

Classification and analysis with Deep Learning for Personalized Product Recommendation in E-Commerce

Dr. Parth Gautam
Associate Professor,
Mandsaur University, Mandsaur
Department of Computer Sciences and Applications
parth.gautam@meu.edu.in

Abstract—"Environmental e-commerce" refers to an online business model that prioritises environmental responsibility. Environmental e-commerce could benefit from a fresh strategy that takes into account a large number of social interactions in order to address the problems with data sparsity and variety that plague conventional e-commerce recommendation algorithms: combining filters. This study introduces a recommendation framework based on Bidirectional Encoder Representations from Transformers (BERT), designed to capture contextual relationships in customer reviews for accurate product suggestions. The Amazon product review dataset was utilized, and preprocessing steps included handling missing values, tokenization, stemming, stop-word removal, and TF-IDF. The proposed BERT model was fine-tuned on the preprocessed dataset and compared against traditional models such as CNN, GRU and VADER. A higher accuracy rate of 89% was attained by the BERT-based method, which substantially outperformed the comparison models, according to the experimental evaluation. These results demonstrate the effectiveness of transformer-based architectures in understanding semantic meaning and user intent, thereby enabling more reliable and scalable product recommendation systems in e-commerce platforms.

Keywords—E-commerce, Personalized Product Recommendation, Sentiment Analysis, Deep Learning, Amazon Reviews.

I. INTRODUCTION

In the past few years, the fast growth of e-commerce and modern transportation has changed the way people shop. With the convenience and efficiency of online platforms, purchasing products via the Internet has become a preferred choice across all age groups. Unlike traditional shopping, however, online customers cannot physically evaluate products before purchase, leading to challenges such as discrepancies between product descriptions and actual goods[1][2][3]. Reviews of products have grown in importance as a crucial indicator for assessing goods, helping consumers overcome this shortcoming and ultimately making more informed purchases. Sentiment analysis and related techniques applied to these reviews provide valuable insights for both consumers and producers.

The expansion of e-commerce has coincided with the era of big data, where vast amounts of structured and unstructured information ranging from user demographics and browsing behaviors to purchase histories and product attributes—are generated daily[4]. While this data presents significant opportunities, it has also created the problem of "information overload," making it difficult for users to efficiently identify products that meet their needs[5]. Consequently, the development of intelligent solutions capable of processing and interpreting large-scale data has become indispensable.

The rise of product recommendation systems, which use big data analytics to provide customers with tailored shopping experiences, is a key reaction to this problem [6]. By analyzing user preferences, behaviors, and transaction histories, these systems generate tailored product suggestions, reducing search time and improving customer satisfaction[7].

Beyond enhancing convenience, personalized recommendations exert a substantial commercial influence, increasing conversion rates, boosting AOV, and driving platform revenue[8]. For instance, studies indicate that even a single interaction with a personalized recommendation can raise AOV by 369%, while such systems may account for up to 31% of an e-commerce platform's total sales. As a result, recommendation systems have become a strategic necessity in the highly competitive online marketplace.

The combination of physical and online purchasing strategies has increased the efficacy of recommendation systems. Modern e-commerce platforms combine the extensive variety of online products with offline channels that deliver tangible product experiences and rapid fulfillment[9]. This integration establishes a feedback loop where offline interactions enrich online datasets, and online analytics optimize offline services, thereby enabling a more accurate and dynamic understanding of consumer behavior.

Personalized recommendation systems are being powered by deep learning, a new and groundbreaking technology [10][11]. Unlike traditional algorithms that depend on manual feature engineering, deep learning models, including ANNs, CNNs, and RNNs, can automatically learn intricate, high-dimensional patterns from complex datasets[12][13]. These models excel in classification, where users and products are categorized based on behavioral and contextual features, as well as in analysis, where hidden relationships and insights are derived from massive amounts of user-generated data. This dual capability enables the delivery of precise, context-aware, and real-time product recommendations. Deep learning's incorporation into recommendation systems not only helps with information overload but also boosts engagement,

retention, and long-term platform growth—all of which are crucial in the increasingly competitive e-commerce market.

A. Motivation and Contribution of Study

The explosive growth of e-commerce has created unprecedented opportunities for personalized shopping experiences, yet it has also intensified the information overload problem, where users face difficulty identifying relevant products from vast online catalogs. While traditional recommendation algorithms have improved personalization, they often struggle with the scale, diversity, and dynamic nature of modern e-commerce data, especially when combining online and offline consumer interactions. This gap highlights the need for more advanced methods capable of automatically extracting deep and latent features from complex datasets to deliver highly accurate and context-aware recommendations. Deep learning offers a powerful solution, enabling effective classification of user preferences and intelligent analysis of behavioral patterns in real time. Online marketplaces can gain a competitive edge in today's data-driven industry, increase sales, raise engagement, and satisfy customers more effectively by utilising these features. The key contributions are discussed below:

- Development of a robust sentiment analysis framework by leveraging a large-scale, multi-category Amazon reviews dataset, ensuring balanced representation across diverse product domains.
- Improve text quality and model readiness by the use of thorough preparation procedures, such as handling missing data, stop-word removal, and tokenisation.
- Comprehensive evaluation of model efficacy utilising many metrics (accuracy, precision, recall, F1-score) to ensure thoroughness and reliability.
- Personalised product recommendations in online shopping could be a reality with the help of the suggested BERT model, which has shown to outperform baseline classifiers.

