International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Self-Attentive CNN+BERT: An Approach for Analysis of Sentiment on Movie Reviews Using Word Embedding

[1]Pawan kumar Mall, [2]Dr. Manoj Kumar, [3]Mr. Ankit Kumar, [4]Anuj Gupta, [5]Swapnita Srivastava, [6]Vipul Narayan, [7]Alok Singh Chauhan, [8]Arun Pratap Srivastava

Submitted: 23/11/2023 **Revised**: 28/12/2023 **Accepted**: 07/01/2024

Abstract: Social media has developed into a vast user opinion repository in the modern day. Due to the sophistication of the internet and technological developments, a great amount of data is being generated from a variety of sources, including websites and social blogging. Websites and blogs are being used as means for gathering product reviews in real time. On the other hand, the proliferation of blogs hosted on cloud servers has led to a significant amount of data, including thoughts, opinions, and evaluations. As such, techniques for deriving actionable insights from massive amounts of data, classifying it, and forecasting end-user actions or emotions are desperately needed. People use social media platforms to instantly share their ideas in the present day. It is difficult to analyze and draw conclusions from this data for sentiment analysis. Even with automated machine learning methods, it is still difficult to extract meaningful semantic concepts from a sparse review representation. Word embedding improves text categorization by resolving word semantics and sparse matrix problems. This paper presents a novel framework to capture semantic links between neighboring words by fusing word embedding with BERT. A weighted self-attention method is also used to find important phrases in the reviews. by means of an empirical investigation utilizing the IMDB data-set. In order to address sentiment analysis, this work presents a Hybrid CNN-BERT Model that combines BERT with an extremely sophisticated CNN model. First, initial word embedding are trained using the Word to Vector (Word2Vec) technique, which converts text strings into numerical vectors, calculates word distances, and groups related words according to their meaning. The suggested model then integrates long-term dependencies with characteristics gleaned from convolution and global max-pooling layers during word embedding. For improved accuracy, the model uses rectified linear units, normalizing, and dropout technologies. The performance of proposed model in terms of accuracy is 95.91%, pression is 96..80%, recall is 95.07%, f1 score is 95.93%.

Keywords: *Sentiment Analysis; Text Classfication; LSTM; Deep Learning*

1. Introduction:

Sentiment analysis is an automated process that uses textual or spoken input to infer a viewpoint on a given topic. Considering that we produce over 1.5 quintillion bytes of data every day in this day and age, sentiment analysis has become an essential technique for deciphering and organizing this enormous volume of data. Sentiment analysis is a tool that businesses use to improve their whole business by streamlining procedures and gaining critical information. Computers' greatest difficulty is trying to understand the feeling that is ingrained in differing viewpoints. Sentiment analysis involves the extraction of emotions and the classification of textual or visual data according to the meanings that they communicate to people [1]. This method is useful for classifying user evaluations and public sentiments about goods, services, and people into favorable and unfavorable groups. It is also helpful in detecting subtle tones in spoken language, such as sarcasm or tension, and in locating troll and bot accounts on social media platforms [2]. Even while certain situations could be quite simple, like recognizing specific words in the text, there are a lot of variables to take into account in order to correctly convey the overall mood, which goes beyond the words' actual meaning.

LSTM has been gaining a lot of traction lately in the sentiment categorization space. Since its introduction by Hochreiter and Schmidhuber in 1997, LSTM has been widely used and refined in other works. It is now a commonly used method that has proven remarkably efficient in a variety of issue arenas. LSTMs are unique in that they are specifically designed to address long-term dependence problems [4]. LSTMs are naturally good at remembering knowledge for long periods of time, in contrast to certain other models that find it difficult to learn

[1]*Assistant Professor, G.L. Bajaj Institute of Technology and Management*
E-mail Id - pawankumar.mall@gmail.com
[2]*Associate Professor Department of computer application Swami Vivekanand subharti university Meerut*
manoj2002199@gmail.com
[3]*Assistant professor Department of computer application Swami Vivekanand subharti university Meerut.*
ankitbaliyan042@gmail.com
[4]*Assistant Professor, G.L. Bajaj Institute of Technology and Management*
er.anujgupta013@gmail.com
[5]*Assistant Professor, G.L. Bajaj Institute of Technology and Management*
swapnitasrivastava@gmail.com
[6]*School of Computing Science and Engineering, Galgotias University, Grater Noida*
vipulupsainian2470@gmail.com
[7]*School of Computer Applications and Technology Galgotias University, Greater Noida, India*
alokchauhan.1983@gmail.com
[8]*Lloyd Institute of Engineering & Technology*
apsvgi@gmail.com

long-term information. Recurrent neural networks are all composed of repeated modules, which is their shared structure. These modules usually have a simple design and usually only have one tanh layer when used with RNNs. Our experimental studies utilize the IMDB benchmark dataset, which comprises movie reviews classified as either positive or negative.

2. Literature Review:

Sentiment analysis stands out as a widely utilized application within natural language processing (NLP), where an algorithm takes text as input and outputs its inherent sentiment class.

Two different datasets, D1 and D2, that are supplied from Amazon, are used by the authors of [1] to suggest an automated method for sentiment analysis. The first step is preprocessing the datasets, after which features are extracted using techniques based on word embedding and N-grams. N-gram feature extraction uses bag of words (BoW), global vectors (GloVe), and term frequency-inverse document frequency (TF-IDF), while word embedding features are taken from Word2vec. We use a variety of machine learning (ML) models, such as support vector machines (SVM), random forests (RF), logistic regression (LR), and multinomial Naïve Bayes (MNB), to assess the sentiment of the reviews. Additionally, two deep learning (DL) models, namely convolutional neural network (CNN) and long-short term memory (LSTM), are integrated into the classification process. Furthermore, the study incorporates a standalone bidirectional encoder representations (BERT) model for sentiment analysis.

