Comparative Study of Sentiment Analysis on Text data using various Deep Learning Models

Milind Chakraborty
Master of Computer Application (MCA)
Vellore Institute of Technology
Vellore, Tamil Nadu, India
milind.chakraborty2023@vitstudent.ac.in

Sayantan Bhattacharyya
Master of Computer Application (MCA)
Vellore Institute of Technology
Vellore, Tamil Nadu, India
sayantan.2023@ vitstudent.ac.in

Rohan Raj
Master of Computer Application (MCA)
Vellore Institute of Technology
Vellore, Tamil Nadu, India
rohan.raj2023@ vitstudent.ac.in

Prof Nagalakshmi Vallabhaneni
Assistant Professor
Vellore Institute of Technology
Vellore, Tamil Nadu, India
nagalakshmi.v@vit.ac.in

***Abstract*— This paper addresses the need to automate sentiment analysis due to the rising volume of digital textual data. Its primary goal is to analyse sentiment analysis techniques, with a focus on assessing emotional tone in sources like Amazon Product Reviews and IMDb movie reviews. We systematically evaluate the effectiveness of various Natural Language Processing (NLP) and Deep Learning models, such as LSTM, CNN, CNN-LSTM, BERT-, across diverse datasets. We also investigate how the choice of dataset influences model accuracy. Our methodology involves data preprocessing, feature extraction, model selection, and performance assessment through cross-validation. We provide a comprehensive evaluation of sentiment analysis models, considering metrics such as accuracy and F1-score. This study advances sentiment analysis techniques, assists in selecting models for specific applications and datasets, and serves as a valuable resource for future research, deepening our understanding of NLP and deep learning techniques in sentiment analysis.**

Keywords*— *Sentiment analysis, Emotional tone, Natural Language Processing (NLP), Deep Learning models, CNN, CNN-LSTM, BERT, LSTM

I. Introduction

The advent of social media applications has revolutionized the way people access information online. These platforms provide individuals with a means to express their views on various issues like social events, product reviews, and film critiques. As such, they generate a wealth of emotionally charged content that can be useful in capturing public opinion and user interests. However, with the vast amount of data available, manual processing alone cannot accomplish this task. This led to the development of text sentiment analysis technology, which has since become increasingly sophisticated.

While traditional machine learning models like Naïve Bayes and SVM were once prevalent, deep learning models have become more widespread due to their superior performance in various fields. There are several deep learning models available for sentiment analysis, each with its unique strengths and weaknesses, depending on the data and the task. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and hybrid models that combine different techniques are some of the most popular models for sentiment analysis.

In recent times, with the advent of advanced technology, text sentiment analysis has become more refined and accessible. Pre-trained models like BERT have further enhanced the effectiveness of sentiment analysis, making it more powerful than ever before. Each model has its specific strengths and weaknesses, emphasizing the need for a comparative study of their effectiveness. Our research paper aims to evaluate the accuracy and effectiveness of deep learning models in analysing social sentiments, using the well-established IMDB dataset and Amazon Product review as our benchmark.

The IMDB dataset is widely recognized in the field of sentiment analysis and is an excellent resource for evaluating the effectiveness of different models. Our study will compare the performance of various deep learning models in analysing social sentiments and provide insights into the strengths and weaknesses of each model. By doing so, our research aims to contribute to the development of better sentiment analysis models and improve our understanding of public opinion and user interests.

II. Related Works

We have studied the works in previous research and observed the advancement in the subject comparatively.

Muhammet Sinan BASARSLAN and Fatih KAYAALP [1] conducted a sentiment analysis study using seven deep learning models employing various techniques. The most promising outcome was achieved by a model comprising two Bidirectional Long Short-Term Memory (Bi-LSTM) layers, attaining an accuracy of 88.21%. Their observations revealed that both Gated Recurrent Unit (GRU) and LSTM layers contributed positively to the model's performance, even in the presence of additional Dense layers. The Adam optimizer in combination with Activation functions yielded favourable results. Surprisingly, using only Bi-GRU layers produced the best results, surpassing the expected benefits of Dropout in enhancing performance.

Lenz Baron S. Balita and et al[2] set out to develop a hybrid sentiment analysis model that combines CNN and LSTM to investigate social sentiments expressed in book reviews on Goodreads. The proposed hybrid model, utilizing Word2Vec, Part-of-Speech (POS) tagging, and SenticNet, achieved an accuracy of 89.53%.

