Twitter Sentiment Classification using Machine Learning Techniques

Nirmal Godara[1], Sanjeev Kumar[2]
Department of CSE, Guru Jambheshwar University of Science & Technology, Hisar
nirmalgodara@gmail.com

Abstract-Twitter (www.twitter.com) remains an operational long range unceremonious communication and smaller scale blogging management that vests its clients to send and peruse contented based presents of up on 140 -character text messages, called "tweets". Twitter Sentiment analysis has achievedlot ofattention in recent years. With constantly changing and the development of Internet marketing sites, the reviews and blogs have been obtained among them. These posts give a more extravagant and progressively shifted asset of assessments and conclusions for future. These posts can bepositive or negative. Machine learning (ML)techniques help to categorize these posts. MLtechniques such as Multilayer Perceptron(MLP), Naïve bayes,Fuzzy classifier, Decision Tree and Support Vector Machine(SVM)are used to classify of theseposts. Performance of these techniques is analyzed performance measures such as accuracy, precision, recall and F measure on Twitter dataset. Results showed that SVM outperformed all others in case of Twitter Sentiment analysis.

Keywords-Twitter Sentiment analysis,ANN, NBand SVM

I. INTRODUCTION

Informal communication is online administrations or destinations which attempt to imitate social connections among individuals who know one another or share a typical intrigue. Person to person communiqué locales enable consumers to share thoughts, exercises, events, and comforts classified their separate systems.Along these lines posts give us a more extravagant and progressively shifted asset of assessments and conclusions.Twitter (www.twitter.com), ranked among the top 10 most visited website from last two years(ref. Alexa's). It has ~250+ million monthly lively users who conveyregarding ~500+ million tweets per day (ref. Wikipedia). Twitter is accessed on a mobile device by 78% users and 77% of its users are living exterior US. 35+ languages are supported by Twitter. Its enormousworldwide usage has made it aextraordinaryproposal for analysing and accepting trends on a worldwide or a confinedlevel. The volume, velocity and variety of data that is produced on twitter hysterics the depiction of big data. Here we willconsider some approaches on analysing Role of Big Data in Twitter:

The ban on election campaign (now called as Online Election campaign) using Internet services such as blogs, social networking services (SNS), among others was lifted from the 2013 Upper House elections. For quick response to the online election campaigns, NHK came to a decision to initiate big data investigation in the news based on Twitter information. The estimationof Internet users was envisioned in well-organized and originalway and then telecasted. This progressive initiative has attracted industry-wide attention as a new attempt for news telecast.Attempt the analysis with a new viewpoint that is different from the earlier exit polls and opinion polls in the news report after 2013 Upper House election wherein the ban on online election campaign was lifted for the first time.Implement a new concept in news telecasting and coverage using SNS that people commonly use, since Internet usage is widespread.

Analysing whole volume of Twitter data facilitatedby different opinions of Internet users as well as exposure the Upper House elections from a perspective of existing surveys. Viewing the suggestion of community transformation through Twitter data facilitated an successfulexposure that captures preliminary responses.Classification is used to classify data into different classes based on some common properties that are set in one class and slightly different from the properties found in the erstwhile classes.Sentiment classifications may be of two types: contextual Twitter Sentiment analysiswhich deals with specific parts of tweets and general Twitter Sentiment analysisworks with the universal sentiment of the complete text. For example tweet is: “4 more years of being in shithole Australia then I move to the USA :D”. In contextual sentiment technique Australia will be recognized as negative sentiment and USA will be recognized as positive sentiment. But general Twitter Sentiment analysis deals with whole text/tweet and will recognize it as positive.

The Tasks for Sentiment Analysis as shown in Fig1, is a perplexing assignment and includes a few separate errands, viz:

- Subjectivity Classifications
- Sentiment Classifications
- Complimentary Tasks(Object Holder Extraction, Object/ Features Extraction)

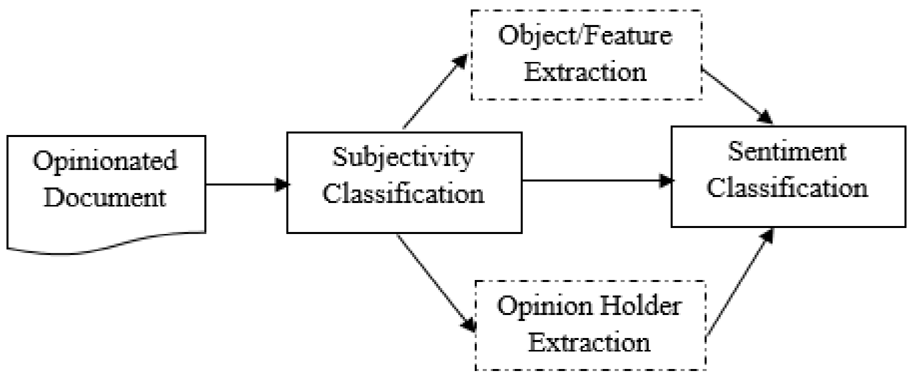

Figure 1: Tasks of Sentiment Analysis

The subsequent subsets explain the minutiae of the major tasks in Sentimentality Analysis:

A. Subjectivity classifications.

Ordinarily, some random archive will contain verdicts that prompt feeling and some that don't. That is, an archive is a gathering of target sentences, verdicts that state true, and abstract judgments, verdicts that speaks to the creator's sentiment, perspective or feeling. Subjectivity grouping is the assignment of arranging condemnations as obstinate or not stubborn. Tang et al. [19], expressed prejudice arrangement as pursues: Let S = {s1, . . ., sn} be a lot of verdicts in report D. The issue of partiality characterization is to recognize verdicts recycled to display assessments and different types of subjectivity (abstract sentences set Ss) from s used to unbiased exhibit real data (target sentences set So), where SsU So= S.

B. Sentiment Classification

When the errand of discovering whether a bit of content is stubborn is over we need to discover the extremity of the content i.e., regardless of whether it communicates an optimistic or destructive sentiment. Conclusion order can be a twofold characterization (optimistic or destructive), multi-class grouping (very destructive, destructive, impartial, optimistic or amazingly optimistic) relapse or else positioning.

C. Opinion Holder Abstraction

Notion Analysis likewise includes optional assignments like conclusion holder extraction, for example the revelation of conclusion frames or fonts. Identification of conclusion holder is to perceive immediate or roundabout wellsprings of sentiment. They are indispensable in news stories and other formal records in light of the fact that numerous suppositions can be communicated in a similar article comparing to various conclusion holders. In archives like these, the numerous feeling holders may unequivocally be referenced by name. In interpersonal organizations audit locales and web journals the feeling holder is generally the creator who might be recognized by the login accreditations.

D. Objects /Feature Extractions

An extra undertaking is the disclosure of the objective substance. Interestingly with survey destinations, sites and online life locales tend not have an established goal or predefined theme and are in this way, slanted to talk about arranged points [20]. In such stages it winds up important to know the objective element. Additionally, as referenced before target elements can have highlights or segments that are being assessed. A commentator can have varying suppositions about the various highlights or segments of the objective substance. Thus, highlight based conclusion investigation, for example, extraction of item include and the related assessment, is a discretionary errand of opinion examination.

Twitter Sentiment analysis can be performed at Feature, Document, and Sentence or at Word level.The various issues in Sentiment Analysisare:

A. Requirement of Keyword Collection:

In feeling investigation, there is a need to differentiate the content in to two classes (optimistic and destructive) which are so not quite the same as one another. This is on the basis that slants can frequently be communicated in a easily broken way making it uncertain to be distinguished.

B. Sentimentality is Domain Specific:

The expression go recite the book would be careful positively in a book audit, however whenever communicated in a motion picture survey, it proposes that the book is favoured over the film, and in this manner have a contrary outcome.

C. Manifold Opinions in a Verdict:

Single decision can contain various opinions close by conceptual and undeniable parts. It is helpful to withdraw such stipulations.

D. Negation Control andSarcasm:

mockery and disjointedness are outstandingly quiet difficult to recognize.

F. Comparative Sentences:

A close to sentence imparts an association reliant on similarities or differences of more than one . For instance, the sentence, ―Car X is better than Car Y passes on an immediate opposite supposition from ―Car Y is better than Car X.

G. Implicit Opinion:

Feeling that shows up in content can be described as: express where the abstract sentence legitimately passes on an assessment ―We had a magnificent time, and certain where the sentence infers a supposition "The battery went on for 3 hours". Present feeling investigation models won't most likely identify this understood sentiment as a negative assessment.

H. Opinion Spam:

Close spam alludes to phony or false sentiments that attempt to purposely deceive peruses or mechanized frameworks by giving undeserving positive feelings to some objective items so as to advance the articles.

