# Analysis of twitter data using Machine learning algorithms

**[1] Miriyala Devi Sai Kiran**
[1]M.tech, Department Of Computer Science and Systems Engineering,
Andhra University, Visakhapatnam, AP, India.

***ABSTRACT***: Social media allows to share the experiences with many best suggestions and provides opportunities to share the ideas about any topics at any time. In the current trending, twitter is used to gather different kinds of information as user need and it is a social network service which enables the user for better communication and gaining of knowledge. Accurate representation of the user interactions can be done based on the facts sematic content. The pre-processed tweets which are stored in database are been identified and classified whether it relates to the user keywords related posts. The best suggestion using polarity can be predicted using the user keywords. For the interactive automatic system which predicts the tweets posted by the user this system deals with the challenges that appears during the sentimental analysis. It deals with effective study prior to the subjective information. The basic task in this is to identify the polarity of a given tweet in the sentence whether it is positive, negative or neutral. However the polarity of the tweets has been identified, it was difficult for us to check with the meaningless data. To address this challenge the extracted tweets are been pre-processed by replacing the full form instead of short term words. The better performance can be achieved using more training data. However the analysis was frequently done using the previously stored data, it was a challenging task to do it using the streaming data. There are very few works related to the sentiment analysis using online streaming data. In this paper, we propose that the sentiment analysis can be improved using the online streaming data. For online streaming data all the data related to the given topic will be collected according to the current data in the twitter. For better up-to-date analysis, the streaming data is used and can achieve better results. In contrast by conducting the continuous learning from the streaming data, this approach provides better results than the traditional way of using the training data and it achieves the overall performance and computational efficiency. The main objective of the work presented with in this paper was to design and implement twitter data analysis and visualization in R Language platform. Our primary approach was to focus on real-time analysis rather than historic datasets. Twitter API allow for collecting the sentiments information in the form of either positive score, negative score or neutral. We show the application of sentimental analysis and how to connect to Twitter and run sentimental analysis queries. We run experiments on different queries from politics to humanity and show the interesting results. We realized that the neutral sentiment for tweets are significantly high which clearly shows the limitations of the current works. this study focuses mainly on sentiment analysis of twitter data which is helpful to analyze the information in the tweets where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases. Sentiment analysis in Twitter has become a good research topic in recent years. As Twitter allows only 140 characters, it is very challenging to determine whether the tweet was a positive or negative tweet., we provide research on twitter data streams. We have also discussed general challenges and applications of Sentiment Analysis on Twitter.

***KEYWORDS****:* Twitter, Sentiment analysis (SA), Opinion mining, Machine learning, Naive Bayes (NB), Support Vector Machine (SVM).

## I. INTRODUCTION

Nowadays, the age of Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media, etc. Nowadays, millions of people are using social network sites like Facebook, Twitter, Google Plus, etc. to express their emotions, opinion and share views about their daily lives. Through the online communities, we get an interactive media where consumers inform and influence others through forums. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. People mostly depend upon user generated content over online to a great extent for decision making. For e.g. if someone wants to buy a product or wants to use any service, then they firstly look up its reviews online, discuss about it on social media before taking a decision. The amount of content generated by users is too vast for a normal user to analyze. So there is a need to automate this, various sentiment analysis techniques are widely used [1] [2].

In the analysis of twitter sentiment, the sentiments in the form of positive and negative feedbacks, where sentiments from all over the world express in relation to various topics linked with people opinion in business, health care, government policies etc... Extraction of information from those data is of great importance[1].Micro blogging in today's life has become a famous communication tool amongst users of Internet, hundreds and lakhs of messages are emerging progressively in well-liked web sites with the intention to offer services for forums and blogs consisting of twitter or messages about lifestyles, critique on the cutting-edge issues. [2][3]. A dataset kaggle is used to collect the messages from Twitter. It contains huge variety of brief opinions hosted by the users of this micro blog platform. There are large numbers of methods involved in sentiment analysis and most of them depend on Machine Learning and Natural Language Processing [1]. Sentiment classifier that can evaluate positive or

negative and neutral opinions for a document is built using the corpus. For most of the online company products, people give positive or Negative review. People can be able to see the comments and rating to buy the good product and also able to identify the comments as positive and negative

**Dataset Constructs and Sentiment Labelling:** The training datasets are used as the application of the domain knowledge for training the lifetime learning model. When a new task comes, that first checks for the interrelation between the new and every previous tasks. If the two tasks are identical then we combine the knowledge of both the tasks into a single task and we acknowledge it as a new task. It stores the knowledge collected from all the previous tasks. Finally it compares the knowledge from the previous tasks and predict on the future tasks for a better learning technique. It recollects the knowledge form the previous learning and it helps for better future learning.

