iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VI **Month of publication:** June 2024

DOI: https://doi.org/10.22214/ijraset.2024.63403

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VI June 2024- Available at www.ijraset.com

Evaluate the Sentiment Analysis Performance of Several Classifiers Using Ensemble Feature Selection Method

Anupriya Singh[1], Shyamol Banerjee[2]
[1]*Research Scholar,* [2]*Faculty, Dept. of Computer Science and Engineering, SRCEM COLLEGE Banmore, Morena, M.P. India*

Abstract**: The growing availability of digital text data has sparked a need for effective sentiment analysis methods, which enable the automatic extraction of sentiment from text for various purposes. This study investigates the application of sentiment analysis using the Amazon Reviews Polarity Dataset, a curated compilation of texts categorised into positive and negative attitudes based on review ratings. Despite the dataset containing a substantial amount of labelled data, its narrow focus and classification technique have difficulties in effectively capturing nuanced expressions of sentiment. Using the Amazon Reviews Polarity Dataset as a foundation, this work does a thorough investigation of the performance of several machine learning models for sentiment analysis tasks. We learn a lot about how well different models can detect sentiment patterns by comparing them using a wide variety of performance metrics like recall, accuracy, precision, and F1 score. These metrics include K-Nearest Neighbours (KNN), Random Forest (RF), Logistic Regression (LG), and Ensemble Classifier (ECLF). Logistic Regression stands out as the most effective model, with the Ensemble Classifier coming in a close second, highlighting their promise for practical sentiment analysis applications. These findings emphasise the usefulness of the dataset and the success of the models, showing the crucial significance of careful model selection and evaluation methods in guaranteeing dependable and accurate outcomes in the field of natural language processing applications.
Keywords: Deep learning, Machine learning, Classifier, Exploratory Data Analysis (EDA), Classification.

I. INTRODUCTION

Various classifiers' performance evaluations with its comprehensive description of the research area, "Ensemble Feature Selection Scheme for Sentiment Analysis" emphasizes the significance of sentiment analysis, the challenges it poses, and the necessity of robust classification methods. With its many uses in areas as diverse as social media monitoring, marketing, and consumer feedback research, sentiment analysis—a subfield of natural language processing (NLP)—has attracted a lot of attention recently. through elaborating on how sentiment analysis has grown in importance in this age of data abundance[1]. It emphasizes the need to comprehend public opinion and sentiment towards various entities such as products, services, or events, as it plays a vital role in decision-making. The statement underscores the increasing amount of material created by users on social networking platforms, review websites, and other online forums. It emphasizes the necessity for automated sentiment analysis tools to rapidly extract valuable insights from these extensive data sources. The introduction outlines the inherent difficulties in sentiment analysis, such as the vagueness of language, sarcasm, dependence on context, and the constantly changing character of linguistic phrases[2]–[4]. These issues highlight the intricate nature of sentiment analysis jobs and need the creation of advanced computational models that can accurately capture subtle linguistic characteristics. After setting the stage, the introduction delves into the function of classification algorithms and other machine learning techniques as they pertain to sentiment analysis. Traditional classifiers like Support Vector Machines (SVM), Naive Bayes, & Decision Trees are briefly reviewed in the text, with an emphasis on their pros and cons when applied to sentiment analysis jobs. In addition, it discusses the emergence of ensemble learning as a viable method to improve classification performance by integrating many base classifiers[5]–[8]. An important contribution presented in the paper is the ensemble feature selection strategy, which seeks to enhance the performance of sentiment analysis by choosing discriminative features from a vast feature space. The introduction offers an understanding of the reasoning behind feature selection, highlighting the significance of choosing informative features while addressing the problem of having too many dimensions. The introduction elucidates the organization of the article, offering a clear plan for the next sections. The text describes the methods used to evaluate the proposed ensemble feature selection approach, which includes selecting the dataset, setting up the experiment, and determining the performance measures[9]–[11].

This study summarises the literature by outlining the significance of sentiment analysis and conducting performance evaluations of several classifiers using an ensemble feature selection scheme for sentiment analysis, discussing the difficulties it presents, and introducing the proposed method to improve classification accuracy. This article offers readers a comprehensive comprehension of the research goals, which encourages additional investigation into the effectiveness of ensemble learning and feature selection methods in sentiment analysis[12]–[15].

II. LITERATURE REVIEW

Loureiro 2023 et.al a comprehensive performance assessment system (PAS) to identify problems and offer recommendations for improving communal irrigation systems' water and energy efficiency. Then, to increase a gravity system's efficiency, the PAS is used to rank options and identify issues.

The results indicate that leaks in low-pressure pipelines and canals, along with releases from intermediate and canal reservoirs, result in significant losses of water. The gravity system's inadequate flow control and monitoring, as well as the network's advancing age, are to blame for these issues.

Due to factors such as low non-revenue water performance, water losses from releases, energy efficiency of pumping stations, and excess system energy, the gravity network area most in need of intervention might be determined by ranking the areas. It was determined that infrastructure solutions involving canal restoration and water discharge management would significantly improve overall performance, despite the high cost, after extensive investigation into several choices for this network area. After that, the PAS is employed to compare pressurised and gravity systems.

Pressurised systems make good use of water resources, but they aren't quite as efficient as gravity when it comes to running pumps. To add insult to injury, the high energy bills highlight the need for pressurised system energy optimisation strategies. Not only can the new PAS evaluate water and energy efficiency, but it may also assist managers and lawmakers in determining the pros and cons of each system[16].

