Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 15 No. 1
jtec.utem.edu.my

An Ensembled Based Machine Learning Technique of Sentiment Analysis

Muhammad Garzali Qabasiyu[1, 2], Musa Ahmed Zayyad,[1] and Shamsu Abdullahi[1]
[1]*Department of Computer Studies, College of Science and Technology, Hassan Usman Katsina Polytechnic, PMB 2052, Katsina State, Nigeria.*
[2] *Department of Computer Science and Information Technology, Al-Qalam University, Katsina State, Nigeria*
garzaliqabasiyu@hukpoly.edu.ng

Article Info

Article history:
Received Sep 20th, 2022
Revised Jan 19th, 2023
Accepted Jan 20th, 2023

Index Terms:
Sentiment analysis
Machine learning
Ensemble
Social networking

Abstract

User evaluations on social networking sites such as Twitter, YouTube, and Facebook have grown rapidly due to their widespread use in sentiment analysis, providing valuable insight for both governmental and non-governmental organizations. Analyzing these evaluations not only helps improve the effectiveness of products and services, but also aids in the developing operational and management strategies. Although various analysis models have been proposed, challenges remain in processing, classifying, and accessing user evaluations, such as dealing with complex sentences, requiring more than sentiment words and achieving adequate accuracy and performance based on limited labeled data. This study primarily examines the performance of three commonly used machine learning algorithms proposes an Ensemble method, which combines Naive Bayes, Support Vector Machines and K-nearest Neihbor algorithms. The proposed method was tested on a Twitter dataset.The Ensemble method creates a classification model by applying the three classification algorithms: Naive Bayes, Support Vector Machines and K-nearest Neighbor, to for the prediction of unknown example and assigns the predicted class receiving the most votes. According to the results, the Ensemble method has an accuracy of 83.28% at 60/40 test split, 83.27% at 70/30 test split, 83.50% at 80/20 test split, and 86.12% at 90/10 test split, and F-measure of 84.15% at 60/40 test split, 83.72% at 70/30 test split, 84.41% at 80/20 test split, and 86.58% at 90/10 test split. In terms of individual performance, k-nearest Neighbor has better accuracy and F-measure than Support Vector Machine and Naive Bayes while the Ensemble method proves to be the most efficient in terms of accuracy and F-measure.

I. INTRODUCTION

Before the advent of the internet, gathering public opinion of people was a cumbersome and time-comsuming process [1]. Traditional methods, such as administering questionnaires or even verbally asking friends and family, were commonly employed by individuals and organizations alike to gather customers' feedback on products and services. However, the widespread growth and adoption of social media platforms have revolutionized this process. Nowadays, people freely share their unsolicited opinions on wide range of topics through comments on blogs and various social media platforms [2]. The raw data available on the internet can come in various forms, such as numerical or text files, and can be structured, semi-structured, or non-structured [3]. Numerous research efforts have been dedicated to developing methods for interpreting and utilizing information [4]. Sentiment analysis gained popularity as people increasingly turned to social media for sharing their opinion[5].

Social media platforms serves as a digital spaces where users can interact and connect with others without being physically present. According to [6], "Social media is an application that allow users to create and share content or participate in a social network". While this definition is familir to most people, it is important to note that the term 'social' originates from the concept of society, reflecting the role of the platform in fostering virtual community and connection.

Today's generation has embraced Twitter as one of the most popular social networking platforms, where users can post their thoughts, expressions, or beliefs [7]. Each post, or tweet, is limited to a maximum of 280 characters and can be viewed by the user's followers. Twitter users write about a wide variety of topics, including social issues, news, events, politics, and personal feelings. Twitter, an American microblogging and social networking platform, allows users to post and interacts with short messages known as tweets. Registered users can tweet, retweet and like tweets, while unregistered users can only view them [8]. Serving as a treasure trove of sentiments for users worldwide, Twitter captures thousands of actions, and opinions, on every topic imaginable, every second of the day [9]. It is also regarded as one of the biggest psychological databases, which is constantly being updated and can be used to analyze millions of data points [10].

