

“ENHANCING THE ACCURACY OF TWITTER SENTIMENT USING CLASSIFICATION ALGORITHM”

¹Shraddha Singh, ²Mr. Pradeep Tripathi

¹Research Scholar, CSE Department, Vindhya Institute Of Technology & Science, Satna (M.P)

²HOD of CSE Department, Vindhya Institute Of Technology & Science, Satna (M.P)

Abstract

SA (Sentiment Analysis) as well as OM (Opinion Mining) are methods for performing such a comprehension. SA problems may be solved adequately via manual training. But, entirely automatic systems for analysing feelings which need no manual interventions have not been formulated thus far because of the several problems present within the field.



Sentiment Analysis provides huge number of opportunities by uncovering the opinions and views from the unstructured Twitter data set. To conclude, this research has explained that an efficient sentiment analysis can be performed on an event, Demonetization in India 2016. Throughout the continuation of this research various data analysis tools were applied to gather, clean, mine and determine review from the dataset. This analysis can help them to spot a positive turn in viewer's opinion of their brand image. Uncovering positive trends early on can permit them to make educated decisions. It is shown in this research the approach of supervised machine learning classifier 'SVM' & Naïve Baise so SVM has a major effect on the overall accuracy of the analysis. This approach has an accuracy of around 89.03% for classification.

Keywords:—Twitter; sentiment; Web data; text mining; SVM; Bayesian algorithm; hybrid; ensembles

INTRODUCTION

Content information in the world can be broadly classified into two main types: facts and opinions. Facts are objective aspects about entities, cases and their properties. A significant piece of our assembled data conduct has consistently been to discover what others think.

With the developing prevalence and accessibility of opinion-rich assets, for example, online audit sites and individual sites, new chances and difficulties emerge as individuals can do, effectively use data advances to search out and comprehend the opinions of others. Opinion mining is a sort of characteristic language handling for following the state of mind conduct of general society about a specific item. For analyzing the human thoughts, the area of opinion mining and sentiment analysis is considered as the essential one. Amazingly, there has been a little confusion between researchers and practitioners regarding the dissimilarity between opinion and sentiment and whether the field must be called Opinion mining or Sentiment analysis. Sentiment analysis and Opinion mining is nearly similar one however there is slight variation between them i.e. opinion mining extracts and analyze people's opinion about an entity whereas Sentiment analysis look for the sentiment words/expression in a text and next analyze it.

Applications of sentiment analysis

Opinion-based or feedback-based applications are more popular, currently natural language processing domain is displaying interest in sentiment analyses as well as opinion mining systems. Main application of opinions mining as well as sentiment analyses are given below:

- a) **Purchasing Products or Services:** When making a decision to purchase products or services, making accurate decisions is not a troublesome job anymore. Through this method, individuals may appraise others' opinions and experiences regarding all products and services and contrasts competitive brands.
- b) **Marketing Research:** Outcomes of sentiment analyses may be used for the purpose of market research. Through sentiment analyses methods, recent trends of customers regarding particular products or service may be examined. Likewise, current attitudes of the common public regarding new state policies may also be easily examined.
- c) **Quality Improvement in Product or Service:** Through opinions mining as well as sentiment analyses, manufacturers may gather critics' opinions as well as positive reviews regarding their products or services and therein enhance quality of their services.
- d) **Recommendation Systems:** Through the classification of individuals' opinions as positive or negative, the system can determine which gets recommended and which does not.

- e) **Opinion Spam Detection:** As the internet is equally accessible to everybody, anybody can upload anything onto the internet. This means that the probability that content is spam is increasing day by day. Individuals could upload spam content for the purpose of misleading people. OM as well as sentiment analyses are capable of classifying internet content into spam and not spam.
- f) **Detection of “flame”:** Supervising news sites, blog posts as well as social networks are made simple through sentiment analyses. OM and sentiment analyses are capable of detecting arrogant, overheated, hatred inducing, swear words utilized in emails or blogs or news sites in an automated manner.
- g) **Policy Making:** By utilizing sentiment analyses, policy makers are able to take into consideration citizens’ perspectives regarding certain policies and this knowledge may be utilized to create fresh policies favoring citizens.