Sentiment Analysis is an expression that surfaces when you need to legitimize your social spend or when you are attempting to make sense of Internet based life. Investigation is basic for each brand that keeps up a web-based social networking record and plans to profit by it. Any computerized advertising aide will weight on how focal testing out web based life technique is. Online networking investigation is the initial move towards distinguishing the best seminar via web-based networking media for your image.Electronic life examination (SMA) implies the strategy of social event data from online life districts and comprises and estimating that data to settle on commercial decisions [1] [2]. This technique goes past the standard checking or a fundamental examination of retweets or "inclinations" to develop a through and through idea of the social client. This is seen as the major foundation for enabling an endeavour to:

- Execute jogged duty like adjusted and one-to-many.
- Enhance societal joint exertion over an arrangement of business limits, for instance, customer organization, advancing, support, etc.
- Maximize the customer knowledge

Web based life is a nice standard to see consistent customer adoptions, objectives and estimations. The maximum prevalent utilization of web based life inspection is to end up aware with the customer base on an inexorably excited dimension to help better target client association and publicizing. The primary stage in an electronic life learning undertaking is to make logic of which commercial targets the data that is aggregated and analyzed with benefit. Business estimations got from online life examination may fuse customer responsibility, which could be assessed by the amount of followers for a Twitter record and number of retweets and notification of an association's name. With electronic long range informal communication checking, associations can in like manner look at what number of people seeks after their essence on Twitter.

II. RELATED WORK

G. Shobana, et.al [1] dissected the celebrated individual's id's (@realdonaldtrump) or hash labels (#IPL2018) for understanding the mentality of individuals in every circumstance when the individual has tweeted or has followed up on certain occurrences. The proposed framework is to break down the supposition of the general population utilizing python, twitter API, Text Blob (Library for preparing content). As the outcomes it serves to investigation the post with a superior precision.

Nick Jennings, et.al [3] investigated how people and AI frameworks can cooperate. In such associations, the people and the AI frameworks supplement each other's qualities and shortcomings, prompting an ascent in the people, just as in the machines. Drawing on multi-disciplinary work in the regions of AI, self-ruling frameworks, AI, publicly supporting and universal figuring, this discussion investigates the logical supporting of such frameworks, the applications they have been connected to, and the societal ramifications of their far reaching reception.

[4]To discover which approach is best for which dataset which will help to specialists to choose approach and dataset. Plotted Word haze of specific occasion which feature the regular term from tweets and furthermore determined quantities of positive, negative and impartial tweets from every occasion.

Anchal Kathuria et.al. (2014) attempts to provide comparative study of existing techniques for opinion mining as well as an elaborate view of machine learning techniques. The performance of machine learning strategies, like SVM and MLP with their advantages and disadvantages are also discussed vividly. According to the study, machine learning approaches are efficient and provide better results. Extensive research work is also carried out in the field of lexicon-based techniques to prove the credibility of their results and also efforts are being made to improve the machine learning techniques in terms of accuracy up to bigram and trigram [11].

[6] Clarified that Twitter speaks to a microblogging webpage where individuals post and read sees about different themes. The size of the information acquired from the twitter is humungous. To deal with such information, the Hadoop system is utilized to store, process and oversee it so it very well may be time proficient.

Junseok Song et al. (2018) proposed two methods to resolve the problems that showed positive and negative words in the estimation of the weight. Work eliminated the unimportant words in the characteristic selection phase using the Multinomial Naïve Bayes (MNB) technique. The assessment of the results showed that the proposed systemconsiderably increases the precision with respect to the existing BNB algorithm and the MNB scheme [7].

Brian Heredia, et.al [8] directed an experimental investigation utilizing opinion information from two sources, online surveys and tweets. We first test the presentation of notion examination models manufactured utilizing a solitary information hotspot for both in-space and cross-area order. At that point, we assess classifiers prepared utilizing examples arbitrarily examined from the two sources. Furthermore, the specialists assessed inspecting various amounts of occasions from the two information sources to decide what number of occurrences ought to be incorporated into a preparation informational collection. We apply factual tests to confirm the noteworthiness of our outcomes and find that utilizing a blend of cases from audits and tweets is like, or superior to anything any model prepared from a solitary space

Bong-Hyun Back et al. (2019) proposed a system using Naive Bayes algorithm and natural language processing (NLP) to gather information regarding sentiment from large unstructured social media big data. Additionally, work analyzed the efficiency of the proposed method through various experiments [9]. Bayes' algorithm provided an accuracy of 63.5%, which was lesser than the NLP method used.

