Complex System
(6CS014)

Portfolio Task 2

Sentiment Analysis using Machine Learning

Student Id : 2039211
Student Name : Samir Husen
Module Leader : Mr. Martin Jackson
Module Lecturer : Mr. Rupak Koirala
Word Count : 2748
Submitted on : Dec 18th 2020

Table of Contents

Table of Figures

1. Introduction:

Sentiment analysis is one of the most developing and arising research areas in the field of Natural language processing (NLP). Sentiment analysis (SA) classifies the opinion expressed in a text and then analyzes it. Sentiment analysis was acknowledged when the rapid development started in the area of the social network, business, and organizations. The main aim of sentiment analysis is to classify and categorize the feedbacks and opinions provided by the user and state whether the sentiment is positive, negative, natural, or irrelevant. (Zhang, et al., 2018)

The study on sentiment analysis has emerged as the use of social media and social networks has increased according to the current context of the world. Twitter, Facebook, Blog, Product reviews are some of the popular social media platforms where a tremendous number of peoples share and exchange their opinions and sentiment analysis classifies those opinions. The two different approaches to perform sentiment analysis are, “Machine learning approach and the Lexicon-based approach where the machine learning approach is divided into supervised and unsupervised learning”. Decisions Tree, Linear, Rules-based, and Probabilistic are the classifiers of supervised learning. Support vector machine (SVM) and Neural Network (NN) are the algorithms used in linear classifiers and Naive Bayes (NB), Bayesian Network (BN), and Maximum Entropy (ME) are the algorithms used in Probabilistic classifiers. (Yang & Chen, 2017)

The sentiment analysis is the classification process that consists of different levels. The three main levels of classification are Document-level, Sentence-level, and Aspect-level. The document level intends to classify the opinion document whether it is positive or negative sentiment. Sentence-level focus on classifying the sentiment of each sentence in the document or a wholesome text. The only difference between the document and sentence level is document level classifies the whole text but the sentence classifies one sentence at a time. "Aspect-level intends to classify the sentiment concerning the specific aspects of entities” according to (Medhat, et al., 2019).

2. Aims and Objectives:

Aim:

- Research on different machine learning techniques and algorithms that are used to perform sentiment analysis and learn how the distinct supervised learning classifiers distinguish positive and negative sentiments.

Objectives:

- Research on different similar techniques that are used for sentiment analysis and compare them.
- Compare various machine-learning algorithms and check the accuracy.
- Analyze what are the impacts of sentiment analysis in today's world.

3. Literature Review:

There are different techniques implemented in machine learning to perform sentiment analysis. There are several types of research performed regarding methods and techniques used in machine learning.

