iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI: http://doi.org/10.22214/ijraset.2017.11011

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI, November 2017- Available at www.ijraset.com

Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner

Priyavrat[1], A. J Singh[2]
[1,2] *Department of Computer Science, Himachal Pradesh University, Shimla, India.*

Abstract: ***Sentiment Analysis is an important and a very active area of research. It is being used by various public and non-public organizations to find out sentiments of the web users about their product and services, which results in making some important and effective decisions. The sudden growth of social media applications has resulted in the generation of a huge amount of opinionated data which is mostly being used in research work. Sentiment Analysis is a sub-discipline of natural language processing, where the main idea is to understand polarity of a sentence, paragraph or whole document by analysis of textual data gathered from various sources. This paper is giving a comparative analysis of four supervised machine learning techniques (Support Vector Machine, Naive Bayes, Decision Tree and Neural Network)used for sentiment analysis on the basis of different performance parameters. In this comparative study,it is analyzed that SVM (Support Vector Machine) has greater performance than other three supervised machine learning techniques.***

Keywords : Sentiment, Sentiment Analysis, Opinion Mining, Machine Learning, Support Vector Machine, Naive Bayes, Decision Tree, Neural Network, Spyder, Rapid Miner.

I. INTRODUCTION

Sentiment Analysis is the analysis of public thoughts and their opinions. In SA common public opinions are used to define polarity of the textile. Whether the given text has a positive, negative or neutral attitude, therefore it is also called as Opinion Mining. Sentiments can also be categorized into n-point scale like very good, good, bad, very bad, and satisfactory. The polarity of the text reflects sentiment or attitude of the public or an individual. These public opinions are gathered from various web 2.0 andfeatures like micro blogs (tweets), blogs, online data sets, movie reviews and product review sites[1].Sentiment Analysis is a stepwise step process and these steps are data extraction, pre-processing, sentiment identification, feature selection, sentiment classification and polarity report[2]as shown in Fig.1.This paper emphasis on Sentiment classification process.

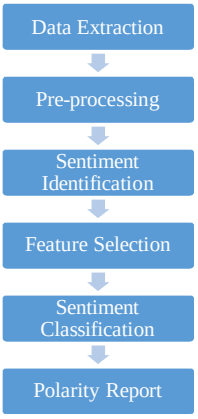


Fig. 1: Sentiment Analysis Process

Sentiment Analysis is different from text mining and it concentrates on attitude/opinion whereas text mining focuses on the analysis of facts. Section II is giving a review of related work done on a comparative study of SA approaches. Section III describes various SA approaches. These approaches are Machine Learning, Lexicon Based and the combination of both i.e. Hybrid approach. This paper only focuses on supervised machine learning approaches used for Sentiment Analysis and also giving a short introduction to other approaches. Section IV explains methodology adopted. Section V describes results and analysis of the experimental work. Section VI is concluding the present research work and Section VII is going to describe the future scope of the study.

II. RELATED WORK

Sentiment Analysis started its roadmap around the year 2000. It has a rapid growth in research area from the last ten to fifteen years. The Google trend chart shown in Fig. 2 is showing results of its popularity. SA is sub area of NLP but it has been seen that SA have a moregrowing popularity than NLP in the previous years.

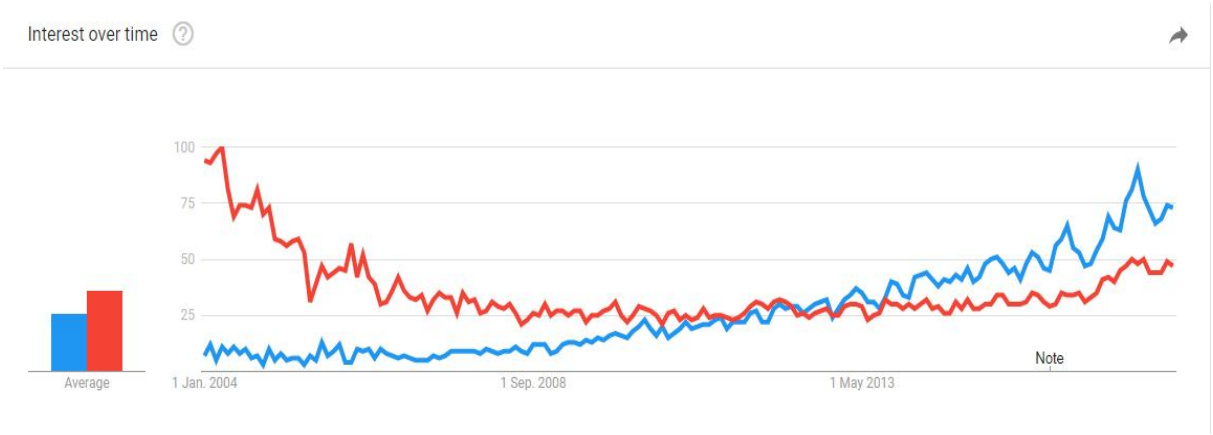


Fig. 2: Google trends result for Sentiment Analysis (Blue Line) and Natural Language Processing (Red Line) [25]

Many research and survey papers have been published related to SA research work and some of the important papers from which the present study has been motivated are discussed below:

Rehab M. Duwairi and Islam Qarqaz [3] gave a comparative analysis of three classifiers: Naive Bayes, SVM, and K-NN. They generated their own training dataset by collecting tweets and facebook comments from the internet. They make use of Rapidminer tool for this comparative study and found that SVM gives the highest precision while K-NN gives the highest recall.

