Kelvin Open Science Publishers
Connect with Research Community

Research Article *Volume 1 | Issue 2*

KOS Journal of AIML, Data Science, and Robotics
https://kelvinpublishers.com/journals/aiml-data-science-robotics.php

Enhancing Customer Experience through ML-Based Chatbots in E-Commerce Platforms

Pradyumna Kumar*

Independent Researcher, India

***Corresponding author:** Pradyumna Kumar, Independent Researcher, India, E-mail: pradyumnak292@gmail.com

Received: December 02, 2025; **Accepted:** December 24, 2025; **Published:** December 26, 2025

Citation: Pradyumna Kumar. (2025) Enhancing Customer Experience through ML-Based Chatbots in E-Commerce Platforms. *KOS J AIML, Data Sci, Robot.* 1(2): 1-7.

1. Abstract

The extensive proliferation of e-commerce systems has seen a tremendous surge in consumer reviews that have given valuable information on customer satisfaction levels as well as product performance. This paper covers the increasing demand of smart and receptive chatbots in the e-commerce realm to ameliorate customer experience by properly interpreting user queries and moods. The issue is that it is difficult to interpret subtle customer reviews and create text answers in a given context with traditional machine learning methods. The main objective is to design and test chatbots based on machine learning that can provide stable intent recognition and sentiment analysis during the e-commerce conversation. Based on the Amazon Reviews dataset available at Kaggle, data were preprocessed then a class balancing technique was applied to ADASYN. Feature extraction employed TF-IDF, and models including BERT and Random Forest (RF) were trained and evaluated using metrics like “accuracy (acc), “precision (prec), ROC curves, “F1-score” (f-measure), confusion matrices, and “recall” (rec). Findings show that BERT model was superior to RF, with an accuracy of 94.87 and AUC of 0.9604, which requires better context-awareness, whilst RF obtained 90.72 and 0.9198 accuracy and AUC respectively. The proposed BERT has a better accuracy and robustness as compared to the previous transformer-based methods, such as FK-BERT, LSTM, SVM, and RoBERTa-A, which are revealed by a comparative analysis. The importance of the study is that it aims to create the best reliable frameworks of chatbots that can be implemented in e-commerce platforms to interact with customers better, understand queries accurately, and respond to them intelligently in a more effective and efficient manner compared to traditional frameworks.

2. Keywords

E-Commerce, Chatbots, Natural Language Processing (NLP), Sentiment Analysis, Customer Experience, Machine Learning, Amazon Reviews Dataset

3. Introduction

The boom of e-commerce witnessed worldwide throughout the last ten years has seen the world of e-commerce businesses grow in a manner never experienced before and this has changed the shopping behavior of consumers and their interaction with the brands in a significant way [1]. It is estimated that online sales will amount to trillions of dollars across the globe and customer experience (CX) is a significant factor in ensuring that an organization has a competitive edge [2]. Consumers in the modern world require fast, individualized and immediate communication, and they tend to prefer those platforms that have real-time customer support and customized suggestions [3]. Delays or poor quality of service result in abandoned carts since the customers find it frustrating and as a source of low brand loyalty [4,5]. In this stiff market race [6], organizations are becoming quick to resort to innovative technology solutions in order to attain customer engagement and high level of customer satisfaction [7,8].

These demands are limiting conventional customer service systems e.g. emails, telephone support and fixed FAQs. They are also slow and costly, and are not scalable when there is high traffic and offer canned answers which cannot cater to particular customer needs [9,10]. Growth in scale and user base diversifying on e-commerce solutions introduce these inefficiencies and, therefore, the necessity to substitute them with smart, flexible, and scalable solutions. Chatbots, in particular, the ones that operate artificial intelligence (AI) have hence offered a viable means of enhancing customer support in turn. Machine learning (ML)-based chatbots can be compared to rule-based systems, which are characterized by prescriptive dialogues [11,12], use data-driven models to process user intent, solve advanced queries and provide context-sensitive, natural user interfaces, therefore improving operational efficiency and customer satisfaction[13-15].

Despite all the potential of the ML-based chatbots, multiple problems as well as research gaps that need to be addressed still exist, and it is not an easy task to correctly process the customer responses that are subtle, extract the sentiment and give consistent answers under different circumstances of e-commerce. The literature is typically split between standard ML models [16] and isolated NLP [17] models, and LLM models [18,19], despite a lack of consideration of unified models with continuous learning and adaptation. The necessity to fill these gaps with the creation of an integrated framework of the chatbot based on ML motivates this research so that the intent recognition, sentiment analysis, and individual response generation could be improved. By solving these issues, the research will enhance automated customer service, culturally interact, and offer scalable solutions to satisfy the changing needs of the contemporary e-commerce platforms.

