ANALYZING TWITTER SENTIMENTS WITH BIGDATA

Ayesha Azeeza
Dept. Of MTech CSE
SSIT, Tumkuru

Dr. H Venugopal
Professor, Dept. Of MTech CSE
SSIT, Tumkuru

***Abstract*— Sentimental analysis is the process of defining the opinion of the individual by extracting the data. Sentimental analysis is done by collecting the data from variety of resources like internet and social media platforms like (Facebook, twitter) and it stored in different formats. The data extracted from these social media platforms can be used for different purposes like prediction, marketing and sentiment analysis. Politicians use sentimental analysis to know how people feel about themselves and their parties. Twitter is a widely used platform for posting comments through short messages. Every year millions or billions of tweets extracted are subjected to sentiment analysis. But handling such a huge amount of unstructured or structured data is a tedious task to take by the traditional storage system (such as RDBMS, Relational Database System). Therefore, we need a tool that can process and analyze this unrelated data. So; we have utilized big data technology to handle different formats of unstructured and structured data from different sources.**

Keywords—*component, formatting, style, styling, insert* (key words)

I. *Introduction of Big Data*

Big Data is the popular buzzword in the IT Industry. Apache's Hadoop is a leading Big Data idea used by IT giants (such as Yahoo, Face book & Google). 'Big Data' is also a **data** but with a huge size. 'Big Data' is a term used to express collection of data that is huge in size and till now data is increasing exponentially with time. In short, Big Data is a data which is so big and complex that none of the traditional data management tools are capable to store it or process it efficiently.

1.1 Examples of Big Data

Followings are the some example of Big Data

The New York Stock Exchange generates about 1TB of latest trade data per day.

1. Statistic shows that *500TB* of new data get inserted into the databases of social media site like Face book, Twitter, every day. This data is mainly produced in terms of photo and video uploads, exchanging message, putting comments etc.
2. Particular Jet engine can generate *10+terabytes* of data in *30 minutes* of a flight time with many thousand flights for every day, production of data reach up to many *Petabytes.*

1.1.1 Categories of Big Data

Big data is categorized in three forms:

1. Structured Data
2. Unstructured Data
3. Semi-structured Data

1. *Structured Data :*Every data that can be stored, accessed and processed in the form of unchanging format is termed as a 'structured' data. More than the era of time, talent in computer science have achieved bigger success in evolving techniques for functioning with such kind of data (where the format is well famous in advance) and also deriving worth out of it. Still nowadays, we are foreseeing issues as soon as size of such data grows to huge level usual sizes are being in the rage of multiple zettabyte. Examples of Structured Data**:** A table in a database is an example.

2. *Unstructured Data:* Every data with unfamiliar form or the structured form is classified as unstructured data. In adding together to the size being huge, un-structured data pose numerous

3. *Semi-structured Data***:** Combination of structured and unstructured form of data is classified as Semi-structured Data. It's a form of data in relational DBMS which gives table description is defined properly in the ordered form. Example of Semi-structured data is data within the XML file.

1.1.2 Characteristics of Big Data

- Volume
- Velocity
- Variety
- Variability
- Value

1.2 Sentiment Analysis

To know whether the slice of text is positive, negative or neutral is determined by using the concept of Sentiment Analysis. It's used to get the opinion or attitude of a speaker it's similar as opinion mining. To discover what's the feeling of a people on a

particular topic is identified by this sentimental analysis technology. If we want to know the feedback of an food Masala Dose by twitter whether the food is good or bad. We can get the answer for this question by Twitter sentiment analysis. By extracting the words that shows why the food is liked or didn't liked by people we come to know why people had given the opinion that whether the food is good or bad. From this we come to know why the people are happy and not happy. This is one of the way to conduct a on the particular interested topics without using huge amount of budget and also man power to conduct a survey.

1.3 Advantages of Sentiment Analysis

- In the business purposes, it provides a good benefit.
- It helps the people to identify about products good or bad feature they want to buy.
- It helps the organization or company to know about their product limitations or bad features to improve.
- It is beneficial for the competitive party so that they can know the weakness of the opposition party.
- It decreases the effort of human being to evaluate the product as good or bad.
- It is beneficial for social media analysis.

