International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

IMPLEMENTATION OF SUPERVISED LEARNING TECHNIQUES FOR SENTIMENT ANALYSIS OF CUSTOMER TWEETS ON AIRLINE SERVICES

Zainab Iqbal, Manoj Yadav
Department of Computer Science and Engineering
Al Falah University,
Faridabad, Haryana, India

Sarfaraz Masood
Department of Computer Engineering
Jamia Millia Islamia
New Delhi, India

Abstract— **Machine learning has evolved a long way over the decades. Classification of data, which is a supervised learning based task has been utilized for various purposes, for instance Sentiment Analysis or Opinion Mining. Sentiment Analysis is the process of gauging the views and sentiments of people in a text document as of positive, negative or of neutral polarity. The accurate and predictive classification of data of social-media, discussion forums, reviews and other platforms can be of vital significance for product analytics, recommendations and customer feedback. Micro-blogging has become a popular medium for expressing one's views and opinions and has emerged as a suitable data source for sentiment analysis because of its rich corpora. Accuracy in prediction is the ultimate goal in analysing sentiments. Therefore this paper compares the predictive performance of some of the prominent supervised learning based classification techniques such as Gaussian Naïve Bayes, Decision Tree, Support Vector Machines and Random Forest among others with the purpose of determination of the most suitable classification algorithm for sentiment analysis of a particular dataset. The dataset that is used in this research work consists of the tweets collected from the users of US airlines. Our work analyses this corpora which can contribute to enhance the quality of the airline services by finding out the best technique for its sentiment classification.**

Keywords— **Bag of words, classification, feature extraction, Supervised learning**

I. INTRODUCTION

Sentiments are the emotions and perspectives which are expressed by people. Internet is a wide source of information of such experiences and feedback of people regarding products and services. Though this data is very useful for the purpose of brand monitoring and voice of customer, but the concern lies that how to analyse the productive information from this large volume of data. Appropriate sentiment analysis is then required in such application areas which helps to categorize the sentiments from this bulk of reviews and feedback. Thus Sentiment Analysis is a task of Natural Language Processing which analyzes the underlying sentiments in the subjective text. Classification of the positive, negative or neutral orientation of the text is done through various algorithms which are divided into Machine Learning and Lexicon based approaches, as shown in Figure 1. Machine learning approach can be undertaken either using supervised or unsupervised learning. In supervised learning, partition is done of the dataset into a labelled training set and a test set. Naïve Bayes, Decision Tree, Random Forest and Support Vector Machines are some of the prominent techniques which are included in supervised learning based sentiment classification .This paper throws light on these supervised learning techniques for classification of sentiments .By the implementation of these classifiers, the sentiments of the airline customers is classified which is beneficial to improve the airline service quality. This research work has been conducted using Python and the performance is measured in the form of accuracy of the classifiers to determine the best algorithm for this dataset.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

