

# Applications of Social Media Analytics Using Machine Learning

Manpreet kaur

Department of Computer Engineering & Technology  
Guru Nanak Dev University  
Amritsar, Punjab, India

Amit Chhabra

Department of Computer Engineering & Technology  
Guru Nanak Dev University  
Amritsar, Punjab, India

**Abstract-** Now a day's with the increased use of social networking, more people are showing curiosity towards social media for information and opinion of other users. Facebook is the most popular social networking site that records large amount of user behavior expressed in various activities such as facebook likes, status updates, comments, photos. As the use of social media increases the importance of analyzing its produced data also increases. Machine learning is the application of artificial intelligence that provides systems to automatically learn from experience without being explicitly programmed. The data on facebook and other online websites can be analyzed using different machine learning techniques.

**Keywords:** Social media, Sentimental Analysis, Machine learning techniques.

## I. Introduction

Machine learning is a subfield of artificial intelligence. The goal of machine learning is to understand the structure of data and fit that data to models that can be understood and utilized by the people. Social network sites are known for information spreading, product reviews, advertisements, sentiment expression. Social networking sites are commonly used as a means of communication and interaction between friends. In social network analysis from the product ratings and reviews generated by the user we may come to know about the positive and negative reviews. Sentimental analysis can be analyzed from the opinion expressed in content generated by user. It is observed that for real time information people are more dependent on social media.

Different machine learning techniques can be used for the social network analysis. Opinion mining is also called sentimental analysis to know the opinion of users through opinion finder tool which is used to analyze different sentences related to any topic and classification is done based on their polarity.

## II. Literature Survey

Gargi kulkarni et al.[1], describes the system that summarizes text messages regarding transportation in California on social media platform based on machine learning. The system uses unsupervised statistical topic modeling algorithm to cluster public messages.

Xinyue wang et al.[2], proposes readability features, topic features and apply supervised machine learning algorithms for performing real life data. Isidoros perikos et al.[3], a graph based method is utilized to model the emotional level of a topic based on the emotions recognized and then it creates the topic's emotional graph visualizing public emotions and mood on the topic. Sneh Paliwal et al.[4], analyze the sentiment analysis using artificial neural network. By using text blob library, the polarity and sentiment of tweets was recorded. Banita verma et al.[5], uses lexicon based approach and machine learning based approaches to analyze the sentiments. Ashar khan et al.[6], analyze the mental state of user to predict depression by analyzing the opinions of the users using machine learning techniques. P.Srinivas Rao et al.[7], identify the fake profile of the user by using support vector machine and naive bayes machine learning algorithms.

Fatih ertam [8], the data taken from facebook is analyzed by machine learning approaches and compare with performance metrics, logistic regression, random forest were used for machine learning approaches. Rita georgina et al.[9], For classification of adult age group, convolution neural network had the best performance among different machine learning algorithms. Xiaoxue zang et al.[10], in this convolution neural network is used to extract visual features and proposes a new method to rank users attractiveness from their online dating interactions. Anees UI Hassan et al.[11], this paper presents how to find depression level of a person by extracting emotions from text using different machine learning techniques.

Elsharif elmurungi et al.[12], uses weka tool to analyze the dataset of movie reviews. supervised classifiers are used to detect fake reviews. Devendra K Tayal et al. [13], uses naive bayes and maximum entropy machine learning algorithms to analyze the sentiments of the users. Peng yang et al.[14], analyze the sentiments of the users by using artificial neural network by back propagation method. Sanketi P.Raut et al.[15], describes chinese spam filtering approach delivered in text of emails. Naive bayes and apriori algorithm are used to detect spam reviews.

Perna mishra et al.[16], collected sentiments and classified polarity of sentiments as positive negative and neutral. Twitter API is used for analysis of twitter data. D.Y.Patil et al.[17], a novel technique is proposed of opinion mining of

product reviews which encourage the customer to choose the right product based on natural language processing. Magesh.S et al.[18], presents a survey of big data analytics techniques applied on social network and also presents the pros and cons of each algorithm. Vikash singh rajput et al. [19], on the basis of supervised and unsupervised learning opinions are judged. This paper gives a comparative study of naïve bays and SVM on the sentiment expressions of reviewers of stock market. Anuja P Jain et al.[20], step by step detail is given about the process of sentimental analysis. Text analysis of twitter data is done using naïve bayes and decision tree learning algorithm. Rajesh KM et al.[21], proposes a method to detect face recognition and face emotions using support vector machine. Suheel Yousuf wani et al.[22], uses weka tool to predict fake profiles. By using training data, a model is prepared to make predictions. Rohit joshi et al.[23], with the help of twitter API from twitter collected a corpus of positive, negative, and neutral tweets. Twitter data is analyzed using supervised classifiers.

Vinay A et al.[24], a survey of concept of cloud computing, big data , social network and machine learning is done. Deep neural network is used for face recognition on social media. Girija V Attigeri et al.[25], analyses of stock prediction is done with technical analyses and fundamental analyses and with five modules i.e data gathering, preparation, sentiment analysis, aggregation. Cao Xiao et al.[26], describes approach to find group of fake accounts. By using raw list of accounts, cluster builders builds clusters of accounts. Hao wang et al.[27], analyze frequency of emoticons to demonstrate prevalence of emoticons on social media. Sentiment polarity of groups of participants are surveyed. Abinash Tripathy et al.[28], uses support vector machine and naïve bayes to classify sentimental reviews, a separate text file is maintained for each review. Kolli Shivagangadhar et al.[29], uses linguistic features and review length to identify fake reviews and build model to detect online reviews.