G. Vinodhini and RM Chandra sekaran[4] have given a detailed survey on SA and OM in their paper. This survey paper discusses various sentiment classification techniques and its challenges. This studycompared various other papers written on SA and OM techniques on the basis of four different factors like accuracy, precision, recall, and F1.

Chetashri Bhadane et al[5] used a hybrid approach for the classification ofa product review. The hybrid approach combined SVM with domain specific lexicons and their experimental result indicates 78 % accuracy.

Hemlata et al[6] have discussed various machine learning algorithms used for Sentiment analysis. They proposed new system which uses a hybrid approach, combining NB and Maximum Entropy technique for the classification of text.

Doaa Mohey El-Din Mohamed Hussein [7] evaluated forty-seven research papers on sentiment analysis in their survey paper and gave various challenges facing in the field of Sentiment Analysis. This survey paper discussed these challenges in detail and introduces another important factor called domain dependence to recognize the sentiment challenges.

Ashutosh Bhatt et al. [8] proposed a novelsystem that performs classification of customer reviews on iPhone5 from Amazon. The proposed system follow an algorithm as discussed in their paper. The proposed system work on various rules, designed for feature selection of the user reviews.

Haseena Rahmath and Tanveer Ahmed [9] studied a theoretical comparative analysis of various techniques used in sentiment analysis. In this paper, various feature selection methods like opinion words, POS, Terms and frequency and negations has been discussed. After studying various research papers they concluded that SVM technique from supervised machine learning approach has the highest accuracy among all the other techniques.

Mudinas et al. [10] developed a new sentiment classification system called pSenti. pSenti uses both machine learning and lexicon approaches i.e. a hybrid approach. The system has taken advantage of stability from lexicon based and advantage of high accuracy from machine learning approach. They extracted sentiment words and considered them as features in machine learning algorithm. This hybrid approach achieved an accuracy of 82.30%.

Prabowo and Thelwall [11] considered Part of Speech (POS)as a feature in their studies and constructed a feature set with the help of adjectives and adverbs to classify sentiments of the text.

Pang et al. [12] used three supervised machine learning techniques to classify the text named as SVM, Naive Bayes, and Maximum Entropy. In this paper, different feature selection method was used such as unigram, ngram and POS.Results show that performance of Naive Bayes classifier is good on small feature set while the performance of SVM is good for the large feature set. They also found that Maximum Entropy hasa better result than Naive Bayes forthe large feature set.

Ashmeet Singh and R Sathyaraj [13] compared three supervised machine learning algorithms using Rapid Miner named Naive Bayes, Random Forest and Decision Tree on the basis of six parameters such as accuracy, precision, recall, true positive, false positive and f-measure. Two datasets havinga small and large number of instances was used for each classifier. In this paper, it has been found that Naive Bayes resultsin better performance than other two in smaller dataset whereas Decision Tree is best suited for the large dataset and Random forest acts as an average in both the cases.

Ashok Badresiya et. al. [14] compared five review spam detection supervised learning techniques for their performance. The five techniques were Naive Bayes, SVM, K-NN, Logistic Regression and Decision Tree.In this comparative study, they found that SVM has greater accuracy (83.19%) than other techniques and Decision tree have very low accuracy only 51.00 %.

Gurneet Kaur and Abhinash Singla [15] presented an empirical study of the efficiency of classifying product review by semantic meaning in their paper. They propose some different approaches including spelling correction in review text and then classifying comments by implementinga hybrid algorithm combining decision tree and Naive Bayes algorithm.

Jeevanandam Jotheeswaran [16] used a hybrid technique for classification, including decision tree to select features from a movie review training dataset from IMDb and a multilayer perceptron to classify the feature extracted from the movie reviews. After selecting 70 to 90 features they achieved 81.25 % classification accuracy.

III. SENTIMENT ANALYSIS APPROACHES

There are mainly three approaches which are used in the classification of text [17]. These three approaches are Machine learning, Lexicon based and Hybrid approaches. Classification of text is done at various levels such as sentence based, document-based and aspect or opinion based [18]. These three approaches exhibit different techniques as shown in Fig. 3. Among all these approaches and techniques, this paper shows a comparison of supervised machine learning techniques, which are described in this section below.

