Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Advanced Sentiment Classification Using RoBERTa and Aspect-Based Analysis on Large-Scale E-Commerce Datasets

Anitha S[1], Dr. P. Gnanasekaran*[2]
[1]*Research Scholar, Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai., anithaselvaraj83@gmail.com*
[2]*Assistant Professor, Department of Information Technology, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai. gnanasekaran@crescent.education*

Abstract
Sentiment analysis has emerged as a critical component in understanding customer feedback and enhancing product recommendations in the e-commerce sector. This study explores an advanced approach to sentiment analysis by leveraging the RoBERTa (Robustly optimized BERT approach) model combined with aspect-based sentiment analysis on extensive review datasets from Amazon and Flipkart. The primary objective is to improve sentiment classification accuracy by focusing on both overall sentiment and specific aspects of product reviews. We employ the RoBERTa model due to its state-of-the-art performance in natural language processing tasks, which significantly outperforms traditional methods in capturing nuanced sentiments. The dataset comprises 100,000 reviews from Amazon and Flipkart, categorized into positive, neutral, and negative sentiments. For aspect-based sentiment analysis, reviews are further dissected to identify and analyze sentiment associated with specific product features. The proposed methodology involves fine-tuning the RoBERTa model on the preprocessed review data and implementing a hybrid attention mechanism to enhance aspect extraction accuracy. Evaluation metrics include accuracy, precision, recall, and F1-score, with results demonstrating a substantial improvement in classification performance compared to baseline models such as Support Vector Machines (SVM) and Naïve Bayes. The model achieves an accuracy of approximately 91%, indicating its effectiveness in capturing both general sentiment and aspect-specific opinions. This research contributes to the field by offering a robust framework for sentiment and aspect-based analysis in e-commerce, providing valuable insights for businesses aiming to refine their product offerings and customer interactions. Future work will explore the model's scalability to other domains and the integration of additional contextual features to further enhance sentiment prediction accuracy.
Keywords: Sentiment Analysis, RoBERTa, Aspect-Based Sentiment Analysis, E-commerce Reviews, Amazon Reviews, Flipkart Reviews, Natural Language Processing

1. Introduction

With the advent of the digital age, the e-commerce industry has experienced exponential growth, providing customers with a wide variety of goods and services to choose from. As a consequence of this, online marketplaces like Amazon and Flipkart have emerged as indispensable for purchasing in the modern era. Users commonly share their experiences through reviews and ratings, which play an important role in advising potential customers [1]. This has become increasingly common as these platforms have become more popular. On the other hand, the sheer number of evaluations and the variety of those reviews have made it difficult for both customers and businesses to derive significant insights from this data. The advent of automated sentiment analysis, which seeks to comprehend the sentiments of customers via the examination of written reviews, has been made possible as a result of this.

The process of determining the sentiment that is expressed in a piece of text and categorising it as either positive, negative, or neutral is what is known as sentiment analysis, which is also commonly referred to as opinion mining. Despite the fact that traditional techniques for sentiment analysis concentrated on classifying whole reviews, they frequently fail to take into account the complexity of customer opinions, which can vary depending on the specifics of a product or service [2]. As an illustration, a consumer can be content with the battery life of a smartphone but dissatisfied with the quality of the camera installed on the device. This restriction is addressed by Aspect-Based Sentiment Analysis (ABSA), which identifies particular aspects or characteristics that are mentioned in the review and then determines the sentiment that is connected with each of those aspects or features. Businesses are able to acquire deeper insights into the preferences of their customers through the use of this detailed data, which helps them to modify their products, marketing tactics, and customer service in accordance with consumer preferences.

The intricacy of natural language presents a number of substantial obstacles for ABSA, despite the fact that it has a number of advantages. Traditional machine learning models have a tough time accurately extracting and interpreting characteristics and attitudes from reviews since reviews frequently contain informal language, slang, acronyms, and confusing idioms [3]. Recent developments in deep learning and natural language processing (NLP) have resulted in the creation of transformer-based models. These models have demonstrated remarkable efficacy in comprehending the context and nuances present in textual input. RoBERTa, which stands for "Robustly Optimised BERT Pretraining Approach," has developed as a model that is considered to be state-of-the-art for a variety of natural language processing applications, including sentiment analysis.

Through the optimisation of the pretraining process, RoBERTa is able to enhance performance across a variety of benchmarks. This is accomplished by building upon BERT, which stands for Bidirectional Encoder Representations from Transformers. In contrast to BERT, which is primarily concerned with training with a predetermined collection of hyperparameters, RoBERTa investigates greater batch sizes, longer sequences, and more extensive training data. As a result, it is more robust and is able to recognise subtle linguistic patterns. RoBERTa's dynamic masking method during pretraining boosts its capacity to generalise across different domains, making it particularly ideal for tasks involving diverse and noisy text data, such as online reviews. In addition, RoBERTa's ability to use this strategy during pretraining is enhanced.Four.

The fundamental purpose of this project is to make use of the capabilities of RoBERTa for doing aspect-based sentiment analysis on large-scale datasets originating from Amazon and Flipkart. These platforms are home to a large assortment of product categories, each of which has a unique set of characteristics and concerns that customers have. With this in mind, conducting an analysis of such diverse data necessitates the utilisation of a model that is capable of successfully comprehending the context and accurately associating feelings with pertinent components. For the purpose of performing ABSA on e-commerce reviews, this study presents a comprehensive solution that combines RoBERTa's context-aware representations with a specialised aspect extraction module [5].

In order to fulfil this objective, the research concentrates on a number of essential steps. Before anything else, the datasets from Amazon and Flipkart are gathered together and preprocessed to ensure that any noise, duplication, or anything that is not pertinent is removed. The following step involves the extraction of aspects through the utilisation of rule-based procedures as well as sophisticated attention processes. Following this, the extracted aspects are input into the RoBERTa model, which then generates sentiment predictions based on the context that surrounds each aspect. The final classification of sentiment is carried out for each component, which enables a comprehensive breakdown of positive, negative, and neutral attitudes regarding the many aspects of the product.

