iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42009

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: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Empirical Analysis of Support Vector Machine and Multinomial Naive Bayes

Aastha Sachdeva[1], Dr. Indu Kashyap[2]
[1]*Student,* [2]*Professor, Department of Computer Science and Engineering, Manav Rachna International Institute of Research and Studies, Faridabad, India*

Abstract: ***As everyone is free to give their opinions about people or products they buy on social media and E-Commerce platforms respectively So, Opinion Mining is extensively used for classifying the opinions into different polarities- positive and negative. Its use has significant overlap with the domain of Machine Learning. The objective of this paper is to compare two Supervised Learning Algorithms- Support Vector Machine (SVM) and Multinomial Naïve Bayes (MNB) and Opinion Mining on twitter dataset. Opinion Mining is a Natural Language Processing (NLP) task which aims to determine the views of people by identifying and extracting the data. The performances of both the models are evaluated and compared their accuracy, precisions, recall values and f scores. The measurement accuracy is measured by Confusion Matrix and ROC curve. In results, it is observed that SVM and MNB both show almost same performance when compared.***

Keywords: Confusion Matrix, Multinomial Naïve Bayes, Natural Language Processing, Opinion Mining ROC curve, Support Vector Machine.

I. INTRODUCTION

Social Media has emanated as an efficient communication tool and is progressively used by more than 4.5 billion people among which 84% are active users. Social media is nowadays used by people to share their thoughts, reviews and discussing over any issue that impacts their personal lives. Social Networking Sites like Twitter allows the end users to easily collaborate and share information which makes social media an interesting platform for data mining. Twitter spectators indeed varies from company representatives, luminaries, politicians, celebrities, athletes etc. which makes twitter a source of breaking news and other high-value information. The focus of the paper is on the performance of two classification techniques namely, Support Vector Machine (SVM) and Multinomial Naïve Bayes (MNB) classifier by calculating the accuracies, efficiency, precision and recall values using Opinion Mining

A. Opinion Mining

Literary data from the twitter can be enormously arranged into two fundamental cases: opinions and facts. Opinion mining which is also known as Sentiment Analysis is Natural Language Processing (NLP) strategy, which is ordinarily used to characterize the polarity of the text content into Positive and Negative. The process of tracing the moods of society against certain topics, products and people is referred to as Sentimental Analysis which can be defined in two categories positive (value=1) and negative (value=0) depending upon the comments of people. [3] Opinions can be comparative or regular. Standard conclusions of regular opinions are regularly alluded to just as "suppositions". In a near conclusion, at least two substances are looked at as far as their likenesses or contrasts; often utilizing similar or superlative descriptors or verb modifiers gave one of the first investigations of sentiment analysis for Twitter information, utilizing AI calculations to classify message assumption and utilizing far off supervision and preparing information comprising of Twitter messages with emojis, which are utilized as loud marks. Individual's opinions are the most important in decision making and arriving at the results because they are based on their past experiences.

Sentiment Mining is an important aspect of Data Mining, a process of mining the valuable data from the corpus data. Data mining provides various tools for analysis like regression, classification, clustering etc.

B. Natural Language Processing

Natural Language Processing (NLP) is a pioneer technology that assist the computers to understand human's natural language. NLP is contemplated as sub-field of Artificial Intelligence concerned with the interactions between humans & machines and programs the computer to process and analyze the huge amount of natural language data. NLP is driving force of many popular applications like: Language Translation Application (Google translator), Personal Assistant Application (Siri, Alexa, OK Google), Word Processors such as Grammarly and Microsoft Word to check the grammatical accuracy of the texts, chatbots and many more. Also, developers can perform various tasks like automatic summarization, sentimental analysis, entity recognition and much more.

©IJRASET: All Rights are Reserved

[13] The NLP has two components: Natural Language Understanding (NLU) whose main task is to convert the natural language of humans to representations that are easily manipulated by the machines. The other one is Natural Language Generation (NLG) which is opposite of NLU and translates the information of Computer Database into readable human language.

C. Support Vector Machine

Support Vector Machine (SVM) is a non-probabilistic binary linear classification algorithm which is one of the most robust prediction methods based on statistical learning. It can also be used for regression problems but mainly classification. SVM algorithm necessitates plotting of each data items as a point. It is basically used to find the N-dimensional hyperplanes, where N is number of features that differentiates the two classes. The decision boundaries that help to classify the data points are known as Hyperplanes. SVM classification is the perimeter which best segregates the two classes. The data points falling on either side of the plane are assigned to be of different classes as shown below. Data point(s) passing through the marginal hyperplanes are the Support Vectors.

