper [5] presents approach for analysing the sentiments of users using data mining classifiers by Anurag P. Jain and Vijay D. Katkar. It also compares the performance of single classifiers for sentiments analysis over ensemble of classifier. An experimental result obtained demonstrates that k-nearest neighbour classifier gives very high predictive accuracy. Dua'a Al-Hajjar and Afraz Z. Syed proposed lexicon-based approach to extracting sentiment and emotion from tweets for digital marketing purposes. In this paper [6] perform sentiment analysis using SentiWordNet while we detect emotions using the NRC Hashtag Emotion Lexicon and compare and combine the scores obtained from the two lexicons into one result per tweet.

In this Paper [7] two visual vocabularies are built from colour compositions and SIFT (scale-invariant feature transform) descriptors. Thereafter, the pLSA (probabilistic latent semantic analysis)-learning is employed to predict the human sentiment hidden in social images from visual words. The proposed system was evaluated to the images collected from Photo.net and Google and 15 Kobayashi's sentiments were considered to label the images.

In this Paper [8] F. Ciullo proposed a brief overview of some concrete examples of applying sentiment analysis to social networks for healthcare purposes, present the current type of tools existing for sentiment analysis, and summarize the challenges involved in

this process focusing on the role of high performance computing.

III. METHOD FOR SENTIMENT ANYALASIS

Sentiments are the words or sentences that represent view or opinion that is held or expressed that can be positive, negative or neutral. We are going to propose a novel hybrid approach involving both corpus-based and dictionary-based techniques, which will find the semantic orientation of the sentiments words in tweets. We will also consider features like emoticons, neutralization, negation handling and capitalization as they have recently become a huge part of the internet language.

The proposed Sentiment Analysis on twitter data is based on two important parts viz Data Extraction, pre-processing of extracted data and classification.

To uncover the sentiments, we will first extract the data from the twitter topic is highly disorganized and contains different types of emojis, stop words and is not specific to a language hence automation involved in analysis of data is the best approach for text classification.

The proposed system ,system architecture is given below, with its various modules.
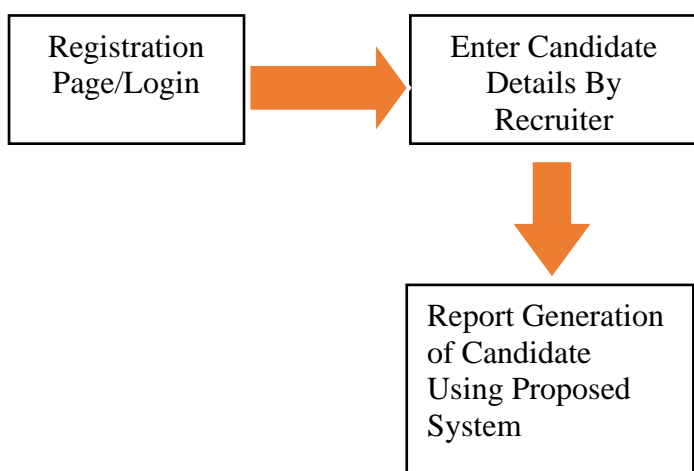
Figure 1 : Flow Diagram of Company Recruitment Process

A. Registration Page/Login

This is the login page wherein the new candidate has to register and already registered candidate can login directly. All the information is stored in database. Here we have provided two types of login that is the user/candidate login and the admin/company login.

Here Company will provide user name to proposed system to search his/her social behavioural details. Here proposed system will generate and provide a report of respective candidate in the form percentage and Bar graph.