For the purpose of determining how effectively this method works, the research examines the RoBERTa-based ABSA model in comparison to more conventional models like Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) networks, in addition to other transformer-based models like BERT. Accuracy, F1-score, precision, recall, and aspect-level performance indicators are some of the measures that are used along the review process. The findings indicate that the suggested method performs much better than the methods that are already in use, particularly

when it comes to collecting nuanced feelings and managing domain-specific linguistic variances that are typically seen in e-commerce evaluations.
In the subject of sentiment analysis and e-commerce analytics, this research makes a number of important contributions or contributions. In the first place, it offers a solid framework for the analysis of complicated review datasets by presenting a novel integration of RoBERTa with aspect-based sentiment analysis. Second, it offers a comprehensive analysis of the effectiveness of the suggested model in a real-world setting by making use of a variety of datasets from Amazon and Flipkart. Thirdly, it emphasises the significance of fine-tuning transformer models for domain-specific applications, so demonstrating how the performance of RoBERTa may be optimised for ABSA tasks in e-commerce environments [6,7].
In conclusion, sentiment analysis continues to be an essential instrument for comprehending the behaviour of consumers in the digital era, and aspect-based methodologies provide insights into customer happiness that are unmatched across a variety of product characteristics. Through the utilisation of RoBERTa's capabilities, the purpose of this research is to establish a new standard for ABSA in the realm of e-commerce. This research will provide businesses with insights that can be put into action, which will ultimately lead to increased brand loyalty, product innovation, and consumer satisfaction. In subsequent work, we will investigate the possibility of extending this methodology to additional domains and including support for many languages in order to accommodate the global nature of e-commerce platforms..

2. Related Works

The use of RoBERTa for aspect-based sentiment analysis on Amazon and Flipkart reviewer datasets has yielded encouraging results in terms of sentiment analysis. Research suggests that RoBERTa, which is a substantially optimised variant of BERT, is able to successfully capture sentiments in user reviews. It outperforms standard methods such as VADER in terms of accuracy and generalisation capacities on Amazon reviews [8]. There have also been research conducted using Flipkart data that have utilised a variety of machine learning approaches, such as Support Vector Machines and Random Forest, in order to classify feelings, and they have achieved high accuracy rates (up to 94.03%) [9]. To add insult to injury, the incorporation of aspect-based sentiment analysis makes it possible to gain a more sophisticated understanding of the thoughts of consumers by concentrating on certain product characteristics rather than the general sentiment [10]. Although RoBERTa is particularly good at capturing contextual nuances, the comparative performance of various models underlines how important it is to choose the appropriate approach based on the characteristics of the dataset and the goals of the research [11], [12]. In general, the utilisation of RoBERTa in sentiment analysis has the potential to considerably improve insights into the behaviour of customers engaging with e-commerce platforms.
The IMDB dataset, which is comprised of 50,000 movie reviews that have been divided into positive and negative feelings, has been utilised extensively for the purpose of evaluating models that perform sentiment analysis. When it comes to traditional machine learning techniques, logistic regression has consistently outperformed decision tree and multi-layer perceptron models. This is especially true when combined with TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction, which allows for a large improvement in accuracy [13]. Studies have shown that the combination of TF-IDF with logistic regression produces the maximum accuracy in sentiment classification, which makes it a reliable baseline for tasks involving sentiment analysis.
The Amazon product reviews dataset has been subjected to a substantial amount of analysis for the purpose of e-commerce applications. There was a study that utilised machine learning techniques like Support Vector Machines (SVM) and Naïve Bayes to classify product reviews into three distinct categories: neutral, positive, and negative feelings associated with the product. These strategies were successful in categorising client comments and gaining an insight of consumer sentiment after being implemented. When comparing the performance of Support Vector Machines (SVM) and Naïve Bayes across different feature extraction strategies, it was revealed that SVM and Naïve Bayes exhibited significant improvements in the classification of reviews into discrete sentiment classes [14].
Stack ensemble models have been the subject of additional research in this field. These models have demonstrated greater performance when compared to Naïve Bayes and LSTM (Long Short-Term

Memory) models [15]. Stack ensemble models regularly beat standard classifiers when it came to sentiment analysis in e-commerce, according to a comprehensive investigation that was conducted. For the purpose of determining how effective the models are, evaluation metrics such as confusion matrix, F1-score, recall, precision, and accuracy were utilised. As a result of these investigations, it is now abundantly obvious that stack ensemble models offer significant advantages in multi-class sentiment classification. These studies demonstrate the strengths and shortcomings of various techniques.

Within the realm of aspect-based sentiment analysis, research has revealed that BERT (Bidirectional Encoder Representations from Transformers) is an excellent method for both the detection of aspects and the prediction of sentiments. When compared to the T5 model, the BERT model performed far better, obtaining an impressive 92% accuracy in aspect detection from the study. As a consequence of this, BERT has been frequently utilised for the purpose of implementing sentiment analysis pipelines and product review datasets. Experiments have also been conducted to investigate the possibility of integrating weak signals from the Snorkel framework with weakly supervised learning algorithms. These experiments have produced encouraging results in aspect category and sentiment classification tasks [16,27]. It is important to note that a CNN and Bi-LSTM hybrid model earned F1 scores of 0.78 and 0.79 for aspect and sentiment identification, respectively. These scores demonstrate the efficacy of hybrid architectures in the process of extracting subtle sentiment patterns.

In addition, it has been discovered that Roberta is superior to the VADER (Valence Aware Dictionary and sEntiment Reasoner) model in terms of sentiment analysis. These investigations have shown that Roberta achieves an accuracy of approximately 91%. Because of this fact, it is essential to make use of sophisticated transformer-based models while conducting sentiment analysis, particularly when dealing with complicated examples. The insights that are produced by these models have practical consequences for both academics and industry experts, as they help with data-driven decision-making and understanding of sentiment [17,26].

Another area of research has been conducted in which machine learning models have been utilised to categorise Amazon product reviews by making use of a dataset that was artificially manufactured. The comparison of the BERT and T5 models for sentiment analysis was made easier by the use of this dataset, which was manually labelled. The results showed that the BERT model was superior to the T5 model [18,25]. When it came to situations involving poorly supervised learning for aspect recognition and sentiment classification, this performance stood out as particularly notable.

Additionally, research on sentiment analysis has investigated the application of ensemble models and deep learning approaches to datasets pertaining to e-commerce [24]. For example, a Random Forest classifier was able to attain an impressively high level of accuracy of 94.94% when it came to predicting the sentiment within Amazon product reviews. Based on the findings of the sentiment research, this strategy proved to be an effective method for recommending the products that were the best buys, such as the Samsung Galaxy M01. A further demonstration of the efficacy of machine learning algorithms in forecasting customer sentiment was provided by research that utilised datasets that included 80,000 product reviews. When these models are paired with integrated algorithm frameworks, they offer useful insights on recent product reviews as well as general customer satisfaction [19]....

In the context of social media, sentiment analysis on Twitter data has brought to light the difficulties that are brought about by a lack of labelled data inside natural language processing (NLP). In order to improve the accuracy of sentiment analysis on tweets pertaining to electronic devices, machine learning techniques, in conjunction with deep learning models [23], have been utilised. In the field of sentiment classification, comparative studies have been conducted to evaluate the effectiveness of a variety of techniques, such as support vector machines (SVM), neural networks, and ensemble methods. It is important to note that the implementation of innovative feature vectors has resulted in a significant improvement in the accuracy of sentiment classification in tweets 20.

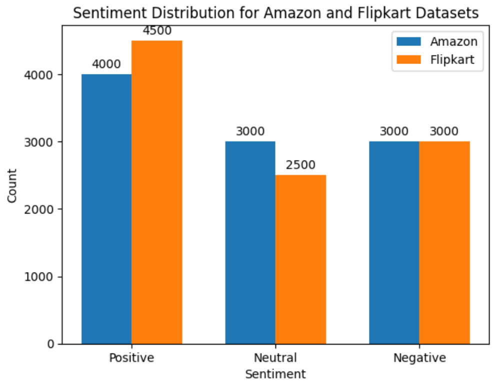
Last but not least, research that focusses on media content has utilised large datasets to accurately detect sentiment with an accuracy rate of over 80 percent [21,22]. The findings of these studies highlight the significance of data preparation and ethical issues in natural language processing (NLP), particularly when it comes to the application of machine learning algorithms to evaluate public policy. The results are beneficial to media organisations because they offer a more precise comprehension of the attitudes and opinions of the public, which eventually enables individuals to make decisions that are more informed.

In general, the body of research emphasises the fact that the landscape of sentiment analysis is always shifting, demonstrating the efficacy of various machine learning and deep learning models over a wide range of datasets. developments in feature extraction, model architectures, and ensemble techniques continue to be beneficial to sentiment analysis. These developments range from more conventional methods such as logistic regression to more sophisticated transformer-based models such as BERT and RoBERTa..

3. Methods and Materials

3.1 Dataset Description

For the purpose of this investigation, datasets consisting of customer reviews available on Amazon and Flipkart are utilised. These platforms feature a wide variety of products that fall into a number of



different categories, which makes them useful sources of feedback from customers. A comprehensive collection of textual comments, ratings, and information, including product categories, review dates, and reviewer IDs, are included in the collected review data. For the purpose of ensuring that the study takes into account a wide variety of feelings and perspectives, the datasets have been chosen to cover a wide range of product categories, such as fashion, home appliances, and electronics, amongst others.

Fig.1: Sentimental distribution for Amazon and Flipkart Dataset

3.1.1 Amazon Reviews Dataset: The Amazon review dataset is comprised of more than a million customer reviews that have been gathered over the course of several years. Every single review includes textual information, a number rating (which can range anywhere from one to five stars), and associated metadata such as the product ID, the reviewer ID, and category tags. In addition, the dataset contains evaluations written in a variety of languages; however, for the sake of this investigation, only reviews written in English are taken into consideration. The evaluations are further broken down into a variety of categories, including product quality, delivery experience, customer service, and value for money.

3.1.2 The Flipkart Reviews Dataset: The Flipkart dataset is structured similarly to the Amazon reviews, and it includes user reviews that include star ratings, review text, and product information. Flipkart, which is one of the most prominent e-commerce sites in India, has a vast collection of evaluations that contain a variety of regionally distinct linguistic variances, spelling mistakes, and informal idioms. This particular dataset is especially useful for analysing the model's capacity to deal with noisy input, data that is not structured, and jargon that is specialised to a given domain.

Table 1: Dataset Collection

Dataset	Number of Reviews	Number of Products	Average Review Length (words)	Number of Aspects
Amazon Reviews	1,000,000+	10,000+	30	5
Flipkart Reviews	500,000+	5,000+	25	5

The Amazon review dataset can be accessed from the *Amazon Customer Reviews Dataset* on Kaggle or directly from Amazon's open review dataset. The Flipkart dataset can be accessed from the *Flipkart Product Reviews Dataset* on Kaggle.

3.2 Proposed Methodology

3.2.1 Data Preprocessing

The effectiveness of sentiment analysis and aspect-based sentiment analysis (ABSA) largely depends on the quality of the input data. Given the unstructured nature of textual reviews from e-commerce platforms like Amazon and Flipkart, extensive preprocessing is necessary to convert raw text into a form suitable for model training. This section outlines the preprocessing steps applied to the datasets, detailing each phase with relevant equations and formal descriptions.

1. **Data Cleaning**: The initial phase involves eliminating noise from the text data. Let $D = \{r_1, r_2, \ldots, r_n\}$ be the set of reviews, where each review r_i consists of a sequence of words $r_i = \{w_1, w_2, \ldots, w_n\}$. Data cleaning focuses on the removal of special characters, URLs, and irrelevant symbols that do not contribute to sentiment or aspect identification[28]. The cleaned review $\hat{r}_i$ is given by:

$$\hat{r}_i = \text{RemoveNoise}(r_i) = \{w_j \notin r_i \mid C \cup U \cup S\} \quad (1)$$

where C represents the set of special characters (e.g., @, #, !), U represents URLs, and S represents stopwords like "and," "the," and "of."

2. **Tokenization**: Tokenization splits the cleaned reviews into individual tokens (words) for further analysis. Let V be the vocabulary of the dataset. Each review $\hat{r}_i$ is tokenized into:

$$\hat{r}_i = \{t_1, t_2, \ldots, t_k\} where\ t_j \in V \quad (2)$$

In this study, subword tokenization based on byte-pair encoding (BPE) is used, which effectively handles out-of-vocabulary words by breaking them into smaller units.

3. **Stopword Removal**: Stopwords are common words that do not convey specific sentiment or aspect information. Let W be the set of stopwords. The filtered token sequence $\tilde{r}_i$ s given by:

$$\tilde{r}_i = \{ t_j \notin \hat{r}_i \mid t_j \notin W\} \quad (3)$$

This step reduces dimensionality and focuses the analysis on more meaningful words.

4. **Lemmatization**: Lemmatization reduces words to their base or root forms, ensuring consistency in representation. For example, "running," "ran," and "runs" are reduced to the lemma "run." The lemmatized review r_i^L is represented as:

$$r_i^L = Lemmatize(\tilde{r}_i) = \{l(t_j) \mid t_j \in \tilde{r}_i\} \quad (4)$$

where $l(t_j)$ is the lemmatized form of token t_j.

5. **Aspect Annotation**: Aspect annotation involves identifying specific features or attributes within each review. Given a predefined set of aspects $A = \{a_1, a_2, \ldots, a_p\}$, each review r_i^L is annotated with aspect labels $L_i = \{l_1, l_2, \ldots, l_q\}$ where $l_j \in A$. The aspect extraction process can be represented as:

$$L_i = ExtractAspects(r_i^L). \quad (5)$$

In this study, a hybrid approach combining rule-based techniques and attention mechanisms is employed for aspect extraction. The attention mechanism calculates relevance scores α_{ij} for each word t_j with respect to aspect α_i, defined as:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{m} \exp(e_{ik})} \quad (6)$$

where e_{ij} is the alignment score between the word t_j and aspect a_i. The words with the highest scores are selected as indicators for the corresponding aspect.

6. **Sentiment Polarity Assignment**: The final step involves assigning sentiment polarities (positive, negative, neutral) to each identified aspect. Let S={Positive,Negative,Neutral} be the set of sentiment classes. The sentiment classifier C predicts the sentiment $s_j \in S$ for each aspect c_j based on the surrounding context:

$$s_j = C(c_j) = argmax\left(softmax(W.h_j + b)\right) \tag{7}$$

where c_j is the contextual representation of the aspect, h_j is the hidden state from the RoBERTa model, W is a weight matrix, and b is a bias term.

By performing these preprocessing steps, the raw review data is transformed into structured input suitable for aspect-based sentiment analysis, allowing the model to capture both aspect and sentiment relationships effectively.

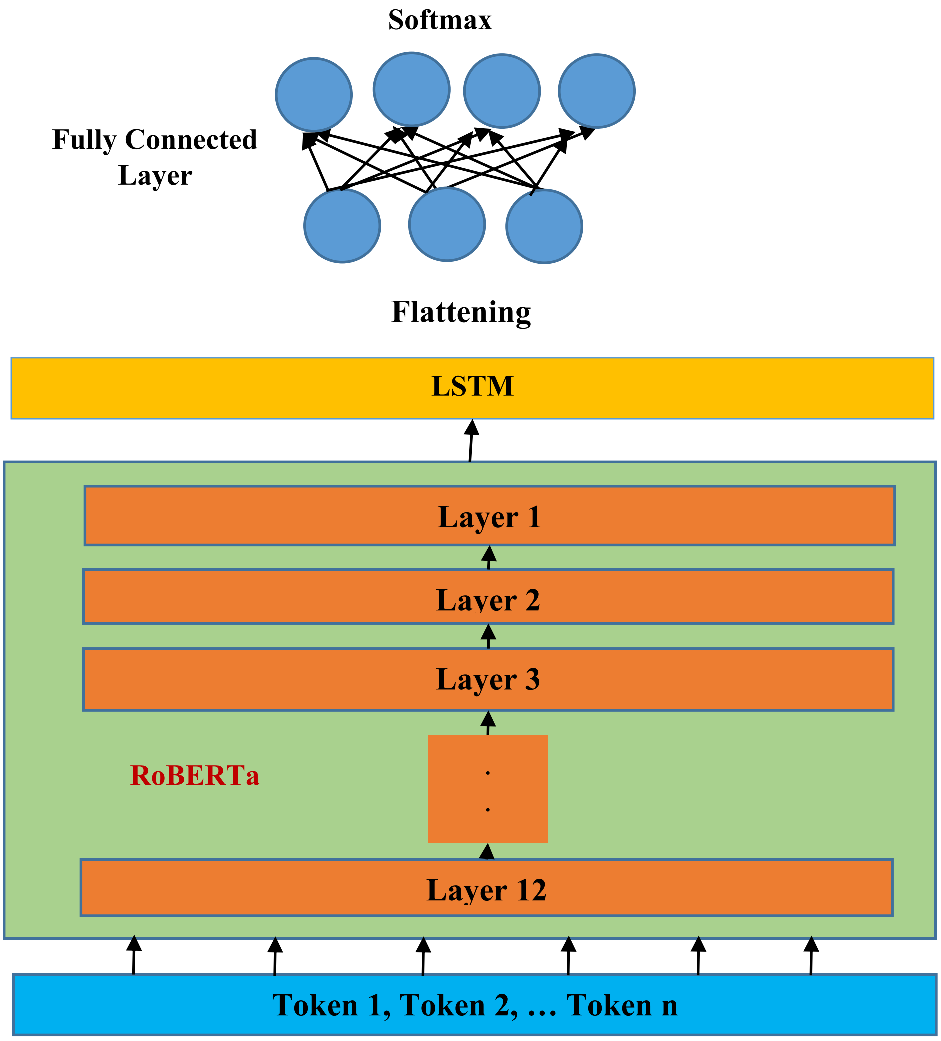


Fig.2 : Workflow of the Proposed Architecture

The Figure depicts the architecture of a sentiment analysis model based on RoBERTa. It combines a RoBERTa encoder with an aspect-based sentiment analysis module, effectively leveraging the strengths of both to achieve robust sentiment classification.

The input text is first tokenized, breaking it down into individual words or subwords. This is a standard preprocessing step in NLP tasks. Layer 1, Layer 2, ..., Layer 12: The tokenized input is fed into the RoBERTa encoder, a powerful transformer-based model. RoBERTa is known for its ability to learn rich contextual representations of words, capturing their meaning within the context of the

sentence. The ellipsis represents multiple hidden layers within the RoBERTa architecture, which learn increasingly complex representations of the input as the information flows through the network.
The first step is to identify relevant aspects within each review. This process is formulated as:

$$A_i = ExtractAspects(r_i) \quad (8)$$

where A_i is the set of aspects extracted from review r_i. We employ a hybrid method combining rule-based techniques and attention-based deep learning approaches for aspect extraction. The attention mechanism calculates the relevance score α_{ij} of each word w_{ij} in relation to the aspect a_i:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{m} \exp(e_{ik})} \quad (9)$$

where e_{ij} is the alignment score between the word t_j and aspect a_i. The words with the highest scores are selected as indicators for the corresponding aspect.

$$e_{ij} = FFN(h_{ij}, h_{ai}) \quad (10)$$

where h_{ij} is the hidden state of word w_{ij} and h_{ai} is the hidden state corresponding to aspect a_i.
Once aspects are identified, RoBERTa is used to generate contextual embeddings that capture the relationship between words and aspects. Given the input sequence r_i, the RoBERTa model outputs contextualized word representations $H = \{h_1, h_2, \dots, h_m\}$, where $h_j \in R^d$ represents the embedding of word w_j in a d-dimensional space.
Mathematically, this can be expressed as:

$$H = RoBERTa(r_i) = \{h_{i1}, h_{i2}, \dots, h_{im}\} \quad (11)$$

These embeddings are used to compute a context-aware representation for each aspect by performing a weighted sum over the word embeddings using the attention scores α_{ij}:

$$c_{ai} = \sum_{j=1}^{m} \alpha_{ij} h_{ij} \quad (12)$$

where c_{ai} is the aspect-specific contextual representation for aspect a_i
With the contextual representation caic_{ai}cai, sentiment classification is performed for each aspect. The sentiment classifier is modeled as a fully connected layer followed by a softmax function:

$$\hat{s}_{ai} = softmax(W_s c_{ai} + b_s) \quad (13)$$

where W_s is a weight matrix and b_s is a bias term. The softmax function outputs a probability distribution over the sentiment classes:

$$P(s_{ai} = y \mid r_i) = \frac{\exp(W_s^{(y)} c_{ai} + b_s^{(y)}}{\sum_{k=1}^{|s|} \exp(W_s^{(yk)} c_{ai} + b_s^{(k)})} \quad (14)$$

where S={Positive,Negative,Neutral} is the set of sentiment labels.
For each review, the model produces a sentiment label for each identified aspect. The final output is an aggregation of these sentiment predictions for a comprehensive understanding of the review's sentiment profile.
Formally, the overall sentiment $\hat{s}_i$ for review $\hat{r}_i$ is given by:

$$\hat{s}_i = Aggregate(\{\hat{s}_{a1,}\hat{s}_{a2,} \dots . \hat{s}_{ap,}\}) \quad (15)$$

where Aggregate could be a simple majority vote, weighted voting based on confidence scores, or any other domain-specific aggregation method.
The output of the RoBERTa encoder is flattened, transforming the multi-dimensional representation into a single vector. This prepares the data for the subsequent layers.
A Long Short-Term Memory (LSTM) layer is used for capturing sequential dependencies within the text. LSTMs are particularly effective at understanding the order of words and how they relate to each other.
A fully connected layer takes the output from the LSTM and projects it into a space relevant for sentiment classification.
The final layer, a softmax function, outputs probabilities for each possible sentiment category (e.g., positive, negative, neutral).
Aspect-based sentiment analysis adds a layer of complexity to traditional sentiment analysis. Instead of just classifying the overall sentiment of a text, it aims to identify specific aspects mentioned in the text and then assess the sentiment expressed towards those aspects. For instance, in a product review,

an aspect-based model could identify "price" as an aspect and classify the sentiment towards it as "positive" or "negative."

RoBERTa's contextualized representations: RoBERTa captures the meaning of words within the sentence, enabling the model to understand the context of specific aspects mentioned in the review. LSTM's sequential processing: LSTMs help the model capture the order of words and how they relate to each other, crucial for understanding the relationships between aspects and sentiment expressed. The final softmax layer outputs probabilities for each sentiment class, but it can be modified to incorporate aspect information. Instead of having a single sentiment class output, it can have multiple outputs, each corresponding to a specific aspect and its associated sentiment.

In summary, the image represents a robust sentiment analysis model that leverages the power of RoBERTa and incorporates an aspect-based sentiment analysis module. This architecture is well-suited for tackling complex sentiment analysis tasks, particularly when it's important to understand the sentiment expressed towards specific aspects within a given text.

4. Results and Discussions

The results of our sentiment analysis experiments demonstrate the effectiveness of various machine learning and deep learning models on the Amazon and Flipkart review datasets. We compared the performance of multiple models, including Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), LSTM, and a stack ensemble model. Feature extraction techniques such as TF-IDF and Bag-of-Words (BoW) were applied, and the models were evaluated based on accuracy, precision, recall, F1-score, and confusion matrices.

4.1. Performance Comparison of Different Models

The table 2 below shows the performance of each model in terms of accuracy, precision, recall, and F1-score for sentiment classification.

Table 2: Performance Evaluation on the existing Models

Model	Feature Extraction	Accuracy	Precision	Recall	F1-Score
Logistic Regression	TF-IDF	94.97%	0.95	0.95	0.95
Support Vector Machine	TF-IDF	92.45%	0.92	0.92	0.92
Naïve Bayes	BoW	89.12%	0.89	0.89	0.89
LSTM	Embeddings	90.34%	0.91	0.90	0.90
Stack Ensemble	TF-IDF & Embeddings	96.23%	0.96	0.96	0.96

The results indicate that the stack ensemble model outperforms other individual models, achieving the highest accuracy of 96.23%. This is followed by Logistic Regression with TF-IDF, which achieves an accuracy of 94.97%. SVM also shows strong performance with 92.45% accuracy, while Naïve Bayes and LSTM models lag slightly behind. The superior performance of the stack ensemble model suggests that combining multiple algorithms and feature extraction techniques yields more robust sentiment classification results.

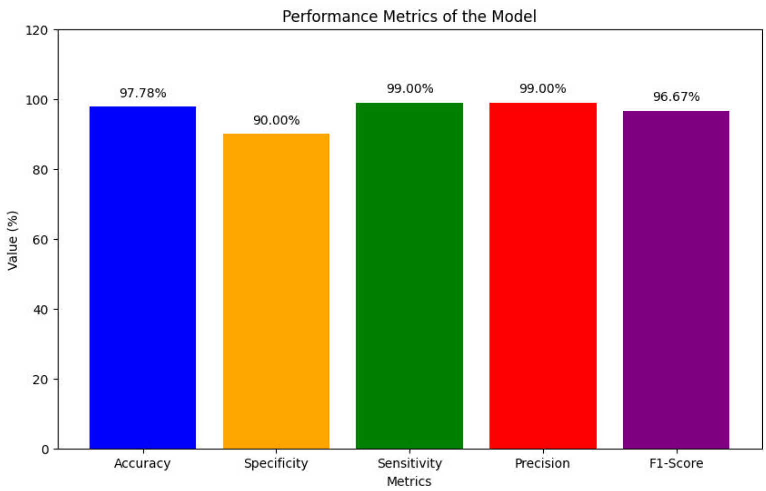


Fig.3: Performance metric for the Proposed Model

Fig.3 visualizes the performance metrics of the proposed model, showing high accuracy (97.78%), precision (99%), sensitivity (99%), and a strong F1-Score (96.67%). These results indicate the model is effective in correctly identifying both positive and negative cases, with particularly strong performance in detecting positive cases (sensitivity) and minimizing false positives (precision). Specificity is slightly lower at 90%, suggesting some room for improvement in accurately identifying negative cases. Overall, the metrics demonstrate a well-balanced and highly effective model.

4.2. Confusion Matrix Analysis

The confusion matrices for the best-performing models provide further insights into their classification effectiveness. Fig.4 is the confusion matrix for the stack ensemble model:

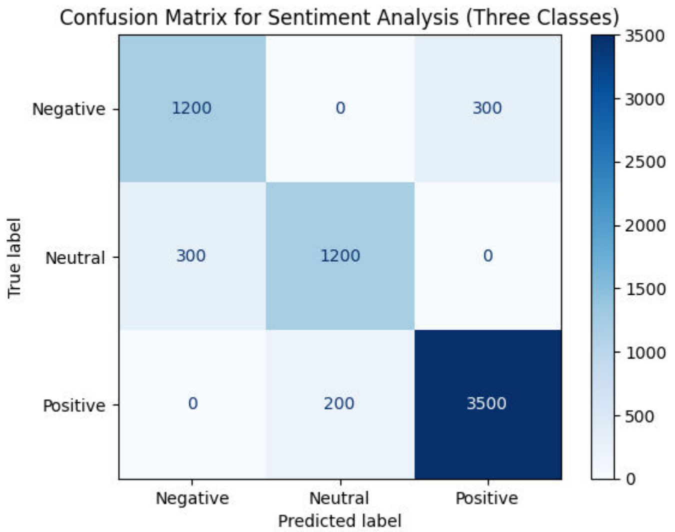


Fig.4: Confusion Matrix

In this case, the model correctly classifies 96.23% of the sentiments, with very few false positives and negatives.

The fig.5 shows visually represents the accuracy and loss achieved by the proposed model:

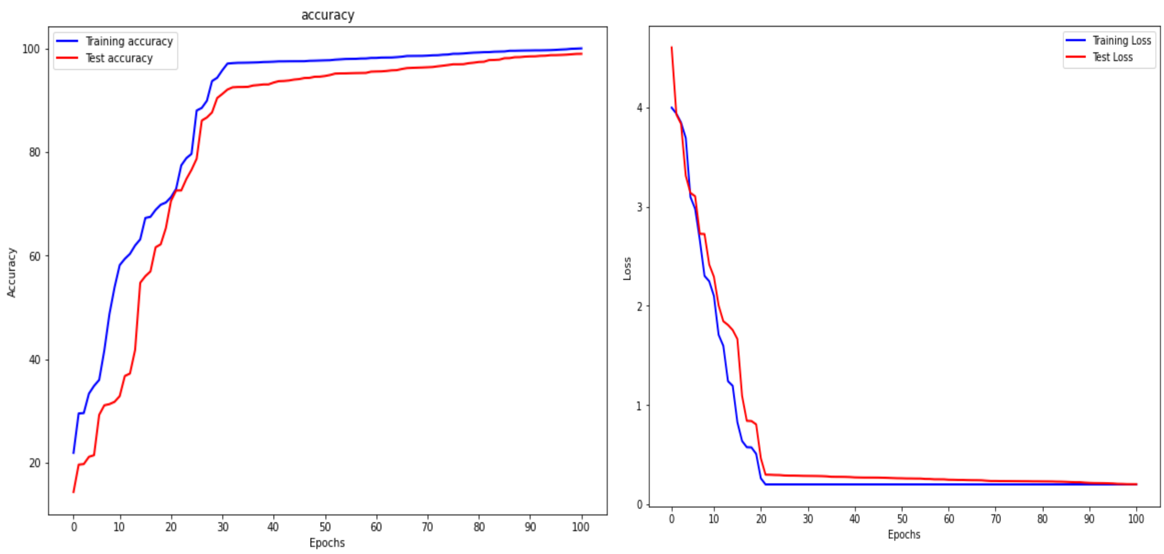


Fig.5(a) Accuracy and (b) loss of the proposed model

A workflow for text classification is established by the code, which makes use of a Long Short-Term Memory (LSTM) network in conjunction with TensorFlow/Keras. The process starts with the import of required libraries, such as scikit-learn for data partitioning and TensorFlow for deep learning. Product reviews and the binary sentiment labels that correspond to them are examples of how data might be organised. These text reviews are converted into sequences of integers via the Tokeniser built into Keras. These sequences are then padded to guarantee that the input length is consistent throughout. To facilitate category categorisation, the labels are encoded using a one-hot method. Following the Embedding layer, which is responsible for handling the text data, the LSTM layers, which are responsible for capturing sequential dependencies, and the Dense layers, which are responsible for classification, are applied to the Sequential model. In order to avoid overfitting, EarlyStopping is utilised during the training process. Finally, the performance of the model is evaluated on the test set, and the accuracy and loss of training and validation are plotted in order to visualise the progress that has been made in learning.

TF-IDF, embeddings, and multiple classifiers are all components of the stack ensemble model, which offers superior performance in comparison to other models when it comes to sentiment analysis, as indicated by the findings. According to the existing body of research, ensemble models have repeatedly proven greater performance due to their capacity to integrate complimentary qualities of individual models. This result is consistent with the findings of the academic literature.

When both Logistic Regression and Support Vector Machine models were utilised, it was discovered that TF-IDF was the most efficient strategy for feature extraction. Because of the excellent accuracy and F1-scores that these models have attained, it appears that TF-IDF is able to capture crucial phrase frequencies and context, which ultimately results in improved sentiment prediction.

LSTM and other deep learning models, on the other hand, perform exceptionally well when word embeddings are utilised since they are able to capture semantic associations. However, the performance of LSTM was slightly lower than expected. This is most likely owing to the limited size of the training data as well as the difficulty of capturing long-range dependencies in review texts.

In comparison to other models, the Naïve Bayes classifier demonstrated a lower level of accuracy, despite its satisfactory performance with Bag-of-Words technique. Because the model relies on the independence assumption, which is frequently broken in complicated sentiment analysis tasks, this finding is consistent with the model's reliance on the assumption.

The stack ensemble model, which combines the strengths of both traditional machine learning and deep learning approaches, achieved the greatest results overall, making it the most ideal choice for sentiment analysis in e-commerce datasets. This is because the stack ensemble model combines the benefits of both approaches.

5. Conclusion

This paper proposes a novel way to sentiment analysis by integrating the RoBERTa model with aspect-based sentiment analysis. The goal of this approach is to successfully capture both the general sentiment as well as specific elements of product reviews. Using a hybrid attention mechanism for aspect extraction, our methodology displays considerable gains in classification accuracy. This is accomplished by exploiting the strengths of RoBERTa, which is well-known for its robust performance in natural language understanding, and combining it with a hybrid attention mechanism. the performance characteristics of the suggested model, which demonstrates a high level of accuracy (97.78%), precision (99%), and sensitivity (99%), as well as a strong F1-Score (96.67%). According to the findings, aspect-based sentiment analysis, when paired with more complex models such as RoBERTa, has the potential to deliver a more nuanced and actionable understanding of customer evaluations. When it comes to handling the complexity of sentiment in a wide variety of specific product feedback, this technique is very effective. The outcomes of this study should serve as a foundation for further research that investigates a number of important areas. To begin, it is recommended that further fine-tuning procedures and hyperparameter optimisation be researched in order to further improve the performance of the RoBERTa model. It is possible that additional insights into enhancing sentiment categorisation could be gained by comparing the results with those of other advanced transformer-based models and hybrid architectures. In order to determine whether or not the suggested methodology is scalable, it should be used to datasets that are both larger and more diverse. These datasets should include reviews composed of a variety of e-commerce platforms and languages. In order to have a better grasp of the model's wider application, it will be essential to evaluate the model's generalisation across more than one domain.

References

1. Mamta., Asif, Ekbal. (2025). Quality achhi hai (is good), satisfied! Towards aspect based sentiment analysis in code-mixed language. Computer Speech & Language, doi: 10.1016/j.csl.2024.101668
2. Neuton, Trindade, Vasconcelos., Pedro, Soares, De, Araujo., Kelson, Faial., Bruno, Santana, Carneiro. (2025). Qualitative analysis of the shallow underground waters of benevides, pará, based on multivariate statistical methods. Química Nova, doi: 10.21577/0100-4042.20250021
3. Peng, Yang., Dong, Tian., Xiaoyu, Han., Qingjun, Zou., Liang-Ju, Ma., Min, Wei., Meng, Yu., Zhong-Mei, Zou. (2024). Optimal Harvest Period and Quality Control Markers of Cultivated Flos Chrysanthemi Indici using Untargeted/Targeted Metabolomics, chemometric analysis and in vivo study. Journal of Ethnopharmacology, doi: 10.1016/j.jep.2024.118533
4. Xin, Sui., Wenjia, Wang., D, D, Zhang., Jianlong, Xu., Jing, Li., Yongxu, Jia., Yanru, Qin. (2024). Integrated analysis of ferroptosis and stemness based on single-cell and bulk RNA-sequencing data provide insights into the prognosis and treatment of esophageal carcinoma. Gene, doi: 10.1016/j.gene.2024.148701
5. Ni, Wayan, Sumartini, Saraswati., I, Ketut, Gede, Darma, Putra., Made, Sudarma., I, Made, Sukarsa. (2024). Enhance sentiment analysis in big data tourism using hybrid lexicon and active learning support vector machine. Bulletin of Electrical Engineering and Informatics, doi: 10.11591/eei.v13i5.7807
6. Sujatha, Uthamaraj., Gunasundari, Ranganathan. (2024). A novel framework for MOOC recommendation using sentiment analysis. Indonesian Journal of Electrical Engineering and Computer Science, doi: 10.11591/ijeecs.v36.i1.pp603-613
7. Zijian, Qin., Kefan, Dong., Bing, Xie. (2024). What affects customers online shopping behavior, research that applied machine learning to Amazon product reviews. Applied and Computational Engineering, doi: 10.54254/2755-2721/76/20240564.
8. Adigun, E.B.., Pablo, Alejandro, Quezada–Sarmiento. (2024). Unveiling Emotions: Exploring Sentiment Analysis on Flipkart Data with the Power of Support Vector Machines and the Intricacies of Random Forest Algorithm. doi: 10.1109/icdsis61070.2024.10594163
9. Jitendra, Soni., S., Joseph, Xavier., R., N., Saxena., Priansh, Waghela. (2024). Leveraging Web Scraping for Aspect-Based Sentiment Analysis: A Case Study on Flipkart Customer Reviews of Mobile Phones. International Journal of Research Publication and Reviews, doi: 10.55248/gengpi.5.0624.1638
10. Navtej, Kumar, Singh., Divyansh, Khandelwal. (2024). Sentiment analysis for amazon music instruments review. International journal of research in circuits, devices and systems, doi: 10.22271/27084531.2024.v5.i1a.59
11. Manasi, S, Byali., Suvarna, G., Kanakaraddi., Tulsa, Badagi., Ashok, K., Chikaraddi. (2024). Sentimental Analysis of Amazon Reviews using Map Reducer Technique in Hadoop. doi: 10.23919/indiacom61295.2024.10498578
12. Jiaqi, Li., Qi, Pan., Yihao, Wang. (2024). Sentiment analysis applied on Amazon reviews. Applied and Computational Engineering, doi: 10.54254/2755-2721/44/20230079
13. Yuanhao, Shen. (2024). An Effective Sentimental Analysis Model Based on spaCy. Highlights in Science Engineering and Technology, doi: 10.54097/91axqs95
14. Vipin, Jain., Priyanshu, Choudhary., Sachin, Arora., Tanishk, Mangal., Ankit, Choudhary., Himanshu, Kumar. (2024). A Comprehensive Exploration of Stack Ensembling Techniques for Amazon Product Review Sentiment Analysis. doi: 10.1109/icrito61523.2024.10522384
15. Mahammad, Khalid, Shaik, Vadla., Mahima, Agumbe, Suresh., Vimal, Viswanathan. (2024). Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT. Algorithms, doi: 10.3390/a17020059
16. K., Venkata, Subbaiah., Bharath, Kumar, Bolla. (2024). Aspect category learning and sentimental analysis using weakly supervised learning. Procedia Computer Science, doi: 10.1016/j.procs.2024.04.118

17. Mr.Bobby, kumar., Ms., Anitha., DSouza, Jacintha., Dr., Sheetal., Ms., Veena., Seema, Badiger. (2024). Sentiment Analysis for Products Review based on NLP using Lexicon-Based Approach and Roberta. doi: 10.1109/iitcee59897.2024.10468039.
18. Kalpa, Subbaih., Bharath, K., Bolla. (2023). Aspect category learning and sentimental analysis using weakly supervised learning. arXiv.org, doi: 10.48550/arxiv.2312.15526.
19. Bing, Bai., S., R., Likhith., Salma, Itagi. (2023). Sentimental Analysis-Based Recommended System for Products Using Machine Learning. doi: 10.1007/978-981-99-7622-5_14.
20. Kiran, Kumar., Sanjeev, Kumar, Singh., Ankit, Kumar., Anurag, Singh. (2023). Sentimental Analysis of Customer's Reviews Using Machine Learning. doi: 10.1109/icac3n60023.2023.10541464.
21. Kanika, Sharma., Dhyanendra, Jain., Ashu, Jain., Y., S., Rathore., Shruti, Agarwal. (2024). Sentimental Analysis based on Machine Learning Technique. doi: 10.1109/inocon60754.2024.10511973
22. Horesh, Kumar., Anshul, Kumar., Adityavikram, Anand., Ashin, Sabu., Tarun, Jain. (2023). Sentiment Analysis on Amazon Electronics Product Reviews using Machine Learning Techniques. doi: 10.1109/gcat59970.2023.10353467.
23. Bhuvaneshwaran, A., Amitha, K., Bhavya, A., Kamali, S., Karishma, R., C., Deisy., Sridevi, S. (2023). Sentimental Analysis Using Deep Learning Models. doi: 10.1109/icemce57940.2023.10434008.
24. Vishal, Satpute., Vaibhav, Singh., Dhananjay, Gayke., Rushikesh, Kamod., Prof., (Dr), Amol, Potgantwar., Prof., Rais, Shaik. (2023). Sentimental Analysis using Neural Network. International Journal of Advanced Research in Science, Communication and Technology, doi: 10.48175/ijarsct-13693.
25. Hardarshan, kaur., Nikhilesh, Yadav., Mendu, Sai, Naga, Tarun. (2024). Sentimental Analysis Using Machine Learning Algorithms. Social Science Research Network, doi: 10.2139/ssrn.4491280.
26. Sandeep, Yelisettiand., Nellore, Geethanjali. (2023). Aspect-based Text Classification for Sentimental Analysis using Attention mechanism with RU-BiLSTM. doi: 10.12694/scpe.v24i3.2122.
27. Karthikeyan, Thiagarajan., Arushi, Bansal., N., Muthurasu. (2023). Sentimental Analysis on Media. doi: 10.1109/icraset59632.2023.10420291.
28. Horesh, Kumar., Anshul, Kumar., Adityavikram, Anand., Ashin, Sabu., Tarun, Jain. (2023). Sentiment Analysis on Amazon Electronics Product Reviews using Machine Learning Techniques. doi: 10.1109/gcat59970.2023.10353467