In their study, Mahesh Mishra and Amol Patil [3] developed a hybrid model that combines CNN and LSTM on the IMDB dataset. Notably, the LSTM architecture exhibited superior performance when compared to CNN and CNN-LSTM architectures.

In their research, Ling Zhang et al[4] explore deep learning techniques for sentiment analysis, specifically by fusing Long Short-Term Memory (LSTM) networks with an adaptive boosting algorithm known as Adaboost. They introduce a novel text sentiment analysis method called LSTM-Adaboost and evaluate its performance on the IMDB movie review dataset. Their findings demonstrate that the LSTM-Adaboost model outperforms both LSTM and CNN, achieving 87.47% accuracy.

Rui Man et al [5] introduce a sentiment analysis algorithm that merges BERT and CNN, achieving the highest accuracy of 90.5% in the BERT-CNN model.

K. Amulya et al [6] conducted a comparison between Machine Learning (ML) and Deep Learning (DL) approaches using IMDB movie reviews as their dataset. Their findings revealed that DL methods outperformed ML algorithms in terms of accuracy. Among the DL algorithms, including CNN, RNN, and LSTM, RNN demonstrated the highest accuracy at 88%.

Kavita Arora et al [7] applied word2vec for word embedding and employed the BERT (Bidirectional Encoder Transformers) model for classification. Their results demonstrated a notable accuracy of 92.40%, indicating the effectiveness and reliability of their proposed approach for sentiment analysis in movie reviews.

Dr M Anusha and R Leelavathi[8] made a comparative study comparing deep learning models CNN-LSTM, CNN, and CNN-RNN on datasets including Sick SST, Twitter, FNC, Sentence Polarity, and IMDb. CNN-LSTM achieved an accuracy of 86.60% (Sick SST), 96.80% (Twitter), 97.80% (FNC), 98.60% (Sentence Polarity), 90.26% (IMDb), and 82% (English Data). Similarly, CNN achieved accuracy of 93% (sent-strength), 99.07% (sentence polarity), 94.80% (hostel data), 97.70% (IMDb), and 87% (SemEval), while CNN-RNN achieved accuracy of 89.67% (IMDb sentiment 140), 95% (KBP37), 86.60% (SST/DM/CS), and 94.60% (IMDb). The study addressed detecting sentiment polarity, pooling, ranking, and extraction.

Lirong Yao and Yazhuo Guan[9] introduce an enhanced Natural Language Processing (NLP) method founded on a Long Short-Term Memory (LSTM) structure. This novel approach involves randomly discarding parameters during backward propagation in the recursive projection layer. In their experiments, the Street Baseline achieved an accuracy of 86.51%, Classic LSTM reached 88.6%, Classic BiLSTM attained 89.3%, and the improved LSTM outperformed all, achieving the highest accuracy of 90.17% when tested on the Wall Street Journal Dataset..

K. Mouthami et al. [10] utilized BiLSTM and GRU models on the IMDB dataset, finding that BiLSTM achieved the highest accuracy score of 99.65% after 10 epochs, while GRU resulted in an accuracy of 98.24%.

Methods Proposed	Accuracy	Dataset
Bi-LSTM using 2 Bi-LSTM and 2 dropout layers[1]	88.21%	IMDB dataset
CNN-LSTM hybrid sentiment analysis model[2]	89.53% accuracy outperformed CNN (81.07%), LSTM (84.27%), and CNN-LSTM hybrid (86.07%)	Goodreads
CNN-LSTM[3]	53% (GRU), 85% (CNN), 87% (LSTM) and 85% (CNN-LSTM)	IMDB dataset
LSTM-Adaboost[4]	87.47% accuracy	IMDB dataset
BERT-CNN[5]	90.50%	Hotel review corpus (ChnSentiCorp)
RNN[6]	88%	IMDB dataset
BERT[7]	92.40%	IMDB dataset
CNN-LSTM[8]	86.60% (Sick SST), 96.80% (Twitter), 97.80% (FNC), 98.60% (Sentence Polarity), 90.26% (IMDb)	Sick SST, Twitter, FNC, Sentence Polarity, IMDb
Improved LSTM[9]	86.51% (Street Baseline), 88.6% (Classic LSTM), 89.3% (Classic BiLSTM), and 90.17% (improved LSTM)	Wall Street Journal Dataset
BiLSTM and GRU[10]	99.65% (BiLSTM), 98.24% (GRU)	IMDB dataset

III. Methodology

The datasets used encompass Amazon Reviews and IMDB datasets, the criterion considers that the review is either positive or negative. IMDB Dataset has 25000 positive and 25000 negative reviews, whereas out of the Amazon dataset, A sample of 100,000 positive and negative reviews has been considered.

A. Models Used

LSTM: LSTM, short for Long Short-Term Memory, is a variant of recurrent neural network (RNN). RNNs are neural networks designed to effectively handle sequential data, like text or audio. Nevertheless, conventional RNNs can encounter challenges when learning extended dependencies within the data.

LSTMs tackle this issue by incorporating a gating mechanism to regulate the information flow within the network. This gating mechanism empowers LSTMs to discern the critical segments of the input sequence and determine which portions can be safely disregarded.

LSTM architectures typically consist of the following components:

An input gate, which controls how much of the input sequence is allowed to flow into the cell state.

A forget gate, which controls how much of the previous cell state is forgotten.

An output gate, which controls how much of the cell state is output from the LSTM.

A cell state, which stores the current state of the LSTM.

LSTMs have been shown to be very effective for a variety of tasks, including text classification, machine translation, and speech recognition.

CNN: CNN stands for convolutional neural network, and is a type of neural network that is well-suited for processing spatial data, such as images or videos. CNNs work by extracting features from the input data using a series of convolutional layers.

Convolutional layers learn to identify specific patterns in the data, such as edges, corners, and textures. The output of each convolutional layer is a feature map, which is a representation of the input data that highlights the features that have been learned.

CNNs typically consist of the following components:

A convolutional layer, which extracts features from the input data.

A pooling layer reduces the size of the feature maps and makes the network more robust to noise.

A fully connected layer, which classifies the input data or performs some other tasks.

CNNs are very effective for a variety of tasks, including image classification, object detection, and natural language processing.

CNN-LSTM: CNN-LSTM models combine the strengths of CNNs and LSTMs to create a powerful model for processing sequential data. CNN-LSTM models typically work by first passing the input sequence through a CNN to extract features. The output of the CNN is then fed into an LSTM to learn the long-term dependencies in the data.

CNN-LSTM models have been shown to be very effective for a variety of tasks, including text classification, machine translation, and speech recognition.

BERT: BERT stands for Bidirectional Encoder Representations from Transformers, and is a type of neural network that has been trained on a massive dataset of text and code. BERT can learn the contextual meaning of words and phrases, which makes it very effective for a variety of natural language processing tasks.

BERT architectures typically consist of the following components:

An embedding layer converts each word in the input sequence to a numerical representation.

A stack of encoder layers, which learn the contextual meaning of the words and phrases in the input sequence.

A pooling layer reduces the size of the output of the encoder layers.

BERT can be used for a variety of natural language processing tasks, such as text classification, question answering, and sentiment analysis.

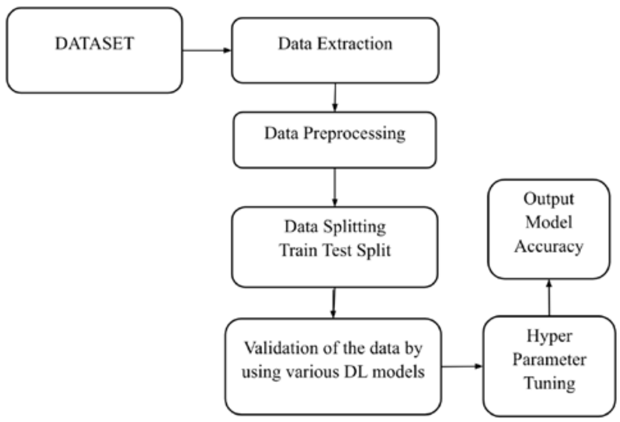
B. Dataset Description

The dataset used in this study includes two primary sources: Amazon Reviews and the IMDB dataset. Both datasets are instrumental in the field of sentiment analysis, enabling researchers to gain a deeper understanding of customer or reviewer feedback.

In the IMDB dataset, there are a total of 50,000 reviews, equally divided into 25,000 positive and 25,000 negative reviews. The feature component of this dataset encompasses the review text, which serves as the primary source of textual data. The label assigned to each review is binary, categorizing the reviews as either positive or negative based on the expressed sentiment.

