



PERFORMANCE EVALUATION OF MACHINE LEARNING ALGORITHMS FOR SENTIMENT ANALYSIS IN WEKA

Ms. Puja M. Dadhe¹, Dr. Dileep .S. Sadhankar², Dr. Ravikant .N. Jugele³

¹ Assistant Professor, Department of Computer Science,
Shivaji Science College, Nagpur, India

² Assistant Professor, Department of Computer Science,
SFS College, Nagpur, India.

³ Professor and Head, Department of Computer Science,
Shivaji Science College, Nagpur, India.

poojadadhe@gmail.com¹, dileep.sadhankar@gmail.com², rn_jugele@yahoo.com³

Abstract: The process of finding the emotional tone of a given data through textual information is considered sentiment analysis. It is a vital process in natural language processing. In this research, we will be using the WEKA tool to examine the effectiveness of different algorithms for sentiment analysis. For this purpose, we have used a balanced dataset consisting of 1000 positive and 1000 negative instances. In this paper, different algorithms are selected that include different approaches to sentiment analysis. Some of these approaches are well-known algorithms and some are recent advancements in this field. Naive Bayes, Support Vector Machine (SVM), Decision Tree, Random Forest, k-Nearest Neighbor (k-NN), and Logistic Regression are some of the algorithms that are considered in this study. In order to understand the effectiveness of these algorithms in distinguishing between positive and negative sentiments, different evaluation metrics are considered in this paper. These evaluation metrics include accuracy, precision, recall, and F1 score. From the experiments conducted using these algorithms, their effectiveness are found to be varied.

Keywords: Sentiments, Machine Learning, WEKA, NLP, SVM.

I. INTRODUCTION

User-generated data is expanding daily at a very fast rate as a result of the worldwide surge in social media and web usage. An estimated 175 Zettabytes of data will have been produced and gathered by 2026. The advent of IOT also greatly facilitates the creation of real-time data in various formats. Text, audio, video, and image are just a few of the formats in which the data is collected and monitored. One of the most important and useful types of data that can be examined to produce insightful findings is text data. Social media and the internet are constantly producing massive amounts of text data because people are eager to share their ideas, opinions, and thoughts [8].

One branch of NLP is termed sentiment analysis or opinion mining and is defined as a process of finding the sentiment or emotional information present in a given text. It is considered a vital means of understanding public opinion and customer feedback that has emerged as a result of the rise of user-generated information found on social media sites and blogs [7]. In this context, the application of machine learning algorithms to sentiment analysis has gained a lot of traction since it is considered a powerful means of dealing with a massive quantity of user-generated information [13]. A free data mining tool termed WEKA is considered a powerful platform for applying sentiment analysis algorithms. In this paper, we seek to examine different algorithms using a dataset containing 1000 positive and negative sentiments.

In order to ensure that a comprehensive evaluation of the algorithms is conducted for both polarities, the dataset is meticulously curated to ensure that it has a balanced distribution of both positive and negative sentiments. By employing different machine learning algorithms, it is possible to determine which algorithms are best suited for sentiment classification. This, in turn, will provide valuable insights into the effectiveness of these algorithms.

The next section of this paper presents a general overview of the machine learning algorithms that were selected for this research, highlighting their varied approaches and uses. Following this, a discussion on how the experiment was conducted is provided. Subsequent sections of this paper will present the results, a discussion of these results, and a conclusion that highlights observations on the effectiveness of the selected algorithms.

II. BACKGROUND

Sentiment analysis, also known as opinion mining, is a vital aspect of natural language processing (NLP) technology. It involves the use of computational techniques to identify the sentiment expressed in the text data. Understanding the emotional tone of online information is crucial for businesses, governments, and researchers to make informed decisions, forecast market trends, and gauge public sentiment (Hasan et al., 2018) [12]. Sentiment analysis has applications in diverse fields, including social media monitoring, customer sentiment analysis, and political discourse analysis (Pang & Lee, 2008) [2].

Traditionally, rule-based approaches and lexical approaches have dominated in sentiment analysis. However, with the arrival and popularity of machine learning, there has been a paradigm shift in favor of data-driven approaches in sentiment analysis (Liu, 2022) [9]. Machine learning algorithms have gained popularity in sentiment classification tasks. These algorithms use features from data to predict the sentiment.

WEKA is an open-source, widely used, and very effective data mining tool that provides an interface for data mining. Machine learning algorithms can be implemented in WEKA for sentiment analysis (Hall et al., 2009) [11]. The tool is very effective for data mining. As it is a data mining tool, it has been very rigorously used for data mining. But it has not been explored for sentiment analysis. The objective of this paper is to conduct a comparative study of various machine learning algorithms using the WEKA tool.

III. RELATED WORK

Early sentiment analysis was heavily based on lexical and rule-based systems. Pang and Lee (2008) gave the foundational basis for shifting towards machine learning and framing sentiment analysis as a text classification rather than just a keyword search [2]. Medhat et al. (2014) further studied and reviewed this transition, suggesting that as rule-based systems struggle with context and sarcasm, machine learning algorithms have ability to learn complex patterns from large datasets, making them the good and well preferred choice for modern "opinion mining." WEKA (Waikato Environment for Knowledge Analysis) has been widely emerging as a corner stone for researchers because of its huge collection of preprocessing filters and algorithms [1].

Frank et al. (2016) proposed WEKA as an beneficial and effective tool that simplifies the data mining process that is starting from data cleaning and feature extraction (like StringToWordVector) to model evaluation [4]. Study by Kaur and Gupta (2013) put mostly its significance on WEKA's strength that lies in its capability to integrate N-gram models with multiple classifiers, which is important for apprehending the nuances of sentiment in small text like social media posts and shares [6].

Ibrahim et al. (2023) explored various classifiers using WEKA, which Naive Bayes is efficient although in terms of computation and works fair with smaller datasets, SVM on the other hand typically achieve greater accuracy in high-dimensional text data due to its ability to find the optimal hyperplane separating sentiment classes [5].

A research conducted by Ankit and Saleena (2018) uses WEKA to highlight that ensemble algorithms like Random Forest often surpass individual classifiers by minimizing variance and preventing overfitting, which frequently happens in noisy text datasets. For sentiment analysis in regional languages [3]. Puttaramaiah et al. (2022) applied WEKA to implement J48, SVM, and NB, finding that the selection of an algorithms to be used some times relies on the distinctive linguistic attributes of the dataset [10].

Fresh studies, including Ibrahim (2023) and Ankit & Saleena (2018), suggest that there is no such universal algorithm that applies to every condition. While Support Vector Machines (SVM) are widely recognized for their superior classification accuracy, Decision Trees, in WEKA, provide exceptional transparency making decision making process easy to visualize and understand. These studies agree that this WEKA platform serves as a dependable skilled tool that bridges nautical data mining with practical sentiment analysis [5][3].

IV. METHODOLOGY

The present paper will be using the Application of various Machine Learning Algorithms to classify the text in the dataset into positive and negative polarity using WEKA. The purpose and objective of this research are to Compare and Evaluate various Machine Learning Models and Determine the most Accurate and Reliable Model. Machine learning classification methods are significantly important in data analysis and decision-making systems.

With the availability of data, it is critical in choosing a proper classification model to achieve high accuracy and optimal performance. In this research paper, multiple machine learning algorithms are compared using the WEKA tool to determine the most efficient binary classification model.

The research will be using a dataset of 1896 instances for binary positive/negative classification. The dataset has high dimensionality and sparsity, which is characteristic of a text-based or frequency-weighted representation for ensuring correct and fair evaluation of the model.

Methodology involves the following Steps

4.1 Data Preprocessing:

The dataset was preprocessed using the WEKA data mining tool to check the quality and suitability of the data for the machine learning models. Since the dataset has string attributes, the StringToWordVector filter was applied to convert the dataset into a numerical form. The conversion of the dataset into a numerical form means that the dataset is converted into a vector based on the frequency of the words for each document in the dataset. The following steps were taken for the data preprocessing: 1. Converting the text into a numerical form using the StringToWordVector filter. 2. Tokenizing the words using the word tokenizer filter 3. Converting the text into lower case for uniformity. 4. Removing irrelevant and noise data. After applying the above steps for the dataset, it was converted into a high-dimensional sparse form, where each attribute was represented based on the frequency of the words.

4.2 Experimental Setup:

In this section, all experiments were performed using the WEKA platform, and the data set was tested using the 10-fold cross-validation method, where the performance of the model was estimated using this method. In this method, the data set is divided into ten equal sized sets. In this method, nine sets are used for training, and one set is used for testing the model. This procedure is repeated ten times, and the performance of the model is obtained by averaging all iterations. The main purpose of this method is to obtain accurate results by avoiding overfitting during training.

4.3 Algorithms implemented:

Following Classification Algorithms were implemented using the tool WEKA

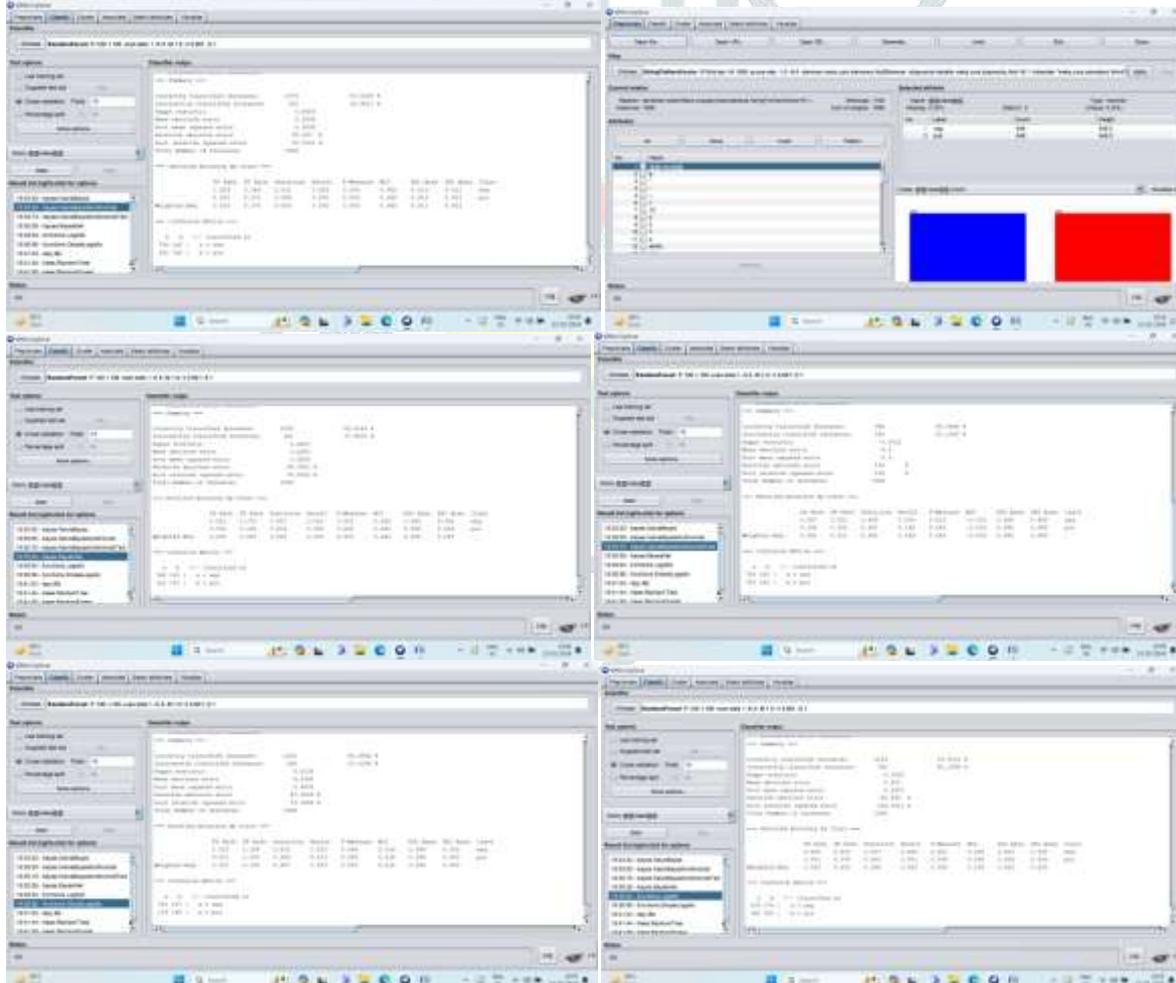
1. Naive Bayes
2. Naive Bayes Multinomial
3. BayesNet
4. Random Forest
5. Random Tree
6. IBK
7. Simple Logistic
8. Logistic Regression

4.4 Evaluation Metrics:

The performance of each model is measured using the following evaluation metrics:

Accuracy: Measures the accuracy of correct classifications. Precision: Measures the accuracy of correct predictions. Recall: Measures the accuracy of actual positives. F1-Score: Measures the harmonic mean of precision and recall. Kappa Statistic: Measures the accuracy of actual positives. ROC Area: Measures the accuracy of classification. Confusion Matrix: Measures the accuracy of classification.

Following are the screenshots of all algorithms implemented in weka:



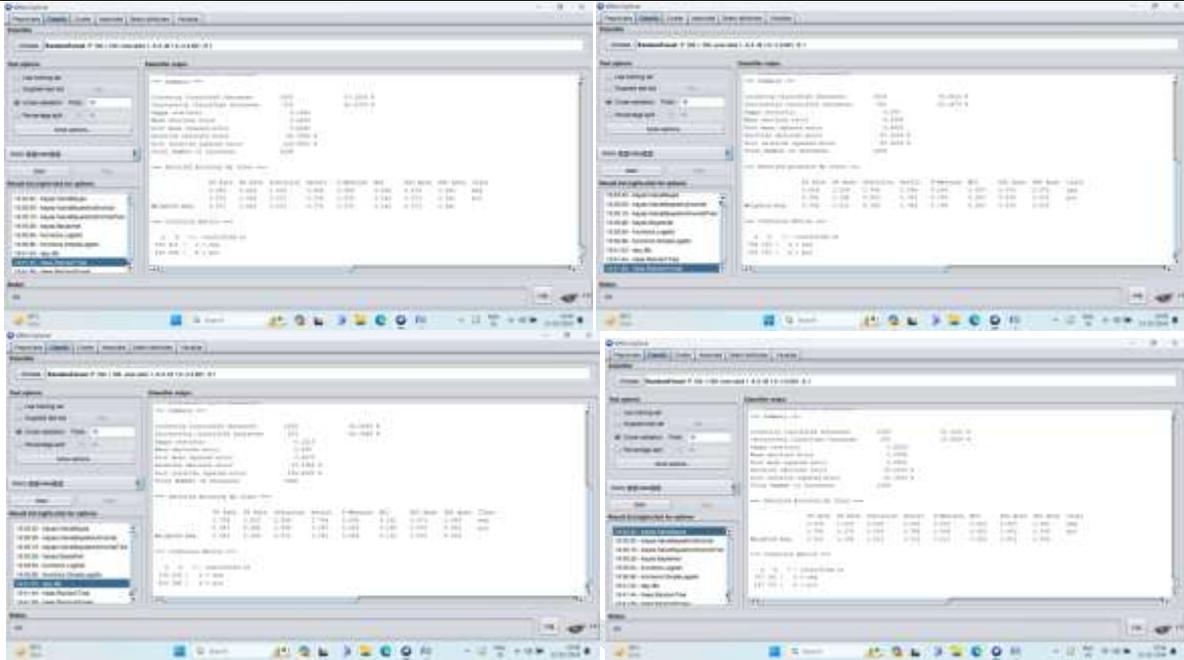


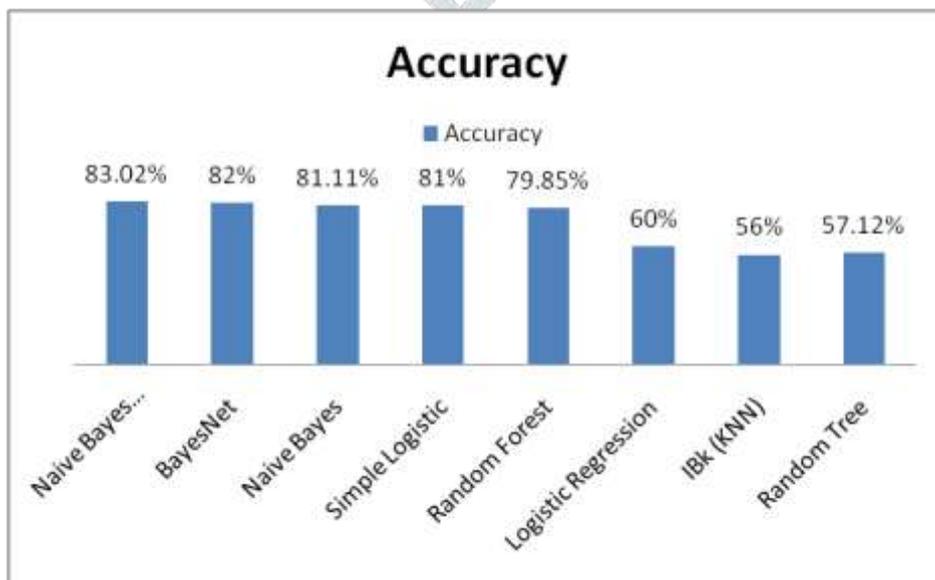
Figure 1: Output Screenshots of all Algorithms in WEKA Tool
RESULT AND DISCUSSION

The experimental results indicate that different models perform differently based on dataset characteristics. Naive Bayes Multinomial achieved the highest accuracy of 83.02% and ROC area of 0.91, making it the best-performing model. BayesNet also performed well with approximately 82% accuracy. Naive Bayes and Simple Logistic showed strong and balanced performance (81%). Random Forest achieved moderate performance (79.85%). IBk, Logistic Regression, and Random Tree showed poor performance due to high error rates and limitations with high-dimensional data.

Table 1: Accuracy, Kappa and Performance of algorithms

Model	Accuracy	Kappa	Performance
Naive Bayes Multinomial	83.02%	0.66	Best
BayesNet	82%	0.64	Excellent
Naive Bayes	81.11%	0.62	Very Good
Simple Logistic	81%	0.61	Good
Random Forest	79.85%	0.59	Good
Logistic Regression	60%	0.19	Weak
IBk (KNN)	56%	0.12	Poor

Graph 1: Plotted graph for Accuracy, Kappa and Performance of algorithms



V. CONCLUSION

The experimental evaluation of different machine learning algorithms shows that Naive Bayes Multinomial performs better than all other algorithms by achieving the highest accuracy of 83.02% and the best ROC score. It can be concluded that this high performance is due to its strong performance in handling high-dimensional and sparse feature spaces. The performance of the BayesNet classifier is also good in terms of accuracy and robust probabilistic modeling. It can be concluded that its strong ability to model dependencies between variables makes it a strong performer compared to Naive Bayes-based approaches. In addition to this, Simple Logistic and Naive Bayes also demonstrated good performance in terms of stability and consistency in different evaluation metrics.

However, the performance was slightly compromised in comparison to the best models. On the other hand, IBk (KNN), Logistic Regression, and Random Tree models demonstrated poor performance. This can be justified by the fact that these models are not effective in dealing with high-dimensional sparse data, resulting in higher error rates. The results demonstrated the significance of choosing models based on the characteristics of the underlying data. On the basis of extensive evaluation, Naive Bayes Multinomial has been selected as the final model due to its superior accuracy, efficiency, and performance. BayesNet has also been identified as a strong model, offering competitive performance with the capability to perform probabilistic reasoning.

REFERENCES:

- [1] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
- [2] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- [3] Ankit, R., & Saleena, N. (2018). Ensemble machine learning approach for sentiment analysis. *Procedia Computer Science*, 132, 815–825. <https://doi.org/10.1016/j.procs.2018.05.093>
- [4] Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques" (4th ed.). Morgan Kaufmann.
- [5] Ibrahim, A., Alzahrani, A., & Alghamdi, M. (2023). Performance evaluation of machine learning classification methods for sentiment analysis. *IEEE Access*, 11, 45210–45225.
- [6] Kaur, A., & Gupta, V. (2013). A survey on sentiment analysis and opinion mining. *International Journal of Innovative Technology and Exploring Engineering*, 2(3), 86–92.
- [7] Wang, S., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 90–94.
- [8] Monika, P., Kulkarni, C., Harish Kumar, N., Shruthi, S., & Vani, V. (2022). Machine learning approaches for sentiment analysis: A survey. *International Journal of Health Sciences*, 6(S4), 1286–1300. <https://doi.org/10.53730/ijhs.v6nS4.6119>.
- [9] Liu, B., 2022. *Sentiment Analysis and Opinion Mining*. Springer Nature.
- [10] Puttaramaiah, Monika & Kulkarni, Chaitanya & Kumar, N & Shruthi, S & Vasudevan, Vani. (2022). Machine learning approaches for sentiment analysis: A survey. *International journal of health sciences*. 1286-1300. [10.53730/ijhs.v6nS4.6119](https://doi.org/10.53730/ijhs.v6nS4.6119).
- [11] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11(1), 10–18. <https://doi.org/10.1145/1656274.1656278>
- [12] Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Mathematical and Computational Applications*, 23(1), 11. <https://doi.org/10.3390/mca23010011>
- [13] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.