II. LITERATURE SURVEY

In the previous chapter we have discussed about the sentimental analysis and its type. There are different methods to perform sentimental analysis and different methods will provide different accuracy. Sentiment analysis is very beneficial for customers and for business organization and for different political parties and also for the individual persons of the party. It also offers the reason of success or failure of a particular entities in their aspect. For example, "SAMSUNG does not have better camera quality". By taking the reason of disliking the product, it helps to business association to know negative points of their product and what are the precaution can be taken to improve the product. But to perform sentimental analysis by computer it's very challenging because the reviews are not directly understood by the computer. Because the review contains noise those should be eliminated before performing the sentimental analysis.

Early Lee and Pang [1] used different types of methods to recognize the polarity of product reviews and movie reviews respectively. They operated at document level. By using the support vector machine and naive Bayes as classifiers they did the sentimental analysis as positive and negative. But these classifiers are not directly applied on the collected dataset
from the reviews because reviews are in unstructured format and also reviews contains noisy data which are not necessary for the sentiment analysis first we have to do the preprocessing of the data later feature extraction is done. In these feature extraction, necessary data is extracted otherwise if we directly apply these classifier methods on dataset then we will not get accurate results. Relations which show the opinions of the review holders are considered as important for sentiment analysis.

There are numerous approaches for opinion mining and accordingly delivers different outcomes. Many authors provided the different methods for sentiment analysis by using different feature extraction method. A lot of work has been done in sentiment analysis by using different methods.

2.1 Supervised Learning Method

These methods use the training and testing set for classification of the reviews. From the training set it will uses the input vector and the respective class labels as input. Training set will be order the input feature vectors into individual class labels. Then finally a test set tests the model by producing the class labels from feature vectors which are not yet seen. Several machine learning methods like Support Vector Machine (SVM), Maximum Entropy (ME) and NB (Naive Bayes) are beneficial for the classification of reviews [2].For the opinion mining some features are used those are negation, Term frequency, Term presence, Part-of-Speech and n-grams [3]. These features govern polarity of single words, sentences and also the documents. Polarity may be of negative, positive or neutral.

The finest classification technique for the text is SVM (Support vector machines)
According to the computational learning technique the statistical classification method called support vector machine and it is also considered as discriminative classifier [4]. SVM also uses the principal of structural risk minimization. Support vector machine also discovers the decision surfaces which help to categorize the training set into two different classes as negative and positive. Among these alternatives Multi class support vector machine is mostly used for opinion mining. On the source of support vectors conclusions are made. Support vectors are the active data points between the training set.

Cui [6] projected that when the size of the training data is very small for the sentimental analysis naïve Bayes classifier will gives the appropriate outcome when compared to SVM because SVM will provide the better outcome when the review is consisting of both positive and negative terms because Support vector machine needs a large set of training data set.
Domingo's [6] determined that Naïve Bayes gives improved outcome. Naive Bayes offers effectual outcomes in case of certain complications. Naive Bayes takes basic statement that features must be independent to each other.

Zhen Niu [6] projected a new model which includes effectual approaches for totaling the weight, classification and feature extraction is applied to increase the effectiveness of Naive Bayes. In his Bayesian algorithm is used for this new model using the term frequency weight is calculated. In this, information which represents a class is known as Representative feature because exclusive feature and illustrative features are used to alter the weights of the classifiers. One more classes is known as Unique feature because information which helps in distinguishing on the basis of weights, likelihood for specific classification is calculated and this feature expands the Bayesian algorithm. Pak [8] created a huge amount of twitter data by gathering the tweets automatically.

He collected the tweets with the help of API (Application Programming Interface) then he started to analyze them by using emoticons. Naive Bayes classifier uses POS-tags and N-gram features. In this method, training data set considered as only on the basis of polarization of emoticons so there is high probability of error because sentiments of tweets. Training set is

considering only those tweets which have emoticons this the reason for less efficiency.
Xia et al. [9] planned sentiment classification by using ensemble architecture.

combination of classification methods and feature sets is called as ensemble framework. In this approach, to create an ensemble framework they used three base classifiers and two kinds of feature sets. By using Word relations and Part of speech information two forms of feature sets are generated. The three base classifiers are selected as Maximum Entropy, Naive Bayes and Support Vector Machine. Opinion mining and accuracy was better as compared to previous one because they used dissimilar ensemble techniques like weighted combination, fixed combination.