On the other hand, the Amazon dataset comprises a substantial sample of 100,000 reviews. In this dataset, the labelling approach is slightly different. Reviews receiving a rating of 4 or 5 stars are labelled as "1," signifying a positive sentiment. Conversely, reviews with ratings of 1 or 2 stars are labelled as "0," indicating a negative sentiment.

C. Model Architecture



D. Data Preprocessing

With Data Preprocessing, intermediary steps are performed before applying models fitting to models, there are two main ways of performing sentiment analysis, they are done by text sentiment analysis and sentiment dictionary which are used to acquire polarity. Appropriate features are selected, from each review, data is labelled and fit to a model. Where feature vectors are extracted.[3]

Various steps were performed to make the data more legible for the model, they are as follows,

Text Preparation:

This initial phase focused on cleansing the textual data extracted from both the IMDB and Amazon Reviews datasets. During this stage, various tasks were executed:

HTML tags and non-alphabetical characters were systematically removed from the text. All textual content was uniformly converted to lowercase, ensuring consistency and ease of processing.

Stop Word Removal:

To diminish the influence of non-informative words, often referred to as "stop words," a comprehensive stop word removal process was carried out. This step significantly reduced dataset noise and enhanced the clarity of the text. We used the NLTK library for the task.

Tokenization:

The dataset was meticulously tokenized, whereby the text was dissected into individual tokens or words. This facilitated subsequent analysis and model training by breaking down the text into its fundamental components.

Encoding and Embedding:

The final step in data preprocessing involved converting the textual reviews into a numerical format. This transformation was achieved by encoding the textual content and embedding it into vector representations. This process was essential for facilitating the analysis and application of various deep learning and transformer models. The numerical representation of the text allowed for seamless integration with the chosen models, ensuring they could effectively process and learn from the textual data. Here we have used the embedding layer.

An embedding layer is a type of neural network layer that converts discrete variables, such as words or integers, into dense vectors of a fixed size. This process is known as embedding. Embeddings are useful for representing the relationships between discrete variables. For example, in natural language processing, embedding layers are used to represent the relationships between words.

Embedding layers are typically implemented as a lookup table, where each word or integer is mapped to a unique dense vector. The size of the embedding vectors is typically a hyperparameter that is tuned to optimize the performance of the neural network.

E. EXPERIMENTAL RESULTS AND ANALYSIS

We used four neural network architectures, namely Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), a combination of CNN and LSTM (CNN-LSTM), and BERT models. We assessed the performance of these models on two distinct datasets: the IMDB Movie Reviews Dataset and the Amazon Product Reviews dataset. Before model training, we meticulously pre-processed the data, employing Keras' preprocessing tokenizer to convert individual token words into vectorized representations, using embedding layers to map each word or integer to a unique dense vector.

For the CNN model, we designed a network with one convolutional layer, one max-pooling layer, a dense layer, and a dropout layer, followed by another dense layer. In the LSTM model, we employed one LSTM layer, a dropout layer, and a dense layer. In the case of the CNN-LSTM model, our architecture included one convolutional layer, one max-pooling layer, a dropout layer, an LSTM layer, another dropout layer, and a dense layer. All of these models were optimized using the Adam optimizer.

Additionally, we integrated the BERT model into our experiments, leveraging a pretrained bert-base-uncased model and a corresponding tokenizer. We tokenized the dataset and fine-tuned the pretrained model with a reduced sample of data due to the resource-intensive nature of fine-tuning. The accuracy results for each model on both the Amazon reviews and IMDB datasets are as follows:

CNN: Amazon reviews accuracy - 84.49%, IMDB accuracy - 87.5%

CNN-LSTM: Amazon reviews accuracy - 82.304%, IMDB accuracy - 87.69%

LSTM: Amazon reviews accuracy - 81.755%, IMDB accuracy - 87.25%

BERT: Amazon reviews accuracy - 92.9%, IMDB accuracy - 93.65%

These results demonstrate the effectiveness of the various models in sentiment analysis tasks, with BERT outperforming the other architectures in terms of accuracy.