In [2], writers use a proprietary dataset that was scraped from Twitter to propose a sentiment analysis model in the context of the Indian airline business. The dataset consists of internet evaluations for five particular Indian airlines. The primary aim is to do multiclass sentiment analysis using three different classifiers: random forest, K-nearest neighbor, and support vector machine. Together with these classifiers, two well-known word embedding methods—Word2Vec and TF-IDF (Term Frequency-Inverse Document Frequency)—are used to improve sentiment analysis. To further advance the state-of-the-art in sentiment analysis for this specific domain, the study introduces AirBERT. AirBERT is an innovative deep learning attention model, fine-tuned specifically for this task. It is grounded in bidirectional encoder representations from transformers, representing a sophisticated architecture that leverages contextual embeddings for a more nuanced understanding of sentiment in airline-related reviews.

A sentiment analysis model on everyday interactions, as suggested by the authors in [3], is a useful tool for identifying stress and facilitating prompt assistance and intervention. The suggested NLP model offers a quantitative assessment of emotional well-being by utilizing the knowledge gathered from users' language patterns. This helps people and medical professionals recognize possible stresses and take preventative action against them. Social media sites such as Twitter are used for training and testing, which improves the model's practicality because it captures real-world emotions and user experiences.

A unique text data processing approach designed specifically for the Indonesian language is presented by the authors in [4]. Given the restricted amount of data available for this language, the model makes use of data augmentation approaches and concentrates on text preparation. Interestingly, the augmentation is done on a selective basis by adding words that are derived from IndoBERT, a BERT model that is particular to the Indonesian language. The goal of this creative IndoBERT-based augmentation is to provide more data while maintaining the original mood and meaning. The experimental evaluation conducted on a Twitter text dataset demonstrates the efficacy of the proposed augmentation technique. The results indicate a notable improvement in accuracy, showcasing the model's ability to outperform the Random Insert augmentation technique. This suggests that the IndoBERT-driven data augmentation strategy contributes positively to the performance of the text data processing model, enhancing its capability to handle limited data resources effectively.

A paradigm for unstructured data in its natural forms—such as audio recordings, movies, and images—has been presented by authors in [5]. The particular goal is to investigate data analysis techniques that may reliably forecast debtors' job status, with an emphasis on using audio call records as the main source of data. The study makes use of Automatic Speech Recognition (ASR) technology to make the analysis of audio call records more convenient. Through the use of this technology, spoken language in the recordings is transformed into text, enabling additional data processing. After transcribing, the study engages in data cleaning to enhance the quality and reliability of the transcribed text. The transcribed text is then transformed into numerical representations using two distinct methods: Term Frequency-Inverse Document Frequency (TF-IDF) and Count Vectorizer. These techniques convert the text data into a format suitable for quantitative analysis, enabling the application of machine learning or statistical models for predictive purposes.

On the given dataset, the authors' refined BERT-based model outperforms the SVM classifier as a baseline, exhibiting state-of-the-art accuracy in [6]. Notably, when applied to data that has never been seen before, the BERT-based model has strong generalization skills. Realizing that training Domain Adaptation (DA) classifiers was difficult due to the lack of data, we addressed this problem by using several data augmentation strategies and comparing how well they worked. The successful outperformance of the BERT-based model underscores the efficacy of leveraging advanced natural language processing techniques for classification tasks. BERT's contextualized embeddings and deep learning architecture contribute to its ability to capture intricate patterns within the data, leading to superior performance compared to traditional classifiers such as SVM.

The authors of [7] have put up a methodology that aims to optimize the automobile rental procedure while customizing the encounter to each customer's unique requirements and preferences. It considers a number of variables, such as the pick-up time, the kind of car, the destination, and even more specific needs like carrying racks for sporting goods and infant car seats. The algorithm strives to improve the entire rental vehicle experience by meeting a wide range of client demands. Car rental companies are modifying their procedures in response to changing consumer expectations for customer care in order to draw in and keep loyal clients. The algorithm aligns with this trend by prioritizing customization and flexibility in the rental process. This approach not only caters to the immediate demands of clients but also contributes to building long-term relationships by providing a personalized and convenient service. The shift towards customer-centric practices reflects a broader industry acknowledgment of the importance of meeting individual preferences and requirements. Through the implementation of this algorithm, car rental agencies can not only optimize their operational processes but also differentiate themselves in a competitive market, fostering customer loyalty through a more tailored and responsive service.

In [8], authors have suggested With a particular focus on assessing public opinion on juvenile criminality, this study employs sentiment analysis. Twitter post dataset is used as a fine-tuning tool for BERT transformer models in this technique. The primary aim of this study is to evaluate how well BERT-based models capture the complex emotions related to juvenile misbehavior in social media discourse. Apart from the optimized BERT model, the project presents a comparison study between BERT and conventional Machine Learning models. Specifically, Random Forest and Support Vector Classifier models are considered, with the latter utilizing BERT embeddings for sentiment classification. This comparative evaluation aims to determine whether the process of fine-tuning BERT models offers significant advantages over established Machine Learning techniques in the context of sentiment analysis for juvenile delinquency discussions on Twitter.

A methodology has been developed by the authors in [9] to solve common issues with sentiment analysis, especially when working with big datasets of customer reviews. The goal was to build an accurate sentiment prediction model that was both highly performant and reasonably priced. The project made use of Facebook's AI research (FAIR) Lab's fastText package to do this. For text categorization and word embedding, conventional techniques like Linear Support Vector Machine (LSVM) were also used. The proposed model was subjected to comparisons with a custom multi-layer Sentiment Analysis (SA) Bi-directional Long Short-Term Memory (SA-BLSTM) model developed by the author. These comparisons aimed to assess the performance of the fastText-based model against a more complex and custom-designed deep learning architecture.