And this report can be refer by company for Employee promotion as well as to define Employee personality.
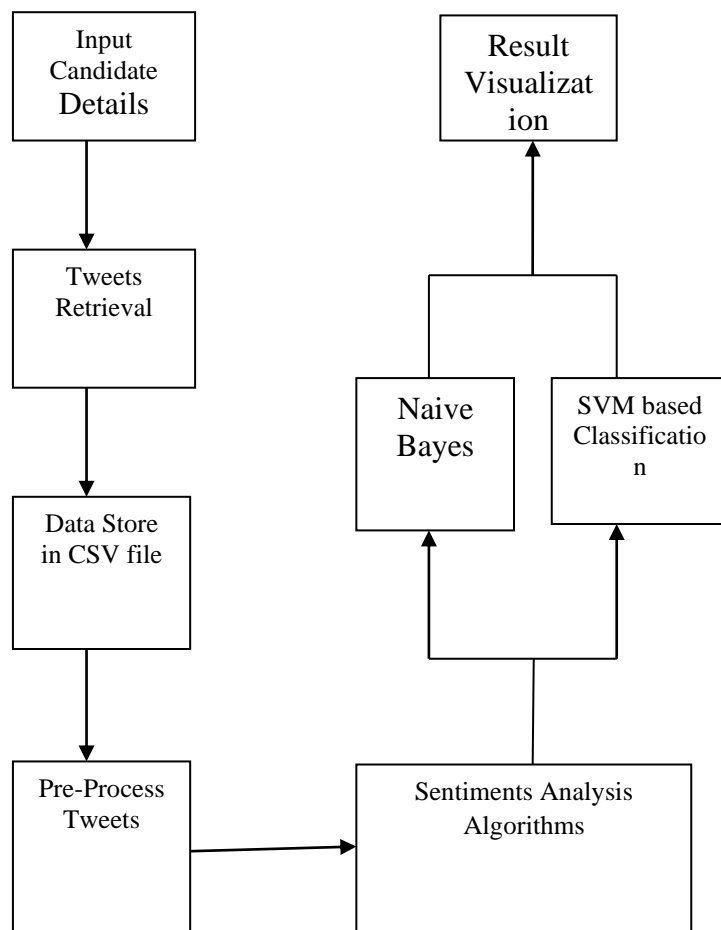

Figure 2 : Flow of Proposed System

B. User Details

In this phase the user can be asked to fill in some required information for further process.

1) Data Gathering: To gather public opinion based on collected tweets related to views about interests including Twitter top trends, we used Tweepy API. We have created an account on Tweepy API linked to our Twitter account. To retrieve the tweets, Tweepy API accepts parameters and provides the Twitter account's data in return.
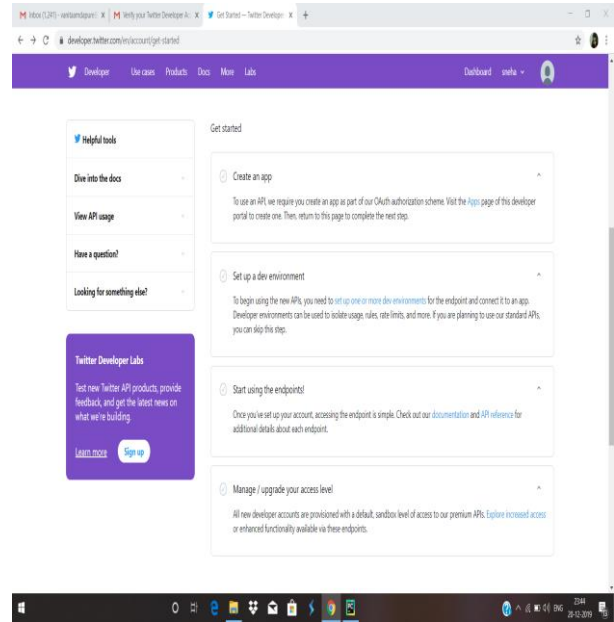
Collecting Data from twitter with Help of Twitter API
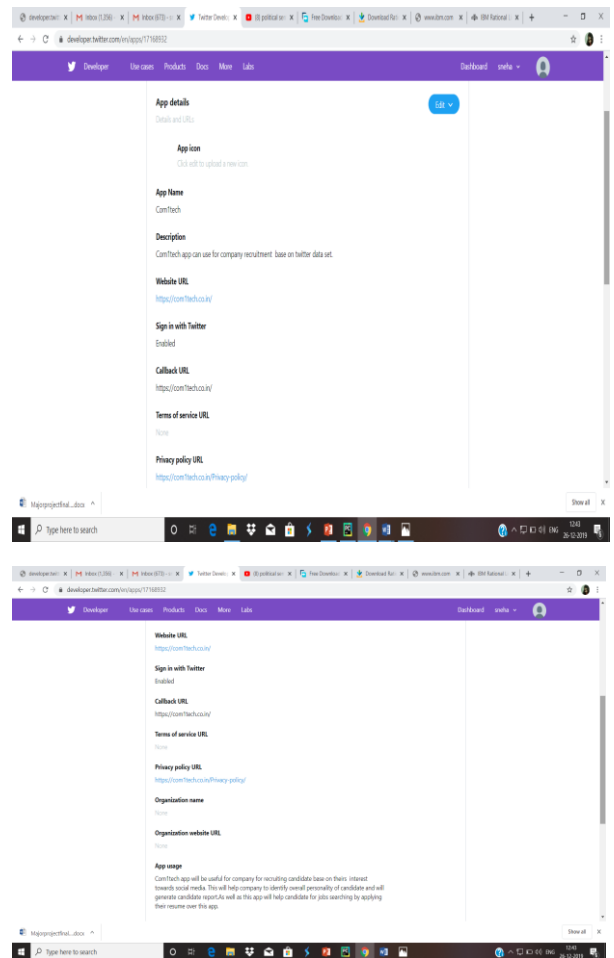
Step 1: Confirmation Message from Twitter
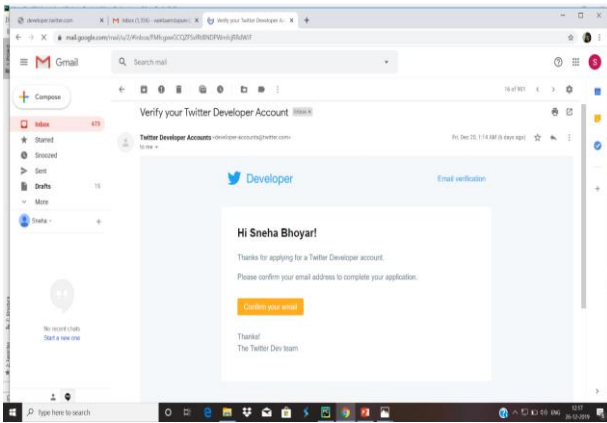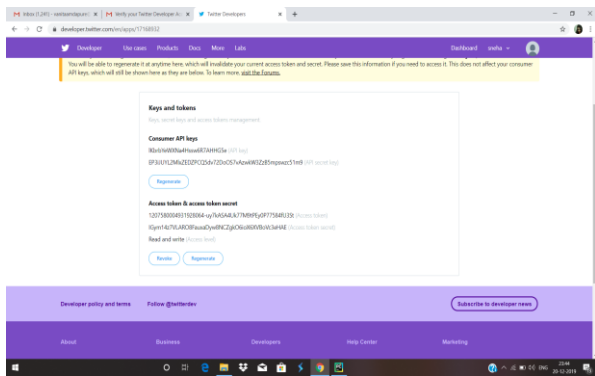