After creating hyperplane two parallel lines are created known as Margin Lines (Positive hyperplane & Negative Hyperplane) from the most positive and most negative data points. The data points falling above the Hyperplane are positive and below are negative. Marginal Distance is denoted by '$d^{'}$. d^{+} is the shortest distance between Maximum Margin Hyperplane and closest Positive point, d^{-} is the shortest distance between Maximum Margin Hyperplane and closest Negative point. Positive and negative hyperplanes are also known as Marginal Hyperplanes. The significance of this is to create generalized model to get better accuracy. Best hyperplane is formed by maximizing the marginal distance as small value of d can give lot of errors.

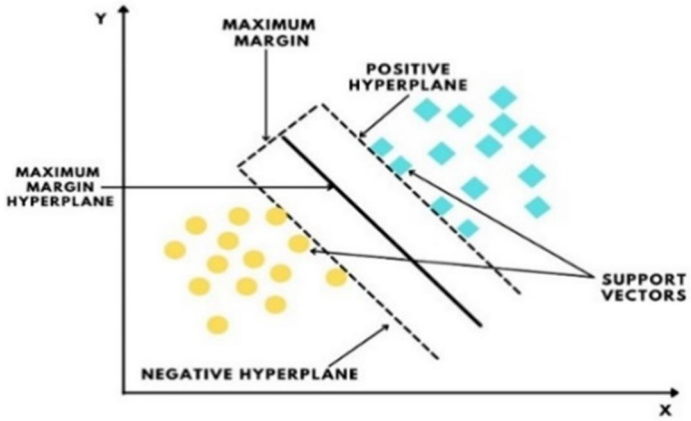


Figure 1. Linearly Separable Hyperplane

Maximization of marginal distance is done by Loss Function which helps in maximizing the margin in Hinge Loss. The use of loss function is to measure the loss or cost of a particular model by telling the error rate so that we could find how well our model is performing. Hingle loss is a type of cost function that is specifically used for Support Vector Machines. But this is only possible with the Linearly Separable SVM.

Linearly Separable SVM: In which data points are categorized in two types by and can be visualized on a single straight line.

Non-Linearly Separable SVM: In which data points cannot be located on a single straight line and are not linear. For maximization in Non-Linearly Separable, SVM kernels are used to convert the Low Dimensional data to High Dimensional data.

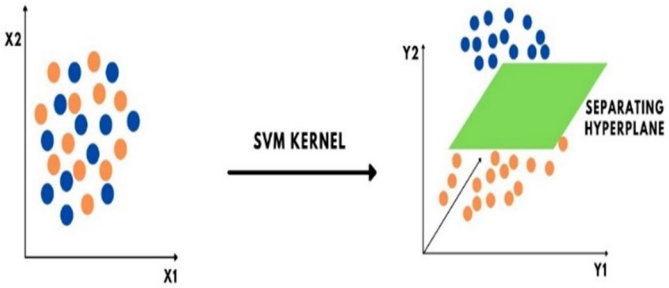


Figure 2. Non-Linearly Separable Hyperplane

The equation for Hyperplane is:

$W^T x+b=0$ (1)

When predicted class is +1 then:

$W^T x+b= +1$ (2)

When predicted class is -1 then,

$W^T x+b= -1$ (3)

Where, W is the orthogonal vector to the planes +1 and -1, x is the point on the respective planes and b is any bias term.

D. Multinomial Naïve Bayes

In this, we will further discuss the theory behind the Naïve Bayes Classifiers and their implementation. Naïve Bayes classifier is the assembly of algorithms that is based on Bayes theorem, which states that each pair that must be classified is independent of others. Naïve Bayes classifier is the probabilistic approach which assumes that features of a particular class are independent of other features of that class. Multinomial Naïve Bayes (MNB)is a naïve Bayes classifier mainly used in text classification. MNB considers the language as Bag full of words and every message is a random handful of them that is why it is mostly used in NLP.