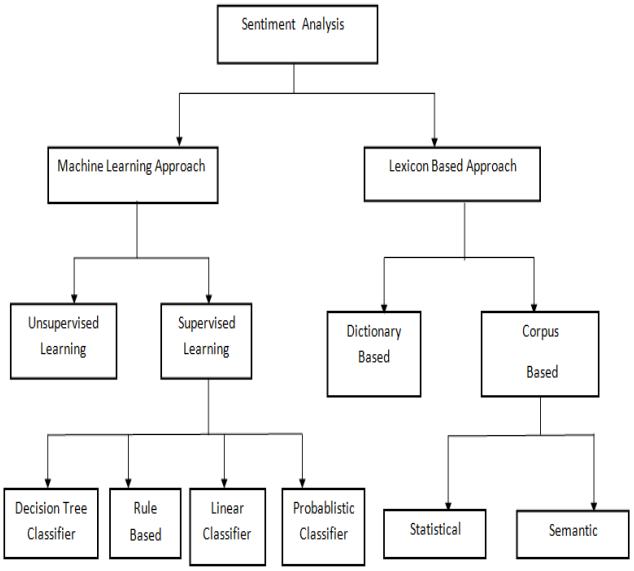

Fig 1: Sentiment Analysis Approaches

II. RELATED WORK

Many researches have been done in the past with respect to sentiment analysis which have led to several approaches and models being proposed. The work by Pang et al. (2008) includes a detailed analysis of the major aspects of sentiment analysis along with a study of methods for opinion based gathering of information.S. Bahrainian et al. (2013) proposed a technique for sentiment summarization and sentiment analysis of short text with the help of hybrid polarity detection system. There was another research work where it was proposed that performing classification of sentiments by the same indivisual ,better accuracy could be achieved using Naïve Bayes technique and this perception was also negated that Naïve Bayes is less effective as compared to Support Vector Machine[3] . A procedure for an automatic collection of a corpus has been put forth which used TreeTagger for POS-tagging and discussed the difference in distributions among positive, negative and neutral sets[4]. Another research by M. I. Zul et al. (2018) shows that Naïve Bayes with the exclusion of K-Means provides greater productivity with 80.526%-82.500% accuracy than with the combination of K-Means which had 80.323%- 81.523% accuracy . Chi-Square was used to select features for classification which indeed improved the accuracy [6]. The study discussed put forth a Poisson model for text classification through Naïve Bayes[7]. Pang etal. [8] studied the relation between finding subjectivity an classification of polarity, showing that detection of subjectivity can reduce the text into much shorter length that still hold information about polarity at the same level as that of the complete text.

III. METHODOLOGY

A. Loading the dataset –

The dataset that we have used is Twitter US airline sentiment dataset, obtained fom Kaggle [9].Twitter has become a popular social-media platform with its tweets which are referred as micro-blogs. The users have expressed their opinions and experiences for six prominent US airlines which is compiled in this dataset. This large amount of raw data when analysed in terms of sentiments can give an idea to the US airline service about its flaws and the preferences of the customers. The dataset consists of positive, negative and neutral tweets of users of US airlines along with their tweets, negative reason and confidence value of negative reason .The implementation of the methodology used has been carried out in Python using PyCharm, which is an Integrated Development Environment for Python. Figure 2 shows the first five entries of this airline dataset as displayed on output screen which shows the tweet_id and timezone of the user.

	tweet_id	...	user_timezone
0	570306133677760513	...	Eastern Time (US & Canada)
1	570301130888122368	...	Pacific Time (US & Canada)
2	570301083672813571	...	Central Time (US & Canada)
3	570301031407624196	...	Pacific Time (US & Canada)
4	570300817074462722	...	Pacific Time (US & Canada)

Fig 2 : Screenshot of Twitter US Airline Sentiment dataset

We conducted an exploratory data analysis of the dataset wherein the available data is analysed with the intent of examining the data in the form of graphs.The following graph in Figure 3 gives the polarity of tweets which can be extracted as this is a labelled dataset.The plot indicates that the positive reviews are too less in number as compared to the negative reviews.

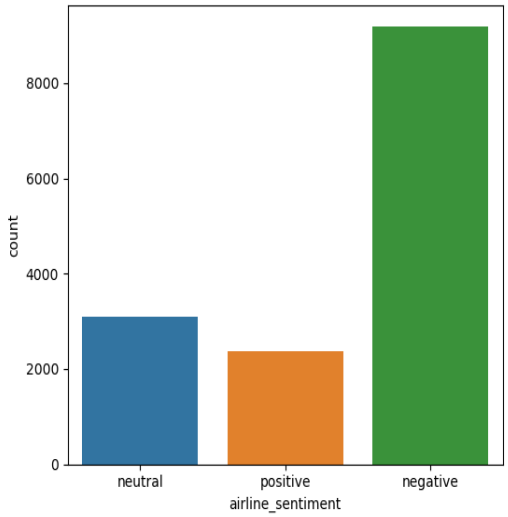

Fig 3: Polarity of tweets in the airline dataset

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

Another analysis done for the polarity of tweets per airline in Figure 4 illustrates that majority of the airlines have higher number of negative tweets than positive and negative .