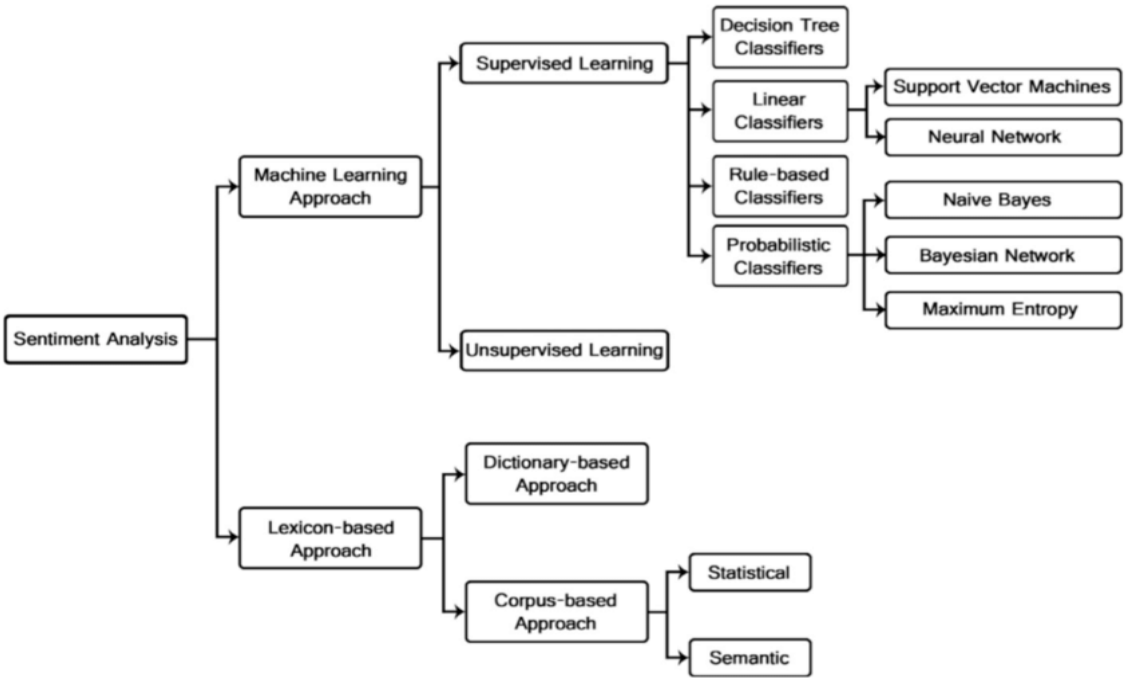

Figure 1: "Classification of Sentiment Analysis" (Yousef, et al., 2014, May)

3.1 Sentiment Classification using C4.5 (DT) Decision Tree and Naive Bayes (NB) classifiers:

In this research paper, (Prasad, et al., 2015) used decision tree classifiers of supervised learning approaches to perform a task named as SAIL (Sentiment Analysis in Indian Languages). This system was proposed to detect the sentiments of Hindi language tweets for identifying whether the tweets are negative, positive, or neutral. J48 java implementation of the Weka data-mining tool that was used here to implement the C4.5 decision tree algorithm to complete the task. The decision tree

algorithm is an unconventional version of the ID3 algorithm, which was introduced by Ross Quinlan, and practices the Shannon Entropy.

Decision tree algorithm is a classification based, which uses information gain ratio and estimated by entropy. "Pre-processing of Tweets, Model Creation and Training, and Sentiment Classification" were the systematic followed methodology in this proposed system. (Prasad, et al., 2015) Proposed Constrained and Unconstrained runs and choose two classifiers NB and C4.5 DT models. The result after the experiment on the training datasets was, "0.527 and 0.804 for Naive Bayes and C4.5 decision tree individually" according to the above researchers. The C4.5 DT algorithm performed well. Therefore, for the rest measure, they implemented C4.5 classifiers.

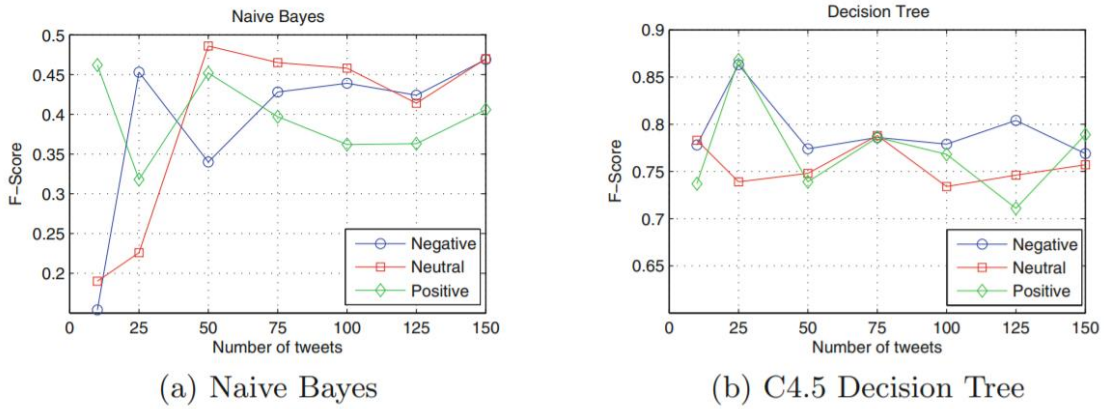

Figure 2: "Performance comparison of Naive Bayes and C4.5 decision tree" (Prasad, et al., 2015)

3.2 Sentiment Analysis using Support Vector Machine (SVM) Classifier:

The uses of text mining tools are growing rapidly according to the current context of the world. Different social network websites are implementing these tools for improvement and maintaining their service. In this paper, (Ahmad, et al., 2017) implemented the support vector machine classifiers in the Weka data mining tool for sentiment analysis. The two different datasets different company and Precision, Recall, and F-Measure measures were used throughout this experiment. The SVM is one of the most used supervised machine learning algorithms that, was introduced by Boser, Guyon, and Vapnik, which is mostly used, for the intention of classification.