Hazaa 2023 et al. Sentiment analysis is a must for gauging how people feel about items and events due to the surge in user feedback. By simplifying the data and enhancing the efficiency of learning algorithms, feature selection in sentiment analysis is a powerful tool. Arabic, in particular, calls for extensive study due to its intricate structure. Decision trees, Naive-Bayes, K-NN, and meta-ensemble approaches are combined in this research to provide a unique hybrid filter-genetic feature selection strategy. There is a considerable improvement in the performance test on Arabic datasets when compared to baseline models. The macro-average F1 score rises by 5%, indicating improved classification accuracy[9].

Balaji 2023 et al. Sentiment analysis assesses user sentiments based on textual data, which is vital in the context of online social media and networking platforms. Due to the lack of information, analysing short texts is challenging. In order to deliver accurate real-time sentiment analysis, this research presents an Ensemble Multi-Layerered Sentiment Analysis Model (EMLSA) that integrates VADER with Recurrent Neural Networks (RNNs). VADER aids in training and uses input datasets to generate sentiment predictions.

Finding the term-frequency and inverse document-frequency are both part of feature extraction. Both Word-Level Embedding (WLE) and Character-Level Embedding (CLE) enhance the analysis process. Testing EMLSA on datasets from IMDB, Amazon, eBay, and TripAdvisor is part of its performance evaluation. Its efficacy is evaluated by looking at its F1-score, sensitivity, specificity, precision, and accuracy[11].

Mohammad 2022 et al. Robust software engineering is necessary in healthcare systems to prioritise patient safety. Identifying flaws is crucial in the development of healthcare applications.

When relevant features are present, the Software Defect Prediction model (SDP) performs well; when irrelevant features are included, it performs poorly. In order to optimise feature selection in binary classification, this study demonstrates how to apply Multi objective Harris Hawk Optimisation (HHO). Combining this method with Adaptive Synthetic Sampling (ADASYN) improves classification performance.

The goal of the multi-objective Harmony Search Algorithm (HHO) is to improve the model's performance while simultaneously decreasing the amount of features. Experiment validation with healthcare data proved the model's usefulness, with a high accuracy of 0.990 and an AUC score of 0.992. Using several search techniques as RF, SVM, bagging, adaptive boosting, voting, and stacking, the proposed model demonstrates its superiority in software defect prediction[17].

TABLE 1 LITERATURE SUMMARY

Author/ year	Title	Method/ model	Parameters	References
Tasnim/2022	Performance Evaluation of Multiple Classifiers	NLP	accuracy of 94%	[18]
Pratiwi/2018	Document Sentiment Analysis Using Information Gain-Based Feature Selection and Classification	Deep learning method	Accuracy of 96%	[19]
Ghosh/2018	Evaluating the Effectiveness of Various Ensemble Feature Selection Schemes for Sentiment Analysis Classifiers	feature selection technique	Accuracy of 87%	[20]
Ankit/2018	An Ensemble Classification System for Twitter Sentiment Analysis	SVm	Accuracy of 85.3 %	[21]
Araque/2017	Enhancing deep learning sentiment analysis with ensemble techniques in social applications	deep learning sentiment analysis	F-Score of 77.73%	[22]

III. PROPOSED METHODOLOGY

The research methodology section outlines the systematic approach utilized to investigate sentiment analysis using the Amazon Reviews Polarity Dataset. The purpose of this research is to compare and contrast how well different machine learning algorithms extract emotion from plain text. Reliability and robustness are fostered by the methodology's successive processes. Obtaining the Amazon Reviews Polarity Dataset, a carefully selected corpus divided into positive and negative feelings, is the first step. As a part of the preprocessing steps to improve the dataset, text cleaning and normalisation procedures are employed. Next, the train_test_split function in scikit-learn is used to divide the dataset into two parts: one for training and one for testing. This keeps the data independent and eliminates bias. In sentiment analysis, a number of machine learning models are used, including Decision Trees, Naive Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbours. To objectively evaluate the efficacy of the model, evaluation entails computing performance indicators like as accuracy, precision, recall, and F1 score. The purpose of this work is to shed light on sentiment analysis and machine learning by means of a detailed methodology.

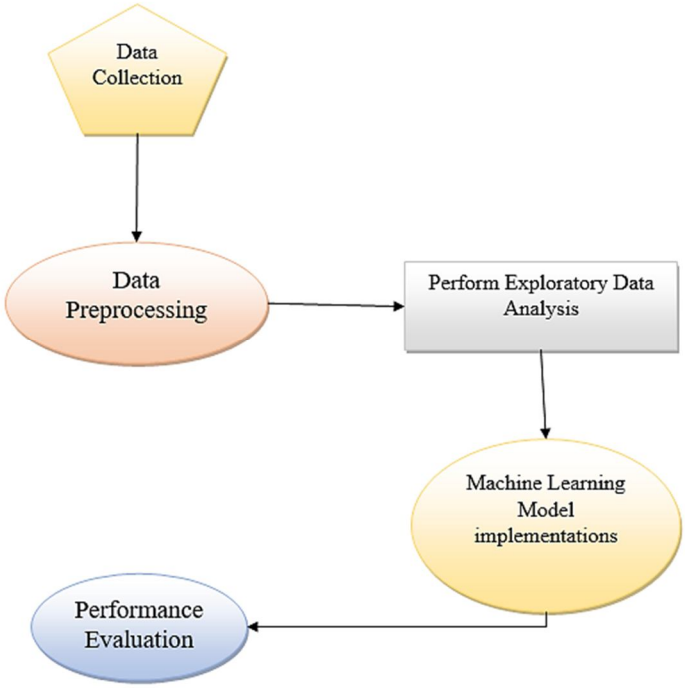


Fig. 1 Proposed Flowchart

A. Data Collection

The Amazon Reviews Polarity Dataset is a meticulously curated collection designed primarily for sentiment analysis tasks. It categorizes reviews into clear positive and negative sentiments, offering insights into consumer opinions. Reviews with scores of 1 and 2 are classified as negative, while those with scores of 4 and 5 are considered positive; instances with a score of 3 are excluded for clarity. The dataset comprises 1.8 million training samples and 200,000 testing samples, ensuring robust model evaluation and validation. It enhances sentiment analysis algorithms' reliability and efficiency. The dataset includes two crucial CSV files: train.csv and test.csv, formatted in comma-separated values. Each entry contains polarity, review title, and content. Polarity values of 1 and 2 denote negative and positive sentiments, respectively. Textual content is enclosed in double quotes, with proper escaping and newline representation for seamless parsing and analysis. Standardized formatting ensures consistency and simplifies data handling, facilitating valuable insights with minimal preprocessing.