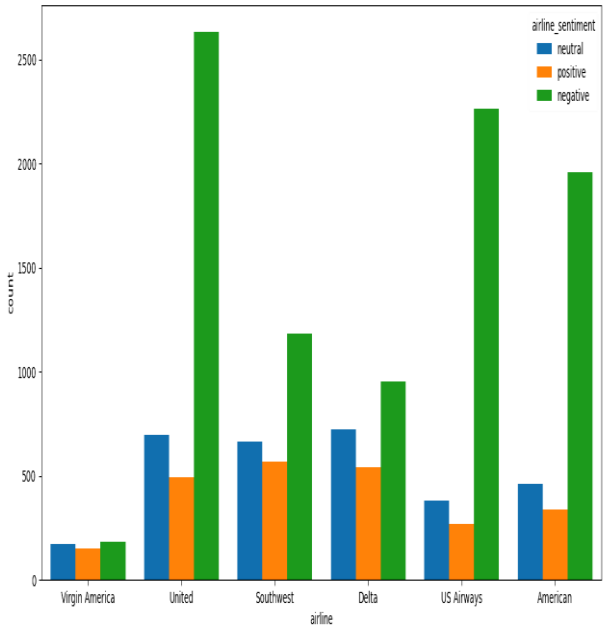

Fig 4: Type of reviews per airline

B. Preprocessing –

Tweets contain URLs, hashtags , emoticons which do not contribute to its polarity and thereby add noise to the data. The data is therefore cleaned and preprocessed into a useful and efficient format after which further steps follow. The special characters and empty spaces are removed from the data. The single characters are then left which are omitted from the corpora. Stop words are removed in the preprocessing step which include pronouns, articles which do not contribute to the meaning of a sentence and hence they are omitted to enhance the classifier accuracy. Also, the text is converted into lower case in order to maintain uniformity in the tweets. The methodology that we have worked upon is represented in Figure 5.

.

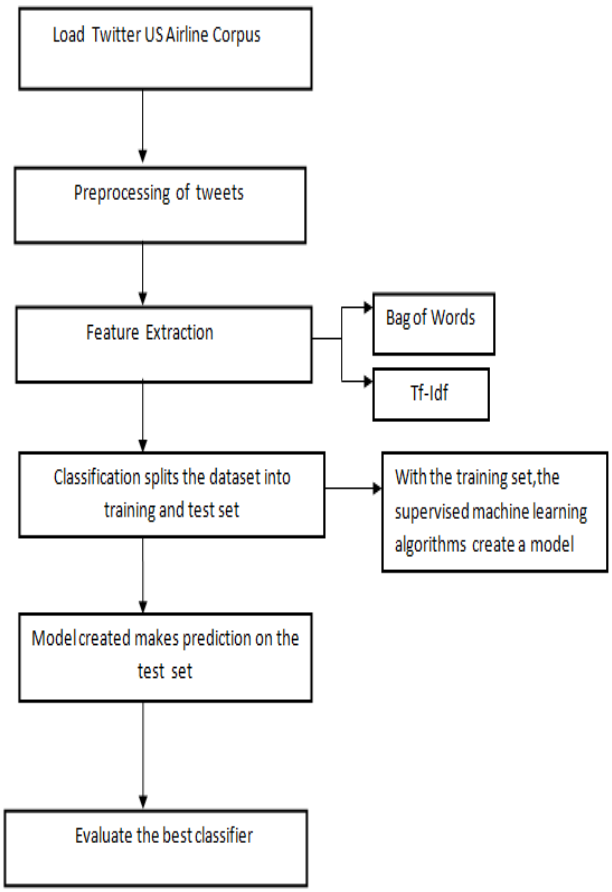

Fig 5: Proposed Methodology

C. Feature Extraction –

After the data is preprocessed and URL's, stopwords, emoticons are removed, the tweets undergo tokenization. A token represents a word in a sentence and an array is constructed for each sentence. These arrays are referred to as vectors and the words in it are known as features. A feature is an information which can be collected from the dataset. Words and their frequency, parts of speech, position of terms are some of them which help the classification algorithms to gauge the polarity of the text.

The methods for feature selection that we have used in our work are:

1. **Bag of Words Approach**: The machine learning models are trained with numbers. Hence the text needs to be converted into numeric form for which a feature matrix is constructed. Each column in a feature matrix is a distinct word of the corpus and each row represents a vector.This

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

model represents the occurrence of words within a document by representing each text document as a feature vector. Therefore if there are N distinct words throughout the tweets in the corpus, then a vector of dimension N would be created per tweet.

2. **Term frequency - Inverse document frequency (TF-IDF):** This approach also converts the preprocessed text into numbers. This scheme describes how important a word is to a review in our dataset by assigning weight to it. Unlike the bag of words method which lays emphasis on the word frequency, in term frequency, the weight of a word is proportional to its count of its occurences in the document.

TF is calculated as:

$$T\,F(t,d) = \frac{\text{Number of times term t appears in a document d}}{\text{Total number of terms in a document}}$$

Where t: term in a document

d: document in the corpus

Whereas in inverse document frequency, the words which occur in some of the documents have more weight than words which appear in all the documents of corpus.