Class	Precision	Recall	F-Measure
Very Negative	0.224	0.1	0.138
Slightly Negative	0.247	0.184	0.211
Neutral	0.708	0.841	0.769
Slightly Positive	0.428	0.305	0.356
Very Positive	0.278	0.237	0.256
Irrelevant	0.225	0.136	0.17
Average	0.558	0.599	0.572

Class	Precision	Recall	F-Measure
Negative	0.732	0.602	0.661
Neutral	0.729	0.859	0.789
Positive	0.548	0.376	0.446
Irrelative	0.318	0.173	0.224
Average	0.702	0.712	0.699

Datasets	Accuracy %
Self-Driving Cars	59.91%
Apple	71.2%

Figure 3: "Class wise Precision, Recall and F-Measure for First and Second Dataset and SVM Accuracy" (Ahmad, et al., 2017)

Weka data mining is a tool produced by java programming language, which is very popular among the developer for its portability and General Public License. The two different datasets were of the self-driving car and the apple products. The first dataset contains 7156 tweets and, the second dataset contains a total of 3884 tweets combination of all Negative, Neutral, Positive, and Irrelevant sentiments. After the collection of datasets, SVM operates on the normalized data for classification. The result from the first self-driving car dataset was, "The average Precision, Recall and, F-Measure is 55.8%, 59.9%, and 57.2% individually and from the second apple product dataset was the average Precision, Recall and, F-Measure is 70.2%, 71.2%, and 69.9%". The result stated that less input data accuracy was higher and further more experiments can lead towards better performance. (Ahmad, et al., 2017)

3.3 Sentiment Analysis Using Convolutional Neural Network (CNN):

(Ouyang, et al., 2015) discussed that sentiment analysis is one of the most significant measures to understand how human beings deal with textual information. The deep learning models are in use from the very beginning to obtain better performance. They used a framework introduced by Google called Word2vec for computing the vector illustration and matching the extent of a word. Word2vec and CNN were, combined to perform the sentiment analysis. Seven layers of architecture model with word2vec and three pairs of convolutional and pooling layers was used for the first time to develop this system. To increase the precision of the particular model they used, "Parametric Rectified Linear Unit (PReLU), Normalization, and Dropout technology".

The dataset, which was used during the experiment, was of the movie review excerpts. RNN and MV-RNN are also two different models of the neural network. The dataset includes 10662 sentences of movie reviews. Firstly, the words present in the datasets are converted into vectors, and then the words were identified in the set of “Google-News word vectors”, and finally, dimensional sentence vectors were applied. The accuracy from the result was about 45.4%. (Ouyang, et al., 2015) The dataset contained huge number of sentences, which also affected the outcome result of this system.

Model	Fine-gain(%)
NB	41.0
SVM	40.7
BiNB	41.9
VecAvg	32.7
RNN	43.2
MV-RNN	44.4
Our Framework	45.4

Figure 4: “CNN accuracy result compared to different machine learning classifiers” (Ouyang, et al., 2015)

3.4 Aspect-based Sentiment Analysis Using Naïve Bayes (NB):

In this research, (Mubaroka, et al., 2017) conducted research and test with sentiment analysis for analyzing the divergence and opinions of product reviews. The test was performed with the assistance and combination of data pre-processing with (POS), Chi-Square for feature assortment and, Naïve Bayes classifier for sentiment classification. Pre-processing is an initial stage in processing data for the objective of uniformity, readability, and classification method as well as for making the data mining process proper and more accessible.