Neethu and Raajsree [10] produces group classifier in their research. Initially preprocessing is done later feature extraction was done. First, twitter specific features are retrieved. To analyze the sentiments in the tweets hashtags and emoticons are considered. weights are assigned to difference Emoticons can be positive or negative. For the Emoticons which are positive are assigned a weight of "1" and the negative ones are assigned"-1". Here might be negative as well as positive hashtags. Therefore, hashtags are taken as two different features in the feature vector based on the number of positive negative.

Therefore, there are 8 relevant features which feature vector. It includes speech (tag) and the occurrence of the negation, sum of positive keywords and negative keywords,19 emoticons, amount of positive, negative hash tags are the features. Naive Bayes, and Maximum entropy and SVM are the base classifiers used .At this time, a passed by vote rule is cast-off to produce a united classifier. The classifier purpose to do classification on the basis of output. By using Twitter API, tweets of product related gathered. Electronic products related dataset is made by assembling 1200 twitter posts. Datasets divided such that there are 200 in the test set and 1000 tweets in the training set. From the collected tweets, they used Stanford postagger1 for recovering a POS tag. Subjective and objective tweets can be analyzed together since a product domain is selected. Equally potentials pay similarly to recognize the product's quality.

This shows that sentiment analysis affects domain information. 3 types of basic classifiers are SVM, Nave Bayes, Maximum Entropy used by cooperative classifier for the resolve of sentiments of classification. Similar performance is given by these classifiers. Naive Bayes has improved precision but accuracy and recall is lower as compared to others, other classifiers like SVM, Maximum Entropy Classifier and Ensemble classifiers like in these evaluation metrics. whereas Naive Bayes has the accuracy of 89.6% While their accuracy is 90%. product domain chosen quality feature vector. sentiment analysis in spite of the classifier used this feature vector helps in improved analysis. for sentiment analysis it was all about the supervised methods but lexical based techniques are also available. The lexical based techniques are also known as symbolic techniques.

2.2 Lexical Based Method

There are various lexical resources available for sentiment analysis**.** In lexicon based unsupervised learning method, there is an impression of sentiment dictionary. For sentiment analysis Senti word net is one of the lexical resources. for each synset of WordNet it provides the three-numerical sentiment score as negative, positive and neutral. Synset means synonyms of the words or words which are related to each other like automobile and car.
Different scores are given so in the state where similar term has dissimilar senses then it will give different sentiment scores. So, it offers synset based representation means for each synset. So, to get the most precise sense of the word, it needs to be coupled with WSD (Word Sense l Disambiguation) algorithm. Kamps et al. [11] sentiment analysis performed by using WordNet as lexical database [12]. from a review with different dimensions it is used to discovery the sentiment words. To find out the polarity of the adjectives they formed a distance metric on the basis of WordNet. The words which are related to each other by synonym relations are extracted by lexical resource WordNet database.

Turney [13] did sentiment analysis by consuming sack of words. polarity of each term or word in a document is calculated. To analyze the overall polarity of all the sentiments some aggregation functions are applied on these values, He resolute polarity of an opinion on the basis of average polarity of words which are extracted from the opinions. Those words are considered which are adverbs or adjectives. It is used for sentiment analysis at conceptual level using lexical resource. It includes roughly composite or progressive ideas like to complete the goal and to celebrate special occasion therefore it is different from all debated resources. Now, sentic net allocates the sentiment scores between the range -1 and 1. Approximate 14,000 common sense concepts are assigned to these scores. On the origin of sixteen basic emotions Sentiment score is assigned to each term. These sentiments are defined in a model known as hourglass of emotions.

Yun quing xia [14] For contextual concept polarity disambiguation sentic net and Bayesian model are used. Mauro [16] to extract key concepts from a sentence he proposed a merged framework that merges wordnet, concept net and sentic net. A classification was proposed [16] that generate their own sentiment analysis framework. They combine sentic net, sentiword net and other sentiment analysis method in their framework. Another projected work [16] features for subjectivity detection and other sentiment analysis task they used senti net to abstract sack of concepts and polarity. Jay kuan [18] to acquire more thoughts and polarity scores used sentic net concepts as seeds by giving a concept of random walk. Other have projected the combined practice of information bases and machine learning for twitter sentiment analysis, [19] classification done by using short text message [20] and also frame based opinion mining [21].