- The paper suggests an extensive architecture of e-commerce chatbots that incorporate methods of machine learning to comprehend customer requests and provide context-sensitive replies.
- It follows systematic preprocessing procedures, such as normalization, tokenization, lemmatization, and class balancing with ADASYN, to facilitate high quality input in model training and minimize bias during classification.
- The study conducts an evaluation of the efficacy of transformer-based architecture such as BERT compared to traditional ML models like RF, and identifies the relative strengths and weaknesses of these models in intent recognition.
- The use of TF-IDF characteristics and the transformer-based representations enhances the capabilities of the chatbot to perceive the subtle customer emotions and deliver correct and context-related answers.
- The framework provides a scalable and trusted way of deploying chatbots in e-commerce sites that can be used to develop actionable recommendations to enhance user interaction and automated customer support systems.

The work is noteworthy because it contributes to the further evolution of smart context-sensitive chatbots in e-commerce sites whose task is to properly interpret the subtle reviews and requests. It is innovative in that it makes use of transformer-based models like BERT with efficient preprocessing and class-balancing methods, which makes it capable of better sentiment and intent recognition than traditional methods. A mixture of advanced NLP models and the practical implementation of chatbots, the current study provides a scalable, robust framework that will promote improved customer service, boost user interaction, and serve as a framework guide to the research on topic of AI-based chatbots in future.

3.1. Structure of the paper

The structure of the paper is as follows: Section II covers the literature addressing similar topics as the study such as ML-based chatbots and customer experience within e-commerce websites. Section III presents the methodology, datasets, preprocessing and model development. Section IV presents experimental results and the results assessment. Section V outlines the major findings and suggestions regarding further research.

4. Literature Review

This section incorporates surveys and critiques of publications on predictive modeling for enhancing customer experience in e-commerce platforms through machine learning-based chatbot frameworks.

Pande, et al. (2025) developed features, along with LIWC-22-based characteristics, are used to categorize feedback as positive or negative. The XGBoost model and Bidirectional Long Short-Term Memory (CapsNet-BiLSTM) and Capsule Network model are used for the categorization. This study has achieved the highest scores with 98.39% acc, 98.81% prec, 98.01% for rec, and 98.41% f1-score, the results demonstrate that suggested method is highly accurate and delivers superior performance [20].

Sharma and Mishra (2024) broad dataset, including records of 50,000 customer interactions collected from different e-commerce platforms along with customer satisfaction rating, response time, and resolution success. Response speed, customer happiness, resolution rate, engagement rate, and cost savings were among the performance metrics used to assess the chatbot after it was taught using a supervised learning technique. The results show that the chatbot achieved an average response time of 1.5 seconds, an engagement rate of 78.5%, a resolution rate of 92.3%, an average customer satisfaction rating of 4.6 out of 5, and a monthly cost savings of $45,000 [21].

Wei, (2024) uses RF (Random Forest) as a tool to study and implement a user behavior scrutiny system based on online shopping platforms. The feasibility of using RF algorithm for user behavior analysis on e-commerce platforms is verified through experiments (when the number of decisions reached 150, the accuracy of RF algorithm was 0.875, higher than other algorithms) [22.

Bawono, Purwitasari, and Purnomo, (2023) revealed that the algorithm's first accuracy score of 0.37 was modest. In the meantime, the accuracy considerably improved to 0.62 following the use of hyperparameter adjustment. A standardized Normed Fit Index (NFI) of 0.707 was obtained in the Smart PLS questionnaire analysis aspect, which was just below the predefined limit of 0.90. The SRMR, or standardized root mean square residual, it demonstrated a successful model fit and was measured at 0.071, falling below specified value of 0.08. But at 0.240, the RMS theta value was higher than the 0.10 cutoff [23].

Zulfadli, Ilham, and Indrabayu (2023) integrated the probability values of three classifiers SVM (Support Vector Machine), RF (Random Forest), and Naive Bayes (NB) for each sentiment category (positive, neutral, and negative). Dataset of Tokopedia product reviews is used for the evaluation. The results show that SS (Sentiment Selector) method is less effective than the SV (Soft Voting) model. The suggested SV model attains a 69% acc, 70% prec, 69% recall, and 69% F-measure, in that order [24].

Table I provides a comparative scrutiny of previous studies on ML-based chatbots in e-commerce, highlighting datasets used, findings, limitations, and directions for recommendations.

Table 2: Comparative review of ML approaches for customer experience in e-commerce sector.