**Challenges in Sentiment Analysis:** Sentiment analysis is useful nowadays in many ways. The method in which delivering the information from person to person plays a important role in customer buying decisions. In commercial ways like sharing the opinions or attitude about any business or product or in any other social issues. Applications such as movie reviews, business, political leaders, government officials, stock market, consumer market and social issues. The challenges in the sentiment analysis are:

**Indirect Sentiment**-Certain sentence may have an indirect sentiment even without any presence of the sentiment words. Example: "How can you afford to buy this product?", "One should question the writer's point of view who wrote this novel". Here both the sentences do not express any negative words but both the sentences contains negativity and it denotes a negative sentence. Hence identifying the semantics is very essential in the sentiment analysis.

**Impede Expectations**-Certain sentence builds up the sentence at the first and refuse it at the end of the statements. Example: "The product is amazing, it has a very good specification and there are many colors available and has very good usages, however it can't hold up". Inspire of the word which has a positive orientation, the total sentiment is negative since the second statement is crucial and it plays an important role. The term frequency of the sentence is more important than the term presence.

**Practical Sentiment**-It is very important to understand the practicality of the sentence because, sometimes it changes the whole meaning of the sentence. It changes the sentiment of the sentence completely. Example: "I just finished watching the football match. The finals completely destroyed me". In this case of practical sentiment, the first statement describes about the watching of football match and the negative orientation is discussed in the second statement. There are many ways to represent the practical of the sentiment. As per the practical knowledge the overall orientation inferred here is negative.

**Negation**-In sentiment analysis, handling the negation task is challenging and is very difficult. The negation terms can be expressed even without the usage of the negative word. In the sentence "I do not like the product", there is a negation operator (like not). But in the sentence "I do not like the acting but I like the location of the movie", there is a combination of "not" and "only". The combination of "not" and "only" deter-mines the orientation of the sentiment. Such negated words should be handled carefully.

**Role of Expressions:** Role of expressions is a type in which the word is assigned using its symbolic functions. It plays a crucial role in assigning each word with its functions. Adjectives-It is a very important parts of speech used and is used very frequently. There is a relation between the adjectives and the subjectivity. In general, people most commonly use the adjectives to express their sentiment towards the topic. Expressing the thoughts using the adjectives has high accurate results and the user can express their own point of view about the topic in an expressive manner. Example for wordlist of positive adjectives: "Brilliant, Excellent, Awesome, Fantastic, Cool, Exciting". Example for wordlist of negative adjectives: "Bad, Slow, Terrible, Stupid". These examples indicates the list of words containing the positive and the negative adjectives.

**Adjective Adverb Mixtures**-In sentiment analysis the adjective-adverb mixture performs the crucial role in analyzing the sentiment in the sentence. To find the overall polarity of the tweets, it is necessary to evaluate all the sentence of the tweets. Each sentence plays an important role in finding the sentiment (polarity). In general the users posting the tweets may use the adjective/adverb in a sentence to represent the thoughts about any topic. So it is necessary to calculate even the adjectives and adverbs for finding the sentiment. There are many types in adverbs representing the sentiment. There are declaration adverbs where the words used are affirmative and proclaimed words. In adverbs with doubts, there are certain words which represents as a doubt statement and does not represent a declarative sentence. There are also strong adverbs, weak adverbs and negation adverbs. For each type of adverbs there is each sentiment and the sentiment for each type is also calculated. Few types are: Adverbs with declaration: Certainly, Completely. Adverbs with doubts: Maybe, Probably. Strong adverbs: Extremely, Distinctly. Weak adverbs: Slightly, Hardly. Negation usage: Never, Certainly not. In the adjective-adverb mixture the sentiment is been calculated as, if a word in a sentence denotes the adverb and other word in the same sentence denotes an adjective, the sentiment score value of the sentence is altered by the adverb adjoining with it. And if the sentence containing more than one adverb and adjective, the score is calculated by altering the score of the adjective as each adverbs gets added to that. To calculate the score of the adjective-adverb mixture, a score value is allocated based on the adjective and adverb. There are certain cases like the sentence representing the weak adverb and adverb with doubts has the score less than or equal to the strong adverbs and adverbs with declaration. Example for strong adverb

