International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 09 | Sep 2019 www.irjet.net p-ISSN: 2395-0072

Improved Real-time Twitter Sentiment Analysis using ML & Word2Vec

Ankita Pal[1], Mr. Neelesh Shrivastava[2], Mr. Pradeep Tripathi[3]

[1]M.Tech Scholar, Department of CSE, Vindhya institute of Technology & Science, Satna (M.P.), India
[2]Assistant Professor, Department of CSE, Vindhya institute of Technology & Science, Satna (M.P.), India

---

ABSTRACT:- The Today's World is known as Data World. Generation of Data is abundant in volume. Demand of useful data is every-where every business domain works upon data this day. Data mining played very important roles from last decade now this data mining is migrated to Machine Learning. Machine Learning comes by Artificial Intelligence and Mathematical Statistics. Machine learning is categorized into supervised, unsupervised and reinforcement. Supervised machine is using the learning algorithm to detect the discovery that is clearly due to the examples supplied to produce general interpretation, which then predicts future scenarios or events. In this Dissertation explains how we fetch data from twitter using Twitter App (API) which contains different Tokens and key. Here we extract real Data from twitter in JSON Format then that converted into csv format then apply ML Algorithm for analysing sentiment analysis using polarity.

In this Dissertation, we have discussed the about the sentiment Analysis of twitter data available today, how they work.

Keywords: *Classification Algorithm, Machine Learning, Polarity, subjectivity, sentiment analysis, positive sentiment negative sentiment, NB, LR and Decision Tree & Random Forest.*

I INTRODUCTION

Social Twitter has become one of the most popular websites on the web. Currently, Twitter maintains more than 100 million users, which generate more than 50 million updates (or "tweets") in one day. While most of these tweets are vain baba or simple conversations, about 3.6% of them are subject to mainstream news. Apart from this, even during the simple conversation of friends, information is being circulated in large quantities which can serve various types of data mining applications.

Unfortunately, many devices available to users to find and use microblogging data in this vast amount are still in their relative infancy. For example, Twitter provides a search engine for the search of those posts that contain a set of key words. However, the result is a list of the positions returned by the regency rather than the relevance. Therefore, it is not uncommon to get spam in large quantities, post in other languages, rent, and other sources when wrong information is received. Another service provided by Twitter is currently a list of trending topics.

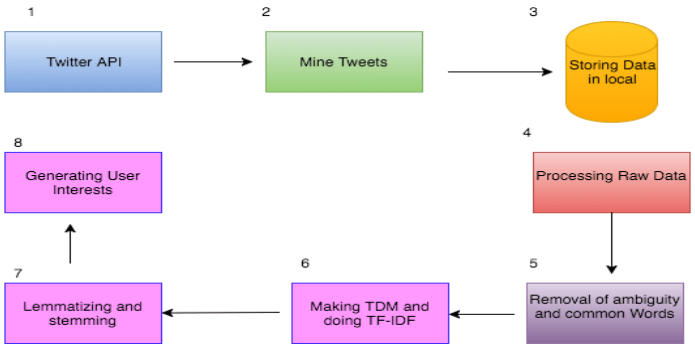


Figure 1: Twitter Architecture

1.1 Sentiment Analysis

Sentiment Analysis play very important roles during product analysis with the help of this sentiment prediction can be defined very accurately. That will give benefits for any business model.

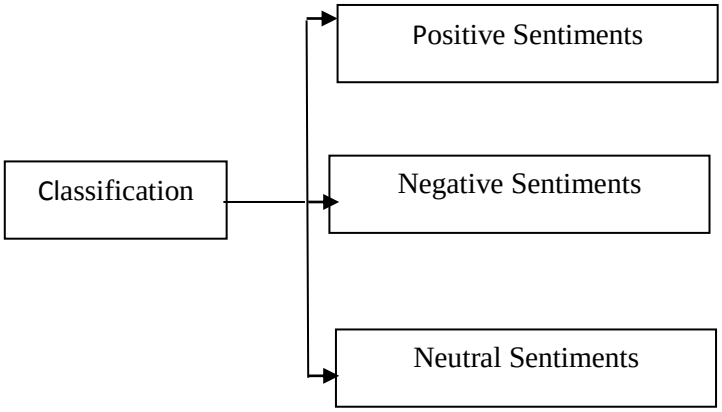


Figure 2: Sentiment Categorization

Positive Sentiments: The number of positive words is counted more than that review is considered as a positive review.