III PROPOSED SYSTEM

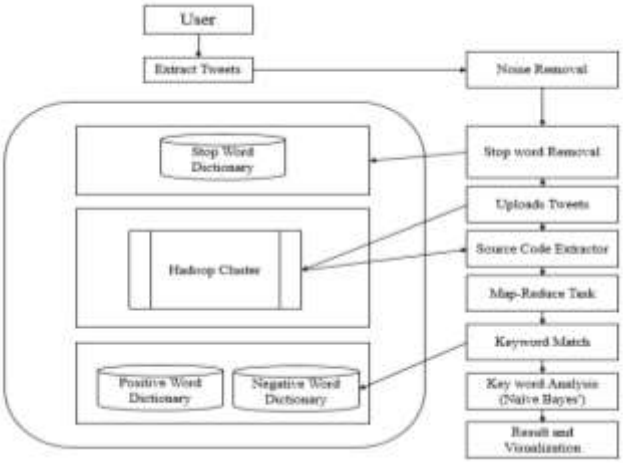
The main objectives of the proposed system focus on the analyzing sentiments of the tweets extracted from the Narendra modi's public twitter account. In collected tweets they are discussed about different topics like politics, or some particular party, china war, sports. Here our main concentration on some main categories of tweets like politics, war, sports, technology, entertainment.

- Tweets Extraction
  - First create the twitter developer account
  - Confirm the Twitter Account and generate consumer and authentication key
  - Provide Permission like Read or Read/Write to access the Account.
  - Extracted tweets are stored in a document file format.
- Clustering Process
  - In Big Data with Hadoop platform we are going to use HDFS cluster.
  - The document file containing tweets is pushed into HDFS cluster using the command
  - $hadoop fs -put /user/cloud era/directory name/Filename.
- Map-Reduce Task
  - Preprocessing of tweets is done, later we are going to perform the map-reduce task based on the particular topic chosen.
- Classification

Using the Naïve Bayes classification algorithm analysis of the percentage of the negative and positive tweets of the particular topic is determined.

IV SYSTEM DESIGN and FRAMEWORK

4.1 System Model



Analyzing Twitter sentiments with Big Data works in two phases. First phase is creating the twitter developer account and extracting the tweets by giving the authentication and consumer key and later extract the tweets and remove the noise like white space, null, special characters and stop word from the tweets by comparing the words with keywords present in the stop word dictionary list which are created by using the English grammar. Second phase is the most important and main phase in this project first the tweets extracted into the hadoop cluster and later the tweets are read by line by line and the source code extractor is formed. Map task is performed and the set of words are outputted with the value, from this task and in the reduce task these set of words are given as input after the reduce task we are going to get the similar words with the counts. These words are matched with the key words lists in the positive and negative word dictionaries. The dictionaries are formed by taking the positive and negative words from the English grammar. Later Naïve Bayes algorithm is applied to get the probability of positive and negative tweets feedback. The proposed system model consists of the following features: Tweets Extraction, Tweets Loader, Source Code Extractor, Noise Removal, Stop Word Cleaning, Positive Word Cleaning, Negative Word Cleaning, Map-Task, Reduce-Task, Keyword Matching, Sentimental Analysis.

Give the document file containing tweets as input and remove the noise and stop word later load the file into hadoop cluster and perform the map-reduce task and compare the set of similar words with the positive and negative dictionary and find the probability of the positive words and negative words based on the topics like politics, war, entertainment, sports, technology.

4. Technologies

Software Requirements

Operating System: Window 6, window 8, window 10.
Language : Java
IDE : Net Beans 6.3.1 and Eclipse,
Cross-platform : Oracle VM Virtual Box
Cloudera : Apache Hadoop based software

Hardware Requirements

Processor	:	Dual core
Speed	:	1.1 GHz
RAM	:	8 GB
Hard disk	:	40 GB

4.2.1 Apache Hadoop ecosystem

For parallel, distributed data processing over several clusters an open source framework was developed that is called as Apache Hadoop Architecture. This architecture design helps to scale efficiently from pair of servers to hundreds by means of additional commodity hardware.
Hence, for computing given dataset the processing time is condensed by assigning the jobs into
different several nodes in the cluster and also distributed storage is offered by this platform.
Within the Apache ecosystem there are many technologies allowed by this computational capabilities. Key technologies that are essential of Apache Hadoop are Hadoop Distributed File System (HDFS), Hadoop YARN and Hadoop MapReduce.

4.2.2 HDFS

HDFS system is developed with the skills to identify and overcome the errors in the data processing process and avoid hardware failure, which might links to electricity blackout also. The huge data which we going to get from the social media like twitter it's usually of more than terabytes volume this data to be stored and processed using distributed feature but traditional system lack this feature. But HDFS to overcome the problem of data loss it offers consistent data storage and also provides data duplication. HDFS permits to store data in different formats like structured or semi-structured or unstructured format. Within HDFS name space there will be a checksum algorithm to prove the data integrity in secured way. In a distributed file system the data will be stored in a block of fixed size. Each block size will be of 64 MB,within the cluster on different nodes the data blocks will be store. The location and the type of the data stored and all keep tracked by the Name Node.

4.2.3 Apache MapReduce

To attain parallel batch processing of huge data over cluster is done by a framework called Apache MapReduce. Algorithm implementation is done but the other development within Apache Hadoop ecosystem is done by using different tools such as (Apache Hive or Apache Pig) to accomplish the similar goal. Benefit of MapReduce is primary data set is further divided into blocks because dataset could go closer to data for the computation. It needs applying two functions to continue further progress with MapReduce. The map function will start to read the input dataset and later it start to process the dataset and finally it yields key and value pairs later sorting process will start it will yields output file for reduce function by sorting creating a key pairs. MapReduce architecture stands mainly established on a single master node called Job Tracker, MapReduce tasks job scheduling and monitoring and keeping track of the failed task and also rescheduling of the task all these tasks are done by Job Tracker. All master tasks are done by the Task Tracker which is included in this architecture. MapReduce task one output is can be given input for other task and formerly the ultimate outcomes are warehoused on distributed file system.

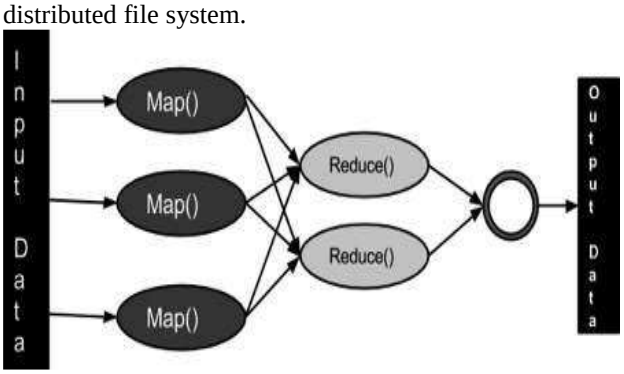


Fig 4.2: Map-Reduce Methodology

4.3 Algorithm
4.3.1 Naïve Bayes Theorem
Naïve Bayes algorithm is used to predict the result of unlabeled data and it's based on the probability of Bayes theorem. Bayes theorem accepts the individuality between predictors (features) and class using the classification technique. There are several models in Naïve Bayes models they are further they are divided based on the feature they will handle. Boolean feature vector is used by the Bernoulli model it defines that each model should have binary variables it doesn't bother about the occurrence of the individual word in the document. In Multinomial model it gives attentions to counts as features, thus in the document word frequencies matters than the variables. To classify the dataset with high volume is easy to implement because it works efficiently by using Naïve Bayes classification algorithm.

$$P(c \mid x) = [P(x \mid c) * P(c)] / P(x)$$

- $P(c \mid x)$ – Posterior Probability
- $P(x \mid c)$ – Likelihood
- $P(c)$ – Class Prior Probability
- $P(x)$ – Predictor Prior Probability

$$P(c \mid X) = P(x1 \mid c) \times P(x2 \mid c) \times \ldots \times P(xn \mid c) \times P(c)$$

For the given information probability of outcome is calculated by Posterior probability. In P (c | x), c denotes class then x is predictor. Within the class Likelihood probability of predictor is the class and class and predictor are belongs to other two probabilistic values. Directive to know the working of Naïve Bayes theorem, to show probability prediction of a tweets occurrence based on the topics War(W), Entertainment(E), politics(P), Technology(T), Sport(S) here is an example.

4.3.2 Naïve Bayes Twitter Classification
Training set will consist of topic and sentiment, which will indicate whether an event occurs. An event will occur, variable indicate Y (Yes), otherwise N(No) as shown in Fig 4.3.

Topics	P	W	P	S	P	E	P	E	E	T	T	W	T	T	P	S	W	S	W
Sentiment	Y	Y	Y	N	N	Y	Y	N	Y	Y	Y	N	N	Y	Y	Y	N	N	Y

Fig 4.3: Training Set

Frequency distribution figure Fig 4.4 shows number of event occurrences by particular topics. Moreover, the total sum for each possible event outcome is calculated.

Topics	Yes	No
Politics	4	1
War	2	2
Sports	2	2
Technology	3	1
Entertainment	2	2
Sum	13	8

Fig 4.4: Frequency distribution

Probability for each row and column is calculated. Total sum of topics occurrences is 21, thus probability for the topic Politics is as shown in Fig 4.5:

P (Politics) = [Y (4) + N (1)] / 21=~ 0.23 %. Hence, the probability for the topic Politics is 23%.

Topics	Yes	No	Probability
Politics	4	1	~0.23
War	2	2	~0.19
Sports	2	2	~0.19
Technology	3	1	~0.19
Entertainment	2	2	~0.19
Sum	13	8	Sum=21
Probability	~0.61	~0.38	

Fig 4.5: Probabilities calculations

With Naïve Bayes theorem, we can now calculate posterior probability that event will occur on the topic Politics:

P (Politics | Yes) = 4 / 13 = ~ 0.30

P (Yes | Politics) = P (Politics | Yes) * P (Yes) / P (Politics)

P (Yes | Politics) = 0.30 * 0.61 / 0.23 = ~ 0.7956

Calculation results show that probability is 79.56% for event to occur on the topic politics.

As Naïve Bayes is machine-learning technique, be applied for Twitter and sentiment analyses.

V RESULT and ANALYSIS

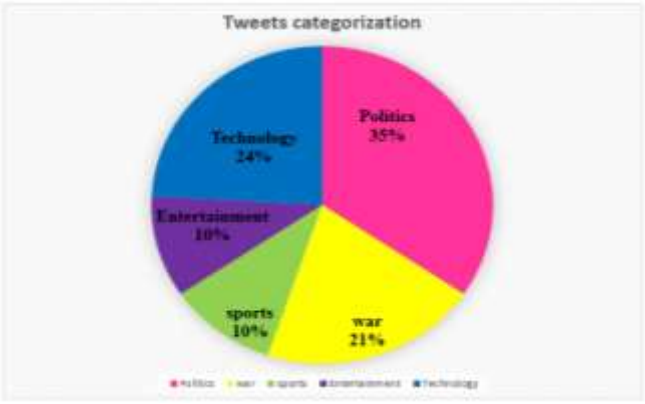


Fig 5.1: Pie chart graph visualization of different topics in tweets categorization with positive scores

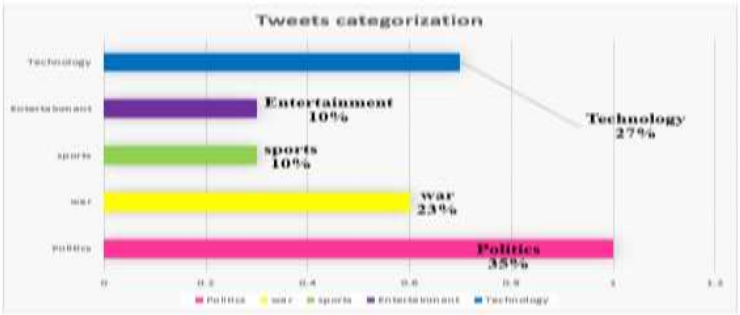


Fig 5.2: Bar chart graph visualization of different topics in tweets categorization with positive scores

VI CONCLUSION and FUTURE WORK

This paper has discovered the impressions of big data and sentiment analysis. The ideas are permitted to know more issues and challenges in the field of sentiment analysis on big data for the further research. Usually, the machine learning methods are very vital to classify the text or reviews from an uncertain data. However, the concepts are supportive to the researcher and whoever newfangled to the area of applications, methods and procedures in Sentiment Analysis. The complete works covers analytics of big data, quantity of big data, matters, approaches, methods for predict the accuracy with a assistance of valuation metrics/statistical analysis and ecosystem of the sentiment analysis on big data.

Till now only analyzing sentiments of a Narendra modi's twitter account has been done in future work we can perform sentimental analysis of a different party members and find the positive and negative feedbacks of each members and compare the analysis of all the members to get to know who has best positive feedback when compared to others.

ACKNOWLEDGEMENT

Firstly, I express my gratitude to my institution **Sri Siddhartha Institute of Technology**, Tumkuru. I express my deep and sincere gratitude to **Department of Computer Science and**

Engineering which provided me an opportunity to fulfill the desire of reaching the goal.
I like to extend my sincere thanks to **Dr. M K Veeraiah,** Principal, SSIT, Tumkuru for providing facilities required to complete this project.
I owe immense debt of gratitude to **Dr. M Siddappa**, Prof and Head, Dept. of CSE (PG), for his valuable guidance and constant support.
I extend my sincere thanks to my project guide **Dr. H Venugopal**, professor, SSIT, Tumkuru for his valuable guidance and constant support and for helping me throughout the project work. I also thank her for her immense support, valuable guidance, inspiration and ideas which have helped me in the completion of my project.
I take this opportunity to thank the entire teaching and non-teaching staff of Dept. of CSE (PG), SSIT, Tumkuru.
My heartfelt thanks to my family and friends who have contributed for the accomplishment of this project by their constant moral and material support.

REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up Sentiment Classification using Machine Learning Techniques". In Proceedings of the Empirical Methods on Natural Language Processing, Pennsylvania, 2002, pp. 69-86.
[2] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: A survey," International Journal, vol. 2, no. 6, 2012.
[3] Y. Mejova, "Sentiment analysis: An overview," Comprehensive exam paper, http://www.csuioedu/˜ymejova/publications/CompsYelenaMejova. pdf
[2010-03], 2009.
[4] [Rui Xia, Chengqing Zong, Shoushan Li, "Ensemble of feature sets and classification algorithms for sentiment classification", Information Sciences 181 (2011) 1138–1162.
[6] H. Cui, V. Mittal, and M. Datar, "Comparative Experiments on Sentiment Classification for Online Product Reviews." In Proceedings of AAAI-06, 2006, pp.1266-1260.
[6] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," Machine Learning, vol. 29, no. 2-3, pp. 103–130, 1996
[6] Z. Niu, Z. Yin, and X. Kong, "Sentiment classification for microblog by machine learning," in Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on, pp. 286–289, IEEE, 2012.
[8] Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in Proceedings of LREC, vol. 2010, 2010.
[9] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Information Sciences: An International Journal, vol. 181, no. 6, pp. 1138–1162, 2011.
[10] Neethu M S and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques" 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India IEEE – 31661.
[11] J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," 2004.
[12] C. Fellbaum, "Wordnet: An electronic lexical database (language, speech, and communication)," 1998.
[13] P. D. Turney, "Thumbs up or thumbs down: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 416–424, Association for Computational
Linguistics, 2002.
[14] Y. Xia et al, "Word polarity disambiguation using Bayesian model and opinion level features" Cognitive Computation, vol. 6, no.3,2016.
[16] M. Dragoni, A.G. Tettamanzi and C. da Costa Pereira, "Combined system for concept-level sentiment analysis" Semantic web evaluation challenge, springer,2014, pp,21-26.
[16] M. Araujo "iFeel: A System that compares and combines sentiment analysis methods," Proc.23 International Conf. World Wide Web,2014, pp.66-68.
[16] J.M Chenlo and D.E. Losada, "An Empirical Study of Sentence Features for Subjectivity and Polarity Classification", Information Sciences, vol.280,2014, pp.266- 288.
[18] J. K, C. Chung, C.E. Wu, and R.T.H. Tsai, "Improve Polarity Detection of Online Reviews with bag of Sentiment Concepts" Proc,11 European Semantic Web Conferrence,2014.
[19] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-Level Sentiment Models for Big Social Data Analysis," Knowledge-Based Systems, vol.69,2014, pp,86-99.
[20] G, Gezici, "SU-Sentilab: A Classification System for Sentiment Analysis in Twitter," Proc. International Workshop Semantic Evaluation,2013, pp.461-466.