International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

LEVERAGING LARGE LANGUAGE MODELS FOR SENTIMENT ANALYSIS: RECENT BREAKTHROUGHS IN TEXT CLASSIFICATION

Akshay Rajshekar Shiraguppi
Irvine, California
akshayrs1993@gmail.com

Abstract

The greatest collection of Text contains human wisdom gathered over thousands of years. If this information is exploited for deeper insights, it will impart even more importance. Traditional machine learning (ML) is offered by sentiment analysis (SA). Social networking sites serve as well-liked venues for users to share a variety of emotional states through brief text messages. Short sentences illustrate a range of emotions, including fear, anxiety, and melancholy in addition to happiness. Both the positive and negative aspects of reviewers' thoughts about certain films are determined by the Sentiment Analysis algorithm on IMDb. Using big language models to evaluate sentiment categorization on IMDb movie reviews, BERT-NN is the model chosen for this investigation. Additional evaluation models, which included Random Forest (RF) and Long Short-Term Memory (LSTM) in conjunction with Support Vector Machines (SVM), were examined by the study team. Out of all the models examined, BERT NN performs the best, exhibiting 90.67% accuracy, 88% precision, 92% recall, and 88% F1-score. BERT NN is the best model for sentiment analysis tasks because of its remarkable contextual connection detecting capabilities.

Keywords: Sentiment analysis, text classification, LLM, Machine learning, NLP, Word Stemming.

I. INTRODUCTION

The digital age has brought forth a number of difficulties in analyzing textual material across various web platforms as well as circumstances for comprehending human emotions[1]. Sentiment analysis is a job in NLP that captures opinions and analyses how emotions are expressed in written language[2]. Sentiment analysis that is automated Through financial news emotions, investors may gain crucial information about market movements that have a significant impact on their decision-making as well as those of other market players.

The three primary steps in sentiment analysis are opinion evaluations to account for subjectivity, emotion detection using three different classification methods, and binary and multi-class categorization[3][4]. The sentiment of entire papers or individual lines, as well as attitudes regarding certain entities or elements, may all be analyzed at different levels[5]. The majority of text mining work is done by solving classification issues using supervised or semi-supervised learning techniques.

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

OM (SA) represents a computational method for studying public emotions and perceptions about themes, individuals, and entities, as well as social matters[6]. The terms SA and OM are frequently confused with each other but they differ in various ways. The OM field collects and analyzes entity-related opinions from people but SA focuses on analyzing the sentiment expressed through textual content before categorizing it as either neutral, negative, or positive[7]. The analysis of sentiment serves as a valuable instrument for different management fields[8], social sciences[9], and industry. TC represents a fundamental task that has increased in popularity because of recent NLP and text mining developments[10][11]. The objective of text classification systems consists of assigning predetermined labels to input text while different text classification techniques serve unique domains.

The sentiment analysis capabilities received substantial enhancement through recent developments in LLMs which included GPT-3. Processed text data classification through sentiment orientation becomes possible because LLMs develop learned representations[12][13]. Automatic systems now have new possibilities to quantify public sentiment through research advancements in sentiment analysis; financial institutions, along with political analysts, leverage their applications to achieve useful outcomes. However, prompt-based approaches with inference on large-scale language models face certain drawbacks. The number of in-context training examples is inherently limited by the maximum prompt length allowed by the architecture of the language model[14][15]. Furthermore, prompt-based approaches require online inference on resource-intensive large-scale models, making them less scalable for real-world use cases due to slow processing and significant memory overhead.

It is frequently possible to use machine learning techniques to categorize and predict if a text conveys a favorable or negative mood[16]. The two categories of machine learning algorithms are supervised and unsupervised [17]. A labeled dataset, in which every training document is labeled with the proper sentiment, is used by the supervised algorithm. Unsupervised learning, on the other hand, uses unlabeled datasets, in which the text is not appropriately sentiment-labeled.

A. Aim and contribution of the study

This study looks at how well LLMs, namely BERT NN, Analyze the sentiment of the dataset of IMDb movie reviews. The goal is to evaluate BERT NN's capacity to precisely classify sentiment and show that it is superior in comprehending the contextual subtleties of text data for sentiment analysis tasks by contrasting it with more conventional ML models like RF, SVM, and LSTM. The main key contributions are:

- Applying text categorization to movie reviews using the IMDb dataset.
- Implements comprehensive data preprocessing techniques including HTML tag removal, tokenization, stop-word removal, and stemming to get the dataset ready for model training.
- Reduces dimensionality and eliminates redundant features, improving model efficiency and performance.
- Developed Bert NN, SVM, LSTM, and RF ML and DL models to improve prediction skills.