Sentiment analysis, also known as opinion mining, is a method for evaluating spoken or written languages to identify its sentiment as positive, negative, or neutral, as well as the degree of articulation [11]. With the help of sentiment analysis, organizations and corporations can gauge customer satisfaction or dissatisfaction with products or their company's image. Furthermore, sentiment analysis plays a significant role in helping corporations, organizations and companies in making informed decisions based on data analysis [12]. Sentiment analysis employs machine learning algorithms to examine information and determine the sentiment levels in various types of data, such as news articles, blog posts, tweets, or reviews [13]. The Ensemble method for sentiment analysis has received relatively little attention. This study examines the performance of three commonly used machine learning algorithms and proposes an Ensemble method comprising Naive Bayes, Support Vector Machines, and K-nearest Neighbor algorithms. The study collected tweets related to Nigeria from Twitter, with a particular focus on tweets about the #EndSARS protest, which occurred in Nigeria from October to November 2020. The protest generated numerous tweets from both within and outside the country, providing a rich dataset for analysis.

II. LITERATURE REVIEW

The study of [14] compared the performance of three machine learning algorithms for sentiment analysis on a Nigerian Twitter dataset. The increasing use of social media platforms, such as Facebook and Twitter, has led to the generation of massive amounts of data at a high rate. This data often consists valuable information that can be used for decision-making both local and international businesses[15]. The three machine learning algorithms used in this study were Naïve Bayes, Support Vector and Maximum Entropy. The results of the experiment show that Naïve Bayes outperformed the others, with the accuracy of 83%, a precision 99%, and a recall of 83% using the Twitter datasets.

In another study, [16] presented an Ensemble classification system for Twitter sentiment analysis. The reserachers used an Ensemble classifier that combined the learning algorithms (Naïve Bayes, Random Forest, Support Vector Machine and Logistic Regression) to form a single classifier with the goal of improving sentiment classification accuracy. The results showed that the proposed Naïve Bayes had an accuracy of 73.65, Random Forest with accuracy of 70.61, Support Vector Machine with the accuracy 74.36, Logistic Regression with the accuracy 73.44 and the proposed ensemble with accuracy 74.67. This shows that the proposed ensemble classifier performed better than each stand-alone classifier.

Similarly, the study of [17] implemented the Naïve Bayes algorithm with feature selection using Genetic Algorithm for sentiment review analysis of online fashion companies. The study combined Naïve Bayes and Genetic Algorithms for feature selection with the goal of categorizing text in the of negative and positive reviews. The results were taken based on the accuracy of naïve bayes, before and after the addition of Genetic Algorithms as feature selection. The results indicated that merging the Naïve Bayes with Genetic Algorithm feature selection improved accuracy. The accuracy of Naïve Bayes algorithm was 68.50%, before using feature selection, while the accuracy after using genetic algorithm feature selection was 87.50%.

Another study by [18] used Naive Bayes and k-NN techniques to analyze the accuracy of sentiments polarity. Tweets about government activities were extracted from Twitter website using a Python-based scraper. The tweets were divided into two datasets: one is with 50 tweets and another with 200 tweets. For the dataset with 50 tweets, Naïve Bayes shows 72% accuracy, and outperformed the k-NN method. When the dataset was increased by 200, both Nave Bayes and k-NN with a value of 5 yielded the same accuracy, recall, and precision. However, when the value of k was greater than 5, Naïve Bayes showed better result than the k-NN method.

Finally, [19] conducted a study and published a paper on “Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using K-Nearest Neighbor or K-NN and Naïve Bayes”. The aim of the research was to assess customer satisfaction with digital payment services in Indonesia. Data samples were obtained from the Twitter accounts of three distinct digital payment service providers. The results revealed that the K-NN classifier algorithm had better accuracy than Naive Bayes.

A review of related works indicates that classifier performance is the most challenging issue in sentiment analysis. According to a comprehensive review on sentiment analysis, the key limitation and issue that has not been adequately addressed and evaluated is the performance of classifiers [20]. Numerous solutions have been proposed to address the classification problem in sentiment analysis. Some notable solutions that have effectively improved performance issues in sentiment analysis include [17], which uses the Naïve Bayes algorithm and Genetic algorithms for feature selection to improve classification performance. Similarly, the works of [21] improves sentiment classification by combining Support Vector Machine and Naive Bayes for sentiment analysis. Furthermore, [22] use term frequency to improve classification performance for sentiment polarity in sentences.