To understand MNB, we must know the concept behind Baye's Rule:

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)*P(\frac{B}{A})}{P(B)} \quad (4)$$

Here, P(A) is prior probability of occurring A, P(B) is prior probability of occurring B, $P\left(\frac{B}{A}\right)$ is posterior probability of B when A has occurred and $P\left(\frac{A}{B}\right)$is posterior probability of A when B has occurred.

Prior probability is the rational evaluation of the probability of an outcome that relies on the contemporary knowledge before, and experiment is performed. Whereas the posterior probability is the probability of an event occurring when another event has already occurred. It is also known as Conditional Probability.

Multinomial Naïve Bayes contemplates a feature vector where a given term indicates the number of times it appears i.e., frequency. It uses the word Count Vectorizer to perform this task. MNB has a low computation cost and can effectively work with large datasets.

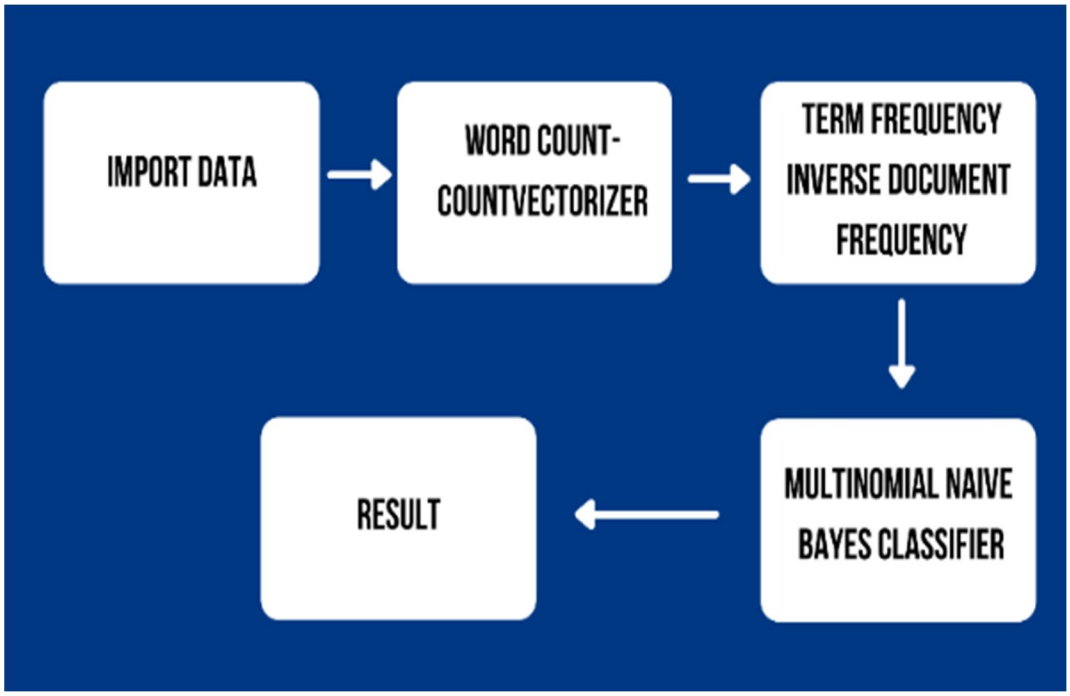


Figure 3. Working of MNB

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

II. LITERATURE REVIEW

S. Ruan et al. [1] focused on feature weighting approaches for Multinomial Naïve Bayes (MNB) classifier. According to their research, the feature weighting is divided into two categories: General Feature Weighting, in which equal weight is assigned to all the classes and class-specific feature weighting approachesinequitably assign each feature a specific weight for each class. The authors contemplateda new Class-Specific Deep Feature Weighting (CDFW)for each class in which every feature is assigned with a specific weight. It also estimated the conditional probability of the text classifier by determining feature weighted recurrences from training data. The accuracy is obtained by using 10-fold cross-validation method.

The authors [6] have done research about finding reasons and statistical analysis of airline accidents in Pakistan. They have used a supervised learning model, i.e., Naïve Bayes Classifier which according to them produces better predictions when there is fewer data in the dataset and in their case, dataset contains only 22 data entries. Their study helped in finding out causes of accidents and suggests how accidents can be reduced. The results have been shown in percentages by using some factors like pilot/crew error, systems failure, weather, unknowns, and stall/runways which are causing airline accidents.

In this paper [10], the primary idea of authors was to know how people feel about specific topics that can be considered as a classification task. The researchers have used Natural Language Processing which may tell the significance of words as well as symbols like sad and happy emoticons.