IDF is calculated as:

$$IDF(t) = \log_e \frac{\text{Total number of documents}}{\text{Total number of documents with term t in it}}$$

Therefore, tf-idf(term frequency inverse document frequency) gives emphasis to those words for classification which appear more frequently in one document and less in all the documents of corpus.

TF-IDF is calculated as

$$TF\text{-}IDF = Tf(t,d) \times IDF(t)$$

A bi-gram model has been used along with the above two approaches where feature consists of a bi-gram, instead of a uni-gram; thereby constituting a sequence of two consecutively occuring terms of the corpus .Thus bi-gram based features are used in the numeric vectors. It has been discussed that using bigrams , the quality of features in a corpus is also improved [10].

D. Classification –

Classification step builds a model for the given input set by the use of feature vectors. Firstly, the data is divided into a training data and test data. Here in our work we have taken 70% portion of the dataset for training purpose and the rest 30% data is implemented in testing. The features extracted and the given labels are fed as training data into the supervised machine learning algorithm to create a model. Then after, the classifier model makes predictions on the test data which implies that the model will be trained on the training data and then predictions will be done on the test set. Since we are using supervised learning approach , we have applied Gaussian Naïve Baye's , Support Vector Machine, K Nearest Neighbour and Random Forest algorithm among others for sentiment classification .We have implemented the following classifiers in our research:

1. Random Forest Classifier:

Random forest algorithm constructs a group of decision trees . It then decides the final class label which is the mode of the classes of different decision trees [11]. The trees must be strong and independent so as to decrease the error rate while creating a large number of trees increases the robustness of this classifier.
This classifier is imported from the Python's sklearn library by the following code :
from sklearn.ensemble import RandomForestClassifier

2. K-Nearest neighbour Classifier:

In this non-parametric learning algorithm , the input set consists of the k nearest data points of the object to be classified.Then assignment of object is done to the most frequent class among its k closest neighbours.
K-Nearest neighbour classifier is imported from the Python's sklearn library by the following code:
From sklearn.neighborsimport KneighborsClassifier.

3. Gaussian Naive Bayes Classifier:

Naive Bayes methods use Baye's theorem and assume that all the characteristics are independent[12].This classifier performs prediction of the probablity for a feature set being associated with a particular label.

$$P\left(\frac{\text{label}}{\text{features}}\right) = \frac{P(\text{label}) * P\left(\frac{\text{features}}{\text{label}}\right)}{P(\text{features})}$$

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

Where P(label)= prior probability of label

$P\left(\frac{\text{features}}{\text{label}}\right)$ = prior probability that feature set is a label

P(features)= prior probability of occurrence of feature set

The classifier that we have chosen for our work is Gaussian Naïve Bayes which is a modification of the traditional Naïve Bayes algorithm that extends the scope of Naïve Bayes to attributes which are real valued, with a Gaussian distribution assumption. The distribution can be summarized by calculation of real-valued inputs, the mean and standard deviation of input values for each class. We studied that Gaussian Naïve Bayes responds well to text classification.
This classifier is imported from the Python's sklearn library by the following code:
from sklearn.naive_bayes import GaussianNB.

4. Decision Tree Classifier:

Decision Trees undetakes prediction of class by using yes or no conditions where each decision node contains a class label and edges represent the combination of features that result in those labels. Tweets pass through nodes and finally reach the decision node which gives appropriate sentiment classification.
The Decision Tree classifier is imported from the Python's sklearn library by the following code:
from sklearn.tree import DecisionTreeClassifier

5. Support Vector Machine

This classifier represents each data item with a point in a dimensional space of size N, when N denotes the cardinality of features. The classifier creates a hyperplane which differentiates the data points, which on lying on the hyperplane's either side, can be assigned to different classes.
This classifier is imported from the 'sklearn library' of python by the following code:
from sklearn.svm import SVC

6. Ada Boost Classifier:

Adaboost is an ensemble boosting based classifier where weights are assigned to the trained classifiers and the data sample is trained in each iteration so as to be sure of the prediction accuracy . It builds a strong classifier by combining output of the multiple weak learners which results in a strong, boosted classifier with high accuracy. AdaBoost has the highest success rate among all the boosting algorithms [13].

