

Sentiment Analysis of Product using Tweets through Hybrid Approach

Himanshu Namdev¹, Bhavana Gupta²

¹M.Tech Scholar, ²Assistant Professor

^{1,2}Department of CSE, Sagar Institute of Science & Technology, Bhopal (M.P.), India

Abstract - Recent Era is known as Data Era. Every business is totally depending upon the previous stored data. Generation of Data is abundant in volume. Many industries are storing & tracking their previous data for next upcoming business. We know that Data mining is very promising industries which play very important roles from last many years. In recent past years Data Mining is upgraded into a new domain or industries that is Machine Learning. Basically, Machine learning comes under Data science where many different domains work simultaneously i.e. statistics, AI, Business Analysis and many more. For doing any analytics in machine learning Data play very important roles that data may be stored Data or fetching data from any different mechanism. In this Dissertation we fetch data online using Twitter App (API) which contains different Tokens and key. We are applying different ML Algorithm for analysing sentiment analysis. Here we are explaining algorithms by confusion matrix where we have number of parameters like TP, TN, FP & FN. Finally, we calculate accuracy % for each algorithm.

Keywords:Data Mining Algorithm, Machine Learning, sentiment Analysis, confusion matrix, performance parameters, Naïve bayes, Decision Tree & many classification machine learning algorithms.

I. INTRODUCTION

In this internet world many social networking sites play important roles in any business. Twitter is also a networking site which has millions of active users which write millions of text messages on their platform. A message writing on twitter platform is known as tweet which has limited characters (140). In this platform many imminent personality or organization write something about market, social impact, rating and their experience about any scenarios and product. This platform is also known as micro blogging site where many people can react accordingly. 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. Another service provided by Twitter is currently a list of trending topics.

The different components in Twitter where a recommender system can be used to recommend are described as follows:

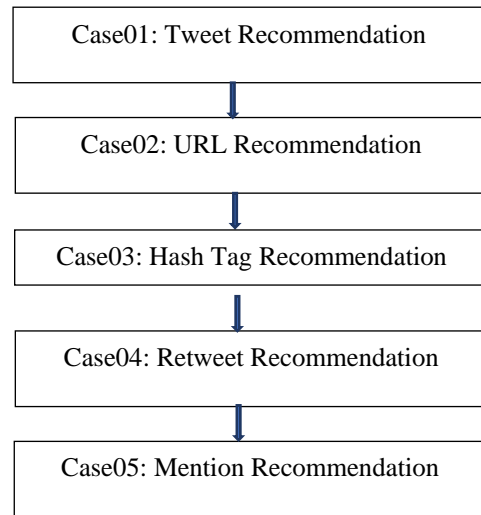


Figure 1: Twitter Recommender System [1]

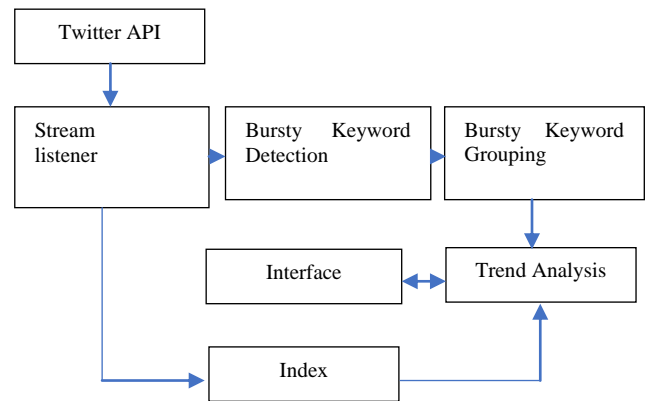


Figure 2: Twitter Architecture [1]

1.1 Sentiment Analysis

With the help of this sentiment prediction, Sentiment Analysis Perfect Important Aspiring Product Analysis can be accurately defined. Which will help for any business model.

Positive Sentiments:More than the number of positive words it is estimated that the review is considered a positive review.

Negative Sentiments:In the case of a product, if the number of negative words is estimated higher than the estimate, it is considered a negative review.

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

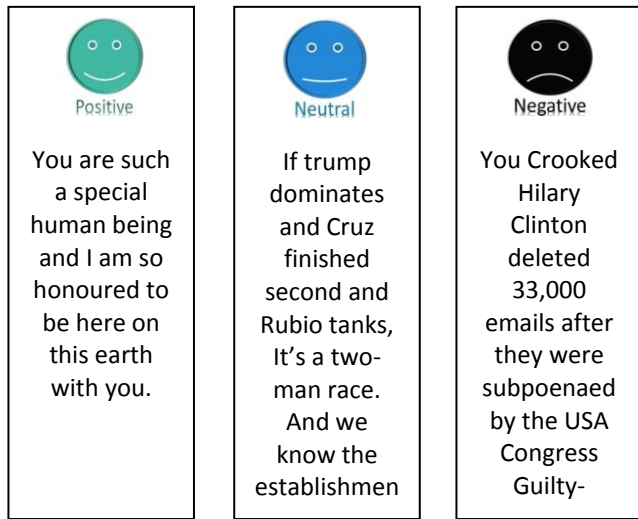


Figure 3: Sentiment Categorization [3]

1.2 Data Mining Process

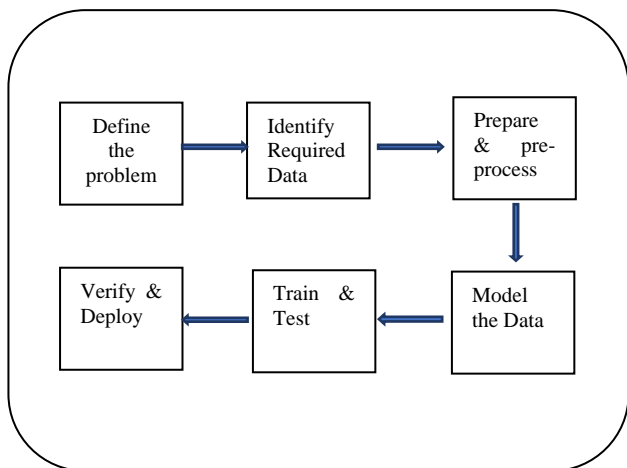
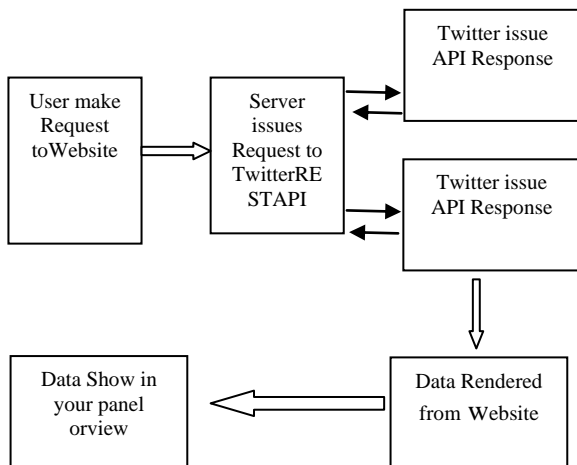


Figure 4: Process of Data Mining [4]

Define the problem: In this section we define our problem domain.



Identify Required Data: In this step we will select which type of Data set will suit the above problem domain.

Prepare & pre-process: In this step we will do the previous task for further analysis.

Train & Test: In this step data will be divided into two major part Training and Testing. At Training Data, we will create a model or classifier. At Testing Data, we will verify the model.

Verify & Deploy: In this step we will deploy a selected model for any new Data set and try to find the prediction from given Data set.

1.3 Sentiment Finding Process

In every sentence is initial classified as subjective or objective. 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.

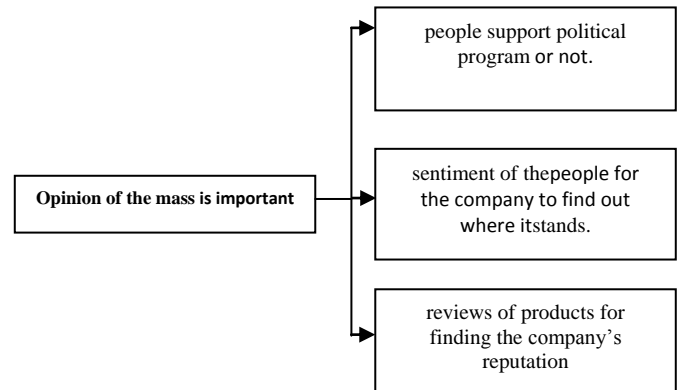


Figure 5: Microblogging as Tool

1.4 Tweeter Application Interface

Twitter provides us 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.

User

Process Tweet : Positive, Negative, Neutral words

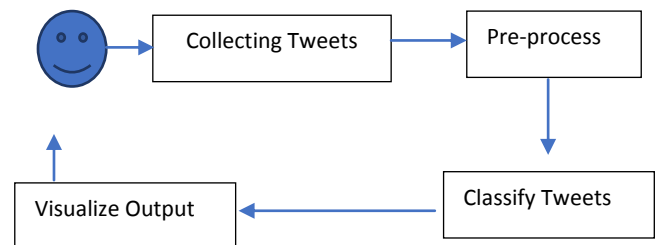


Figure 5: Framework of sentiment Analysis [8]

In the Figure 5 we explained how we fetched data from twitter using different API. API play important roles to extract data from twitter web Engine. We know that in twitter very frequently Data is changed so many researchers rely on this architecture for data variety.

1.5 Data Mining Techniques

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

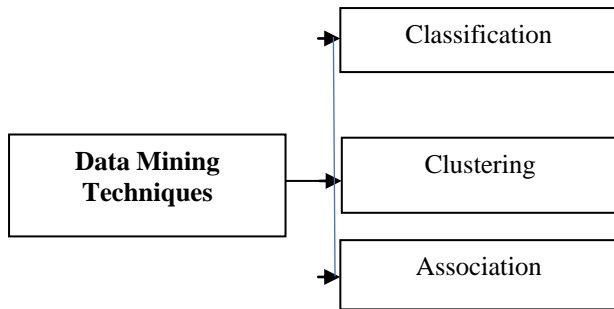


Figure 6: Framework of sentiment Analysis [8]

Classification

In this mechanism we have predefined levels on that basis we can divide our given data into different parts.

Clustering

In this mechanism we don't have a level or class on that basis we divide our data into different categories. We can say that clustering support unsupervised mechanism.

Association

Here we concentrate over trends and facts in this mechanism we concentrate over maximum properties which is similar to one another.

II. LITERATURE SURVEY

We are in Networking era which is growing day by day in different domain and industries. At every place its acceptance is growing. Nowadays in this field research is growing. In the field like Natural Language Processing which is really promising domain where analysis can give us many benefits. In different social Network millions of customers share their view or giving review for different products.

In this work, tweets of different film reviews are classified on the basis of their polarity of opinion. Which are using many different features in separate or combination. Here Author's calculated the performance of various machine learning algorithms like NaiveBayes (NB), Support Vector Machine (SVM). Here main goal of Autor's is to evaluate how the performance of a classifier may affected by using different feature combinations in sentiment analysis. Experiments setup had done with data from Stanford Twitter Sentiment 140 dataset and the second source is IMDb Movie Review dataset. Mainly Four different evaluation metrics: accuracy, precision, recall, and F1 score are used for our results. [6].

In 21st century Social Networks are used for finding the opinions of users for different products and services utilize by them in day-to-day life. In fraction of second users or

customers shares their opinion in different scenario or environment. Our world is growing in terms of companies and organizations and everyone is contributing their experience with help of different platform. Every platform is made for catering different issues or you can say that they required for different requirement. The motto of this

paper is to provide an automated system that gives us review of different people posted on social Networking site using Hadoop, Hadoop may process huge amounts of data. Now, Here Authors work with different issues like emotion classifications, and negatives related to operations. Authors wants to use accurate mechanism which gives better prediction. Authors are dealing with the different challenges arises during the process of sentiment analysis. The main motto of this paper is to find-out real-time emotional analysis on the tweets that are extracted from Twitter's platform and give us time-based analysis to our customer or people [7].

Analysis of sentiment comes under Natural Language Process Domain. Here we can analyse computational linguistics, process to find identification using biometrics, Analysis of Text and find out the information in terms of subject. we have many Social Networking Platform out of then Tweeter has very important impact in many industries where customers give the real time and their real emotion in term of Text Writing. Which is very consist in terms of counting. People are giving their opinions as positive, negative or neutral which may give different information regarding product and services. Trends is also very important for any products and services which may analyse with help of those Analysis. The results are illustrated using visualization techniques such as Bar Graph, pie Graph and histograms charts to make positive, negative, and neutral comments about their opinions [8].

Nowadays, people from all over the world use social media sites to share information. For example, Twitter is a platform in which users read posts known as 'platform tweets' and interact with various communities. Users share their daily lives, posting their opinions on everything such as brands and locations. Companies can benefit from this huge platform by collecting opinion related data on them. The purpose of this paper is to present a model that can perform sentiment analysis of actual data collected from Twitter. API, then cleaning and searching the performed data. The data were then fed into several models for training purposes. Each tweet is classified based on its sentiment whether positive, negative or neutral. Data were collected on two restaurants, McDonald's and KFC, to show which restaurant has more popularity. Various machine learning algorithms were used. The results of these models were tested using various test matrices such as cross validation and F-score. In addition, our model

performs strongly on mining texts extracted directly from Twitter [9].

III. PROBLEM IDENTIFICATION

We are living in internet world here every second generated millions of records. Because generation of record is so fast and frequent so our main motto is to capture or fetch a useful data and apply a good mechanism or algorithm to find out better results. We know that Facebook, Instagram and twitter is reliable social websites. Here Authors shows trust over Twitter Web Engine to extract online Data for finding any reviews.

V Block Diagram & Methodology

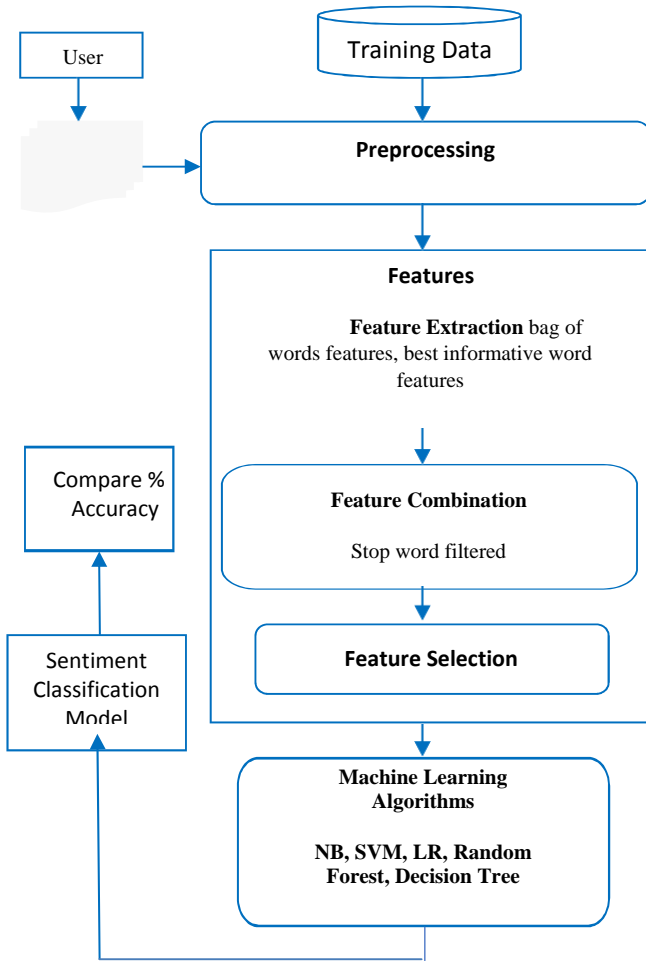


Figure 6: Block Diagram of Flow of Operation

Explanation: In the above figure 6 Authors explained that any twitter user first signup and create an App for Authentication process then fetch some data from twitter. That data is divided into two parts training & Testing Data. First Authors used Training Data for Model creation in this process Data pre-processing is mandatory because it helps to find better result. Then Authors applied number of Machine learning Algorithms for sentiment classification. That model classifies Training Data into Negative & positive class. Finally, we find Accuracy %.

IV. 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:

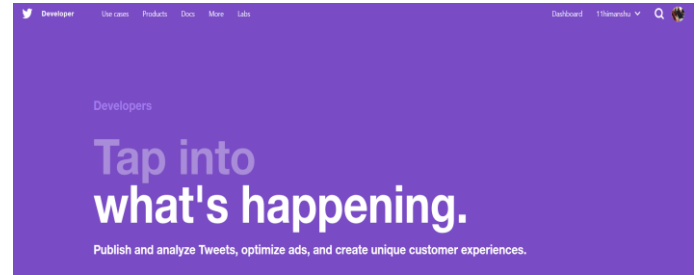


Figure 7: Twitter App to create Authentication

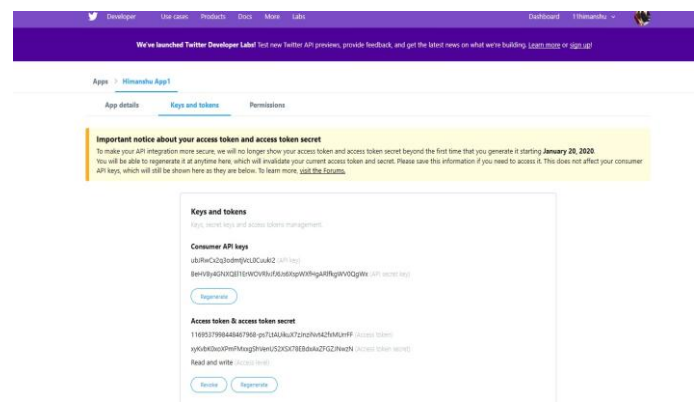


Figure 8: Tweeter Key & Tokens

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

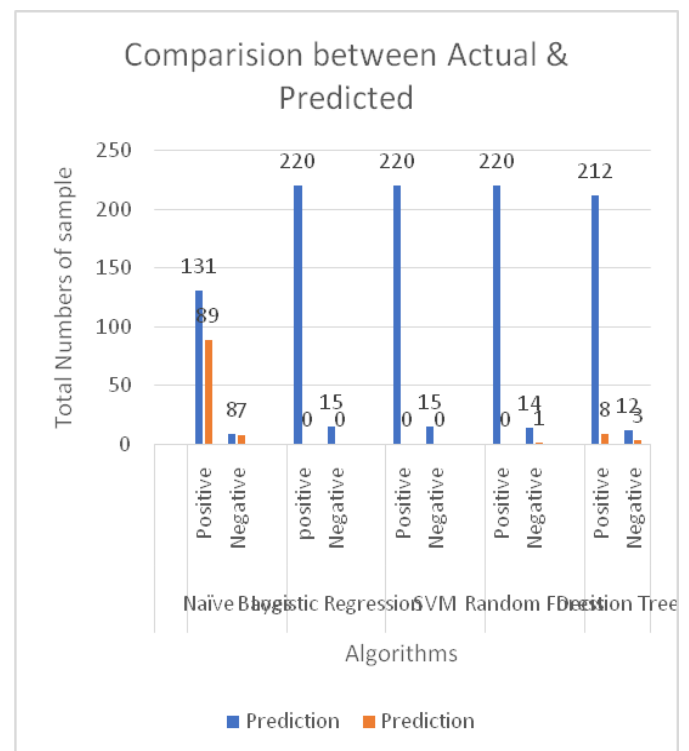


Figure 9: Comparison of Algorithm

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

V. CONCLUSION

Here Author's conclude that many machine learning algorithms are used to classify the Text on the basis of their requirement. Authors did their work with the help of machine learning algorithms like NaiveBayes, Decision Tree & Random Forest and shows their comparison on the basis of their Accuracy. Finally, by result analysis they conclude that Random Forest gives better result.

REFERENCES

- [1] Kamaljit Kaur, Kanwalvir Singh Dhindsa, "Study of Existing Design Architectures and Frameworks in Twitter". 2019 International Conference on Automation, Computational and Technology Management (ICACTM) Amity University, 978-1-5386-8010-0/19 ©2019 IEEE.
- [2] Rohan D.W Perera, S. Anand, K. P. Subbalakshmi and R. Chandramouli, "Twitter Analytics: Architecture, Tools and Analysis", The 2010 Military Communications Conference - Unclassified Program - Cyber Security and Network Management, 978-1-4244-8180-4/10 ©2010 IEEE.
- [3] Metin Bilgin, "Sentiment Analysis on Twitter data with Semi Supervised Doc2Vec". UBMK-17, 2nd International conference on computer science & Engineering, 978-1-5386-0930-9/17 ©2017 IEEE.
- [4] <https://digitaltransformationpro.com/data-mining-steps/>
- [5] Andreea Salinca, "Business reviews classification using sentiment analysis," 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 978-1-5090-0461-4/16 © 2016 IEEE.
- [6] Nazma Iqbal, Afifa Mim Chowdhury, "Enhancing the Performance of Sentiment Analysis by Using Different Feature Combinations," International Islamic University, Chittagong Chittagong- 4318, Bangladesh.
- [7] M. Trupthi, Suresh Pabboju, G. Narasimha, "Sentiment Analysis on Twitter Using Streaming Api," 2017 IEEE 7th International Advance Computing Conference, 978-1-5090-1560-3/17 © 2017 IEEE.
- [8] Prakruthi V, Sindhu D, Dr S Anupama Kumar, "Real Time Sentiment Analysis of Twitter Posts," 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018, ISBN: 978-1-5386-6078-2 © 2018 IEEE.
- [9] Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi, Wejdan Abdullah AlShehri, "Sentiment Analysis of Twitter Data," 978-1-5386-8125-1/19 ©2019 IEEE.
- [10] Miss. Shital Anil Phand, Mr. Jeevan Anil Phand, "Twitter Sentiment Classification using Stanford NLP," 978-1-5090-4264-7/17 ©2017 IEEE.
- [11] Paramita Ray, Amlan Chakrabarti, "Twitter Sentiment Analysis for Product Review Using Lexicon Method," 2017 International Conference on Data Management, Analytics

and Innovation (ICDMAI) Zeal Education Society, Pune, India, Feb 24-26, 2017.

- [12] Prerna Mishra, Dr. Ranjana Rajnish, Dr. Pankaj Kumar, "Sentiment Analysis of Twitter Data: Case Study on Digital India," 2016 International Conference of Information Technology (InCITE)-The Next Generation Summit, 978-1-5090-2612-8/16 @ 2016 IEEE.