Step 2: Website Of Twitter Developer
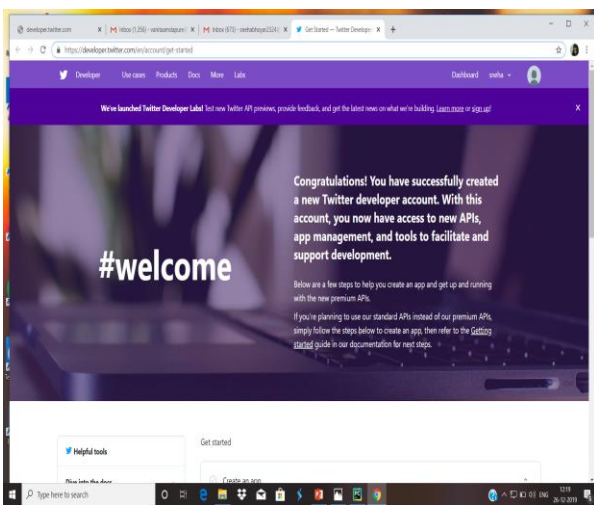
Step 3: Create App For Twitter API

Step 4: Enter App Details

Step 5: Got Access of Twitter API

2) Data Pre-processing : In the pre-processing step, we removed the irrelevant Twitter data. First, we just kept tweets that were in English. We then remove all the hashtags and other special characters and only save the data in text format. Retrieved tweets, from Twitter accounts, were saved in the database under the following fields: twitter_id, tweet_created, user_id, screen_name, tweet_text, retweet_count, follower_count, and favourite_count of each tweet. The gathered tweets numbered 100,00.

3) Polarity Calculation and Sentiment Analysis : Sentiment analysis can provide valuable insights from social media platforms by detecting emotions or opinions from a large volume of data present in unstructured format. Sentiment analysis includes three polarity classes, which are negative, neutral and positive. The polarity of each tweet is determined by assigning a score from –1 to 1 based on the words used, where a negative score means a negative sentiment and a positive score means a positive sentiment while the zero value is considered a neutral sentiment. A score of subjectivity assigned to each tweet is based on whether it is representing a subjective meaning or an objective meaning; the range of subjectivity score is also from 0 to 1 where a value near to 0 represents objective and near to 1 subjective.