Adaboost classifier is imported from the Python's sklearn library by the following code:
from sklearn.ensemble import AdaBoostClassifier

IV. RESULT

The classifiers were applied on the dataset and their performance was evaluated through classification metrics. Four effectiveness measures that have been used in our work are True Positive(TP), False Positive(FP), True Negative (TN), and FalseNegative(FN).

Positive (P) : Sentiment is positive

Negative (N) : A negative sentiment

True Positive (TP) : Sentiment is positive, and is predicted as positive.

False Negative (FN) : Sentiment is positive, but is predicted as negative.

True Negative (TN) :Negative sentiment and is predicted as negative.

False Positive (FP) : Sentiment is negative, but is predicted as positive.

In our work,the metrics mentioned below determine the performance of the classifiers:

1. **Accuracy:** the ratio of all true predicted instances to all the predicted instances

$$\text{Accuracy}(A) = \frac{(TP+TN)}{(TP + TN + FP + FN)}$$

2. **Precision:.** The ratio of true positives to the total of true positives and false positives.

$$\text{Precision}(P) = \frac{TP}{(TP+FP)}$$

3. **Recall:** Recall is the portion of positives that were identified correctly.

$$\text{Recall}(R) = \frac{TP}{(TP+FN)}$$

4. **F1-score:** A harmonic average of precision and recall

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 3, ISSN No. 2455-2143, Pages 351-357
Published Online July 2020 in IJEAST (http://www.ijeast.com)

$$\text{F-Measure(Micro-averaging)} = \frac{2.(P.R)}{(P+R)}$$

The accuracy obtained after implementing all the classification algorithms on the 'Twitter US Airline sentiment' dataset is shown in Table 1. It is observed that Random Forest classifier outperforms the other classifiers with an accuracy of 77.1% using Bag of Words approach and 73.6 % using Tf-idf scheme. This results in Random Forest as the most suitable classifier for determining the sentiments for tweets of this US Airline dataset. On implementation of Random Forest classifier through Bag of Words, the values of precision, recall and F1-score is shown in Table 2 and through Tf-idf approach, the values of these metrics are shown in Table 3.

Table -1 Accuracy of classifiers

Classifier	Accuracy (With Bag of Words)	Accuracy (With Tf-idf)
RandomForest	**0.771**	**0.736**
Kneighbors Classifier	**0.557**	**0.667**
Gaussian Naive Bayes	**0.503**	**0.549**
Decision Tree Classifier	**0.699**	**0.656**
Support Vector Machine	**0.641**	**0.679**
Ada Boost Classifier	**0.725**	**0.692**

Table -2 Random Forest through Bag of Words

	Precision	Recall	F1-score
Positive Tweet	**0.92**	**0.79**	**0.85**
Negative Tweet	**0.53**	**0.78**	**0.63**
Neutral Tweet	**0.44**	**0.61**	**0.51**

Table -3 Random Forest through Tf-idf

	Precision	Recall	F1-score
Positive Tweet	**0.84**	**0.79**	**0.82**
Negative Tweet	**0.42**	**0.61**	**0.50**
Neutral Tweet	**0.48**	**0.47**	**0.47**

The values for accuracies of classifiers with both the approaches are compared in Figure 6, which shows that Random Forest which is found to have the highest accuracy among all the classifiers ,performs more efficiently with Bag of Words than with Tf-idf.

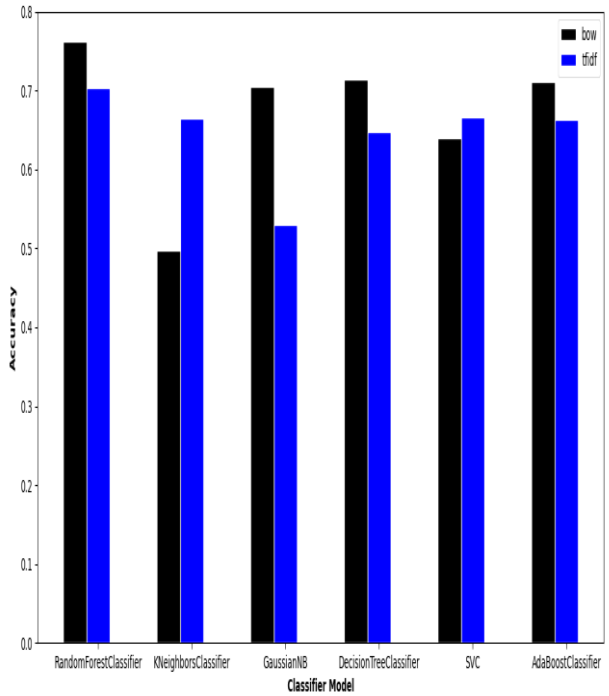

Fig 6: Comparison of Bag of Words and Tf-idf approach on Twitter US Airline Sentiment dataset