- Evaluate the suggested BERT NN against both classic ML (RF, SVM) and DL (LSTM) models in terms of important performance measures, including recall, accuracy, precision, and F1-score.

B. Structure of the paper

The research is organized as follows: The methods and resources used are described in Section III, whereas Section II presents pertinent sentiment analysis research. Section IV presents the experimental findings of the proposed system. Section V wraps up the inquiry and presents a summary of its results.

II. LITERATURE REVIEW

This section discusses some A survey of the literature on sentiment is included in this section. The Text categorization system uses IMDB movie review data for analytical purposes.

Verma et al. (2022) In order to identify favourable or negative feelings in film reviews, the researcher suggested an approach based on sentence-level analysis. The use of Linear SVM, LR, and Multinomial NB, three supervised machine methods, in the evaluation of data from the IMDb dataset highlights the research. The hyper parameter values for the three supervised machine learning algorithms were chosen to maximize their performance. F1-Scores and Accuracy Scores, and AUC Scores were used for a comparative examination of the retrieved model results. This experimental application was achieved. An AUC-ROC score of 0.97 and an F1-Score of 0.914 were obtained during a 10-fold cross-validation procedure. In comparison to other state-of-the-art models, their model performed better in terms of F1 score and AUC-ROC score evaluation[18].

Singh, Kumar and Kumar (2022) The procedure that is being explained consists of three separate steps. The text data is pre-processed in the first stage to remove unnecessary information and make it more refined. Step two involves extracting features using the TF-IDF approach. In addition, the classifier is fed its predictions using the retrieved characteristics in the third step. The US Airlines dataset is made freely available on Twitter and used for the experiments. Analysis and classification are two areas where ML approaches see extensive use[19].

Asmawati, Saikhu and Siahaan, (2022) look at four different sentiment analysis approaches and evaluate their efficacy on text and image-based Indonesian memes. NB, SVM, DT, and CNN were the SVM algorithms used to classify the retrieved text memes after their extraction. The results showed that the NB approach, when applied to meme text for sentiment analysis, yielded the best results (65.4% accuracy)[20].

Sabba et al. (2022) addressed the problem of user sentiment analysis by offering a cutting-edge method that makes use of natural language processing and deep convolutional neural networks. The proposed method has been tested on a dataset of 50,000 movie reviews from IMDB. With an

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

accuracy of 89% in the testing phase and 89% in the training phase, the results were quite compelling[21].

Nabiha, Mutalib and Malik (2021) categorization method Using 1200 examples from a non-structured Malay text dataset, the performance of three machine learning classifiers was tested to see which one was the most accurate SA classifier. They analyzed and discussed the outcomes of DT (J48), SVM, and NB. Ten-fold cross-validation achieves a maximum accuracy of 69.7 percent, while the Percentage Split approach reaches a maximum accuracy of 70.9 percent. It demonstrates that, among various sentiment-based text classifiers, SVM outperforms the competition[22].

Below, Table I provides a summary of a literature review with dataset approaches, results and limitations for movie dataset classification.

TABLE I. SUMMARY OF BACKGROUND STUDY FOR SENTIMENT ANALYSIS USING MACHINE LEARNING

Authors	Methodology	Dataset	Key Findings	Limitation & Future Gap
Verma et al., (2022)	Three supervised machine learning algorithms—Multinomial Naïve Bayes, Logistic Regression, and Linear SVM—are used to do sentiment analysis at the sentence level.	IMDb dataset	The best 10-fold cross-validation score was 0.97 with an AUC-ROC and 0.914 with an F1-Score. Comparing the model to the state-of-the-art, performance improvements were found.	Only the IMDb database is allowed. They did not compare the models using BERT or other deep learning methods. To enhance generalization, future studies should examine deep learning models like BERT and evaluate them on other datasets.
Singh, Kumar, and Kumar (2022)	Three-step process: text data preprocessing, TF-IDF feature extraction, and classification using different machine learning criteria.	US Airlines Twitter dataset	Used multiple ML techniques for classification. TF-IDF feature extraction yielded useful features for text classification.	Limited to Twitter data from US Airlines. TF-IDF may not capture the full context of complex sentences. Future work could explore other platforms and more advanced models like LSTM or Transformers for improved context understanding.
Asmawati, Saikhu, and Siahaan (2022)	Comparison of four supervised ML algorithms (Naïve Bayes, SVM, Decision Tree, CNN) for sentiment analysis on text extracted from memes.	Indonesian memes (text & images)	Naïve Bayes achieved the best accuracy (65.4%). CNNs also performed well for image-based text sentiment analysis.	Limited to Indonesian meme data, which may not generalize well to other languages or types of text. Future research could explore multimodal sentiment analysis (text + images) and expand to other languages and meme types for better generalization.

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

Sabba et al., (2022)	NLP techniques combined with a deep CNN for sentiment analysis of film reviews. A big dataset was used to assess the proposed model.	IMDb dataset (50,000 reviews)	Attained 89% accuracy in testing and accuracy in training. Excellent results using a CNN-based approach.	Needs testing on diverse datasets and real-world applications. Overfitting on the training data could be a concern. Future work should focus on improving generalization by testing on additional datasets and improving the testing phase accuracy.
Nabiha, Mutalib, and Malik (2021)	Evaluation of Naïve Bayes, SVM, and Decision Tree (J48) classifiers for sentiment analysis on a text dataset in Malay.	Informal Malay textual dataset	SVM outperformed other models with the highest accuracy (70.9%) using the Percentage Split method. Decision Tree also showed promising results.	The dataset may not generalize well to formal or other languages because it is limited to informal Malay text. Future studies might concentrate on evaluating other content genres, enhancing classification accuracy, and multilingual datasets.

III. METHODOLOGY

The objective of this study is to use LLMs for text classification in order to examine recent advancements in sentiment analysis. To improve the accuracy and efficacy of text categorization, developments in model designs, training strategies, and methodology are investigated. The procedure begins with collecting information from sources such as IMDb movie reviews, as seen in Figure 1, then preprocessing the text to make it cleaner by deleting HTML tags, noise, special characters, stemming words, and stop words. The cleaned text is then tokenized and transformed into numerical representations using word embeddings. There are three sections to the dataset: training 70%, validation 10%, and test 20%. The proposed model, BERT NN, is fine-tuned on the training data, while comparative models, including NN, LSTM networks, RF, and SVM, are trained and evaluated. The performance of all models is assessed employing common classification metrics, such as recall, accuracy, precision, and F1-score, to determine which sentiment analysis model performs the best.

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

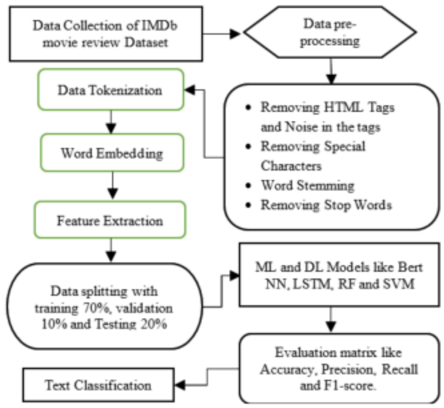

Fig. 1. Flowchart for Sentiment Analysis on text classification.

The steps of the suggested technique are described in short below:

A. Data Collection

There are 50,000 reviews of films in the IMDb movie reviews dataset. The three columns in this data set are id, sentiment, and review, and it includes 25,000 reviews that are classified as either good or negative. It prevents biases towards either attitude by guaranteeing balanced training. The equal distribution makes it easier to evaluate performance accurately and compare models in a trustworthy manner. Both attitudes' underlying patterns are captured, which enhances model generalization. Furthermore, it reduces prejudice and encourages impartial sentiment analysis findings. The overall balance of the data set improves sentiment analysis algorithms' efficacy and dependability.

Fig. 2. Word Cloud of Positive Reviews

The predominance of positive phrases in the 25,000 comments in the dataset is shown by the word cloud seen in Figure 2. It's interesting to see that the phrases "great," "really," and "good" make up the bulk of the positive-tagged data.

The frequency of negative phrases in the 25,000 comments in the dataset is depicted by the word cloud in Figure 3. Notably, terms like "bad," "even," and "though" predominate in the negative-tagged data.

International Journal of Business Quantitative Economics and Applied Management Research
Volume-7, Issue-7, 2023 ISSN No: 2349-5677

Fig. 3. Word Cloud of Negative Reviews.