For detecting the polarity and subjectivity of tweet reviews, and to give a clear view of the most accurate analyser for the polarity and subjectivity calculator, we used Textblob, SentiWordNet and Word Sense Disambiguation (WSD) sentiment analyzers. Textblob comes with the basic features of natural-language processing essentials. We used this analyzer for the polarity and subjectivity calculation of tweets. Similarly, we used SentiWordNet, which is a publicly available analyzer of the English language that contains opinions extracted from a wordnet database. In addition to that, W-WSD has the ability to detect the correct word sense within a specified context. Baseline words/Unigrams are clear representatives for the calculation of polarity.

C. Experimental Observation

When we calculated the polarity and subjectivity of the Twitter dataset, the resultant files were in comma separated value (csv) format. To make them compatible for validation, we converted the csv files into attribute-relation file format (arff). For model building, we applied supervised machine-learning algorithms, Naïve Bayes on the training dataset, and percentage split for SVM. Steps of the analyzers' validation through Naïve Bayes and SVM validation are often viewed from the pseudocode "Naive Bayes Classifier" and "SVM Classifier". These machine-learning algorithms (Naïve Bayes and SVM) were applied on the training set to create an analysis model. On the idea of the model constructed for every analyzer, the test set was evaluated. After test set evaluation, we recorded the accuracy of every analyzer under each model.

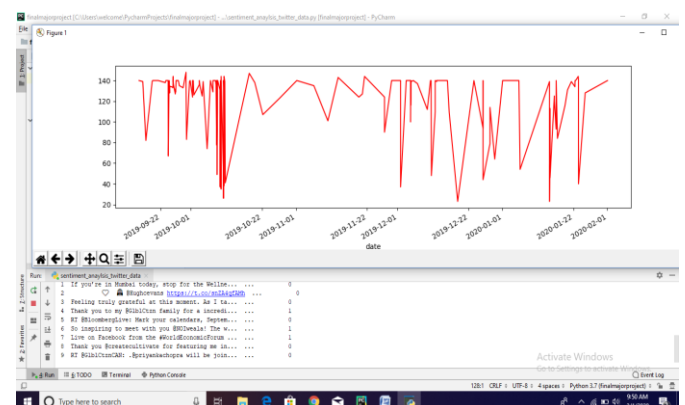The outputs and graphs for the user twitter:

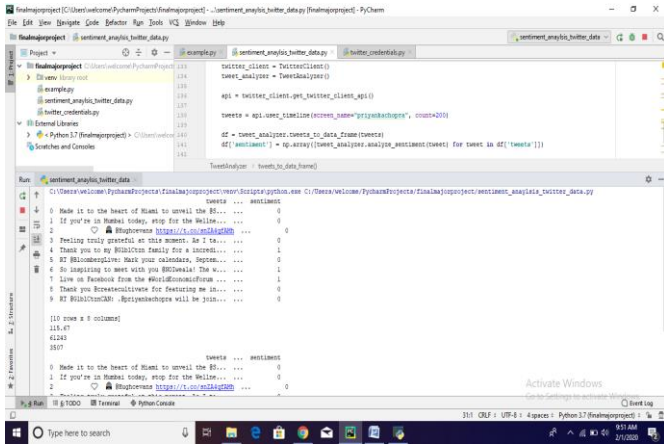Figure 3: Length v/s Like v/s Retweets

Figure 4: The tweets and its polarity of sentiment analysis

Results are display using chart, Bar, graph etc. and performance tuning is done accordingly .In this proposed system we generate candidate report in the form of Bar graph.

Table below shows the number of positive , negative and neutral tweets in percentage .

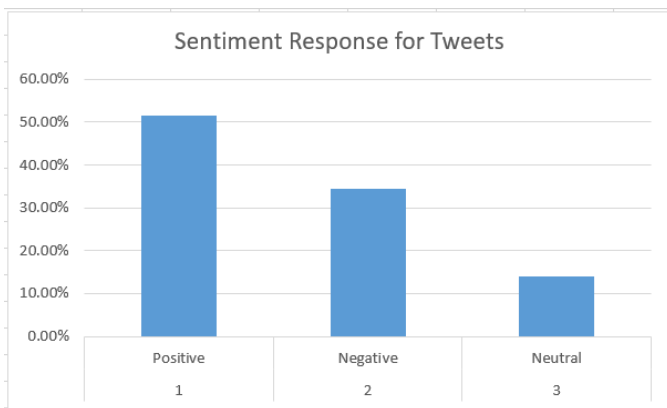|	Tweets	Percentage
1	Positive	51.50%
2	Negative	34.50%
3	Neutral	14.00%