This stage also contained case folding, tokenization, stop word extraction, and originating. Next was the feature selection process, where all words used in the classification phase were extricated. Aspects and sentiments were the variables for the classification phase. After training the dataset and several tests, the result showed

that the Naïve Bayes classifier performed better with the highest F1-Measure of 78.12% result. (Mubaroka, et al., 2017)

3.5 Sentiment Analysis based on Maximum Entropy (ME) and K-Means Clustering (KMC):

The study of Natural language processing is rapidly increasing according to the current context of the world where sentiment analysis is an important field to work on. (Xie, et al., 2017) performed a task where they improved the performance of the algorithm and, the innovative maximum entropy-PLSA model is used for the experiment. The model was designed as firstly, the PLSA (Probabilistic Latent Semantic Analysis) is used to obtain the seed emotion words from Wikipedia and the training corpus. Then from the above tweets, the features are obtained where the words are the input of the ME model and for training the ME model. Then the test set is concocted for emotional classification. The K-fold method divides the training set and test set. After the test it stated that the, “Precision, recall, and F-measure of MEP are 87.11%, 91.42%, and 89.21% individually” according to (Xie, et al., 2017).

	Precision (%)	Recall (%)	*F*-measure(%)
LSA	82.51	85.71	84.08
MI	80.63	81.34	80.98
PLSA	85.6	90.5	87.9

Figure 5: “Results based on different similarity calculation” (Xie, et al., 2017)

(Hamzah & Widyastuti, 2016) also performed the experiment using Maximum Entropy (ME) and K-Mean Clustering (KMC). The dataset to be classified was Indonesian textual comments, which was 2000 comments with different sentiments. TF/IDF weighting scheme was used for clustering. The result showed that K-mean clustering performed better and fast than Maximum Entropy with a precision accuracy of 0.87%.

Training	Test	Precision		Recall		Accuracy	
Data	Data	ME	K-MC	ME	K-MC	ME	K-MC
70%	30%	0.86	0.89	0.75	0.76	0.81	0.87
60%	40%	0.85	0.88	0.72	0.71	0.82	0.85
50%	50%	0.82	0.86	0.73	0.77	0.81	0.83
Average		**0.84**	**0.87**	**0.73**	**0.75**	**0.80**	**0.81**

Figure 6: "Precision, Recall and Accuracy of ME and KMC" (Hamzah & Widyastuti, 2016)

3.6 Sentiment Analysis using Stochastic Gradient Descent (SGD) and Linguistic Features:

(Gunther & Furrer, 2019) developed a sentiment analysis system for SemEval-2013 shared task on Twitter. The linear classifier was used and trained using stochastic SGD with hinge loss and elastic net regularization to make the predictions. The system also uses features like word stems, word clusters, and negation handling. During data, pre-processing simple regular expression was used for tokenization of messages. In the feature extraction stage, the extended category of linguistic and lexical features was discovered. A total of 9419 training samples was used to train the model combining all positive, negative, and neutral sentiments. The result showed that performance could improve with the use of linguistically motivated features and the precision accuracy was about 86.09 %.

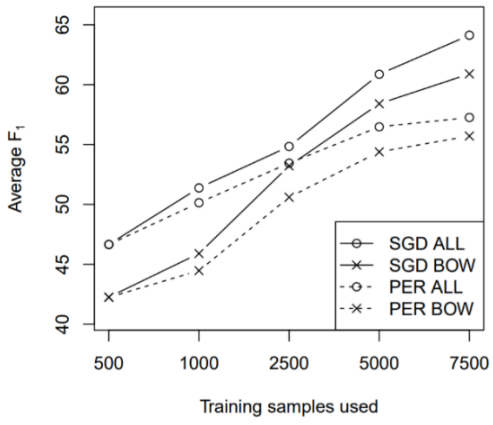

Figure 7: "Effect of training set size on different classifiers" (Gunther & Furrer, 2019)

3.7 Sentiment Analysis Implementing Random Forest (RF):

(Munshi, et al., 2020) used Random Forest for the classification with sentiment analysis in this project. On the daily basis, millions of users post comments and feedback on social networks and, then the machine learning classifiers are used to categorize the sentiments. Random forest is also known as one of the best algorithms for text classification. The steps involved in this system were followed as firstly the Hindi tweets were translated into the English language datasets. The usernames, stop-words, punctuations, and symbols were removed, from the sentences after the data visualization and analysis. For the feature extraction process BOW and TF-IDF were used and finally, the model was built using the random forest classifier. The random forest performed better than all previous techniques such as SVM and Neural Networks (NN) and more and result, from those techniques were 70% to 80% but the accuracy result of this system was about 90.24%.