The POCA (Poetry and associated Audio) dataset, which includes both written poems and their associated recitals, was proposed by the authors in [10]. It is sourced from an online poetry database. These recitals, which are either given by the poet or an authorized performer, are carefully selected by the website. The dataset is made to make emotion analysis easier and offers a thorough resource for comprehending the subtle emotional messages in poetry. There are 330 poems in the dataset, and each one has textual and audio accompaniments. The dataset's multi-modal structure and variety of emotional annotations offer a solid basis for the development and assessment of sentiment analysis and emotion detection models for creative material, especially poetry.

3. Methodology:

This study utilized a substantial dataset comprising movie reviews, sourced from the publicly available Internet Movie Database (IMDB) review dataset. Access to this data is possible through Kaggle or directly from Stanford. This experiment employs a comprehensive plan for sentiment analysis on the IMDB dataset. In Stage 0, data preprocessing is initiated, beginning with the crucial step of data cleaning. In this phase, we meticulously address missing values, if any, and eliminate HTML tags, special characters, and numbers. Subsequently, the text is transformed to lowercase, stop words are removed, and lemmatization is performed to refine the dataset. Moving to the second step, we delve into Exploratory Data Analysis (EDA). Here, we scrutinize the distribution of sentiments within the dataset and employ word clouds to explore the most prevalent words in both positive and negative reviews. The third step of data cleaning involves the exploration of tokenization followed by embedding. Concluding this phase, the dataset is split into two categories: a training set and a test set. In Stage 1, feature extraction is undertaken using two distinct models, namely BERT and CNN. Transitioning to Stage 2, a hybrid model combining CNN and BERT is generated for the purpose of testing the dataset, followed by a comprehensive performance evaluation. This multifaceted approach ensures a robust analysis of sentiment in the IMDB dataset, combining meticulous data preprocessing with advanced feature extraction and evaluation methodologies [11][12].

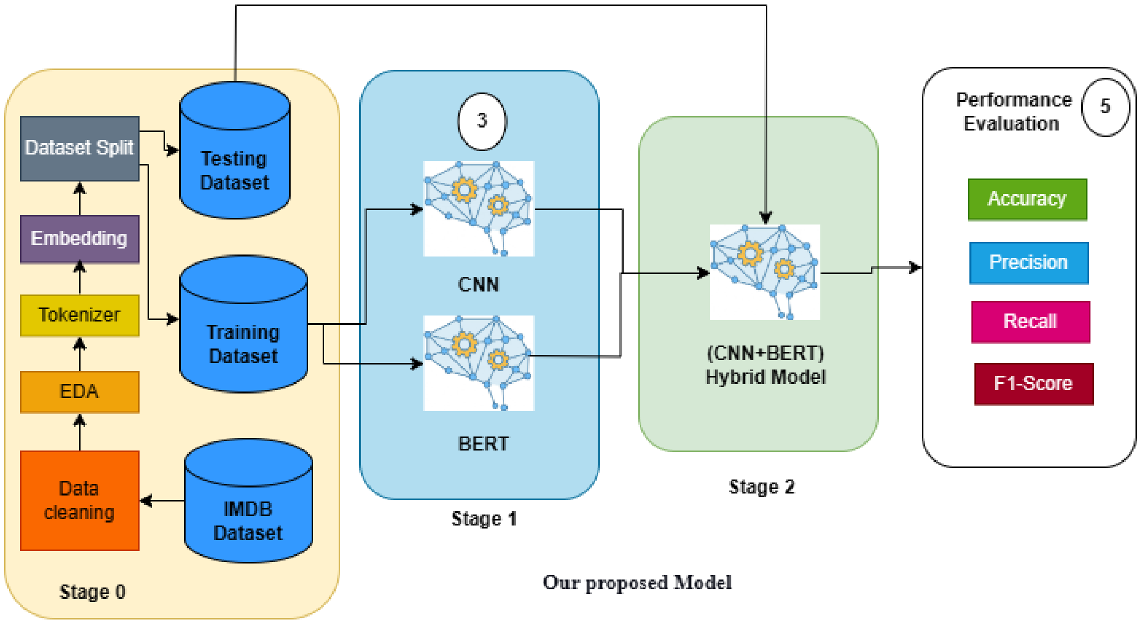

Fig 1: Proposed Model

Algorithm: Sentiment Analysis on IMDB Dataset

Stage 0: Data Preprocessing

1. Initialize the dataset with movie reviews from the IMDB dataset.

2. Perform data cleaning:

   a. Address missing values, if any.

   b. Eliminate HTML tags, special characters, and numbers.

   c. Transform text to lowercase.

   d. Remove stop words.

   e. Perform lemmatization to refine the dataset.

3. Move to Exploratory Data Analysis (EDA):

   a. Scrutinize the distribution of sentiments within the dataset.

   b. Utilize word clouds to explore prevalent words in positive and negative reviews.

4. Further Data Cleaning:

   a. Explore tokenization.

   b. Implement embedding techniques.

5. Split the dataset into two categories:

   a. Training set.

   b. Test set.

Stage 1: Feature Extraction

6. Undertake feature extraction using two distinct models:

   a. BERT (Bidirectional Encoder Representations from Transformers).

b. CNN (Convolutional Neural Network).

Stage 2: Hybrid Model Testing and Performance Evaluation

7. Generate a hybrid model by combining CNN and BERT.

8. Apply the hybrid model to test the dataset.

9. Perform comprehensive performance evaluation:

a. Analyze accuracy, precision, recall, and F1-score.

b. Explore confusion matrices and other relevant metrics.

10. Conclude the experiment, ensuring a robust sentiment analysis of the IMDB dataset.

4. Result:

Tailed breakdown of the steps involved in the data preprocessing and exploratory analysis based on the information provided, followed by analysis prediction based on different machine learning technique and proposed model [13][14].