Abinash Tripathy et al. (2015) presented a comparison of the results obtained with the Naive Bayes (NB) and Support Vector Machine (SVM) classification algorithms. These algorithms are used to classify a sentimental check that can be positive or negative. The polarity film dataset was used for training and testing, and a comparison was made with the results available for a critical examination [10].

Ali Hasan et al. (2015) aims to examine the views or texts on various social media platforms through machine learning with sentiment, analysis of subjectivity or polarity calculations. Regardless of the use of diverse machine learning techniques for mood analysis through the elections, there is a immensecall for an inventive approach. To address these challenges, the contribution of this document includes adopting a hybrid approach with a feeling analyzer that includes machine learning [12]. Results also provided a comparison of sentiment analysis techniques when analyzing political views using Naive Bayes and SVM.

MaiteTaboada, et.al [13] given an exploration goal to extricate data on the notoriety of various writers, in view of works concerning the writers. The venture means to make a database of writings, and computational instruments to concentrate content naturally. This paper depicts the underlying phases of an undertaking following the artistic notorieties of six creators somewhere in the range of 1900 and 1950, and the appropriateness of existing procedures for separating assessment from writings that examine and scrutinize these writers. Bing Liu, et.al [14] concentrated on online client audits of items. It makes two commitments. In the first place, it proposes a novel structure for breaking down and looking at buyer feelings of contending items. A model framework called Opinion Observer was additionally actualized. The framework is to such an extent that with a solitary look of its perception, the client had the option to obviously observe the qualities and shortcomings of every item in the brains of buyers regarding different item includes. This correlation is helpful to both potential clients and item makers.

III. MACHINE LEARNING TECHNIQUES FORTWITTER SENTIMENT ANALYSIS

In this section various ML Techniques used for classification of tweets are discussed.

A. Decision Tree

The decision tree technique recursively divides the training set into smaller subsets. The decision tree techniquecalculates information gain by taking feature used at the partitionusing Claude Shannon's information theory. This attribute provides minimal arbitrariness in the partition. This practice of selecting attributes and minimizing arbitrariness is repeated recursively. The expected information necessary to classify a data set D is given by:

$$\text{Info (D)} = -\sum p_i \log2 (p_i) \quad \text{Where } i = 1, 2 \ldots\ldots\ldots\ldots n$$

Where pi is the probability of the instance ithatbelong to D with class Ci. Information (D) is required to recognize the class name of a record in D[15].

B. Artificial Neural Network

The ANN as shown in Fig 2, is also known as Multilayer perceptron. In this architecture the input layer is layer which is used to provide input. The input layer data fed to the next hidden layer and hidden layer processes this data. This architecture can have many hidden layers and each hidden layer have many neurons. After processing of data at the hidden layer,the data sent to the to the output layer. Then the data from the output layer and the target values are compared. If difference between these two values are found then the weightadjustment is performed and repeated until the error is reduced to the given limits [2][15].

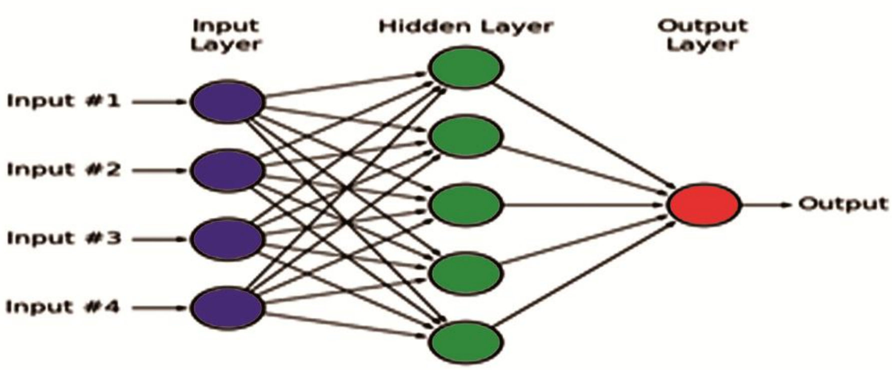

Fig. 2: Multi-Layer Perceptron

C. Support vector machine(RBF)

The Support Vector Machine as shown in Fig 3 examines the data, defines the assessmentlimits. Kernels are used for the calculation that is carried out in the input space [11][15]. Therefore, all input data is divided into a class. Therefore, distance between the two classes was found which is used to define classifier scope. By maximizing leeway, non-critical decisions are reduced. SVM supports classification and regression, used in statistical learning theory.