3.8 Sentiment Analysis with Multi-Layer Perceptron (MLP) and Meta-Heuristic Optimization:

(Alboaneen, et al., 2017) developed a method to classify tweets using two main stages feature extraction and tweets classifications. Mutual information is the first stage that, aims to choose the most suitable conventional of features to decrease the feature extents. In the next stage, the Meta-heuristic algorithm optimizes weights and preferences of MLPs network and then applied to categorize the sentiments. The pre-processing is the stage where all the outcast URLs, #, parenthesis, slash, dash, punctuations, space, all stop word, and lower case digit or characters are detached from the words in dataset. Then in the feature mining, TF-F is used to extract the features.

After feature extraction, using Mutual information the features are selected. Finally, the Glowworm Swarm Optimization (GSO), Genetic algorithm (GA) and Biogeography-Based Optimization (BBO) are implemented with MLP to train the

tweets datasets. Therefore, this is how the system stages are arranged for the experiment. The datasets were obtained from four different companies Apple, Google, Microsoft, and Twitter and the tweets were about 1050. All three algorithms were executed, about ten times for the comparison. After the test, it shows that GSO has the best accuracy with 54%, and BBO performed poorly with a 48.86% accuracy result. (Alboaneen, et al., 2017)

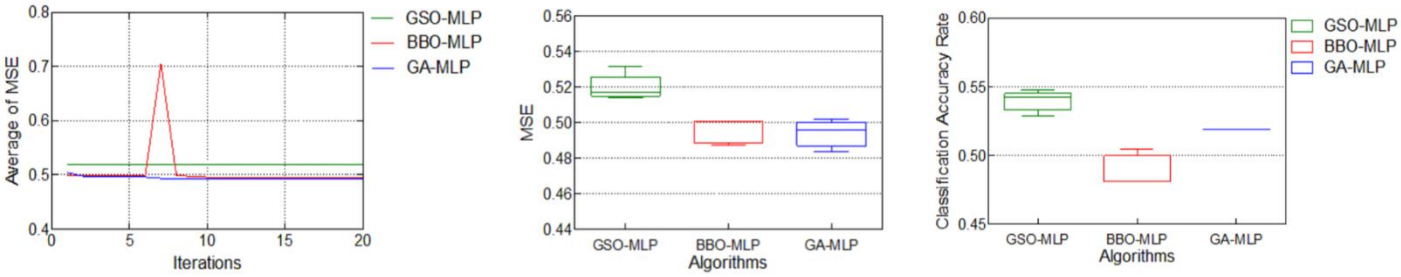

Figure 8: "Convergence curves and Boxplot charts of all algorithms" (Alboaneen, et al., 2017)

3.9 Sentiment Analysis using Multinomial Naive Bayes (MNB) classifier:

(Farisi, et al., 2019) designed a sentiment analysis system using Multinomial Naïve Bayes Classifier to analyze the hotel reviews provided by tourists so that the hotel can improve their management. The dataset for this experiment was around 3949 sentences. During the data pre-processing techniques, "Reduction of dataset dimension, case folding, removing punctuation, removing stop word, lemmatization, and tokenization are some of the data pre-processing techniques where Lemmatization is the process of converting a word into a root word for each word of tokenization." K-Fold Cross Validation is a cross-validation technique that is implemented to calculate a model where data is divided into two subdivisions which is also known as learning process data and validation data. The best experimental result was an average F1-Score of more than 91%.The result was higher than any previous method compared to Random Forest (RF). (Singh, et al., 2019) also performed a test where, “Multinomial Naïve Bayes performs slightly better than Bernoulli Naïve Bayes on the dataset”. Therefore, MNB performs better than NB classifier.