Negative Sentiments: In case of a product if the number of negative words is counted more than that review is considered as a negative review.

Neutral Sentiments: Here we will consider as a neutral sentiment.

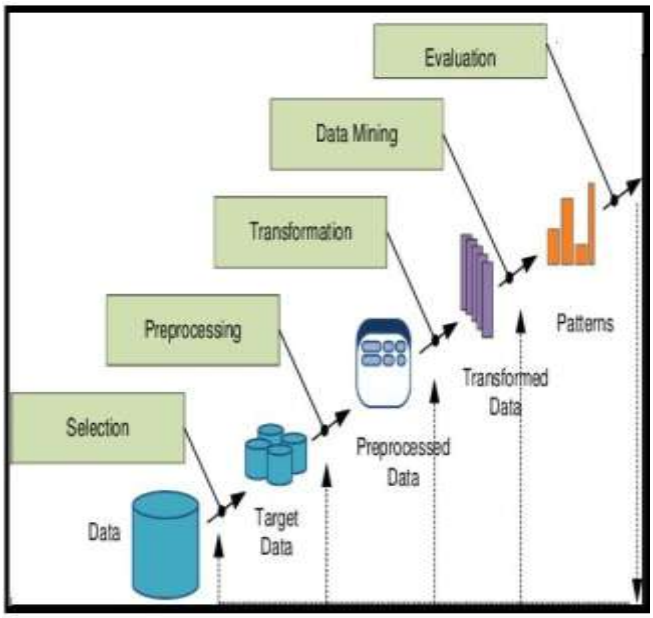


Figure 3: Process of Data Mining

Data cleaning: It is a process of removing noise and inconsistent data.
Data integration: In this step data from multiple sources are combined.
Data selection: In this step data relevant for mining task is selected.
Data transformation: In this step data will be transformed into form that is appropriate for mining.
Data mining: In this step some intelligent methods are applied for extracting data patterns.
Pattern evaluation: In this step we concentrate upon important patterns representing knowledge based on some measure are identified.
Knowledge presentation: In this step visualization and knowledge representation techniques are used to present the mined knowledge to the user.

1.3 Classification Process

In every sentence is initial classified as subjective or objective. solely subjective sentences square measure helpful for sentiment classification. Hence, the target sentences square measure discarded and therefore the polarity of subjective sentences is calculated. in line with the polarity.

Microblogging has popular communication tools in the figure 4 we explained how multiple organization works with previous data some effective example is given in above figure.

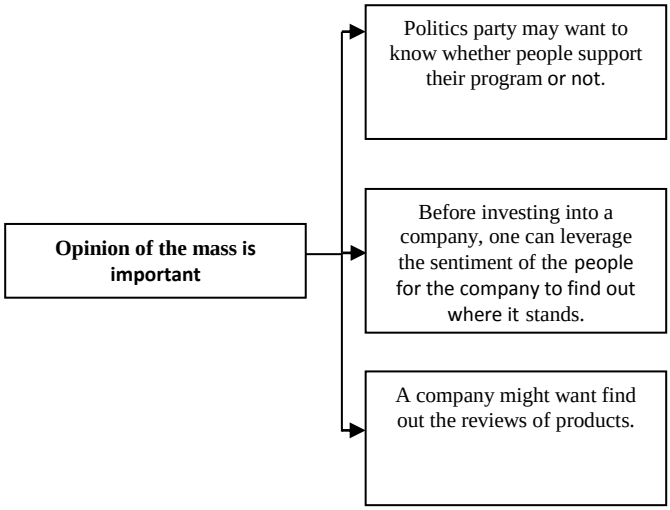
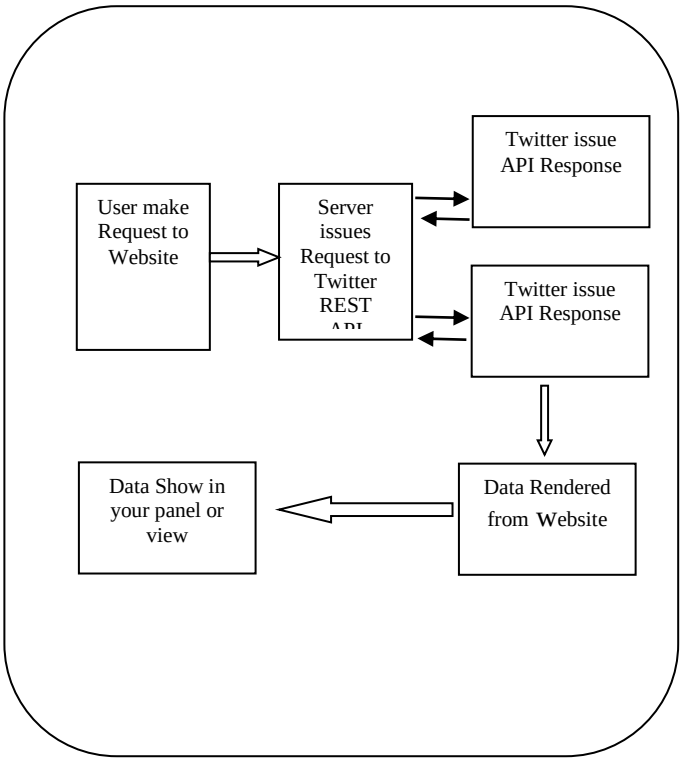