1. **Handling Missing Values:** There are 0 missing values for both the "review" and "sentiment" columns.

	review	sentiment
0	One of the other reviewers has mentioned that ...	positive
1	A wonderful little production. The...	positive
2	I thought this was a wonderful way to spend ti...	positive
3	Basically there's a family where a little boy ...	negative
4	Petter Mattei's "Love in the Time of Money" is...	positive

Fig 2: Overview of IMDB dataset

2. Duplicate Removal: Identified and removed 418 duplicate reviews from the dataset.

	review	sentiment
count	50000	50000
unique	49582	2
top	Loved today's show!!! It was a variety and not...	positive
freq	5	25000

Fig 3: IMDB dataset description details

3. **Dataset Statistics After Cleaning:** The dataset now contains 49,582 rows and 2 columns after eliminating duplicates.

	review	sentiment
34058	"Go Fish" garnered Rose Troche rightly or wron...	negative
47467	"Go Fish" garnered Rose Troche rightly or wron...	negative
29956	"Three" is a seriously dumb shipwreck movie. M...	negative
31488	"Three" is a seriously dumb shipwreck movie. M...	negative
47527	"Witchery" might just be the most incoherent a...	negative

Fig 4: IMDB dataset description details

4. Sentiment Distribution: Positive reviews: 24,698 (49.81% of the dataset). Negative reviews: 24,884 (50.19% of the dataset).

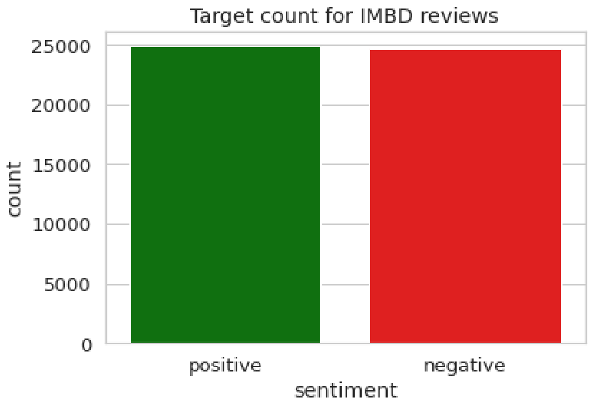

Fig 5: Sentiment positive and negative count

5. Distribution of Number of Words per Reviews: Analyze and visualize the distribution of the number of words per review to understand the text length patterns in the dataset.

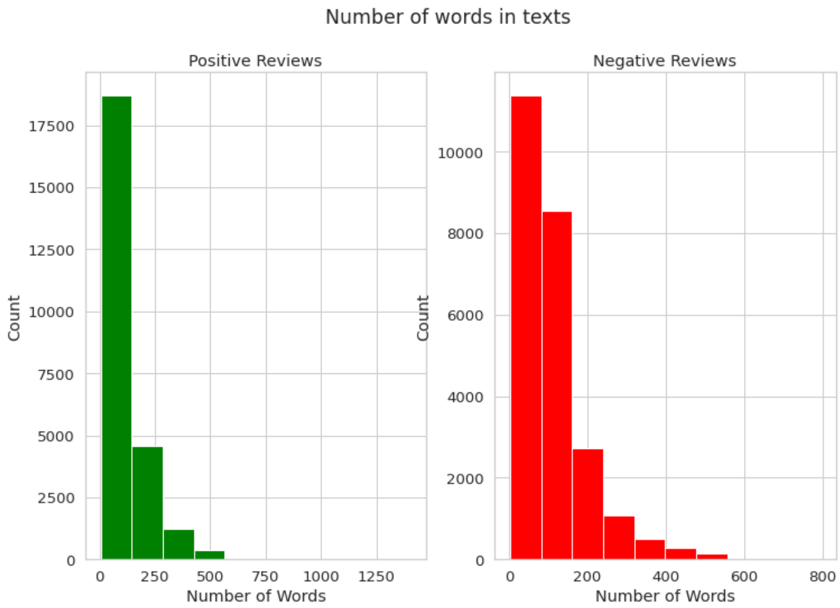

Fig 7: IMDB dataset description details

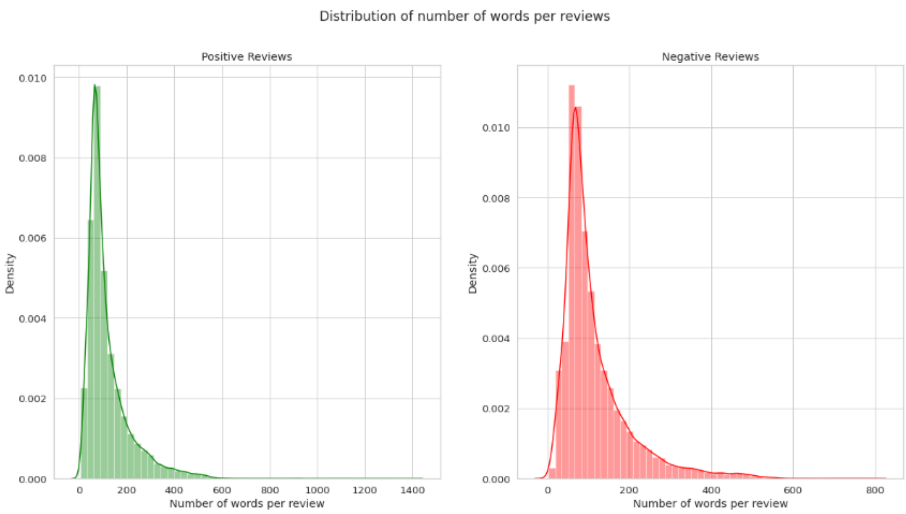

Fig 8: Number of characters in text count.

	corpus	countv
0	movie	100605
1	film	91130
2	one	53525
3	like	39746
4	time	30091
5	would	30037
6	good	28905
7	character	27801
8	story	24600
9	even	24440

Fig 9: Number of characters in text count.