Reference	**Methods**	**Results**	**Advantages**	**Limitations**	**Recommendation**
Pande et al. (2025)	CapsNet-BiLSTM, XGBoost; features with LIWC-22-based characteristics	98.39% accuracy, 98.81% precision, 98.01% recall, and 98.41% F1-score	High accuracy, superior performance in sentiment classification, effective consumer behavior forecasting	May require high computational resources due to CapsNet-BiLSTM complexity	Suitable for applications requiring highly accurate sentiment analysis and consumer behavior prediction
Sharma & Mishra (2024)	Supervised learning-based chatbot trained on 50,000 e-commerce interactions	Response time: 1.5s, Resolution rate: 92.3%, Customer satisfaction: 4.6/5, Engagement: 78.5%, Cost savings: $45,000/month	Fast response, high resolution rate, improved customer satisfaction, significant cost savings	Dependent on labeled data quality and coverage	Ideal for large-scale e-commerce platforms seeking efficient automated customer support
Wei (2024)	Random Forest for user behavior analysis on e-commerce platforms	Accuracy: 0.875 (for 150 decision trees)	High feasibility for user behavior prediction, outperforms other algorithms	Accuracy sensitive to number of decision trees; may not capture sequential patterns	RF suitable for scalable, interpretable user behavior analysis in e-commerce
Bawono, Purwitasari & Purnomo (2023)	Algorithm with hyperparameter tuning; Smart PLS questionnaire analysis	Accuracy improved from 0.37 → 0.62; NFI: 0.707, SRMR: 0.071, RMS theta: 0.240	Hyperparameter tuning improves accuracy; SRMR indicates good model fit	NFI below threshold; RMS theta above threshold; modest initial performance	Recommend hyperparameter tuning and further model refinement for better fit
Zulfadli, Ilham & Indrabayu (2023)	Soft Voting (SVM, RF, NB) vs Sentiment Selector (SS) on Tokopedia reviews	Soft Voting: Accuracy 69%, Precision 70%, Recall 69%, F1 69%	Ensemble approach improves performance over single classifiers	Overall performance moderate; lower than deep learning methods	Soft Voting suitable for moderate-performance requirements where simplicity and interpretability are preferred

4.1. Research gaps

A number of papers have researched on machine learning techniques to improve e-commerce customer service and consumer feedback. The fields of customer sentiment classification, behavior analysis and chatbot optimization have previously been examined using the latest features and models of NLP (Capsule Networks with BiLSTM) and other (Random Forest, XGBoost and ensemble techniques). Although these approaches succeeded to different extents of accuracy and efficiency, issues of capturing subtle contextual semantics, dealing with imbalanced and large-scale data, and offering flexible and real-time responses in different e-commerce contexts still exist. Most of them are aimed at sentiment classification or simple analysis of user behavior without an integrated system, which is a combination of transformer-based models with strong preprocessing, class balancing, and continuous learning. The proposed work is addressing these gaps by means of a finetuned BERT model, as well as extensive preprocessing and ADASYN-based balancing to enhance intent recognition, sentiment analysis, and chatbot-based response generation, which is a more scalable, context-aware, and trusted solution to the improvement of customer experience on the e-commerce platform.

5. Methodology

This paper trains and tests chatbots based on AI to enhance customer experience in e-commerce. Using the Amazon Reviews dataset from Kaggle, data is preprocessed with case normalization, stopword removal, tokenization, and lemmatization, followed by class balancing via ADASYN. TF-IDF features are extracted, and dataset is partitioned into 80/20 for training and testing. BERT and Random Forest models are implemented, with performance evaluated using acc, prec, rec, f-measure, and confusion matrices. The framework effectively interprets customer reviews to enhance chatbot responses (**Figure 1**).

5.1. Data collection and analysis

The Amazon Reviews dataset on Kaggle contains 34.7 million reviews. Using review scores of 1 and 2 as negative and 4 and 5 as positive, the Amazon reviews polarity dataset is created.

Figure 2: Bar graph of Distribution of Review Polarities.

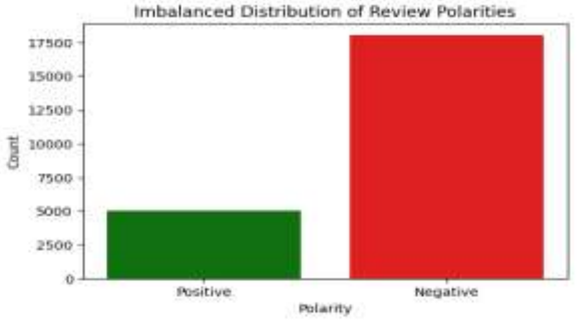


Figure 2 shows the bar graph of review polarities, highlighting a significant class imbalance with many more negative than positive reviews. The "Positive" polarity is represented by a green bar, indicating approximately 5,000 reviews. In contrast, the "Negative" polarity, shown in red, has a much larger count, exceeding 17,500 reviews.

Figure 3: Word Cloud of Customer Review Keywords.

Figure 3 is a word cloud generated from cleaned customer review text, displaying the most frequent and significant words such as “amazon”, “device”, “use”, “fire”, “read”, “kindle”, and "great." Size of each word is proportional to its frequency, with larger words indicating higher counts, providing a quick way to identify the main topics, product features, and common sentiments in the customer feedback.