The dataset consists of 50,000 reviews from the IMDB website; there are 25,000 good and 25,000 negative comments, with an equal distribution of both types of feedback. Figure 4 shows how the sentiment classifications are distributed throughout the dataset.

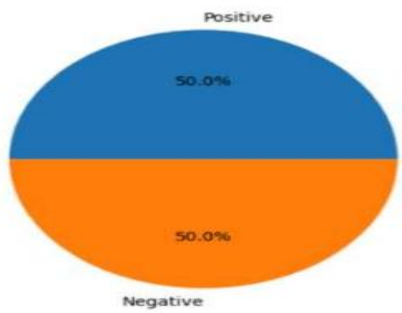

Fig. 4. Distribution of Positive and Negative Reviews

B. Data Preprocessing

The sentiment analysis can be performed, but text data must undergo extensive preparation because it is primarily unstructured. Many excellent resources exist for understanding preprocessing; however, they just briefly discuss the main phases before going on to this study's objective is to evaluate sentiment analysis algorithms. Before doing sentiment analysis, the following procedures must be taken to pre-process text data:

- Removing HTML Tags and Noise in the tags: Unnecessary elements such as HTML tags, URLs, and hashtags were eliminated to retain only meaningful text.
- Removing Special Characters: Punctuation, symbols, and numbers were removed, and to maintain uniformity, all text was changed to lowercase.
- Word Stemming: Using the NLTK library, Suffixes and prefixes were removed from words to reduce them to their basic forms, which helped standardize word variants.
- Removing Stop Words: Words that don't significantly deepen the text's meaning, such as him, they, it, and how, were excluded to enhance processing efficiency.

C. Data Tokenization

The process of tokenizing text involves dividing it into more manageable chunks, such words or sentences, to facilitate analysis. This study employed word-level tokenization to capture sentiment-related words effectively. Example:

- Original Review: "I felt this was a great way to entertain oneself when sitting in the movie theatre on a hot summer weekend. The movie was fantastic."
- Tokenized Review: "I," "thought," "this," "was," "a," "great," "way," "to," "spend," "time," "on," "a," "too," "hot," "summer," "weekend," "sitting," "in," "the," "movie," "theatre," "enjoying," "myself," "I," "truly," "liked," "the," "film".

D. Word Embedding

Generally, data consisting of strings or plain text in its native format cannot be analyzed by machine learning algorithms. Without numerical inputs, they cannot finish the job. Word embedding is an approach to emotion prediction that involves mapping lexical terms to their matching vector of integers. [23]. Two-word embedding models, glove and word2vec, have already undergone training. Use an embedding layer provided by Keras, though, and employ a vocabulary of 8,000 distinct words, each of which is embedded in a vector space of dimension 100. Rather than Their embedding layer was developed using training samples from the IMDb movie review data set, taking advantage of a pre-trained embedding word model.

E. Feature Extraction

Feature extraction is a powerful tool for reducing resource requirements while preserving critical data. The effectiveness and precision of ML models are greatly enhanced by feature extraction. Emotion counts (both positive and negative), question marks, hashtags, and exclamation points are some of the key characteristics.

F. Data Splitting

Data splitting is a stepping stone on the road to ML model performance evaluation. The fundamental idea is to use percentages of 70:10:20 to separate the dataset into separate parts for testing, validation, and training.

G. Classification of BERT NN Model

BERT is a neural network model used to analyse the emotions of the IMDb movie dataset. Classifying the reviews as either good or negative is the goal of this study. This deep learning model has already been trained and may be adjusted for a particular task.

The dataset of labeled movie reviews needs to be used to fine-tune the BERT neural network. To minimize the loss function, the fine-tuning procedure consists of simply modifying the weights of the model by the use of back-propagation and gradient descent. The difference between the expected and actual sentiments of the labeled data is known as the loss function. The Pytorch optimizer is used to carry out the fine-tuning procedure. The fine-tuning is mathematically expressed as follows in Eq. (1):

$$f(x\_i;\theta)=\mathrm{softmax}(w\_2*(\mathrm{Relu}(W\_1*h\_i+b\_2)) \quad (1)$$

The following is an explanation of the equation's terms:

- The function f (X_i; θ) outlines a neural network model that use a pre-established BERT model and is fed a sequence of tokens that each represent a movie review. The input sequence in this study is a movie review, and it derived contextualized embedding using the Bert-based-uncased model.