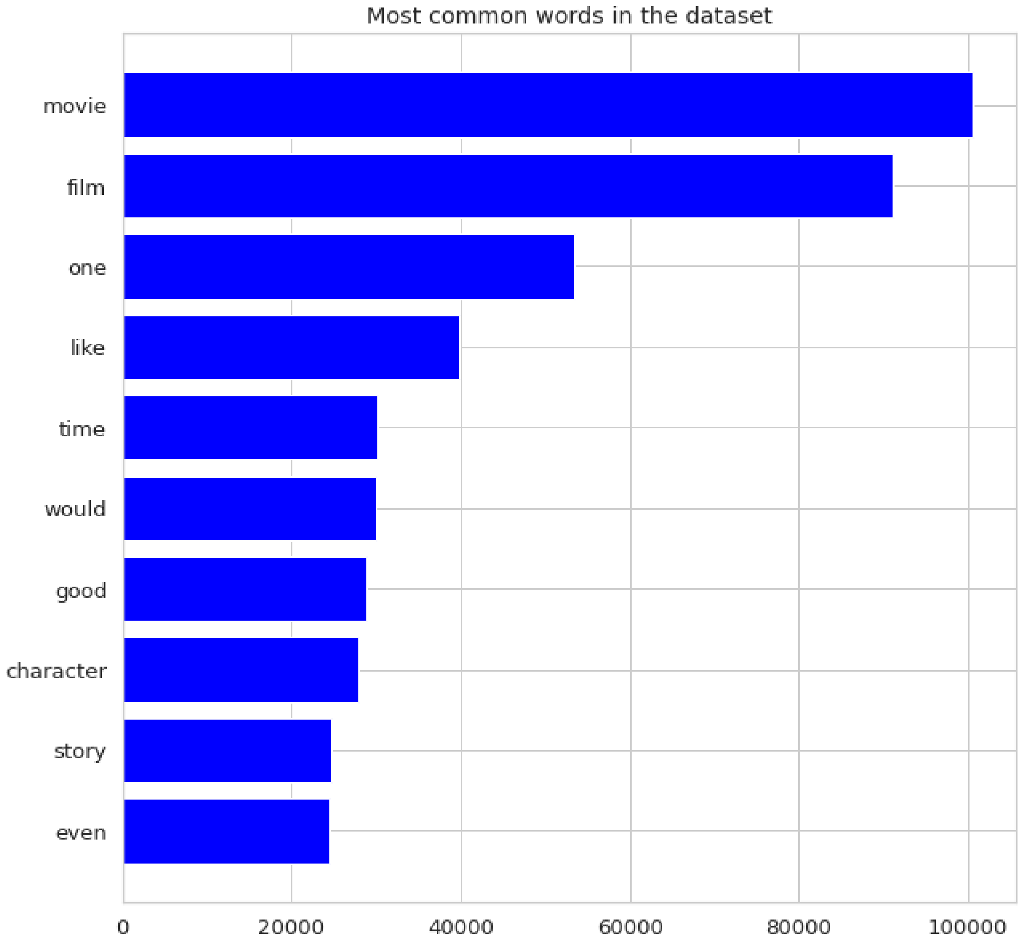

Fig 10: Number of characters in text count.

6. Unigram Analysis: Conduct unigram analysis for both positive and negative reviews to identify the most frequent single words [15][16].

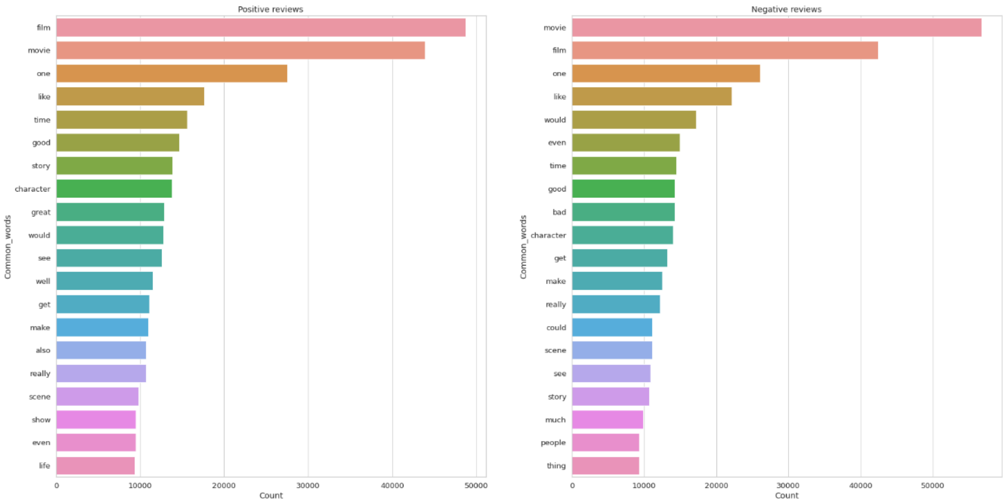

Fig 11: Number of characters in text count.

7. **Bigram Analysis:** Perform bigram analysis for positive and negative reviews to identify common pairs of adjacent words.