Feelings of people can be neutral, positive,or negative. In their study, the authors have focused on three classifiers: Perceptron, Naive Bayes and Rocchio. To compare the performance of classifiersFacebook status updates are predicted as negative or positive. They have collected data of 90 users. The statuses were then accordingly labelled as positive or negative. The results are shown by comparing Precision values, Recall values and F-scores for each classifier.

[11] The authors have taken some random medical data to perform statistical analysis using the Principal Component Analysis (PCA) and the Data Clustering. The PCA was used in decrementing the number of studied parameters (i.e., feature extraction) and Data Clustering to analyse the relation between diagnosed data and the patient conditions data. The authors have performed Agglomerative Hierarchical Clustering whichbegins with each element as a separate cluster and merge them into successively larger clusters. Theresults of hierarchical cluster analysis are shown graphically which is known as Dendogram, is a structure represented as a fusion of two branches of the tree into one.

As the use of sentiment analysis in business environment is increasing[14] The authors have elaborated mainly two statistical supervised machine learning algorithms in the paper: K-Nearest Neighbour (KNN) and Naïve Bayes. They havefocused onweb crawling framework of sentimental contents of movie reviews and hotel reviews and analysis. Theresults of comments, feedbacks or critiques sentiments providesuseful measures for many different purposes and can be categorized by polarity: positive one or a negative one. In results,Naïve Bayes gave better results than K-NN but for hotel reviewsalgorithms gave lesser, almost same accuracies.

[16] Mochamad Wahyudi and Dinar Ajeng Kristyanti have done research on a diverse dataset of smartphone products reviews which produced classifications of text in the negative or positive reviews using a Supervised Classification Technique, Support Vector Machine (SVM). According toauthors, the electing and setting parameters in SVM significantly affects the results of accuracy, therefore in this research they have used particle Swarm Optimization in-order to increase the classifications accuracy Support Vector Machine. The evaluation was done by using 10-Fold Cross Validation and accuracy was measured by Confusion matrix & ROC curves.

[17] K. Singh et al. have formulated a metric based on the ordinary words used in the social networking sites (like Instagram, Facebook, LinkedIn, Twitter, Flickr etc.). They have compared the performance of Simple and Spectral K-means algorithms to find textual similarity, WordNet to group words together based on their significances.

The authors have performed different methods of data mining like pre-processing, stop words removal, stemming, lexical analysis and data extraction. Spectral clustering has been used to map the original data into a vector space spanned by a few Eigen vectors and applied the K-means algorithm in the scope. The results have shown that both algorithms have given almost similar output, but more accurate algorithm was Simple K-Means.

As the social media has taken commendable pace, D. D. Das et al. [20] have accomplished a research study to establish the opinion mining for the airlines twitter data focusing on two airlines especially, using a machine learning algorithm i.e., Naïve Bayes Classification (NB) through R Studio for cleaning of data, Rapid Miner for classification of data and used Natural Language Processing (NLP) classes & methods. The tweets have been extracted and pre-processed, then the results have been shown by categorizing them into Positive, Negative and Neutral.

©IJRASET: All Rights are Reserved

Table 1. Comparison Table of Literature Review

	Name of Paper	Technique Used	Dataset	Results
1.	Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers.	Multinomial Naive Bayes, 10-Fold Cross Validation	19 random datasets	CDFW shows accuracy of 86.05%
2.	PIA Accidents Analysis Using Naïve Bayes Classifier.	Naïve Bayes Classifier, K-Means Clustering	Airlines Accidents Data	18.02% of crashes occur due to landing or runways.
3.	Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning.	Naïve Bayes, Rocchio and Perceptron Classifiers	Facebook Statuses	Rocchio classifier have best performance and Perceptron the least.
4.	Feature Extraction for Agglomerative Clustering.	Agglomerative Clustering	Amino Acids	Classified the amino acids into 3 families: GLY, THR, VAL
5.	Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier.	K-Nearest Neighbors (KNN), Naïve Bayes	Movie and Hotel Reviews	Naïve Bayes performed better than KNN in both datasets.
6.	. Sentiment Analysis pf Smartphone Product Review using Support Vector Machine Algorithm-Based Particle Swarm Optimization.	Support Vector Machine	Data Amazon Product Reviews	The results show an increasing accuracy of SVM from 82% to 94.05%.
7.	Clustering of people in social network based on textual similarity	Simple K-means and Spectral K-means Clustering	Data from Social Networking Sites	Both algorithms have shown almost same performance.
8.	Sentimental Analysis for Airline Twitter data.	Naïve Bayes Classifier	Twitter data of Jet Airways, Emirates and United Airlines.	It was classified that 395 tweets were negative, 187 neutral and 716 positives for United Airlines

III. RESEARCH METHODOLOGY

A. Feature Extraction

The NLP technique works well and emphasizes rare words rather than treating all words the same in the case of the binary bag of words model.

Steps in NLP:

1) *Lexical Analysis:* Identifies and analyse the structure of words and splits the whole text data into words, paragraphs, and sentences.
2) *Syntactic Analysis:* Analysethe words in a sentence and arrange them in-order to that showthe relationship between them.
3) *Semantic Analysis:* Extracts the most appropriate meaning of sentences from the text.
4) *Discourse Integration:* It operates with meanings of sentences based on the sentences used before.
5) *Pragmatic Analysis:* Analyses and extracts the meaning of text data in the context.

The tasks performed by NLP are as follow:

a) *Tokenization:* In this, the strings are broken into tokens which in-turn are small structures or units that can be used for tokenization.
b) *Stemming:* The process of normalizing the words into its original form or root form is known as stemming. It works by changing the end or beginning of the word considering the common prefixes that can be found in effective word. But this can be successful on some occasions not all.
c) *Lemmatization:* Lemmatization focuses on the Morphological analysis of words. To do so, it is mandetory to have detailed lexicon which the algorithm can work through to link the firm to its original or root word which is known as Lemma.
d) *POS Tags:* The paths of speech like adverbs, verbs, nouns, articles, adjectives etc. It checks the word grammatically to the sentence. But as the word can have more than one part of speech based on the context, which is used, so to solve this problem we have Named Entity Recognition.
e) *Named Entity Recognition:* It is used in detecting name entities such as person's name, organization, location, quantities, and monetary values. It has three steps: The first one is Non-Phrase Identification, second is The Phrase Classification and the last one is Entity Disambiguation.
f) *Chunking:* The individual pieces of information when grouped together are known as Chunks. In words of NLP, chunking means grouping the words into chunks that helps in determining meaningful information and insights from the sentences.

Natural Language Processing is imperative for our system because tweets are categorized by a noisy text containing undesired data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The tweets are classified into two polarities: positive and negative. For research, a set of data was made which is then prepared for processing. Subsequently, the model was trained and evaluated.

A. Data Collection and Pre-processing

Firstly, the dataset has been downloaded from a web source (KaggleInc), then it is further processed in Jupyter Notebook from Anaconda Platform. The dataset consists of 6 columns (class_label, id, date, flag, userid, tweet) and about 1.6 million rows. The creators of the dataset used the Twitter search API to collect the tweets by using search keyword.

	class_label	id	date	flag	userid	tweet
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467812723	Mon Apr 06 22:20:19 PDT 2009	NO_QUERY	TLeC	@caregiving I couldn't bear to watch it. And ...
2	0	1467813985	Mon Apr 06 22:20:37 PDT 2009	NO_QUERY	quanvu	@alydesigns i was out most of the day so didn'...
3	0	1467816749	Mon Apr 06 22:21:20 PDT 2009	NO_QUERY	scarletletterm	ok I'm sick and spent an hour sitting in the s...
4	0	1467819650	Mon Apr 06 22:22:05 PDT 2009	NO_QUERY	antzpantz	@Viennah Yay! I'm happy for you with your job!...
5	0	1467820906	Mon Apr 06 22:22:24 PDT 2009	NO_QUERY	voyage2k	@localtweeps Wow, tons of replies from you, ma...
6	0	1467824967	Mon Apr 06 22:23:28 PDT 2009	NO_QUERY	playboybacon	Mo jobs, no money. how in the hell is min wag...
7	0	1467833736	Mon Apr 06 22:25:45 PDT 2009	NO_QUERY	MagicalMason	No new csi tonight. FML
8	0	1467836583	Mon Apr 06 22:26:29 PDT 2009	NO_QUERY	homeworld	@stark YOU don't follow me, either and i work...
9	0	1467838630	Mon Apr 06 22:27:02 PDT 2009	NO_QUERY	ashleyirvin	@salancaster hope your ok.