5.2. Data preprocessing

Preprocessing of text is the first step in natural language processing processes. Preprocessing [25] is essential in order to make the textual data of Amazon product reviews useful and high-quality. At the first stage, case normalization is implemented, that is, all the text is converted to lower characters and this is used to reduce redundancy such that words such as Good and good are considered the same. Then stopword removal is used which means removing the words that are common and which are not informative like the, is, and which do not have any contribution to the sentiment or context of a review. The text is then tokenized into word or token and then it can count the frequency and patterns of words utilized. Last, lemmatization simplifies words to their base forms whereby words such as running and ran will be put down to run assisting in capturing any real meaning at the same time curtailing the feature space. These over procedures serve to increase the consistency, accuracy and understandability of Amazon review data prior to a further analysis.

5.3. Data balancing

Once the dataset has been preprocessed, it is analyzed to test whether there is a class imbalance, which can often occur in review data where certain sentiments might dominate. In order to overcome this, ADASYN (Adaptive Synthetic Sampling) is used. ADASYN creates synthetic minority classes through data distribution which is useful to balance data. This guarantees that the model will not be biased to majority classes and enhance predictive performance especially on the underrepresented sentiment fines.

Figure 4: Bar graph of Balanced Distribution of Review Polarities.

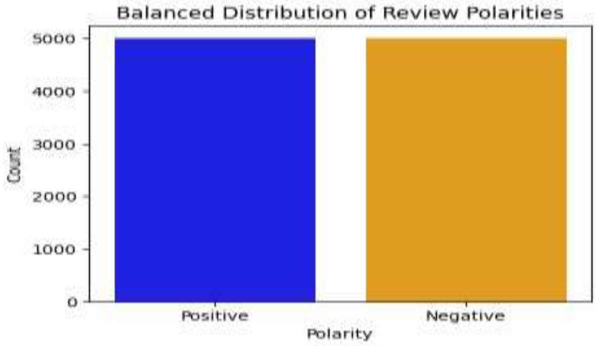


Figure 4: illustrates distribution polarities of reviews after using a balancing strategy. The balanced distribution of review polarities after ADASYN, with both Positive and Negative classes having 5,000 reviews each, ensuring a fair dataset for model training. This balance prevents model bias and improves accuracy in sentiment analysis.

5.4. Feature Extraction using TF-IDF for E-commerce reviews

Following preprocessing, textual input must be converted into numerical features which may be processed by ML models. A popular method for this in e-commerce review analysis is “TF-IDF” (Term Frequency-Inverse Document Frequency). A term's frequency of occurrence in a document (TF) is multiplied by the inverse of its frequency of occurrence in all documents (inverse document frequency, or IDF) to get TF-IDF [26]. Such scores attract the attention of terms in document that are unique and information laden. Equations (1) and (2) show that TF is ratio of number of terms in document to total number of terms. Alogarithm of total number of documents divided by number of documents containing the phrase, on other hand, is known as IDF. The relevance of a phrase is more accurately represented by the ensuing TF-IDF scores.

$$\mathrm{Tf}(w) = \frac{\text{Number of times word } w \text{ appears in a document}}{\text{Total number of terms in the document}} \quad (1)$$

$$\mathrm{Idf}(w) = log_e \frac{\text{Total number of documents}}{\text{Number of documents with term } w} \quad (2)$$

Term frequency and inverse document frequency work together to ensure that TF-IDF minimizes the weight of frequently occurring terms throughout the corpus in addition to capturing how frequently a word appears in a particular document.

5.5. Data splitting

There are several methods for separating data into test and training sets. One method is a fixed split, in which the test set receives the remaining data while a specific percentage of the data is assigned to the training set. As an illustration, the data can either be partitioned into 80/20, whereby 80 percent of the information is addressed during training and 20 percent forming part of testing. It used an 80/20 partitioning method in the proposed detection model.

5.6. Classification of machine learning models

E-commerce chatbots can also be enhanced by BERT and Random Forest (RF) by analyzing customer text in the context and intent.

5.6.1. Bidirectional encoder representations from transformers (BERT): BERT is a bidirectional model, this is what allows it to be able to comprehend the meaning of words within a sentence using both the words that the words precede and follow [27]. This is found to be powerful over the past language models which were only able to reflect the meaning of a word after consideration of the previous words [28]. BERT is trained over a tremendous volume of text and code and is suitable in a wide range of NLP applications, including question answers, sentiment extraction, NLP. In this work, they train BERT on the preprocessed TF-IDF feature set to learn about semantic subtleties in customer feedback. The major hyperparameters used in this implementation will be pretrained model of bert-base-uncased, a maximum of 128 tokens on the input scene with batch size set to 16 and learning rate at 2e-5. The training is done on 10 epochs on the AdamW optimizer with drop out of 0.1 percent on the classification layers to help avoid overfitting. The bidirectional attention mechanism of the BERT model gives it the ability to examine the context in both directions, and this aspect enables it to recognize sentiments in product reviews correctly.

5.6.2. Random forest (RF): RF stands for ensemble learning algorithm, which improves classification accuracy and generalization by combining decision trees. Only a bootstrap sample— Each decision tree in the ensemble is trained using a portion of the original data, where samples are generated by drawing at random with replacement. In this study, RF serves as an interpretable baseline model for sentiment classification of Amazon reviews. The key hyperparameters include n_estimators = 100 for number of trees, max_depth = None to allow trees to expand fully, min_samples_leaf = 1, min_samples_split = 2, and the gini criterion for evaluating split quality. Bootstrap sampling is enabled to create diverse trees from random subsets of the training data. RF efficiently handles the high-dimensional TF-IDF feature vectors while providing a robust classification baseline alongside BERT.

5.7. Model evaluation

It is pivotal to calculate specific performance metrics and visualize the confusion matrix in order to evaluate any data mining classification system. These will assist in evaluating each method's effectiveness and contrasting the results of other approaches. It shows how many samples are in each quadrant. It facilitates comprehension of the model's expected "False Positives" (FP), "True Positives" (TP), "False Negatives"(FN), and "True Negatives"(TN). As a result, it is useful in determining how well the model completed the categorization. Acc, prec, rec and f-measure are important assessment measures for stratification models defined as Equations (3) to (6).

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

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

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

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

The ratio of accurately predicted observations (counting TP and TN) to all predictions is how accuracy gauges a model's overall correctness. It could, however, be deceptive for unbalanced datasets. Among all projected positives, the percentage of genuine positives is known as precision. A model's recall is percentage of true positives that it properly identifies. The F1-score is a balanced metric based on a harmonic combination of precision and recall, which is especially useful, when unequal classes are distributed. A closer F1-score to 1 signifies an improved model performance. Also, ROC and AUC score graphically display a model's capacity to differentiate between classes at various relevant thresholds. serving to provide both a graphical and quantified assessment of classification performance.

5.8. GUI-based Chatbot deployment

The most efficient model will be incorporated into Graphical User Interface (GUI) to develop an interactive chatbot. The GUI facilitates end-user query input and query response in real-time simulating a live e-commerce customer support experience. This action shows applicability to the proposed framework in action and offers a friendly user environment to use machine learning-powered chatbots to improve customer interactions and conversational engagement.

6. Result Analysis and Discussion

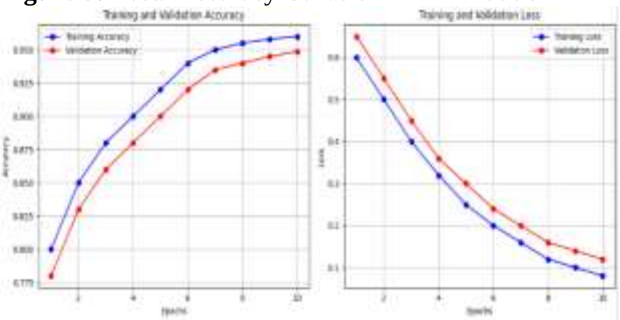
The study was conducted and tested within a controlled experimental setup on a computer with the core i5-8250U processor with the frequency of 1.8 GHz, 12GB RAM, and Windows Professional 64-bit. The ML and DL models were implemented in Python 3 as the development framework. On Amazon Reviews dataset, the evaluation was carried out through the provision of strong benchmark to intent recognition, sentiment analysis, and response generation. In order to provide thorough analysis, transformer-based and conventional models were used, as well as proposed BERT, Random Forest, FK-BERT, and RoBERTa-A. These models were included so that the comparative study could be taken with respect to the performance variations measured in the terms of acc, prec, rec, and f-measure to indicate the efficacy of current approach in ameliorating customer experience with the assist of ML-based chatbots within the e-commerce platforms.

Table 4: Performance Evaluation Of The Proposed Models For Ml-Based Chatbots In E-Commerce Platforms.

Measures	BERT	RF
Accuracy	94.87	90.72
Precision	96.21	92.54
Recall	93.73	87.94
F1-score	94.95	90.18

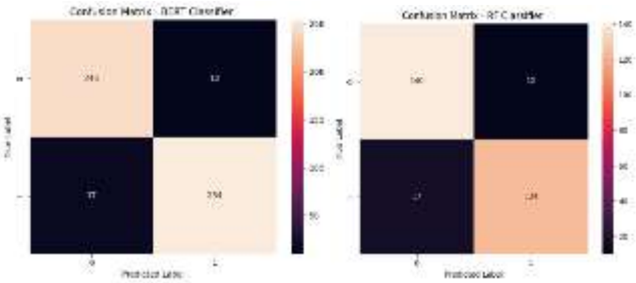
The discussion of the proposed models has revealed that it is effective in facilitating the e-commerce sales chatbot features. The BERT model did not perform poorly and had acc of 94.87%, prec of 96.21%, rec of 93.73%, and an f-measure of 94.95% shows in Table II, indicating that it could capture the semantics in context and produce a reliable intent classification. On the same note, Random Forest (RF) model did not perform poorly, as it recorded acc of 90.72%, prec of 92.54%, rec of 87.94%, and f-measure of 90.18%, thus demonstrating its strength as a baseline machine learning algorithm. In aggregate, these findings corroborate the predictive aptitude of the postulated models in the framework of chatbot-facilitated query understanding, response generation, and that transformer-based models are robust with respect to applicative expressiveness and the lesser traditional classifiers both show stable performance exhibits less contextual sensitivity.

Figure 5: Loss/ Accuracy Curve of BERT Model.



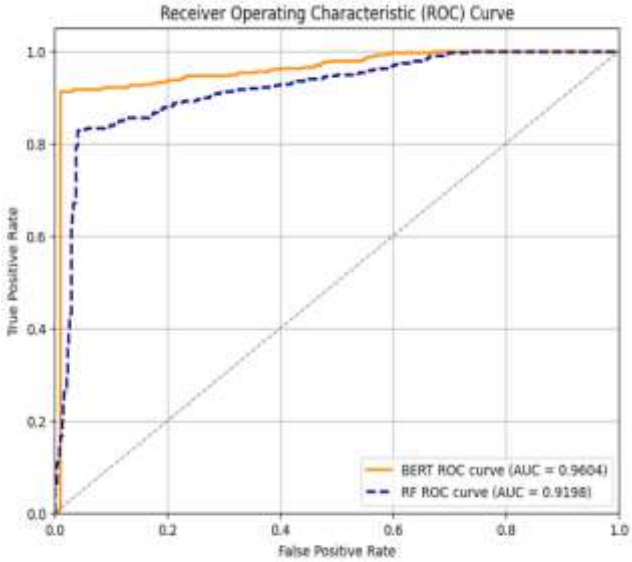
In the **Figure 5**, the training and validation of BERT model compared after 10 epochs. The accuracy curves in the left graph indicates that, the training accuracy rises steadily through the epochs, form around 0.80 at the first epoch to about 0.956 at the 10th. Accuracy of model validation is also increasing steadily and the generalization power can be summarized as high towards unobserved data at 94.87 percent. The right figure shows the loss curves, where training loss ranges between roughly 0.62 to 0.09 across the 10 epochs and validation loss begins at roughly 0.65 to 0.12 also across the 10 epochs. The auspicious trends show that the BERT model is successful at learning the Amazon review data with minimal overfitting, which makes it extremely accurate and acquires a minimal loss on training and validation sets.

Figure 6: Confusion Matrices of the Proposed BERT and RF on the Amazon Reviews Dataset.



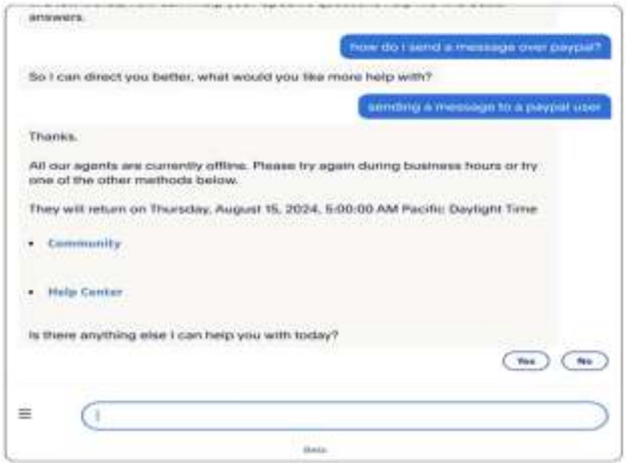
The confusion matrix of the proposed BERT and RF classifiers on the dataset of Amazon Reviews are presented in Fig. 6. BERT model provided a good performance with 245 true negatives, 254 true positive and 10 false positive and 17 false negative that demonstrated that BERT model was effective in incorporating the semantic variations and minimizing mislabeling. Comparatively, the SRF model was found to have 140 TN, 124 TP, 10 FP, and 17 FN showing that its overall predictive ability is weaker. These results affirm the assumption that transformer-based models like BERT offer increased dependability in the areas of intent recognition and sentiment perception in e-commerce chatbots.

Figure 7: ROC Curves of the Proposed BERT and RF Classifiers on the Amazon Reviews Dataset.



The ROC curves on the Amazon Reviews dataset comparing the proposed BERT and Random Forest (RF) classifier are shown in Fig. 7. BERT model had the best discriminative performance with an AUC equal to 0.9604, good sensitivity and specificity that are close to optimal. On the other hand, the RF model achieved a strong result with an AUC of 0.9198 which is relatively lower. These findings affirm that BERT is a better predictive model to distinguish sentiment and intent in e-commerce reviews, which proves useful in improving chatbot-based assignments.

Figure 8: Interaction with the E-Commerce Chatbot Interface.



The following Fig. 8 is an example of a transaction with the e-commerce chatbot interface, in which the user asks to send a message using PayPal. The chatbot will further improve the customer experience, as it clarifies the demand, informs about the availability of services, and proposes alternative support options, which will enable the efficient guidance and the provision of a consistent help in the e-commerce platform.

6.1. Comparative analysis

The relative comparison of the suggested BERT model with the previous transformer-based models to consider their performance in chatbot-related activities. The discrepancies of the accuracy between the models are also mentioned, where the higher potential of the proposed strategy is described in **Table 3.**

Table 3: Comparative Analysis Of The Proposed Bert Model Against Prior Transformer-Based Approaches.

Reference	Models	Accuracy
Proposed	BERT	94.87
[29]	FK-Bert	94.70
[30]	LSTM	90.00
[31]	SVM	76.00
[32]	RoBERTa-A	70.80

The comparative analysis demonstrates better performance of proposed BERT model that achieved the highest acc of 94.87% being just a bit higher than the FK-BERT with 94.70%. Out of the baselines, LSTM performed well with the accuracy of 90.00% meaning, it is appropriate in the modeling of sequential patterns in texts, but with the SVM the result was quite poor (accuracy of 76.00%), which shows that the traditional classifiers are not sufficient in modelling text with complicated semantic dependencies. RoBERTa-A was found to be the worst with the score of 70.80 in terms of relevance of model optimization and fine-tuning of transformer-based architecture. Generally, the results encourage the suitability of the given BERT framework in contextual embeddings to deliver satisfactory WORDsEN earned accuracy in sentiment analysis and intent detection in e-commerce reviews.

7. Conclusion and Future Work

In this investigation study, author has proven the opportunities of advanced NLP in simplifying customer experience in e-commerce context through the implementation of sentiment analysis and intelligent chat robot systems. The results with preprocessing and Random Forest and BERT models showed that transformer based frameworks, including BERT provide better results in relation to conventional classifiers and can achieve up to 94.87 percent to identify the contextual sense and sentiment. The effects of such results are that the context-aware models are able to lead to a more accurate interpretation of the language and hence greater customer satisfaction, efficiency of responses, and personalization. More than technical benefits, the incorporation of sentiment insights enhances trust and decision making of online retailers. Finally, the paper points to the fact that customer engagement is not limited to transactions, and smart systems are capable of producing adaptive, responsive, and customer-focused e-commerce systems. Future research can focus on developing lightweight transformer models for real-time deployment, incorporating multimodal data (text, audio, and images) to improve chatbot interactions, hybrid models (CNN-LSTM or XGBoost-RNN etc.) and exploring explainable AI approaches (LIME and SHAP) to increase transparency and trust in automated systems. Additionally, integrating reinforcement learning for adaptive dialogue management could further personalize customer engagement and drive more effective decision-making in e-commerce environments

8. References

1. Al-Ebrahim MA, Bunian S, Nour AA (2023) Recent machine-learning-driven developments in e-commerce: Current challenges and future perspectives. *Eng Sci.* 28: 1–32. doi:10.30919/es1044.
2. Seetharaman KMR, Pandya S (2023) Importance of artificial intelligence in transforming sales, procurement, and supply chain processes. *Int J Recent Technol Sci Manag.* 8(7): 1–9.
3. Aisha MA, Jamei RB (2025) Conversational AI revolution: A comparative review of machine learning algorithms in chatbot evolution. *East J Eng.* 1(1): 1–18.
4. Murugandi K, Seetharaman R (2023) Digital transformation in retail sales: Analyzing the impact of omni-channel strategies on customer engagement. *J Glob Res Math Arch.* 10(12).
5. Modalavalasa G, Kali H (2023) Exploring big data role in modern business strategies: A survey with techniques and tools. *Int J Adv Res Sci Commun Technol.* 3(3): 431–441. doi:10.48175/IJARSCT-11900B.
6. Neeli SSS (2023) Amazon Redshift performance tuning for large-scale analytics. *J Artif Intell Mach Learn Data Sci.* 1(4).
7. Majumder RQ (2025) Evaluating the correlation between leverage and profitability in the retail sector: A comparative study of listed companies across five years. *Int J Res Anal Rev.* 12(2): 82–88.
8. Bilipelli AR (2022) End-to-end predictive analytics pipeline of sales forecasting in Python for business decision support systems. *Int J Curr Eng Technol.* 12(6): 819–827.
9. Pandya S (2024) Comparative analysis of large language models and traditional methods for sentiment analysis of tweets dataset. *Int J Innov Sci Res Technol.* 9(12): 1647–1657. doi:10.5281/zenodo.14575886.
10. Verma V (2021) Improving product recommendations in retail with hybrid collaborative filtering and LSTM. *Int J Eng Sci Math.* 10(8).
11. Baloska S, Mechkaroska D, Domazet E (2025) AI-driven chatbot in e-commerce. *J UIST.* 1(1).
12. Marjerison RK, Dong H, Kim JM, et al. (2025) Understanding user acceptance of AI-driven chatbots in China's e-commerce: The roles of perceived authenticity, usefulness, and risk. *Systems.* 13(2): 71. doi:10.3390/systems13020071.
13. Johnson E (2024) Machine learning for optimizing customer service in retail environments. *Am J Mach Learn.* 5(3): 23–33.
14. Prajapati N (2025) The role of machine learning in big data analytics: Tools, techniques, and applications. *ESP J Eng Technol Adv.* 5(2): 16–22. doi:10.56472/25832646/JETA-V5I2P103.
15. Patel C (2023) A survey of data-driven customer segmentation methods for targeted marketing campaigns. *ESP J Eng Technol Adv.* 3(3). doi:10.56472/25832646/JETA-V3I7P119.
16. Prajapati V (2025) Enhancing supply chain resilience through machine learning-based predictive analytics for demand forecasting. *Int J Sci Res Comput Sci Eng Inf Technol.* 11(3). doi:10.32628/CSEIT25112857.
17. Rongala S, Pahune SA, Velu H, et al. (2025) Leveraging natural language processing and machine learning for consumer insights from Amazon product reviews. In: *Proc ICSSES 2025.* IEEE: 1–6. doi:10.1109/ICSSES64899.2025.11009528.
18. Pahune S, Chandrasekharan M (2023) Several categories of large language models (LLMs): A short survey. *Int J*

Res Appl Sci Eng Technol. 11(7): 615–633. doi:10.22214/ijraset.2023.54677.

19. Dudhipala A, Karne R, Pativada PK (2025) Prompt2Graph: Leveraging LLMs to construct knowledge graphs from technical manuals. In: *Proc ICIMIA 2025.* IEEE: 912–919. doi:10.1109/ICIMIA67127.2025.11200177.

20. Pande SD, et al. (2025) A novel chatbot driven by sentiment analysis using a capsule network and BiLSTM for online user feedback on consumer electronics. *IEEE Trans Consum Electron.* 1–1. doi:10.1109/TCE.2025.3593110.

21. Sharma R, Mishra A (2024) Advanced NLP and ML techniques in e-commerce: Enhancing customer experience with AI chatbots. In: *Proc WCONF 2024.* IEEE: 1–5. doi:10.1109/WCONF61366.2024.10692186.

22. Wei W (2024) User behavior analysis of e-commerce platforms under random forest recommendation algorithm. In: *Proc ICDOT 2024.* IEEE: 1–5. doi:10.1109/ICDCOT61034.2024.10516196.

23. Aji Bawono MW, Purwitasari D, Purnomo MH (2023) Measuring credibility level of e-commerce sellers with SEM and Naive Bayes on tweets of service quality. In: *Proc TENCON 2023.* IEEE. doi:10.1109/TENCON58879.2023.10322485.

24. Zulfadli, Ilham AA, Indrabayu (2023) Sentiment analysis with soft-voting method on customer reviews for purchasing transactions of e-commerce. In: *Proc IAICT 2023.* IEEE: 295–299. doi:10.1109/IAICT59002.2023.10205954.

25. Malali N (2025) AI-powered data preprocessing and transformation platform for autonomous data cleaning and advanced feature engineering. *Patent* 202521035175.

26. Tusher EH, Ismail MA, Rahman MA, et al. (2024) Email spam: A comprehensive review of optimized detection methods, challenges, and open research problems. *IEEE Access.* 12: 143627–143657. doi:10.1109/ACCESS.2024.3467996.

27. Khan SS, Alharbi Y (2024) Sentiment analysis of movie review classifications using deep learning approaches. *Int J Adv Appl Sci.* 11(8): 146–157. doi:10.21833/ijaas.2024.08.016.

28. Devlin J, Chang MW, Lee K, et al. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proc NAACL-HLT 2019.*

29. Rajasekar AA, et al. (2025) Flippi: End-to-end GenAI assistant for e-commerce. *arXiv.* 1(1).

30. Coyne E, Smit J, Güner L (2019) Sentiment analysis for Amazon.com reviews. *Big Data Media Technol.* doi:10.13140/RG.2.2.13939.37920.

31. Monsoor RS, et al. (2023) Sentiment analysis of Amazon reviews using machine learning classifier. In: *Proc ICCIT 2023.* IEEE: 1–6. doi:10.1109/ICCIT60459.2023.10441259.

32. Rajasekar A, Garera N (2021) Answer generation for questions with multiple information sources in e-commerce. *arXiv.* 1(1).