B. Justification and novelty

Traditional machine learning models often struggle to capture the semantic richness and contextual subtleties of consumer evaluations, highlighting the need for work to address the growing demand for intelligent recommendation systems in online shopping. To determine user intent and product relevance, existing techniques mostly use shallow text representations like TF-IDF or bag-of-words. BERT is a novel component of the suggested framework that fills this need; it is used for recommendation tasks to improve prediction accuracy by utilizing its deep contextual embeddings. Unlike conventional classifiers, BERT effectively models long-range dependencies in textual feedback, enabling the system to provide highly personalized product suggestions. The proposed approach not only achieves superior performance compared to baseline models such as CNN, GRU and VADER but also demonstrates the potential of transformer-based architectures as a scalable and robust solution for modern e-commerce recommendation systems.

C. Structure of paper

The structure of this paper is as follows: Section I introduces the significance of recommendation systems in e-commerce. Section II reviews related work on traditional and deep learning approaches. Section III presents the research methodology, including dataset description and preprocessing

steps. The experimental setup and results evaluation are detailed in Section IV. Section V wraps up the report and provides suggestions for further research.

II. LITERATURE REVIEW

This section presents research on Personalized Product Recommendation in E-Commerce that utilizes diverse deep learning techniques; the summary of these studies is provided in Table I.

Chen *et al.* (2025) A multi-layer perceptron (MLP) structure is built using a deep neural network (DNN) as its core model. The input layer receives user and commodity information; the hidden layer extracts high-dimensional data using a multi-layer nonlinear transformation. The output layer then returns the prediction score or preference probability of users for the commodities. In the aspect of feature extraction and selection, the user's basic information is digitally encoded, the historical behavior data is transformed into vector form, the interest tag is embedded with words, the commodity basic information category is uniquely encoded, the price is normalized, the text is transformed into vectors by pre-trained BERT model, and the image is extracted with pre-trained ResNet model[14].

B and M (2025) the goal of creating a personalized recommendation system that uses both user- and item-based collaborative filtering and the Apriori algorithm for association rule mining to suggest goods based on groups of frequently occurring items. An anonymized e-commerce dataset including user interactions like ratings, purchases, and product metadata used to train the algorithm. Performance metrics including F1Score, Precision, and Recall evaluated the efficacy and precision of the recommendation system. To help e-commerce platforms boost client engagement and revenue, the system built to give personalized recommendations in real-time based on user preferences [15].

L. Li (2024) forecasts the items that customers most likely buy this month. As an example, the proposed strategy was empirically tested on the fast-moving consumer goods e-commerce platform. The performance of the algorithm was simulated using offline experiments. Experiments show that compared with the collaborative filtering method based on cosine similarity, the accuracy of this method can be 3.74% higher, and the recall rate reaches 3.91%. This method cannot only improve the user's selection efficiency, but also improve the performance of collaborative filtering[16].

Li (2024) uses deep learning models, that is, models, to analyze user browsing, purchasing, and feedback behaviours, and optimize recommendation algorithms. The research process mainly includes user portrait design and model construction, model description recommendation and optimization, system integration, and application. According to the experiment, the recommendation accuracy of the rural e-commerce user portrait recommendation system increased by 20% or more, and the purchase conversion rate increased by 44.23%. It can be seen that deep learning[17],

Lee and Kim (2023), product recommendation was defined as making suggestions for new products to be sold and showcased in the store. This recommendation system suggests products that specific stores haven't stockpiled yet but are likely to sell well. Sales revenue increased by 1.75 times compared to the overall store average for these products, and data showed that 88% of the newly recommended

products in individual stores sold out within a week, based on the system's test installation. A score of 4.2 out of 5 was obtained from the survey that measured the satisfaction of business owners, suggesting a good degree of contentment [18].

Shi (2022) proposed a solution to the issue of poor recommendation accuracy in electronic gadgets: a BERT-BiLSTM-based customized recommendation model for online store products. The two BERT model pre-training challenges were used to develop the bidirectional language model. As a result, the BLSTM bidirectional neural network could be introduced later on, which merged the results of the text's forward and backward hidden layers to give contextual semantic information. With an RMSE value of only 0.82 for the personalised recommendation model for e-commerce products based on BERT-BiLSTM, the experimental results showed that the proposed model outperformed the benchmark models (BERT-SVM, BERT-RNN, and BERT-LSTM) in terms of recommendation accuracy (0.82). Hence, tailored product suggestions for online retailers are a practical application of the suggested methodology [19].

Ge (2022) modern e-commerce experiences a lack of localised user interest data, leading to inaccurate personalised product recommendations. build an item-specific recommendation model called LSPCNN, which is based on CNN's improved local similarity prediction. To begin, local characteristics are extracted using a convolutional neural network (CNN). Next, build the item score matrix to help first-time users localise their interests, and add a regulating layer on top of the CNN network. Finally, convolutional neural networks (CNNs) are used to predict the missing score in order to accomplish personalised recommendations. The experimental results show that the proposed LSPCNN model outperforms the upgraded CNN network model and the collaborative filtering recommendation model built on hybrid neural network with respect to data sparsity and mean absolute error (MAE). Users' interest in local feature data may be more precisely extracted by the proposed method, leading to more accurate e-commerce customised suggestions. There are also some practical considerations [20].

TABLE I. COMPARATIVE ANALYSIS OF PERSONALIZED PRODUCT RECOMMENDATION IN E-COMMERCE

Author	Methodology	Dataset	Key Findings	Limitations/Future Work
Chen et al., (2025)	Deep Neural Network (DNN) with MLP; feature extraction via BERT for text, ResNet for images, encoding user/item data	E-commerce user & commodity dataset	Achieved effective feature extraction and preference prediction	Requires scalability testing on large-scale real-world datasets
B and M, (2025)	The recommendation system is a hybrid that uses Apriori for association rule mining and Collaborative Filtering for both user-based and item-based recommendations.	Anonymized e-commerce dataset with user interactions (purchase history, ratings, product metadata)	Improved recommendation accuracy using Precision, Recall, and F1-Score; real-time personalized recommendations	Needs optimization for real-time large-scale deployment
L. Li, (2024)	Predictive model for monthly product purchases; empirical analysis on FMCG e-commerce	FMCG e-commerce platform dataset	Accuracy improved by 3.74% and recall by 3.91% over cosine similarity CF	Limited to FMCG domain; extension to other product categories needed
Li, (2024)	Deep learning for user portrait design & optimization of recommendation algorithms	Rural e-commerce user dataset	Accuracy improved by 20%; purchase conversion rate increased by 44.23%	Focused on rural e-commerce; needs extension to broader contexts
Lee & Kim, (2023)	Store-level recommendation for unIntroduced products	Retail store sales data	88% of recommended products sold in a week; 1.75× higher sales; owner satisfaction = 4.2/5	Pilot study; requires broader deployment & long-term analysis
Shi, (2022)	Personalized recommendation using BERT-BiLSTM	E-commerce product dataset	Lowest RMSE = 0.82 compared to BERT-SVM, BERT-RNN, BERT-LSTM	High computational cost; scalability for large-scale real-time recommendation
Ge, (2022)	Item recommendation using CNN with local similarity prediction (LSPCNN)	E-commerce dataset with sparse user interest data	Reduced sparsity, lower MAE, improved accuracy	Needs testing with more diverse datasets; efficiency in real-time systems

III. METHODOLOGY

The methodology for this study, as illustrated in Figure 1, begins with the Amazon product reviews dataset, which is subjected to multiple preprocessing steps such as handling missing values, stemming, tokenization, conversion to lowercase, and removal of stop. The dataset is subsequently partitioned into training and testing subsets after cleaning. Afterwards, the TF-IDF method is used to extract features. For categorization and generating individualized product suggestions, the retrieved features are loaded into the suggested BERT model. In the end, a performance matrix including recall, accuracy, precision, and F1-score is used to evaluate the system's usefulness. This matrix ensures that the model's predictive potential is evaluated thoroughly.

The following sections describe each step in detail, as illustrated in the methodology and proposed flowchart.

A. Data Collection

The dataset used for sentiment analysis comes from Amazon product reviews on the Kaggle website and is open to the public. Phones, furniture, cameras, groceries, and watches are just a few of the varied product types represented in the dataset. While the original dataset contains over 100 million reviews, extracted a balanced subset of 80,000 reviews per category, resulting in a total of 400,000 records. The extensive metadata that is included with every review entry includes details about the marketplace, the review, the product, the parent product, the title, the category, the star rating, the number of helpful votes, the total number of votes, the status of the purchase verification, the review body, the review headline, and the review body itself. This comprehensive set of attributes provides a robust foundation for analysing customer opinions and sentiments across varied product domains.

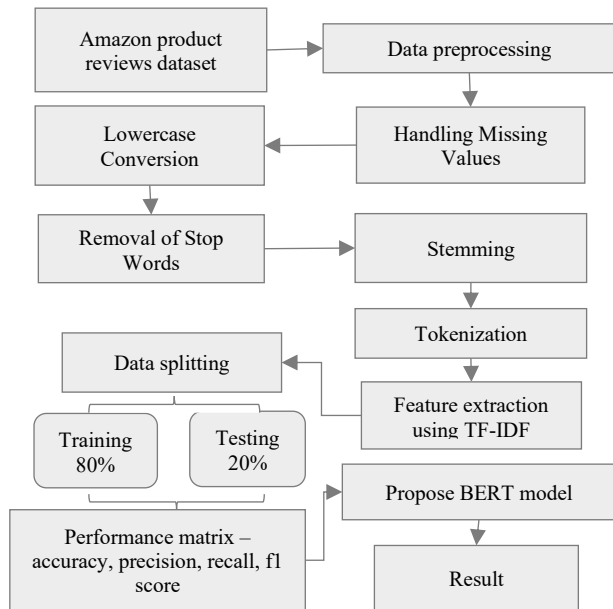


Fig. 1. Flowchart of the Product Recommendation in E-Commerce

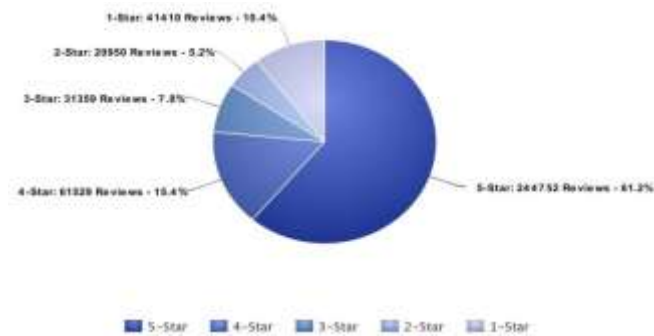


Fig. 2. Star rating distribution of reviews

The figure 2 presents a pie chart depicting the distribution of product review ratings. A significant majority of the reviews are rated 5 stars (61.2%), corresponding to 244,752 entries. The next largest proportion is 4-star reviews, comprising 15.4% (61,529 entries). Reviews rated 3 stars account for 7.8% (31,359 entries), while 1-star reviews constitute 10.4% (41,410 entries). The smallest proportion is represented by 2-star reviews, contributing 5.2% (20,950 entries). The distribution highlights a predominantly positive sentiment across the dataset, with higher star ratings (4-star and 5-star) forming the majority.

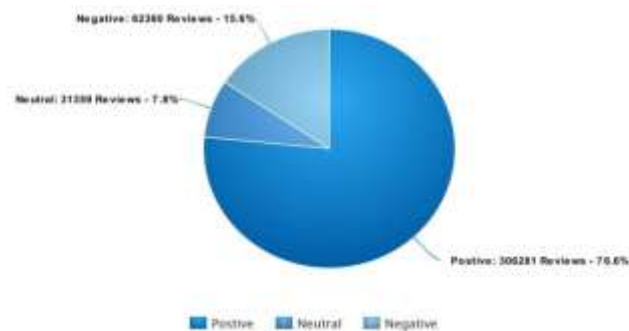


Fig. 3. Sentiment classification of star ratings

Figure 3 illustrates the sentiment analysis of product review ratings, where positive reviews constitute the majority at 76.6% (306,281 reviews). Neutral reviews represent 7.8%

(31,359 reviews), whereas negative reviews account for 15.6% (62,360 reviews). The distribution demonstrates a clear skew towards positive sentiment, indicating a predominantly favorable consumer response.

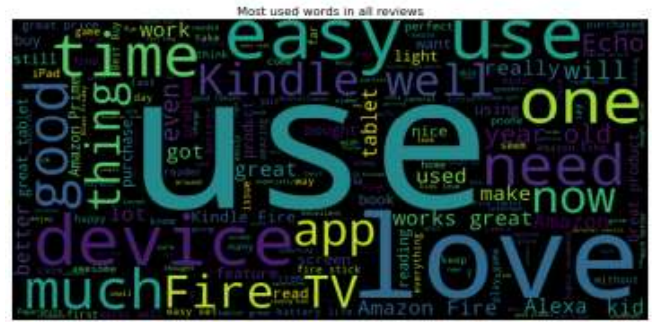


Fig. 4. Word cloud view of most used words in all reviews

Figure 4 shows a word cloud that displays the most common words used in all product reviews. Prominent terms such as *use*, *love*, *device*, *easy*, *time*, *app*, and *Fire* appear in larger font sizes, indicating their higher frequency of occurrence. These terms reflect recurring themes in customer feedback, emphasizing usability, satisfaction, and product-related features.

B. Data pre-processing

A number of critical actions were performed during data preprocessing in order to get the Amazon review dataset ready for sentiment analysis. Data cleansing, integration, transformation, and reduction are all part of it, with the goal of making the dataset better so that analysis can be more successful. every one of the procedures used to prepare the data for this study.

1) Handling Missing Values

Considering the importance of the "review body" and "star rating" characteristics in sentiment analysis, they were given priority for resolving missing values. utilised Python's `fillna()` method to populate null values for features with an object data type. used the Interpolate approach, which looks at the average of nearby values to fill in missing entries, for the "star rating" function. This ensured data completeness without introducing bias.

2) Lowercase Conversion

To keep things consistent, all review sentences were changed to lowercase. One example is the change from "aMazIng" to "great" and "amazing" from "Great." Removing case sensitivity from word processing and ensuring that differences in capitalisation do not impact analysis results is an important step in reducing data complexity.

3) Removal of Stop Words

The corpus was cleaned up by removing HTML tags, stop words, and punctuation. For the majority of natural language processing (NLP) tasks, common terms like "the," "is," and "and" don't have much semantic weight. Getting rid of them makes the sentiment analysis model more efficient and accurate by reducing noise.

4) Stemming

The process of reducing inflected words and expressions to their basic or root forms Stemming is the standard form for written phrases. One example is the transformation of "cleaned" into "clean."

5) Tokenization

The term "tokenization" refers to the act of dividing a string of text into a collection of discrete pieces of information [21]. Tokens are small pieces of text that form larger chunks. Words, sentences, paragraphs, and the like can be included. Ngrams is a tokenization method that we've employed in investigation. Another definition of an n-gram is a continuous string of n elements drawn from a text or audio sample. Phonemes, syllables, letters, words, or base pairs are the objects that are often included in the appliance.

6) Feature Extraction using TF-IDF

The TF-IDF is a phrase originality metric that compares the number of occurrences of a phrase in an extremely report to the number of files that include that term [22]. According to equations 1 and 2, if the TFIDF weight of terms in the material is high, the content always rank highly in search results, making it easy for everyone to find. Consequently, everyone can find phrases with lower competition and better search volumes without worrying about stop-words.

$$TF(t) = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}} \quad (1)$$

$$IDF = \log_e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right) \quad (2)$$

C. Data Splitting

Utilising stratified sampling to guarantee uniform class distribution, the dataset was divided in half: 80% for training and 20% for testing.

D. Classification with BERT model

A supervision-free pre-training paradigm for natural language processing, BERT stands for Bidirectional Encoder Representations from Transformers [23]. The BERT framework is illustrated in Fig. 5 [24]. Model pre-training and model fine-tuning are the two phases of the model. The model's fine-tuning feature was utilized for classification in this research, whereas it is often employed for masked language modeling objectives and next sentence prediction. To understand the contextual relations between words and tokens (Tok1.....TokM) in text, the BERT employs the bidirectional training of transformer. A different mechanism called an encoder (E1..., EN) reads the text input into the transformer, and a separate mechanism called a decoder creates predictions for the task at hand [25]. Transform encoders (T1...TN) read the entire bidirectional string of words all at once, unlike RNNs and LSTMs, which read the input text sequentially (from left to right or right to left). The machine learning community is abuzz with the publishing of state-of-the-art findings in various NLP (natural language processing) tasks, among others.

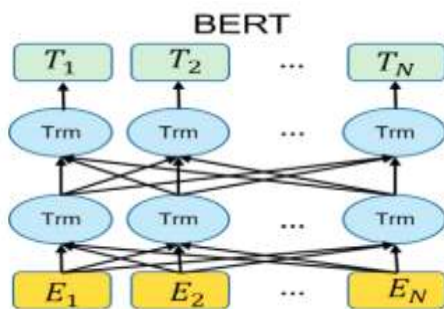


Fig. 5. The framework for BERT

The "masked language model" (MLM) pre-training method for natural language processing has problems in both directions. BERT [26] fixes these problems. Masking some input tokens at random allows the masked language model to learn the original vocabulary of a word based on its context. By using this two-way training methodology, BERT is able to gain a more nuanced understanding of linguistic context and master word meaning in context.

E. Performance Matrix

Precision, accuracy, F-score, and recall are some of the measures used to evaluate categorisation algorithms. When testing the efficacy of supervised DL algorithms, these parameters are priceless [27]. The confusion matrix is trusted for its accurate breakdown of TPS, TND, FN, and FG, which enables a comprehensive evaluation of the model's performance.

- **True Positive (TP):** This describes situations where the classifier maintains a good evaluation.
- **False Positive (FP):** A classifier produces a false positive when it mislabels a positive perspective as a negative one.
- **False Negative (FN):** Incorrect positive labelling of negative evaluations as positive by a classifier is known as False Negatives.
- **True Negatives (TN):** A review is considered "True Negative" if the classifier has correctly identified it as being negative.

1) Accuracy

The correctness of a model is defined as its probability of being correct; the formula for this likelihood is given by Eq. (3), which takes into consideration the sum of the True Positive and True Negative values in the numerator and all elements of the confusion matrix in the denominator:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (3)$$

2) Precision

A classifier's precision, or the percentage of correct return documents, is known as its precision [28]. When the accuracy is high, the number of false positives is low, and vice versa when the precision is low. P, or precision, is the percentage of instances that were properly classified out of the total. As an equation, it looks like this: (4):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

3) Recall

A high score indicates that the model correctly detects every positive case in the dataset. It entails dividing the sum of all positive predictions (including TP and FN) by the number of actual positive predictions (TP/FN). within the context of equity (5):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

4) F1_Score

The F1-Score is a weighted mean of the recall and accuracy scores; a perfect score is 1 and a poor score is 0. In order to assess the algorithm's success, as demonstrated in Eq. (8), it is necessary to determine the correlation between the number of correct predictions and the total number of predictions. These results prove that these metrics are reliable:

$$F1 - \text{score} = 2 * \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (6)$$

F. LIME

This model-neutral By offering interpretable explanations for complex black-box models, the Local Interpretable Model-agnostic Explanations (LIME) technique aids users in comprehending the decision-making process. This approach shows how accurately a model's predictions correspond to the needs of a given position [29]. It is especially helpful when comprehending the reasoning behind a model's choices is just as crucial as knowing how accurate its outcomes are. In Eq. 7, the goal of the LIME methodology is to identify a readily comprehensible model, denoted as $\hat{g}$, from a given set of models G . The objective is to lessen the loss L , which measures the disparity between the model g 's predictions and the more complicated model f , taking into account the impact of the locality kernel π_x . On top of that, the complexity of the model g is represented by $\Omega(g)$, which shows that simpler models are preferred for reasons like better interpretability and simplicity.

$$\hat{g} = \arg \min_{g \in G} \mathcal{L}(f, g, \pi_x) + \Omega(g) \quad (7)$$

IV. RESULTS AND DISCUSSION

Deep learning algorithms were used to get these outcomes. An Intel Core i7, a 7th-generation computer running Windows, was used to conduct the experiments. Jupiter notebook and Python version 3+ were utilised. The BERT model's performance indicators for e-commerce personalised product suggestion are displayed in Table 2, and all evaluation measures demonstrate consistently high values. A precision indicates that the model can generate highly relevant suggestions with few false positives, and an accuracy of 89 shows that it has good overall prediction performance. A recall demonstrates its effectiveness in correctly identifying relevant products without missing many true positives, and the F1-score confirms a well-balanced trade-off between precision and recall.

TABLE II. PARAMETERS PERFORMANCE BERT MODEL ON AMAZON PRODUCT REVIEWS DATASET

Measures	BERT
Accuracy	89
Precision	88
Recall	89
F1-score	88

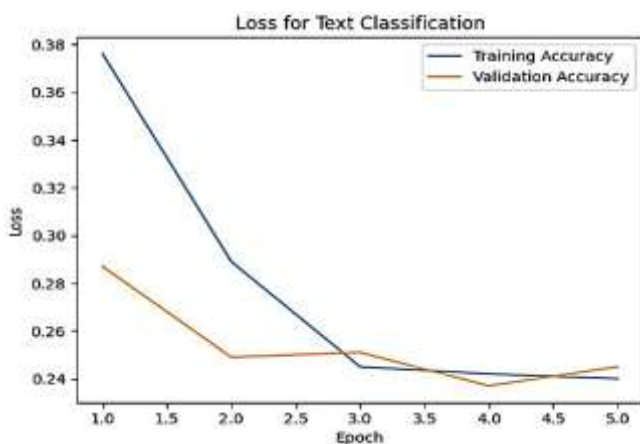


Fig. 6. Loss curve of BERT model

The BERT model's loss curve for text classification across five epochs is shown in Figure 6, which also shows patterns in

validation loss. A somewhat large training loss (around 0.38 at the outset) that gradually drops with each iteration shows that the model is picking up useful information from the data. Starting at around 0.28 and stabilising near 0.24 after the third epoch, the validation loss similarly follows a similar declining trend. The convergence of training and validation loss values in later epochs suggests that the model achieves good generalization without significant overfitting, as the gap between the two curves remains small throughout the training process.

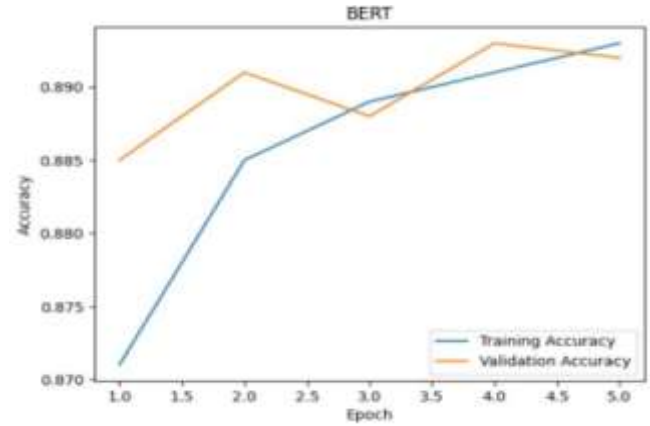


Fig. 7. Accuracy curve of BERT model

Figure 7 presents the accuracy curve of the BERT model across training and validation datasets over five epochs. The training accuracy shows a steady upward trend, starting around 87% in the first epoch and reaching above 89% by the fifth epoch, indicating consistent learning and improvement. Validation accuracy remains slightly higher than training accuracy throughout, fluctuating mildly but maintaining a generally upward trajectory, peaking near the fourth epoch.

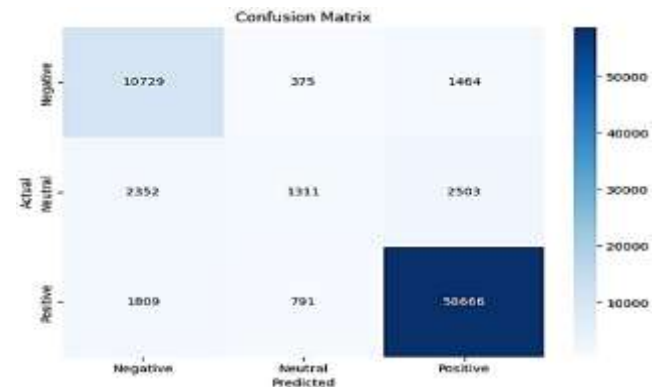


Fig. 8. Confusion matrix of BERT model

Figure 8 presents the confusion matrix of the BERT model, illustrating its classification performance across Negative, Neutral, and Positive sentiment categories. The model correctly predicted 10,729 Negative, 1,311 Neutral, and 58,666 Positive instances, showing particularly strong performance in the Positive class. However, there are notable misclassifications, such as 2,352 Neutral instances labeled as Negative and 2,503 Neutral instances labeled as Positive, indicating challenges in accurately identifying Neutral sentiments. Additionally, some Negative and Positive samples are confused with each other, though these errors are relatively smaller compared to Neutral misclassifications.

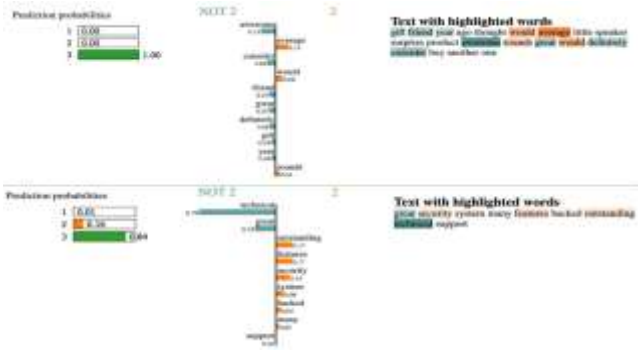


Fig. 9. LIME visualization for positive reviews

Figure 9 illustrates the LIME visualization for positive reviews, highlighting how different words contribute to the model's predictions. The bar plots on the left show the prediction probabilities across classes, while the highlighted text on the right identifies words that strongly influence classification. Positive sentiment words such as *awesome*, *sounds great*, *outstanding*, and *security* are emphasized in green, indicating their significant role in driving the prediction toward a positive class. Conversely, words like *average* and *technical* are marked in orange, showing a weaker or contrasting influence. This visualization effectively explains the decision-making process of the model by linking prediction outcomes with specific textual features, thereby enhancing interpretability and demonstrating the model's ability to capture meaningful cues from customer reviews.

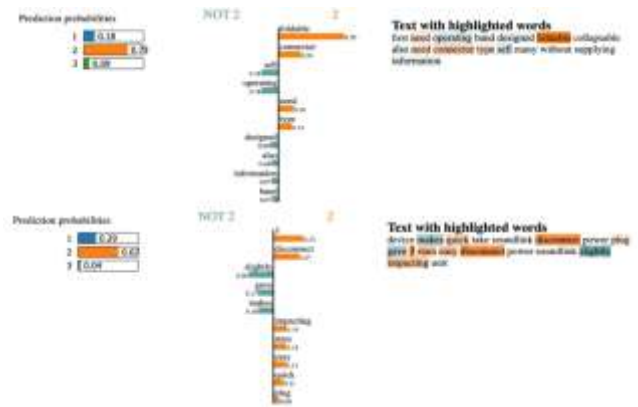


Fig. 10. LIME visualization for neutral reviews

Figure 10 LIME visualization for neutral reviews illustrates how the model interprets and classifies customer feedback by identifying influential words that contribute to neutral sentiment. In the first example, the review is predicted as class "2," where terms like *foldable* and *connector* strongly drive the classification towards neutrality, while words such as *operating* and *sell* exert a weaker influence towards the opposite class. In the second example, also classified as "2," the words *disconnect*, *makes*, and *slightly* emerge as key contributors, with *disconnect* having the strongest impact on the prediction outcome. These LIME visualizations provide interpretability by clearly indicating which textual features are shaping the model's decision, thereby validating the reasoning process and offering insights into how specific words influence neutral sentiment detection.

The model's decision-making process for negative product reviews is illustrated in Figure 11 by two cases where LIME (Local Interpretable Model-agnostic Explanations) is used. In the first case, the review is classified into class "1" with a very

high probability, primarily influenced by negative terms such as *broke* and *within first week*, which significantly contribute to the negative sentiment classification. The second case, also classified as class "1" with high confidence, presents a more nuanced explanation, where strongly negative words such as *horrible* and *chipped* dominate the prediction, while comparatively neutral or positive terms like *decent* and *favor* provide opposing influence, slightly mitigating the negative classification.

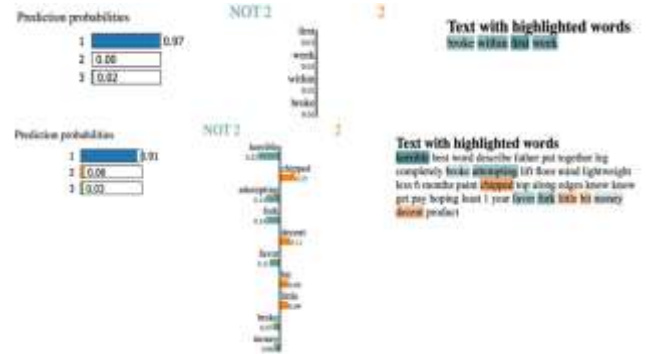


Fig. 11. LIME visualization for negative reviews

A. Comparative analysis

A comparison of current methods with the suggested BERT-based method for online product suggestion is shown in Table 3. found that the suggested BERT model beat more conventional approaches like CNN, VADER, and GRU on every single metric. In particular, BERT's exceptional performance in achieving the best possible accuracy (89%), precision (88%), recall (89%), and F1-score (88%), reveals its remarkable resilience in acquiring contextual and semantic data. In contrast, while CNN and GRU show moderate performance with accuracies of 84.9% and 81.82% respectively, VADER performs significantly lower, particularly in recall (47%) and F1-score (41%), highlighting its limitations in handling complex recommendation tasks.

TABLE III. PERFORMANCE COMPARISON BETWEEN EXISTING AND PROPOSED MODEL FOR PRODUCT RECOMMENDATION IN E-COMMERCE

Matrix	BERT	CNN[30]	VADER[31]	GRU[32]
Accuracy	89	84.9	46	81.82
Precision	88	82.1	50	59.45
Recall	89	84.9	47	42.52
F1-score	88	83.5	41	-

The proposed BERT-based model demonstrates clear advantages over traditional approaches by leveraging its deep contextual understanding of text to generate more accurate and personalized product recommendations. Unlike conventional models that rely heavily on shallow features or sequential dependencies, BERT effectively captures semantic nuances, user preferences, and product attributes within the recommendation process. This leads to improved adaptability, stronger generalization to diverse datasets, and enhanced precision in predicting user interests. As a result, the BERT-based model ensures highly relevant and reliable recommendations, making it particularly suitable for dynamic and large-scale e-commerce environments.

V. CONCLUSION AND FUTURE SCOPE

E-commerce sites offer a huge variety of goods and services, which has changed the way people shop. Customers may experience decision fatigue due to the overwhelming

number of alternatives. To combat this, recommendation systems are vital in improving the browsing experience and increasing sales by screening and proposing products that match user tastes. This study demonstrates that the proposed BERT-based personalized product recommendation system effectively leverages sentiment analysis of Amazon product reviews to deliver accurate and relevant recommendations across diverse product categories. By employing comprehensive preprocessing techniques and exploiting BERT ability to capture long-term dependencies, the model achieved superior performance compared to CNN, VADER AND GRU baselines, with high accuracy, balanced precision-recall scores. These results highlight the potential of DL-driven sentiment analysis for enhancing user satisfaction and engagement in e-commerce platforms.

Future research may focus on extending the BERT-based recommendation framework to multi-modal learning by incorporating product images, videos, and structured metadata alongside textual reviews for richer feature representation. Additionally, efforts can be directed towards addressing challenges such as the cold-start problem for new users and products, enhancing interpretability through explainable AI techniques, and adapting the system for real-time recommendations in dynamic environments. Scalability remains another important area, where integrating distributed training and big data platforms can enable deployment on large-scale e-commerce systems.

REFERENCES

- [1] X. Li, X. Sun, Z. Xu, and Y. Zhou, "Explainable Sentence-Level Sentiment Analysis for Amazon Product Reviews," in *Proceedings - 2021 5th International Conference on Imaging, Signal Processing and Communications, ICISPC 2021*, 2021. doi: 10.1109/ICISPC53419.2021.00024.
- [2] V. Verma, "The Role of Data Migration in Modern Business Intelligence Systems," *Int. J. Res. Anal. Rev.*, vol. 11, no. 2, pp. 1–11, 2024.
- [3] K. M. R. Seetharaman and S. Pandya, "Leveraging AI And IoT Technologies For Demand Forecasting in Modern Supply," *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 6, pp. 66–73, 2024, [Online]. Available: <https://ijrtsm.com/wp-content/uploads/2025/05/June-2024-Karthika-66-73.pdf>
- [4] L. Liu, "e-Commerce Personalized Recommendation Based on Machine Learning Technology," *Mob. Inf. Syst.*, vol. 2022, pp. 1–11, Apr. 2022, doi: 10.1155/2022/1761579.
- [5] R. Patel and P. Patel, "A Machine Learning-Based Detection and Recognition Approach for ID Barcode Scanning Through Visible Light Communication," *Int. J. Res. Anal. Rev.*, vol. 8, no. 2, 2021, doi: 10.56975/ijrar.v8i2.315834.
- [6] D.-N. Nguyen, V.-H. Nguyen, T. Trinh, T. Ho, and H.-S. Le, "A personalized product recommendation model in e-commerce based on retrieval strategy," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 2, p. 100303, Jun. 2024, doi: 10.1016/j.joitmc.2024.100303.
- [7] V. Verma, "Optimizing Database Performance For Big Data Analytics And Business Intelligence," *Int. J. Eng. Sci. Math.*, vol. 13, no. 11, pp. 56–75, 2024.
- [8] H. Kali, "Diversity , Equity , and Inclusion Analytics in HR : How Workday Enables Data-Driven Decision-Making," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 2, pp. 162–170, 2023.
- [9] L. Kang and Y. Wang, "Efficient and accurate personalized product recommendations through frequent item set mining fusion algorithm," *Heliyon*, vol. 10, no. 3, p. e25044, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25044.
- [10] A. R. Bilipelli, "End-to-End Predictive Analytics Pipeline of Sales Forecasting in Python for Business Decision Support Systems," *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 1–9, 2022.
- [11] G. Maddali, "Enhancing Database Architectures with Artificial Intelligence (AI)," *Int. J. Sci. Res. Sci. Technol.*, vol. 12, no. 3, pp. 296–308, May 2025, doi: 10.32628/IJSRST2512331.
- [12] C. Li, I. Ishak, H. Ibrahim, M. Zolkepli, F. Sidi, and C. Li, "Deep Learning-Based Recommendation System: Systematic Review and Classification," *IEEE Access*, vol. 11, pp. 113790–113835, 2023, doi: 10.1109/ACCESS.2023.3323353.
- [13] G. M. and H. Kali, "Exploring Big Data Role in Modern Business Strategies: A Survey with Techniques and Tools," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, pp. 1–11, 2023.
- [14] H. Chen, H. Chen, Y. Wang, and J. Li, "Research on Optimization of Personalized Algorithm of E-Commerce Recommendation System Based on Deep Learning," in *2025 International Conference on Digital Analysis and Processing, Intelligent Computation (DAPIC)*, IEEE, Feb. 2025, pp. 591–595. doi: 10.1109/DAPIC66097.2025.00115.
- [15] S. P. B and S. K. M., "OSE: Optimizing User Segmentation in E-Commerce Using APRIORI Algorithm for Personalized Product Recommendations," in *2025 International Conference on Visual Analytics and Data Visualization (ICVADV)*, IEEE, Mar. 2025, pp. 1376–1383. doi: 10.1109/ICVADV63329.2025.10961010.
- [16] L. Li, "Research on Personalized Recommendation System for E-Commerce Products Based on Collaborative Filtering Algorithm," in *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, IEEE, Feb. 2024, pp. 876–880. doi: 10.1109/EEBDA60612.2024.10485710.
- [17] J. Li, "Construction of E-Commerce User Profile Design Model Based on Deep Learning Algorithm," in *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, IEEE, Dec. 2024, pp. 1–5. doi: 10.1109/ICMNWC63764.2024.10872316.
- [18] J. Lee and J. Kim, "Developing a Convenience Store Product Recommendation System through Store-Based Collaborative Filtering," *Appl. Sci.*, vol. 13, no. 20, p. 11231, Oct. 2023, doi: 10.3390/app132011231.
- [19] J. Shi, "E-Commerce Products Personalized Recommendation Based on Deep Learning," in *2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT)*, IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/ACAIT56212.2022.10137959.
- [20] Q. Ge, "E-Commerce Personalized Recommendation Based on Convolutional Neural Network," in *2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT)*, IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/ACAIT56212.2022.10137920.
- [21] B. Kamalesh and S. Vijayalakshmi, "Sentiment Analysis on Amazon Product Review," in *Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022*, 2022, pp. 846–848. doi: 10.1109/ICAC3N56670.2022.10074287.
- [22] J. C. Gope, T. Tabassum, M. M. Mabur, K. Yu, and M. Arifuzzaman, "Sentiment Analysis of Amazon Product Reviews Using Machine Learning and Deep Learning Models," in *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, IEEE, Feb. 2022, pp. 1–6. doi: 10.1109/ICAEEE54957.2022.9836420.
- [23] X. Zhao and Y. Sun, "Amazon Fine Food Reviews with BERT Model," *Procedia Comput. Sci.*, vol. 208, pp. 401–406, 2022, doi: 10.1016/j.procs.2022.10.056.
- [24] 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 (ICSSSES)*, 2025, pp. 1–6. doi: 10.1109/ICSSSES64899.2025.11009528.
- [25] 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.
- [26] Y. Liu, J. Lu, J. Yang, and F. Mao, "Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax," *Math. Biosci. Eng.*, vol. 17, pp. 7819–7837, 2020, doi: 10.3934/mbe.2020398.
- [27] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, p. 100267, Nov. 2024, doi:

- 10.1016/j.jjime.2024.100267.
- [28] S. Johar and S. Mubeen, "Sentiment Analysis on Large Scale Amazon Product Reviews," *Int. J. Sci. Res. Comput. Sci. Eng.*, vol. 8, no. 1, pp. 7–15, 2020, doi: 10.26438/ijsrcse/v8i1.715.
- [29] E. Hashmi and S. Y. Yayilgan, "A robust hybrid approach with product context-aware learning and explainable AI for sentiment analysis in Amazon user reviews," *Electron. Commer. Res.*, Aug. 2024, doi: 10.1007/s10660-024-09896-5.
- [30] H. Ali, E. Hashmi, S. Yayilgan Yildirim, and S. Shaikh, "Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques," *Electron.*, vol. 13, no. 7, pp. 1–21, 2024, doi: 10.3390/electronics13071305.
- [31] O. Shobayo, S. Sasikumar, S. Makkar, and O. Okoyeigbo, "Customer Sentiments in Product Reviews: A Comparative Study with GooglePaLM," *Analytics*, vol. 3, no. 2, pp. 241–254, Jun. 2024, doi: 10.3390/analytics3020014.
- [32] N. Shrestha and F. Nasoz, "Deep Learning Sentiment Analysis of Amazon.Com Reviews and Ratings," *Int. J. Soft Comput. Artif. Intell. Appl.*, vol. 8, no. 1, pp. 01-15, 2019, doi: 10.5121/ijscai.2019.8101.