Figure 4. Dataset

The dataset has been split into two sets: Training and Test set in the 70-30 ratio, respectively. The emoticons were replaced with Text Expressions like Happy, Sad, Playful, Love and Shock. The last step of data pre-processing was removing stop-words, numbers, punctuations, white spaces, and stemming. Then every tweet is labelled as positive(value=1), or negative(value=0) class as shown below:

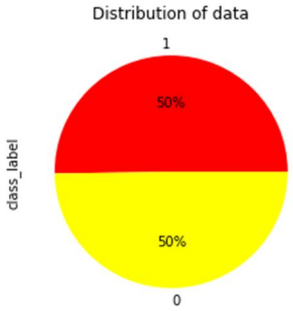


Figure 5. Sentiment Analysis

V. RESULTS AND EVALUATION

The performance of the proposed algorithm is analysed using four parameters i.e., Accuracy, Recall, Precision, and f-score for evaluating the performance of Opinion Mining.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$F1\,Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (8)$$

Here, TP: True Positive, TN: True Negative, FN: False Negative, and FP: False Positive.

	PRECISION		RECALL		F1-SCORE		ACCURACY	
Classifier	0	1	0	1	0	1	Weighted Average	Macro Average
SVM	0.78	0.77	0.76	0.79	0.77	0.78	0.78	0.78
MNB	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77

Figure 6. Performance Metrics of Algorithms

The above table shows the performances of the Algorithms by comparing their precisions, f1 score, recall values for both positive and negative class. It has been found that SVM and MNB shows almost same accuracy of 78% and 77% respectively.
In the last step, Confusion Matrix and Receiver Operating Characteristics (ROC) curves have been made.
Confusion Matrix shows the actual values targeted and the predicted values by the models. It is a N×N matrix used for evaluating performances of models which is made by using TP, TN, FP, and FN. For a Binary Classification problem, we would have a 2×2 matrix.

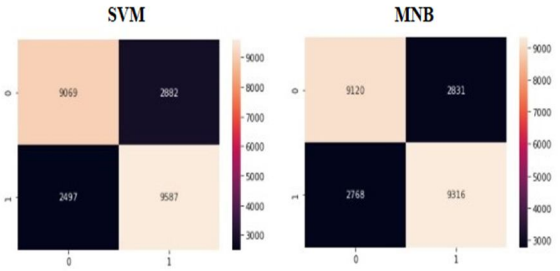


Figure 7. Confusion Matrix of SVM and MNB

ROC curves are used to measure the Area Under the Curve (AUC) [16] and divides the negative result on x-axis and positive result on y-axis. It is basically used to evaluate and compare the accuracy of classification models. Larger the area under the curve, better the predicted results.

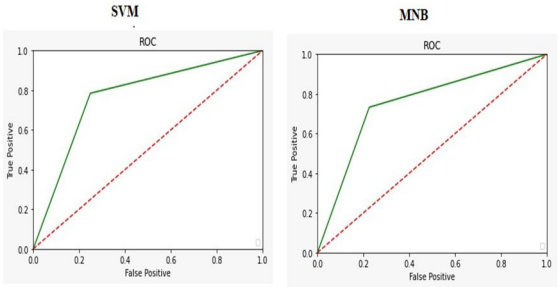


Figure 8. ROC curve of SVM and MNB

VI. CONCLUSION AND FUTURE WORK

As twitter these days became a powerful tool for communication, people share their thoughts and opinions there with the public. Opinion mining, one of the common and important analyses has been implemented on a large dataset of tweets. In the proposed work, the comparison of Support Vector Machine and Multinomial Naïve Bayes classifications has been accomplished for an automated collection of corpuses that can be used to train the model and classified the tweets into positive and negative. The tweets were pre-processed by using Natural Language Processing. The proposed system is implemented by using Python and its libraries. In the future work, it is planned to increase the accuracy of our classifiers and will focus on some more text classifier methods and techniques. Also, we will improve the pre-processing with NLP.