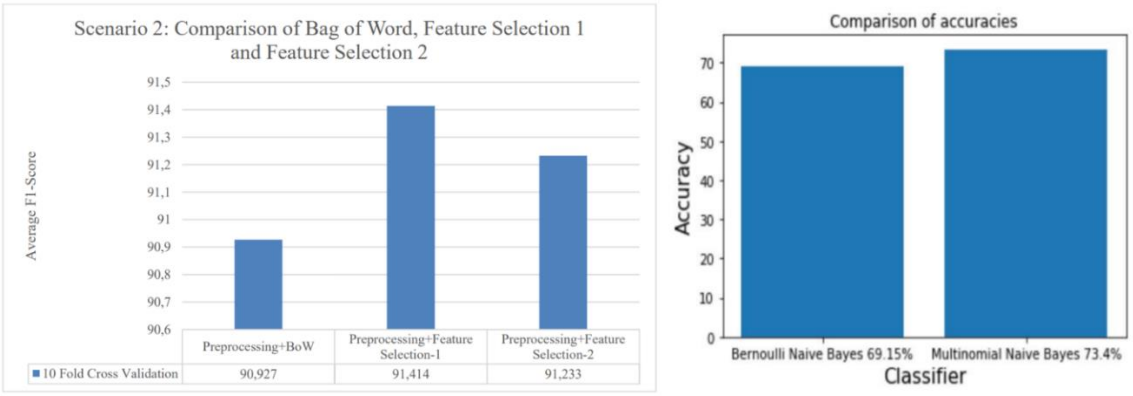

Figure 9:"MNB accuracy of first test" (Farisi, et al., 2019) "Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes" (Singh, et al., 2019)

3.10 Sentiment analysis with Natural language Processing (NLP) on Twitter:

Twitter also lies in the category of one of the biggest social media platform after Facebook. The number of tweets per hour in twitter is about 21 million expressing multiple sentiments. So, to classify those tweets this system was proposed. (Hasan, et al., 2019) developed a system using NLP based framework to classify the tweets whether it is positive or negative polarity. BoW and TF-IDF models were used to implement the system. Firstly, the API of Twitter was implemented to fetch the twitter data. Then in NLP pipeline was created where tokenization, stemming, lemmatization, POS such tasks were performed and data frame was created using BoW and TF-IDF models. After the classifier model is developed the dataset is trained and sentiment analysis was performed. The result from the test stated that the accuracy was about 85.25%.

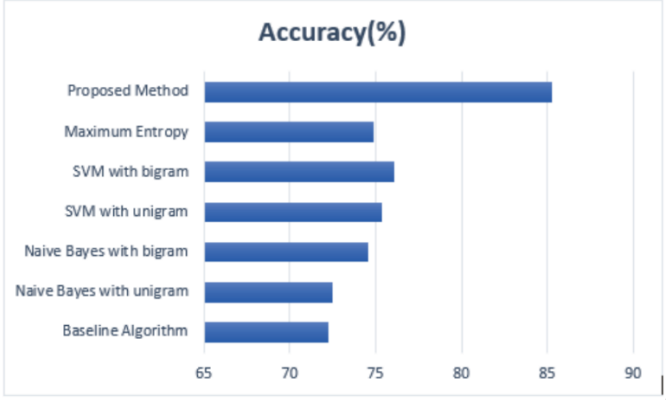

Figure 10: "Comparison of Performance Evaluation" (Hasan, et al., 2019)

4. Analysis of Findings:

After studying and analyzing different machine learning techniques from the above research papers, it states that every technique has somehow similar processes only the difference between them are the classifiers or tools used in the experiment and the accuracy result rate. Every test begins from the data pre-processing process, tokenization, lemmatization, where all the redundant words are extracted, after that feature extraction and feature selection is done, and finally, the algorithm or classifier is applied. Among all the algorithms Multinomial Naïve Bayes (MNB), using K-Fold Method and Random Forest (RF) perform well and given the highest accuracy result. The multinomial bias gain about 91% accuracy result.

Therefore, according to my research Multinomial Naïve Bayes (MNB) is the best algorithm for sentiment analysis. However, in some cases, the dataset types and numbers also make an impact on the accuracy rate. K-Means Clustering (KMC), Maximum Entropy (ME), and Stochastic Gradient Descent (SGD) also performed well. More research in this study can lead to better performance and high accuracy result rate.

5. Conclusion:

The term sentiment analysis is making an immense change in today's world, where every individual human being is using social networks and their services. It is a type of data mining tool, which differentiates the feedbacks, provided by users and categorized them according to the sentiment it expresses such, as Negative, Neutral, Positive, and Irrelevant. There are different approaches and techniques, which are used to perform the sentiment analysis. In the above sections, we discussed how different researchers performed experiments and researches regarding this topic with machine learning approaches. The comparison of different algorithms and classifiers are above with their accuracy result. However, distinct techniques have distinct accuracy results. More research and experiment can lead this topic to better performance in the future. Sentiment analysis has become a basic need software system for every business and social networks. I am glad that I get to learn and explore these machine learning techniques and their working mechanism.

6. References:

Ahmad, M., Aftab, S. & Ali, I., 2017. Sentiment Analysis of Tweets using SVM. *International Journal of Computer Applications,* 177(5), pp. 0975-8887.

Alboaneen, D. A., Tianfield, H. & Zhang, Y., 2017. Sentiment Analysis via Multi-Layer Perceptron Trained by Meta-Heuristic Optimisation. *2017 IEEE International Conference on Big Data (BIGDATA),* pp. 4630-4635 IEEE.

Farisi, A. A., Sibaroni, Y. & Faraby, S. A., 2019. Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series,* Volume 1192, p. 012024 IOP Publishing.

Gunther, T. & Furrer, L., 2019. GU-MLT-LT: Sentiment Analysis of Short Messages using Linguistic Features and Stochastic Gradient Descent. *Second Joint Conference on Lexical and Computational Semantics (*SEM),* Volume 2, p. 328–332.

Hamzah, A. & Widyastuti, N., 2016. Opinion Classification Using Maximum Entropy and K-Means Clustering. *2016 International Conference on Information, Communication Technology and System (ICTS),* pp. 162-166 IEEE.

Hasan, M. R., Maliha, M. & Arifuzzaman, M., 2019. Sentiment Analysis with NLP on Twitter Data. *International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2),* pp. 1-4 IEEE.

Medhat, W., Hassan, A. & Korashy, H., 2019. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal,* 5(4), pp. 1093-1113.

Mubaroka, M. S., A. & Aldhic, M. D., 2017. Aspect-based sentiment analysis to review products using Naïve Bayes. *AIP Conference Proceedings,* Volume 1867, p. 020060.

Munshi, A., Sapra, S. & Arvindhan, M., 2020. *A Novel Random Forest Implementation of Sentiment Analysis,* Greater Noida, India: EasyChair Preprint.

Ouyang, X., Zhou, P., Li, C. H. & Liu, L., 2015. Sentiment Analysis Using Convolutional Neural Network. *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing,* pp. 2359-2364.

Prasad, S. S., Kumar, J., Prabhakar, D. K. & Pal, S., 2015. Sentiment Classification: An Approach for Indian Language Tweets Using Decision Tree. *International Conference on Mining Intelligence and Knowledge Exploration,* pp. 656-663.

Singh, G., Kumar, B., Gaur, L. & Tyagi, A., 2019. Comparison between Multinomial and Bernoulli Comparison between Multinomial and Bernoulli. *2019 International Conference on Automation, Computational and Technology Management (ICACTM),* pp. 593-596 IEEE.

Xie, X. et al., 2017. An improved algorithm for sentiment analysis based on maximum entropy. *Soft Computing,* 23(2), pp. 599-611.

Yang, P. & Chen, Y., 2017. *A Survey on Sentiment Analysis by using Machine,* Nanjing, China : Nanjing University of Posts and Telecommunications.

Yousef, A. H., Medhat, W. & Mohamed, H. K., 2014, May. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal,* Volume 5.

Zhang, L., Wang, S. & Liu, B., 2018. Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery,* 8(IV), pp. 1-25.