Table 01: Positive v/s Negative v/s Neutral tweets

Figure 5 : Positive v/s Negative v/s Neutral

IV. CONCLUSION

The increase of various social platforms of twitter where people can use short messages that comes under the category of text and opinion mining .It helps us to create technologies which can help in analyzing for sentiments of the tweets and feeding the data to a machine learning model in order to train it and then check its accuracy. So that we can use this model for recruit an employee for company according to the results. It comprises of steps like data collection, text pre-processing, sentiment detection, sentiment classification, training and testing the model.

In this paper we use the sentiment classification approaches can be classified in (i) Machine learning (ii) Lexicon based .The machine learning approach is used for predicting the polarity of sentiments. In which, we are using algorithms such as Naive Bayes, Support Vector Machine to increase performance. While the lexicon based approach uses a predefined list of words, where each word is associated with a specific sentiment. Text classification is one of the most important areas in text data mining, and sentiment analysis can be performed based on text classification.

Along with the sentiments Classification of the tweets we are also able to extract various characteristics of the tweets e.g. Likes, retweets. The most liked tweet and the number of times it is retweeted. The process thus defined is exploratory and we can improve it by using better approaches and algorithms.

V. REFERENCES

[1]. Parinya Sanguansat," Paragraph2Vec-Based Sentiment Analysis on Social Media for Business in Thailand", 2016, IEEE, 348765342-43565-65765-555, vol. 9, pp. 231-236

[2]. Nehal Mamgain, Ekta Mehta, Ankush Mittal, Gaurav Bhatt," Sentiment Analysis of Top Colleges in India Using Twitter Data", 2016, IEEE, 978-1-5090-0082-1, vol.14, pp. 67-72

[3]. Rincy Jose, Varghese S Chooralil," Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense

Disambiguation", 2015, IEEE, 978-1-4673-7349-4, vol. 4, pp. 123-128

[4]. Manju Venugopalan, Deepa Gupta," Exploring Sentiment Analysis on Twitter Data", 2015, IEEE, 978-1-4673-7948-9 vol.6, pp. 34-39

[5]. Anurag P. Jain, Mr. Vijay D. Katkar," Sentiments Analysis Of Twitter Data Using Data Mining", 2015 International Conference on Information Processing (ICIP), 978-1-4673-7758-4, vol. 19, pp. 432-438

[6]. Dua'a Al-Hajjar, Afraz Z. Syed," Applying Sentiment and Emotion Analysis on Brand Tweets for Digital Marketing", 2015, IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), vol. 7, pp. 153-159

[7]. Eunjeong Ko and Eun Yi Kim," Recognizing the Sentiments of Web Images using Hand-designed Features", 2015, IEEE, 918-1-4613-1290-9, vol. 14, pp. 784-788

[8]. F. Ciullo, C. Zucco, B. Calabrese, G. Agapito, P. H. Guzzi, M. Cannataro," Computational Challenges for Sentiment Analysis in Life Sciences", 2016, IEEE, 978-1-5090-2088, vol. 10, pp. 478-482

[9]. Eunjeong Ko, Chanhee Yoon, Eun Yi Kim," Discovering Visual Features for Recognizing User's Sentiments in Social Images", 2016, IEEE, 978-1-4673-8796, vol. 42, pp. 773-779

[10]. Dang, Y.; Zhang, Y.; Chen, H. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. IEEE Intell. Syst. 2010, 25, 46–53. [CrossRef]

[11]. Cambria, E. Affective computing and sentiment analysis. IEEE Intell. Syst. 2016, 31, 102–107. [CrossRef]

[12]. Jagdale, O.; Harmalkar, V.; Chavan, S.; Sharma, N. Twitter mining using R. Int. J. Eng. Res. Adv. Tech. 2017, 3, 252–256.

[13]. Anjaria, M.; Guddeti, R.M.R. Influence factor based opinion mining of twitter data using supervised learning. In Proceedings of the 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 6–10 January 2014; pp. 1–8.

[14]. Dubey, G.; Chawla, S.; Kaur, K. Social media opinion analysis for indian political diplomats. In Proceedings of the 2017 7th International Conference on Cloud Computing, Data Science & Engineering, Noida, India, 12–13 January 2017; pp. 681–686.