Sentiment analysis, like any other technology, faces technical issues. Performance is one of the most difficult aspects of implementing sentiment analysis as an alternative for traditional methods [23]. The primary technical hurdle for sentiment classification adaptation is identifying the most appropriate sentiment classifier capable of accurately classifying sentiment polarity [24]. Generally, base classifiers like Naive Bayes classifier, Random Forest classifier, Support Vector Machine, K-nearest Neighbor and Logistic Regression are being used in sentiment analysis [25]. Despite the recent increase in research on sentiment analysis, there is a need for more sentiment analysis classifiers to substantially enhance the performance of sentiment analysis in social media data. Current sentiment analysis classifiers face issues such as the inability to handle complex sentences requiring more than sentiment words and insufficient accuracy and performance due to inadequate data. The goal of this study is to propose an Ensemble classifier using majority voting to combine three base learning classifiers (Naïve Bayes, Support Vector Machine and K-nearest neighbor) to form a single classifier, with the aim of improving the performance and accuracy of sentiment classification in sentiment analysis.

III. RESEARCH METHODOLOGY

In this study, tweets related to the #EndSARS protest that took place in Nigeria were collected using R Studio software. The accuracy of the polarity of the sentiment was determined using the Rapid Miner tool, which employed Naïve Bayes, Support Vector Machine and K-nearest Neighbor classification techniques. These techniques were analyzed to determine the most efficient one among them.

Regardless of the chosen model, every data mining activity follows a specific process, which includes data collection, data pre-processing, modeling, and evaluation. The proposed method for sentiment analysis in this study involves the combination of Naïve Bayes, Support Vector Machine and k-nearest Neighbor machine learning algorithms.

A. Data Collection

Data collection is the stage of data processing where the data is collected and analyzed to determine its potential use [26]. In this study, data is collected from Twitter, while RStudio software was used to retrieve the Tweets from Twitter. At this stage, the initial data is collected, analyzed and understood so that conclusions can be drawn. Data in the form of tweets from Twitter's social media platforms is collected. Twitter offers two types of accounts: one for regular users and another for developer accounts using the Application Programming Interface (API). While regular Twitter users share and read tweets, those with developer accounts have access to the Twitter data through the API. Data can be collected using unique keys provided by Twitter, including consumer key, consumer secret key, token key, and token secret key. These keys can be used in various multitude programming languages to collect data [27].

In this study, approximately fifteen thousand tweets related to the #EndSARS protest in Nigeria from 19th October, 2020 to 21st October, 2020, were extracted from the Twitter platform. After analyzing the data, using RapidMiner tools, the resulting dataset contained 2,452 tweets ,consisting of 888 positive, 1009 negative, and 555 neutral classified sentiments. The tweet dataset used is imbalanced, meaning it does not contain an equal number of negative, positive and neutral tweets.

B. Data pre-processing

This stage encompasses all activities required to prepare data for entry into a modeling tool, derived from the initial raw data. In the text mining process, the obtained data will undergo the pre-processing stage. After data collection, the next step is to prepare it for analysis. This step involves data cleaning, data reduction, and transformation, with the following purposes:

i. Remove duplicate data and tweets that contain 'mention (@), retweet (RT)' or links that are unnecessary for the research.
ii. Create uniform data, such as changing all data to lowercase characters.
iii. Speed up the classification process due to the reduced volume of data.

C. Feature Extraction

Feature extraction is a crucial step in classification, as data mining techniques need numerical vectors as they cannot learn from raw data. This step extracts key features from raw data, typically reflecting the extracted features in a numerical form [28]. In this study, feature extraction was performed using term frequency-Inverse Document Frequency (TF-IDF), a popular technique for feature extraction in text mining research. The term frequency indicates the number of times a term has been repeated in a text, while inverse document frequency is represented by IDF.

TF-IDF is caculated as follows:

$$\mathrm{TF-IDF} = \mathrm{TF}\ (\mathrm{w}) * \mathrm{IDF}\ (\mathrm{w}) \tag{1}$$

where:

$$\mathrm{TF}\ (\mathrm{w}) = \frac{\text{No. of times word 'w' appears}}{\text{Total No. of words}}$$

$$\mathrm{IDF}\ (\mathrm{w}) = \log\ \left(\frac{Total\ No.\ of\ docs}{No.\ of\ docs\ with\ 'w'}\right)$$

D. Modeling

Modelling begins once the data pre-processing stage is complete. A modeling technique is selected from several available models, with Naïve Bayes, Support Vector machine and K-nearest Neighbor used in this study.