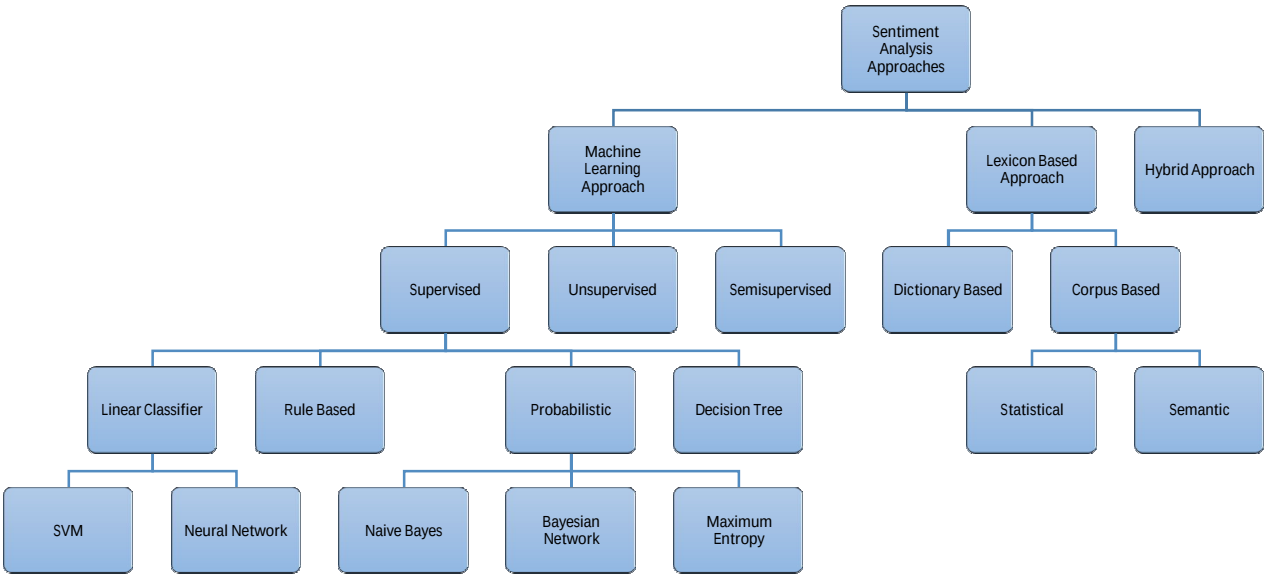


Fig. 3: Sentiment Analysis approaches

A. Machine Learning

Machine learning enables computers to grow, modify and learn by them self when they are exposed to a new data[19]. Machine learning algorithms use many computational methods to learn information directly from data without relying on predetermined equations as a model. The algorithms always try to improve their performance as the number of samples available for learning increases. There are different types of machine learning methods categorized as supervised, unsupervised and semi-supervised learning described below:

In supervised learning, the training set(Xi, Yi) and an algorithm train a model which is capable to predict the output for every new input as shown in Fig. 4. Supervised learning uses classification and regression techniques to develop a predictive model.Supervised learning techniques use labelled data. Labelled data is data that is augmented with sort of some meaningful tag, label or class that is somehow informative.

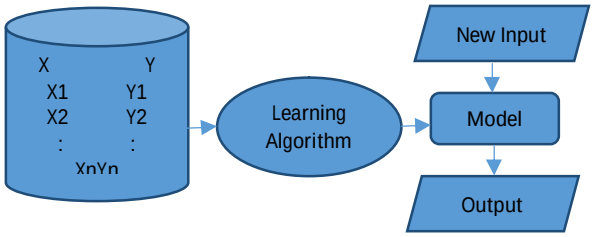


Fig. 4: Supervised Learning

Various supervised machine learning techniques which have compared in this paper are described below:
A linear classifier is one which classifies the objects into different classes to which they belong according to their features or characteristics. A linear classifier uses alinear combination of X and Y as (aX + bY) (where as and b are constant) to make the classification decision. Two linear classifiers used in this study are:

1.) *SVM:* Support Vector Machine classifier is considered as the most accurate text classifier. SVM tries to find a hyperplane which separates data intotwo categories, i.e. positive or negative with the maximized margin. Support Vector Machine technique is called so because it makes use of support vector. Support vector is an array of the data points, which is used to find out the boundary of each plane. When a new unknown input is given to SVM classifier, it predicts that which side of the margin it falls on.

2.) *Neural Network:* In a neural network, there is a network of neurons or nodes where each neuron is a basic processing unit of the network. Neural Network follows a layered structure. The network is arranged such that one layer is the input layer which receives inputs that are to be classified. These inputs cause some of the neurons in the input layer to fire, and these neurons, in turn, pass signals to the neurons which they are connected, some of which also fire, and so on. In this way, a complex pattern of firings is arranged throughout the network, with the final result being that some neurons in the final output layer fire. The connections between neurons are weighted and the weight is associated with each neuron which is used to compute a function for the given inputs. Implementation of Neural networks is given in [20,21]Probabilistic techniques make use of conditional probabilities. Conditional probability is given as P (H|E) and it is read as, the probability of hypothesis H oversupportive evidence E. To compute this we need to take into account the prior probability of H and the extent to which E provide evidence of H. The probabilistic technique from supervised learning used in this paper is discussed as:

3.) *Naive Bayes:* Naive Bayes sentiment classification technique makes use of Naive Bayes theorem, which was given by Thomas Bayes: P (C|X) = P(X|C).P(C)/P(X),where X= document and C = Class (positive or negative). In this classifier, it is assumed that the probability of one word in the document being in a specific category is unrelated to the probability of the other words being in that category. This technique uses some steps to classify a document or text. In the first step, the dataset is converted into a frequency table. In second step PRIOR is calculated using formula $P(C) = N_c/N$ where P(C) is the possibility of the class, N_c is the total count of a particular class in the training dataset, and N is the total count of class in the training dataset. In the thirdstep, conditional probability/likelihood of each word attribute is computed. And in last step posterior probability is computed using formula C_{map} = argmax P(X1,X2,X3.....Xn) P(c).

4.) *Decision Tree:* A decision tree is a nonlinear function. DT is in the form of a tree having two types of nodes: Decision nodes and Leaf nodes. Decision node specifies a choice or a direction based on features while leaf nodes specify classification or value of the example. If in a decision node all the examples are positive or all the examples are negative, then all the examples belong to the same class otherwise they belong to different classes. DT categorize a document in a top-down fashion i.e. it starts from the root of the tree and moves downwards via the branches until a leaf node is reached. After that, the document is classified in the category that is labelled with the eaf node. In unsupervised learning as shown in Fig. 5, there are no pre-classified training samples or training datasets. This type of learnig tries to find out hidden patterns in data. Unsupervised learning method uses unlabelled data. Unlabelled data is the data that can easily get from natural sources e.g. photos, tweets, audio recordings, videos etc. There is no explanation for each piece of unlabelled data. It just contains data and nothing else.