and adverb with declaration: "Extremely good is more positive and has high score when compared with the adjective good". The word extremely good has more positive score when compared with good. Example for weak adverb: "Hardly good has more negative score when compared with the adjective good which has a positive score". In this, hardly good indicates a negative orientation and good indicated a positive orientation. Example for adverb with doubt: "Perhaps good has less score point when compared with the word extremely good". The first word "perhaps good" indicates a word containing doubt and the second word "extremely good"

## 2. SENTIMENT ANALYSIS

Sentiment analysis can be defined as a process that automates mining of attitudes, opinions, views and emotions from text, speech, tweets and database sources through Natural Language Processing (NLP). Sentiment analysis involves classifying opinions in text into categories like "positive" or "negative" or "neutral". It's also referred as subjectivity analysis, opinion mining, and appraisal extraction.

The words opinion, sentiment, view and belief are used interchangeably but there are differences between them.

Opinion: A conclusion open to dispute (because different experts have different opinions)

View: subjective opinion

Belief: deliberate acceptance and intellectual assent

Sentiment: opinion representing one's feelings

An example for terminologies for Sentiment Analysis is as given below,

<SENTENCE> = the story of the movie was weak and boring

<OPINION HOLDER> =<author>

<OBJECT> = <movie>

<FEATURE> = <story>

<OPINION >= <weak><boring>

<POLARITY> = <negative>

Sentiment Analysis is a term that include many tasks such as sentiment extraction, sentiment classification, and subjectivity classification, summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as, products, individuals, topics, organizations, and services.

Mathematically we can represent an opinion as a quintuple (o, f, so, h, t), where o =object; f =feature of the object o; so=orientation or polarity of the opinion on feature f of object o; h=opinion holder; t =time when the opinion is expressed.

**Object:** An entity which can be a, person, event, product, organization, or topic

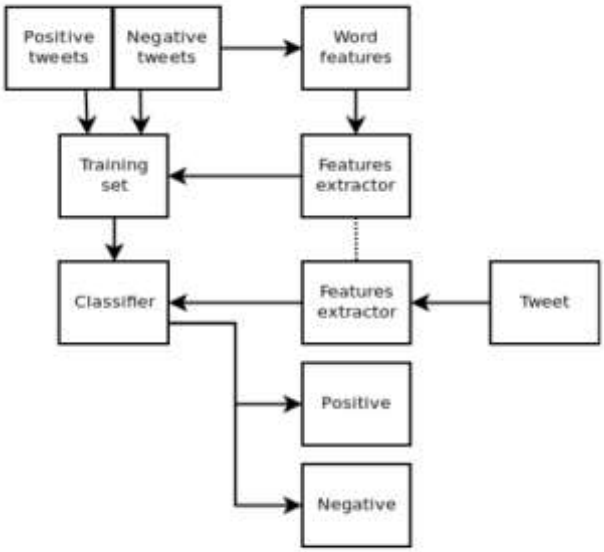
**Feature:** An attribute (or a part) of the object with respect to which evaluation is made.

**Opinion orientation or polarity:** The orientation of an opinion on a feature f represent whether the opinion is positive, negative or neutral .

**Opinion holder:** The holder of an opinion is the person or organization or an entity that expresses the opinion .

In recent years a lot of work has been done in the field of "Sentiment Analysis on Twitter" by number of researchers. In its early stage it was intended for binary classification which assigns opinions or reviews to bipolar classes such as positive or negative only.

Social networks is a rich platform to learn about people's opinion and sentiment regarding different topics as they can communicate and share their opinion actively on social medias including Facebook and Twitter. There are different opinion-oriented information gathering systems which aim to extract people's opinion regarding different topics. The sentiment-aware systems these days have many applications from business to social sciences. Since social networks, especially Twitter, contains small texts and people may use different words and abbreviations which are difficult to extract their sentiment by current Natural Language processing systems easily, therefore some researchers have used deep learning and machine learning techniques to extract and mine the polarity of the text.