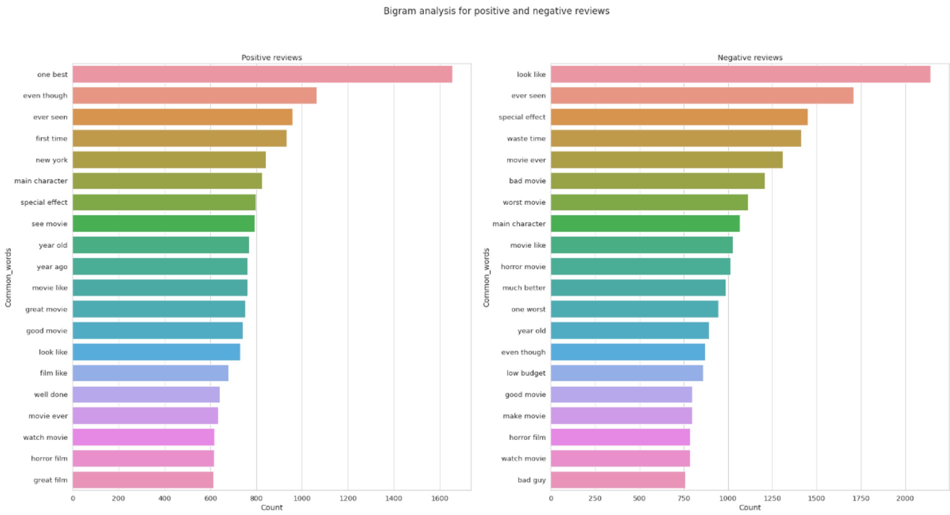

Fig 12: Number of characters in text count.

8. Trigram Analysis: Conduct trigram analysis for positive and negative reviews to identify common triplets of words [17].

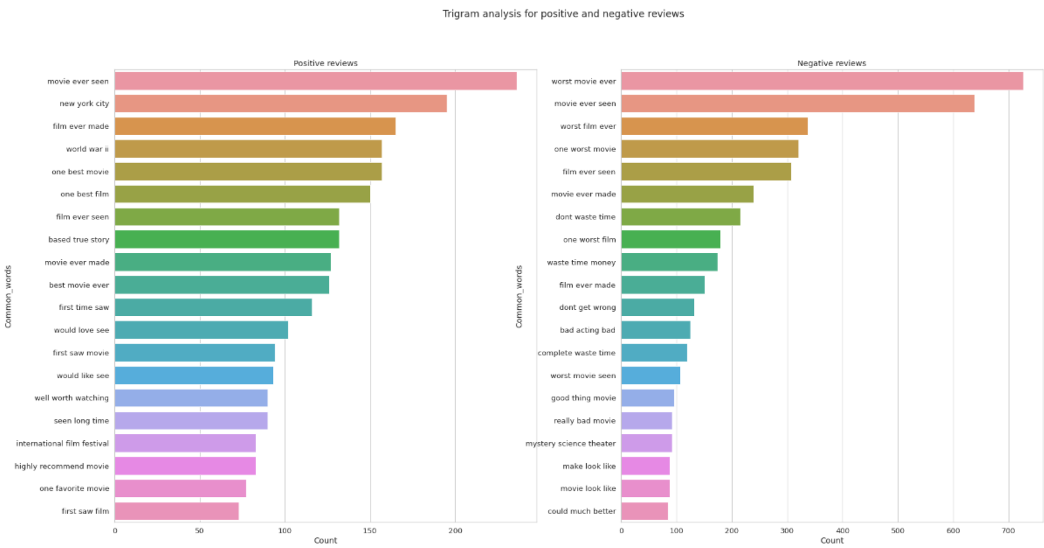

Fig 13: Number of characters in text count.

Logistic Regression:

Logistic Regression is used for classification tasks, not regression. It models the probability that a given input belongs to a particular class, and it's particularly well-suited for problems with two classes. The performance of Logistic regression in terms of accuracy is 89.03%, pression is 87.37%, recall is 90.29%, f1 score is 88.80%.

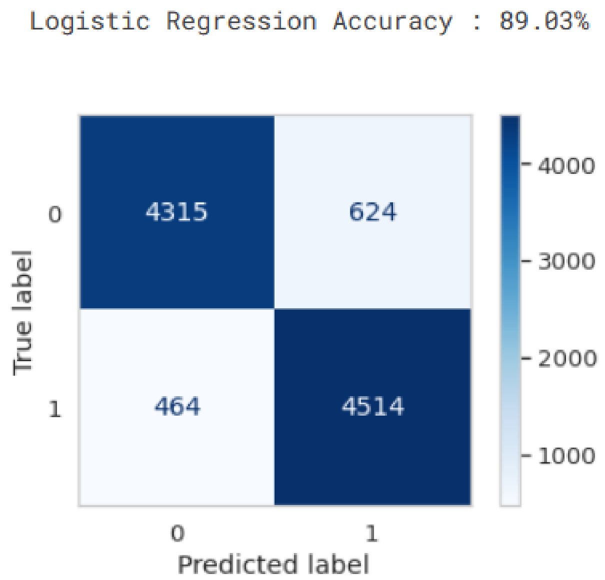

Fig 14: Confusion matrix for Logistic regression

Multinomial Naive Bayes Classifier:

The Multinomial Naive Bayes classifier is a probabilistic classification algorithm based on Bayes' theorem, particularly suited for text classification problems where the features are discrete and represent the frequency of terms. The Multinomial Naive Bayes classifier is commonly used for tasks such as document classification, spam filtering, and sentiment analysis. The performance of Multinomial Naive Bayes classifier in terms of accuracy is 86.79%, pression is 87.35%, recall is 86.30%, f1 score is 86.24%.