Clustering is the most commonly used unsupervised learning technique used in the classification of text. There are several variants of clustering and one of the popular variants among those is K-means clustering.

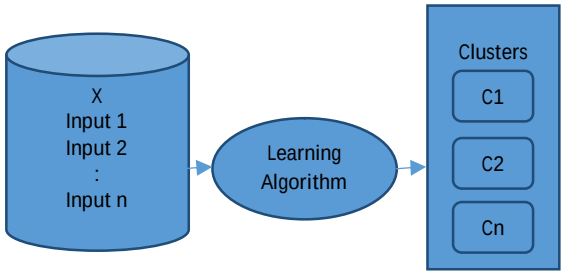


Fig. 5: Unsupervised learning

Semi-supervised learning technique is a combination of supervised and unsupervised learning where the classification of data usesa large amount of labelledas well asunlabelled data.

B. Lexicon Based Approach

Lexicon Based approach makes use of opinion bearing words. These opinion bearing words are used to classify sentiments of the text.In this approach, the input text is converted to tokens by the tokenization process. After that, each newly generated token is matched in the dictionary. When there is a positive match, the score is incremented. Otherwise, the score is decremented and the word is tagged as negative. Two methods are used under Lexicon based approaches first one is a dictionary-basedmethod and the second one is the corpus-basedmethod. In the dictionary-based method, a dictionary is built manually in a text editor with some basic words. These basic words, collected inside a dictionary are also called as seed words and are assigned with positive or negative values. After that, the anonymous and synonymous for these words are find out online using WordNet or any other online dictionary. In this way, newly founded words are iteratively added to the seed list. This iterative process ends when no new words could be found. On the other side,the corpus-based method makes use of given seed list plus a general purpose sentiment lexicon used for a specific domain. The corpus-based approach is performed using statistical or semantic approach.

C. Hybrid Approach

The hybrid approach uses both machine learning and lexicon based techniques. Some researchers [22,23,24] have developed sentiment classification systems which are more accurate than machine learning or lexicon based approaches.

IV. METHODOLOGY

This paper compares four machine learning approaches namely SVM (Support Vector Machines), NB (Naive Bayes), DT (Decision Tree) and NN (Neural Networks) on the basis of five parameters (accuracy, precision, recall, classification error, and F- measure). In this study, Rapid Miner tool is used to compare all these techniques. Since the supervised learning uses training dataset and test dataset, therefore another tool called Spyder is used in this work which collects tweets from Twitter with keyword “Donald Trump” to prepare training and test dataset. The work has considered only two classesi.e. positive or negative to classify the sentiments of the tweets.

Stepwise step process that is used for comparison of these techniques is given in the Figure below:

A. Crawling of data (tweets) from twitter. (Anaconda (spyder) is used for this) Data is gathered regarding the keyword “Donald Trump” in a CSV (Comma Separated Values) file format.
B. Convert CSV data file into Excel data file format and write sentiment values i.e. positive or negative, manually for eachtweet and save this example file. This excel file is used as a training dataset to train the model.
C. Prepare another excel file called test dataset using Step 1 which contain tweets whose sentiments have to predict.
D. Design a process in Rapid Miner with the help of various operators as shown in Fig. 6.
E. Add training datasets and test dataset into the local repository of RapidMiner.
F. Run the designed process for different Supervised Machine Learning Techniques that we have chosen and
G. Find results for each technique and compare those results.