**Sentiment Analysis Architecture**

**2.1 Pre-processing of the datasets**

A tweet contains a lot of opinions about the data which are expressed in different ways by different users .The twitter dataset used in this survey work is already labeled into two classes viz. negative and positive polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features. The raw data having polarity is highly susceptible to inconsistency and redundancy. Preprocessing of tweet include following points,

1. Remove all URLs (e.g. www.xyz.com), hash tags (e.g. #topic), targets (@username)
2. Correct the spellings; sequence of repeated characters is to be handled
3. Replace all the emoticons with their sentiment.
4. Remove all punctuations ,symbols, numbers
5. Remove Stop Words
6. Expand Acronyms(we can use a acronym dictionary)
7. Remove Non-English Tweets

**2.2 Feature Extraction**

The preprocessed dataset has many distinctive properties. In the feature extraction method, we extract the aspects from the processed dataset. Later this aspect are used to compute the positive and negative polarity in a sentence which is useful for determining the opinion of the individuals using models like unigram, bigram [18].

Machine learning techniques require representing the key features of text or documents for processing. These key features are considered as feature vectors which are used for the classification task. Some examples features that have been reported in literature are:

**1. Words and Their Frequencies**

Unigrams, bigrams and n-gram models with their frequency counts are considered as features. There has been more research on using word presence rather than frequencies to better describe this feature. Panget al. [23] showed better results by using presence instead of frequencies.

**2. Parts Of Speech Tags**

Parts of speech like adjectives, adverbs and some groups of verbs and nouns are good indicators of subjectivity and sentiment. We can generate syntactic dependency patterns by parsing or dependency trees.

**3. Opinion Words and Phrases**

Apart from specific words, some phrases and idioms which convey sentiments can be used as features. e.g. cost someone an arm and leg.

**4. Position of Terms**

The position of a term with in a text can effect on how much the term makes difference in overall sentiment of the text.

**5. Negation**

Negation is an important but difficult feature to interpret. The presence of a negation usually changes the polarity of the opinion

**6. Syntax**

Syntactic patterns like collocations are used as features to learn subjectivity patterns by many of the researchers.

**2.3 Training**

Supervised learning is an important technique for solving classification problems. Training the classifier makes it easier for future predictions for unknown data.

**2.4 Classification**

**2.4.1 Naive Bayes**

It is a probabilistic classifier and can learn the pattern of examining a set of documents that has been categorized [9]. It compares the contents with the list of words to classify the documents to their right category or class. Let d be the tweet and c* be a class that is assigned to d, where

$$C^* = \arg mac_c P_{NB}(c \mid d)$$

$$P_{NB}(c \mid d) = \frac{(P(c))\sum_{i=1}^{m} p(f \mid c)^{n_{i(d)}}}{P(d)}$$

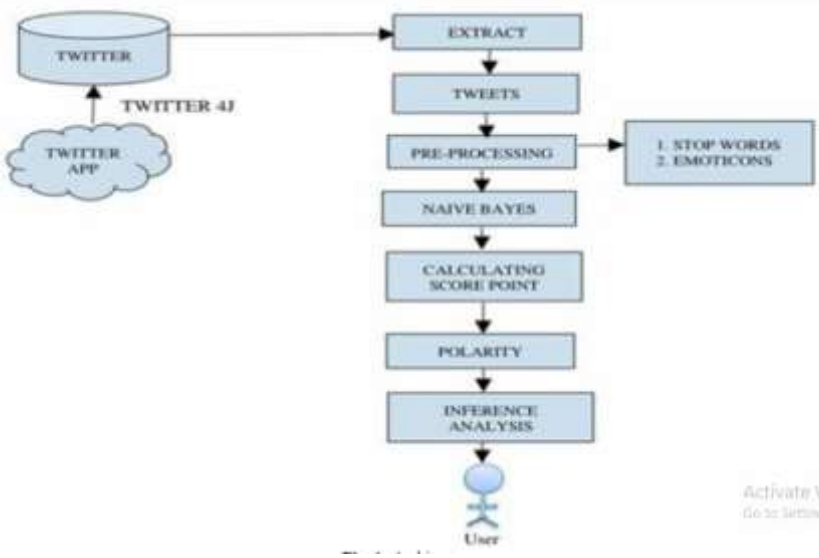
From the above equation, f‟ is a „feature‟, count of feature (fi) is denoted with ni(d) and is present in d which represents a tweet. Here, m denotes no. of features. Parameters P(c) and P(f|c) are computed through maximum likelihood estimates, and smoothing is utilized for unseen features. To train and classify using Naïve Bayes Machine Learning technique, we can use the R Language NLTK library [09]