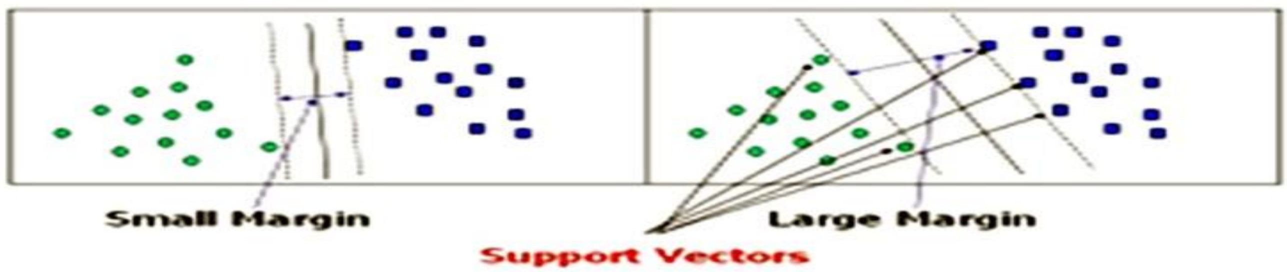

Fig. 3: SVM Topology

D. Naïve Bayes method

Bayes' theory is the base of the classifier on which unforeseen probability that an event y has a place with class n can be resolved from the prohibitive probabilities of finding explicit event in each class and the boundless probability of the event in each class. It learns the pattern of investigative set of items and compares the contents with the listing of words which is used to classify the objects to their accurate class. Consider d as given tweet and c* is an assigned class of given tweet d, where:

$$C^{*} = \arg mac_{c} P_{NB}(c|d) \ldots \ldots \ldots \ldots$$

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

Here f is a object, number of object is (fi) and is denoted with ni(d) , m shows number of objects. Parameters P(c) and P $(f|c)$ are calculated using maximum likelihood estimation[7][9].

E. Fuzzy classifier

A Fuzzy classifier assigns a class value to an item, based on the item explanation. The item explanation contains vector containing values of the features (attributes) related to the classification task. The fuzzy rule-based classifier is a set of fuzzy if-then else rules used for specification of classification rules .

IV. METHODOLOGY USED FOR TWEETSCLASSIFICATION

Methodology consists of two Modules, namely Pre-processing Module and Classification Module.

A. Pre-processing Module

First, an application account in Twitter created streaming API whichis allowed to retrieve real-time tweets.After collecting the data, an inspection of the tweets is performed. 50 tweets of each review is taken to examine.Hash tags, usernames, Punctuation characters, English stop words and Emojisarereplaced with a space.Pre-processed cleansed data set is obtained in output to provide input to the classifier.

B. Classification Module

This module consists of the following steps:

Tweets are converted into a sequence file containing a pair of <key,value> using subdirectory utility.Then <Text, Vector>pairs, containing term frequencies for every tweets document are obtained.Tweets dataset is divided into two parts, 70% for testing and 30% for training.Train the classifier with the training dataset using ML Techniques.

V. RESULT AND ANALYSIS

A. Twitter Datasets

A Twitter dataset which is huge and complex in naturecan have 140 characters in a tweet. Tweet is a combination of text data and metadata. Tweets text dataset are collected by REST API and these are used for classification. Metadatacontains users states, hashtags, URLs and user ID which are used to discover the domain of tweet.

B. Performance Measures and Results

Confusion matrix is used forexplanation of classification results as shown in Table 1. The higher left cornertells the number of items classified as sensitiveas they were in factsensitive, and the inferior right cell tells the number of items classified as non sensitiveas they were in factnon sensitive. The remaining two columns show the number of misclassified items[15].

Table 1: Confusion Matrix

	Classified as sensitive	Classified as not sensitive
Actuallysensitive	A	B
Actual not sensitive	C	D

Below formulae are used to calculate comparison parameters:

$$ACCURACY = \frac{\mathrm{A + D}}{\mathrm{A + B + C + D}}$$

$$PRECISION = \frac{D}{B + D}$$

$$Recall = \frac{D}{C + D}$$

$$FSCORE = 2.\frac{Precision\,.Recall}{Precision + recall}$$

Classification Method	Accuracy	Precision	Recall	F-score
Naïve Bayes	72.66	72.16	72.81	72.4
Decision Tree	80.45	79.45	80.14	79.6
MLP(ANN)	89.44	88.45	87.95	88.1
SVM-RBF(SR)	92.34	90.12	94.34	92.3
FUZZY(FU)	89.23	90.12	91.23	90.3

Table.2 Analysis of Classification methods based on Performance measures

In Table 2, Naïve Bayes, Decision Tree, Artificial Neural Networks (ANNs), FUZZY(FU) and Support Vector Machine (SVM) are analyzed on basis of various performance measures.

