|  |
| --- |
|  |



**Product Sentiment Comparison in E-Commerce (Amazon)**

Contents

[Abstract ii](#_Toc153146427)

[1. Introduction 1](#_Toc153146428)

[2. Related Works 2](#_Toc153146429)

[3. Methodology 3](#_Toc153146430)

[3.1 Dataset 3](#_Toc153146431)

[3.2 Exploring the Data 3](#_Toc153146432)

[3.3 Rating Analysis 4](#_Toc153146433)

[3.4 Wordcloud 6](#_Toc153146434)

[3.5 Preprocessing for the RNN model 7](#_Toc153146435)

[3.5.1 Word Embedding 7](#_Toc153146436)

[3.5.2 Data Sampling 7](#_Toc153146437)

[3.6 Preprocessing for BERT model 8](#_Toc153146438)

[3.7. LSTM Model Development 8](#_Toc153146439)

[3.7.1 LSTM Model Training 9](#_Toc153146440)

[3.8 Tuned LSTM Model Development 10](#_Toc153146441)

[3.8.1 Tuned Model Training 11](#_Toc153146442)

[3.9 Bert Model 11](#_Toc153146443)

[4. Result 13](#_Toc153146444)

[4.1 LSTM Model Evaluation 13](#_Toc153146445)

[4.2 Tuned LSTM Model Evaluation 14](#_Toc153146446)

[4.3 BERT Model Evaluation 15](#_Toc153146447)

[4.4 Models Comparison 16](#_Toc153146448)

[5. Discussion 17](#_Toc153146449)

[6. Conclusion 18](#_Toc153146450)

[References 19](#_Toc153146451)

# Abstract

In the ever expanding world of e-commerce, understanding customer sentiments towards products is important for both consumers and businesses. This study focuses on the analysis and classification of product sentiments within the e-commerce giant Amazon, with the aim of categorizing customer reviews into either positive or negative sentiments. By using natural language processing (NLP) techniques a dataset of Amazon product reviews is analyzed to extract valuable insights into customer satisfaction and dissatisfaction. The primary objectives of this research include collecting, preprocessing, and structuring the ecommerce review data, then performing LSTM (Long Short Term Memory) and a pre-trained model, BERT (Bidirectional Encoder Representations from Transformers), to classify the reviews, with the aim to predict the overall sentiment of customers towards various products. This study explores the impact of sentiment classification in improving customer experiences, aiding purchasing decisions, and helping businesses better understand consumer preferences.

The analysis of the dataset revealed a predominant distribution of 5-star ratings, followed by 4-star ratings, with 2-star ratings being the least frequent. Text preprocessing, including cleaning, stop word removal, and lemmatization, was performed to enhance the quality of the review text for subsequent analysis. This study explores sentiment polarity and its correlation with ratings, revealing a positive relationship between mean sentiment and higher ratings. The subsequent word analysis showcased the most common words associated with different ratings, providing valuable insights into customer sentiments. Visual representations, such as word clouds, offered a comprehensive overview of frequently used words in positive, negative, and overall reviews.

Two models were used for the sentiment prediction, a Long Short-Term Memory (LSTM) model was constructed, incorporating layers like Embedding, Bidirectional LSTM, GlobalMaxPool1D, BatchNormalization, Dropout, and Dense layers. The model, compiled with the RMSprop optimizer, achieved an accuracy of 87.94%. Further tuning of LSTM model units, dropout rates, and the optimizer resulted in an enhanced accuracy of 89.76% and the BERT model, a powerful transformer-based architecture fine-tuning for text classification yielded remarkable accuracy at 93.91%, surpassing both the baseline and tuned LSTM models. The BERT model's proficiency in understanding contextual meaning and relationships between words showcased its superiority in sentiment analysis. Comparing the models, BERT demonstrated the highest accuracy at 93.91%, followed by the tuned LSTM at 89.76%, and the baseline LSTM at 87.94%. The detailed evaluation metrics, including confusion matrices and classification reports, provided a nuanced understanding of each model's strengths and areas for improvement. The BERT model's superior performance underscores its potential as a robust solution for sentiment analysis tasks in the domain of e-commerce reviews.

# 1. Introduction

Sentiment analysis also known as opinion mining or emotion AI, is the process of computationally identifying and categorizing opinions expressed in a piece of text, Tanishka, et al., 2022. The goal of sentiment analysis is to classify the text based on the mood or mentality expressed in the text, which can be positive, negative, or neutral and sometimes further categorize it into specific sentiments like happiness, anger, sadness, etc. Understanding consumer sentiment and feedback has become essential to corporate success in the continually changing marketplace of today, Harish, et al., 2022. Unprecedented opportunities for businesses to learn about how customers perceive their goods and services have been created by the explosion in digital communication and online reviews. The field of Natural Language Processing (NLP) has enabled the extraction of valuable information from text data, paving way for advanced sentiment analysis methods, Zie, et al., 2023.

NLP is a subfield of artificial intelligence (AI) and linguistics that focuses on the interaction between computers and human language. It deals with the development of algorithms and models that enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful NAOMI, 2023. The main objectives of this research are to analyze, evaluate and interpret the results. To accomplish the sentiment classification task this research employs Long Short-Term Memory (LSTM) units. There are different approaches to product sentiment analysis, one common approach is to use natural language processing (NLP) techniques to extract features from the text of the reviews, such as the number of positive and negative words, the use of certain keywords and the overall tone of the review.

Transfer learning is a machine learning technique that involves training a model on one dataset and then using that model to make predictions on another dataset Daniel, et al., 2023. In this report, pre-trained model/transfer learning will be used to develop a model for product sentiment comparison in e-commerce. Specifically, a pre-trained NLP model called BERT and LSTM will be used to extract features from the text of Amazon product reviews. These features will then be used to train a machine learning model to classify the reviews as positive or negative. The performance of the model will be evaluated on the test set of Amazon product reviews. Comparison will be done to compare the performance of our model to the performance of a baseline model that uses a simple NLP feature extraction technique.

# 2. Related Works

In recent years, the rise of digital platforms and online communication has reshaped the dynamics of consumer behavior and market competitiveness Franco, et al., 2021. Consumers now have the power to voice their opinions, experiences, and preferences on a global scale through product reviews, social media, and online forums Asha, et al., 2022. This wealth of user-generated content presents a treasure trove of insights that businesses can tap into to refine their strategies and offerings Marcelo, et al., 2020. Marouane, et al., 2021 in their paper said Sentiment analysis has become a critical tool for comprehending client sentiments and opinions in this field. Opinion mining and sentiment analysis both use Natural Language Processing (NLP) methods to ascertain the sentiment contained in textual data Tanishka, et al., 2022. Businesses can acquire a better grasp of client feedback and make data-driven decisions by automatically determining if a text message reflects a good, negative, or neutral attitude. Yifan, et al., 2021 then proposes a self-updating iterative algorithm for sentiment analysis of commodity evaluation text in online shopping. The algorithm aims to address the mismatch between commodity evaluation and scoring. The experiment conducted in the paper demonstrates that the algorithm is simple and efficient, with an accuracy of commodity evaluation reaching more than 99.17%.

Although individual sentiment analysis has been beneficial, the idea of comparative sentiment analysis adds a fresh perspective Arnd, et al., 2020. The capacity to determine which products are more favorably received by customers can be a game-changer in competitive markets where numerous products compete for consumer attention Arnd, et al., 2020. Norhaslinda, et al., 2021 with comparative sentiment analysis seeks to answer issues like Which smartphone brand has received more positive reviews? Which streaming platform has a higher proportion of client satisfaction? Businesses may get a more complete picture of their position in the market by examining sentiment toward items that are competing with theirs. Krishna, et al., 2021 highlight the use of machine learning and Natural Language Processing (NLP) methods for examining opinions expressed on social networking platforms, where they used deep learning methods, specifically (RNN) and (CNN), was used to improve sentiment analysis results. The distinctive advantage of these LSTM models, as emphasized in the paper, lies in their superior ability to handle sentiment impact compared to other types of Recurrent Neural Networks (RNNs). The LSTM model used in the paper achieved an accuracy of 93.66% in predicting customer review opinions.

Philippe, et al., 2023. proposes four deep learning models that combine BERT with Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Gated Recurrent Unit (BiGRU) algorithms to enhance sentiment analysis accuracy. These models are trained on three datasets and compared to pre-trained BERT models and other classical machine learning models. The architectures with BiGRU layers show the best results. The paper also examines the impact of changing the number and location of BiLSTM and BiGRU layers to enhance performance in short text classification. It proposes a hybrid BERT-based multi-feature fusion model for emotion classification. The study compares the accuracies of different pre-trained BERT classifier models, such as RoBERTa (BERTBase) and DistilBERT (BERTMini), and identifies the best-performing models for different datasets. For example, the RoBERTa model "3G" achieves 91.72% accuracy, and the DistilBERT model "GLG" achieves 88.04% accuracy.

In conclusion, comparative sentiment analysis represents an evolution of sentiment analysis that aligns well with the current demands of businesses operating in competitive markets. This project seeks to contribute to the growing body of research and application in the field of sentiment.

# 3. Methodology

## 3.1 Dataset

The data was gotten from Kaggle.com <https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products>. The data contains over 67,992 consumer reviews for 68 different Amazon products. The dataset includes basic product information; id, name, asins, brand, categories, keys, manufacturer, reviews.date, reviews.dateAdded, reviews.dateSeen, reviews.didPurchase, reviews.doRecommend, reviews.id, reviews.numHelpful, reviews.rating, reviews.sourceURLs, reviews.text, reviews.title, reviews.userCity, reviews.userProvince, reviews.username. The data preparation process includes selecting specific columns from the initial dataset, namely 'reviews.rating' and 'reviews.text', to focus on review ratings and review text content then analyzing the distribution of review ratings in the selected data, Showed a significant class imbalance, with a majority of ratings being 5.0. The class imbalance was addressed by adding more data from two additional datasets and selecting just rows that contain reviews with ratings of 3 or lower. After which the three datasets are concatenated (data1, data2, and data3) to create a unified dataset with a broader range of review ratings.

## 3.2 Exploring the Data

Figure 1. shows the distribution of the review rating scores and it show that there are significant number of 5 rating, followed by 4 rating, the rating with the least number is 2 rating

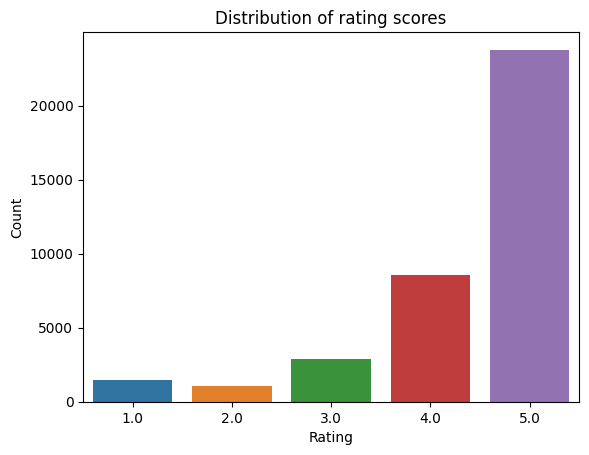


Figure 1. Distribution of rating scores after combining the datasets

Figure 2. compares the lengths distribution of positive and negative review text

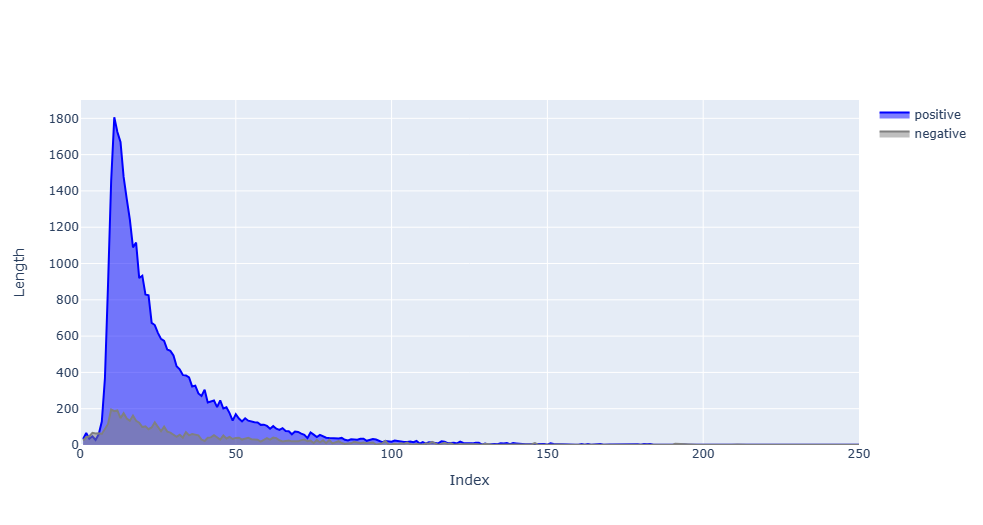


Figure 2. Length of positive and negative review text

Text preprocessing is performed as part of natural language processing (NLP) to clean and structure raw text data, removing noise, irrelevant information, and ensuring uniformity, which facilitates more effective analysis and modeling. The following preprocessing are done:

Clean Text - A function that takes a text as input and performs several text cleaning operations. It converts the text to lowercase, removes leading and trailing spaces, removes numbers, HTML tags, and special characters, and replaces multiple consecutive spaces with a single space.

Remove Stopwords - A function that removes common English stopwords (e.g., "the," "is," "in") from a text. It tokenizes(process of breaking down a text into individual words or tokens for analysis and understanding) the text into words, filters out stopwords, This helps reduce noise in the text data.

Lemmatization - It uses the WordNetLemmatizer from NLTK to reduce words to their base or dictionary form. Lemmatization takes into account the context of the word in the sentence.

After defining the text preprocessing functions, the functions are applied to a the dataframe to process the 'reviews.text' column. It sequentially cleans the text, removes stopwords, and then lemmatizes the text.

## 3.3 Rating Analysis

Table 1: Most Common Words for Rating 1:

|  |  |
| --- | --- |
| Word | Count |
| Battery | 1115 |
| Buy | 509 |
| Amazon | 478 |
| Use | 457 |
| Last | 401 |

Table 2: Most Common Words for Rating 2:

|  |  |
| --- | --- |
| Word | Count |
| Battery | 527 |
| Use | 394 |
| Tablet | 297 |
| Amazon | 293 |
| Buy | 289 |

Table 3: Most Common Words for Rating 3:

|  |  |
| --- | --- |
| Word | Count |
| Tablet | 1105 |
| Use | 756 |
| Get | 668 |
| Amazon | 576 |
| Good | 842 |

Table 4: Most Common Words for Rating 4:

|  |  |
| --- | --- |
| Word | Count |
| Tablet | 3142 |
| Use | 2748 |
| Get | 1660 |
| Good | 2154 |
| Great | 2906 |

Table 5: Most Common Words for Rating 5:

|  |  |
| --- | --- |
| Word | Count |
| Use | 7038 |
| Tablet | 5938 |
| Easy | 5066 |
| Love | 8165 |
| Great | 8643 |

## 3.4 Wordcloud

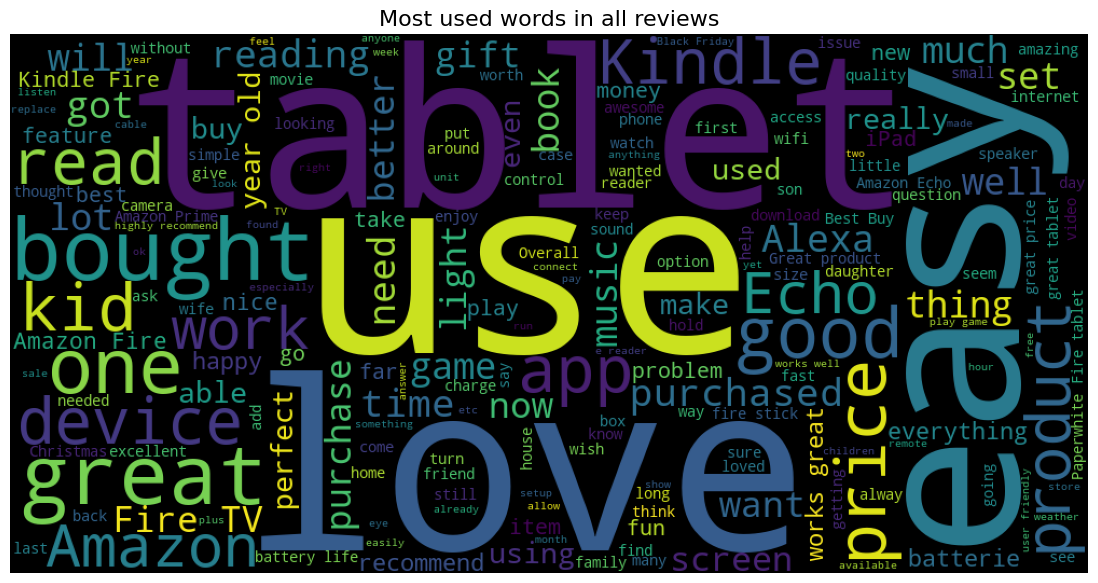


Figure 3a. Most used words in all reviews

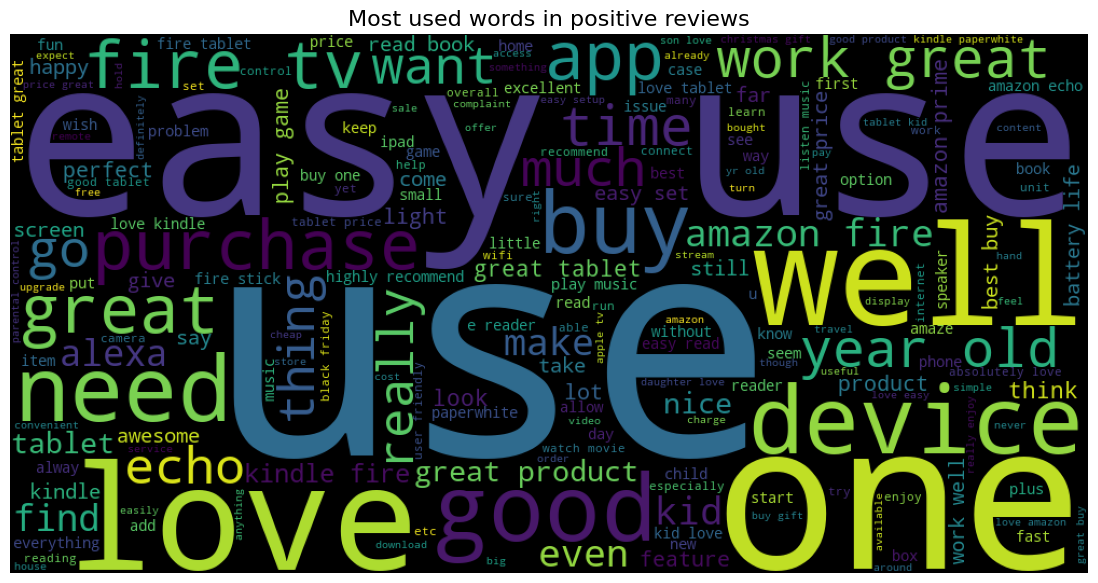


Figure 3b. Most used words in positive reviews



Figure 3c. Most used words in negative reviews

## 3.5 Preprocessing for the RNN model

### 3.5.1 Word Embedding

A technique used to represent words in a numerical form suitable for machine learning and natural language processing tasks (Meiying, Liu., et al., 2023). These embeddings can capture semantic meaning and relationships between words, which can be useful for various NLP tasks (Dongsuk, Oh., et al., 2022).

Tokenizing the cleaned review text involves assigning it to the text variable, calculating a vocabulary length of 12,115 unique words. Sequences are padded to the longest sentence length using the embed function, ensuring uniformity. GloVe, a pre-trained word embedding method, is loaded from Stanford's GloVe 100d word embeddings, capturing semantic word relationships. An embedding matrix aligning text data vocabulary with GloVe vectors is created, initializing with zeros for unmatched words (adapted from the method by Meiying, Liu., et al., 2023).

### 3.5.2 Data Sampling

To balance the class imbalance, the dataset undergoes oversampling. The text data is split into X\_train, y\_train (training), and X\_test, y\_test (testing) using an 80-20 ratio via train\_test\_split. Class imbalance is addressed using RandomOverSampler from imbalanced-learn, exclusively applied to the training set. Focusing on the minority class ('Negative' sentiment) to prevents data leakage, have model generalization, and gauges robustness against imbalanced data. This strategy ensures accurate measurement of the model's performance on unseen, imbalanced data.

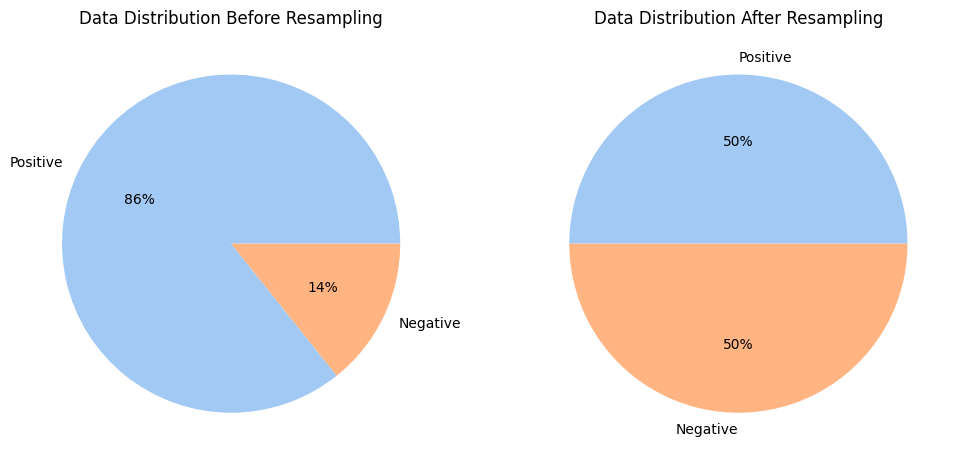


Figure 4. Data Distribution before oversampling and after oversampling

## 3.6 Preprocessing for BERT model

BERT is loaded, encompassing a tokenizer for text preprocessing and a model for contextual understanding in NLP. Token lengths are calculated to grasp text length distribution and set an optimal sequence length for BERT, avoiding efficiency issues with overly long sequences. The dataset is shuffled and split into 80-20 training and testing sets. To tackle class imbalance, RandomOverSampler balances sentiment classes, and one-hot encoding enhances model performance. Further, a stratified split ensures class distribution preservation in training, validation, and test sets. Tokenization and padding defined by Patawee, et al., 2022, involve a tokenize function for input IDs and attention masks, maintaining a consistent 128-token length for all sets.

## 3.7. LSTM Model Development

Long Short-Term Memory model for sentiment analysis is a type of recurrent neural network (RNN) that is known for handling sequential data effectively (Gaurav, Dubey., et al., 2023).

LSTM, a potent recurrent neural network (RNN) for sentiment analysis, excels in sequential data processing (Gaurav, Dubey., et al., 2023). The 'glove\_lstm' architecture in Keras involves a Sequential Model, where layers are sequentially added:

1. Embedding Layer: Converts input text data into a suitable deep learning format, with parameters like input\_dim (vocabulary size), output\_dim (embedding vector size impacting semantic relationships), weights (initialized with pre-trained embeddings from 'embedding\_matrix'), and input\_length (sequence length).

2. Bidirectional LSTM: Processes input sequences bidirectionally, with key parameters such as the number of LSTM units (32 for computational efficiency), return\_sequences (True to output full sequence), and recurrent\_dropout (set at 0.2 to prevent overfitting in recurrent connections).

3. GlobalMaxPool1D Layer: Reduces data dimensionality by taking the maximum value over the time dimension, resulting in a 2D tensor.

4. BatchNormalization Layer: Enhances training stability by normalizing activations from the previous layer.

5. Dropout Layer: Mitigates overfitting by randomly setting a fraction of input units to zero during training.

6. Dense Layers: Perform the final classification, with two dense layers having 64 and 32 units, ReLU activation for the first two layers, and a sigmoid activation for the final output layer.

**The model is compiled with these settings:**

1. Optimizer - 'RMSprop': An optimization algorithm for neural network training, RMSprop adjusts learning rates individually for each parameter. Its adaptability based on historical gradient information enhances convergence speed and training stability.

2. Loss - 'Binary Cross-Entropy': This metric quantifies dissimilarity between predicted class probabilities and actual labels. For sentiment analysis, it gauges how well predicted probabilities align with true binary labels. Lower binary cross-entropy values signify a better model fit.

3. Metrics - 'Accuracy': A widely used classification metric, accuracy measures the percentage of correctly classified data points among the total.

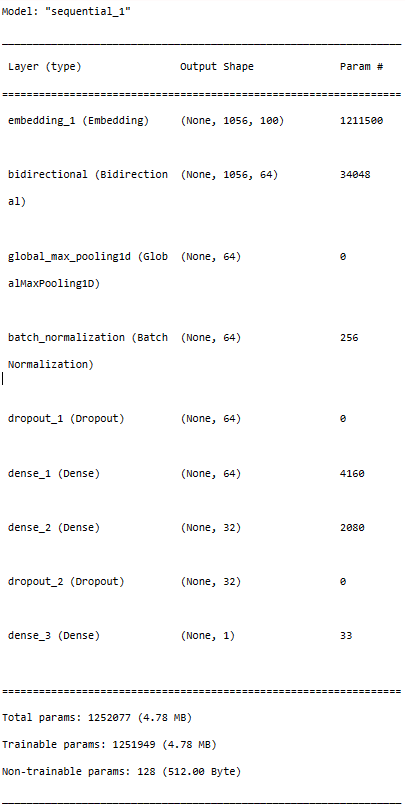


Fig 5a. Baseline Sequential Model Architecture

## 3.7.1 LSTM Model Training

The LSTM model was trained using the following key parameters:

Training Setup:

- Data and Labels: `train\_resampled` contains input text data, and `y\_train\_resampled` holds corresponding sentiment labels (positive or negative).

- Epochs: Set at 10, determining complete passes through training data.

- Batch Size: Affects efficiency and memory usage, fixed at 1024 for this model.

- Validation Data: (X\_test, y\_test) monitors model generalization, evaluating performance on unseen data after each epoch.

- Callbacks: Two crucial callbacks enhance training:

- ReduceLROnPlateau: Adjusts learning rate (`lr`) if 'val\_loss' (validation loss) plateaus, refining model performance.

- EarlyStopping: Halts training if 'val\_accuracy' (validation accuracy) stalls, preventing overfitting and saving time.

## 3.8 Tuned LSTM Model Development

To tune the LSTM model to get improved results the number of LSTM units in the Bidirectional LSTM layer was increased from 32 to 64. This increase in LSTM units provide the model with more capacity to capture complex patterns in the text data, potentially improving its ability to understand and predict sentiment.

The dropout rate in the model has been adjusted from 0.5 to 0.4. Instead of using the rmsprop optimizer as in the baseline model the tuned model uses the Adam optimizer with a custom learning rate of 0.001. Customizing the learning rate can help control the step size during gradient descent and potentially lead to faster convergence or better optimization.

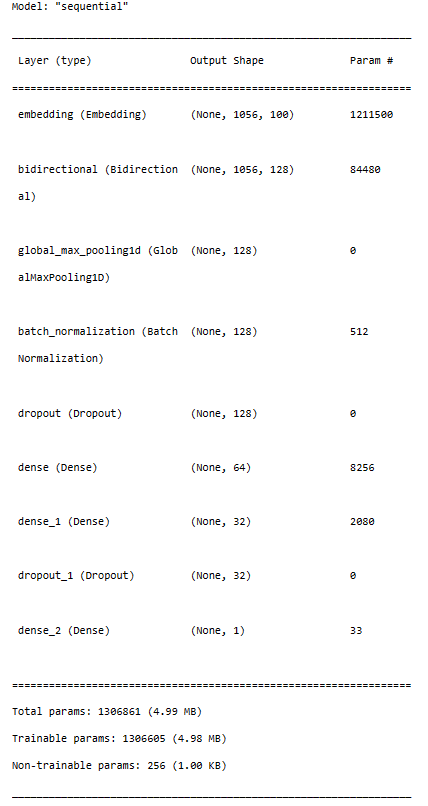


Fig 5b. Tuned Sequential Model Architecture

## 3.8.1 Tuned Model Training

The model was trained using the same parameters as in the baseline model

## 3.9 Bert Model

BERT is a deep learning model that processes text by creating contextual embeddings for each token in the input sequence (Harsh, Darji., et al., 2023). A BERT model called bert-base-uncased was loaded. This model is pre-trained on a massive amount of text data and is known for its ability to understand the contextual meaning of words in natural language (Md., Kamrul, Hasan., et al., 2023). It's a part of the "transformers" library for NLP. A sequence length is defined with the variable MAX\_LEN. This value specifies the maximum length for the input sequences. Text sequences longer than this length will be truncated, and sequences shorter will be padded to match this .

A function was created to create a custom text classification model using the BERT model. It takes the BERT model and max\_len parameter as inputs.

Optimizer, Loss, and Metrics: In this part, the code defines crucial components for training and evaluating the model. The optimizer, specifically Adam, is responsible for adjusting the model's internal parameters during training to minimize the loss. The loss function, Categorical Crossentropy, quantifies the error between predicted and actual class labels. Lastly, the evaluation metric, Categorical Accuracy, measures how well the model's predictions align with the true labels during training and evaluation. These components are essential for model training and assessing its performance.

Input Layers: Two input layers were defined input\_ids and attention\_masks. These layers are fundamental for processing text data with BERT. input\_ids contain the tokenized input text, where each word or subword is mapped to a unique identifier. attention\_masks help BERT focus on the relevant parts of the input while ignoring padding. These input layers ensure that the model receives structured text data for its operations.

The embedding sends the input through the BERT model, and the embeddings (vectors) generated represent the meaning and context of the words in the text. These embeddings are used for downstream tasks like classification.

An output layer was added to the model which is responsible for performing binary classification in this case. The output layer consists of two units and employs softmax activation, which converts the model's raw outputs into probability scores for each class. This allows the model to provide class probabilities, which can be used to make predictions.

The model is then compiled by bringing together all the defined components. This includes specifying the optimizer Adam, the loss function Categorical Crossentropy, and the metric for evaluation Categorical Accuracy.

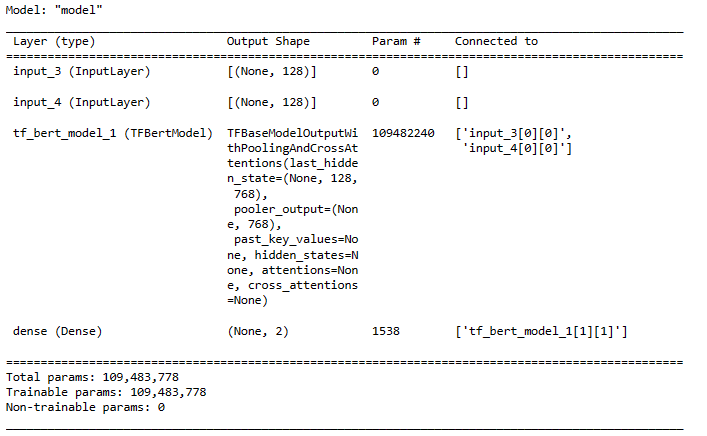


Fig 5c. Bert Model Architecture

# 4. Result

## 4.1 LSTM Model Evaluation

From Figure 6a. Accuracy is a measure of the model's overall correctness. It indicates the proportion of correctly classified instances out of the total instances in the test dataset. the accuracy is 87.94%. This means that the LSTM model correctly predicted the sentiment for about 87.94% of the samples in the test dataset.Loss is a measure of how well the model's predictions align with the true labels in the test dataset. It quantifies the difference between predicted values and actual values. The loss value is 0.2706. Lower loss values indicate better alignment between predictions and true labels.

Figure 7a. presents a confusion matrix illustrating the model's true positives, true negatives, false positives, and false negatives, accompanied by a concise Classification Report in Figure 8a. offering a detailed model performance summary.Precision, gauging the correctness of predicted positive cases, is 0.56 for the "Negative" class, signifying 56% accuracy, and notably higher at 0.97 for the "Positive" class, denoting 97% accuracy.Recall, alias sensitivity or true positive rate, measures correct predictions of actual positive cases. "Negative" class recall is 0.85 (85% identification accuracy), while "Positive" class recall is 0.88 (88% correct classification).The F1-score, a harmonic mean of precision and recall, provides a balanced accuracy measure, crucial for imbalanced class scenarios. The "Negative" class F1-score is 0.67, contrasting with the "Positive" class at 0.93, where higher F1-scores indicate superior balance between precision and recall.

|  |  |
| --- | --- |
| Fig 6a. Training-Validation Accuracy | Fig 6b. Training-Validation loss |

|  |  |
| --- | --- |
| Fig 7a. Confusion Matrix | Fig 8a. Classification Report |
|  |  |

## 4.2 Tuned LSTM Model Evaluation

From Fig 6c. and Fig 6d. above it is evident that there is improvement after tuning the LSTM model. The primary metrics considered in the report are accuracy and loss, as they are fundamental indicators of a model's performance. Below is the comparison of the model's performance before and after tuning to highlight the enhancements gained through this process.

|  |  |
| --- | --- |
| Scores After Tuning:  Accuracy: 0.8976  Loss: 0.2858 | Scores Before Tuning:  Accuracy: 0.8794  Loss: 0.2706 |

The accuracy of the LSTM model significantly improved after the tuning process.

|  |  |
| --- | --- |
| Fig 6c. Tuned LSTM Training-Validation Accuracy | Fig 6d. Tuned LSTM Training-Validation loss |

|  |  |
| --- | --- |
| Fig 7b. Confusion Matrix and | Figure 8b classification report |

## 4.3 BERT Model Evaluation

Using the BERT model for the prediction gives improvements in model performance when compared to the LSTM models both before and after tuning. The BERT model has achieved a notably higher accuracy score of 0.9391 and a slightly increased loss of 0.2808. These results signify a substantial boost in the model's ability to make accurate predictions and are a testament to the power of leveraging pre-trained transformer models like BERT for natural language understanding tasks.

|  |  |
| --- | --- |
| Fig 6e. Bert Train - Validation Accuracy | Fig 6f. Bert Train - Validation loss |

|  |  |
| --- | --- |
|  | Fig 8c. Classification Report |
| Fig 7c. Confusion Matrix |  |

## 4.4 Models Comparison

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model | Accuracy | Loss |
| 0 | LSTM | 0.8794 | 0.2706 |
| 1 | Tuned LSTM | 0.8976 | 0.2858 |
| 2 | BERT | 0.9391 | 0.2808 |

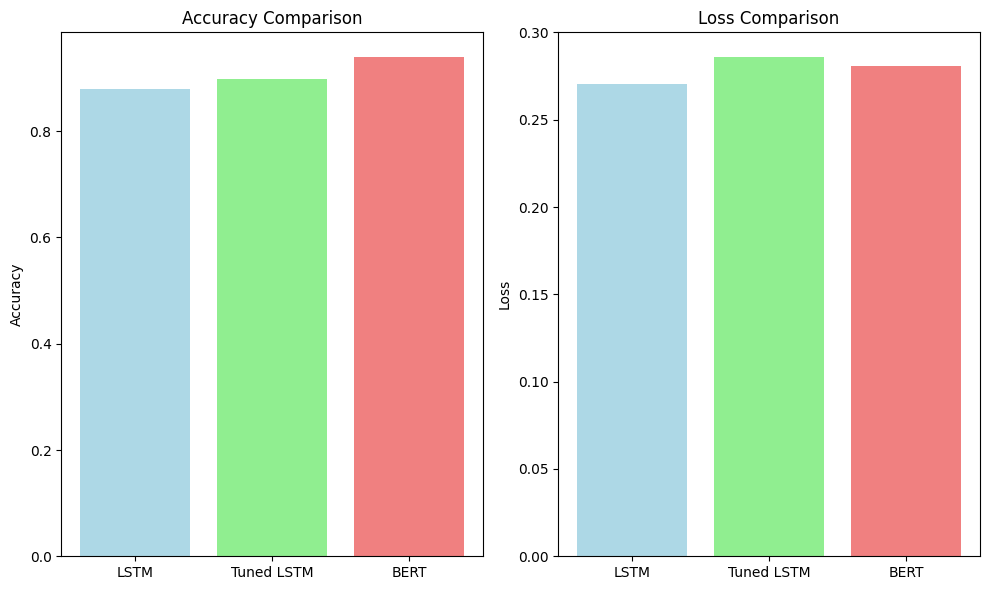


Fig 9. Models Results

# 5. Discussion

The sentiment analysis models demonstrate a comprehensive approach to text classification, making use of both traditional LSTM-based architectures and a BERT model. The initial LSTM model, named glove\_lstm, exhibits a well-structured architecture which uses an Embedding Layer, Bidirectional LSTM, GlobalMaxPool1D Layer, BatchNormalization Layer, Dropout Layer, and Dense Layers and provides a sequential architecture for processing sequential data effectively. The model's performance is evaluated using the RMSprop optimizer, Binary Cross-Entropy loss function, and Accuracy as the metric. The results of the glove\_lstm model training and evaluation show an accuracy of 87.94%. The confusion matrix and classification report provide additional insights into the model's performance, revealing precision, recall, and F1-score metrics for both positive and negative sentiments. However, to further improve the LSTM model's performance, a tuning process is performed. The tuned LSTM model shows improvements with an accuracy of 89.76% showcasing the effectiveness of adjusting hyperparameters such as the number of LSTM units, dropout rate, and optimizer.

The BERT model has a higher performance, leveraging the transformers library for NLP the BERT model processes text by creating contextual embeddings, and capturing intricate semantic relationships. The model achieves an accuracy of 93.91%, outperforming both the baseline LSTM and the tuned LSTM models. The BERT model's utilization of pre-trained contextual embeddings and its ability to understand the contextual meaning of words contribute significantly to its superior performance in natural language understanding tasks. The inclusion of the sentiment prediction function using the BERT model adds a practical aspect to the study showcasing how these models can be employed for real-world applications. The function utilizes the BERT model because it is the best performing model to predict sentiment based on input text.

In comparison to related work, this study aligns with the evolving landscape of sentiment analysis, acknowledging the transformative impact of digital platforms on consumer behavior. The notion of comparative sentiment analysis, as discussed by Arnd et al. and Norhaslinda et al., adds a valuable dimension, addressing the importance of understanding how products fare relative to competitors. The studies by Krishna et al. and Philippe et al. further expands the discourse, showcasing the advancements in machine learning and NLP methods, particularly the hybrid models combining BERT with BiLSTM and BiGRU algorithms. This study contributes to the evolving field of sentiment analysis by providing insights into both traditional and state-of-the-art models and aligning with the broader research landscape.

# 6. Conclusion

Two distinct approaches Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) were employed for sentiment classification. The LSTM model both in its baseline and tuned versions, demonstrated effective performance, achieving an accuracy of 87.94% and 89.76% respectively. Tuning involved adjusting parameters such as the number of LSTM units, dropout rate, and the optimizer. The BERT model a transformer-based architecture, outperformed the LSTM models, achieving an accuracy of 93.91%. BERT's contextual embeddings and advanced understanding of language nuances contributed to its superior performance.

The use of real-world examples demonstrates the BERT model's effectiveness in capturing nuanced sentiments from user reviews. This practical application underscores the model's potential for sentiment analysis in various contexts, aiding businesses in understanding customer feedback and making informed decisions based on user sentiments, the comparison of the three models highlights the progression from traditional LSTM architectures to advanced transformer-based models like BERT. While the LSTM models demonstrate good performance the BERT model performs better, emphasizing the importance of leveraging pre-trained models for complex natural language processing tasks.

Finally, the project successfully showcased the application of sentiment analysis in the e-commerce domain, demonstrates the importance of understanding customer sentiment for both consumers and businesses. The comparison of LSTM and BERT models demonstrated the impact of advanced techniques in natural language processing on model performance. This study contributes valuable insights for businesses seeking to enhance customer experiences, make informed decisions, and gain a deeper understanding of consumer preferences in the competitive realm of e-commerce.

# References

1. Tanishka, Balkrishna, Tiwatane., Vrushali, Balaji, Zampalkar., Gauri, Sunil, Bhavsar., Sonali, Rajendra, Bachute. (2022). Sentiment Analysis and Opinion Mining Applied to Scientific Paper Reviews. International Journal For Science Technology And Engineering, Available from: 10.22214/ijraset.2022.48240

2. Harish, U.C.., Dhanya, N.M.. (2022). Deciphering Voice of the Customer using Text Analytics and Sentiment Analysis: An Interpretable Review Rating Prediction using RoBERTa. International Journal on Recent and Innovation Trends in Computing and Communication, Available from: 10.17762/ijritcc.v10i12.5840

3. Zie, Eya, Ekolle., Ryuji, Kohno. (2023). GenCo: A Generative Learning Model for Heterogeneous Text Classification Based on Collaborative Partial Classifications. Applied Sciences, Available from: 10.3390/app13148211

4. NAOMI. (2023). Machine Learning for Natural Language Processing: Techniques and Applications. Available from: 10.59646/csebookc6/004

5. Daniel, Jakubovitz., Miguel, R., D., Rodrigues., Raja, Giryes. (2023). An information-Theoretic Approach to Semi-supervised Transfer Learning. arXiv.org, Available from: 10.48550/arXiv.2306.06731

6. Franco, B., Berbeglia, G., & Pascal, V. H. (2021). Market segmentation in online platforms. European Journal of Operational Research, 10.1016/J.EJOR.2021.03.056.

7. Marcelo, G. A., Armentano, E., Bagheri, E., Kiseleva, J., & Takes, F. W. (2020). Foreword to the special issue on mining actionable insights from online user-generated content. Journal Name, 10.1007/S10791-020-09380-2.

8. Marouane, B., Mohammed, K., & Abderrahim, B. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge Based Systems, 10.1016/J.KNOSYS.2021.107134.

9. Umarani, V., Julian, A., & Deepa, J. (2021). Sentiment analysis using various machine learning and deep learning techniques. Journal Name, 10.46481/JNSPS.2021.308.

10. Florack, A., Egger, M., & Hübner, R. (2020). When products compete for consumers' attention: How selective attention affects preferences. Journal of Business Research, 10.1016/J.JBUSRES.2019.05.009.

11. Yifan, Xu., Yong, Ren. (2021). Research on Sentiment Analysis Model of Online Shopping Product Evaluation Based on Machine Learning. 482-486. Available from: 10.1109/ICAICA52286.2021.9498066

12. Krishna, Kumar, Mohbey. (2021). Sentiment analysis for product rating using a deep learning approach. 121-126. Available from: 10.1109/ICAIS50930.2021.9395802

13. Philippe, Huneman. (2023). Sentiment analysis classification system using hybrid BERT models. Journal of Big Data, 10(1) Available from: 10.1186/s40537-023-00781-w

14. Asha, Patel., Bhavesh, Patel., Meghna, Patel. (2022). A Proposed Sentiment Analysis Model for Product Reviews on Social Media. Available from: 10.1109/ICAC3N56670.2022.10074561

15. Tanishka, Balkrishna, Tiwatane., Vrushali, Balaji, Zampalkar., Gauri, Sunil, Bhavsar., Sonali, Rajendra, Bachute. (2022). Sentiment Analysis and Opinion Mining Applied to Scientific Paper Reviews. International Journal For Science Technology And Engineering, Available from: 10.22214/ijraset.2022.48240

16. Arnd, Florack., Martin, Egger., Ronald, Hübner. (2020). When products compete for consumers attention: How selective attention affects preferences. Journal of Business Research, Available from: 10.1016/J.JBUSRES.2019.05.009

17. Norhaslinda, Kamaruddin, et.al. (2021). Comparative Study on Sentiment Analysis Approach for Online Shopping Review. Available from: 10.17762/TURCOMAT.V12I3.907

18. Meiying, Liu., Wendi, Luo., Zangtai, Cai., Xiujuan, Du., Jiliang, Zhang., Shuai, Li. (2023). Numerical‐discrete‐scheme‐incorporated recurrent neural network for tasks in natural language processing. CAAI Transactions on Intelligence Technology, Available from: 10.1049/cit2.12172

19. Dongsuk, Oh., Jungwoo, Lim., Heuiseok, Lim. (2022). Neuro-Symbolic Word Embedding Using Textual and Knowledge Graph Information. Applied Sciences, Available from: 10.3390/app12199424

20. Patawee, Prakrankamanant., Ekapol, Chuangsuwanich. (2022). Tokenization-based data augmentation for text classification. Available from: 10.1109/jcsse54890.2022.9836268

21. Gaurav, Dubey., Richa, Khera., Ashish, Grover., Amandeep, Kaur., R., Rajkumar., Harsh, Khatter., Somya, Srivastava. (2023). A Hybrid Convolutional Network and Long Short-Term Memory (HBCNLS) model for Sentiment Analysis on Movie Reviews. International Journal on Recent and Innovation Trends in Computing and Communication, Available from: 10.17762/ijritcc.v11i4.6458

22. Harsh, Darji., Jelena, Mitrović., Michael, Granitzer. (2023). German BERT Model for Legal Named Entity Recognition. Available from: 10.5220/0011749400003393

23. Md., Kamrul, Hasan., Md., Saiful, Islam., Sangwu, Lee., Wasifur, Rahman., Iftekhar, Naim., Mohammed, Ibrahim, Khan., Ehsan, Hoque. (2023). TextMI: Textualize Multimodal Information for Integrating Non-verbal Cues in Pre-trained Language Models. arXiv.org, Available from: 10.48550/arXiv.2303.15430