**2.4.2 Support Vector Machine**

Support vector machine analyzes the data, define the decision boundaries and uses the kernels for computation which are performed in input space [12]. The input data are two sets of vectors of size m each. Then every data which represented as a vector is classified into a class. Neatly we find a margin between the two classes that is far from any document. The distance defines the margin of the classifier, maximizing the margin reduces indecisive decisions. SVM also supports classification and regression which are useful for statistical learning theory and it also helps recognizing the factors precisely, that needs to be taken into account, to understand it successfully

**3 Methodology**

The methodology includes the Twitter Extraction, Classification of the tweets, Sentiment Analysis, Polarity Prediction. Module Description: Twitter Extraction-It facilitates interrelation among the system and the client. The client has privileges to create account and access his/her feeds from the system. Classification of the tweets-Using Naïve baye's algorithm, the data can be analyzed and can recognize the patterns used for the classification and regression analysis. After pre-processing the extracted data, it is then classified into keyword related tweets. It classifies and predicts the group and clusters according to the user group. The same clusters are grouped under a classification. All the positive tweets are clustered into a positive classification, all the negative tweets are clustered into a negative classification, tweets containing a mixture of both the positive and the negative clustered into a mixed classification and all the neutral tweets are clustered into a neutral classification. The classification of the tweets makes easier to find the score point for each classification. And hence the polarity of each classification can be represented. Sentiment Analysis-The sentiment analysis is used comparatively to categories the positive, negative, mixed and neutral comments related with the text categorization. Sentiment analysis has the complexity like conveying the assumptions in different ways. In opinion texts, the lexical content might get misloaded. Intra-textual and sub-sentential reversals and negation topics can be commonly interpreted. The below are the possibilities that need to be classified as Users, Texts, Sentences (paragraphs, chunks of text), Predetermined descriptive phrases(<ADJN>,<N N>,<ADV ADJ>, etc), Words, Tweets/updates. Sentiment-oriented data sets are scope sensitive and it is challenging to create/collect the data from large domain. Representations of the sentiment needs more attention on elements to classify and scale the domain-appropriate annotated data is available or not. This work deals with the analysis of the tweets and it checks for the behavior of the tweets posted by the user. Polarity Prediction-The analysis of the tweets is done and then it is classified based on polarity of the words as positive, negative, mixed and neutral etc. The number of positive, negative, mixed and neutral tweets are identified based on the polarity. Naïve Bayes algorithm finds the polarity of the tweets by classifying it based on the positive, negative, mixed and neutral tweets. For correct predictive accuracy, set of training data is necessary. Bayes algorithm finds the polarity of the tweets by classifying it based on the positive, negative, mixed and neutral tweets For correct predictive accuracy, set of training data is necessary. Example: Analyzing of the hotel reviews, government official feedback or any social reviews. The task is to predict the overall polarity based on the users comment about a topic. Example: If the specific topic is been represented using a hashtag, it extracts the tweets which are related to that topic and after the pre-processing technique the overall polarity is been determined. For finding the polarity, first the score point for each classification should be calculated. According to the threshold value assigned for each word, similarly all the words belonging to each classification are been calculated with its threshold value and the overall score point is represented as a percentage




**4. APPROACHES FOR SENTIMENT ANALYSIS**

There are mainly two techniques for sentiment analysis for the twitter data:

**4 .1 Machine Learning Approaches**

Machine learning based approach uses classification technique to classify text into classes. There are mainly two types of machine learning techniques

**4.1.1. Unsupervised learning:**

It does not consist of a category and they do not provide with the correct targets at all and therefore rely on clustering.

**4.1.2. Supervised learning:**

It is based on labeled dataset and thus the labels are provided to the model during the process. These labeled dataset are trained To get meaningful outputs when encountered during decision-making. The success of both this learning methods is mainly depends on the selection and extraction of the specific set of features used to detect sentiment. The machine learning approach applicable to sentiment analysis mainly belongs to supervised classification. In a machine learning techniques, two sets of data are needed:

1. Training Set
2. Test Set.

A number of machine learning techniques have been formulated to classify the tweets into classes. Machine learning techniques like Naive Bayes (NB) and support vector machines (SVM) have achieved great success in sentiment analysis. Machine learning starts with collecting training dataset. Neatly we train a classifier on the training data. Once a supervised classification technique is selected, an important decision to make is to select feature. They can tell us how documents are represented.

The most commonly used features in sentiment classification are

- Term presence and their frequency
- Part of speech information
- Negations
- Opinion words and phrases

**5. EVALUATION OF SENTIMENT CLASSIFICATION**

The performance of sentiment classification can be evaluated by using four indexes calculated as the following equations:

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

F1 = (2×Precision×Recall)/ (Precision + Recall)

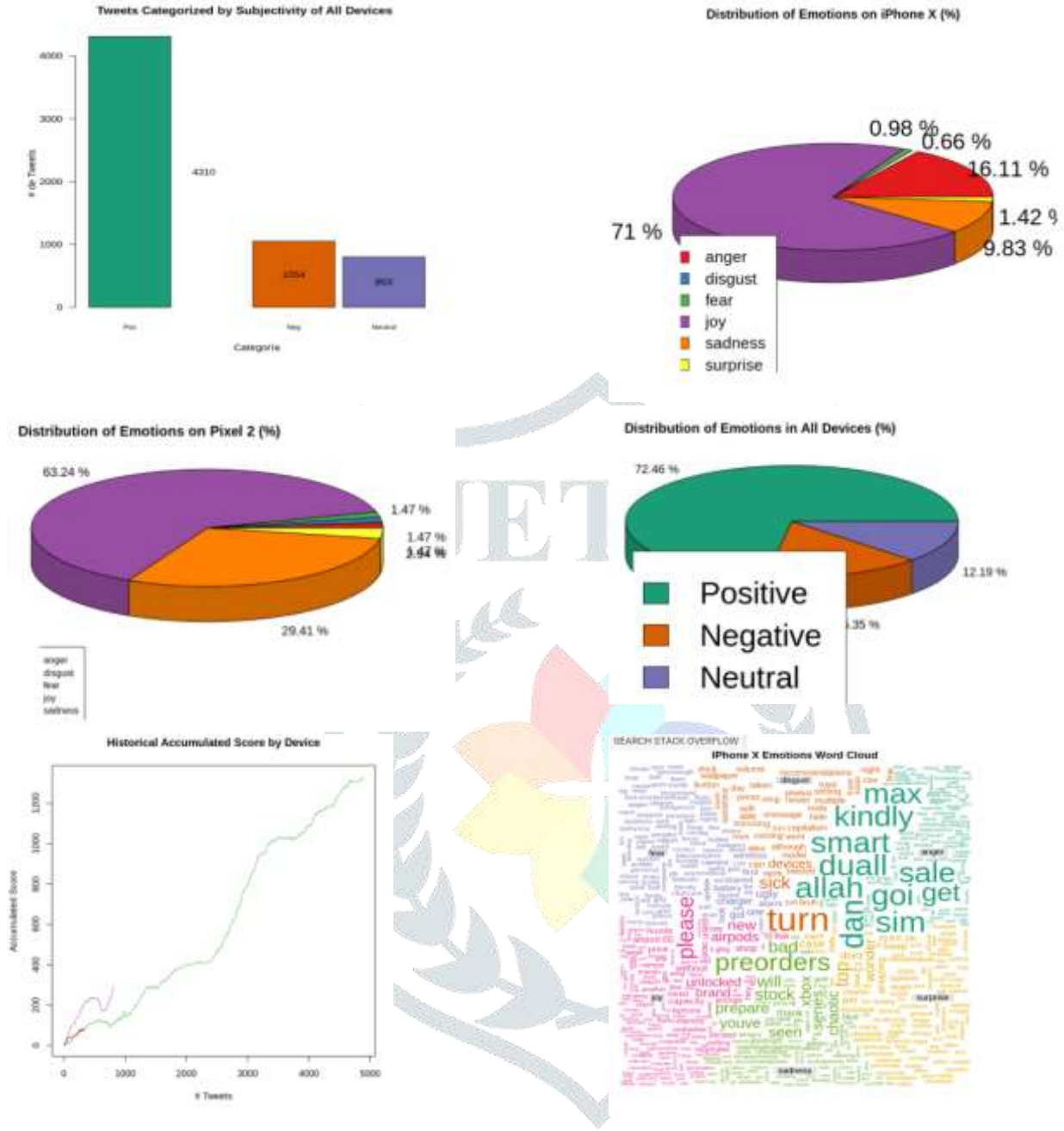
In which TP, FN, FP and TN refer respectively to the number of true positive instances, the number of false negative instances, the number of false positive instances and the number of true negative instances,