E. Ensemble Method

The objective of this study is to combine Naïve Bayes, Support Vector Machine and K-nearest Neighbor algorithms using majority voting ensemble methods. The method employes three base classifiers that predict the same input data. The predictions can then be used as new data to derive a single prediction of sentiment polarity. This method optimally improves the sentiment classification performance that each base classifier can achieve, yielding better results.

The majority voting ensemble machine learning model incorporates predictions from multiple models [29], increasing the performance and efficiency of the model. For classification, the predictions for each label are tallied, and the label with the highest votes is predicted.

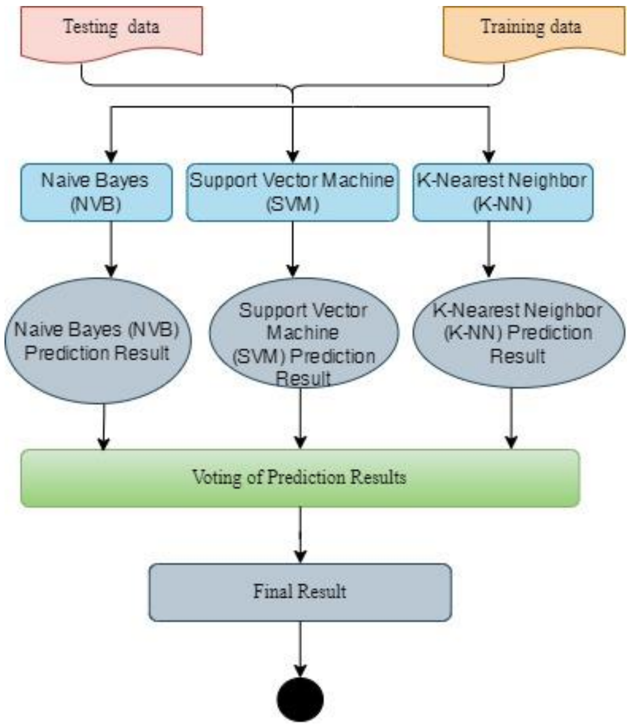

Figure 1: Proposed ensemble method

F. Performance Evaluation

The best way to measure the performance of text classification models involves measuring key metrics accuracy, precision, recall and F-measure, which are defined based on predicted and actual classes label [30]. To calculate these metrics, we must first understand the following notations, known as the confusion matrix.

Table 1
The confusion matrix

		Predicted class	
Actual class		Class= Yes	Class= No
	Class = Yes	True positive	False positive
	Class = No	False negative	True negative

Table 2
Performance metrics

Accuracy	Precision	Recall	F-measure
$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FP}$	$\frac{TP}{TP+FN}$	$2*\frac{Precision*Recall}{precision+Recall}$

IV. RESULT AND DISCUSSION

Experiments were conducted using Naïve Bayes (NVB), Support Vector Machine (SVM), and K-nearest Neighbor (KNN) classifiers to classify sentiments into positive, negative, and neutral classes. The dataset was divided into training and testing sets using 60:40, 70:30, 80:20, and 90:10 splilts, respectively. Even though the proposed Ensemble combines Naïve Bayes, Support Vector and K-nearest Neighbor, each method was also evaluated individually as well to assess their performance and efficiency. Consequently, the three methods were compared against the suggested method.

The accuracy, precision, recall, and F-measure for each techniques were calculated and analyzed in order to determine the relative performance.

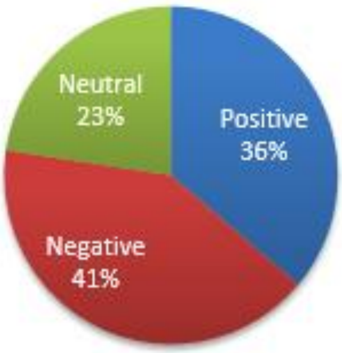

Figure 2: Distribution of sentiments

A. Evaluation result

Experiments were conducted on Twitter tweets containing #EndSARS for sentiment classification. To evaluate our approach, we first assigned a polarity for each user comment by using VADER (Valence Aware Dictionary for Sentiment Reasoning), a tool available in RapidMiner that shows the positive or negative sentiment of a sentence. Then, we used RapidMiner as a data mining tool to classify and evaluate the results of user tweets. Table 3 displays the accuracy, precision, recall, and F-measure for each split validation using the four data mining methods: Naïve Bayes, Support Vector Machine, k-nearest Neighbor, and Ensemble.

Table 3
Performance of Machine Learning algorithms and ensemble method

Test split	Classifier	Accuracy	Recall	Precision	F-measure
60/40	Naïve Bayes	64.93%	64.79%	71.73%	68.08%
	Support Vector Machine	70.44%	60.28%	52.74%	56.26%
	k-nearest Neighbor	80.73%	82.88%	83.72%	83.30%
	Ensemble	83.28%	79.89%	88.88%	84.15%
70/30	Naïve Bayes	65.44%	64.71%	72.88%	68.55%
	Support Vector Machine	71.97%	61.65%	53.18%	57.10%
	k-nearest Neighbor	79.86%	82.07%	83.00%	82.53%
	Ensemble	83.27%	79.23%	88.76%	83.72%
80/20	Naïve Bayes	65.58%	65.32%	72.93%	68.92%
	Support Vector Machine	70.47%	60.30%	52.74%	56.27%
	k-nearest Neighbor	79.43%	81.86%	83.39%	82.62%
	Ensemble	83.50%	79.90%	89.46%	84.41%
90/10	Naïve Bayes	68.16%	68.21%	75.09%	71.48%
	Support Vector Machine	72.24%	61.80%	53.25%	57.23%
	k-nearest Neighbor	79.18%	81.40%	83.02%	82.32%
	Ensemble	86.12%	82.91%	90.58%	86.58%

As shown in Table 3, when the dataset is split into a 60% for training set and a 40% a testing set, the accuracy (the ratio of correctly predicted tweets on total tweets) of NVB is the lowest among the four at 64.93%, with a recall of 64.79% (indicating that Naïve Bayes yields a 64.79% value in identifying positive tweets). The precision (the ratio of true positives to total positives) is 71.73%, while the F-measure (a balance between recall and precision) is 68.08%. The SVM has an accuracy of 70.44%, a recall of 60.28%, a precision of 52.74% and the lowest F-measure among the four at 56.26%. For the KNN classifier, the accuracy is 80.73%, with a recall of 82.88%, precision of 83.72% and F-measure of 84.30%. In contrast, the Ensemble method, has an accuracy of 83.28%. a recall of 79.89%, precision of 88.88%, and an F-measure of 84.15%.

When the dataset is split into 70% percent for training and 30% percent for testing, as shown in Table 3, the accuracy of NVB is the lowest among the four at 65.44, with a recall of 64.71%. The precision is 72.88% and the F-measure is 68.55%. The accuracy of SVM is estimated at 71.97%, with a recall of 61.65%, precision of 53.18%, and the lowest F-measure among the four at 57.10%. For the KNN, the accuracy is 79.86%, with a recall of 82.07%, precision of 83.00% and F-measure of 82.53%. The Ensemble method has an accuracy of 83.27%, a recall of 79.23%, precision of 88.76% and an F-measure of 83.72%.

Table 3 demonstrates that when the dataset is split into 80% for training and 20% for testing, the accuracy of NVB is the lowest among the four at 65.58%, with a recall of 65.32%. The precision is 72.93%, and the F-measure is 68.92%. The accuracy of SVM is measured at 70.47%, with a recall of 60.30%, precision of 52.74%, and the lowest F-measure among the four at 56.27%. For the KNN, the accuracy is

79.43%, with a recall of 81.86%, precision of 83.39% and F-measure of 82.62%. The Ensemble method achieves an accuracy of 83.50%, with a recall of 79.90%, precision of 89.46%, and an F-measure of 84.41%.

Table 3 indicates that when the dataset is split into 90% for training and 10% for testing, the accuracy of NVB is the lowest among the four at 68.16%, with a recall of 68.21%. The precision is 75.09%, and the F-measure is 71.48%. The accuracy of SVM is set at 72.24%, with a recall of 61.80%, precision of 53.25%, and the lowest F-measure among the four at 57.23%. For KNN, the accuracy is 79.18%, with a recall of 81.64%, precision of 83.02% and an F-measure of 82.32%. The Ensemble method achieves an accuracy of 86.12%, with a recall of 82.91%, precision of 90.58%, and an F-measure of 86.58%.

Among the three classifiers (Naïve Bayes, Support Vector Machine and K-nearest Neighbor) used in this study, the results showed that K-nearest Neighbor out performs Naïve Bayes in terms of accuracy. The result corresponds with the research finding of [19]. Similarly, the work of [21] proposed the Ensemble of Support Vector Machine and Naive Bayes for sentiment analysis, obtaining an accuracy of 78.31% and F-measure of 78.20% at 70% training and 30% test split of the used sample. However, when applied to the Twitter dataset in this study, the accuracy was 81.36% and F1 score was 81.66%, while at the same time, the proposed Ensemble method of this study yielded an accuracy of 83.27% and F-measure of 83.72% at a 70% and 30% test split.

The proposed Ensemble method, combining Naive Bayes, Support Vector Machine and K-nearest Neighbor outperformed the approach in [21]. This difference may be attributed to the balance and unbalance dataset used in training the model.

V. CONCLUSION

Sentiment analysis has become a crucial source for decision making, with may relying on it to achieve an efficient outcomes. Despite the large number of posts and writings on social media every day, research in this area is still limited. Analyzing social media domains is challenging due to their unique features and characteristics, which affect sentiment polarity classification.

In this study, a novel technique for analyzing social media domains based on sentiment analysis was introduced. This technique aims to enhance sentiment analysis evaluation and improve the accuracy and understanding of social media sentiment reviews. The developed technique addresses performance limitations related to sentence-level sentiment analysis and offers an automated model for evaluating sentiments. It relies on Term Frequency - Inverse Document Frequency for feature selection.

To evaluate the efficiency of the developed technique, we compared it to three well-known methods. The results demonstrate a comparison of accuracy and performance among the four techniques when applied to a Twitter dataset. The comparison highlights how the proposed technique can improve accuracy and performance. In terms of individual classifier performance, K-nearest Neighbor exhibits better accuracy and F-measure than Support Vector Machine and Naive Bayes. While the developed ensemble method proves to be the most efficient in terms of accuracy and F-measure, outperforming the individual classifiers.

ACKNOWLEDGMENT

Thanks to the authors for their immense contribution towards this study.

REFERENCES

[1] K. Ravi and V. J. K.-b. s. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-based systems,* vol. 89, pp. 14-46, 2015.

[2] D. Höttecke and D. J. S. E. Allchin, "Reconceptualizing nature-of-science education in the age of social media," *Science Education,* vol. 104, no. 4, pp. 641-666, 2020.

[3] C. I. Papanagnou, O. J. C. Matthews-Amune, and O. Research, "Coping with demand volatility in retail pharmacies with the aid of big data exploration," Computers & Operations Research, vol. 98, pp. 343-354, 2018.

[4] D. Gurkhe, N. Pal, and R. J. I. J. o. C. A. Bhatia, "Effective sentiment analysis of social media datasets using Naive Bayesian classification," *International Journal of Computer Applications,* vol. 975, no. 8887, p. 99, 2014.

[5] S. Sohangir, D. Wang, A. Pomeranets, and T. M. J. J. o. B. D. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *Journal of Big Data,* vol. 5, no. 1, pp. 1-25, 2018.

[6] S. Siddiqui, T. J. I. j. o. c. a. t. Singh, and research, "Social media its impact with positive and negative aspects," *journal of computer applications technology and research,* vol. 5, no. 2, pp. 71-75, 2016.

[7] S. F. Waterloo, S. E. Baumgartner, J. Peter, P. M. J. n. m. Valkenburg, and society, "Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp," *New Media & Society,* vol. 20, no. 5, pp. 1813-1831, 2018.

[8] A. Giachanou and F. J. A. C. S. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Computing Surveys (CSUR),* vol. 49, no. 2, pp. 1-41, 2016.

[9] H. Ismail, S. Harous, and B. J. R. C. S. Belkhouche, "A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis," *Res. Comput. Sci.,* vol. 110, pp. 71-83, 2016.

[10] M. M. CHANDIO, "TWITTER SENTIMENTS ANALYSIS," PhD diss., NEAR EAST UNIVERSITY, 2019.

[11] A. Alsaeedi, M. Z. J. I. J. o. A. C. S. Khan, and Applications, "A study on sentiment analysis techniques of Twitter data," *International Journal of Advanced Computer Science and Applications,* vol. 10, no. 2, pp. 361-374, 2019.

[12] M. Čišija, E. Žunić, and D. Đonko, "Collection and sentiment analysis of twitter data on the political atmosphere," in *2018 14th symposium on neural networks and applications (NEUREL)*, 2018, pp. 1-5: IEEE.

[13] S. t. S. Avudaiappan and T. J. I. J. O. S. T. R. V. Jenifer, ISSUE 02, "Twitter sentimental analysis using neural network," *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH , vo/. 9,* pp. 2277-8616, 2020.

[14] M. Tabra, A. J. T. I. J. o. E.-L. Lawan, and E. T. i. t. D. Media, "A Comparative Analysis of the Performance of Three Machine Learning Algorithms for Tweets on Nigerian dataset," *The International Journal of E-Learning and Educational Technologies in the Digital Media (IJEETDM),* vol. 3, no. 1, pp. 23-30, 2017.

[15] A. Telukdarie, E. Buhulaiga, S. Bag, S. Gupta, Z. J. P. S. Luo, and E. Protection, "Industry 4.0 implementation for multinationals," *Process Safety and Environmental Protection,* vol. 118, pp. 316-329, 2018.

[16] N. J. P. c. s. Saleena, "An ensemble classification system for twitter sentiment analysis," *Procedia computer science,* vol. 132, pp. 937-946, 2018.

[17] S. Ernawati and E. R. Yulia, "Implementation of the Naïve Bayes algorithm with feature selection using genetic algorithm for sentiment review analysis of fashion online companies," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 2018, pp. 1-5: IEEE.

[18] S. S. Hanswal, A. Pareek, and A. J. I. J. o. C. A. Sharma, "Twitter Sentiment Analysis using Rapid Miner Tool," *International Journal of Computer Applications,* vol. 975, p. 8887.

[19] H. Wisnu, M. Afif, and Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," in *Journal of Physics: Conference Series*, 2020, vol. 1444, no. 1, p. 012034: IOP Publishing.

[20] M. Birjali, M. Kasri, and A. J. K.-B. S. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems,* vol. 226, p. 107134, 2021.

[21] K. Korovkinas, P. Danėnas, and G. J. B. j. o. m. c. Garšva, "SVM and Naïve Bayes classification ensemble method for sentiment analysis," *Baltic journal of modern computing,* vol. 5, no. 4, pp. 398-409, 2017.

[22] A. Poornima and K. S. Priya, "A comparative sentiment analysis of sentence embedding using machine learning techniques," in *2020 6th international conference on advanced computing and communication systems (ICACCS)*, 2020, pp. 493-496: IEEE.

[23] L. Yue, W. Chen, X. Li, W. Zuo, M. J. K. Yin, and I. Systems, "A survey of sentiment analysis in social media," *Knowledge and Information Systems,* vol. 60, no. 2, pp. 617-663, 2019.

[24] Y. Al Amrani, M. Lazaar, and K. E. J. P. C. S. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Computer Science,* vol. 127, pp. 511-520, 2018.

[25] N. O. F. Daeli, A. J. J. o. D. S. Adiwijaya, and I. Applications, "Sentiment analysis on movie reviews using Information gain and K-nearest neighbor," *Journal of Data Science and Its Applications,* vol. 3, no. 1, pp. 1-7, 2020.

[26] J. Pallant, *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge, 2020.

[27] W. J. I. o. S. S. B. Ahmed, "Using Twitter as a data source: an overview of social media research tools (updated for 2017)," *Impact of Social Sciences Blog,* 2017.

[28] S. Abro, S. Shaikh, R. A. Abro, S. F. Soomro, H. M. J. S. I. J. o. C. Malik, and M. Sciences, "Aspect Based Sentimental Analysis of Hotel Reviews: A Comparative Study," *Sukkur IBA Journal of Computing and Mathematical Sciences,* vol. 4, no. 1, pp. 11-20, 2020.

[29] U. Qamar, R. Niza, S. Bashir, F. H. J. B. Khan, and I. S. Engineering, "A majority vote based classifier ensemble for web service classification," *Business & Information Systems Engineering* , vol. 58, no. 4, pp. 249-259, 2016.

[30] S. L. Marie-Sainte, N. J. J. o. K. S. U.-C. Alalyani, and I. Sciences, "Firefly algorithm based feature selection for Arabic text classification," *Journal of King Saud University-Computer and Information Sciences,* vol. 32, no. 3, pp. 320-328, 2020.