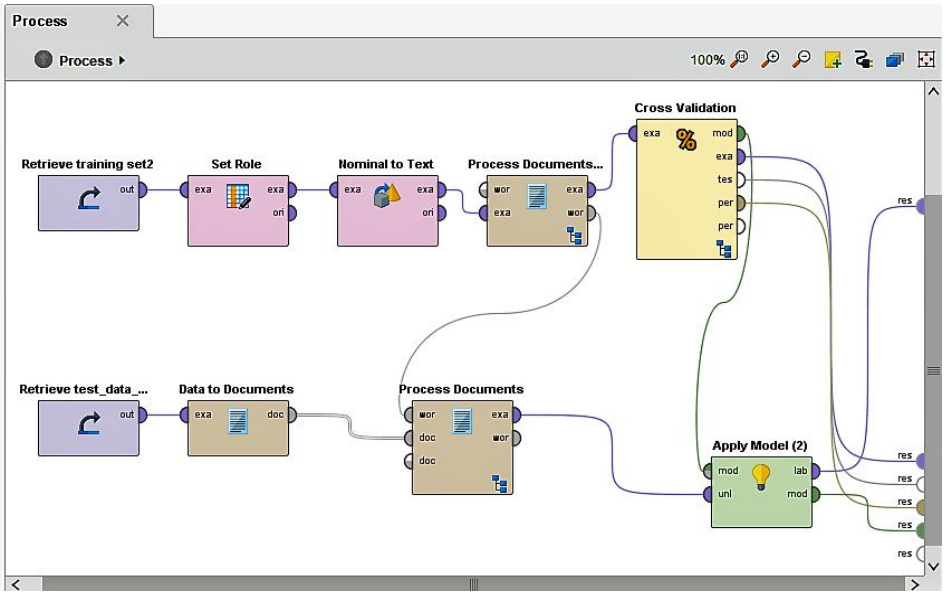


Fig. 6: Sentiment Analysis process in RapidMiner

To accomplish this we have created five different training datasets (contain 50, 100, 200, 300 and 500 tweets) which are used in the training of each supervised learning technique and a test dataset (contain 40 tweets) whose sentiments has to predict.
Each of SVM, NB, NN, and DT supervised learning technique is trained for every training dataset containing 50,100,200,300 and 500 tweets and tested using K-fold cross-validation process. -fold cross-validation allows testingthe accuracy of the techniques by splitting labelled data into k classes and then performing k number of iterative running rounds as shown in Fig. 7.The present work has used 5-fold cross-validation method, i.e. data is divided into 5 divisions such that one is used for testing and four others are used for training in the first run. Accuracy for each round is fond out and the final accuracy is calculated as the average of the accuracy of all five rounds.

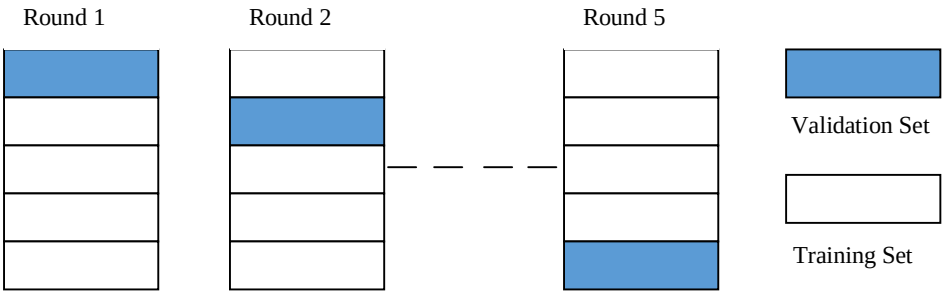


Fig. 7: K - fold validation process.

i.e. Final Accuracy = Average Accuracy (Round 1, Round 2 .... Round 5)
Results for all the parameters are noted and evaluated to conclude that which technique is best suited for theclassification of training datasets. These parameters are explained as:

H. Accuracy

Accuracy is the percentage of correct predictions made by a classifier. It is calculated as Accuracy = Total correct classification / Total number of examples.

I. Precision

High precision value states accurate results and it takes all relevant data but returns only topmost results. Precision is calculated as theratio of correctly positive classified examples among all examples that are predicted as positive.

Precision = Positives Correctly Classified / Total Predicted Positives.
(Total Predicted Positives = True Positives + False Positives)

J. Recall

Recall givessensitivity of the problem and it processes completeness. Recall specifies a relative number of correctly classified positive examples among all positive examples.
i.e. Recall = Correctly Classified positive examples / Total Positives.

K. Classification Error

Classification error is the ratio of a total number of incorrect classification to the total number of training examples and it is calculatedasClassification error = Total incorrect classification / Total number of examples.

L. F-Measure

F-measure depends upon precision and recall values and it is calculated as:
f = 2pr / (p + r), where p=precision and r = recall.

V. RESULTS

This section discusses result for comparative analysis of all the four supervised learning techniques which are compared in this paper for different training datasets.
All the parameters are based on confusion matrix resulted in the execution of each classifier. The confusion matrix is given in Fig. 8.

accuracy: 62.00% +/- 15.03% (mikro: 62.00%)

	true positive	true negative	class precision
pred. positive	90	62	59.21%
pred. negative	14	34	70.83%
class recall	86.54%	35.42%	

Fig. 8: Confusion Matrix

The confusion matrix is used to define all the parameters. The row 'predicted positive' tells about the examples that were classified as positive and 'predicted negative' tells about the examples that were classified as negative. True positive and true negative tells about the examples that were actually labelled positive and actually labelled negative respectively.
For example, in the above confusion matrix, 90 tweets that are actually positive are predicted as positive. 62 tweets which are actually negative are predicted as positive. Similarly, 14 tweets which are actually positive are predicted as negative and 34 tweets which are actually negative are predicted as negative.
The true negative is equal to the examples that were actually labelled negative and were classified as negative.
The true positive is equal to the examples that were actually positive and were classified as positive.

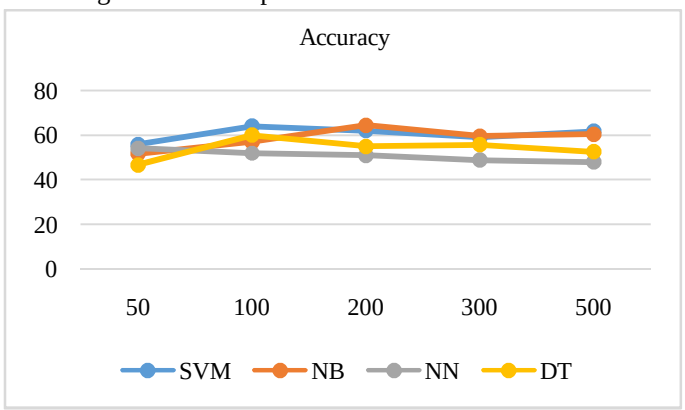
Table 1 is showing a comparison of all the four techniques for different parameters:

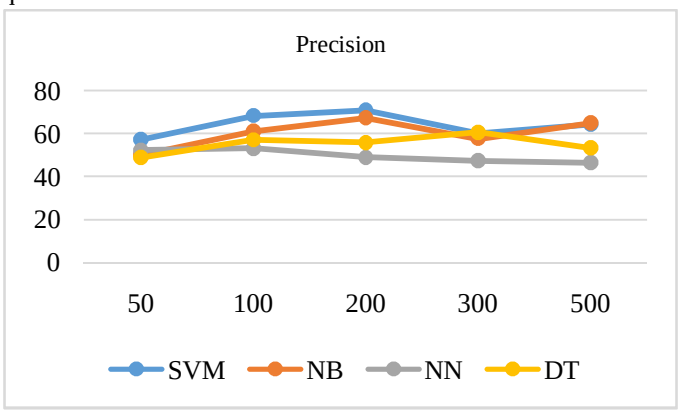
	SVM	Naive Bayes	Neural Nets	Decision Tree
Accuracy (%)				
For 50 Tweets	55.83	51.67	54.17	46.67
For 100 Tweets	64.00	57.00	52.00	60.00
For 200 Tweets	62.00	64.50	51.00	55.00
sFor 300 Tweets	59.00	59.67	49.00	55.67
For 500 Tweets	61.80	60.40	48.03	52.60
Precision (%)				
For 50 Tweets	57.14	50.00	52.17	48.84
For 100 Tweets	68.29	61.11	53.06	57.08

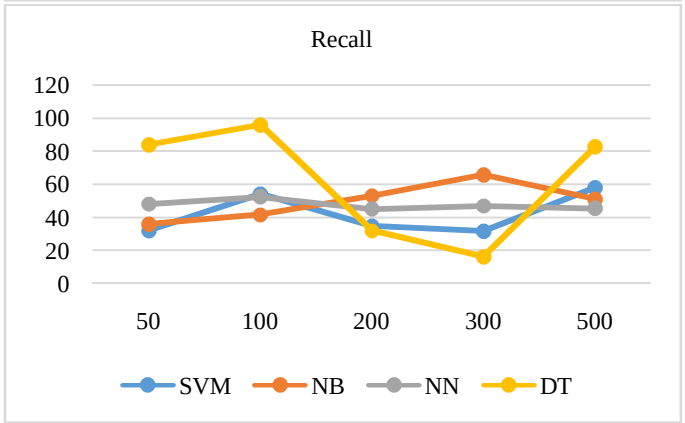
For 200 Tweets	70.83	67.25	48.86	55.77
For 300 Tweets	60.15	57.58	47.23	60.53
For 500 Tweets	64.34	64.76	46.31	53.28
Recall (%)				
For 50 Tweets	32.00	36.00	48.00	84.00
For 100 Tweets	54.17	41.67	52.50	95.83
For 200 Tweets	35.00	53.25	45.00	32.00
For 300 Tweets	31.79	65.89	46.80	16.34
For 500 Tweets	58.08	51.03	45.57	82.60
Classification error (%)				
For 50 Tweets	44.17	48.33	45.83	53.33
For 100 Tweets	36.00	43.00	48.00	40.00
For 200 Tweets	38.00	35.50	49.00	45.00
For 300 Tweets	41.00	40.33	51.00	44.33
For 500 Tweets	38.20	39.60	51.97	47.40
F measure (%)				
For 50 Tweets	41.03	41.86	50.00	61.76
For 100 Tweets	60.87	50.57	52.00	70.48
For 200 Tweets	47.22	57.84	46.74	39.19
For 300 Tweets	42.25	60.07	47.01	25.70
For 500 Tweets	60.08	56.09	45.93	65.52

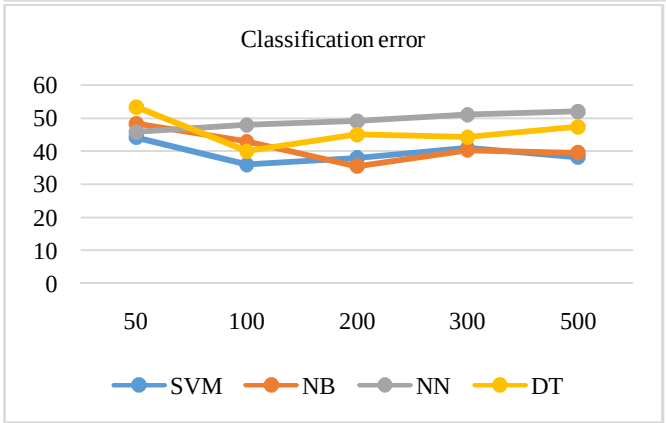
Table 1: Comparison of SVM, NB, NN, and DT on the basis of five parameters

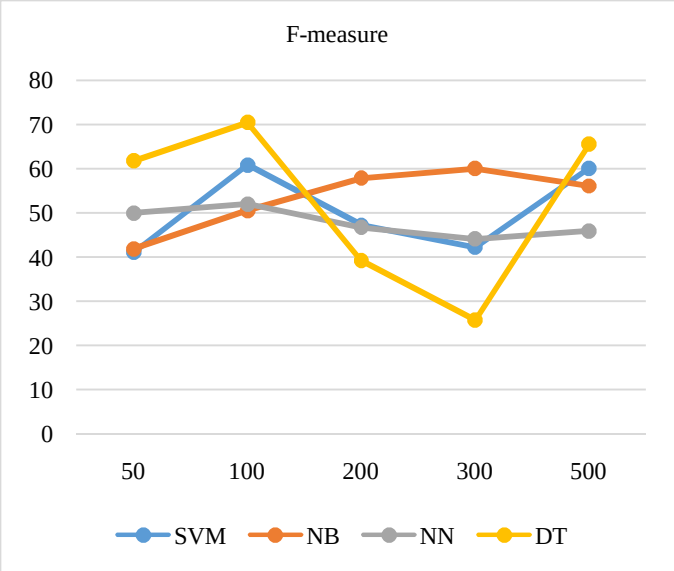
Following charts shows parameters behavior for different techniques and are based on the above table:











A. All these above charts show following results

1) SVM has more accuracy than other techniques for small as well as large training/example datasets. Naive Bayes also has good accuracy for large datasets and DT has an average accuracy result for classification of sentiments.