**Confusion Matrix**

| | Predicted Positives | Predicted Negatives |
|---|---|---|
| Actual Positive | TP | FN |
| Actual Negative | FP | TN |

## 6 . RESULTS AND DISCUSSION

We used the twitter dataset publicly made available by Stanford University. Analyses was done on this labelled datasets using various feature extraction technique. We used the framework where the pre-processor is applied to the raw sentences which make it more appropriate to understand. Further, the different machine learning techniques trains the dataset with feature vectors and then the semantic analysis offers a large set of synonyms and similarity which provides the polarity of the content.




## 6. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment Analysis has many applications in various Fields.

1. Applications that use Reviews from Websites:

Today Internet has a large collection of reviews and feedbacks on almost everything. This includes product reviews, feedbacks on political issues, comments about services, etc. Thus there is a need for a sentiment analysis system that can extract sentiments about a particular product or services. It will help us to automate in provision of feedback or rating for the given product, item, etc. This would serve the needs of both the users and the vendors.

2. Applications as a Sub-component Technology

A sentiment predictor system can be helpful in recommender systems as well. The recommender system will not recommend items that receive a lot of negative feedback or fewer ratings.

In online communication, we come across abusive language and other negative elements. These can be detected simply by identifying a highly negative sentiment and correspondingly taking action against it.

3. Applications in Business Intelligence

It has been observed that people nowadays tend to look upon reviews of products which are available online before they buy them. And for many businesses, the online opinion decides the success or failure of their product. Thus, Sentiment Analysis plays an important role in businesses. Businesses also wish to extract sentiment from the online reviews in order to improve their products and in turn their reputation and help in customer satisfaction.

4. Applications across Domains:

Recent researches in sociology and other fields like medical, sports have also been benefitted by Sentiment Analysis that show trends in human emotions especially on social media.

5. Applications in Smart Homes

Smart homes are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device. Recently there has been lot of research going on Internet of Things(IoT). Sentiment Analysis would also find its way in IoT. Like for example, based on the current sentiment or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment.

Sentiment Analysis can also be used in trend prediction. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.

**Implication:** From the implications of the results, it is proved that for better up-to-date analysis the streaming data can be used. It gives better results when compared to static data. The result proves that the sentiment analysis is been done and can predict the overall polarity of the tweets, the user who wishes to see the review can infer from the analysis of the data using the resultant graph and it gives the user an efficient result about that topic. In this paper, the streaming data is used for the analyzing the tweets and finding its polarity. This paper uses streaming data for social media analysis. For extracting the streaming data, Twitter 4j is used. Twitter 4j: Twitter 4j is an unauthorized java library for the twitter API and has android platform and google app engine. Using this API, integrating the java application can be done using the twitter support. Using this, the streaming data can be easily collected from the Twitter Database (TDB). Extract tweets: Here the tweets are been extracted from the twitter. Extracting the tweets are necessary to do the sentimental analysis. The tweets extracted is based on the users posting the tweets according to the topic. According to the input given, the tweets are collected from the twitter service. Streaming data related to the topic are collected. Extracting the trending tweets based on the latest news in the current social media any amount of tweet can be extracted and can be modified into any form. Our work deals with extracting the streaming data according to the input given. The input by the user can be based on a trending topic or also based on a regular topic. The tweets which are given by the user in the social media using hashtags etc. in the twitter. The input is been represented as a hashtag followed by the topic. Our project has the ability to collect the tweets of 50 users for all topic.

