International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

# Analyzing Customer Sentiments from Product Reviews Using Deep Learning Algorithms in E-Commerce Web Applications

**Dr Prashant Kumar Srivastava**

Associate Professor, SOCT, Sanjeev Agrawal Global Educational (SAGE) University, Bhopal, India

prashant.s@sageuniversity.edu.in

**Publication History:** Received: 18.03.2026; Revised: 10.04.2026; Accepted: 13.04. 2026; Published: 18.04.2026.

**ABSTRACT:** Sentimental Analysis is utilized in all the online product companies' reviews. Other users had these reviews considered as they searched products. This study focuses on analyzing customer sentiments from product reviews using supervised machine learning and deep e-commerce Web application learning algorithms. The review of the Amazon product used a balanced data set of 400,000 reviews of fifty thousand products over five product categories including: mobile electronics, furniture, cameras, groceries, and watches. A preprocessing pipeline of the reviews went through a preprocessing stage of data cleaning and tokenization followed by Features extracted using Bag-of-Words (Bow) and TF-IDF approaches. As a set of models, the employed BERT, a transformer-based deep learning model, Naive Bayes, and Support Vector Machines (SVMs). Among them, BERT's F1-score of 92.41, recall of 92.24, accuracy of 92.14, and precision of 92.14 were the best. The results validate BERT's superior capability in capturing contextual and sentiments, making it a powerful tool for automating sentiment analysis and enhancing user experience in e-commerce platforms.

**KEYWORDS:** E-Commerce Web Applications, BERT Model, Amazon Product Review Dataset, Machine Learning (ML), Deep Learning (DL).

## I. INTRODUCTION

With the exponential growth of digital data in today's era of Big Data, businesses are inundated with massive volumes of unstructured information generated from various online platforms. Among these, product reviews on E-Commerce web review applications are abundant source of customer views and experience [1][2][3]. The reviews not only show the satisfaction of customers but also significantly contribute to making buying decisions of potential buyers who might see them[4]. Nevertheless, such reviews are of high volumes and have different forms of linguistic expression and therefore manual analysis is not feasible and hence automated analytic methods are required[5][6]. It is at this point that the idea of sentiment analysis of customers comes in [7][8]. An NLP technique is sentiment analysis, often known as opinion mining that categorizes the textual information into basic sets like positive, negative, or neutral. It will help businesses to make sense of the mass reviews of customers and comprehend what people think about their products and services [9][10]. Summarizing and analyzing the mass response of customers, the results of sentiment analysis offer practical details about how the marketing, product development, and customer support staff need to act strategically[11][12].

ML algorithms as NB, SVM, and RF have been extensively applied to address the complexities entailed in this analysis [13][14]. Such supervised learning algorithms can learn patterns in labelled data and predict sentiment rather accurately. Although they achieve success, classical ML models usually fail at describing the contextual meaning of the words in the sentence, especially when it comes to sarcasm, negation, or implicit feelings [15][16]. Such shortcomings have necessitated DL algorithms that are more sophisticated, particularly those that rely on transformers like BERT. Because of its robust context understanding function, which takes into account both the left and right context of a word, BERT is very effective in sentiment categorization. Training BERT using domain-specific data (such as Amazon product reviews) makes it more precise and applicable. This is why E-Commerce sentiment analysis is seeing a paradigm shift as a result of deep learning models [17][18]. Therefore, a combination of sentiment analysis and ML and DL approaches can make E-Commerce apps smart enough to create personalized suggestions, predict consumer actions, and promote the general shopping experience. The paper further discusses such methodologies and illustrates the use of supervised learning techniques as a means of

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

analyzing customers' sentiments based on the comments made on products, and assists in making data-based decisions in the online market.

*A.* Motivation and Contribution of Study

The motivation behind this study stems from the exponential growth of unstructured customer feedback on e-commerce platforms like Amazon, where millions of reviews on a product are created daily. It is very difficult to analyse this enormous amount of textual data manually, but this kind of review contains very useful information about preferences, satisfaction and pains of the customer. However, customer sentiment has become an important factor to augment better product recommendations in the online marketplace, where competition is on the rise. The proposed research will fill this gap and apply supervised ML and DL models, especially BERT to automate and enhance the precision of sentiment classification. It aims to equip businesses with pragmatic information and helping consumers make effective decisions when buying products.

- Introduced a dual-stage classification approach separating product/service reviews and analyzing sentiment polarity for more detailed customer feedback analysis.
- Developed a thorough preprocessing pipeline addressing missing data, normalization, and tokenization to prepare text for efficient sentiment modeling.
- Compared traditional ML models with BERT, proving BERT's superior accuracy and contextual understanding in customer sentiment classification tasks.
- Employed Bow and TF-IDF methods to extract meaningful text features, enhancing sentiment classification across diverse product categories.
- Applied methods on Amazon review dataset, showcasing real-world effectiveness and supporting better decision-making for E-Commerce platforms and businesses.

*B.* Structure of paper

The outline of this article is as follows: Section II discusses relevant research on sentiment analysis that makes use of ML and DL models. Section III explains the methodology and BERT model. Section IV displays the outcomes and compares the models. Section V ends by discussing its boundaries and potential applications.

## II. LITERATURE REVIEW

This section presents research on customer sentiments from product reviews is systems utilizing various ML and DL techniques; summary of these studies is provided in Table I.

Krishna et al. (2025) aims to automate the classification of reviews based on both topic classification (product-based review or service-based review) and sentiment (positive review, negative review, neutral review). This study examines the application of ML techniques to analyze customer feedback from Amazon, with a focus on separately classifying product and service reviews, followed by sentiment analysis for each category. The results highlight key trends in customer experiences and demonstrate the efficacy of ML in multi-level sentiment analysis. By utilising the NB Classifier, the model achieves the accuracy of range 80% to 90% [19].

Rongala et al. (2025) online shopping as seen in product reviews like observed in Amazon has grown over the past years. Sentiment analysis or opinion mining concept is founded on identifying the sentiment or opinion of the specific text and the constituents of the text based on reviews or tweets or customer feed-back based on e-learning, computational linguistics, ML, and NLP. This paper explores how ML and NLP techniques can first analyse Amazon product reviews and then extract key attitudes, tastes and trends. Other metrics taken into consideration during the evaluation of the models, including the BERT, are F1-Score, recall, accuracy, and precision. In terms of sentiment categorization as well as review analysis, BERT stood tall performing the best at 89 percent accuracy [10].

Arora et al. (2025) discuss the role of sentiment analysis and forecasting to enhance the performance in the business, particularly in such an e-commerce business as Amazon to evaluate the feedback of customers to improve the work in the future. This paper aims at integrating ML models, such as SVM, RF and NB, to classify product reviews as positive, negative or no sentiment analysis, which has higher accuracy levels of 0.85, 0.87 and 0.85 based on their use in 11,000 or more datasets [20].

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

Okay et al. (2024) the phenomenon of sentiment analysis in the reviews of e-commerce websites and it's the potential to reveal the opinions of consumers and support business and customer decision-making activity. As online shopping flourishes, the ability to read the sentiment of the customers has become the key in improving experiences as well as enhancing product offerings to users. The RF model and the XGBoost model achieve an excellent performance in three sentiments using the TF-IDF with the Count Vectorizer. RF model has an accuracy of 91 % in both TF-IDF and Count Vectorizer, and XGBoost model has 91 % in Count Vectorizer and 90 % in TF-IDF. Also, SGD + TF-IDF and Count Vectorizer were used in two classes of sentiment, compared to an accuracy of 80 per cent in both methods [21].

Ceniza-Canillo et al. (2022) attempts to examine product-review sentiments as customer-advice in Shopee within the Philippines. The reviews of the products on Shopee were first scraped and then pre-processed after which it is annotated using VADER. The sentiments of the customers were analysed by using Multinomial-NB (MNB), and SVM. Thereafter, the value of accuracy, the value of precision, the value of recall and the F-measure of the results of the models were calculated using a confusion matrix and a classification report. And as a last part, their survey results would be employed to rationalize the model result the recommendations to the customers. The conclusion section of this research study reveals the effect of positive, negative, and neutral opinions on the condition of a product, whether to be able to recommend it to other consumers or not. As the review of the product reviews shows, 83.6 of them are positive, 9.1 percent negative, and 7.3 percent neutral. It has been discovered that the SVM model is an improved model as compared to MNB, which obtained an 83 per cent accuracy mark [22].

Mursalin and Khan et al. (2022). This era of internet technology is seeing Bangladeshi internet marketers and retailers go ballistic. With an increasing number of citizens affected by the COVID-19 pandemic, online shopping became the primary means of shopping and was considered the safest option. The businesses were compelled to be online. The present work is an attempt to present a model to assign sentiment to the nature of online tech gadget reviews in Bengali, into three simple classes, viz., positive, negative, and neutral. To this end, data on Bengali tech reviews, totalling approximately 6015 cases, is obtained. Several Feature Extraction techniques are subsequently applied in conjunction with various ML techniques. Having assessed the performance, the significance of the Random Forest outweighs that of the remaining other techniques, and its accuracy reaches a maximum of 86.28% [23].

TABLE I. COMPARATIVE ANALYSIS OF RELATED WORK BASED ON CUSTOMER SENTIMENTS FROM PRODUCT REVIEWS

| Author | Methodology | Data | Key Findings | Limitation | Future Work |
|---|---|---|---|---|---|
| Krishna et al. (2025) | Naïve Bayes Classifier for topic + sentiment classification | Amazon Reviews | Achieved 80–90% accuracy; classified reviews as product/service-based and then by sentiment | Used only Naïve Bayes; lacks deep learning comparison | Extend to deep learning models for improved performance |
| Rongala et al. (2025) | ML and NLP including BERT, evaluated using F1, Recall, Accuracy, Precision | Amazon Product Reviews | BERT outperformed others with 89% accuracy; extracted opinions and trends effectively | Limited feature set and data diversity | Apply to real-time streaming reviews and multilingual datasets |
| Tiwari and Arora (2025) | SVM, Random Forest, Naïve Bayes | 11,000+ Amazon Reviews | Accuracy: SVM (85%), RF (87%), NB (85%); provided business improvement insights | Focus on basic sentiment classes only | Include multi-class emotion detection and aspect sentiment |
| Aydoğan and Okay (2024) | RF, XGBoost, SGD with TF-IDF & CountVectorizer | E-commerce reviews from multiple platforms | RF and XGBoost achieved 91% accuracy with TF-IDF and CountVectorizer | Did not address multilingual or complex sentence structures | Apply in recommender systems and hybrid modeling |
| Ceniza-Canillo et al. (2022) | SVM, Multinomial Naïve Bayes with VADER annotation | Shopee Philippines product reviews | SVM outperformed MNB with 83% accuracy; majority sentiments were positive (83.6%) | Regional dataset; model generalization not verified | Extend to multi-regional data with deep learning models |
| Mursalin and Khan et al. (2022) | Multiple ML models (incl. RF), feature extraction for Bengali sentiment | 6,015 Bengali tech gadget reviews | RF gave highest accuracy (86.28%); effective sentiment classification in local language | Dataset limited to Bengali and tech products | Scale up for cross-language and category sentiment models |

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

## III. METHODOLOGY

The methodology of the study involves a comprehensive pipeline in sentimental examination of product evaluations posted by online shoppers on sites like Amazon. It begins with data collection from a large, publicly available dataset containing over 100 million reviews, from which a balanced subset of 400,000 reviews across five product categories was selected. Preprocessing included handling missing values using interpolation and fill methods, changing to lower-case, dropping the stop words, punctuation and HTML classes, and running tokenization to prepare the text for DL. Feature extraction was performed using Bag-of-Words (BoW), with the CBOW variation capturing contextual relationships, and TF-IDF scoring words by their importance across documents. The result split into two parts: testing and training. The main model, BERT, was fine-tuned using the processed reviews. BERT's strength lies in its bidirectional transformer architecture, leveraging self-attention to understand contextual dependencies in text. The [CLS] token embedding was passed through a fully used SoftMax layer to make a sentiment prediction. The metrics used for assessment reasons, including F1-score, recall, accuracy, and precision, with BERT outperforming other models and achieving 92% accuracy, confirming its ability to capture nuanced sentiment patterns effectively in customer reviews. The Fig. 1 shows the flowchart of Customers Product Review E-Commerce Web Applications are given below:

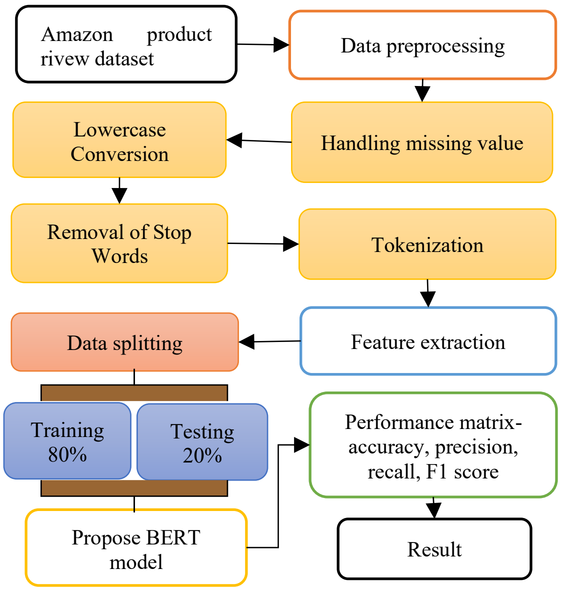


Fig. 1. Flowchart of the Customers Product Review E-Commerce Web Applications

### *C.* Data Collection

The data was collected from a publicly available Amazon product reviews dataset hosted on Kaggle. It includes reviews across five major categories of products: mobile electronics, furniture, cameras, groceries and watches. Based on the wide range of original dataset of over 100 million entries, a balanced subset of 400,000 reviews was selected, comprising 80,000 reviews from each category. Each review entry includes detailed columns like customer ID, product ID, headline of review, body of review, star rating, product category, and purchase verification status. This structured and diverse dataset served as a robust foundation for performing sentiment analysis using the spectrum of ML, DL and transformer-based models.

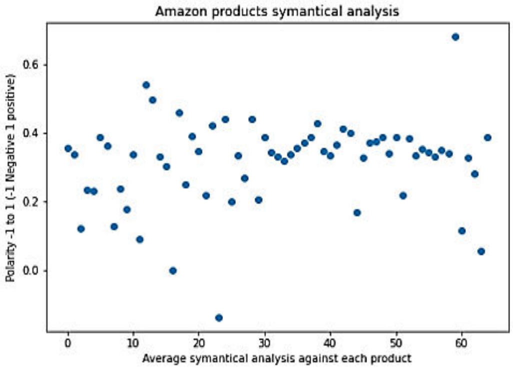


Fig. 2. The Dot Visualization of the Amazon Product Semantical Analysis

Fig. 2 presents a dot plot visualization depicting the semantical (sentiment-based) analysis of various Amazon products. Each dot on the graph represents a product, with the x-axis indicating the average semantical analysis score for that product and the y-axis showing its corresponding polarity value, which ranges from -1 (most negative) to +1 (most positive). The overall trend suggests a moderate concentration of polarity scores between 0.2 and 0.4, indicating that most products receive slightly positive feedback from users. A few outliers can be seen with very low or relatively high polarity, showing that some products are either heavily criticized or exceptionally praised. This visualization effectively captures how customer sentiment varies across different products based on textual review analysis.

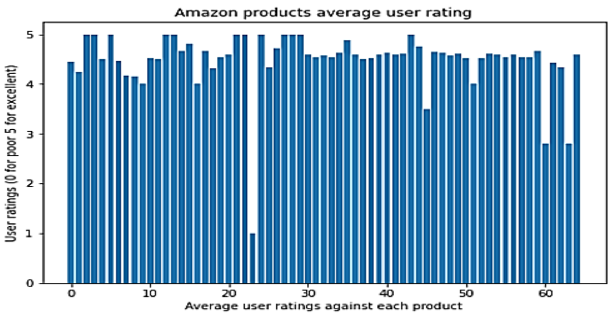


Fig. 3. The Bar Visualization of the Amazon Product User Rating

Fig. 3 presents a bar chart representing the average user ratings for various Amazon products. The x-axis shows different products, while the y-axis denotes the user ratings, ranging from 0 (poor) to 5 (excellent). Most products receive high ratings, with the majority clustering around the 4 to 5 range, indicating generally positive customer satisfaction. A few noticeable dips in the bars highlight products with significantly lower ratings suggesting potential quality or service issues. This visualization helps identify consumer preferences and perceptions, supporting sentiment analysis by quantitatively reflecting how well products are received in the marketplace.

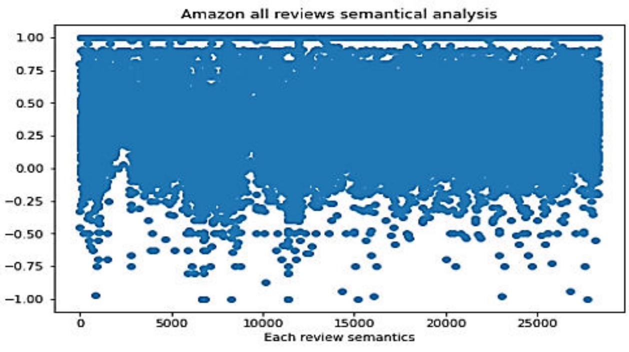


Fig. 4. The Dot Visualization of the Amazon All Reviews Semantical Analysis

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

Fig. 4 presents a dot plot visualizing the semantic polarity of all Amazon product reviews. On one side, we have reviews themselves; on the other, we have their semantic polarity, which might be anything from -1 (very negative) to 1 (extremely favourable)). The dense clustering at the top suggests a large number of reviews with high positive sentiment, while a scattering of points toward the bottom reveals the presence of moderately to strongly negative reviews. This wide distribution reflects the diversity in customer opinions, underlining the importance of sentiment analysis in extracting meaningful insights from massive and unstructured review data in e-commerce platforms.

*D.* Data Preprocessing

Interpolation and fill procedures were used to handle missing values in the data preparation. All text was converted to lowercase and stop words, punctuation, and HTML elements were removed. To round things off, the data was tokenised to separate words and phrases in order to facilitate feature extraction and model training.

- **Handling Missing Values:** At first, numerous characteristics in the dataset were missing. Since the "review body" and "star rating" sections are important for sentiment categorisation, the writers zeroed in particularly on them. Null values were filled using Python's fill () function for string (object) data types. For numeric fields like “star rating”, interpolation was used, which, using nearby data points, attempts to fill up missing values. Table 2 in the paper lists the total missing values per feature.
- **Lowercase Conversion:** A lowercase version of each review text was applied. The equality of treatment of terms like "Great" and "great" is guaranteed by this step, reducing feature space dimensionality and preventing duplication of semantically identical terms due to case sensitivity.
- **Removal of Stop Words:** We cleaned up the text data by removing HTML elements, punctuation marks, and stop words such "and," "the," and "is." These elements generally do not contribute meaningful information to sentiment classification and are considered noise in text mining.
- **Tokenization:** Both sentence-level and word-level tokenization were applied. Tokenisation is a method that breaks down large pieces of text into smaller, more manageable pieces called tokens, which form the basis for further NLP operations like vectorization and parsing.

*E.* Feature Extraction

The method of feature extraction is fundamental to natural language processing (NLP) since it allows ML systems to make sense of unstructured textual material. The authors used well-known methods in their study Backpack of Words (BoW). Ignoring word order and grammar, the Bag-of-Words (BoW) model preserves frequency but portrays each review as a vector of word counts. The frequency of each distinct word in the corpus is transformed into a feature in a review determines the feature’s value. This simple yet effective approach enables the algorithm to understand the presence and frequency of specific words that indicate sentiment. A variation of Bow used is the Continuous Bag-of-Words (CBOW) model, whose mathematical formulation (1) is:

$$P(w_t|\, \text{context}) = \text{softmax}(W_{out}.\frac{1}{n}\sum_{i=1}^{n} W_{in}\,[w_{t-1}] + b_{in}) \quad (1)$$

Where:

- $P(w_t|\, \text{context})$ is the probability of predicting the target word $w_t$ from surrounding words (context),
- $W_{in}$ and $W_{out}$ are the input and output embedding matrices,
- n is the number of context words,
- and the SoftMax function converts logits into probabilities.

*F.* Data splitting

There was a 20% testing set and an 80% training set inside the dataset. The models were optimised and trained on the training set, and subsequently evaluated on the testing set. The results of each model on the validation set were recorded.

*G.* Proposed BERT Model

BERT is an acronym for is an ML-based ML model that was pre-trained on a huge text corpus. Unlike traditional language models, BERT captures both left and right context simultaneously, enabling a deep understanding of word meanings in context an essential capability for sentiment analysis[24][25].

In the study, based on customer feedback on Amazon, we fine-tuned the BERT-base and BERT-large versions. Tokenising and embedding the input text is the initial step before passing it via several transformer encoder levels. Word connections are represented by each layer's self-attention processes, independent of the position of the words in the sentence.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

This attention mechanism in BERT is based on the Equation (2) for Scaled Dot-Product Attention, which is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_K}}\right) V) \qquad (2)$$

Where:

- Q = Query matrix
- K = Key matrix
- V = Value matrix
- $D_K$ = dimension of the key vectors
- $QK^T$ = dot product of query and key
- SoftMax = normalization function to get attention weights

This allows BERT to weigh the relative significance of different words inside a phrase, even if they are far apart.

After multiple layers of attention and transformations, the final embedding corresponding to the special classification token [CLS] is passed to a classification layer that is completely linked and uses the SoftMax activation function into sentiment categories (positive, neutral, negative). The Final Classification Output shown in Equation (3):

$$\hat{y} = \text{softmax}(W_h \,.\, h_{[CLS]} + b) \qquad (3)$$

Where:

- $h_{[CLS]}$ = output vector of the [CLS] token
- $W_h, b$ = learned weights and bias of the classification layer
- $\hat{y}$ = predicted probability distribution over sentiment classes

In the study, BERT achieved the highest performance among all models, with an accuracy of 92%, confirming its effectiveness in capturing complex contextual sentiment patterns in Amazon product reviews.

*H.* Performance Matrix

The confusion matrix is the first metric we use to evaluate the models. The number of correctly identified positive samples (TP) and negative samples (TN) is shown on the primary diagonal of a confusion matrix. In this context, "false positives" refer to negative samples that are mistakenly thought to be positive, while "false negatives" refer to positive samples that are mistakenly thought to be negative, are shown on the counter diagonal.

**Accuracy:** This statistic measures the accuracy of the predictions relative to the number of instances that were actually tested, where FN denotes false negatives, FP denotes false positives, TN denotes true negatives, and TP denotes true positives in Equation (4):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

**Precision:** This statistic determines the accuracy rate of the model's predictions as a percentage of all positive predictions (including false positives and true positives) using Equation (5):

$$\text{Precision} = TP \,/\, TP + FP \qquad (5)$$

**Recall:** This metric assesses how well the model detected cases based on true positive data. within the framework of Equation (6):

$$\text{Recall} = TP \,/\, TP + FN \quad (6)$$

**F1_Score:** A harmonic mean of recall and accuracy, the F-score may take on values ranging from 0 to 1. Model performance is improved with a higher F-score. To get the F-score, one might use the following formula (7):

$$F1 = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (7)$$

## IV. RESULTS AND DISCUSSION

The experiment has been conducted on a computer system that has a Core i5 10 th Generation processor, a 3.1 GHz processor, with 16 GB RAM and a 500 GB hard drive as a storage. The programming language of Python was implemented

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

on Spyder IDE and analyzed to develop and test the model. Spyder, managed by the Spyder Project contributors, is an open-source environment, designed to fit the requirements of scientific computing and data analysis in Python.

TABLE II. PERFORMANCE MATRIX OF BERT MODEL FOR CUSTOMER SENTIMENTS FROM PRODUCT REVIEWS

| Matrix | BERT |
|---|---|
| Accuracy | 92 |
| Precision | 92.14 |
| Recall | 92.24 |
| F1 Score | 92.41 |

Table II shows the evaluation values of BERT model on customer sentiment presented through product reviews. General accuracy of the model is 92%, which means its effectiveness in properly assigning sentiments. The model has an accuracy of 92.14% and thus it can precisely determine the interesting classes of sentiment without causing too much false positive. A high recall of 92.24 shows that it is a powerful model capable of attracting most of the real instances of the sentiment cases hence reducing the false negatives. The model's stability and dependability are shown by its degree of recall and accuracy (F1 score of 92.41%) in all moods. All these measurements are confirming the high efficiency and relevance of BERT used in real-life sentiment analysis in e-commerce platforms.

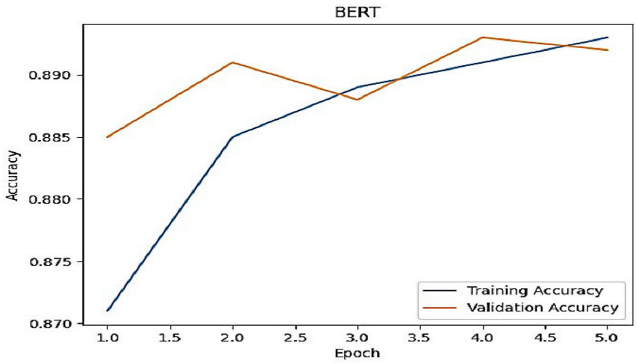


Fig. 5. Accuracy curve of BERT Model.

Fig. 5 shows the accuracy curve of BERT model trained with five epochs compared to both training and validation accuracies. As it can be seen, the values of the training accuracy show a stable and progressively increasing graph of about 87.1-89.2, which implies that the model is still learning during the run. The accuracy of the validation increases rapidly and reaches its maximum about 4th epoch and almost 89.3 per cent, then falls a bit indicating that slight overfitting has begun. In general, the BERT model has shown good and steady results with near similar training and testing curves demonstrating the generalization ability of the model in carrying out the sentiment classification task. As it may be seen, the trend of the training accuracy increases gradually up to about 87.1 the model continues to learn effectively over time, with a range of 89.2% to 100%. The accuracy with which the validation is validated increases rapidly and reaches its height when the 4th epoch is reached and is approximately equal to 89.3 percent and then, it decreases slightly indicating that it is starting to be slightly overfitted. On the whole, the BERT model proved to perform well and robust with values of the training and validation curves being very close and showing the ability of the model to generalize sentiment classification tasks.

International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)

|www.ijrpetm.com | ISSN: 2454-7875 | editor@ijrpetm.com |A Bimonthly, Peer Reviewed & Scholarly Journal|

||Volume 9, Issue 2, March - April 2026||

DOI:10.15662/IJRPETM.2026.0902014

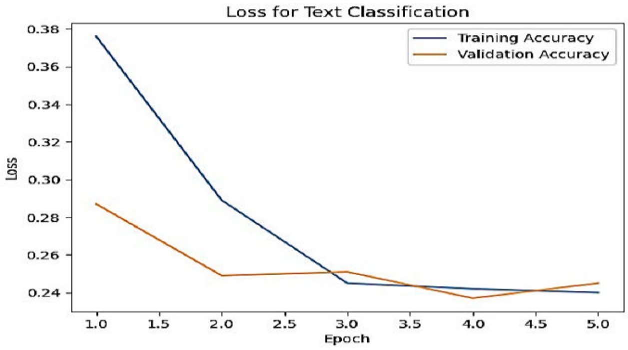


Fig. 6. Loss curve of BERT Model.

Fig. 6 shows the loss graph of the BERT model used in text classification in five epochs. The training loss shows a clear and consistent decline from approximately 0.38 to 0.23, indicating effective learning by the model. Similarly, the validation loss follows a downward trend initially, reaching its lowest point near the fourth epoch, before slightly increasing in the final epoch. This minor rise in validation loss suggests a slight overfitting tendency as the model continues to train. Overall, the loss curves reflect stable and efficient training, with the BERT model maintaining low error rates throughout the learning process.

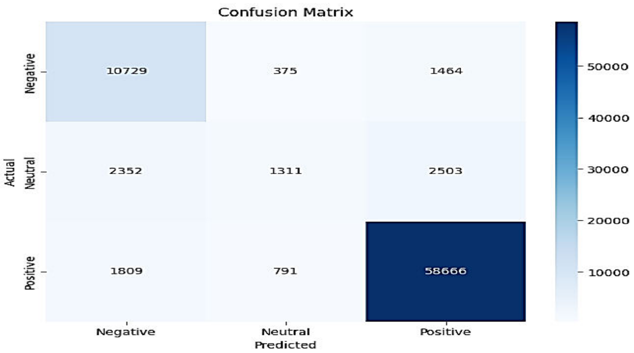


Fig. 7. Confusion Matrix for the BERT Model.

Fig. 7: A confusion matrix for the BERT model, used to assess how well it categorizes emotion into three groups: Negative, Neutral, and Positive. The model accurately predicts a large portion of Positive sentiments (58,666), indicating strong performance in identifying positive reviews. It also performs fairly well in the Negative class with 10,729 correct predictions but shows some misclassification, as 1,464 bad evaluations are erroneously deemed good. The Neutral class is the most challenging, with only 1,311 correct predictions and a noticeable number of misclassifications into both Negative (2,352) and Positive (2,503). Overall, the matrix reflects high accuracy for the Positive class but reveals difficulty in distinguishing Neutral sentiments, suggesting potential improvement in handling ambiguous or mixed reviews.

*I.* Comparison with Discussion

The following comparison between propose and existing model performance are based on evaluation matrix provided in table III.

TABLE III. COMPARISON BETWEEN BASE AND PROPOSE MODELS FOR CUSTOMERS PRODUCT REVIEW E-COMMERCE WEB APPLICATIONS

| **Matrix** | **BERT** | **SVM**[26] | **RF**[27] |
|---|---|---|---|
| accuracy | 92 | 85.44 | 83.33 |

In the context of customer product evaluation, Table III compares the proposed BERT model's performance to those of standard ML models like SVM and RF. research of user sentiment for online shopping platforms. The BERT model outperforms both SVM and RF across all evaluated metrics. BERT achieves the highest accuracy of 92%, significantly surpassing SVM at 85.44% and RF at 83.33%. Moreover, precision, recall, and F1-score of BERT are 92.14%, 92.24%, and 92.41%, respectively, and larger than the values of the other models, which indicates the great capacity to recognize and label sentiments. Contributing to the convincing nature of the results provided by BERT, the lack of precision, recall, and F1-score values related to SVM and RF additionally confirms the overall assessment of the gathered results, making BERT a viable model of learning when it comes to the complex understandings of sentiment in the customer reviews scenario than the classical approaches to the matter.

## VI. CONCLUSION AND FUTURE SCOPE

Nearly every individual depends on social media to a great extent in his or her daily routine. It enables the individuals to express what they believe and feel about the products in E-commerce web site. The purpose and the aim of an opinion or a review is to calculate the mood of the customer, it could either be positive or be negative about the merchandise. To conclude, this research has effectively presented that, the adoption of sophisticated deep learning methods such as the BERT is very effective in improving the accuracy and context of customer sentiment classification based on e-commerce product reviews and the results of the traditional algorithm i.e. SVM and Random Forest. The model achieved 92% accuracy, along with strong precision, recall, and F1-score, confirming its robustness for real-world applications. Future work can extend this framework to multilingual and real-time review streams, incorporate aspect-based sentiment analysis for finer insights, and integrate these systems into recommender engines to provide highly personalized customer experiences in global e-commerce platforms.

## REFERENCES

[1] M. Loukili, F. Messaoudi, and M. El Ghazi, "Sentiment Analysis of Product Reviews for E-Commerce Recommendation based on Machine Learning," Int. J. Adv. Soft Comput. its Appl., vol. 15, no. 1, pp. 1–13, 2023, doi: 10.15849/IJASCA.230320.01.
[2] M. Menghnani, "A Comprehensive Survey on Scalability and Performance in Full Stack Web Applications," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 2, pp. 1–14, 2025.
[3] V. S. Thokala, "A Comparative Study of Data Integrity and Redundancy in Distributed Databases for Web Applications," Int. J. Res. Anal. Rev., vol. 8, no. 4, pp. 383–389, 2021.
[4] H. Ravilla, "Predictive Analytics for Customer Churn in Salesforce Service Cloud," in Learning and Analytics in Intelligent Systems, 2026, pp. 14–24. doi: 10.1007/978-3-032-05377-0_2.
[5] C. Patel, "Effect of Digital Transformation on Customer Engagement in Retail Industries : A Comparative Review," J. Technol., vol. 10, no. 1, pp. 165–174, 2022.
[6] A. Syed, "The Anatomy of Supply Chain Applications," in Supply Chain Software Security, Berkeley, CA: Apress, 2024, pp. 85–125. doi: 10.1007/979-8-8688-0799-2_3.
[7] S. Pandya, "Comparative Analysis of Large Language Models and Traditional Methods for Sentiment Analysis of Tweets dataset for text classification," Int. J. Innov. Sci. Res. Technol., vol. 9, no. 12, pp. 1647–1657, 2024, doi: 10.5281/zenodo.14575886.
[8] D. P. Guda, "Application Security Testing: From Manual Reviews to Automation," Shodh Sagar Int. Publ., 2025.
[9] S. Ainin, A. Feizollah, N. B. Anuar, and N. A. Abdullah, "Sentiment analyses of multilingual tweets on halal tourism," Tour. Manag. Perspect., vol. 34, Apr. 2020, doi: 10.1016/j.tmp.2020.100658.
[10] S. Rongala, S. A. Pahune, H. Velu, and S. Mathur, "Leveraging Natural Language Processing and Machine Learning for Consumer Insights from Amazon Product Reviews," in 2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), 2025, pp. 1–6. doi: 10.1109/ICSSES64899.2025.11009528.
[11] K. Dixit, "Custom Experience Analysis in the Hospitality Service Sector using AI-Based Hotel Review Analytics," Int. J. Adv. Res. Sci. Commun. Technol., vol. 5, no. 1, pp. 1–10, 2025.
[12] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," ESP J. Eng. Technol. Adv., vol. 5, no. 2, pp. 16–22, 2025, doi: 10.56472/25832646/JETA-V5I2P103.
[13] V. S. Thokala, "Integrating Machine Learning into Web Applications for Personalized Content Delivery using Python," Int. J. Curr. Eng. Technol., vol. 11, no. 6, pp. 652–660, 2021, doi: https://doi.org/10.14741/ijcet/v.11.6.9.
[14] A. R. Duggasani, "Scalable and Optimized Load Balancing in Cloud Systems: Intelligent Nature-Inspired Evolutionary Approach," Int. J. Innov. Sci. Res. Technol., pp. 2153–2160, May 2025, doi: 10.38124/ijisrt/25may1290.
[15] A. Dudhipala, R. Karne, and P. K. Pativada, "Prompt2Graph: Leveraging LLMs to Construct Knowledge Graphs from Technical Manuals," in 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2025, pp. 912–919. doi: 10.1109/ICIMIA67127.2025.11200177.
[16] S. P. Kalava, "Revolutionizing Customer Experience: How CRM Digital Transformation Shapes Business," Eur. J.

Adv. Eng. Technol., vol. 11, no. 3, pp. 163–166, 2024.

[17] Y. V. J. Reddy, T. Saibhagath, S. Bhaavana, and M. Janakiram, “A Sentiment Analysis on E-commerce Product Recommendation System using Machine Learning and Deep Learning Algorithms,” Int. J. Res. Publ. Rev., vol. 5, no. 11, 2024.

[18] V. Thangaraju, “Security Considerations in Multi-Cloud Environments with Seamless Integration: A Review of Best Practices and Emerging Threats,” Trans. Eng. Comput. Sci., vol. 12, no. 2, pp. 1–6, 2024.

[19] J. Krishna, A. Lakshmi Narayana, S. M. K. Juber, Y. Mothilal Yadav, G. Siva Karthik, and V. Pavan Kumar, “Analyzing Customer Feedback on Amazon to Reveal Sentiment Trends using Machine Learning,” in 2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE), IEEE, Feb. 2025, pp. 1–7. doi: 10.1109/ICCECE61355.2025.10941446.

[20] N. K. Tiwari and H. Arora, “Sentiment Analysis and Forecasting for Improved Business Performance in E-Commerce using Machine Learning Algorithms,” in 2025 International Conference on Electronics and Renewable Systems (ICEARS), 2025, pp. 1487–1491. doi: 10.1109/ICEARS64219.2025.10940994.

[21] E. H. Aydoğan and F. Y. Okay, “Sentiment Analysis of Reviews for E-Commerce Applications,” in 2024 9th International Conference on Computer Science and Engineering (UBMK), 2024, pp. 1–6. doi: 10.1109/UBMK63289.2024.10773400.

[22] V. M. C. Sagarino, J. I. M. Montejo, and A. M. Ceniza-Canillo, “Sentiment Analysis of Product Reviews as Customer Recommendations in Shopee Philippines Using Hybrid Approach,” in Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications, ICITDA 2022, 2022. doi: 10.1109/ICITDA55840.2022.9971379.

[23] S. D. Bappon, G. Sarwar Md. Mursalin, and M. I. Khan, “Sentiment Analysis of Bengali Texts on Online Tech Gadget Reviews using Machine Learning,” in Proceedings of 2022 25th International Conference on Computer and Information Technology, ICCIT 2022, 2022. doi: 10.1109/ICCIT57492.2022.10055639.

[24] V. Rajendran, D. Besiahgari, S. C. Patil, M. Chandrashekaraiah, and V. Challagulla, “A Multi-Agent LLM Environment for Software Design and Refactoring: A Conceptual Framework,” in SoutheastCon 2025, 2025, pp. 488–493. doi: 10.1109/SoutheastCon56624.2025.10971563.

[25] R. Karne, A. Dudhipala, and P. K. Pativada, “LoRA-Fine Tuned LLaMA-3 for Low-Resource Multilingual Machine Translation,” in 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2025, pp. 39–45. doi: 10.1109/ICIMIA67127.2025.11200663.

[26] J. Li, Q. Pan, and Y. Wang, “Sentiment analysis applied on Amazon reviews,” Appl. Comput. Eng., 2024, doi: 10.54254/2755-2721/44/20230079.

[27] B. Kamalesh and S. Vijayalakshmi, “Sentiment Analysis on Amazon Product Review,” Proc. - 2022 4th Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2022, no. 04, pp. 846–848, 2022, doi: 10.1109/ICAC3N56670.2022.10074287.