- The likelihood distribution of either positive or negative emotion in each class is then obtained by applying a softmax function to the result of two ReLU-enabled linear transformations that are performed on the word embeddings.
- In order to train the model parameters (θ), It minimizes a loss function between the actual movie review labels and the predicted probability distribution.
- The output of the last encoding layer of the BERT model for the preprocessed text is the h_i.
- X_i, W_1, and b_11 are the biases and weights of the first fully linked layer; the weights of the second fully linked layer and biases are denoted by w_2 and b_2 , whereas ReLU is the function for rectified linear activation. The softmax activation function is called softmax.
- Lastly, the BERT approach is designed to utilize two distinct output categories, namely, to categorize emotion that is both positive and negative.

H. Evaluation Metrics

The effectiveness of various categorization techniques was assessed utilizing measures such as F1 score, precision, accuracy, and recall. The effectiveness of ML and LLM models is evaluated using these measures. The TP, FP, TN, and FN in Figure 5 are the sources of the confusion matrix values that make up these metrics.

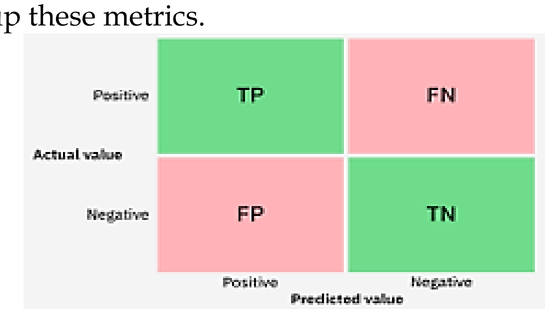

Fig. 5. Confusion matrix

The following is a definition of the performance metrics:

1. Accuracy

The ratio of properly categorized undesirable and normal events to total oil well events in the dataset is known as accuracy. It is stated as follows Eq. (2):

$$Accuracy=(TP+TN)/(TP+FN+TN+FN) \quad (2)$$

2. Precision

According to formula Eq. (3), precision is a percentage of positive observations that accurately forecast a total number of forecasts that came true.

$$Precision=TP/(TP+FP) \quad (3)$$

3. Recall

Recall is a percentage of properly detected positive observations, as computed using Eq. (4):

$$Recall=TP/(TP+FN) \quad (4)$$

4. F1-Score

The following assessment of the F1-Score using Eq. (5) combines both recall and precision values using the following mathematical formula:

$$F1\text{-}Score=(2(P*R))/(P+R) \quad (5)$$

5. ROC

The ROC curve provides a visual representation of how TPR and FPR rates fluctuate when threshold values change. Out of all recorded positive occurrences, the TPR measurement demonstrates how well the system detects current positive instances. This measure is also known as recall or sensitivity. The fraction of falsely predicted positive cases among all actual negative cases is known as the FPR.

These matrices are used to compare various ML algorithms while examining the IMDB movie review dataset.

IV. RESULT ANALYSIS AND DISCUSSION

The section demonstrates experimental outcomes for the proposed BERT NN model and sentiment analysis ML models which were compared for their performance. The experimental platform featured a system that implemented the Python program on Ubuntu OS but also incorporated a GPU along with CPU components and 1TB of RAM powered by an i5 processor to boost speed and performance. The performance measures used in the system evaluation are F1-Measure, Accuracy, Precision, and Recall. The Table II in slot two shows the assessment of BERT NN alongside different ML models applied to sentiment analysis work.

TABLE II. PROPOSED BERT NN MODEL PERFORMANCE ON THE IMDB MOVIE REVIEW DATASET FOR SENTIMENT ANALYSIS

Performance Measures	BERT NN
Accuracy	90.67
Precision	88
Recall	92
F1 Score	88

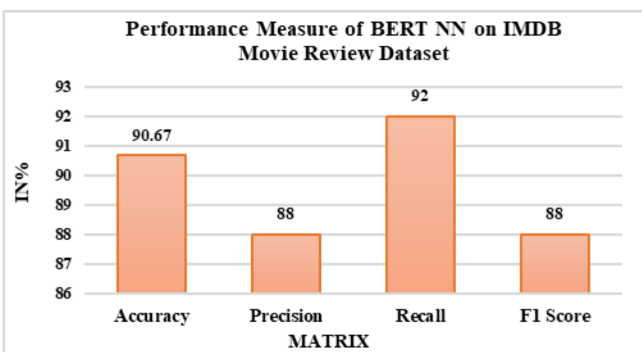

Fig. 6. Bar Graph for Model Performance

The BERT NN model demonstrated the accuracy of 90.67% in the above Table II and Figure 6 showing a high degree of overall correctness. The accuracy statistics of the model displayed an 88% precision and a 92% recall. The F1-score of 88% indicates efficient classification because it balances precision and recall performance metrics.

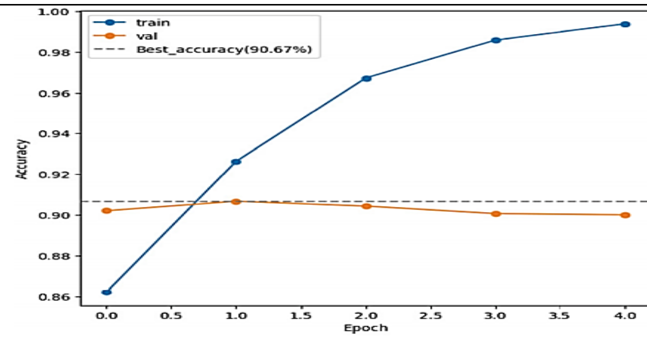

Fig. 7. Training and Validation Accuracy of BERT NN Model

The accuracy of the BERT NN model's training and validation throughout five epochs is displayed in Figure 7. The training accuracy rises from 86% to nearly 99%, while the validation accuracy remains around 90%, below the best accuracy of 90.67%. The growing gap indicates overfitting, suggesting the need for regularization techniques to improve generalization.

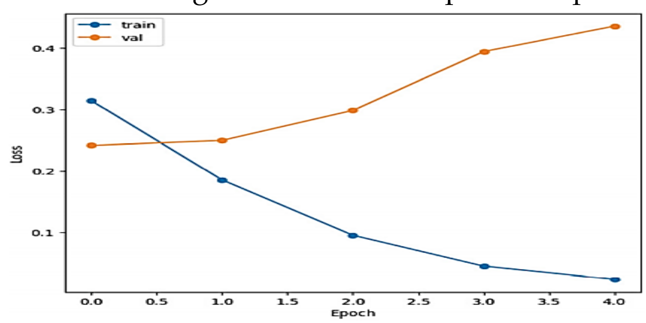

Fig. 8. Training and Validation Loss of BERT NN Model

Figure 8 displays the 5 epochs of training and validation loss for the BERT NN model. While the validation loss increases from 0.26 to 0.43, indicating overfitting, the training loss decreases from 0.32 to 0.05, showing successful learning. This emphasizes how regularization strategies like dropout or early halting are necessary to enhance generalization.

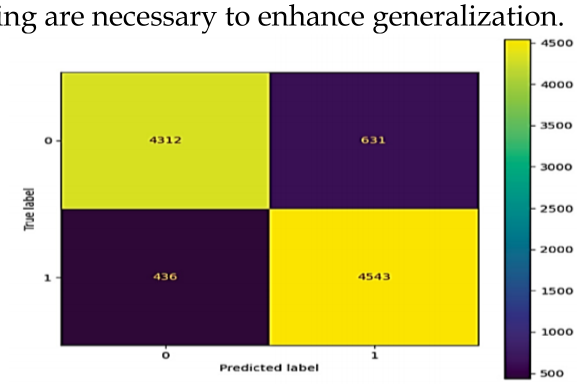

Fig. 9. Confusion Matrix of Bert NN Model on Movie Review Dataset

The BERT NN model's confusion matrix, which includes TP (4312), FP (631), FN (436), and TN (4543), is displayed in Figure 9. Whereas off-diagonal values imply misclassifications, diagonal values show occurrences that were correctly categorized. The color intensity reflects the frequency distribution, demonstrating the model's strong performance.

A. Comparative Analysis and Discussion

This section presents a comparative examination of sentiment analysis using LLM and ML on the IMDB movie review dataset. The suggested model, BERT NN, and LSTM are compared with LLM, ML, and DL models [24], RF[25], and SVM[26] as Table III shows comparable models based on performance parameters such F1 Score, Accuracy, Precision, and Recall.

TABLE III. LLM AND ML MODELS COMPARISON ON THE IMDB MOVIE REVIEW DATASET FOR SENTIMENT ANALYSIS

Models	Accuracy	Precision	Recall	F1 Score
BERT NN	90.67	88	92	88
LSTM	86.75	86.80	86.75	86.75
RF	85.62	85.97	85.39	85.68
SVM	79	75	87	80.56

Table III compares various LLM-based and ML models for evaluations of films on IMDB as a dataset for sentiment analysis. BERT NN achieves the highest accuracy (90.67%), with precision (88%), recall (92%), and an F1-score of 88, demonstrating its superior contextual understanding. F1-score, recall, accuracy, and precision for LSTM are 86.75%, 86.75%, and 86.75%, respectively, putting it in second place, demonstrating its powerful sequential modeling skills. With 85.62% accuracy, 85.97% precision, 85.39% recall, and 85.68% F1-score, RF outperforms traditional ML models. With an F1-score of 80.56%, SVM has the lowest accuracy (79%) but a reasonably high recall (87%) and lesser precision (75%). This suggests a propensity for increased false positives.

Overall, BERT NN, as the proposed model, outperforms the comparative models, including the relevance of transformer-based architectures for sentiment categorization, as demonstrated by the deep learning model (LSTM) and machine learning models (RF, SVM).

V. CONCLUSION AND FUTURE WORK

Sentiment analysis, often called opinion mining, is a method for discovering favorable, negative, and neutral views in text data. As the volume of internet data grows at an exponential rate, sentiment analysis is becoming increasingly significant. Therefore, sentiment research on social media or online reviews is used to anticipate and forecast public opinion. According to the research, the BERT NN model functions as the suggested approach since, when processing the IMDb movie review dataset, it performs better than both DL models, LSTM, and traditional ML models, such as SVM and RF. As evidenced by the performance measures, BERT NN is successful in text sentiment categorization, achieving 90.67% accuracy and 88% precision along with 92% recall and 88% F1-score. The analysis from this study underlines transformer-based models, in particular BERT NN, as they deliver peak performance in natural language processing challenges, thus establishing them as superior sentiment analysis tools.

Future research might concentrate on resolving overfitting by applying regularization strategies such as early halting or dropout, which would improve the model's capacity for generalization. Furthermore, investigates hybrid models that enhance classification performance across several domains by combining BERT NN with conventional machine learning techniques or multilingual datasets. Further investigations into transfer learning with domain-specific data and the incorporation of larger datasets could help refine sentiment analysis models for a broader range of real-world applications.

REFERENCES

1. K. M. R. Seetharaman, “Internet of Things (IoT) Applications in SAP: A Survey of Trends, Challenges, and Opportunities,” Int. J. Adv. Res. Sci. Commun. Technol., vol. 3, no. 2, 2021, doi: DOI: 10.48175/IJARSCT-6268B.
2. A. Nandi and P. Sharma, “Comparative Study of Sentiment Analysis Techniques,” in Interdisciplinary Research in Technology and Management, 2021. doi: 10.1201/9781003202240-72.
3. S. Pandya, “Predictive Analytics in Smart Grids : Leveraging Machine Learning for Renewable Energy Sources,” Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 677–683, 2021.
4. G. Alexandridis, I. Varlamis, K. Korovesis, G. Caridakis, and P. Tsantilas, “A survey on sentiment analysis and opinion mining in greek social media,” Inf., 2021, doi: 10.3390/info12080331.
5. S. Tyagi, “Analyzing Machine Learning Models for Credit Scoring with Explainable AI and Optimizing Investment Decisions,” Am. Int. J. Bus. Manag., vol. 5, no. 01, pp. 5–19, 2022.
6. S. Chatterjee, “Mitigating Supply Chain Malware Risks in Operational Technology : Challenges and Solutions for the Oil and Gas Industry,” J. Adv. Dev. Res., vol. 12, no. 2, pp. 1–12, 2021.
7. A. Bhagat, A. Sharma, and S. K. Chettri, “Machine Learning Based Sentiment Analysis for Text Messages,” IJCAT - Int. J. Comput. Technol., 2020.
8. K. Murugandi and R. Seetharaman, “Analysing the Role of Inventory and Warehouse Management in Supply Chain Agility : Insights from Retail and Manufacturing Industries,” Int. J. Curr. Eng. Technol., vol. 12, no. 6, pp. 583–590, 2022.
9. F. Torres-Cruz, S. Tyagi, M. Sathe, S. S. C. Mary, K. Joshi, and S. K. Shukla, “Evaluation of Performance of Artificial Intelligence System during Voice Recognition in Social Conversation,” in 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Dec. 2022, pp. 117–122. doi: 10.1109/IC3I56241.2022.10072741.
10. A. Gasparetto, M. Marcuzzo, A. Zangari, and A. Albarelli, “ Survey on Text Classification Algorithms: From Text to Predictions,” Inf., 2022, doi: 10.3390/info13020083.
11. Krishna Gandhi and Pankaj Verma, “ML in energy sector revolutionizing the energy sector machine learning applications for efficiency, sustainability and predictive analytics,” Int. J. Sci. Res. Arch., vol. 7, no. 1, pp. 533–541, Oct. 2022, doi: 10.30574/ijsra.2022.7.1.0226.

12. M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," Procedia Comput. Sci., vol. 87, pp. 44–49, 2016, doi: 10.1016/j.procs.2016.05.124.
13. N. Patel, "Sustainable Smart Cities: Leveraging IoT and Data Analytics for Energy Efficiency and Urban Development," J. Emerg. Technol. Innov. Res., vol. 8, no. 3, 2021.
14. K. M. Yoo, D. Park, J. Kang, S. W. Lee, and W. Park, "GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation," in Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021, 2021. doi: 10.18653/v1/2021.findings-emnlp.192.
15. S. Tyagi, T. Jindal, S. H. Krishna, S. M. Hassen, S. K. Shukla, and C. Kaur, "Comparative Analysis of Artificial Intelligence and its Powered Technologies Applications in the Finance Sector," in Proceedings of 5th International Conference on Contemporary Computing and Informatics, IC3I 2022, 2022. doi: 10.1109/IC3I56241.2022.10073077.
16. A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," in Procedia Computer Science, 2015. doi: 10.1016/j.procs.2015.07.523.
17. V. S. Thokala, "Integrating Machine Learning into Web Applications for Personalized Content Delivery using Python," Int. J. Curr. Eng. Technol., vol. 11, no. 06, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.9.
18. S. Verma, A. K. Gautam, S. Gandhi, and A. Goyal, "Improving and Analyzing the Movie Sentiments Using the SVM Approach," in 2022 IEEE Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2022, 2022. doi: 10.1109/IATMSI56455.2022.10119442.
19. S. Singh, K. Kumar, and B. Kumar, "Sentiment Analysis of Twitter Data Using TF-IDF and Machine Learning Techniques," in 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, COM-IT-CON 2022, 2022. doi: 10.1109/COM-IT-CON54601.2022.9850477.
20. E. Asmawati, A. Saikhu, and D. Siahaan, "Sentiment Analysis of Text Memes: A Comparison among Supervised Machine Learning Methods," in International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2022. doi: 10.23919/EECSI56542.2022.9946506.
21. S. Sabba, N. Chekired, H. Katab, N. Chekkai, and M. Chalbi, "Sentiment Analysis for IMDb Reviews Using Deep Learning Classifier," in 2022 7th International Conference on Image and Signal Processing and their Applications, ISPA 2022 - Proceedings, 2022. doi: 10.1109/ISPA54004.2022.9786284.
22. A. Nabiha, S. Mutalib, and A. M. A. Malik, "Sentiment Analysis for Informal Malay Text in Social Commerce," in 2021 2nd International Conference on Artificial Intelligence and Data Sciences, AiDAS 2021, 2021. doi: 10.1109/AiDAS53897.2021.9574436.
23. M. R. Haque, S. Akter Lima, and S. Z. Mishu, "Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews," in 3rd International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2019, 2019. doi: 10.1109/ICECTE48615.2019.9303573.
24. S. Alaparthi and M. Mishra, "BERT: A Sentiment Analysis Odyssey," J. Mark. Anal., vol. 9,

no. 2, pp. 118–126, Jun. 2021, doi: 10.1057/s41270-021-00109-8.
25. S. Tripathi, R. Mehrotra, V. Bansal, and S. Upadhyay, "Analyzing Sentiment using IMDb Dataset," in Proceedings - 2020 12th International Conference on Computational Intelligence and Communication Networks, CICN 2020, 2020. doi: 10.1109/CICN49253.2020.9242570.
26. N. G. Ramadhan and T. I. Ramadhan, "Analysis Sentiment Based on IMDB Aspects from Movie Reviews using SVM," Sinkron, 2022, doi: 10.33395/sinkron.v7i1.11204