## CONCLUSION

In this paper, we provide a survey and comparative study of existing techniques for opinion mining including machine learning and lexicon-based approaches, together with cross domain and cross-lingual methods and some evaluation metrics. We discussed the importance of social network analysis and its applications in different areas. We focused on Twitter as and have implemented the R Language program to implement sentimental analysis. We present a survey of existing techniques for sentiment analysis using machine learning algorithms. As the twitter users are increasing day by day and the posts shared by the users were tweets. Analyzing the twitter dataset can be beneficial. Many techniques were suggested for sentiment analysis, but the problem of sarcasm still not persist. A lot of researches have done the summarization of events, real-time event detection as well as sentence-based sentiment classification accurately and efficient. Naive Bayes classifier is insensitive to un-balanced data which give more accurate results we showed the results on different daily topics. We realized that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis. Research results show that machine learning methods, such as SVM and naive Bayes have the highest accuracy and can be regarded as the baseline learning methods. We also studied the effects of various features on classifier. This work proposes a system for categorizing the results based on the polarity analysis of the data. The data which are extracted from the twitter database will be a streaming data. The prediction gives more accurate results and quick response using the social network based behavioral analysis. The limitation of this work is that the languages other than English are not considered for the sentiment analysis

## FUTURE WORK

We can conclude that more the cleaner data, more accurate results can be obtained. Use of bigram model provides better sentiment accuracy as compared to other models. We can focus on the study of combining machine learning methods in order to improve the accuracy of sentiment classification and adaptive capacity to variety of domains and different languages. This work can be further extended by improvising the sentiment analysis for various other social media like Instagram, Face-book. In future, the sentiment analysis can also be done in the languages other than English. In this work estimating the polarity and the calculation of the sentiment cannot be done if it is in the language other than English. In our work, the sentiment analysis can be done only in English and the future work can be extended by using other languages also

## REFERENCES

[1] Levy, M. (2016). Playing with Twitter Data. [Blog] R-bloggers. Available at: https://www.r-bloggers.com/playing-with-twitter-data/ [Accessed 7 Feb. 2018].

[2] Popularity Analysis for Saudi Telecom Companies Based on Twitter Data. (2013). Research Journal of Applied Sciences, Engineering and Technology. [online] Available at: http://maxwellsci.com/print/rjaset/v6-4676-4680.pdf [Accessed 1 Feb. 2018].

[3] Zhao, Y. (2016). Twitter Data Analysis with R – Text Mining and Social Network Analysis. [online] University of Canberra, p.40. Available at:https://paulvanderlaken.files.wordpress.com/2017/08/rdatamining-slides-twitter-analysis.pdf [Accessed 7 Feb. 2018].

[4] Alrubaiee, H., Qiu, R., Alomar, K. and Li, D. (2016). Sentiment Analysis of Arabic Tweets in e-Learning. Journal of Computer Science. [online] Available at: http://thescipub.com/PDF/jcssp.2016.553.563.pdf [Accessed 7 Feb. 2018].

[5] Qamar, A., Alsuhibany, S. and Ahmed, S. (2017). Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies. (IJACSA) International Journal of Advanced Computer Science and Applications, [online] 8. Available https://thesai.org/Downloads/Volume8No1/Paper_50-Sentiment_Classification_of_Twitter_Data_Belonging.pdf [Accessed 1 Feb. 2018].

[6] R. M. Duwairi and I.Qarqaz, "A framework for Arabic sentiment analysis using supervised classification" , Int. J. Data Mining, Modelling and Management, Vol. 8, No. 4, pp.369-381 , 2016.

[7] Hossam S. Ibrahim, Sherif M. Abdou, Mervat Gheith, "Sentiment Analysis For Modern Standard Arabic And Colloquial", International Journal on Natural Language Computing (IJNLC), Vol. 4, No.2, pp. 95-109, April 2015.

[8] Assiri, A., Emam, A. and Al-Dossari, H. (2016). Saudi Twitter Corpus for Sentiment Analysis. International Journal of Computer and Information Engineering, [online] 10. Available at: http://waset.org/publications/10003483/saudi-twitter-corpus-for-sentiment-analysis [Accessed 1 Mar. 2018].

[9] L. Wasser and C. Farmer, "Sentiment Analysis of Colorado Flood Tweets in R", Earth Lab, 2018. [Online]. Available: https://earthdatascience.org/courses/earth-analytics/get-data-using-apis/sentiment-analysis-of-twitter-data-r/. [Accessed: 01- Mar- 2018].

[10] D. Robinson, "Text analysis of Trump's tweets confirms he writes only the (angrier) Android half", Variance explained, 2016.

[11] _____"A Common Database Interface (DBI)", cran.r-R, 2003. [Online]. Available: https://cran.r-project.org/web/packages/DBI/vignettes/DBI-1.html. [Accessed: 25- Mar- 2018].

[12] V. Kharde and S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications, vol. 139, p. 11, 2016.