Figure 4: Microblogging as Tool

1.4 Tweeter Application Interface

Twitter provides an open API or application programming interface for external developers who designs a technology that relies on Twitter's data. Twitter API is classified based on their design and access method to access data on Twitter [8]. They are the REST APIs and streaming APIs.



In the Figure 5 we explained how we fetched data from twitter using tweepy API. In the above figure first a request comes from user that pass to tweeter App server that gives you authentication then any data can be streamed to your personal location.

1.5 Data Mining Techniques

Data mining Algorithms is categorized into different which is given below:

Classification
Classification is the frequently (most commonly) applied data mining mechanism, which explains a set of pre-classified examples to develop a (procedure) model that can (identifies or categories) classify the population (Dataset) of records at large.

Clustering
Clustering can be said as to find out of similar classes of objects. By using clustering mechanism, we can further find out dense and sparse (n Dimensional Space) regions in object space and can discover overall distribution trends (pattern) and relation among many coordinate points (correlations) among data attributes.

Association
Clustering Association play very important roles when we have abundant data with similar properties means with the above mechanism, we can divide the data which has similar properties associated with another object.

II LITERATURE SURVEY

Opinion mining and sentiment analysis is growing in every second. Aim of this mechanism is extract text present in any sentences. That data is extracted from any social media like twitter, LinkedIn, Instagram and many more. This can be solved with the help of machine learning algorithm. In recent days election is very important in any society. Sentiment Analysis play important roles in this mechanism.

According to the paper, nowadays, we have a witness for various types of review websites. That's why we can share our vision for different products that we have been able to get. The easiest way to analyse meaningful information from different types of reviews, through which we can understand the user's choice towards various products. The conventional recommendation system is used before the recommendation model so that these models can recommend products to users if they search for similar products for websites. Therefore, we can keep this in mind after earning condolences on products [1].

Linguistic options are used to check the sentiment of twitter messages for police messages. Therefore, the three types of datasets gathered together give feedback about the products: hash tag datasets, facial features datasets [3,4].

The growth of techniques of social network analysis is fast at present. These techniques are of interest to many researchers in different areas such as sociology, communication and computer science, social psychologist and so on. Nowadays, by analysing how the members of network interact, share information or establish relationships, useful knowledge about them and their relations can be extracted.

Sentiment mining from sources like Twitter which contain informal texts is needed as there is prominent information and vast amount of data to be analysed, understood and experimented. There has been a lot of research in this area to get the semantic information from this domain and to create better prediction in terms of Sentiment classification. We present a novel approach which provides an ensemble model for Classification taking SVM as base learner and Adaboost as the Ensemble Boosting algorithm. We show the Precision, Recall and F1 score by comparing it with the baseline SVM algorithm [8].

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. In fact, it has spread from computer science to management Twitter is a platform widely used by people to express their opinions and display sentiments on different occasions. Sentiment analysis is an approach to analyse data and retrieve sentiment that it embodies.

III PROBLEM IDENTIFICATION

With the rapid growth of the World Wide Web, people are using social media such as Twitter which generates big volumes of opinion texts in the form of tweets which is available for the sentiment analysis. This translates to a huge volume of information from a human viewpoint which make it difficult to extract a sentence, read them, analyse tweet by tweet, summarize them and organize them into an understandable format in a timely manner. Informal language refers to the use of colloquialisms and slang in communication, employing the conventions of spoken language such as 'would not' and 'wouldn't'. Not all systems are able to detect sentiment from use of informal language and this could hanker the analysis and decision-making process.