Sagar bhutta et al.[30], uses lexicon based approach and learning based methods for sentimental analysis. Geetika Gautam et al.[31], performs pre processing of data, here analysis of data becomes easy if labeled data has negative and positive polarity. Mariam adedoyin-olowe et al.[32], uses nodes and links to represents a network. Clustering technique is used to compress a large network to a smaller group to form community. Neethu M S et al.[33], analyze the twitter posts about electronic products using machine learning approaches. In sentiment classification, the effect of domain information can be achieved. Morgan Madeira et al.[34], uses linear regression and logistic regression techniques to identify facebook interactions and predict a users closest friend with 85% accuracy. Jalel Akaichi [35], extract useful information about users sentiments during Tunisian revolution associated period. Support vector machine and naïve bayes algorithms are used for sentiment analysis.

Balakrishnan Gokulakrishnan et al.[36], proposes an approach from twitter microblogging sites where tweets are preprocessed based on emotional content as positive and negative. Radoslaw Michalski et al.[37], proposes time forecasting and regular classification approach, and two parameter measures prediction error and time of prediction are used. Unnamalai.K et al.[38], describes sentimental analysis from unstructured data that can be explicit, implicit and associative. Pawel Sobkowicz et al. [39], uses three building block of tracking online opinions and uses natural language processing to show similarities between texts and degree of positive, negative opinions. Nan li et al.[40], uses clustering algorithms to group forum into clusters, within the current timespan.

### III. Use of Machine Learning Techniques for Social Media Analysis

Machine learning is used in a wide range of applications today. One of the most well-known is social networking sites. In this section we discuss various machine learning applications to analyze data from social network sites.

#### 1. Community Detection using hierarchical clustering

Clustering is the task of dividing the data points into a number of groups such that data points in the same group are more similar to other data points in the same group than those in the other group.

One of the most important characteristics of social network sites is formation of communities[32]. Similar interest users form one community displaying strong sectional structure. In social network communities are complex and difficult to detect. By the use of clustering technique community formation can be done.

- Nodes represents the users and links represents the relationship with each other.
- Similar interest users form one community using clustering technique.

Nan li et al. [40] uses an approach which contains following steps: data collection, cleansing, text sentiment calculation and marking, uses k means clustering for hotspot detection and SVM for hotspot forecast. Unsupervised clustering algorithm is applied to group forum into clusters, within the current time span the center of each cluster represent a hotspot forum and SVM is used to realize hotspot forecasting.

#### 2. Analysis of data on Social Media

The large amount of data on the internet can be analyzed using different machine learning approaches. Logistic Regression and Random forest[8] are the machine learning approaches used to analyze the data taken from facebook.

**Logistic regression:** It is a classification approach used for binary approaches. Its aim is to find the probability of obtaining another dependent variable by using independent variables. The result is always between 0 and 1.

Balakrishnan Gokulakrishnan et al.[36] proposes an approach from twitter micro blogging sites where tweets are preprocessed based on emotional content as positive or negative. Random forest is used for classifying big data. It acts as a

community learning method using decision tree. Each sample vector is decided on each tree of forest to classify the data. Sequential mining optimization solve the optimization problem by breaking it into series of sub problems.

Sneh Paliwal et al. [4] uses twitter API and facebook API performs sentiment analysis for iphonex and also collect data from news websites using python. then by using text blob library, the polarity and sentiment of tweets was recorded, feed forward back propagation is used. In the ratio of 80:20 the collected data is partitioned. To split the data into train and test data, the train data is used and to determine the accuracy of result, the test data is used which is predicted by the artificial neural network.

To analyze the twitter data using supervised classifier authors in [23] with the help of twitter 4j java from twitter collected a corpus of positive and negative tweets. To make collected corpus free from commas, remove stop words. Then machine learning algorithms are applied: Naive bayes is flexible and can handle number of attributes and it is the simplest probabilistic model. Maximum entropy outperforms naive bayes for text classification. It is used in applications of natural language processing. Unigram, bigram, hybrid features are used. Performance measures precision, recall are used. By using hybrid features SVM outperforms all other classifiers.

### 3. Sentiment analysis on Social network

Opinion mining is also known as sentiment analysis with which content generated by the user can be analyzed such as comments, posts, people thoughts, ratings towards some products and topics. Twitter contains large number of messages and is very convenient for research. Twitter API is used for analyzing the data collected from twitter. Opinion finder tool is used for analysis of sentences related to any topics and based on their polarity classification is done[16].

Now a day's people express their views through posts, online discussion, and product reviews websites. Neethu et al. [33] used Naïve bayes, Support vector machine to classify reviews. Naive bayes classifier uses feature vector which are independent of each other and analyzes them individually. SVM uses hyperplane to separate the tweets. In this sentence level sentimental analysis is done in three phases: Firstly pre processing is done. Then using relevant feature, feature vector is extracted. Finally Tweets are classified into different classes using different classifiers. The feature vector performs well for electronic product domain. Sagar bhutta et al. [30] uses sentimental classification technique for sentimental analysis:

**Lexicon based approach:** In this number of positive and negative words are counted in the text. If more positive words are there then positive score is given to the text. If more negative words are there then negative score is given. Neutral score is assigned If equal number of positive and negative words are there. The words which express opinion are called opinion words.

Syntactic and statistical technique can also be used. Syntactic rules can be used to indicate nouns, adjectives, etc. To focus on categories of words, statistical technique is used which is independent of languages. Banita et al.[5] uses lexicon based approach which uses lexicon and unlabeled data. Based on opinion lexicon, words in the text are evaluated to determine their orientation and sentiment of text. It is performed by manual approach in which opinion words are manually collected based on individual knowledge and Dictionary based approach which is used to determine word sentiment which uses synonyms and antonyms.

**Machine learning based approach:** The data collected for sentiment analysis may be noisy and needs to be preprocessing using various natural language processing techniques and finally classifier is trained and tested on unseen data. Machine learning classifiers used to trained model are naïve bayes, support vector machine, decision tree[20].