V. CONCLUSION

Through our work, we have implemented six supervised machine learning techniques on the tweets dataset and compared the performance of the classifiers. Tweets related to six major US airlines were collected fom the dataset which were further preprocessed ,the features being extracted and converted into feature vector which is trained on these classifiers. Bag of words and Tf-idf approach has been implemented along with bigrams. Through our analysis, we can determine that comparatively, Random Forest classifier,

using Bag of words is having highest accuracy of 77.1% and hence it emerges as the most preferred sentiment classification methods for the tweets by users about US airlines . Also, we found that Bag of Words works better than Tf-idf for feature extraction in this dataset .The analysis conducted in our research can be beneficial for airline service quality enhancement.Sentiment Analysis is a useful tool for users to review a service or product, or for the growth and betterment of a particular domain as in the case of US airlines.In future, other social networking sites can also be a source of data for social media analysis from where more datasets can be choosen for sentiment analysis task and the performance can be evaluated with other supervised machine learning classifiers. Sentiment Analysis has evolved during the past years with models reaching an increasing efficiency over time.Its scope in market analysis, brand monitoring and all those applications which utilize and gauge the sentiments of users is very promising in future .

VI. REFERENCE

[1]Pang, Bo and Lee , L. (2008), "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval: Vol. 2: No. 1–2, (Pg. 1-135). doi:10.1561/1500000011

[2] Bahrainian, S. and Dengel, A. , "Sentiment Analysis and Summarization of Twitter Data," (2013) IEEE 16th International Conference on Computational Science and Engineering, Sydney, NSW, 2013, (Pg. 227-234).

doi: 10.1109/CSE.2013.44

[3]Ibrahim, M.N.M. , Yusoff, M.Z.M. , "Twitter sentiment classification using Naive Bayes based on trainer perception," (2015)IEEE Conference on e-Learning, e-Management and e-Services (IC3e), Melaka, 2015, (Pg. 187-189).

doi: 10.1109/IC3e.2015.7403510

[4] Pak, A. & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of LREC. 10.

[5] Zul,M., I., Yulia.,F., and Nurmalasari, D.,"Social Media Sentiment Analysis Using K-Means and Naïve Bayes Algorithm",(2018) 2nd International Conference on Electrical Engineering and Informatics (ICon EEI), Batam, Indonesia, 2018,(Pg. 24-29).doi: 10.1109/ICon-EEI.2018.8784326

[6] Zainuddin, N. and Selamat, A. (2014). Sentiment analysis using Support Vector Machine. I4CT 2014 - 1st International Conference on Computer, Communications, and Control Technology,Proceedings.(Pg.333-337). doi10.1109/I4CT.2014.6914200.

[7] Kim, Sang-Bum and Han, Kyoung-Soo and Rim, Hae-Chang and Myaeng, Sung-Hyon. (2006). Some Effective Techniques for Naive Bayes Text Classification. Knowledge and Data Engineering, IEEE Transactions on. 18. (Pg 1457-1466). doi:10.1109/TKDE.2006.180.

[8] Pang, Bo & Lee, L. (2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. Computing Research Repository - CORR. (Pg 271-278)

doi:10.3115/1218955.1218990.

[9]https://www.kaggle.com/crowdflower/twitter-airline-sentiment

[10]Tan, Chade-Meng and Wang, Y. and Lee, C. (2002). The Use of BiGrams to Enhance Text Categorization. Information Processing & Management. 38. (Pg 529-546.) doi:10.1016/S0306-4573(01)00045-0.

[11] Ho,T. K., "The random subspace method for constructing decision forests,"(1998) in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, (Pg. 832-844), Aug. 1998

[12]Kang H. , Yoo, S., J., Han D.,(2012) "Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews", Expert Syst Appl ,39.(Pg 6000–6010, 2012.doi:10.1016/j.eswa.2011.11.107

[13] CAO, Ying and MIAO, Qi-Guang and LIU, Jia-Chen and Gao, Lin. (2013). Advance and Prospects of AdaBoost Algorithm. Acta Automatica Sinica. 39. (Pg 745–758). doi:10.1016/S1874-1029(13)60052-X.