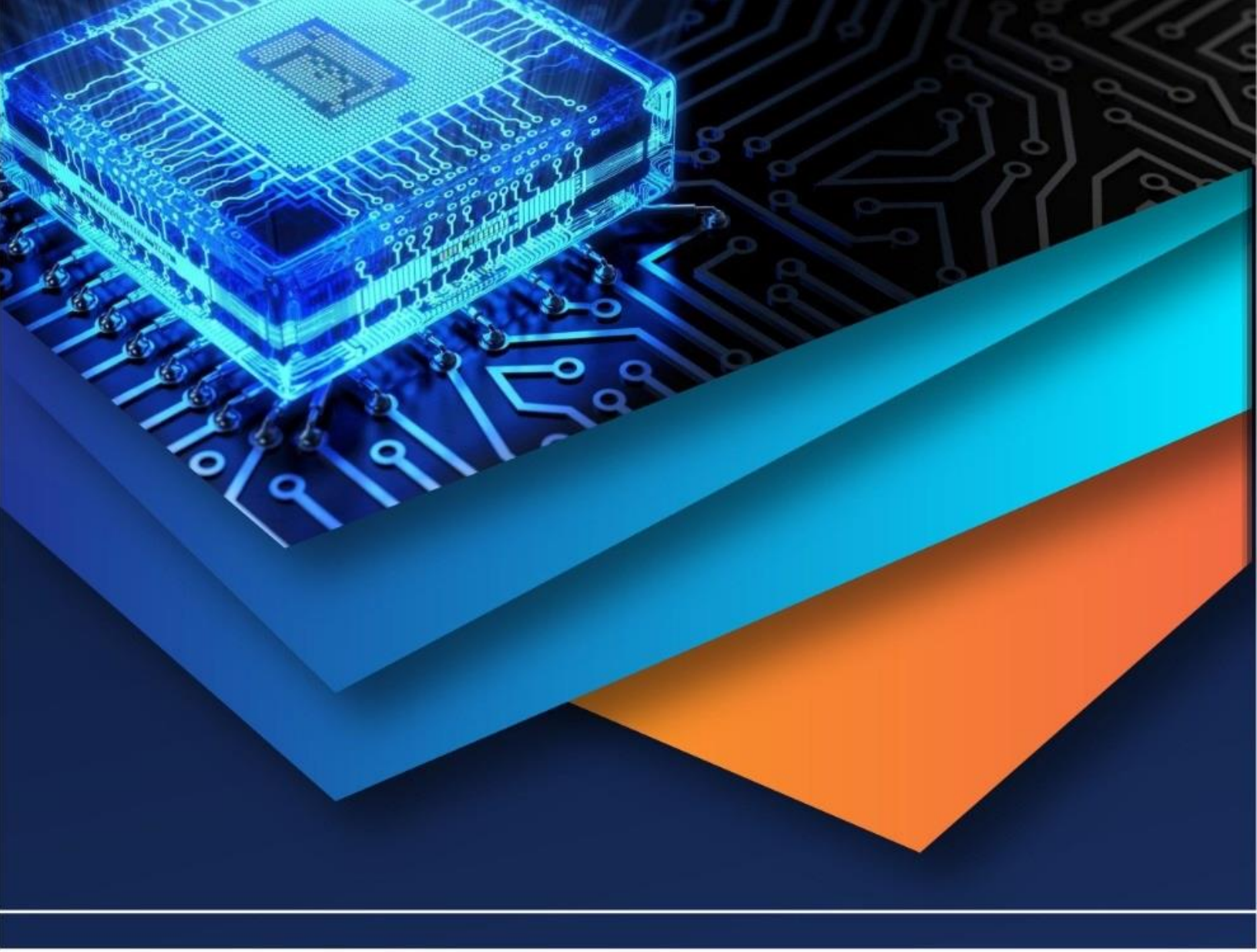
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VII. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Dr. Indu Kashyap for the continuous support of my research work, her motivation, enthusiasm and immense knowledge. Her guidance helped me in all time of research and writing this review paper. Also, my sincere thanks to my family for supporting me spiritually throughout my life.