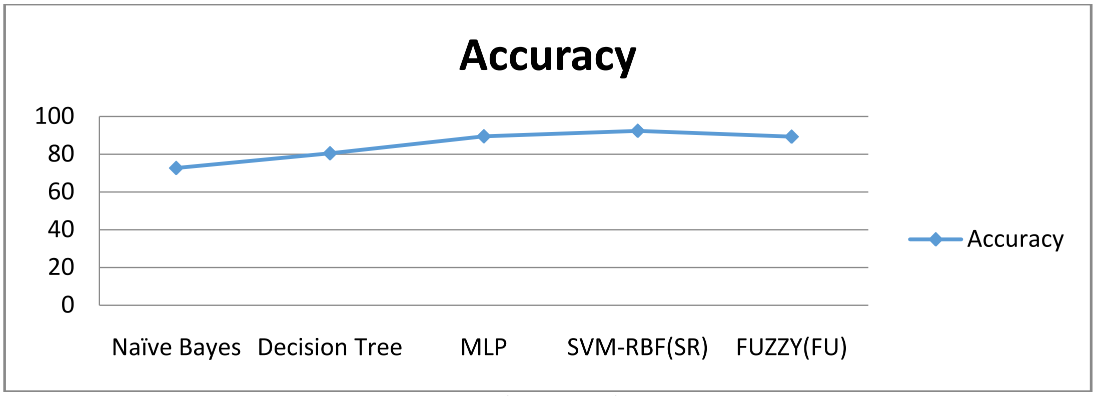

Fig. 4: Comparison of various performance measures

As shown in Fig4Accuracy ranges from 72.66-92.34% for given classification methods. MLP and Fuzzy have attained accuracy of 89.44, 89.23 respectively. SVM has attained 92.34% accuracy which is 3% higher as compared to Fuzzy and MLP.Naïve Bays has attained the accuracy 72.66 which performs worst as compared to other. Support Vector Machine (SVM) performs well as compared to others because SVM kernel maps the features on a high dimension. The basis behind SVM's better performance is that it maximizes the margin by which lesschances of misclassification of unknown objects. In case of fuzzy rules, accuracy is not improved as compared to SVM because of its statistical rules which are not able to handle the overlapping of features.

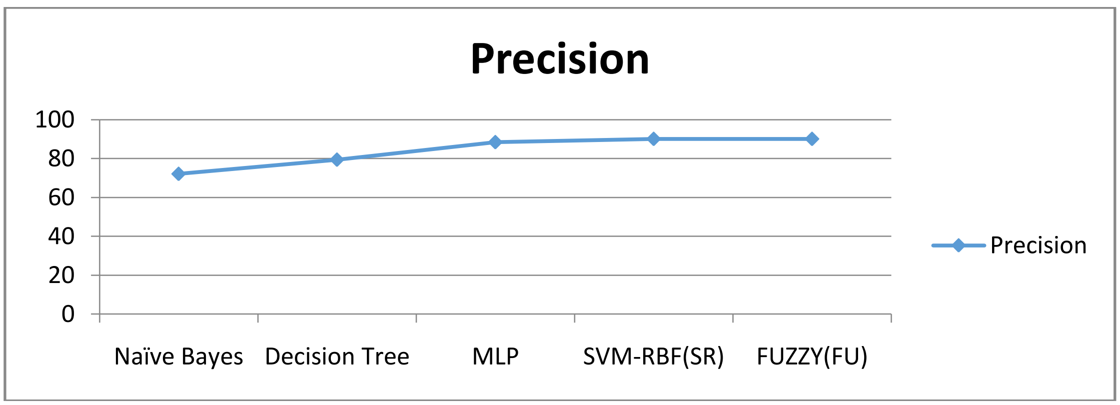

Fig. 5: Comparison of various performance measures

In Fig5,Naïve Bayes, Decision Tree, Artificial Neural Networks (ANNs), FUZZY(FU) and Support Vector Machine (SVM) are analyzed on basis of precision.Precision ranges from 72.16-90.12 for given classification methods. MLP and Fuzzy have attained precision of 88.45, 90.12 respectively. SVM has attained 90.12% precision which is higher as compared to others. Support Vector Machine (SVM) performs well as compared to others because SVM kernel maps the features on a high dimension. The basis behind SVM's better performance is that it maximizes the margin by which less chances of misclassification of unknown objects. In case of Naïve Bayes, Decision Tree, Artificial Neural Networks (ANNs) and fuzzy rules, precision is low as compared to SVM because they are not able to handle the overlapping of features.

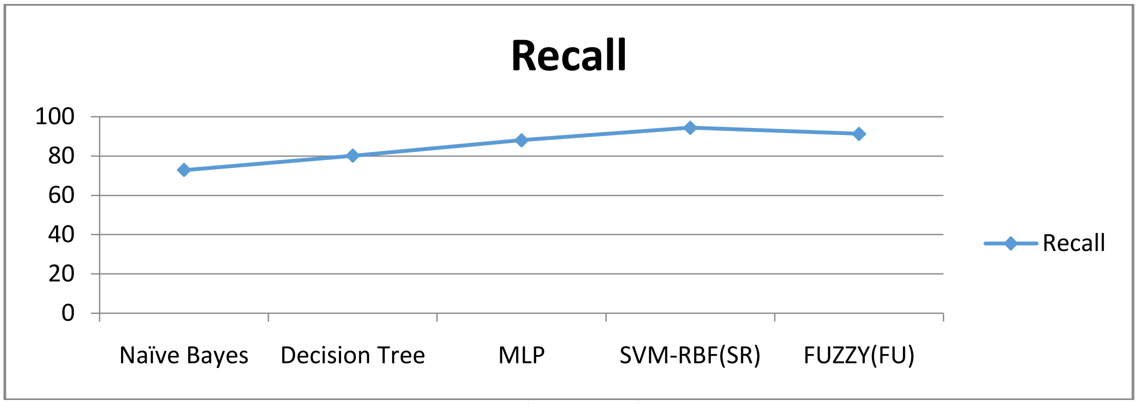

Fig. 6: Comparison of various performance measures

In Fig6 Naïve Bayes, Decision Tree, Artificial Neural Networks (ANNs), FUZZY(FU) and Support Vector Machine (SVM) are analyzed on basis of recall. Recall ranges from 72.81%-94.34%. Recall for SVM is 94.34% which is3% higher from fuzzy and 7% higher from MLP.

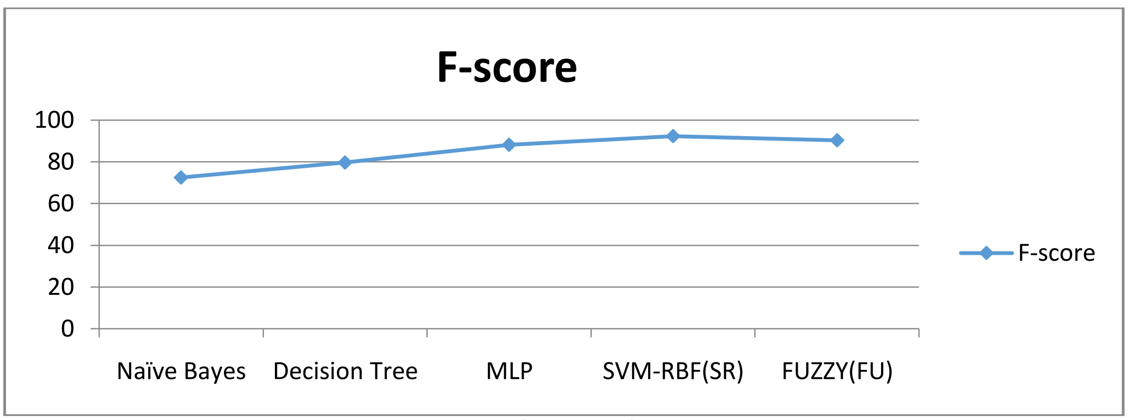

Fig. 7: Comparison of various performance measures

In Fig7 Naïve Bayes, Decision Tree, Artificial Neural Networks (ANNs), FUZZY(FU) and Support Vector Machine (SVM) are analyzed on basis of f-measure. F-measure ranges from 72.4-92.3%.F-measure for SVM is 92.3% which is 2% higher from fuzzy and 5% higher from ANN/MLP. High value of f-measure shows that there is a good agreement between actual and predicted class.

VI. CONCLUSION

In this examination, the idea of Decision Tree, Artificial neural networks (ANNs), Naïve bayes, Fuzzy(FU) and Support Vector Machine (SVM) are utilized for grouping of calculation with paired order process. Such sort of techniques helps in investigating diverse component vectors with a doled-out class so as to distinguish the connection reliance between an assessment and every one of the elements. Here, every one of the vectors is considered as a point of information in vector dimensional space that equivalents to the size of list of capabilities. Performances of these techniques are compared taking Twitter dataset through various performance measures such as accuracy, precision, recall and F measure. All the parameters Accuracy, Precision, Recall and F-score are high in case of SVM. Our study illustrated that SVM technique comes out to be most excellent classifier for Twitter Sentiment analysis.

REFERENCES

[1] Shobana G, Vigneshwara B, Maniraj Sai A. (2018). Twitter Sentimental Analysis. International Journal of Recent Technology and Engineering (IJRTE). 7(4s) (pp: 2277-3878)

[2] G. Shidaganti, R. G. Hulkund, and S. Prakash, "Analysis and Exploitation of Twitter Data Using Machine Learning Techniques," in *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications*, vol. 628, pp. 135–146, Springer Singapore, 2018.

[3] Jennings, N. (2018, December). Human-Artificial Intelligence Partnerships. In Proceedings of the 6th International Conference on Human-Agent Interaction (pp. 2-2). ACM.

[4]Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., ... & Socher, R. (2016, June). Ask me anything: Dynamic memory networks for natural language processing. In International conference on machine learning (pp. 1378-1387).

[5] P. W., & Dai, B. R. (2013, June). Opinion mining on social media data. In 2013 IEEE 14th International Conference on Mobile Data Management (Vol. 2, pp. 91-96). IEEE.

[6] Kotwal, Aishwarya, Jadhav, Dipali & Fulari, Priyanka. (2016). Improvement in Sentiment Analysis of Twitter Data using Hadoop. International Conference on "Computing for Sustainable Global Development. pp: 0973-7529.

[7] kuna;, J., Kim, K. T., Lee, B. J., Kim, S. Y., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis.TIIS, 11(6), 2996-3011

[8] Heredia, B., Khoshgoftaar, T. M., Prusa, J., & Crawford, M. (2016, November). Integrating multiple data sources to enhance sentiment prediction. In 2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC) (pp. 285-291). IEEE.

[9] Back, B. H., & Ha, I. K. (2019). Comparison of Sentiment Analysis from Large Twitter Datasets by Naïve Bayes and Natural Language Processing Methods. Journal of information and communication convergence engineering, 17(4), 239-245.

[10] [65] Tikar, A. P., Jaybhaye, S. M., & Pathak, G. R. (2015, October). A systematic review on scheduling types, methods and simulators in cloud computing system. In 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) (pp. 382-388). IEEE.

[11] Kathuria, A., & Sharma, A. (2018). A Comparative Study of Various Approaches for Sentiment Analysis of Twitter Data.

[12] AAli, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information sciences, 305, 357-383.

[13]Taboada, M., Gillies, M. A., & McFetridge, P. (2006, May). Sentiment classification techniques for tracking literary reputation. In LREC workshop: towards computational models of literary analysis (pp. 36-43).

[14]Liu, B., Hu, M., & Cheng, J. (2005, May). Opinion observer: analyzing and comparing opinions on the web. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.

[15]Sunila Godara and Rishipal Singh(2016).Evaluation of Predictive Machine Learning Techniques as Expert Systems in Medical Diagnosis.Indian Journal of Science and Technology.Volume: 9, Issue: 10, Pages: 1-14

AUTHORS PROFILE

Author-1

Dr. Sanjeev Kumar did his M.Tech and Ph.D in Comp. Sc. & Engg from Guru Jambheshwar University of Science and Technology Hisar(Haryana).He is working as Associate Professor in CSE deptt. GJUS&T,Hisar.He had more than 50 papers in his account.His area of interest are machine learning, computer networks and cloud computing.

A

Nirmal Godarais doing her Ph.D D in Comp. Sc. & Engg from Guru Jambheshwar University of Science and Technology Hisar(Haryana). Her area of interest are machine learning and cloud computing.