When preprocessing of data is done, the analysis of data becomes easy if the labeled data has negative and positive polarity. Then features are extracted and three supervised techniques are used i.e. naive bayes, SVM, maximum entropy are used. Authors in [31] uses semantic analysis in which each term is associated with each other. The words are semantically similar if two words are close to each other.

Radoslaw Michalski et al.[37] proposes time series forecasting and regular classification approach. It forecast value of network by analyzing time series. Two parameter measures are used i.e. prediction error and time of prediction. In regular classification approach there were classes defined in which values which are measured could fit. 10 fold cross validation method is used for validation. It can be predicted that better forecasting results are seen from time series period.

**Decision tree learning approach:** Two types of trees can be built: classification trees and regression trees. Depending on decision tree formed, clustering of users can take place. ID3 and c4.5 are used to obtain the relationship among users and also used to discover interesting pattern among the user.

To analyze the sentiment of users, support vector machine is used as it is a classification model and because of sparse nature of text, Text data is suited for SVM where few features are not relevant and is organized into linearly separable category. Naive bayes is based bayes theorem. Artificial neural network handles real data and is powerful tool for prediction, Back propagation neural network is considered to be the best network[14]. Unnamalai.k [38] describes sentiment analyses from unstructured data that can be of three types i.e. explicit, implicit, associative. Stemming is the first process used to reduce words to their root, it can be viewed as recall-enhancing device. The basic element in information retrieval system, query system, opinion mining, are the stemmers. It also helps in reducing the redundant terms. Sentiment classification orientation is used to determine the polarity of opinions.

#### 4. Face recognition on social media

We upload a picture with a friend and facebook instantly recognizes that friend facebook checks the poses and projections in the picture, notice the unique features and then match them with the people in our friend list. Vinay A et al. [24] uses machine learning technique: Deep neural network which is used to recognize human behavior through face recognition. Deep learning operates by modeling high level abstraction based on how human brain recognize pattern. Deep Neural Networks model high level complex data and multiple features. Deep faces perform effective recognition in the variations in term of camera angle. Deep face utilizes 3D modeling technique to rotate an image in 3 dimensions to visualize various angles of face. Deep face used to recognize human faces.

Rajesh KM et al.[21] uses SVM to recognize face emotions for finding different emotions of face. SVM is used to find minimum possible separation between two classes of data. For face recognition, the input image can be still images, video. The images used will be save in database. To identify users identity, matching of input image is performed. For classification of different users algorithm used is fishers algorithm which generates fisher faces used for recognition. SVM is used to predict emotions of face by comparing features of training and testing data. By increasing the number of images during training, the accuracy of face recognition and emotion detection can be increased.

#### 5. Online dating websites

Online dating websites[10] are popular platform for adults to search for their life partners. An important factor to determine others impression is through users profile image. To analyze visual attractiveness the main focus is on profile image. Users send contact to those whom they find attractive to them. At receiver side, they select those contacts that they want to reply.

**Age Prediction:** Classical age estimation uses two steps: feature extraction and age prediction. Convolution neural network is used in facial Age estimation, feature extraction and classification can combine into one procedure. The result proved that using deep learning models we can gain better performance than traditional methods.

**Facial attractiveness:** In analysis of facial attractiveness the important part is the facial representation. To analyze the geometric and spectral feature for facial attractiveness the convolution neural network performs better for facial beauty prediction.

#### 6. Stock Market sentiment analysis based on machine learning

Vikash singh rajput et al. [19] in this paper author presents that product rating and reviews often contain sentiment expressions of the user generated content. Most online stores allow their customers to review their product. Opinion mining is used for identification of reviews of users about particular product. On the basis of supervised and unsupervised learning, public opinions can be judged. In this support vector machine is used on the opinions of reviews of stock market. Support vector machine is a supervised learning technique used to analyze the data and identifies the pattern for classification by finding the hyper plane that maximizes the margin between two classes. The opinions of the users affect a lot to the service provider. The supervised and unsupervised methods help to find the result in better way[34].

Two types of analysis for stock prediction is used: technical and fundamental[25]. By using technical analysis, the historical data of stock prices is done with machine learning. Five modules are used for prediction of stock market as data gathering, data preparation, sentiment analysis, aggregation, visualization. Logistic regression technique is used whose outcome is determined by independent variables in categorical form.

#### 7. Identification of fake reviews on social media

Now a day's purchasing decision is totally dependent upon online reviews of customers. Fake reviews is a major problem which can be analyzed through behavior features, spammers and non-spammers. For deep analysis spammers post fake reviews, non spammers product truthful reviews. Main focus is on the behavioral features of spammers, as it difficult to collect textual features. Classifiers used to analyze fake reviews are: logistic regression, k-nearest neighbor, naïve bayes and support vector machine[2]. To classify reviews as false or benign, N-gram features are used. It includes unigram, bigram, trigram features with help of which false and true reviews can be analyzed. Classifier of logistic regression performs the best, which achieves up to 97.2% accuracy.

Elshrif Elmurngi et al.[12] uses weka tool to analyze dataset of movie reviews. Sentiment classification algorithms are applied using without stopwords and with stopwords methods. With stopword method is more efficient in text categorization method and to detect fake reviews. Naive bayes classifiers based on applying bayes theorem. It has fast decision making process. SVM is a supervised learning model which identify patterns and examines data used for classification analysis. To find out opinions, sentiment analysis is used which might be positive, negative or neutral. Authors in [13] uses naive bayes, maximum entropy and SVM classifiers. Naive bayes: The classes that have highly dependent feature, this approach is optimal for them. Maximum entropy is used in natural language processing application and is an efficient technique. Based on SVM, an incremental algorithm provides good results.

Authors in [28] for online movie review a separate text file is maintained for each review. Vague information will be eliminated from reviews special characters are also removed, character repetition also gets eliminated. After cleaning,

feature extraction take place. naive bayes and SVM algorithms are used. In naive bayes numeric vectors are used to extract features. Value of precision, recall are found for each model.

In affecting consumers actions and decision making online review system plays an important role. Authors in [15] describes Chinese spam filtering approach delivered in texts of emails and provide method to identify spam reviews. Naive bayes and apriori algorithm are used to detect spam reviews. To detect data Naive bayes and logistic regression is an intelligent approach on the basis of single word. It check keyword to classify a review as spam or not spam. Linguistic features are used which are unigram bigram frequency. Training data is used to train the logistic regression, naive bayes. These techniques are trained separately to generate detection accuracy[29].

### **8. Modeling emotions on social media**

Emotion recognition and public emotion modeling are used for modeling emotions on social media. Authors in [3] have used ensemble classifier. In emotion recognition the data generated by the user is recognized by ensemble classifier schema that analyzes data and recognize the existence of emotions. For deep analysis of natural language sentences, knowledge based tool is used. It stores emotional words that convey emotions. The modeling method uses graph based approach to recognize the emotion level of the user generated content. It is the structural way of representing the public emotions. The result indicates that performance of ensemble classifier schema is 4% better as compared to performance of other classifier.

Authors in [27] analyze frequency of emoticons to demonstrate prevalence of emoticons on social media. Four analysis were carried out i.e. firstly sentiment polarity of group of participants are surveyed, secondly to better understand the meaning of emoticons, clustering of words are examined. Then as emoticons were removed from the text, sentiment polarity of post are compared. the result shows that few emotions are strong and advantage can be taken in sentiment analysis some emoticons shows complicated sentiments. Three building blocks of tracking online opinion: emotion detection in real time, information flow modeling, modeling of opinions[39]. Natural language processing is implicit representation between texts. Semantic web approach is based on semantic annotation and is explicit representation.

### **9. Close friend interaction on social media**

Morgan madeira et al. [17] have used machine learning techniques to predict user's closest friends on facebook. In this weka tool is selected to predict user's closest friend. Three different algorithms used for analysis are: Bayesian network, sequential minimal optimization and naive bayes decision tree.

In Bayesian network the facebook features are considered as attribute variable and existence of close friend relationship represent class variables. The optimization algorithm assigns the correct binary label to vector of facebook. The naive bayes decision tree uses naive bayes classifier which calculates utility of each leaf in the tree. To accurately classify strong and weak ties between facebook users Bayesian network is the best network.

### **10. Measurement of depression on social network**

Support vector machine is the machine learning technique used to measure the features of depression on social networking sites. SVM selects best features from training data and then these features passes to artificial neural network to perform classification. Anees UI Hassan et al. [11] uses SVM model to extract three features of the depressed people: user micro blog text, user profile emotion theories, user behavior to describe user situations. SVM and naive bayes classifiers are used. For text classification, SVM is the best performing method. It is used for data analysis in different domains. Maximum entropy converts labeled feature into vector using coding. Vector is used to calculate weights for each label which then combine to determine most likely label.

The mental state of user is analyzed by analyzing users opinion, news, through social media platforms. Two methods to access behavior by measuring behavior through sensors and by technology. Supervised machine learning algorithms are used to access behavior of humans.

### **11. Facebook status updates mining for sentiment analysis**

For sentiment classification author uses text mining[35]. In this useful information is extracted about user sentiment during Tunisian revolution associated period. At that time users share their feelings and condition of the cities with their friends by sharing videos, pictures. To identify nature and statuses updated own dataset is created and then naive bayes and support vector machine learning algorithms are applied. SVM classifier is used because of its best performance and many text classification studies uses SVM classifier.

### **12. Online community input to transport system**

As the use of online social media increases, it is possible to summarize the rider's opinion in the social media messages. A system based on machine learning is used for the text messages regarding transportation in California. Topic modeling technique is used to cluster public messages on the twitter related to transportation[1]. Then based on the polarity of each message, sentimental analysis is done. The system defines sentiment toward each topic.

### 13. Fake account detection

Authors in [26] describes approach to find group of fake accounts. To classify an entire cluster of fake accounts as malicious or legitimate. Technique used is supervised machine learning. The key features include frequency of patterns and comparisons of text frequencies. To find parameter using maximum likelihood, logistic regression is used. SVM is used to find optimal hyperplane in high dimensional space. Random forest is used to form strong classifier by combining many weak classifiers. By using raw list of accounts, cluster builder builds clusters of accounts. Profile feature have three categories: Basic distribution features, pattern features, frequency features.

By using training data a model is prepared to make predictions. To achieve desired level of accuracy, this process continues. Suheel yousuf wani in [22] uses ensemble classifier to make results. Weka tool is used to predict fake profiles. SVM separate all data points of one class from other by finding best hyperplane. Decision tree build regression model in tree structure form. Artificial neural network is system of interconnected neurons which exchange message each other. Here, 5-fold cross validation is used and efficiency of model will be calculated by false positive and false negative. To minimize indexing record natural language processing is used. In this SVM, naive bayes are used. To find hyperplane the separate information from other classifier, SVM technique is used. Naive bayes is a probabilistic classifier and is used to classifier profile.

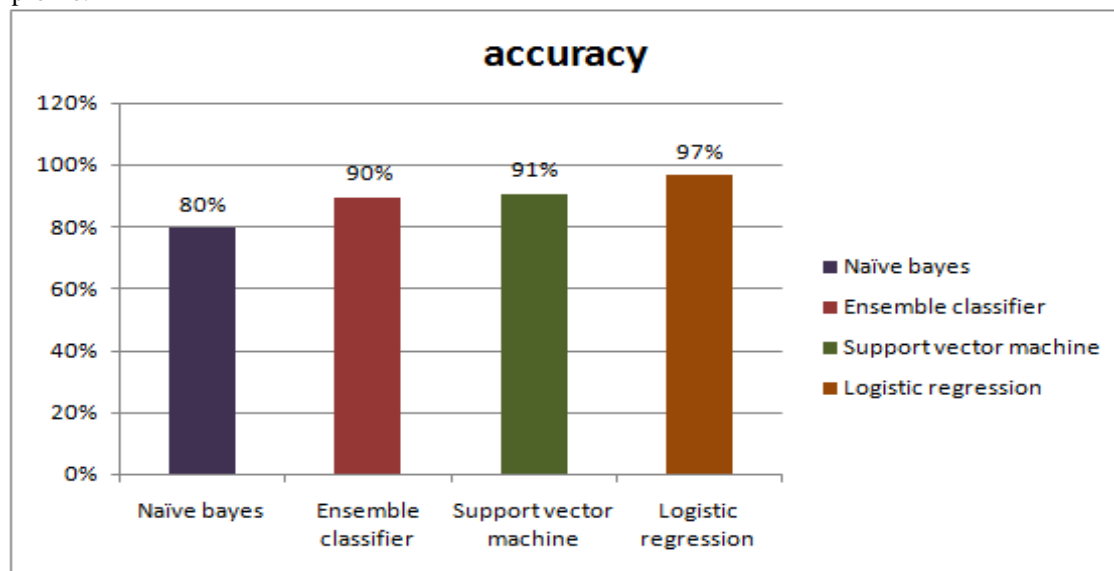


Figure 1: Performance of various techniques

A few inferences that can be made from the graph in the figure2:

- Naive bayes and SVM perform better in case of analysis of data on social network sites for measurement of depression.
- Logistic regression performs best for the identification of fake reviews which acquires 97% accuracy for all features.
- Ensemble classifier perform better than naive bayes and SVM. To do the best classification ensemble classifier make use of features of all base classifiers.
- Ensemble classifier perform 90% accurately to analyze sentiments in twitter data.

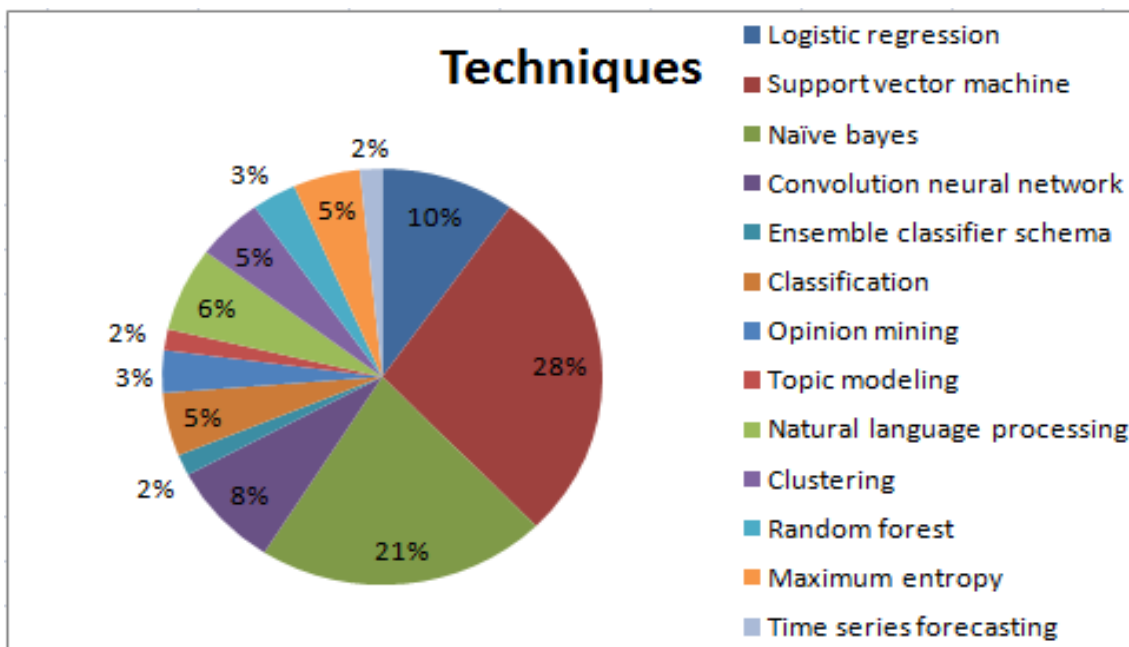


Figure 2: Machine Learning Techniques used in Social Media Analysis

The graphical representation of techniques can be represented using pie chart as shown in figure2. By this we can clearly see that the highest percentage of techniques is 28% of the support vector machine, naive bayes and that of logistic regression is 10%. All other techniques used in this paper are shown in figure2.

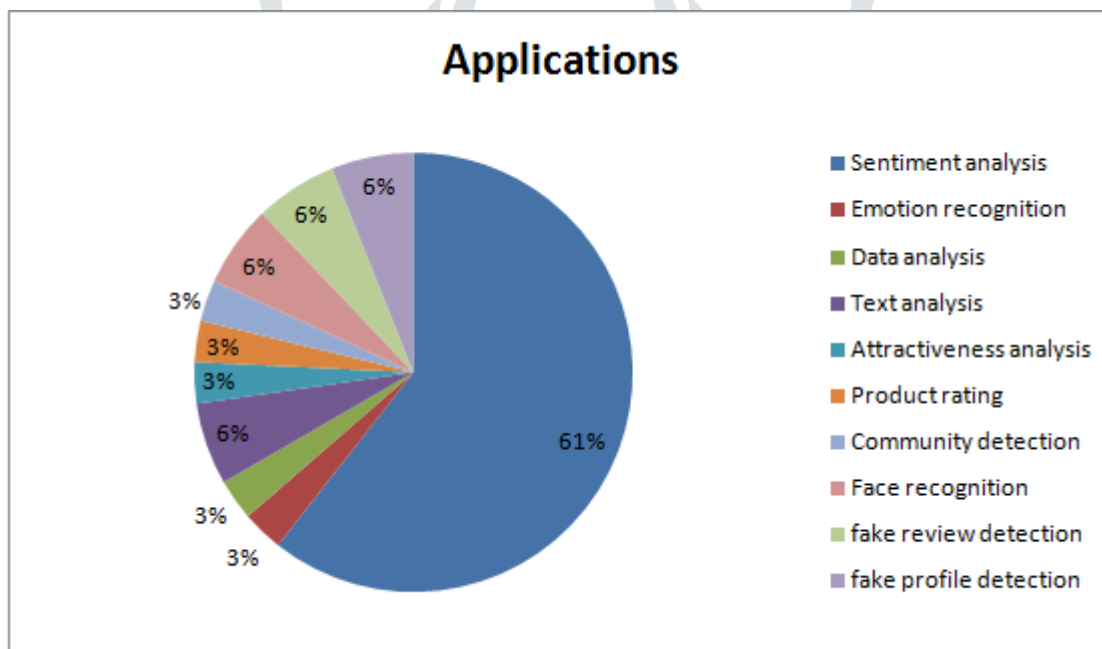


Figure 3: Applications shown in percentage

The graphical representation of applications can be represented using pie chart as shown in figure3. The percentage of sentiment analysis is 61% which is the highest among other applications used in this paper.

Table 1. Different Applications and Machine Learning Techniques used in this paper are listed below:

Title of paper	Reference	Year of publication	Application used	Technique used	Highlights of outcome

Unsupervised classification of online community input to advance transportation system	[1]	2018	Sentiment analysis	Classification, Topic Modeling	Online community analyzes data on social media and provides transport related topic and sentiment of online users.
Identification of fake reviews using semantic and behavioral features.	[2]	2018	Sentiment analysis	Logistic Regression, SVM	Fake reviews can be analyzed through user generated content by using logistic regression technique.
A framework for analyzing big social data and modeling emotions in social media	[3]	2018	Emotion recognition	Ensemble classifier schema	The content generated by user is recognized by classifier that recognizes the existence of emotions. Knowledge base tool stores emotional words that convey emotions.
Sentiment analysis and prediction using neural networks	[4]	2018	Sentiment analysis	Artificial neural network	Feed forward back propagation is used to analyze the sentiment of users.
Sentiment analysis using lexicon and machine learning based approach: a survey	[5]	2018	Sentiment analysis	Lexicon based approach	Lexicon based approach is used to determine sentiment of text.
Analysis of mental state of users using social media to predict depression! a survey	[6]	2018	Sentiment analysis	Support vector analysis	SVM is used to access behavior of humans. Word count tool is used to analyze text.
Fake profile identification in online social network using machine learning technique	[7]	2018	Sentiment analysis	NLP,SVM, NB	Fake profile can be identified by using supervised machine learning algorithms.
Analysis of data using machine learning approaches in social networks.	[8]	2017	Data analysis	Logistic regression, Random forest	RF and LR have been chosen as classifier because of their rapid performance in analyzing big data.
Age group classification in social network using deep learning.	[9]	2017	Text analysis	Classification , Convolution neural network	Based on the users writing characteristics and similar topic, behavior can be analyzed by using convolution neural network.
Prediction of user facial attractiveness on an online dating website	[10]	2017	Attractiveness analysis	Regression, Classification, Convolution Neural Network	The main focus of online dating website is on profile images and to analyze users visual attractiveness.
Sentiment analysis of social network	[11]	2017	Sentiment analysis	Naive bayes, SVM	Measurement of depression can be



sites data using machine learning approaches for measurement of depression					done by extracting features of the depressed people through user profile, user behavior.
An empirical study on detecting fake reviews using machine learning technique	[12]	2017	Fake review detection	Naive bayes, SVM	Fake reviews can be detected using svm and naive bayes, weka tool is used to analyze the data.
Analysis of sentiments and polarity computation of opinions	[13]	2017	Opinion mining	Naive bayes, SVM, Max entropy	To analyze the opinion of the user SVM provides the good result.
A survey on sentiment analysis by using machine learning methods	[14]	2017	Sentiment analysis	Svm ,naive bayes , artificial neural network	Sparse nature of text data is suited for svm classification, supervised technique is used to analyze sentiments.
Efficient and trustworthy review/opinion spam detection	[15]	2017	Text analysis	Naive bayes, logistic regression	Chinese spam filtering approach is described using naive bayes and apriori algorithm
Sentiment analysis of twitter data: case study on digital India	[16]	2016	Sentiment Analysis using opinion finder tool	Opinion Mining	Sentimental analysis uses opinion of the user to analyze data which can be positive, negative and neutral.
Opinion mining of customer reviews based on their score using machine learning techniques.	[17]	2016	Product rating	Opinion mining	Product review contain sentiment expression, and based on the opinions of the reviewers the item can be rated.
A survey on machine learning approaches to social media analytics	[18]	2016	Social Network analysis	Decision Tree	Decision tree learning has been utilized to profile users based on their relationship with other user.
Application of machine learning technique to sentiment analysis	[19]	2016	Sentiment analysis	Natural Language Processing	Lexicon based approach is used based on which polarity of sentiment analysis can be done.
Stock market sentiment analysis based on machine learning	[20]	2016	Opinion mining using Apache spark	SVM, Supervised learning	Based on supervised and unsupervised learning public opinion can be judged about particular product.
A robust method for face recognition and face emotion detection system using support vector machine	[21]	2016	Face recognition	SVM	To find different emotions of face, svm is used to find minimum possible separation between two classes.
Prediction of fake profile on facebook using supervised machine learning technique - a	[22]	2016	Fake profile detection	SVM, decision tree, naive bayes, ANN	Using training data a model is prepared for prediction. Ensemble classifier is used to make result accurate.

theoretical model					
Comparative analysis of twitter data using supervised classifier	[23]	2016	Sentiment analysis	Svm, naive bayes, max entropy	Twitter API is used to collect data from twitter.
Cloud based big data analysis framework for face recognition in social network using machine learning	[24]	2015	Face recognition	Artificial Neural Network	Face recognition make easy for people to select someone when they are sharing photos. It helps in visual impairments.
Stock market prediction: big data approach	[25]	2015	Sentiment analysis	Logistic regression	For stock market prediction technical and fundamental analysis can be done.
Detecting clusters of fake accounts in online social network	[26]	2015	Fake profile	clustering	Describes approach to find group of fake accounts using clustering technique.
Sentiment expression via emoticons on social media	[27]	2015	Sentiment analysis	clustering	Analyze frequency of emoticons to demonstrate prevalence of emoticons on social media.
Classification of sentiment reviews using machine learning technique	[28]	2015	Sentiment analysis	Svm, naive bayes	SVM, Naive bayes is used to analyse the sentiment of the users.
Fraud detection in online reviews using machine learning technique	[29]	2015	Fake reviews	Logistic regression, svm, naive bayes	Unigram, bigram linguistic features are used to analyze online reviews
A review of techniques for sentiment analysis of twitter data	[30]	2014	Sentiment analysis	Supervised learning	For sentiment analysis of text both lexicon and learning based methods can be used to analyze the data.
Sentiment analysis of twitter data using machine learning approaches and sentiment analysis	[31]	2014	Sentiment analysis	Naive bayes, svm max entropy	Preprocessing of data is done with this analysis of data becomes easy if labeled data has positive or negative polarity.
A survey of data mining techniques for social network analysis	[32]	2013	Community detection	Clustering	Nodes and links are used to represent a network. Large network can be compressed to smaller group to form community.
Sentiment analysis in twitter using machine learning techniques	[33]	2013	Sentiment analysis	Naive bayes, SVM	Machine learning techniques are to classify reviews. A sentence level sentiment analysis is done on twitter data.

Analyzing close friend interactions in social media	[34]	2013	Social network analysis Using Weka tool	Bayesian network, Naïve bayes	Bayesian network is considered as the best network for analyzing close friend interaction.
Social network facebook status updates mining for sentiment classification.	[35]	2013	Sentiment analysis	Naïve bayes	Naïve bayes technique is used to extract the useful information during Tunisian revolution associated period.
Opinion mining and sentiment analysis on twitter data stream	[36]	2012	Opinion mining	Random forest, svm, naive bayes	Random forest and svm is used to analyze the sentiments of the users.
Predicting social network measures using ML approach	[37]	2012	Social network analysis	Time series forecasting, classification	Prediction error and time of prediction measures are used.
Sentiment analysis of product using web	[38]	2012	Sentiment analysis	NLP	Unstructured data is taken which can be implicit, explicit, and associative.
Opinion mining in social media : modeling simulation and forecasting political opinion	[39]	2012	Opinion mining	NLP	Three building blocks are used for tracking online opinions of the users.
Using text mining and sentiment analysis for online forums hotspot detection	[40]	2009	Sentiment analysis	Svm	Various steps are performed for hotspot detection. svm is used to realize hotspot forecasting.

#### IV. Conclusion

The use of social media is increases day by day. Different machine learning techniques have been used in different application of social network analysis to analyze the user generated content. Sentimental analyses helps in finding the opinions of the users about particular product or for the measurement of depression. Lexicon based approach is used when number of positive and negative words are counted in the text. Logistic regression technique is used in identification of fake reviews using behavioral features and in analysis of data in social network analysis. Ensemble classifier schema is used in recognizing the emotions from the content generated by users and knowledge base tool is used to store emotional words. Convolution neural network is used to predict users facial attractiveness on an online dating websites. Naive bayes and SVM techniques are used to classify reviews and also used to extract useful information during Tunisian revolution associated period. In this paper Opinion finder tool helps to analyze the polarity of sentiment expressions of the users. For collecting the data used in twitter, the twitter API is used.

Deep neural network is used for the recognizing the behavior of human through face recognition. Deep learning methods give better performance than traditional methods. For the analysis of the fake reviews logistic regression gives up to 97.2% accuracy. Measurement of depression can be done by extracting features of the depressed people through user profile, user behavior. For the measurement of depression support vector machine is the best performing method. To judge the public opinion about particular product supervised and unsupervised learning methods are used. Bayesian network is considered as the best network for analyzing close friend interaction. Logistic regression technique gives 85% accuracy to identify facebook interactions.

#### V. References

- [1] G. kulkarni, L. Abellera "Unsupervised classification of online community input to advance transportation services" 8<sup>th</sup> Annual Computing & Communication Workshop & Conference, IEEE,2018
- [2] X. Wang, X. Zhang "Identification of fake reviews using semantic and behavioral features" 4<sup>th</sup> International Conference on Information Management, IEEE,2018
- [3] I. Perikos, I. Hatzilygeroudis "A framework for analyzing big social data and modeling emotions in social media" 4<sup>th</sup> International Conference on Big data Computing service & Application, IEEE,2018

- [4] S. Paliwal, S. kumar Khatri, M. sharma "Sentiment analysis and prediction using neural network" *International Conference in Inventive Research in Computing Application, IEEE*, 2018
- [5] B. verma, R. S. thakur "Sentimental analysis using lexicon and machine learning based approaches: A survey" *International Journal of Computer Science and Engineering*, 2018
- [6] A. khan, A. khan, M. S husain "Analysis of mental state of users using social media to predict depression! A survey" *International Conference on Advancement in Computer Engineering and Information Technology*, 2018
- [7] P.S. Rao, Dr.J. gyani, Dr.G.Narsimha "Fake profile identification in online social network using machine learning and NLP" *International Journal of Applied Engineering Research*, 2018
- [8] F. Ertam "Analysis of data using machine learning approaches in social network" *International Conference on Computer Science and Engineering*, 2017
- [9] R. Georgina, R. L. Rosa (2017). "Age group classification in social network using deep learning" *IEEE*, 2016
- [10] X. Zang, T. Ymasaki "Prediction of user facial attractiveness on an online dating websites" *International Conference on Multimedia & Expo Workshops, IEEE*, 2017
- [11] A. U. Hassan, J. Hussain "Sentimental analysis of social networking sites using machine learning approach for the measurement of depression" *International Conference on Information and Communication Technology Convergence, IEEE*, 2017
- [12] E. Elmurungi, A. Gherbi " An empirical study on detecting fake reviews using machine learning techniques" *7<sup>th</sup> International Conference on Innovative Computing Technology, IEEE*, 2017
- [13] D. k. tayal, S. k. yadav "Analysis of sentiments and polarity computation of opinions" *2<sup>nd</sup> International Conference on Telecommunications and Networks, IEEE*, 2017
- [14] P. yang, Y. chen "A survey on sentiment analysis by using machine learning methods" *2<sup>nd</sup> Information Technology, Networking, Electronic and Automation control Conference, IEEE*, 2017
- [15] S. P.raut, Prof. C. wasnik "Efficient and trustworthy review/opinion spam detection" *International Journal on Recent and Innovation Trends in Computing and Communication*, 2017
- [16] P. mishra, Dr. R. Rajnish "Sentiment analysis of twitter data: case studt on digital India" *International Conference on Information Technology, IEEE*, 2016
- [17] Miss. L. Kushwaha, Mr. S. D. Rathod "Opinion mining of customer reviews based on their score using machine learning techniques" *International Research Journal of Engineering and Technology*, 2016
- [18] M. S, N. K "A survey on machine learning approaches to social media analytics" *International Journal of Applied Engineering Research*, 2016
- [19] A. P Jain, Asst.Prof P. Danannavar "Application of machine learning technique to sentiment analysis" *2<sup>nd</sup> International Conference on Applied and Theoretical Computing and Communication Technology, IEEE*, 2016
- [20] V. S. rajput, S. M. dubey (2016). "Stock market sentiment analysis based on machine learning" *2<sup>nd</sup> International Conference on Next Generation Computing Technologies, IEEE*, 2016
- [21] R. KM, N. A. K. M "A robust method for face recognition and face emotion detection system using support vector machine" *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, IEEE*, 2016
- [22] S. Y. wani, M. M. Kismani, S. I. Ansarulla "Prediction of fake profile on facebook using supervised machine learning technique- a theoretical model" *International Journal of Computer Science and Information Technologies*, 2016
- [23] R. joshi, R. tekchandani "Comparative analysis of twitter data using supervised classifier" *International Conference on Inventive Computation Technologies, IEEE*, 2016
- [24] V. A, V. S. Shekhar "Cloud based big data analysis framework for face recognition in social network using machine learning" *2<sup>nd</sup> International Symposium on Big Data Cloud Computing, Elsevier*, 2015
- [25] G. v attigeri, M. P. MM, R. M pai, A. Nayak "Stock market prediction: a big data approach", *IEEE*, 2015
- [26] C. xiao, D. M. freeman, T. hwa "Detecting clusters of fake accounts in online social network", *ACM*, 2015
- [27] H. wang, J. A. castanon "Sentiment expression via emoticons on social media" *International Conference on Big Data, IEEE*, 2015
- [28] A. tripathy, A. aggarwal, S. k. rath "Classification of sentiment reviews using machine learning techniques" *3<sup>rd</sup> Conference on Recent Trends in Computing, Elsevier*, 2015
- [29] K. Shivangadhar, S. h, S. sathyan, vanipriya "Fraud detection in online reviews using machine learning techniques" *International Journal of Computational Engineering Research*, 2015
- [30] S. bhutta, A. Doshi "A review of techniques for sentiment analysis of twitter data", *IEEE*, 2014
- [31] G. gautam, D. yadav "Sentiment analysis of twitter data using machine learning approach and sentiment analysis" *7<sup>th</sup> International Conference on Contemporary Computing, IEEE*, 2014
- [32] M. adedoyin-olowe, M. M. gaber "A survey of data mining techniques for social network analysis", 2013
- [33] N. M. S, R. R (2013). "Sentimental analysis in twitter using machine learning techniques", *IEEE*, 2013
- [34] M. Madeira, A. joshi "Analyzing close friend interaction in social media", *IEEE*, 2013
- [35] J. akaichi "Social network facebook status updates mining for sentiment classification", *IEEE*, 2013
- [36] B. gokulakrishnan, P. priyanthan, T. ragavan "Opinion mining and sentimental analysis on twitter data stream" *International Conference on Advances in ICT for Emerging Regions, IEEE*, 2012
- [37] R. michalski, P. K. D. krol "Predicting social network measures using machine learning approach" *International Conference for Advances in Social Network Analysis and Mining, IEEE*, 2012

- [38] U. k "Sentiment analysis of products using web" *International Conference on Modeling Optimization and Computing, Elsevier, 2012*
- [39] P. sobkowiz, M. kaschesky, G. bouchard "opinion mining in social media: modeling, simulation and forecasting political opinions in the web", *Elsevier, 2012*
- [40] N. li, D. d. wu "Using text mining and sentiment analysis for online forums hotspot detection and forecast", *Elsevier, 2009*