B. Data Pre-processing

The preprocessing pipeline begins by utilizing NLTK's stop words corpus and the string library to eliminate common English stop words and punctuation marks, respectively, thereby refining the text's cleanliness and relevance. Subsequently, the Porter Stemmer algorithm is applied to standardize words to their root form, ensuring consistency and simplifying subsequent analyses. Following preprocessing, the TF-IDF vectorization process is initiated using scikit-learn's TfidfVectorizer. This part turns text data into numerical vectors, where each feature shows how important a term is in a document compared to the whole corpus. In order to govern the dimensionality and the scope of feature extraction, parameters like as max_features and ngram_range are used to customise the representation of textual information. The TF-IDF vectorizer is fitted and transformed on the training data to generate TF-IDF representations. Similarly, it's applied to the test data, ensuring consistency and compatibility in feature extraction across both datasets. This meticulous preprocessing and vectorization pipeline prepares text data for subsequent machine learning tasks, enabling efficient model training and evaluation by presenting structured numerical information amenable to analysis and interpretation.

C. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an indispensable preliminary process for understanding the nuances of the Amazon Reviews Polarity Dataset and extracting actionable insights for sentiment analysis. This method employs diverse investigative techniques to grasp the dataset's content, sentiment distribution, and underlying patterns comprehensively. Initially, assessing the dataset's structure involves scrutinizing its dimensions, data types, and summary statistics to grasp fundamental properties. Ensuring balance in sentiment classes is vital for unbiased representation and informs subsequent analyses. Visualizing the distribution of positive and negative sentiment classes enables researchers to discern dataset nature and potential biases effectively. Further exploration delves into review duration dispersion, shedding light on content diversity and depth. Analyzing word frequency distribution across sentiment categories aids in identifying recurring themes and predominant sentiments. Correlation analysis on variables like review ratings and durations uncovers potential connections and dependencies. Visualization techniques such as scatter plots, heat maps, and pair plots facilitate trend detection and outlier identification, unveiling dataset structure intricacies. EDA findings serve as a foundation for informed decisions in modeling and feature engineering, enhancing sentiment analysis results. By comprehending dataset intricacies, researchers can optimize feature and model selection, leading to more reliable and precise sentiment analysis models. EDA is pivotal for deriving meaningful insights and guiding informed decision-making in Amazon Reviews Polarity Dataset analysis.

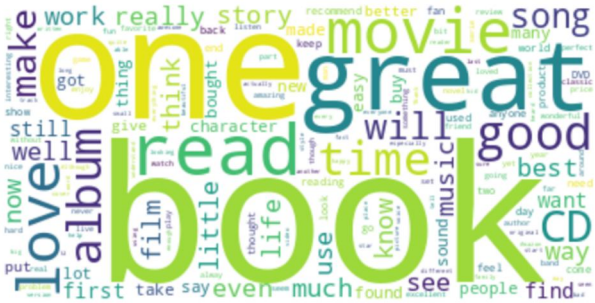


Fig. 2 word cloud for positive Sentiments

Fig 2 displays a word cloud visualization focusing on positive sentiment, visually representing frequently occurring words in the dataset. Larger words indicate higher frequency, offering a concise snapshot of prevalent positive terms within the dataset, aiding in sentiment analysis and interpretation.

Fig. 3 word cloud for negative Sentiments

Fig 3 presents a word cloud visualization emphasizing negative sentiment, showcasing frequently occurring words associated with negativity within the dataset. Larger words signify higher frequency, offering a visual summary of prevalent negative terms, facilitating sentiment analysis and insight extraction.

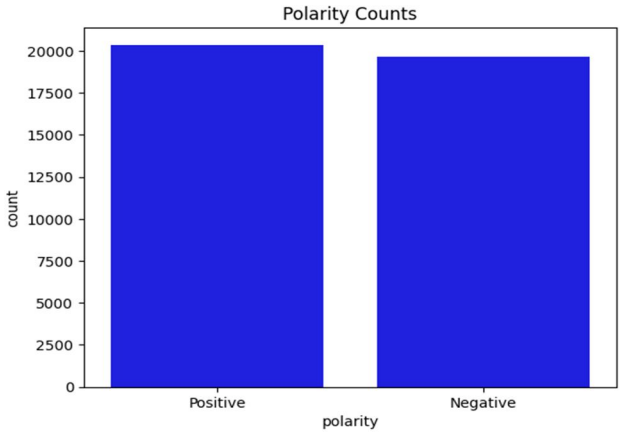


Fig. 4 Polarity counts

Fig 4 illustrates polarity counts, providing a visual representation of the distribution of positive and negative sentiments within the dataset. This bar chart or histogram showcases the frequency of positive and negative instances, offering insights into the overall sentiment distribution and balance, which are crucial for subsequent analysis and interpretation.