2) SVM has the largest Precision while NN has the least.

3) NB increases its recall value as numbers of tweets in example dataset are increased.

4) SVM has lowest classification error whereas NN has maximum classification error.

5) F measure for NB is good among other techniques.

VI. CONCLUSION

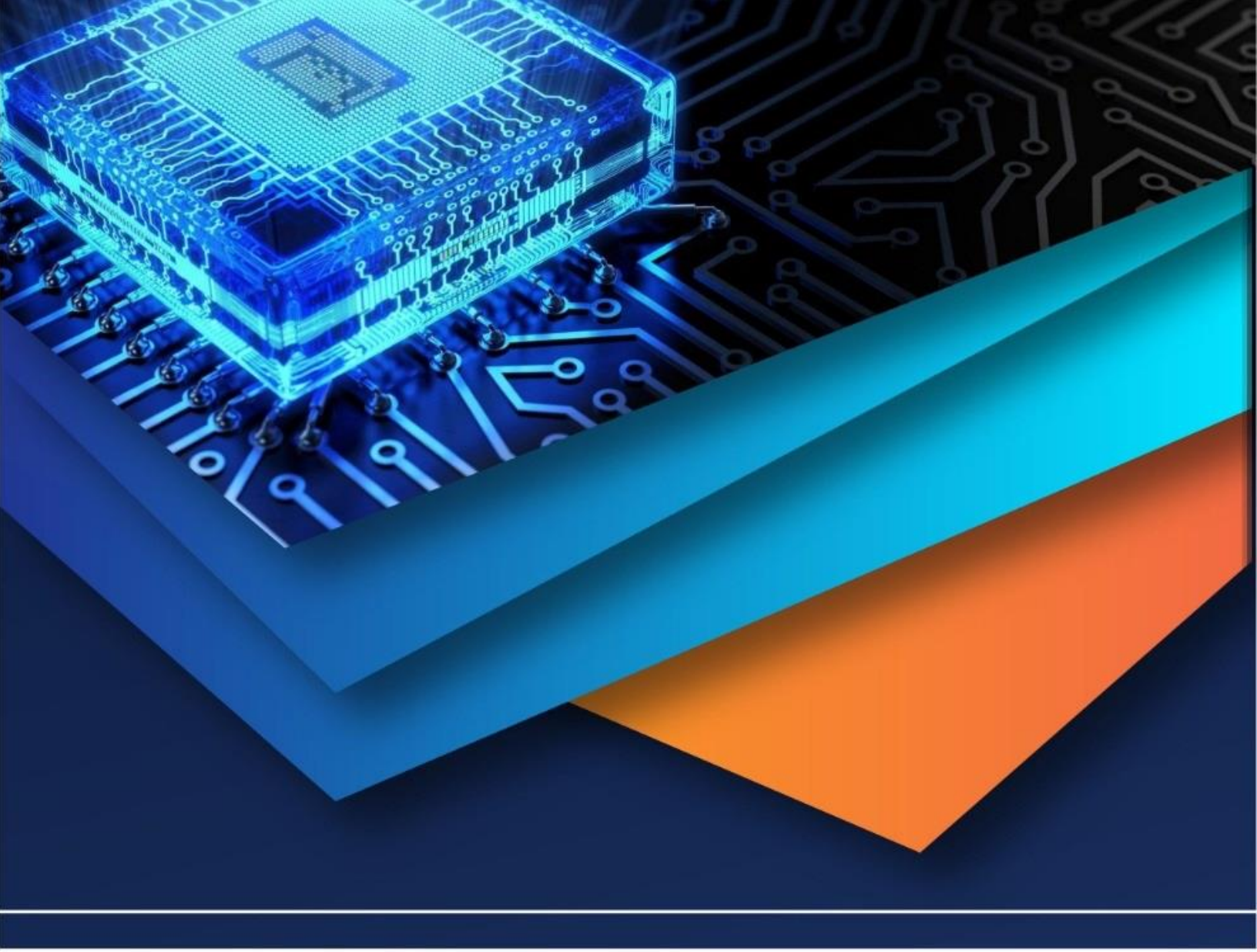
Sentiment Analysis is a process during which polarity of unstructured textual data is determined. Sentiment analysis has several areas of applications including classifying reviews, classifying tweets, summarizing reviews etc. This paper has described a detailed work on Sentiment Analysis and various supervised machine learning algorithms used in Sentiment Analysis. This work has considered four sentiment analysis techniques: Support Vector Machine, Naïve Bayes, Neural Network and Decision tree and has compared them on a GUI based tool called RapidMiner. All these techniques are compared on five parameters named as accuracy, precision, recall, classification error, and F-measure. This work has considered tweets from Twitter, which are used to train and test the various classification models that we have opted to compare. In this comparative study, it is tried to find out that what wouldbethe impact of example dataset on the performance of various Sentiment classification algorithms? We have analyzed various results for all the sentiment classification techniques and concluded that SVM has high performance than other three techniques and if there would be a large number of example dataset then greater accuracy could be achieved by the machine learning techniques.

VII. FUTURE SCOPE

This paper shows a comparative study of different supervised learning techniques which are used in Sentiment Analysis. All these techniques are compared for their performance but more future work is needed on further improving the performance of such techniques. There is a huge need in the industry for applications of Sentiment Analysis because every company wants to know how consumers feel about their services and product. In future work different types of approaches such as machine learning and lexicon based should be combined in order to overcome their drawbacks and enhance their performance by utilizing their merits. Also in future, Sentiment Analysis needs broader and deeper commonsense knowledge bases. A complete knowledge must be combined with reasoning methods that are more deeply inspired by human thought and psychology and this will lead to better understanding of natural language opinions. The concept of Sentiment Analysis can also be added to the systems that people are commonly using nowadays like PDA (Personal Digital Assistance), computer systems, and automatic cars etc. which could detect the negative mood of an individual and respond in a positive way to please him/her.

REFERENCES

[1] A. Kumar and M.T Sebastian,Sentiment Analysis: A Perspective on its Past, Present, and Future, International Journal of Intelligent Systems and Applications 4.10, 2012.
[2] R. Rajput and A.Kumar Solanki,Review Of Sentiment Analysis Methods using Lexicon-Based Approach, International Journal of Computer Science and Mobile Computing 5.2, 2016
[3] R. M. Duwairi and I. Qarqaz,Arabic Sentiment Analysis using Supervised Classification, Future Internet of Things and Cloud (FiCloud), 2014 International Conference on. IEEE, 2014.
[4] G. Vinodhini and R. M. Chandrasekaran,Sentiment Analysis and Opinion Mining: A Survey,International Journal 2.6 pp. 282-292, June 2012.
[5] C. Bhadane, H. Dalal, and H. Doshi, Sentiment Analysis: Measuring Opinions, Procedia Computer Science 45 pp. 808-814, Jan 2015
[6] I. Hemlatha,G.P Saradhi Verma andA. Goverdhan, Sentiment Analysis Tools using Machine Learning Algorithms, International Journal of Emerging Trends of Technology in Computer Science(IJETICS), Volume 2, April 2013
[7] D.M. Hussein, A Survey on Sentiment Analysis Challenges, Journal of King Saud University-Engineering Science, April 2016
[8] A. Bhatt, A. Patel, H. Chedda and K. Gawande, Amazon Review Classification and Sentiment Analysis, International Journal of Computer Science and Information Technologies 6.6,pp.-5107-5110, 2015.
[9] H. Rahmath and T. Ahmed, Sentiment Analysis Techniques-A Comparative Study, in International Journal of Computational Engineering and Mathematics (IJCEM), Vol. 17, July 2014.
[10] A. Mudinas, D. Zhang and M. Levene, Combining Lexicon and Learning Based Approaches for Concept-Level Sentiment Analysis, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, pp. 1-8, 2012.
[11] R. Prabowo and M. Thelwall. SentimentAnalysis: A Combined Approach, Journal of Informatics 3.2,pp. 143-157, 2009.
[12] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs Up?: Sentiment Classification using Machine Learning Techniques, Proceedings of the ACL-02 conference on Empirical methods in natural language Processing-Volume 10. Association for Computational Linguistics, 2002.
[13] A. Singh and R. Sathyaraj, A Comparison between Classification Algorithms on Different Datasets Methodologies Using Rapidminer, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, May 2016
[14] Badresiya, Ashok, and J. Teraiya,Performance Analysis of Supervised Techniques for Review Spam Detection.
[15] G. Kaur, and A. Singla,Sentimental Analysis of Flipkart Reviews using Naïve Bayes and Decision Tree Algorithm, International Journal of Advanced Research in ComputerEngineering& Technology 5.1,pp. 148-153, 2016.
[16] J. Jotheeswaran and S. Koteeswaran, Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis, ARPN Journal of Engineering and Applied Sciences10.14,pp. 5883-94, 2015.
[17] U. Kumari, D. Soni and A.K. Sharma, A Cognitive Study of Sentiment Analysis Techniques and Tools : A Survey, IJCST vol.8, Issue 1, Jan-march 2017
[18] A. D'Andrea, F. Ferri, P.Griffoni, and T. Guzzo,Approaches, Tools and Applications for Sentiment Analysis Implementation, International Journal of Computer Applications 125.3 (2015).
[19] https://www.simplilearn.com/what-is-machine-learning-and-why-it-matters-article
[20] M.E. Ruiz andP. Srinivasan,Hierarchical Neural Networks for Text Categorization, Presented at the ACM SIGIR conference; 1999
[21] H.T. Ng, W.B. Goh, andK.L. Low, Feature Selection, Perceptron Learning, and A Usability Case Study for Text Categorization,Presented at the ACM SIGIR conference; 1997
[22] J. Fang and B. Chen, Incorporating Lexicon Knowledge into Svm Learning to Improve Sentiment Classification, In Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), pages 94–100, 2011
[23] A. Mudinas, D. Zhang, M. Levene, Combining Lexicon and Learning Based Approaches for Concept-Level Sentiment Analysis, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, pp. 1-8, 2012.
[24] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B.Liu, Combined Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis, Technical report, HP Laboratories, 2011.
[25] ImageSource: https://trends.google.com/trends/explore?date=all&q=sentiment%20analysis,natural%20language%20processing.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY