

# A Novel Approach for Sentiment Analysis using Machine Learning for Twitter Web Portal

Amar Kumar <sup>1</sup>, Dr. Avinash Sharma<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Department of CSE, MGI, Bhopal (M.P), India

<sup>2</sup>Head & Associate Professor, Department of CSE, MGI, 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.

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

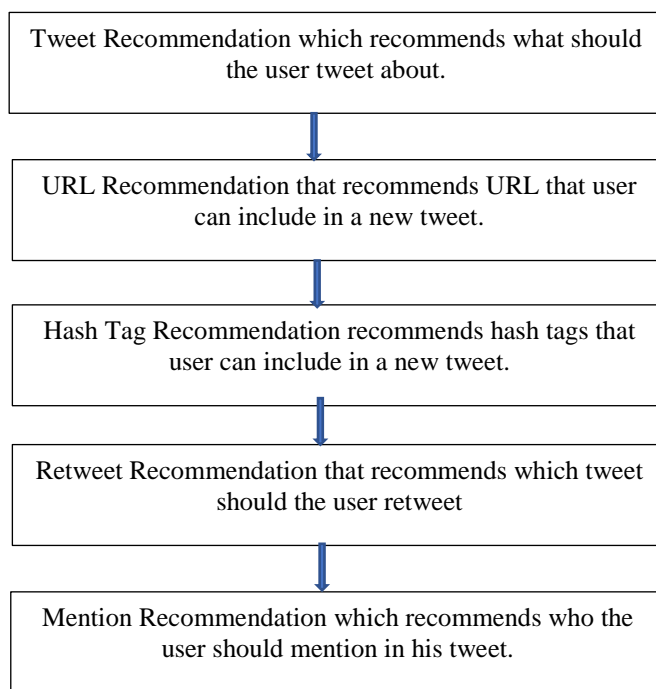


Figure 1: Twitter Recommender System [1]

### 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.

### 1.2 Data Mining Process

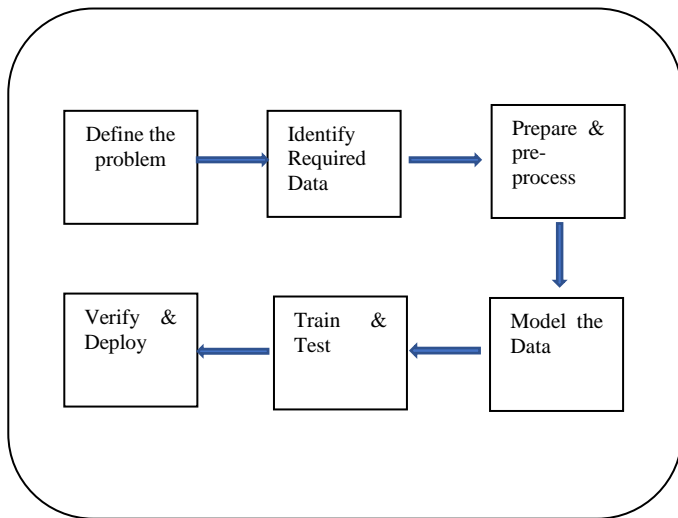


Figure 4: Process of Data Mining [4]

**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 examples are given in above figure.

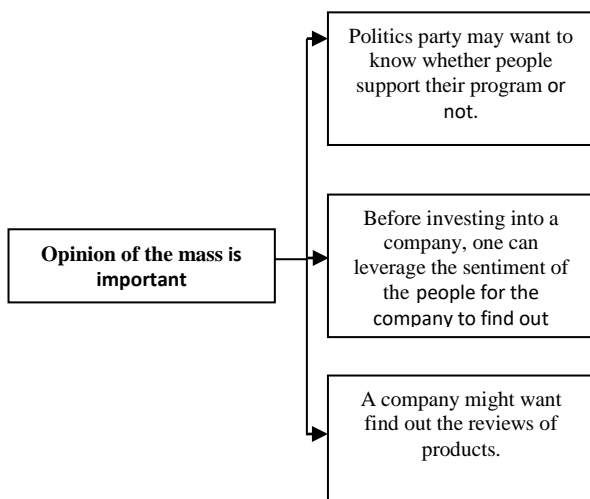


Figure 5: Microblogging as Tool

### 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.

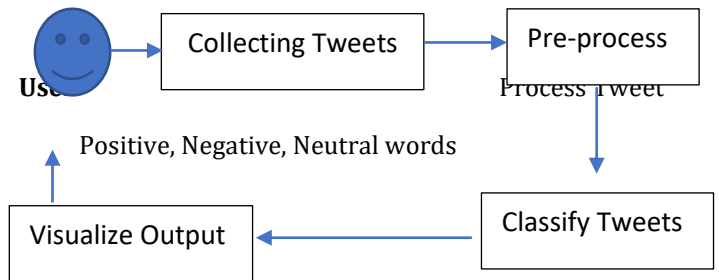


Figure 5: Framework of sentiment Analysis [8]

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:

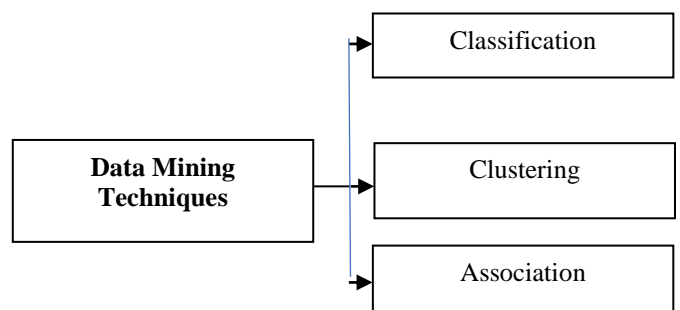


Figure 6: Framework of sentiment Analysis [8]

#### 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

With the increasing reach and acceptance of the social networking era, sentiment analysis has become one of the most prominent research domains in natural language processing. Each day, millions of people share their thoughts and ideas by posting to social media or writing online reviews. This heavy participation, on the one hand, makes these media public; However, on the other hand, it presents some challenges in identifying the dominant opinion. In this work, tweets and film reviews are classified according to the polarity of opinion using several features in combination. The performance of several feature combinations was evaluated by feeding into various machine learning algorithms (NB, SVM, MaxEnt). Therefore, the goal of the work was to evaluate how the performance of a classifier is affected by using different feature combinations in sentiment analysis. Experiments were conducted on data from two different domains such as the Stanford Twitter Sentiment 140 dataset and the IMDb Movie Review dataset. Four different evaluation metrics: recall, precision, accuracy, and F1 score are used to evaluate the test results of our system. This research shows that carefully choosing the correct feature combination can increase classification accuracy while a random feature combination will provide little benefit. [6].

In general, social media platforms have used opinion mining to find out what people think and feel about their products and services. Every day millions of users share opinions on different aspects of life. Inspired by that growth, companies and media organizations are increasingly seeking information. Efficient techniques are needed to collect large amounts of social media data and extract meaningful information from them. The purpose of this paper is to provide an interactive automated system that explains the sentiment of review / tweet of people posted on social media using Hadoop, which can process large amounts of data. Now, this research community has somewhat different problems, namely, emotion classifications, feature-based classifications and negatives related to operations. An accurate method is used to predict sentiment polarity, which helps improve marketing strategies. This paper deals with the challenges seen in the process of sentiment analysis, confirming real-time tweets as they are rich sources of data for opinion mining and sentiment analysis. This paper focuses on Sentiment Analysis, Feature Based Sentiment Classification and Opinion Summarization. The main objective of this paper is to conduct real-time emotional analysis on the tweet being extracted from Twitter and provide time-based analysis to the user. [7].

Sentiment analysis refers to application to natural language, text analysis, computational linguistics, and biometrics processing to identify, extract, quantify, and learn affective state and subjective information. Twitter, one of many popular social media platforms, is a place where people often choose to express their feelings and emotions about a brand, product, or service. An analysis of

the sentiment of a tweet is very helpful in determining people's opinions as positive, negative or neutral. This paper evaluates people's feelings about a person, trend, product, or brand. The Twitter API is used to access tweets directly from Twitter and to create a sentiment classification for tweets [8]. The results of the analysis are illustrated using positive visualization techniques such as histograms and pie charts to make positive, negative, and neutral comments about their opinions.

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].

Twitter is a micro-blogging site where users review or tweet their viewpoints, that is, opinions about service providers' Twitter pages are in words and it is useful to analyse sentiments. Analysis means finding the viewpoint of users or customers where it is between and representing positive, negative, neutral or positive-neutral, or between negative-neutral. Such a system or tool tweet gets information from Twitter about shopping websites, or any other Twitter page like some business, mobile brand, textile brand, live events like sports matches, elections etc. These results will help the service provider to know the attitude of customers towards their products [10].

## 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.

#### IV. Block Diagram & Methodology

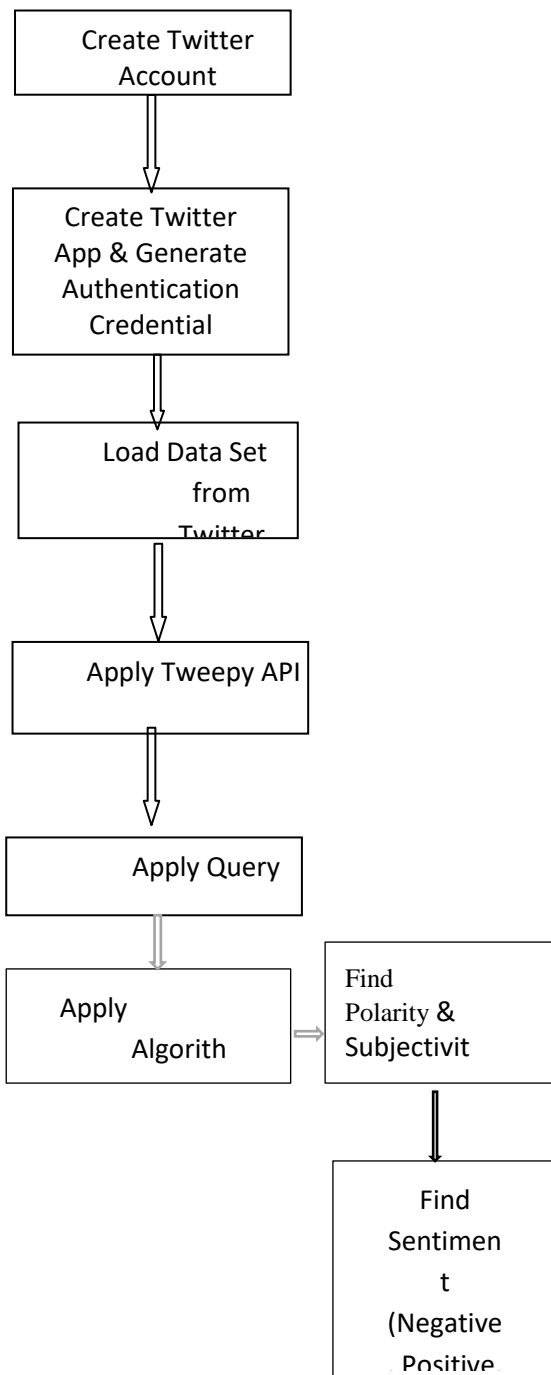


Figure 6: Block Diagram of Flow of Operation

**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 %.

#### 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

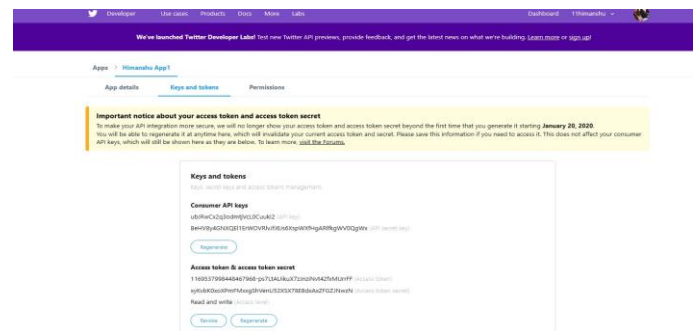


Figure 8: Tweeter Key & Tokens

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

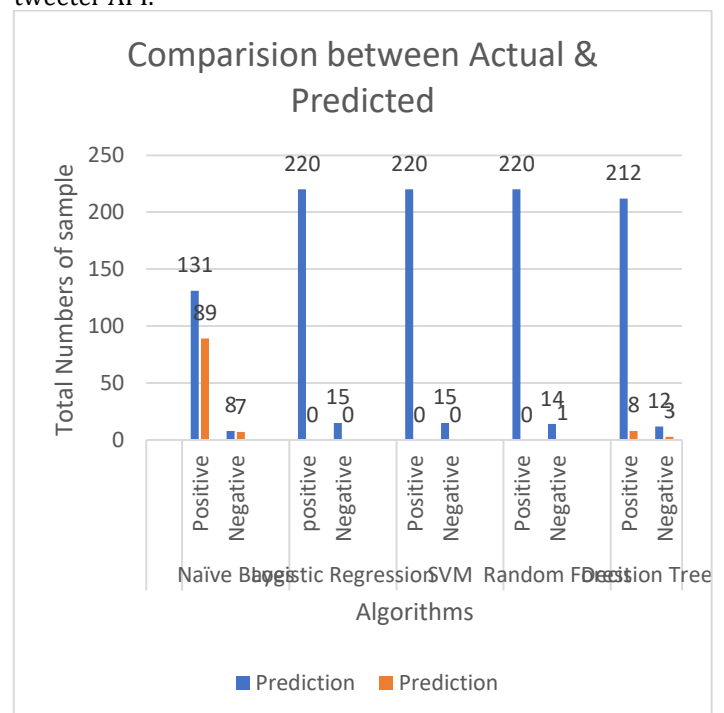


Figure 9: Comparison of Algorithm

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 Naive 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] 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 ,2<sup>nd</sup> 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 .
- [13] Ankur Goel, Jyoti Gautam, Sitesh Kumar, "Real Time Sentiment Analysis of Tweets Using Naive Bayes," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14-16 October 2016, 978-1-5090-3257-0/16 ©2016 IEEE.
- [14] Bharat R. Naiknaware, Seema S.Kawathekar, "Prediction of 2019 Indian Election Using Sentiment Analysis," Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018) IEEE Xplore Part Number:CFP18OZV-ART; ISBN:978-1-5386-1442-6 , 978-1-5386-1442-6/18 ©2018 IEEE.