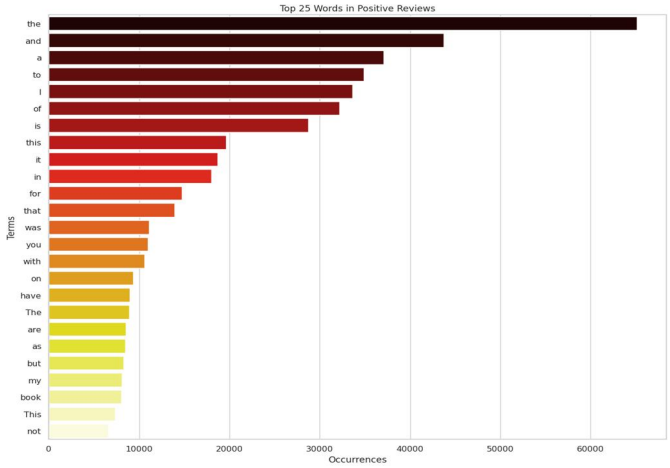


Fig. 5 Words in positive reviews

Fig 5 presents a visualization highlighting words commonly found in positive reviews. This word cloud provides a graphical representation of frequently occurring words, with larger words indicating higher frequency. It offers a snapshot of the prevalent themes and sentiments expressed in positive reviews, aiding in understanding customer perceptions and preferences.

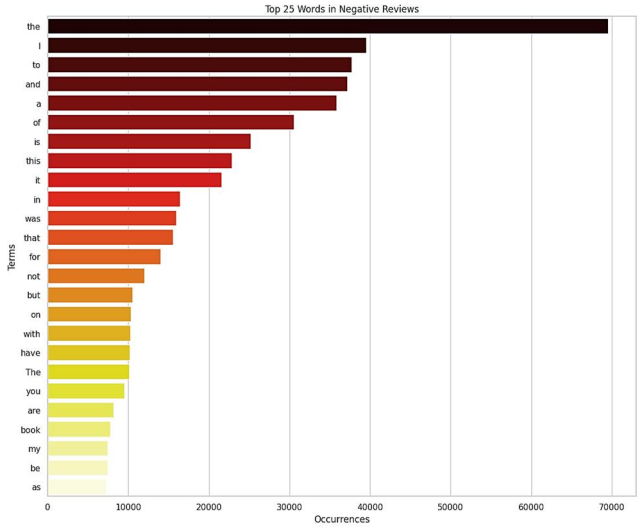


Fig. 6 Words in negative reviews

Fig 6 displays a word cloud visualization focused on words frequently occurring in negative reviews. Larger words represent higher frequency, offering a visual summary of prevalent negative sentiments within the dataset. This visualization aids in understanding common themes and issues expressed in negative reviews, facilitating sentiment analysis and insight generation.

D. Data Splitting

The "train test split" function in scikit-learn efficiently divides datasets for model training and evaluation. By default, it allocates 80% for training and 20% for testing, with options to adjust proportions. Setting "random state=42" ensures result reproducibility across iterations, crucial for scientific rigor. This method enables robust model training on ample data while facilitating unbiased evaluation. The resulting subsets, "X train," "y train," "X test," and "y test," enable comprehensive model construction and assessment. This systematic approach establishes a robust foundation for machine learning model development and evaluation, fostering progress in research and application.

E. Machine learning Modelling

When it comes to classification and regression, Support Vector Machines (SVMs) have a stellar reputation, adeptly separating classes in high-dimensional spaces using optimal hyperplanes. Random Forest, an ensemble learning technique, amalgamates decision trees to mitigate overfitting, yielding robust predictions. Decision Trees offer simplicity and interpretability, dividing feature spaces to maximize purity.

Logistic Regression, though linear, excels in binary classification, providing transparent insights into feature influence. Voting Classifier combines diverse classifiers, leveraging their collective intelligence for enhanced predictions, particularly valuable in uncertain scenarios. These models collectively offer a versatile toolkit for various machine learning tasks, ensuring accurate and reliable performance across diverse datasets and applications.

1) *Support Vector Machines (SVM):* One of the most effective supervised learning algorithms, the Support Vector Machine (SVM) is often used for regression and classification tasks. Using a high-dimensional space as its basis, support vector machines (SVMs) accurately classify data by finding the hyperplane that maximises the margin between distinct classes. SVM is known for its ability to effectively handle both linear and non-linear data patterns by utilizing different kernel functions to convert input data into higher-dimensional spaces, making it easier to separate the data. The versatility of Support Vector Machines (SVM) allows it to effectively handle complicated datasets that have different topologies. This makes SVM a significant tool in several sectors where accurate classification or regression studies are needed.

2) *Random Forest:* When training, Random Forest, an effective ensemble learning method, does a great job at generating several decision trees. The operation involves the generation of random subsets of features and the independent training of several decision trees. The last step of this method entails combining the forecasts from each individual tree to obtain the ultimate projection. This strategy greatly improves the accuracy and resilience of the model by reducing the likelihood of overfitting, which is a common problem with individual decision trees. Random Forest utilises the combined knowledge of several trees to achieve exceptional predicted accuracy in a wide range of classification and regression problems. This makes it a popular choice in machine learning applications that demand high levels of accuracy and generalization.
3) *Decision Trees:* Decision Trees are flexible and easily understandable supervised learning algorithms utilized for tasks including classification and regression. The feature space is divided into areas based on feature values, with the goal of maximizing purity or minimizing impurity within each zone. Decision Trees are renowned for their capacity to make decisions in a logical and straightforward manner, as well as their capability to handle complex interactions that are not linear in nature.
4) *Logistic Regression:* Logistic Regression, despite its misleading name, is actually a linear model that is mainly used for binary classification tasks. The primary objective of this system is to calculate the likelihood that a given input belongs to a particular category. This is accomplished by using a logistic (or sigmoid) function. Logistic Regression is well-known for its simplicity, interpretability, and computing efficiency. It is often used for classification tasks, especially when there is a linear relationship between the features and the objective. The clear interpretation of coefficients in this approach allows practitioners to accurately determine the influence of specific features on the classification outcome, hence improving the transparency and comprehension of the model. Logistic Regression is restricted to binary classification, but it can be expanded to handle multi-class classification problems by using extensions like multinomial logistic regression. Logistic Regression exhibits impressive performance in a range of practical scenarios, showcasing its resilience when the connection between features and the target is roughly linear. Logistic Regression is an essential technique in machine learning due to its high efficiency in both training and prediction. It is particularly valuable when interpretability and computational resources are of utmost importance.
5) *Voting Classifier:* One versatile ensemble learning method in scikit-learn is the Voting Classifier, which enhances the model's performance by combining predictions from many basic classifiers. To do this, we can use either a majority vote (hard voting) or an average of the predicted probabilities (soft voting) to combine the predictions of the several classifiers. Several classifiers, including SVM, Random Forest, and Naive Bayes, are utilized by the Voting Classifier in order to utilize the pooled knowledge of several models. This helps to reduce biases and enhance the overall ability to make accurate predictions. Furthermore, its adaptability permits the customization of voting algorithms and classifier weights, facilitating the use of bespoke approaches for specific applications. The Voting Classifier is very effective in classifying problems that benefit from ensemble learning, which combines the unique insights of multiple models. It is also useful in situations where there is doubt about the best method to take. By amalgamating the capabilities of several classifiers, it guarantees resilient and dependable predictions, rendering it an invaluable asset in the machine learning arsenal for attaining exceptional performance in diverse real-world scenarios.

IV. RESULT & DISCUSSION

The Results and Discussion section critically presents main findings and their implications, serving as a cornerstone for analysis and interpretation. We detail results from the empirical study of the Amazon Reviews Polarity Dataset, probing their broader significance in sentiment analysis. Through comprehensive analysis, we aim to unveil machine learning models' efficacy in sentiment categorization tasks. Evaluating performance measures like accuracy etc. provides numerical insights into model effectiveness. Subsequent discussion delves into factors influencing model performance, guiding practical applications and future research. By integrating empirical evidence and theoretical insights, this section informs decision-making and propels sentiment analysis research forward.

A. Performance Evaluation Metrics

1) Accuracy

An all-encompassing measure of the model's correctness is accuracy, which is found by dividing the total number of instances by the number of correctly categorised examples. However, in datasets that are imbalanced, meaning one class is much more prevalent than the others, accuracy might be deceptive since it may give too much importance to the performance of the majority class while disregarding the accuracy of the minority class.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2) Precision

An important metric in binary classification, precision measures how well the model predicts true positives. This metric measures how well the model avoids false positives by comparing the proportion of positively detected instances to the total number of examples designated as positive. For medical diagnostics and other situations where minimising the probability of false positives is of the utmost importance, this statistic is invaluable.

$$Precision = \frac{TP}{TP + FP}$$

3) Recall (Sensitivity)

An indicator of how well a model detects all genuine positive cases is recall, which is sometimes called sensitivity. This statistic shows how well the model can recognise positive events since it measures the proportion of positive examples that are correctly detected relative to the total number of positive cases. This statistic is crucial in situations where the absence of positive examples results in substantial expenses.

$$Recall = \frac{TP}{TP + FN}$$

4) F1 Score

The F1 score, which combines precision and recall, achieves an equilibrium between the two criteria. By calculating the harmonic mean, it provides a thorough assessment of a classifier's performance, particularly advantageous in datasets with imbalanced class distributions where both false positives and false negatives must be taken into account.

$$F1 - score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

5) Confusion Matrix

By offering a comprehensive breakdown of model predictions against genuine class labels, a confusion matrix is an essential tool for assessing the efficacy of classification models. In this square matrix, the actual classes are on the row and the anticipated classes are on the column. The number of occurrences of a true class being predicted as a given class by the model is represented by each cell in the matrix. Correct predictions are shown by the main diagonal of the matrix, while misclassifications are indicated by entries that are off-diagonal. By utilising the confusion matrix, one may calculate a number of performance metrics. These include recall, accuracy, precision, and F1 score, which allow for a thorough evaluation of the model's ability to forecast across multiple classes.

Table -2 Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 score
KNN	66	69	66	68
Random forest	81	82	82	82
Logistic Regression	85	86	84	85
Voting classifiers	83	85	82	83

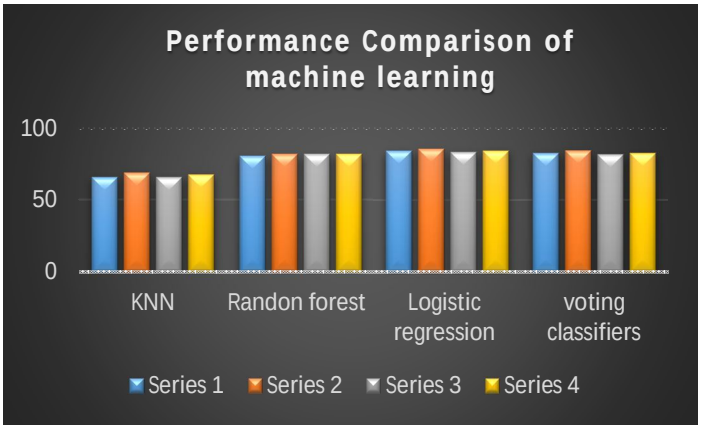


Fig. 7 Performance comparison of machine learning

Key parameters such as precision, recall, accuracy, and F1 score are used to compare the performance of various models in Table 2. We test each model using the Amazon Reviews Polarity Dataset to see how well it can categorise customer sentiment. Greater values signify enhanced efficiency. Random Forest and Logistic Regression demonstrate superior performance across all metrics, with accuracies of 81% and 85%, respectively. While KNN achieves lower scores, Voting Classifiers exhibit competitive results. These findings suggest that Logistic Regression offers the highest overall performance, balancing accuracy and precision in sentiment classification tasks.

B. Confusion Matrix of ML Techniques

In machine learning, the Confusion Matrix is a tabular representation that shows the counts of true positives, true negatives, false positives, and false negatives. It summarizes the performance of classification models. It helps evaluate the model's overall performance by providing precise insights into its capability to appropriately classify cases across multiple classes. Practitioners may evaluate the model's efficacy in classification tasks thoroughly by examining the distribution of predictions and actual class labels within the matrix. This allows them to extract important performance measures including recall, accuracy, precision, and F1 score.

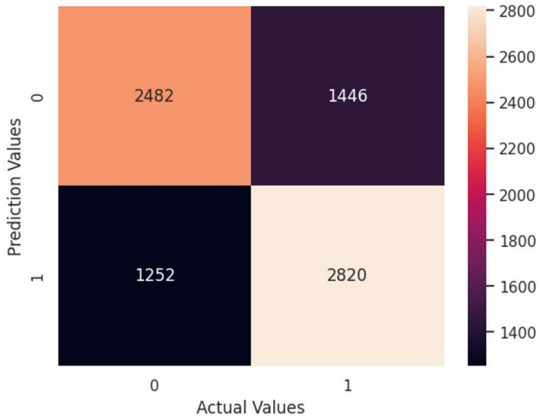


Fig. 8 confusion matrix for KNN

In Figure 8, we can see a KNN (K-Nearest Neighbours) classifier-specific confusion matrix. This matrix shows how well the KNN model performed by contrasting the predicted and real class labels. The accuracy and error patterns of the model can be seen in the matrix, where each column reflects the count of instances classified properly (true positives and true negatives) and erroneously (false positives and false negatives). You can see how well the KNN classifier does at detecting each class and where it could use some work with this visual help.

1) Random Forest Results

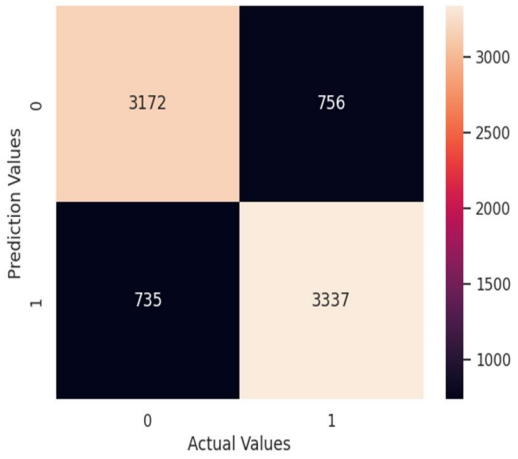


Fig. 9 confusion matrix for Random forest

A confusion matrix specific to the Random Forest (RF) classifier is shown in Figure 9. By comparing the actual and predicted class labels, this matrix provides a graphical view of the RF model's performance. You can see how accurate the model is and how often it misclassifies things by looking at the matrix, where each cell shows the number of instances categorised correctly (true positives as well as genuine negatives) and erroneously (false positives and false negatives). By examining this visualisation, one can gain insight into the RF classifier's accuracy in classifying occurrences and discover areas that should be improved.

2) Logistic Regression Results

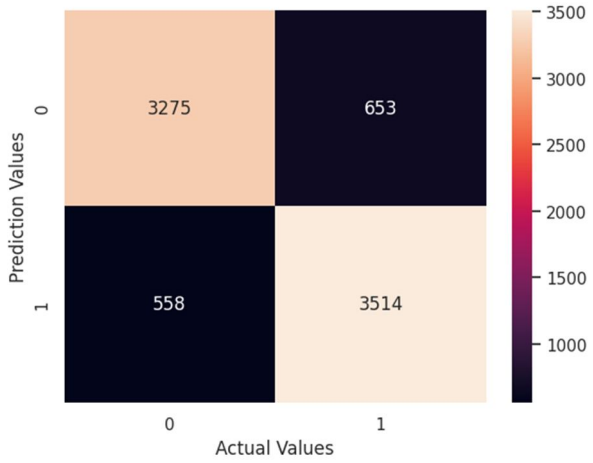


Fig. 10 confusion matrix for Logistic Regression

As shown in Figure 10, the Logistic Regression (LF) classifier's confusion matrix is unique. In this matrix, we can see how well the LF model performed by comparing the predicted and real class labels. In order to understand the model's precision and propensity for misclassification, the matrix displays the counts of occurrences categorised accurately (true positives and true negatives) and incorrectly (false positives & false negatives) in each cell. For the purpose of evaluating and improving the model, this visualisation makes it easier to evaluate the LF classifier's accuracy in classifying examples across multiple classes.

3) Voting Classifiers Results

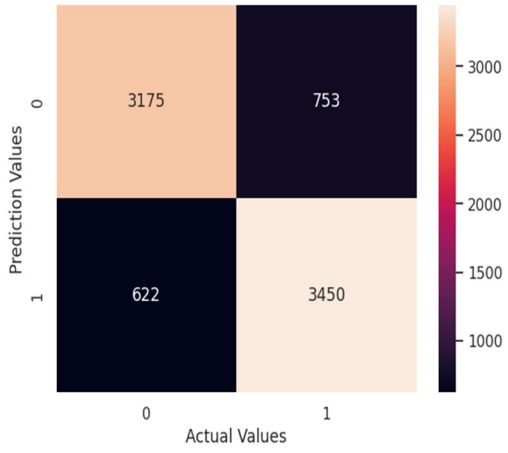


Fig. 11 Confusion matrix for voting classifiers

Figure 11 illustrates a confusion matrix for voting classifiers, showcasing the model's performance in classifying instances. It displays true positive, false positive, true negative, and false negative predictions, offering insights into the classifier's accuracy and error rates across different classes.

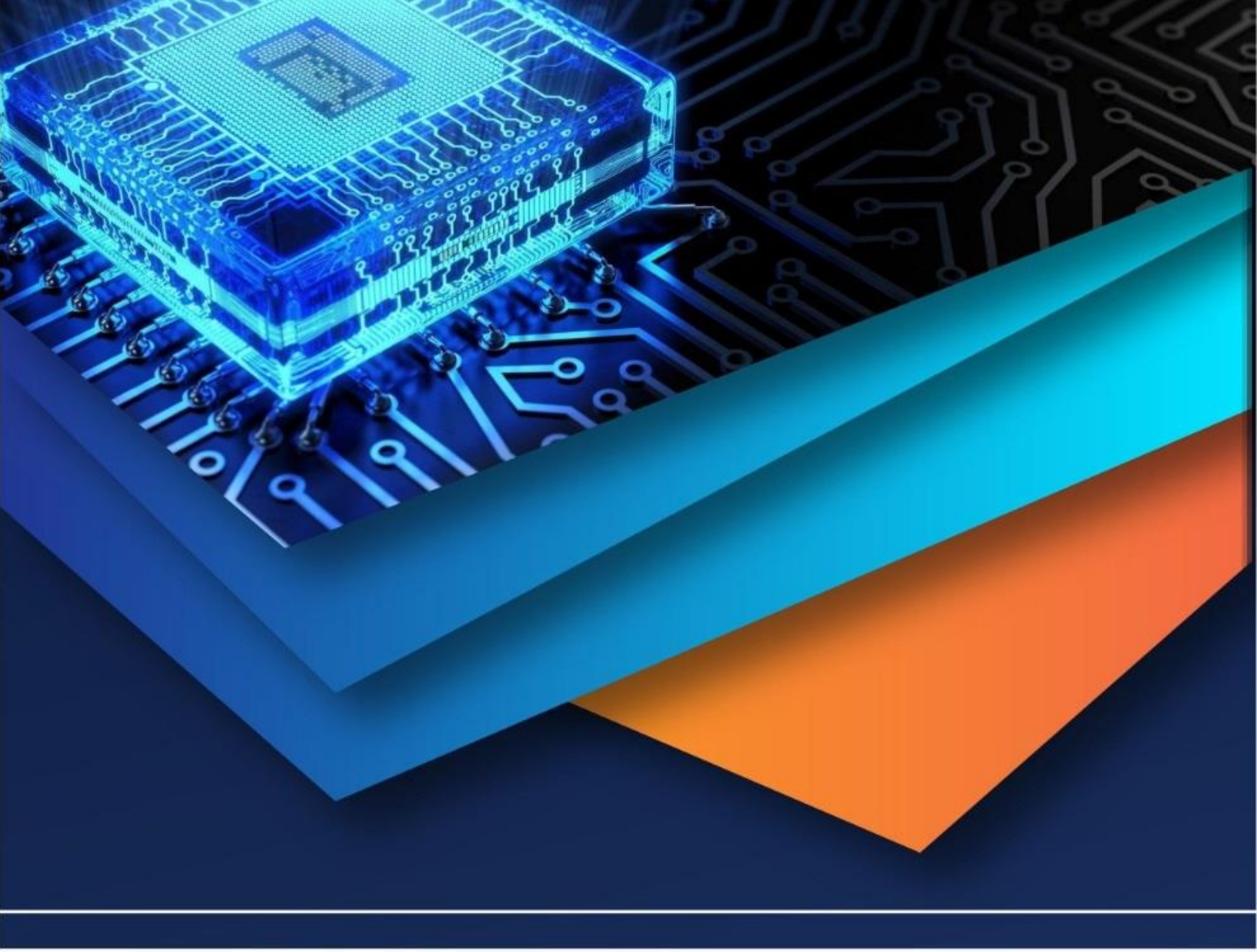
V. CONCLUSION

In this study, we used the Amazon Reviews Polarity Dataset to compare and contrast how well different machine learning models performed on sentiment analysis tasks. Utilising crucial metrics like recall, accuracy, precision, and F1 score, a thorough assessment of models like K-Nearest Neighbours (KNN), Random Forest (RF), Logistic Regression (LG), and Ensemble Classifier (ECLF) sheds light on how well these models analyse sentiment. The findings emphasise the importance of methodological rigour in machine learning research by highlighting the essential role that powerful assessment procedures and prudent model selection play in obtaining precise sentiment classification. As a result of their potential for use in real-world applications of sentiment analysis, Logistic Regression (LG) emerged as the model with the highest performance, closely followed by Ensemble Classifier (ECLF). Putting an emphasis on the effectiveness of these machine learning models, these findings offer practical guidelines for deploying solutions for sentiment analysis. To further improve the outcomes of sentiment analysis, research may investigate sophisticated feature engineering, model optimisation, and ensemble methods in the future. This will help to stimulate innovation and advancement in the field. KNN, Random Forest (RF), Logistic Regression (LG), and Ensemble Classifier (ECLF) are some of the models that are highlighted in the performance metrics table. Their usefulness in sentiment analysis is also highlighted. While RF is able to reach an accuracy of 81% and LG is able to achieve 85%, LG shows superior performance in terms of precision with 86%. With an F1 score of 85%, LG exhibits balanced performance, while RF demonstrates remarkable recall with a score of 82%! It is important to note that ECLF obtains competitive metrics across all evaluation parameters, which indicates that it possesses solid predictive skills. These findings, taken as a whole, highlight the usefulness of the dataset and the efficiency of the models in performing sentiment analysis tasks. Furthermore, they highlight the significance of model selection and evaluation in order to achieve accurate and dependable results in natural language processing applications.

REFERENCES

[1] J. Hossen, T. T. Ramanathan, and A. Al Mamun, “An Ensemble Feature Selection Approach-Based Machine Learning Classifiers for Prediction of COVID-19 Disease,” vol. 2024, 2024.
[2] J. Rufino et al., “Performance and explainability of feature selection-boosted tree-based classifiers for COVID-19 detection,” Heliyon, vol. 10, no. 1, p. e23219, 2024, doi: 10.1016/j.heliyon.2023.e23219.
[3] A. M. Asri, S. R. Ahmad, and N. M. M. Yusop, “Feature Selection using Particle Swarm Optimization for Sentiment Analysis of Drug Reviews,” Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 5, pp. 286–295, 2023, doi: 10.14569/IJACSA.2023.0140530.
[4] P. Thölke et al., “Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data,” Neuroimage, vol. 277, no. June, 2023, doi: 10.1016/j.neuroimage.2023.120253.
[5] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, “A review on sentiment analysis from social media platforms,” Expert Syst. Appl., vol. 223, no. March, 2023, doi: 10.1016/j.eswa.2023.119862.
[6] M. Alruily, “Sentiment analysis for predicting stress among workers and classification utilizing CNN: Unveiling the mechanism,” Alexandria Eng. J., vol. 81, no. July, pp. 360–370, 2023, doi: 10.1016/j.aej.2023.09.040.
[7] J. Shi, W. Li, Q. Bai, Y. Yang, and J. Jiang, “Syntax-enhanced aspect-based sentiment analysis with multi-layer attention,” Neurocomputing, vol. 557, no. July, p. 126730, 2023, doi: 10.1016/j.neucom.2023.126730.
[8] S. Šuster, T. Baldwin, and K. Verspoor, “Analysis of predictive performance and reliability of classifiers for quality assessment of medical evidence revealed important variation by medical area,” J. Clin. Epidemiol., vol. 159, pp. 58–69, 2023, doi: 10.1016/j.jclinepi.2023.04.006.
[9] M. A. S. Hazaa and S. A. A. H. Salah, “Hybrid Filter-Genetic Feature Selection Method For Arabic Sentiment Analysis,” Thamar Univ. J. Nat. Appl. Sci., vol. 8, no. 1, pp. 26–38, 2023, doi: 10.59167/tujnas.v8i1.1487.
[10] S. J and K. U, “Sentiment analysis of amazon user reviews using a hybrid approach,” Meas. Sensors, vol. 27, no. January, p. 100790, 2023, doi: 10.1016/j.measen.2023.100790.
[11] P. Balaji and D. Haritha, “An Ensemble Multi-layered Sentiment Analysis Model (EMLSA) for Classifying the Complex Datasets,” Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 3, pp. 185–190, 2023, doi: 10.14569/IJACSA.2023.0140320.
[12] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, “Sentiment analysis and opinion mining on educational data: A survey,” Nat. Lang. Process. J., vol. 2, no. October 2022, p. 100003, 2023, doi: 10.1016/j.nlp.2022.100003.
[13] A. Alsayat, “Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model,” Arab. J. Sci. Eng., vol. 47, no. 2, pp. 2499–2511, 2022, doi: 10.1007/s13369-021-06227-w.
[14] V. S. Hemachandira and R. Viswanathan, “A Framework on Performance Analysis of Mathematical Model-Based Classifiers in Detection of Epileptic Seizure from EEG Signals with Efficient Feature Selection,” J. Healthc. Eng., vol. 2022, 2022, doi: 10.1155/2022/7654666.
[15] J. Lv and S. Ge, “Industrial Land Performance Assessment Based on Fuzzy Analytic Hierarchy Process,” Math. Probl. Eng., vol. 2022, 2022, doi: 10.1155/2022/1384583.
[16] D. Loureiro, P. Beceiro, M. Moreira, C. Arranja, D. Cordeiro, and H. Alegre, “A comprehensive performance assessment system for diagnosis and decision-support to improve water and energy efficiency and its demonstration in Portuguese collective irrigation systems,” Agric. Water Manag., vol. 275, no. July 2022, 2023, doi: 10.1016/j.agwat.2022.107998.
[17] U. G. Mohammad, S. Imtiaz, M. Shakya, A. Almadhor, and F. Anwar, “An Optimized Feature Selection Method Using Ensemble Classifiers in Software Defect Prediction for Healthcare Systems,” Wirel. Commun. Mob. Comput., vol. 2022, 2022, doi: 10.1155/2022/1028175.

[18] A. Tasnim, M. Saiduzzaman, M. A. Rahman, J. Akhter, and A. S. M. M. Rahaman, "Performance Evaluation of Multiple Classifiers for Predicting Fake News," J. Comput. Commun., vol. 10, no. 09, pp. 1–21, 2022, doi: 10.4236/jcc.2022.109001.
[19] A. I. Pratiwi and Adiwijaya, "On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis," Appl. Comput. Intell. Soft Comput., vol. 2018, 2018, doi: 10.1155/2018/1407817.
[20] M. Ghosh and G. Sanyal, "Performance Assessment of Multiple Classifiers Based on Ensemble Feature Selection Scheme for Sentiment Analysis," Appl. Comput. Intell. Soft Comput., vol. 2018, 2018, doi: 10.1155/2018/8909357.
[21] Ankit and N. Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 937–946, 2018, doi: 10.1016/j.procs.2018.05.109.
[22] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," Expert Syst. Appl., vol. 77, pp. 236–246, 2017, doi: 10.1016/j.eswa.2017.02.002.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY