Sentiment Analysis for E-commerce customer review using BERT

By

Mohammad Kamrul Hasan (C193072)

Munir Uddin Rohan (C201066)

S.M. Anim (C201128)

Thesis report presented in order to complete the prerequisites for the Bachelor of Science in Computer Science & Engineering Degree

Department of Computer Science and Engineering

INTERNATIONAL ISLAMIC UNIVERSITY CHITTAGONG

Approved By

Sazid Zaman Khan Supervisor

Assistant Professor

Department of Computer Science and Engineering

International Islamic University Chittagong

Jamil As-ad Co-Supervisor

Lecturer

Department of Computer Science and Engineering

International Islamic University Chittagong

DECLARATION

We hereby declare that this thesis, "Sentiment Analysis for E-commerce Customer Review using BERT" is entirely original and has not been previously presented to any academic institution for the purpose of obtaining a degree. Our independent research, which was undertaken at the International Islamic University of Chittagong under the supervision of our Supervisor Sazid Zaman Khan and Co-supervisor Jamil As-ad, is the source of the work given in this thesis.

We certify that all sources used in this thesis have been properly cited and acknowledged.

Any support received during the course of the study has been duly recognized in the thesis. We are aware that academic dishonesty, such as plagiarism or fabricating data, carries serious repercussions, including the possibility of having our degree revoked.

Finally, we provide the non-exclusive right to International Islamic University Chittagong to archive and make this thesis accessible for future study and academic purposes.

25 November 2024

Mohammad Kamrul Hasan [C193072]

Munir Uddin Rohan [C201066]

S.M.Anim [C201128]

SUPERVISOR'S DECLARATION

I, Sazid Zaman Khan, serving as an Assistant Professor within the Department of Computer Science & Engineering at International Islamic University Chittagong, do hereby confirm that I have meticulously reviewed the document titled "Sentiment Analysis for E-commerce Customer Review using BERT" authored by Mohammad Kamrul Hasan, Munir Uddin Rohan, and S.M. Anim.

The authors, Mohammad Kamrul Hasan, Munir Uddin Rohan, and S.M. Anim, have shown extraordinary devotion and proficiency. The thesis represents an important contribution to the field of computer science and engineering and demonstrates a thorough mastery of the topic. The applicants' study was also carried out directly under my direction and meets with ethical standards.

After carefully analyzing the study, I can confidently suggest that this thesis be accepted in order to satisfy the criteria for the B. Sc. in Computer Science & Engineering degree.

Signed: ___________________ 25 November 2024

Sazid Zaman Khan

Assistant Professor, Department of CSE

International Islamic University Chittagong

DECLARATION OF

THESIS REPORT AND COPYRIGHT

THESIS REPORT TITLE: Sentiment Analysis for E-commerce customer review using BERT

AUTHORS:

SL NO.	AUTHOR'S NAME	STUDENT ID	SIGNATURE
1			
2			
3			

Sazid Zaman Khan

Assistant Professor

Computer Science and Engineering

International Islamic University Chittagong

We, Mohammad Kamrul Hasan , Munir Uddin Rohan and S.M. Anim, hereby declare that the title of our thesis is " Sentiment Analysis for E-commerce customer review using BERT ".

1. We declare that this thesis will be open to everyone at the university database or archive, allowing it to be freely accessible to the public.
2. The Department of CSE, has the following rights:
   i. This work belongs to the Department of CSE and any reproduction or distribution without prior written permission from the department is strictly prohibited.
   ii. University's library has the right to make copies for academic activities.
   iii. The library can copy this work to share knowledge within the academic community.

We understand the implications of making my thesis publicly accessible and agree to abide by the rules and regulations set forth by the Department of CSE, IIUC, regarding the use and distribution of this work.

ACKNOWLEDGMENT

We would like to extend our sincere appreciation to everyone who has been instrumental in making this thesis feasible. This feat would not have been possible without their assistance, direction, and support.

First of all, we would like to begin by expressing our sincere gratitude and humble submission to Allah, the Almighty, for His divine direction and blessings throughout this journey.

We sincerely appreciate the helpful guidance and advice provided by Assistant Professor Sazid Zaman Khan. We were able to finish this thesis in a significant way thanks to his kind and well-deserved advice. We've had a lot of help from his creative ideas and suggestions as we've navigated new approaches. We are also appreciative of the chance he gave us to work on this wonderful thesis under his guidance, as well as of his ongoing support and insightful advice.

ABSTRACT

The rapid growth of e-commerce platforms has resulted in an immense volume of customer reviews, offering valuable insights into user satisfaction and preferences. Sentiment analysis plays a critical role in understanding these reviews, enabling businesses to enhance decision-making and customer experience. This study explores the application of DistilBERT, a lightweight and efficient variant of the BERT model, for sentiment analysis of e-commerce customer reviews.

We utilized datasets from Amazon and Flipkart, creating three distinct datasets: a multilabel dataset (0, 1, 2) representing different levels of sentiment and two binary datasets (positive/negative). Each dataset comprised 20,000 samples to ensure a robust evaluation. The DistilBERT model was fine-tuned to classify customer reviews across these datasets, addressing the challenges of multilabel classification and cross-platform data variability.

Our results demonstrate that DistilBERT achieves 97% and 92% accuracy in two binary and 86% accuracy in one multilabel sentiment classification tasks, highlighting its effectiveness in handling e-commerce data. This study contributes to advancing sentiment analysis by providing insights into the performance of Transformer-based models in multilabel and binary contexts, offering scalable solutions for analyzing large-scale customer reviews.

Keyword : DistilBERT, BERT, binary, multilabel, sentiment analysis

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SA	Sentiment Analysis
BERT	Bidirectional Encoder Representations from Transformers
DL	Deep Learning
NN	Neural Network
ANN	Artificial Neural Network
LSTM	Long Short Term Memory
SVM	Support Vector Machine
EC	E-commerce

CHAPTER I

INTRODUCTION

1.1 Overview

The proliferation of e-commerce platforms such as Amazon and Flipkart has revolutionized how consumers make purchasing decisions. A critical factor influencing these decisions is customer reviews, which provide direct feedback on products and services. These reviews contain valuable insights into customer satisfaction, preferences, and expectations, making them a vital resource for businesses aiming to improve their offerings and customer experiences. However, analyzing and extracting actionable insights from large volumes of textual reviews is a challenging task that necessitates robust and scalable sentiment analysis techniques.

Sentiment analysis, a subfield of natural language processing (NLP), focuses on identifying and categorizing opinions expressed in text. Traditionally, machine learning techniques such as Support Vector Machines (SVM) and Naïve Bayes have been employed for sentiment analysis. However, these methods often struggle with the nuances of language, such as context and semantics, limiting their effectiveness for complex datasets like customer reviews. With the advent of deep learning and Transformer-based architectures, such as BERT (Bidirectional Encoder Representations from Transformers), the field has witnessed significant advancements in capturing contextual information and improving classification accuracy.

This study leverages **DistilBERT**, a compact and efficient variant of BERT, to perform sentiment analysis on customer reviews from Amazon and Flipkart. DistilBERT retains much of BERT's performance while reducing computational overhead, making it well-suited for large-scale e-commerce applications. To evaluate its effectiveness, three datasets were prepared: one for multilabel classification, categorizing sentiments into three levels (negative, neutral, and positive), and two for binary classification (positive and negative). Each dataset comprises 20,000 samples, ensuring a balanced and comprehensive evaluation.

1.2. Sentiment Analysis using BERT

Sentiment analysis is a vital task in natural language processing (NLP) that focuses on determining the sentiment or emotion expressed in a piece of text. It is widely used in domains such as customer feedback analysis, social media monitoring, and brand reputation management. Traditional approaches to sentiment analysis relied on rule-based methods or machine learning techniques, which often struggled to capture the intricacies of human language, such as context, sarcasm, and word dependencies. However, the introduction of Transformer-based architectures, particularly BERT (Bidirectional Encoder Representations from Transformers), has revolutionized this field by offering state-of-the-art performance in understanding the semantics and context of text.

DistilBERT, a smaller, faster, and more efficient variant of BERT, has emerged as a popular choice for tasks that require computational efficiency without significant loss in performance. By retaining 97% of BERT's capabilities with only 60% of its parameters, DistilBERT provides a scalable solution for applications like sentiment analysis, where datasets are often large and diverse.

1.3. Application of Sentiment Analysis

Sentiment analysis, also known as opinion mining, has become a cornerstone of natural language processing (NLP) applications, offering immense value across a variety of industries. It involves analyzing textual data to determine the sentiment expressed, such as positive, negative, or neutral. This ability to interpret emotions and opinions from text enables organizations to gain actionable insights, improve decision-making, and enhance customer experiences.

1.4. Objective of Sentiment Analysis

The objective of this study is to develop an effective sentiment analysis model using DistilBERT for classifying e-commerce customer reviews into distinct sentiment categories. Specifically, the study aims to:

Analyze Customer Sentiments: Identify and categorize sentiments (positive, neutral, negative, etc.) expressed in e-commerce reviews to understand customer opinions and feedback.

Enhance Business Decision-Making: Enable businesses to leverage sentiment insights for improving product quality, customer service, and marketing strategies.

Evaluate Model Performance: Assess the performance of DistilBERT in comparison to other models for sentiment analysis in terms of accuracy, efficiency, and scalability.

Provide Actionable Insights: Provide businesses with actionable insights derived from customer reviews to enhance customer satisfaction and improve overall business performance.

1.5. Significance of the Thesis

This thesis is significant as it explores the application of advanced Natural Language Processing (NLP) techniques, specifically DistilBERT, in analyzing e-commerce customer reviews. By automating the sentiment analysis of large-scale reviews, this research addresses a key challenge faced by e-commerce businesses: efficiently understanding customer feedback.

The findings of this thesis are crucial for several reasons:

Improved Customer Insight: It provides businesses with deeper insights into customer satisfaction, preferences, and concerns, which can be used to refine product offerings, enhance user experience, and improve customer service.

Data-Driven Decision Making: The automated analysis of sentiment allows for quicker, more accurate decision-making processes, enabling businesses to adapt swiftly to customer needs and market trends.

Scalability and Efficiency: By utilizing DistilBERT, this research demonstrates how to scale sentiment analysis processes to handle large volumes of reviews, making it possible for businesses to analyze massive datasets efficiently.

Practical Application: The thesis bridges the gap between theoretical advancements in NLP and their real-world applications in e-commerce, offering businesses a practical tool to process and analyze customer feedback automatically.

1.6. Objective of the Thesis

The primary objective of this thesis is to develop and evaluate a sentiment analysis model using DistilBERT for e-commerce customer reviews. The specific goals of this research are:

To Implement Sentiment Analysis on E-commerce Reviews: Develop a robust model using DistilBERT to classify e-commerce customer reviews into distinct sentiment categories, such as positive, negative, and neutral.

To Assess Model Performance: Evaluate how DistilBERT perform across different sentiment label and review size.

To Optimize Preprocessing Techniques: Investigate and implement efficient text preprocessing techniques (such as spell checking, word splitting, and tokenization) to improve the quality and accuracy of sentiment classification.

Through these objectives, the thesis aims to demonstrate the practical utility of advanced NLP models in addressing real-world challenges in the e-commerce sector.

CHAPTER II

THEORETICAL BACKGROUND

2.1 Overview

The fields of technology DL, NN, AI, and ML are all interconnected. All of these topics are collectively referred to as AI. The field of AI known as ML focuses on the methods that enable computers to learn from data and form conclusions or predictions. NNs serve as the essential building blocks of DL, a branch of ML**[15]**. All of the theoretical information that will be used throughout our entire thesis will be explained in this chapter. The Venn diagram below demonstrates how DL is a subset of ML, which is employed in many but not all methods of artificial intelligence. An illustration of AI technology may be found in each section of the Venn diagram**[16]**.

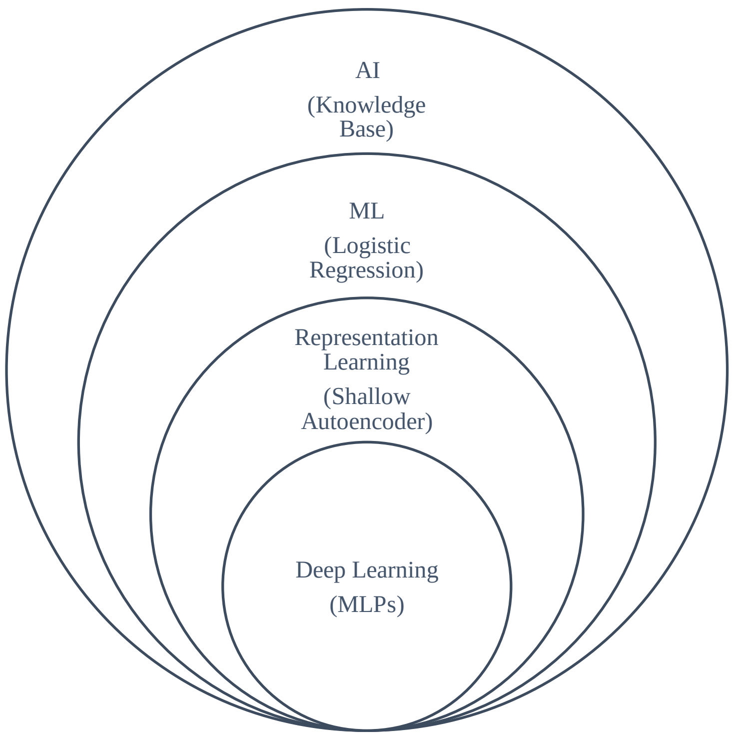


Fig. 2. 1: A Venn Diagram for AI Vs ML Vs Deep Learning

2.2. Deep Learning

DL is a subclass of ML that varies in its learning process and data consumption. Automating feature extraction with deep learning reduces the requirement for human involvement. It is a scalable machine learning approach that excels at handling huge datasets and is useful for evaluating unstructured data, which accounts for a sizable amount of an organization's data. Deep learning models are able to find data patterns and the right group inputs. DL’s models need more data points to increase accuracy. DL is well suited for difficult jobs like text classification, sentiment classification etc.

2.3. Neural Network

One particular branch of ML is NN. For DL algorithms, they serve as the foundation. They are termed "neural" because they mirror the communication process of neurons in the brain. Input, hidden, and output layers are three of the layers that exist. Each node has a threshold value and a weight. A node becomes active and transmits data when its output surpasses the threshold value. When it falls below the threshold, no data is transmitted. Learning from training data allows neural networks to become more accurate over time. Tasks such as recognizing spoken language and identifying images, which often require extended periods of time can be completed in a matter of minutes with neural networks. A well-known example of a neural network is Google's search engine**[15]**.

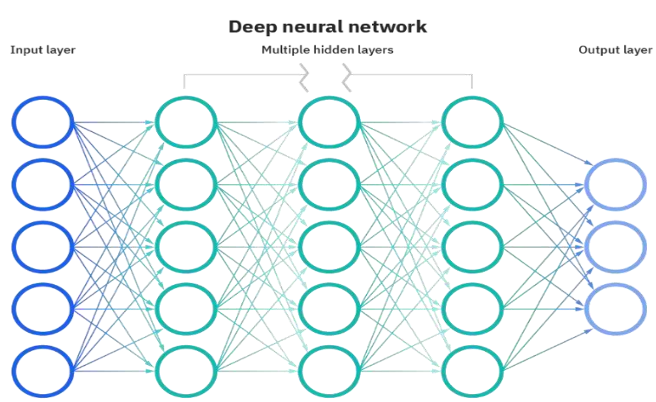


Fig. 2. 2: A typical deep neural network

2.4. BERT Model

BERT apply the bidirectional training of transformer to learns contextual relations between words in text, which transformer includes two separate mechanisms that an encoder that reads the text input and a decoder that produces a prediction for the task. As opposed to single-direction models of RNN (Recurrent Neural Network) and LSTM (Long Short-Term Memory), which read the text input sequentially (left-to-right or right-to-left), the transformer encoder reads the entire bidirectional sequence of words at once. It has caused a stir in the machine learning community by presenting state-of-the-art results in a wide variety of NLP (natural language processing) tasks and others. The novel technique of BERT is the masked language model (MLM) which allows bidirectional training in models. This characteristic allows the BERT model have deeper sense of language context and learn the context of word by all of its surrounding, which it was previously impossible **[7].**

2.5. DistilBERT Model

DistilBERT is a lightweight version of BERT, created by applying knowledge distillation, where a smaller model (DistilBERT) learns to mimic the behavior of a larger, pre-trained BERT model. It retains the core architecture of BERT but reduces the number of parameters and layers, making it faster and more efficient while maintaining most of BERT's performance. By leveraging the same transformer architecture and training objectives, DistilBERT inherits BERT's ability to understand contextual relationships in language tasks.

DistilBERT is significantly more efficient than BERT, achieving 60% fewer parameters and a 40% reduction in model size, which leads to faster inference and lower computational requirements. Despite these reductions, DistilBERT retains approximately 97% of BERT's performance on various natural language processing tasks, making it a highly effective alternative for resource-constrained environments or applications requiring real-time processing **[6].**

2.6. Activation Function

The terms Threshold Function and Activation Function are interchangeable. In ANNs, activation functions play a vital role in converting incoming signals into outputs, which then serve as input for the next layer in the sequence **[10]**. The activation functions that are most frequently utilized are:

- Binary Step Function
- Linear Function
- Sigmoid Function
- Tanh
- ReLU
- SoftMax

2.6.1. ReLU

ReLU has the benefit of avoiding simultaneous activation of all neurons. Until the result of a linear transformation hits 0, a neuron will continue to operate**.** Mathematically, it may be expressed as:

$$f\,(x) = \max(0, x)$$

ReLU stands out as more efficient compared to other functions, as it activates only a subset of neurons at a given time, as opposed to all of them concurrently **[11]**.

2.6.2. GeLU

The Gaussian Error Linear Unit (GeLU) activation function is a smooth approximation of the Rectified Linear Unit (ReLU). An approximate formula for GeLU is:

$$GeLU(x) \approx 0.5x(1 + tanh(\sqrt{\frac{2}{\Pi}}(x + 0.044715x^3))$$

This function allows smoother gradient flow compared to ReLU, making it particularly useful in models like BERT and DistilBERT.

2.6.3. Linear Function

An activation function that grows or shrinks in response to input is called a linear activation function. The primary disadvantage of the binary step function was its lack of gradient, as it did not consider the value of x in the function. The application of a linear function can get rid of that problem. It is described as:

$$F(x) = ax$$

Any constant value selected by the user may be used as the value of variable "a" **[13]**.

2.6.4. Sigmoid Function

The most often used activation function is sigmoid. The sigmoid function transforms numbers between 0 and 1 **[13].** It is described as:

$$f\ (x) = {}^{1}/e_{-x}$$

2.6.5. SoftMax

The SoftMax function combines many sigmoid functions **[13].** The mathematical expression for SoftMax function is:

$$\sigma\ (Z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}\ for\ j = 1, ..$$

2.7. Transformer Architecture

The Transformer architecture is a deep learning model designed for sequence-to-sequence tasks, like language translation and text generation. It uses an **encoder-decoder structure**, but models like BERT and DistilBERT only use the encoder. It is based entirely on self-attention mechanisms, discarding traditional recurrent layers (RNNs) and convolutional layers (CNNs), making it highly parallelizable and efficient for large-scale tasks.

2.7.1. Embedding Layer

The Embedding Layer is a type of layer in neural networks used to convert categorical data, typically words or tokens, into continuous vector representations. These embeddings capture semantic meanings by representing words with dense vectors in a lower-dimensional space, where similar words have similar vector representations.

In natural language processing (NLP), the embedding layer maps each word in the vocabulary to a fixed-size vector of real numbers. Instead of using sparse one-hot encoding, which results in high-dimensional vectors with mostly zeros, word embeddings provide a more compact and meaningful representation.

2.7.2. Transformer Encoder Layer

The Transformer Encoder Layer in DistilBERT is similar to the one in the original Transformer architecture but with some modifications to reduce the model size and improve efficiency. DistilBERT is a distilled version of BERT, designed to retain much of BERT's performance while being smaller and faster. Despite the distillation process, the core Transformer architecture remains intact, including the encoder layers.

2.7.3. Pooling Layer

In the DistilBERT model, the pooling layer is responsible for aggregating information from the token-level outputs of the transformer into a fixed-size representation that can be used for downstream tasks like classification. Typically, the model uses the embedding of the **`[CLS]`** token from the final hidden layer as a summary representation of the entire input sequence. This embedding is designed to capture contextualized information about the input as a whole. For classification tasks, this pooled output is often passed through an optional dense layer for further feature refinement before being sent to a softmax layer to generate the final class probabilities. This process ensures that variable-length input sequences are converted into a fixed-size vector representation suitable for task-specific predictions.

2.7.4. Fully Connected Layer

The fully connected layer in the Transformer architecture plays a crucial role in transforming and enriching the representations of tokens after they have been processed by the self-attention mechanism. Unlike self-attention, which captures contextual relationships between tokens, the fully connected layer operates independently on each token's representation. It consists of a feedforward neural network that applies a sequence of linear transformations, typically increasing the hidden dimensionality in the first layer and reducing it back in the second layer, with a non-linear activation function, such as ReLU, applied in between.

This layer adds non-linearity to the model, enabling it to capture complex patterns and relationships that are essential for tasks such as sentiment analysis, text classification, and language understanding. By complementing the contextualized outputs of self-attention, the fully connected layer ensures that both local and global patterns are effectively learned. This transformation helps refine token representations, preparing them for further processing in subsequent layers of the Transformer.

2.7.5. Dropout Layer

Dropout, a kind of regularization used in deep learning, increases testing accuracy at the cost of training accuracy in order to minimize overfitting. Dropouts randomly disconnect inputs between network design layers with a probability of p during training. This unpredictability prevents several nodes from being primarily responsible for activating in response to a particular input pattern, which encourages generalization. Between fully connected (FC) layers, DO with 0.5 probability are often used, whereas previous levels may use DO with 0.10 and 0.25 probabilities after down-sampling procedures. After the forward and backward passes, the disconnected connections may be reconnected, retaining the regularization effect while allowing for the sampling of a new set of disconnected connections.

Below are two network architectures: one with dropouts and one without.

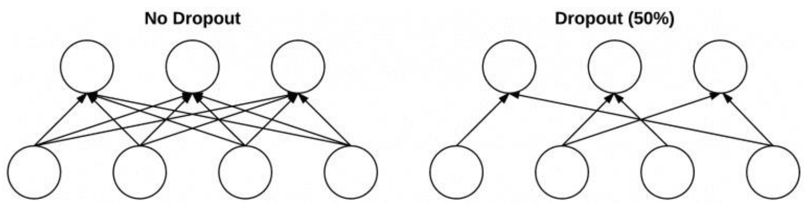


Fig. 2. 3: Two architectures with and without dropout

2.7.6. Output Layer

The output layer in a Transformer-based model, such as DistilBERT, is the final layer that produces predictions for a specific task. In the case of a classification task like sentiment analysis, the output layer typically consists of a fully connected layer followed by an activation function appropriate to the task.

The input to the output layer is usually the representation of the special token `[CLS]` (classification token) from the last hidden layer of the Transformer. This `[CLS]` token is designed to aggregate information from the entire input sequence, making it suitable for tasks where a single output prediction is required.

CHAPTER III

LITERATURE REVIEW

3.1. Overview

This chapter reviews the existing body of work on sentiment analysis, with a specific focus on the application of BERT and related models. Sentiment analysis, a key area within natural language processing (NLP), has evolved significantly from traditional lexicon-based and classical machine learning methods to the adoption of deep learning and Transformer-based models. This transition has been driven by the need for more accurate, context-aware, and scalable solutions for tasks such as analyzing customer feedback.

The review begins with an exploration of traditional sentiment analysis techniques, discussing their limitations in handling the nuances of language, such as sarcasm, context dependence, and complex sentence structures. It highlights the challenges posed by large-scale datasets in e-commerce, which often contain noisy, unstructured text.

The emergence of Transformer-based models, particularly BERT (Bidirectional Encoder Representations from Transformers), marks a significant advancement in the field. BERT's ability to process text bidirectionally and capture intricate contextual relationships has set new benchmarks in sentiment analysis. This chapter examines studies that have applied BERT to sentiment analysis tasks, emphasizing its superior performance compared to earlier approaches.

3.2. Literature Review on SA using BERT's model

We reviewed some literature on SA using BERT's model, where we have seen different types of method. For our model building this literature review has great impact. After reviewing the literature we found some research gap and thought about our model. Summary of literature review are given below.

3.2.1. Study 1: Experimenting performance of DistilBERT with traditional WV model

In **[1],** The experiment compared the performance of DistilBERT with traditional word vector models like Fast Text, Word2Vec, and GloVe. Despite the lack of fine-tuning, DistilBERT showcased promising results, outperforming traditional models. This underscores the effectiveness of leveraging deep learning techniques, particularly BERT models, in enhancing sentiment analysis accuracy and efficiency. In this experiment they used binary sentiment that means positive and negative. But here we used also multilabel sentiment that means positive , negative and neutral.

We also used large dataset than them.

Table 3. 1: The summary of the literature **[1]**

Authors	Domain	Methodology		Findings	Limitation
Yichao Wu 1* ,Zhengyu Jin 1, et al	SA	Dataset	Techniques	92.07%	Evaluated using a small benchmark dataset
		SST2	distilBERT,Word2Vec ,Fast Text,GloVe		

3.2.2. Study 2: Comparing performance of SA using hybrid BERT model

In **[2],** This paper provided hybridizing methods for developing a deep learning model for tweet emotion classifcation using multiple labels for three datasets. Hybrid model such as Random forest, Naive bayes, KNN, Decision Tree, SVM, XGboost, Logistic Regression and BERT model such as DistilBERT , RoBERTa used for comparision. They used Three dataset. Airline, CrowdFlower, Apple here. They worked also with emoji text data and non-emoji text data. In Airline and CrowdFlower dataset they got 83% and 79% accuracy but we go 86% accuracy for our multilabel dataset.

3.2.3. Study 3: SA on amazon review using traditional ML,DL and BERT

In **[3],** This paper two dataset D1 and D2 are used. They used traditional ML model like SVM, LR, RF, MNB, DL model like CNN, LSTM and BERT model for SA. Here D1 is only telecommunication based and D2 is mixture of all product review. So here they got 90% accuracy from D1 using BERT and 97% accuracy from D2 using CNN. In our project we used distilBERT model on flipkart and amazon review and we got 97% and 92% accuracy respectively.

3.2.4. Study 4: SA for EC product using BERT-BiGRU-Softmax

In **[4],** This paper BERT-BiGRU-Softmax model are used against RNN, BiGRU, BiLSTM. Here BiGRU actually enhance the functionality of BERT model. So here BERT-BiGRU-Softmax outperform RNN, BiGRU and Bi-LSTM extreamly. It achieved 95.5% accuracy. We got 97% accuracy in limited time using distilBERT , though our dataset was smaller than them.

3.2.4. Study 5: Study on current technique for SA on EC platform

In **[5],** This paper conducted a review on current technique and future direction for SA on EC platform. They discovered a proportion of machine learning and deep learning model which mostly used in SA for EC platform. They found that SVM, a traditional ML model is used mostly for SA in their mentioned paper. Lowest used model they found is transformer based model. This is the exect gap we found and determined to apply a transformer based model on SA for EC product review.

3.2.5. Study 6: Review of application of BERT on NLP

In **[6],** This paper describe the mechanism of operation of the model. It also review the area or field it applied. Comparision with similar model in each task is also an interesting part of this paper. Bert is a powerful model to handle large amount of data and that’s why picked it for build our SA model on EC product review.

3.3. Summary of the Literature Review

Sentiment Analysis is being a very hot topic at recent time. And everyone is trying to prove the power of their targeted model by fine tuning it. Comparison related research giving us knowledge to know the power of those model in the field of sentiment analysis. Sometimes physical resource is being barrier of using some of those model. Above paper shown us comparison between different model and accuracy.

3.4. Research Gaps

Our proposed strategies addresses various research gap in the field of Sentiment Analysis. In our research we shown that transformer based contextualized word embedding approach is better than traditional word embedding system. We also shown that BERT model always outperform traditional ML based model. We specifically worked with E-commerce product review dataset and shown that how to get perfect accuracy from different dataset. After all we can say that still there are many research gap exist in the field of sentiment analysis and we have to work with them to step forward.

3.5. Problem Statement

With several application sentiment analysis has become a thrilling area. Many effort have been made in sentiment analysis sector, however certain problem have not yet been solved, creating prospects of further study. How good a model will perform , it will depends on the dataset, how long the reviews are? How cleaned the dataset are? How much category are there? All this things has effects on model's performance.

Absence of evaluation on different datasets which causes comparatively low performance while dealing with bigger and unfamiliar dataset. As a result we worked with large and small review dataset to compare the performance of our proposed model.

CHAPTER IV

METHODOLOGY

4.1 Overview

In this chapter, we go into detail about the methods we used to create our thesis, "Sentiment Analysis for E-commerce customer review using BERT". We carefully examined the field of SA with an emphasis on the DistilBERT model to lay the foundation. This chapter will describe the details about how we collected the dataset to how we trained our model. Every step by step process will be listed in this chapter. The outline of this section are

4.2. Experimental Setup

Here, we will explain about the experimental setup for our sentiment analysis model. The use of the right tools and resources has a significant impact on the efficacy and accuracy of our suggested strategy. As a result, we list the key components of our experimental setup, such as the programming languages selected for implementation, the integrated development environment (IDE) used for coding and experimentation, as well as the libraries and frameworks integrated to speed up the design and evaluation of our model.

4.2.1. Integrated Development Environment

We used Google Collaboratory (Colab) as the IDE for our goal to develop our SA model. Colab is a free tool offered by Google Research that executes Python code directly in a web browser. It provides a hosted Jupyter notebook environment that can use processing resources like GPUs and TPUs for faster deep learning library performance without the need for setup. Our SA model were easily integrated into Colab because of its interoperability with Keras and TensorFlow. Additionally, its ability to automatically save files to Google Drive made it simple to import data from Github and Google Drive. Overall, Google Collaboratory proved to be the perfect IDE, allowing for rapid experimentation, easy development, and efficient sharing of our work.

4.2.2. Libraries

Our SA model was implemented using a significant number of key Python libraries. We were able to create and deploy our distilBERT model effectively throughout a range of processors thanks to TensorFlow's end-to-end machine learning environment. Essential numerical transformations and array manipulation features were given by NumPy, while advanced data visualization and analysis were made possible by Matplotlib. Scikit-Learn gave us access to a variety of machine learning methods, enabling feature extraction and model training.

4.3. Flow diagram

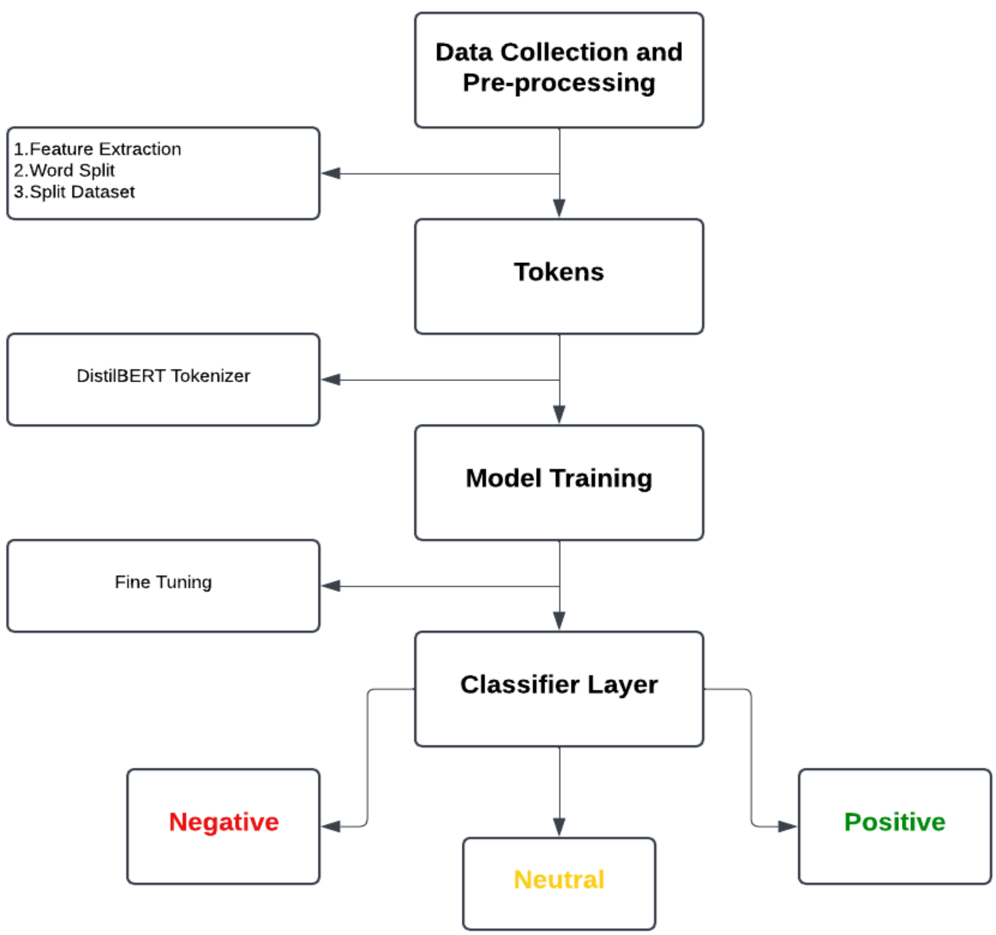


Fig.4.1: Flow Diagram of our Model

4.4. Model Architecture

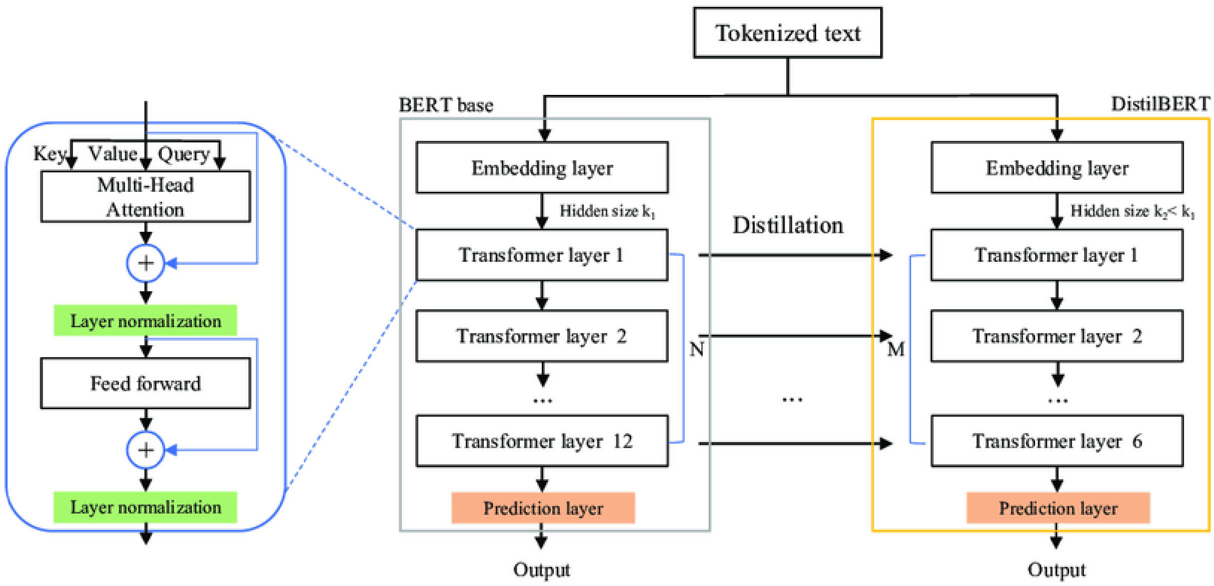


Fig 4.2: DistilBERT model architecture

4.5. Data Collection and Dataset Description

The dataset we used in this study was taken from kaggle. We picked dataset of two well known E-commerce website, they are Amazon **[5]** and Flipkart **[14].** In these dataset there are random reviews of random product. The Amazon product review dataset has binary sentiment, that means only positive and negative and Flipkart product review dataset has multi label sentiment that means, positive , negative and neutral. We further divided Flipkart product review dataset to made a third dataset that contain only positive and negative sentiment for this study purpose.

4.6. Data Preprocessing

Pre-processing of three dataset are given below.

4.6.1. Amazon Product Review dataset

Amazon product review dataset we got from kaggle were too large. After cleaning the dataset we randomly dropped row and took 20,000 data. Review in this dataset are comparatively large. Average review size is 404 word.

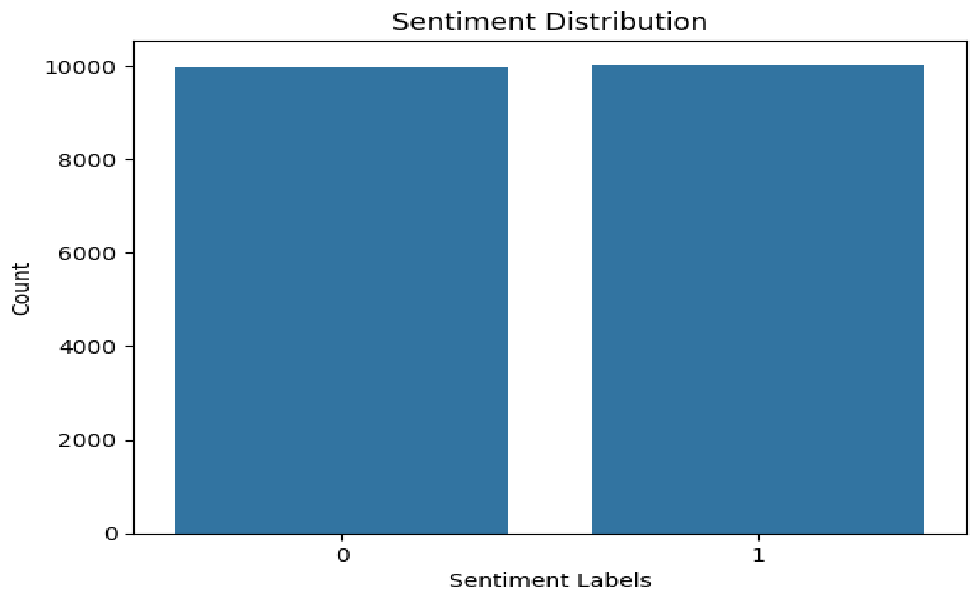


Fig 4.3: Amazon review dataset

4.6.2. Flipkart Product Review dataset

Flipkart product review dataset we also got from kaggle. After cleaning dataset we randomly dropped row and took 20,000 data. Then we have done some pre-processing like word splitting, spelling correction etc. Review in this dataset are comparatively small. Average review size are 45 word. This dataset contain multi label sentiment, that means positive, negative and neutral. For this study purpose, similarly we cleaned, pre-processed and randomly dropped row to make another version of this dataset which contain only positive and negative sentiment.

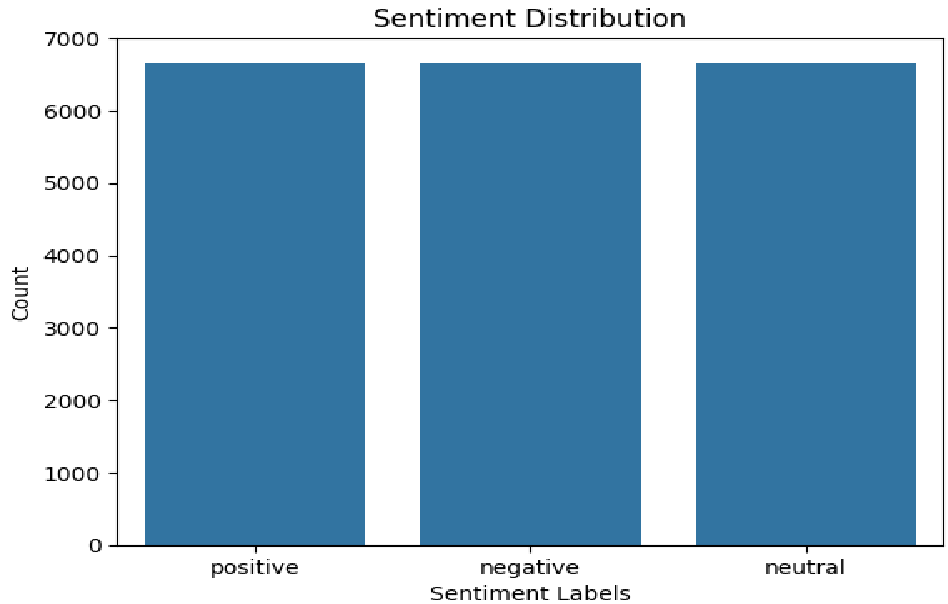


Fig 4.4: Flipkart review dataset (version 01)

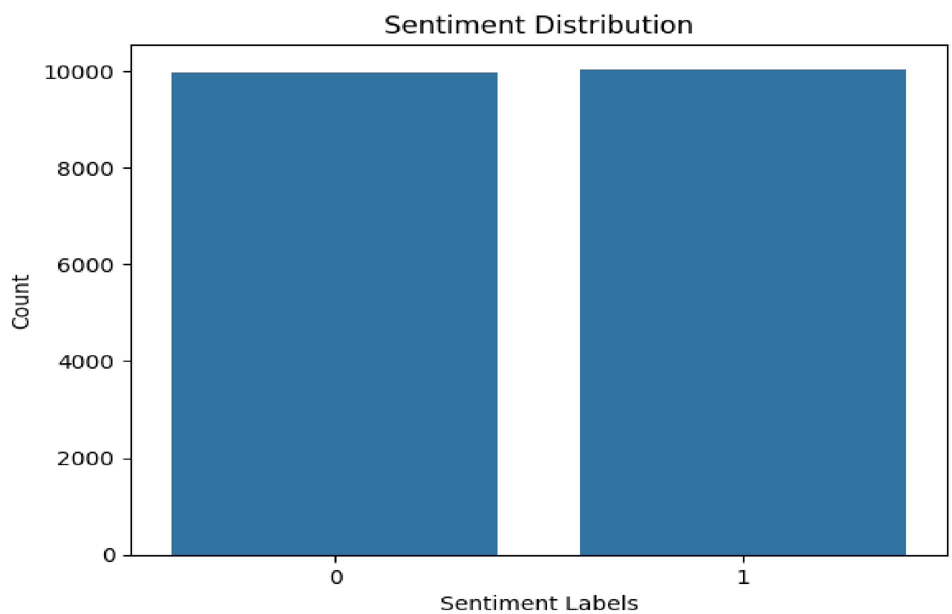


Fig 4.5: Flipkart review dataset (version 02)

4.7. Split the dataset

We splitted the dataset into 0.8 and 0.2, that means 80% for training data and 20% for testing data. Because of that 16,000 data will go for training and 4000 for testing and validation.

4.8. Feature Extraction

After loading the DistilBERT tokenizer we started to extract the feature. We added special tokens like [CLS] and [SEP] , set max review length 512 for amazon dataset's review and 100 for flipkart dataset's review. Set padding to max length and truncation true to truncate the rest if greater than max length. We also used attention mask to fill up the review to max length. We set 16 sample review per batch for 16,000 training data. Then after shuffling the training data , the dataset is ready to fed into the model.

4.9. Loss Function

The loss function quantifies how well the model's predictions match the actual labels. It guides the training process by providing feedback to adjust model weights. We used sparse categorical cross-entropy as loss function.

For sparse categorical cross-entropy, the formula is:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} log(P_{i,}y_i)$$

Where,

- N = Number of samples in the batch
- y_i = True label for the i-th sample
- $P_{i,}y_i$ = Predicted probability for the true class y_i from the model's output.

4.10. Optimization Algorithm

An optimization algorithm minimizes the loss function by adjusting the model's parameters (weights and biases) during training. The Adam optimizer is a popular algorithm that combines the benefits of momentum (for speed) and adaptive learning rates (for efficiency).

Adam updates the model using the gradients' first and second instances. These updates were calculated as:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$$

$$\hat{m_t} = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v_t} = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{\hat{m_t}}{\sqrt{\hat{v_t}} + \epsilon}$$

Where,

g_t = loss gradient

β_1 and β_2= rates of exponential decay

η= LR, and ϵ = small constant.

4.11. Dropout Layer

The dropout layer is a regularization technique used to reduce overfitting by randomly setting a fraction of the neurons to 0 during each training step. This prevents the model from relying too heavily on specific neurons and encourages it to learn more robust and generalizable patterns.

Let:

- h be the input to the dropout layer.
- r be a mask vector where each element $r_i \sim \text{Bernoulli}(1-p)$ (1 with probability $1-p$, 0 with probability p).
- y be the output after dropout.

The dropout formula during training is:

$$y = r \odot h$$

4.12. Training Procedure

After all above setting here comes the training procedure. For three different dataset we set different learning rate, adam optimizer value, epoch value and dropout layer to get maximum performance in minimum time. Hyper parameter tuning of three different dataset are given below.

4.12.1. Amazon Product Review dataset (binary sentiment)

Hyper parameter	**Value**
Learning Rate	0.000002
Epsilon	0.000000001
Hidden Dropout Prob.	0.5
Attention Probability	0.5
Epoch	1

Table 4.1: Tuning Hyperparameter 01

4.12.2. Flipkart Product Review dataset (multi label sentiment)

Hyper parameter	**Value**
Learning Rate	0.000001
Epsilon	0.000000001
Hidden Dropout Prob.	0.4
Attention Probability	0.4
Epoch	2
Step per epoch	1000

Table 4.2: Tuning Hyperparameter 02

4.12.3. Flipkart Product Review dataset (binary sentiment)

Hyper parameter	**Value**
Learning Rate	0.000001
Epsilon	0.000000001
Hidden Dropout Prob.	0.4
Attention Probability	0.4
Epoch	1
Step per epoch	1000

Table 4.3: Tuning Hyperparameter 03

4.13. Summary

Overall, in our methodology section we present a comprehensive approach from collecting dataset to build our model. We begin with a flow diagram and an overall architecture description, followed by insights into data collection and dataset characteristics. Our data preprocessing techniques ensure data quality and compatibility. We offer mathematical explanations for the optimization algorithm, loss function, and dropout layer used in our model. Detailed methodologies for model training, including hyperparameter tuning.

CHAPTER IV

RESULTS AND DISCUSSIONS

5.1. Evaluation Matrix

A matrix known as the confusion matrix is used to assess the performance of the models for classification for a particular set of test data. True and predicted values, as well as the overall number of forecasts, are represented in the matrix. Using this matrix, we perform various calculations for our model.

Classification Accuracy

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

Misclassification Rate

$$Error\ Rate = \frac{FP + FN}{TP + FP + FN + TN}$$

Precision

Out of all the positive predictions the model tried, it measures the proportion of precise predictions it made. In other words, it demonstrates how effective the model was in locating actual positive examples[32]. This can be calculated using the formula:

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

Recall

It is described as the number of accurately predicted positive classes out of all possible positive classes[32]. The recall formula is shown below.

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

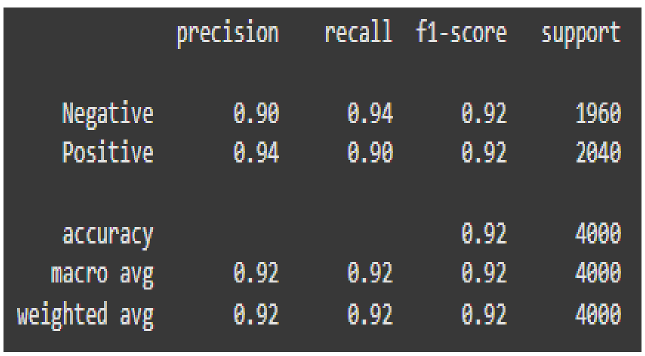
<u>F1-Score</u>

$$F - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

5.2. Evaluating the model

For evaluating the model's performance correctly use we generated classification report, confusion matrix and loss and accuracy curve. For three datasets evaluation report are given below.

5.2.1. Amazon Product Review dataset



	precision	recall	f1-score	support
Negative	0.90	0.94	0.92	1960
Positive	0.94	0.90	0.92	2040
accuracy			0.92	4000
macro avg	0.92	0.92	0.92	4000
weighted avg	0.92	0.92	0.92	4000

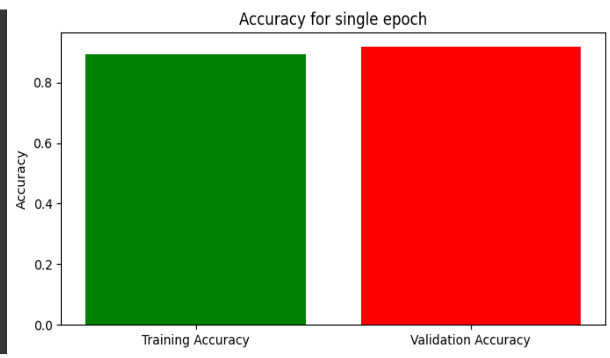
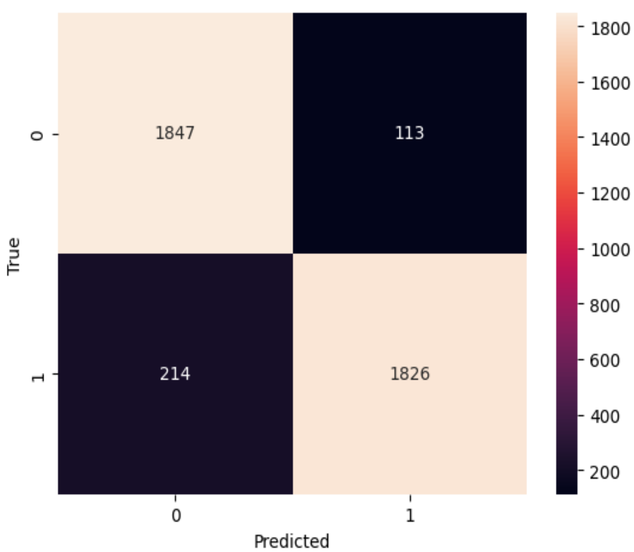


Fig 5.1: Classification Report 01

Fig 5.2: Accuracy Plot



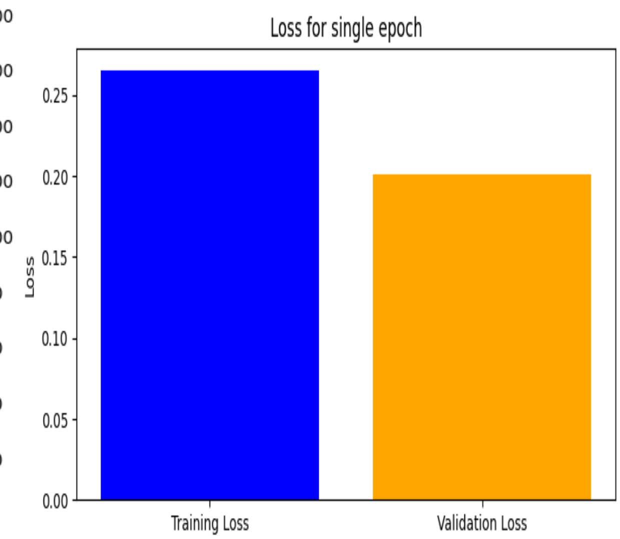


Fig 5.3: Confusion Matrix 01

Fig 5.4: Loss Plot

5.2.2. Flipkart Product Review dataset (multi label sentiment)

	precision	recall	f1-score	support
Negative	0.87	0.88	0.88	1369
Neutral	0.80	0.80	0.80	1301
Positive	0.91	0.90	0.91	1330
accuracy			0.86	4000
macro avg	0.86	0.86	0.86	4000
weighted avg	0.86	0.86	0.86	4000

Fig 5.5: Classification Report 02

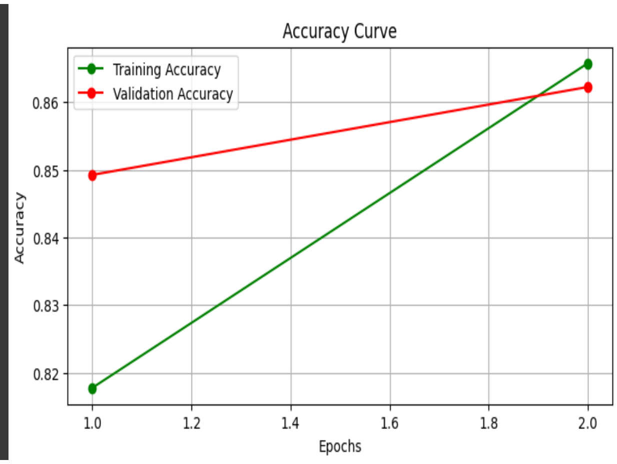


Fig 5.6: Accuracy Plot

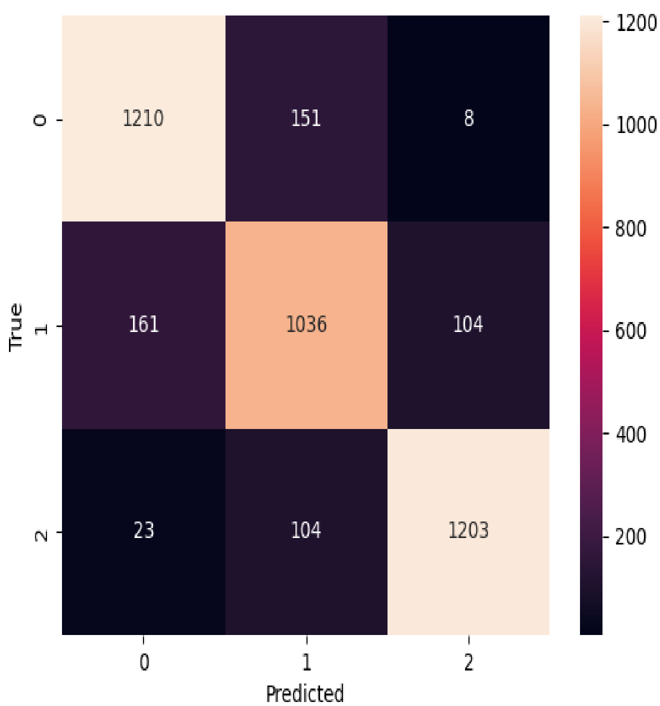


Fig 5.7: Confusion Matrix 02

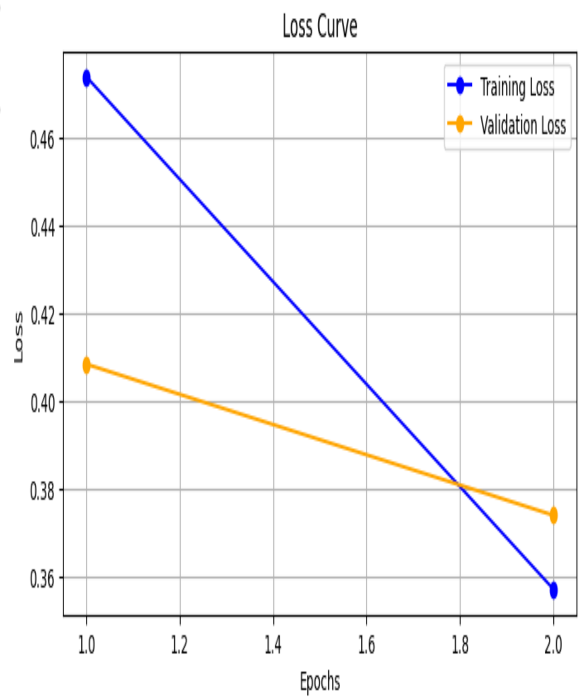


Fig 5.8: Loss Plot

5.2.3. Flipkart Product Review dataset (binary sentiment)

	precision	recall	f1-score	support
Negative	0.96	0.99	0.97	2055
Positive	0.99	0.95	0.97	1945
accuracy			0.97	4000
macro avg	0.97	0.97	0.97	4000
weighted avg	0.97	0.97	0.97	4000

Fig 5.9: Classification Report 03

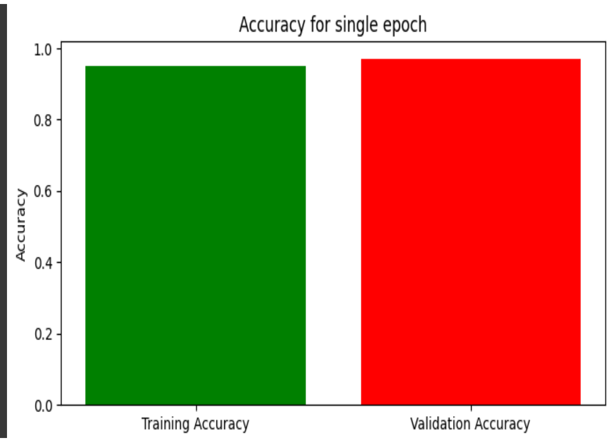


Fig 5.10: Accuracy Plot

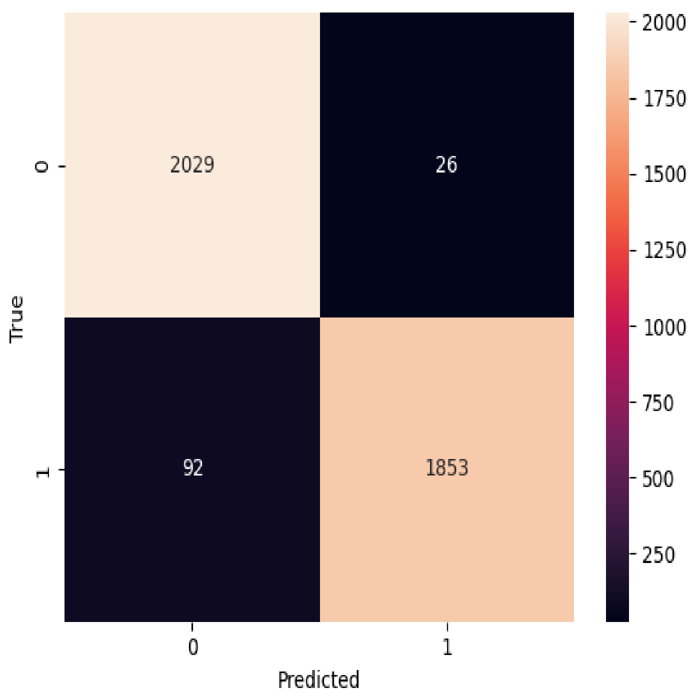


Fig 5.11: Confusion Matrix 03

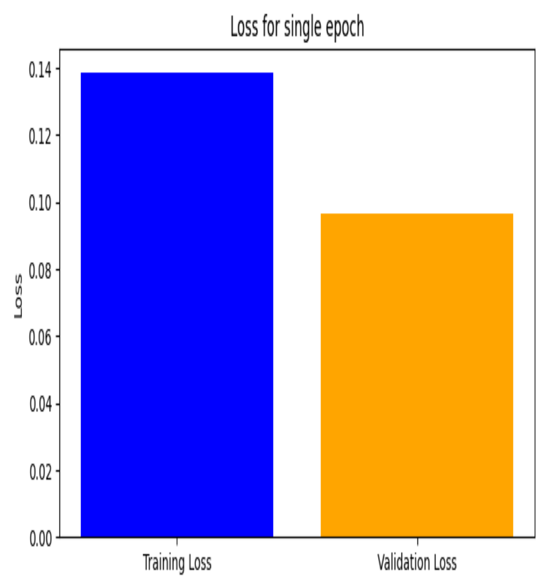


Fig 5.12: Loss Plot

5.3. Discussion and Analysis

We tuned our SA model to get the optimal result. Changed the value of learning rate, dropout layer etc. many times to maximize the training accuracy, validation accuracy and minimize the overfitting. By proper tuning we also got our expected result in very few epoch,

Variants No.	Dataset	Accuracy	Validation Accuracy	F1 Score
1	Amazon dataset (binary sentiment)	88.70%	92.22%	0.92
2	Flipkart dataset (multi label sentiment)	86.71%	86.00%	0.86
3.	Flipkart dataset (binary sentiment)	95.52%	97.15%	0.97

Table 5.1: Result after analysis

CHAPTER VI

CONCLUSION AND FUTURE WORKS

6.1. Conclusion

In summary, our first objective was to build a robust model for E-commerce product review sentiment and classify the sentiment as positive, negative and neutral. We did it by getting 92% accuracy for amazon dataset (binary sentiment), 86% accuracy for flipkart dataset (multi label sentiment) and 97% accuracy for flipkart dataset (binary sentiment). Our second objective was to measure for different sentiment label and different review size how our model perform. Here we found that for multi label sentiment detection our model is less performing than binary sentiment. And we also found that our model is performing very well for smaller review size.

6.2. Future Work

For future work, the sentiment analysis system can be extended to include additional datasets from diverse e-commerce platforms to enhance model generalization. Exploring multilingual support can make the system more versatile for global users. Incorporating fine-grained sentiment categories and analyzing aspects like product features or user satisfaction levels could provide deeper insights. Using more advanced models like BERT-large or fine-tuned domain-specific variants could improve accuracy. Lastly, deploying the model in real-world scenarios and integrating it into recommendation systems or customer service platforms could be a practical extension of this research.

Reference

[1] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis," *arXiv.org*, Mar. 12, 2024. https://arxiv.org/abs/2403.08217

[2] Amira Samy Talaat, "Sentiment analysis classification system using hybrid BERT models," *Journal of Big Data*, vol. 10, no. 1, Jun. 2023, doi: https://doi.org/10.1186/s40537-023-00781-w.

[3] S. Iftikhar, Bandar Alluhaybi, M. Y. Suliman, A. Saeed, and K. Fatima, "Amazon products reviews classification based on machine learning, deep learning methods and BERT," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 21, no. 5, pp. 1084–1084, Oct. 2023, doi: https://doi.org/10.12928/telkomnika.v21i5.24046.

[4] Y. Liu, "Sentiment Analysis for E-commerce Product Reviews by Deep Learning Model of Bert- BiGRU-Softmax," *Journal of Corrosion and Materials*, vol. 48, no. 1, pp. 42–57, Dec. 2024, doi: https://doi.org/10.61336/jcm2023-5.

[5] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," *IEEE Access*, vol. 11, pp. 90367–90382, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3307308.

[6] M. V. Koroteev, "BERT: A Review of Applications in Natural Language Processing and Understanding," *arXiv:2103.11943 [cs]*, Mar. 2021, Available: https://arxiv.org/abs/2103.11943

[7] "Amazon reviews," *www.kaggle.com*. https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews

[8] Nirali vaghani, "Flipkart Product reviews with sentiment Dataset," *Kaggle.com*, 2022. https://www.kaggle.com/datasets/niraliivaghani/flipkart-product-customer-reviews-dataset (accessed Dec. 07, 2024).

[9] "DistilBERT," *huggingface.co*. https://huggingface.co/docs/transformers/en/model_doc/distilbert

[10] B. Muller, "BERT 101 - State Of The Art NLP Model Explained," *huggingface.co*, Mar. 02, 2022. https://huggingface.co/blog/bert-101

[11] Geeksforgeeks, "Neural Networks | A beginners guide," *GeeksforGeeks*, Jan. 17, 2019. https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/

[12] E. Burns, "What is deep learning and how does it work?," *TechTarget*, Mar. 2021. https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network

[13] S. Tiwari, "Activation functions in Neural Networks - GeeksforGeeks," *GeeksforGeeks*, Feb. 06, 2018. https://www.geeksforgeeks.org/activation-functions-neural-networks/

[14] B. Krishnamurthy, "ReLU Activation Function Explained | Built In," *builtin.com*, Oct. 28, 2022. https://builtin.com/machine-learning/relu-activation-function

[15] "Papers with Code - GELU Explained," *paperswithcode.com*. https://paperswithcode.com/method/gelu

[16] M. Ali, "Introduction to Activation Functions in Neural Networks," *Datacamp.com*, Nov. 09, 2023. https://www.datacamp.com/tutorial/introduction-to-activation-functions-in-neural-networks