REFERENCES

[1] S. Ruan, Hongwei Li, Chaoqun Li, K. Song, "Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers, IEEE Access Volume 8,2020.
[2] Harshada Borade, Abhijit Naik, Shraddha Yalgekar, "Twitter Sentimental Analysis", International Research Journal of Engineering and Technology (IRJET) Volume:07 Issue: 02, Feb 2020, Pg. 2280-2283.
[3] Phyu Thwe, Cho Cho Lwin, Yi Yi Aung, "Naïve Bayes Classifier for Sentiment Analysis", IJCIRAS, Vol. 3 Issue. 7, December 2020, ISSN (O) - 2581-5334.
[4] K. M. Azharul Hasan, Mir Shahriar Sabuj, Zakia Afrin, "Opinion Mining using Naïve Bayes", IEEE International WIE Conference on Electrical and Computer Science (WIECON-ECE), December 2015, pg. 511-514.
[5] J. Bhanbhro, Faiez Yousuf, S. Narejo, "PIA Accidents Analysis Using Naïve Bayes Classifier", International Conference on Computational Sciences and Technologies (INCCST'20), 17-19 December 2020, pg. 38-43.
[6] K.Srividya, A.Mary Sowjanya, T.Anil Kumar, "Sentiment Analysis of Facebook Data using Naïve Bayes Classifier", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 1, January 2017.
[7] Fiktor Imanuel Tanesab, Irwan Sembiring, Hindriyanto Dwi Purnomo, "Sentiment Analysis Model Based on Youtube Comment Using Support Vector Machine", International Journal of Computer Science and Software Engineering (IJCSSE), Volume 6, Issue 8, August 2017 ISSN (Online): 2409-4825, Page: 180-185.
[8] Troussas, Maria Virvou, K.J. Espinosa, Kevin Llaguno, Jaime Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning", IISA 2013.
[9] V. Menaka, J. Mary Dallfin Bruxella, "Feature Extraction for Agglomerative Clustering", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 8, August - 2013 IJERT ISSN: 2278-0181.
[10] Suman Rani, Jaswinder Singh, "Sentiment Analysis of Tweets Using Support Vector Machine", International Journal of Computer Science and Mobile Applications, Vol.5 Issue. 10, October- 2017, pg. 83-91, ISSN: 2321-8363.
[11] Abdeljalil El Abdouli, Larbi Hassouni, Houda Anoun, "Sentiment Analysis of Moroccan Tweets using Naive Bayes Algorithm", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 12, December 2017, Pg. 191-200, ISSN 1947-5500.
[12] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bos, Sweta Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier", I.J. Information Engineering and Electronic Business, 2016, 4, 54-62, July 2016.
[13] Bruce Walter, Kavita Bala, Milind Kulkarni, Keshav Pingali, "Fast Agglomerative Clustering for Rendering".
[14] Mochamad Wahyudi, Dinar Ajeng Kristyanti, "Sentiment Analysis pf Smartphone Product Review using Support Vector Machine Algorithm-Based Particle Swarm Optimization", Journal of Theoretical and Applied Information Technology, September 2016. Vol.91. No.1, ISSN: 1992-8645.
[15] Kuldeep Singh, Harish Kumar Shakya, Bhaskar Biswas, "Clustering of people in social network based on textual similarity", ELSEVIER Perspectives in Science (2016) 8, Pg. 570-573, July 2016.
[16] Vikas Malik, Amit Kumar, "Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-169 Volume: 6 Issue: 4, pg. 120 – 125.
[17] Arpita Lasod, Rahul Pawar, "Sentiment Analysis Using Machine Learning Techniques", IJIRT, Volume 6 Issue 7, ISSN: 2349-6002, December 2019.
[18] Deb Dutta Das, Sharan Sharma, Shubham Natani, Neelu Khare and Brijendra Singh, "Sentimental Analysis for Airline Twitter data", IOP Conf. Series: Materials Science and Engineering 263 (2017) 042067.
[19] M. Vadivukarassi, N. Puviarasan and P. Aruna, "Sentimental Analysis of Tweets Using Naive Bayes Algorithm", World Applied Sciences Journal 35 (1): 54-59, 2017 ISSN 1818-4952.
[20] Merin Thomas, Latha C.A, "Sentimental analysis using recurrent neural network", International Journal of Engineering & Technology, 7 (2.27) (2018) 88-92.
[21] Mihir Athale, Sumedh Nakod, Anmol Kumar, "Stock Analysis using Sentiment Analysis and Machine Learning", IJIRT, Volume 6 Issue 11, ISSN: 2349-6002, April 2020.
[22] Pinakee Mishra, Himanshu Pundir, "Twitter Sentimental Analysis", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume: 8, Issue V May 2020, ISSN: 2321-9653.
[23] Dalibor Bužić, Jasminka Dobša, "Lyrics Classification using Naive Bayes", 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018.
[24] https://www.kaggle.com/valkenberg/twitter-sentiment-analysis-v2



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)