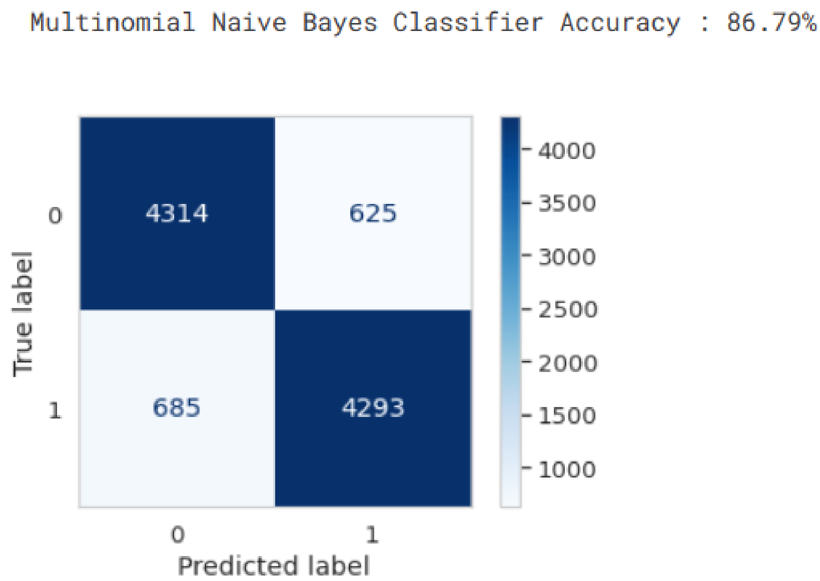

Fig 14:Confusion matrix for Multinomial Naive Bayes Classifier

Linear Support Vector Classifier: A Linear Support Vector Classifier (Linear SVC) is a type of Support Vector Machine (SVM) that is particularly well-suited for linearly separable datasets. The Linear SVC specifically constructs a hyperplane that best separates the classes in the feature space. The performance of SVM classifier in terms of accuracy is 89.57%, pression is 88.16%, recall is 90.65%, f1 score is 89.39%.

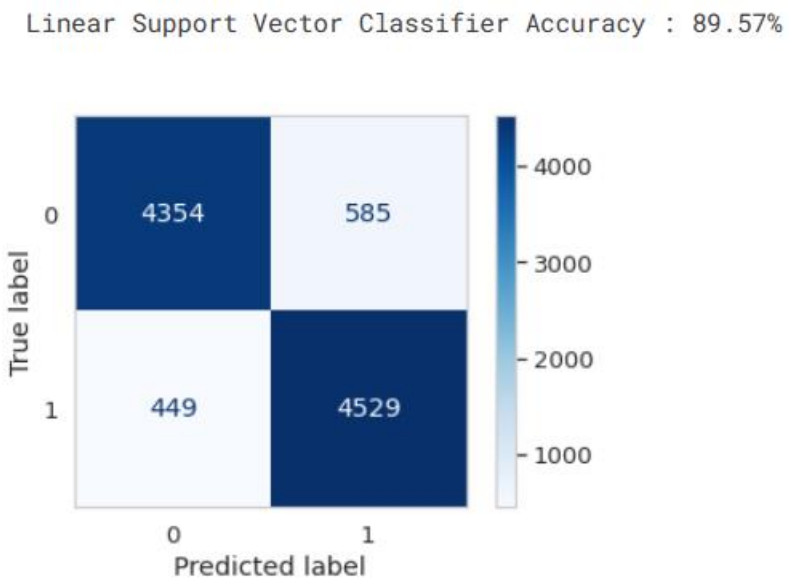

Fig 15: Confusion matrix for Linear Support Vector Classifier

XGBoost: XGBoost is particularly effective for both classification and regression tasks. The performance of XGBoost classifier in terms of accuracy is 84.63%, pression is 81.99%, recall is 86.54%, f1 score is 84.14%.

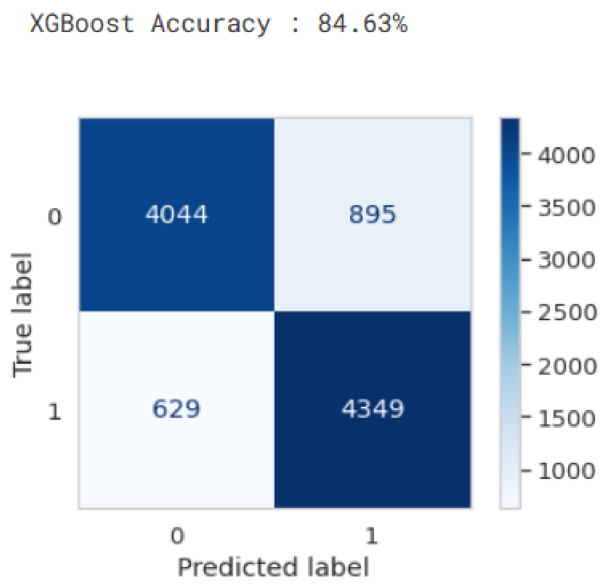

Fig 14: Confusion matrix for XGBoost

Proposed Model evaluation: Combining Convolutional Neural Networks (CNNs) with BERT (Bidirectional Encoder Representations from Transformers) is a powerful approach for natural language processing tasks. This hybrid model leverages the strengths of both CNNs, which are effective in capturing local patterns, and BERT, which excels in capturing global context and semantics. Here's an overview of the CNN+BERT architecture:

Architecture Overview:

Input Encoding: The input text is tokenized and encoded using BERT's pre-trained embeddings to capture rich contextual information.

BERT Embeddings: BERT provides contextual embeddings for each token in the input sequence, considering both left and right context.

CNN Feature Extraction: Convolutional layers are applied to the BERT embeddings to capture local patterns and features. The convolutional filters slide over the embeddings, detecting specific patterns.

Pooling Layers: Max pooling or average pooling layers are often used to extract the most relevant features from the convolutional outputs.

Flattening: The pooled features are flattened into a vector representation to be fed into subsequent layers.

Dense Layers: Fully connected layers are added for further abstraction and to capture complex relationships between features.