PROBLEM STATEMENT

The main intention of opinion mining is to decide the polarity of feedbacks by deriving the features and constituents of the object that have been mentioned in every document. The high-speed development of mobile communication and miniature devices offer an inventive and a novel instructive strategy in the learning system. At any moment and place the learning circumstances presented in the system of Machine learning build up the learning skill of learners. Hence it is essential to develop the system of Machine Learning and to recognize the user’s viewpoint and assessment about them. This encompasses practicing the self-activating text evaluation to derive the opinions and designating self-activating text evaluation to discover the opinion’s sentiment from the viewpoint on which the users are examining or expressing their personal opinions and rating of services. The latest researches on sentiment analysis experience many numbers of restrictions, very small accuracy in classification and also document level views are not discovered. This proposed research work tackles all these issues.

OBJECTIVE

Following are the objectives of the research work:

- Collect Twitter data for sentiment classification of India demonetization 2016.

- Design “proposed algorithm” approach for analysing and classifying sentiment for demonetization.
- Implement the proposed approach on the data obtained from Twitter using „R” programming language.
- Analyze performance of the “proposed algorithm” approach against existing machine learning algorithms for classifying sentiment for demonetization.

IMPLEMENTATION

The implementation and performance evolution of proposed algorithm is ‘R’ language. An event used for sentiment analysis in this research is demonetization in India apply on midnight of 8th November 2016. The proposed ‘SVM, Naive Bayes’ approach had been applied to the tweets that belonged to an event ‘Demonetization in India 2016’, to analyze the sentiment of Twitter users towards this Demonetization . The tweets had been collected for a period of 11/22/2016 10:57:33 AM to 11/23/2016 6:40:30 PM. A database of 8000 tweets was collected for analysis.

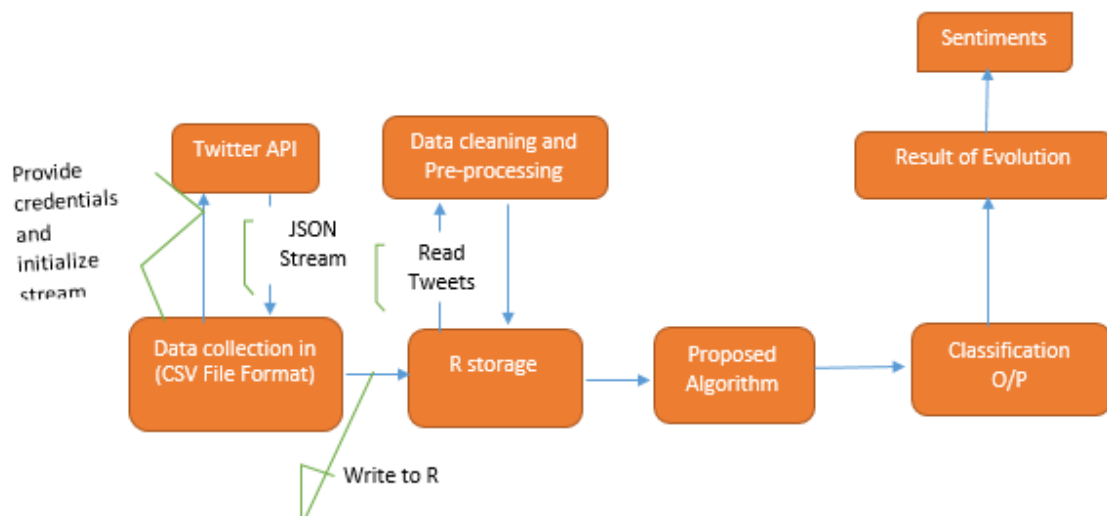


Figure:1 Architecture of Proposed Work

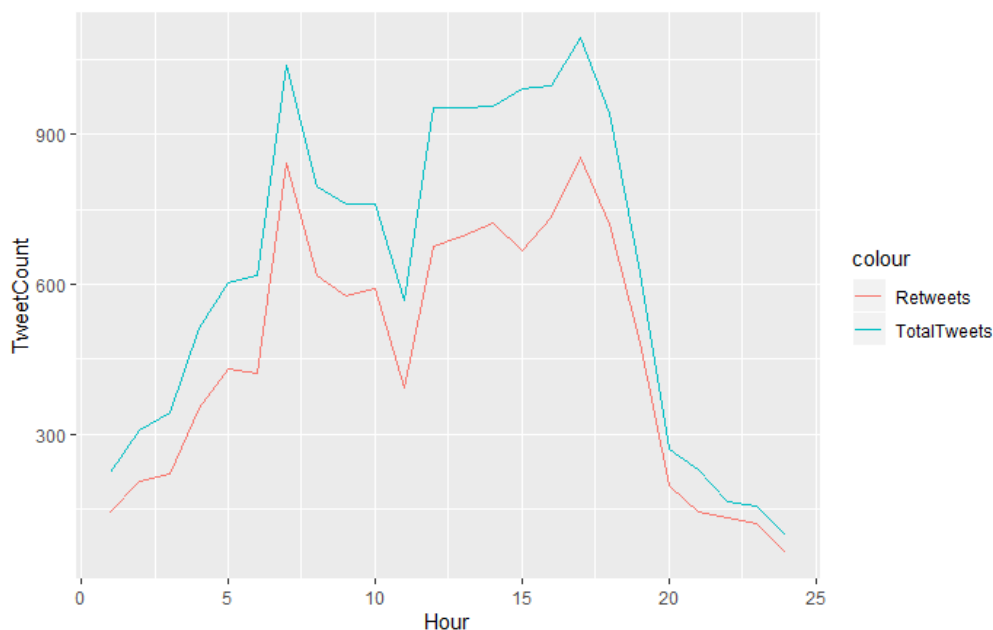
This Process model shown in above Figure 1 represents the data flow between various tools and techniques for processing also represent the process of implementation. The firstly we gather a data using Twitter API using python script ‘Tweepy’ extension library, and storing the resulting tweets in CSV file format. For this dissertation work, only the tweets containing the phrase ‘Demonetization in India 2016’ were filtered and used. This CSV file is loaded into R for next step analysis. The data is preprocessed and cleaned by using various ‘R’

packages and techniques. After this, Supervised machine learning approach (extremely randomized tree) is performed and finally compared with different machine learning techniques. For classification the model with the best accuracy parameters is chosen for sentiment classification.

RESULT

The below chart shows that majority of the tweets at any given hour are retweets. The number of original tweets are very less. This suggests that tweets from popular users are the major forces driving traffic on twitter and not a lot of original content is created by ordinary users.

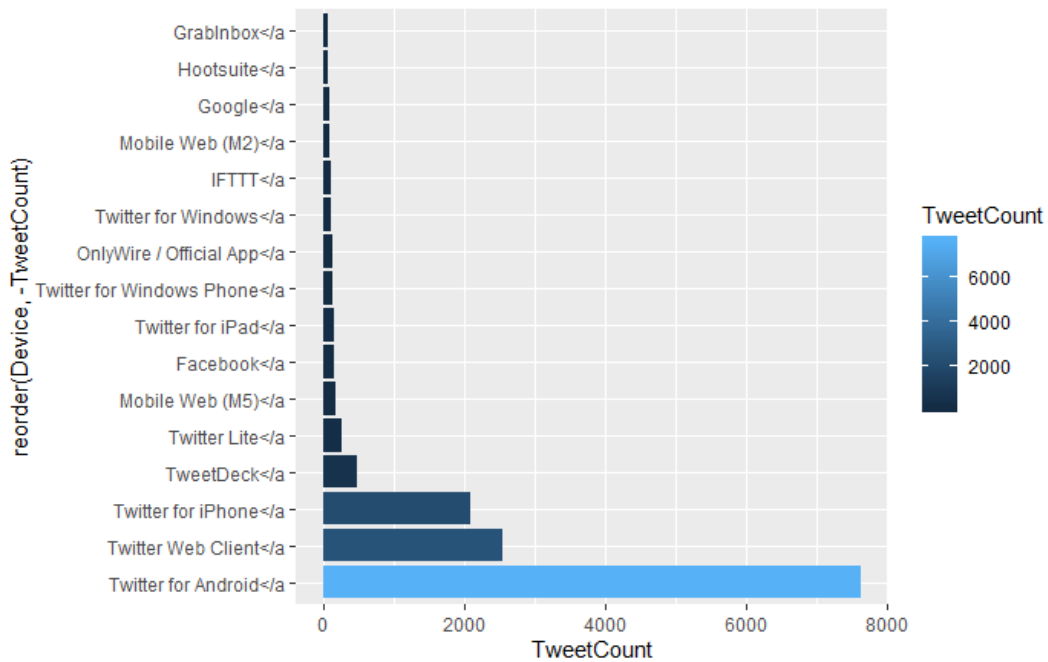
Also, the below chart shows that number of tweets peaks up in the hours between 12 PM to 7 PM in the evening. This is also the time when the number of original tweets is higher. Overall activity for other hours of the day is very small as compared to the peak hours.



Graph 1 Hourly trends for the tweets

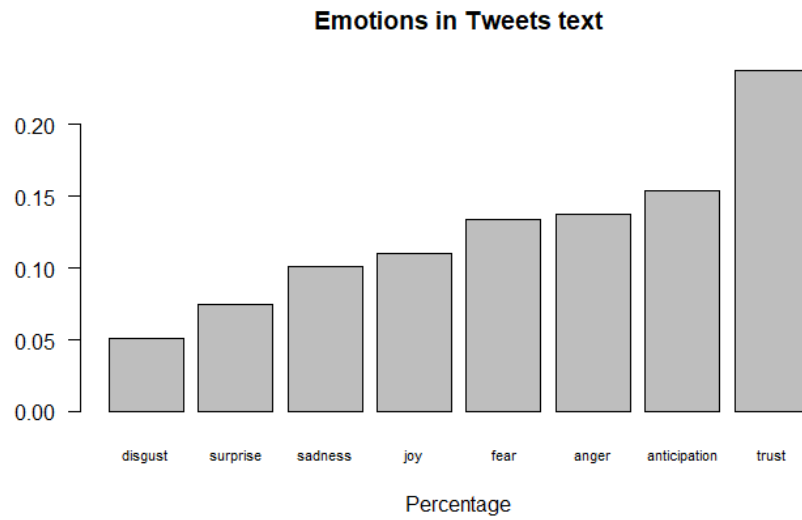
The below chart shows that tweet volume from Android Phones is much higher than any other device. This is expected in a country like India as Android has a huge user base in India and iPhones and Windows Phone do not have a high penetration in the market. Interestingly,

the number of tweets from Android phones is also much higher than number of tweets from Web Client(Desktop). Android accounts for more than 50% of the overall tweets.



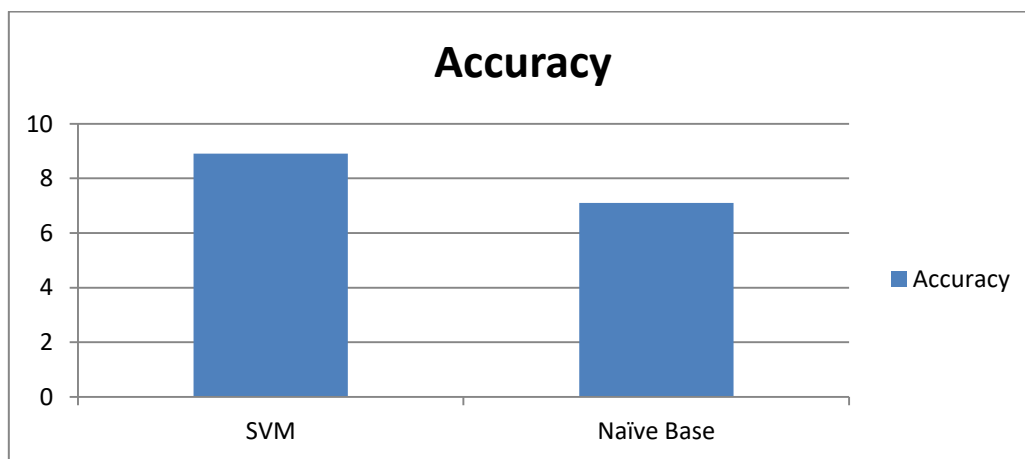
Graph 2 The top devices being used

The most dominant emotion expressed in the tweets is trust, indicating the amount of trust indians had at that time in the prime minister in taking on the black money in country. Also there was fear related to the uncertainty and problems it would cause in the country as 86% of currency was demonetized in one go. A significant percentage of people also showed emotions of disgust, sadness and anger in their tweets.



Graph 3 Sentiments on Demonetization

The accuracy of classification of the sentiment for proposed approach is highest. It means that using the ‘SVM’ approach, 89.03% of the tweets in test data are correctly classified as showing the actual sentiment of the users towards the event ‘Demonetization in India 2016’.



Graph 4 Performance Evaluations of SVM and naïve Base

CONCLUSION

Sentiment Analysis provides huge number of opportunities by uncovering the opinions and views from the unstructured Twitter data set. To conclude, this research has explained that an efficient sentiment analysis can be performed on an event, Demonetization in India 2016. Throughout the continuation of this research various data analysis tools were applied to gather, clean, mine and determine review from the dataset. This analysis can help them to spot a positive turn in viewer's opinion of their brand image. Uncovering positive trends early on can permit them to make educated decisions. It is shown in this research the approach of supervised machine learning classifier 'SVM' & Naïve Baise so SVM has a major effect on the overall accuracy of the analysis. This approach has an accuracy of around 89.03% for classification.

REFERENCE

1. Andrea Esuli and Fabrizio Sebastiani "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining" 2006. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming.
2. Haseena Rahmath P(2014) "Opinion mining and sentiment analysis -challenges and applications", International Journal of Application or Innovation in Engineering & Management, Volume 3, Issue 5, pp 401-403, May 2014
3. SMEUREANU, Cristian BUCUR "Applying Supervised Opinion Mining Techniques on Online User Reviews" Informatica Economică vol. 16, no. 2/2012
4. Chang, CH, Kayed, M, Girgis, MR & Shaalan, KF 2006, „A Survey of Web Information Extraction Systems“, IEEE T. Knowl. Data En., vol. 18, no. 10, pp. 1411-1428.
5. Hong, JL 2010, „Deep Web Data Extraction“, Proceeding of the IEEE International Conference on Systems Man and Cybernetics (SMC). Istanbul, pp. 3420-3427.

6. Jain, A, Jain, S, Shukla P & Bandiya, H 2012, „Towards Automatic Detection of Sentiments in Customer Reviews“, Int. J. Inform. Sci. Tech., vol. 2, no. 4, pp. 103-111.
7. Kamal, A & Abulaish, M 2013, „Statistical Features Identification for Sentiment Analysis using Machine Learning Techniques“, Proceeding of the International Symposium on Computational and Business Intelligence (ISCBI). New Delhi, pp. 178-181
8. Liu, B 2012, „Sentiment analysis and opinion mining“, Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167.
9. Miao, Q, Li, Q & Zeng, D 2010, „Mining fine grained opinions by using probabilistic models and domain knowledge“, Proceeding of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Toronto, ON, vol. 1, pp. 358-365.
10. Ion Smeureanu & Cristian Bucur 2012, „Applying Supervised Opinion Mining Techniques On Online User Reviews“, Informatica Economică, vol. 16, no. 2, pp. 81-91.