IV Block Diagram & Methodology

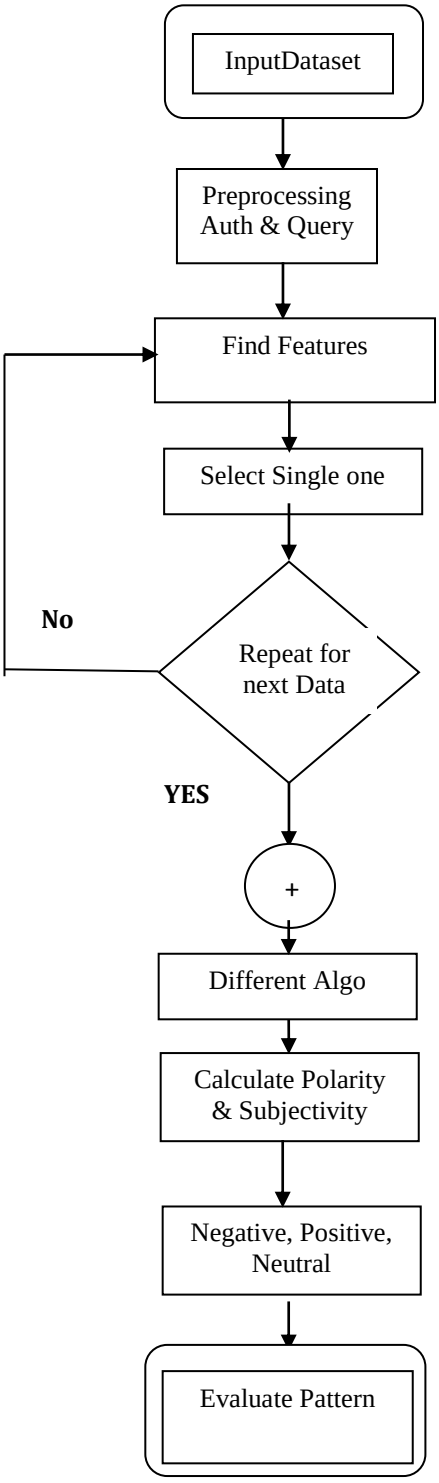


Figure 6: Block Diagram of Flow of Operation

V Experimental Result

We used python programming Language to implement our logic we used number of libraries like NumPy, pandas, tweepy, matplotlib, seaborn and many more. This project is divided into two parts in first part we fetch data in another part we will process our data our overall process is given below:

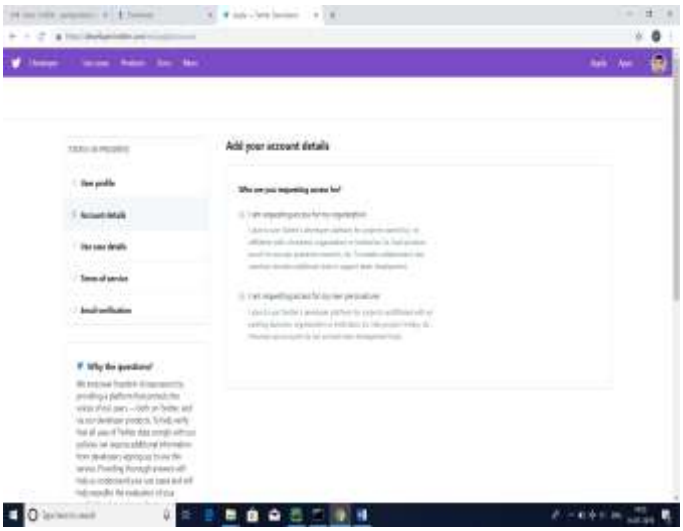


Figure 7: Twitter App to create Authentication

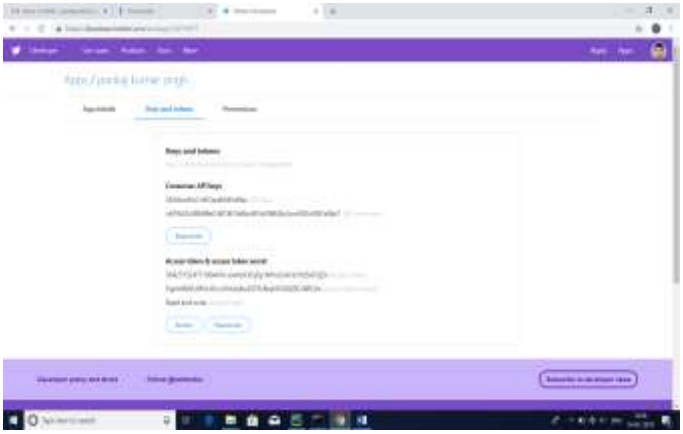


Figure 8: Tweeter Key & Tokens

Description: In the above two figures we configure our tweeter API.

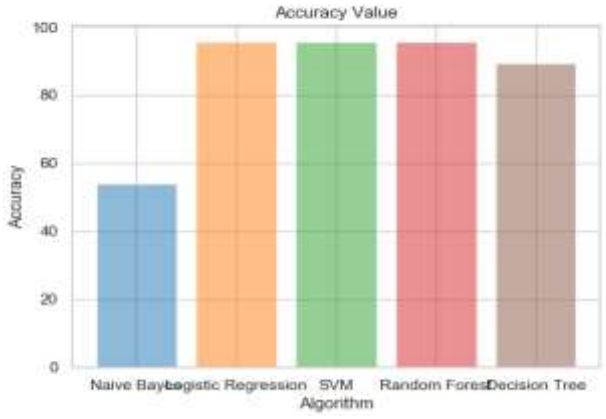


Figure 9: Comparison of Algorithm

Description: In the above figure we compare different Algorithm and gives in term of Bar Graph.

VI CONCLUSION

We the approach of machine learning is basically destined to classify the text by applying algorithms such as Naïve Bayes and SVM on the files. Considerable work has been done in the field of sentiment analysis either from sentiment lexicons or from machine-learning techniques. But, this research is focused on providing a comparison between different type of ML techniques. Experiment analysis shows that Decision Tree outstand in terms of accuracy.

REFERENCES

[1] M.Rambocas, and J. Gama, "Marketing Research: The Role of Sentiment Analysis". The 5th SNA-KDD Workshop'11. University of Porto, 2013.

[2] A. K. Jose, N. Bhatia, and S. Krishna, "Twitter Sentiment Analysis". National Institute of TechnologyCalicut,2010.

[3] P. Lai, "Extracting Strong Sentiment Trend from Twitter". Stanford University, 2012.

[4] Y. Zhou, and Y. Fan, "A Sociolinguistic Study of American Slang," Theory and Practice in Language Studies, 3(12), 2209–2213, 2013. doi:10.4304/tpls.3.12.2209-2213

[5] M. Comesaña, A. P.Soares, M.Perea, A.P. Piñeiro, I. Fraga, and A. Pinheiro, " Author ' s personal copy Computers in Human Behavior ERP correlates of masked affective priming with emoticons," Computers in Human Behavior, 29, 588–595, 2013.

[6] A.H.Huang, D.C. Yen, & X. Zhang, "Exploring the effects of emoticons," Information & Management, 45(7), 466–473, 2008.

[7] D. Boyd, S. Golder, & G. Lotan, "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter," System Sciences (HICSS), 2010

[8] T. Carpenter, and T. Way, "Tracking Sentiment Analysis through Twitter,". ACM computer survey. Villanova:Villanova University, 2010.

[9] D. Osimo, and F. Mureddu, "Research Challenge on Opinion Mining and Sentiment Analysis," Proceeding of the 12th conference of Fruct association, 2010, United Kingdom.

[10] A. Pak, and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Special Issue of International Journal of Computer Application, France: Universitede Paris-Sud, 2010.

[11] G. Rätsch, T. Onoda, and K. R. Müller, "Soft margins for AdaBoost," Mach. Learn., vol. 42, no. 3, pp. 287–320, 2001.

[12] K. Ting, "Precision and Recall," in Encyclopedia of Machine Learning, 2011, p. 1031.

[13] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014, vol. 8725 LNAI,

[14] F. Iqbal, H. Binsalleeh, B. C. M. Fung, and M. Debbabi, "A unified data mining solution for authorship analysis in anonymous textual communications," Information Sciences, vol. 231, pp. 98–112, May 2013 2013.

[15] T. Kucukyilmaz, B. B. Cambazoglu, C. Aykanat, and F. Can, "Chat mining: Predicting user and message attributes in computer-mediated communication," Information Processing & Management, vol. 44, pp. 1448–1466, July 2008 2008.