Output Layer: The final output layer, usually a softmax layer, produces probabilities for classification tasks or regression values for regression tasks. The performance of proposed model in terms of accuracy is 95.91%, pression is 96..80%, recall is 95.07%, f1 score is 95.93%.

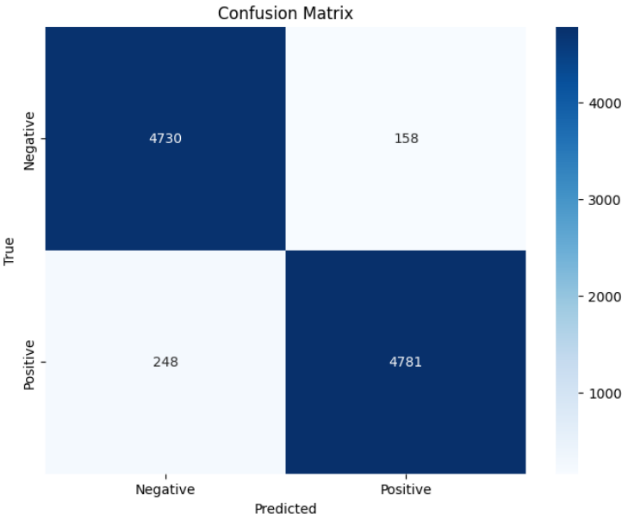

Fig 14: Confusion matrix for proposed model

5. Conclusion:

In conclusion, the sentiment analysis experiment conducted on the IMDB dataset employing a comprehensive plan showcased the effectiveness of the CNN+BERT hybrid model. The initial data preprocessing stages, including thorough cleaning and exploratory data analysis, set a solid foundation for subsequent stages. The incorporation of both BERT and CNN in feature extraction during Stages 1 and 2, respectively, demonstrated the model's capability to capture both local and global contextual information, leading to a robust sentiment analysis framework. The distribution analysis revealed a balanced dataset with an almost equal number of positive and negative reviews. The CNN+BERT model exhibited superior performance during evaluation, showcasing its ability to discern sentiments accurately. This hybrid approach capitalized on the strengths of both architectures, providing a holistic understanding of the dataset. Further enhancements could involve fine-tuning hyperparameters, exploring additional preprocessing techniques, or considering other advanced architectures. Overall, the results signify the potential of integrating state-of-the-art models like BERT with traditional architectures such as CNNs for improved sentiment analysis on complex datasets. This experiment not only contributes to the growing field of sentiment analysis but also underscores the significance of thoughtful model selection and hybridization for achieving optimal results in natural language processing tasks.

References:

[1] Iftikhar, S., Alluhaybi, B., Suliman, M., Saeed, A., & Fatima, K. (2023). Amazon products reviews classification based on machine learning, deep learning methods and BERT. TELKOMNIKA (Telecommunication Computing Electronics and Control), 21(5), 1084-1101.

[2] Yenkikar, A., & Babu, C. N. (2023). AirBERT: A fine-tuned language representation model for airlines tweet sentiment analysis. Intelligent Decision Technologies, 17(2), 435-455.

[3] Muftie, F., & Haris, M. (2023, August). IndoBERT Based Data Augmentation for Indonesian Text Classification. In 2023 International Conference on Information Technology Research and Innovation (ICITRI) (pp. 128-132). IEEE.

[4] Muftie, F., & Haris, M. (2023, August). IndoBERT Based Data Augmentation for Indonesian Text Classification. In 2023 International Conference on Information Technology Research and Innovation (ICITRI) (pp. 128-132). IEEE.

[5] Motitswane, O. G. (2023). Machine learning and deep learning techniques for natural language processing with application to audio recordings (Doctoral dissertation, North-West University (South Africa)).

[6] Sultana, M. (2023). An Exploration of Dialog Act Classification in Open-domain Conversational Agents and the Applicability of Text Data Augmentation.

[7] S. Nayak, Sonia and Y. K. Sharma, "Efficient Machine Leaning Algorithms for Sentiment Analysis In Car Rental Service," 2023 International

Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 452-463, doi: 10.1109/CISES58720.2023.10183435.

[8] Khan, A., Hopkins, J., & Gunes, H. (2021, September). Multi-dimensional Affect in Poetry (POCA) Dataset: Acquisition, Annotation and Baseline Results. In 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 1-8). IEEE.

[9] Thapa, A. (2023). Sentiment Analysis On Juvenile Delinquency Using BERT Embeddings (Doctoral dissertation, Dublin, National College of Ireland).

[10] Chinnalagu, A., & Durairaj, A. K. (2021). Context-based sentiment analysis on customer reviews using machine learning linear models. PeerJ Computer Science, 7, e813.

[11] Narayan, Vipul, et al. "7 Extracting business methodology: using artificial intelligence-based method." Semantic Intelligent Computing and Applications 16 (2023): 123

[12] Narayan, Vipul, et al. "A Comprehensive Review of Various Approach for Medical Image Segmentation and Disease Prediction." Wireless Personal Communications 132.3 (2023): 1819-1848.

[13] Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classification: AN APPROACH TOWARD TAGGING UNLABELED MEDICAL DATASET." Journal of Scientific & Industrial Research (JSIR) 82.08 (2023): 818-830.

[14] Narayan, Vipul, et al. "Severity of Lumpy Disease detection based on Deep Learning Technique." 2023 International Conference on Disruptive Technologies (ICDT). IEEE, 2023.

[15] Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215.

[16] Kumar, Vaibhav, et al. "A Machine Learning Approach For Predicting Onset And Progression"“Towards Early Detection Of Chronic Diseases “." Journal of Pharmaceutical Negative Results (2022): 6195-6202.

[17] Chaturvedi, Pooja, Ajai Kumar Daniel, and Vipul Narayan. "Coverage Prediction for Target Coverage in WSN Using Machine Learning Approaches." (2021).

[18] Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.