Model	Amazon reviews	IMDB
CNN	84.49	87.5
CNN_LSTM	82.304	87.69
LSTM	81.755	87.25
BERT	92	93.65

IV. CONCLUSION AND FUTURE RECOMMENDATIONS

In conclusion, our experiments have demonstrated the remarkable accuracy achieved by pretrained models, notably BERT, with the added advantages of requiring minimal data and training while reducing the need for extensive hyperparameter tuning. However, there is room for further improvement in our models. For instance, we could have expanded the model complexity by incorporating additional layers and fine-tuning hyperparameters to enhance overall accuracy.

Despite our promising results, there are certain limitations to our study. One limitation is the potential for model overfitting when adding more layers and the need for carefully balancing model complexity. Additionally, we could have benefited from implementing advanced text preprocessing techniques such as Part-of-Speech (POS) tagging, stemming, and lemmatization, which have the potential to improve both accuracy and precision.

As we look to the future, it is essential to consider expanding our research to languages like Bengali and Tamil, where robust models are currently lacking. This expansion would allow us to address a broader range of languages and cultural contexts, thereby making our sentiment analysis models more versatile and inclusive.

Furthermore, the applications of our approach extend beyond sentiment analysis. We can leverage this methodology to classify various types of text data, such as reviews, and make informed decisions regarding the quality or worthiness of a product, service, or content, even in cases where a specific rating system is unavailable. This suggests exciting possibilities for the utilization of our models in diverse decision-making scenarios.

In summary, while our experiments have yielded impressive results, there is room for further refinement and expansion of our models, making our research a stepping stone for future endeavours in natural language processing and sentiment analysis.

Acknowledgment

We wish to convey our sincere gratitude to Professor Nagalakshmi Vallabhaneni, a highly esteemed Associate Professor at Vellore Institute of Technology, Vellore. She graciously provided us with the invaluable opportunity to pursue our project at VIT University. Professor Vallabhaneni's consistent guidance and unwavering support were a constant source of motivation, playing a vital role in the successful completion of our project.

Additionally, we extend our profound appreciation to VIT University for entrusting us with the opportunity to undertake this project. Their unwavering support and access to resources were instrumental in our research journey, and we genuinely appreciate the confidence they had in us.

References

[1] Basarslan, Muhammet Sinan, and Fatih Kayaalp. "Sentiment Analysis with Various Deep Learning Models on Movie Reviews." In 2022 International Conference on Artificial Intelligence of Things (ICAIoT), pp. 1-5. IEEE, 2022.

[2] Balita, Lenz Baron S., Kyle Matthew A. Degrano, Andrei Daniel A. Pamoso, and Joel C. De Goma. "Sentiment Analysis on Book Reviews Using Convolutional Neural Network (CNN) Long Short-Term Memory (LSTM) Hybrid." In Proceedings of the 2022 6th International Conference on E-Business and Internet, pp. 125-132. 2022.

[3] Mishra, Mahesh, and Amol Patil. "Sentiment Prediction of IMDb Movie Reviews Using CNN-LSTM Approach." In 2023 International Conference on Control, Communication and Computing (ICCC), pp. 1-6. IEEE, 2023.

[4] Zhang, Ling, Miao Wang, Ming Liu, and Haozhan Li. "Sentiment Analysis of Movie Reviews Based on LSTM-Adaboost." In 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), vol. 5, pp. 186-190. IEEE, 2022.

[5] Man, Rui, and Ke Lin. "Sentiment analysis algorithm based on bert and convolutional neural network." In 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), pp. 769-772. IEEE, 2021.

[6] Amulya, K., S. B. Swathi, P. Kamakshi, and Y. Bhavani. "Sentiment analysis on IMDB movie reviews using machine learning and deep learning algorithms." In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 814-819. IEEE, 2022.

[7] Arora, Kavita, Neha Gupta, and Sonal Pathak. "Sentimental Analysis on IMDb Movies Review using BERT." In 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 866-871. IEEE, 2023.

[8] Anusha, M., and R. Leelavathi. "Analysis on sentiment analytics using deep learning techniques." In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 542-547. IEEE, 2021.

[9] Yao, Lirong, and Yazhuo Guan. "An improved LSTM structure for natural language processing." In 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), pp. 565-569. IEEE, 2018.

[10] Mouthami, K., N. Yuvaraj, K. K. Thilaheswaran, and K. J. Lokeshvar. "Text Sentiment Analysis of Film Reviews Using Bi-LSTM and GRU." In 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1379-1386. IEEE, 2023.