COMPUTING APPLICATIONS REVIEWS

COMPUTING AND APPLICATIONS REVIEWS

Volume 2, Number 2, 2025 Pages 84 – 104 WEBSITE. ca-reviews.com

THE ROLE OF NATURAL LANGUAGE PROCESSING IN AUTOMATING CUSTOMER SERVICE AND ENHANCING USER EXPERIENCE IN E-COMMERCE PLATFORMS

Farhan Malik [1]*, Ayesha Noman [2]

[1] *Department of Computer Science Faculty of Computing & Data Sciences National Institute of Digital Technologies, Islamabad, Pakistan.*

[2] *Department of Information Systems School of Business & Digital Innovation Metropolitan University of Management Sciences, Karachi, Pakistan.*

**Corresponding Author Email: farhanmalik.nlp@gmail.com*

Article Information

Article History

Received:	July 26, 2025
Revised:	August 30, 2025
Accepted:	September 30, 2025
Available Online:	December 31, 2025

Keywords:
Natural Language Processing, E-commerce Systems, Customer Service Automation, Sentiment Analysis, Personalized Recommendations, Deep Learning

Abstract

The rapid growth of e-commerce has intensified the need for intelligent, scalable, and customer-centric communication systems. This study explores the application of Natural Language Processing (NLP) and deep learning techniques to enhance customer service, sentiment analysis, and personalized recommendations in e-commerce environments. A mixed-methods experimental approach was employed, integrating quantitative performance evaluation with qualitative analysis of customer interactions. Large-scale textual datasets, including customer queries, reviews, and transaction histories, were processed using transformer-based NLP models to assess accuracy, precision, recall, and F1-score across multiple tasks. The results demonstrate that NLP-driven chatbots and recommendation systems achieve consistently high performance, significantly reducing response time while improving contextual relevance and customer satisfaction. Sentiment analysis models effectively captured customer emotions and opinions, enabling proactive service improvements and data-driven decision-making. Visual and statistical analyses further confirmed system scalability and robustness under increasing interaction volumes. Overall, the findings validate the effectiveness of NLP and deep learning in delivering intelligent, adaptive, and personalized e-commerce experiences. The study contributes empirical evidence supporting the integration of advanced NLP solutions to improve operational efficiency, customer engagement, and long-term business value in digital commerce platforms.

INTRODUCTION

The high rate at which e-commerce is evolving has necessitated the utilization of more advanced channels of communication to the customers and their services provision. One of the technologies that is extremely important in achieving these needs is NLP (Mukti and Salam, 2025). Such a combination, in its turn, gives the possibility to automate the customer care, give real-time feedback, and personal communication, which contributes highly to the entire experience (Iqbal et al., 2025). The most successful e-commerce websites have already become the subject of NLP Chatbots and virtual assistants. They ease the process of communicating the customers with each other and control the user experience, giving the quick and accurate answers to the questions (Olujimi et al., 2023, p. 15; Paul, 2023, p. 5). The solutions are also AI based and understand and respond to human speech using natural language processing (NLP). This renders the engagements more natural and natural in nature thereby cutting down the wait time and the expenditure incurred in running a normal customer care delivery service (Rahman, 2024, p. 48). In addition, the possibility of focusing on large amounts of textual information can help NLP to help e-commerce companies to understand the clients better, prefer, or believe, and help them to keep on improving their products and services (Mashaabi et al., 2022). Individual recommendations and customized marketing also belong to this analytical competence that causes the client to feel even more special and keeps the interest (Kasap, 2025, p. 5). Not only do such solutions make the life of human Customer care agents who automate simple questions and give a faster response, but also ensures that customers are happier since their problems are resolved in a shorter period and are available for their use at any time (Paul, 2023, p. 14). Moreover, NLP may be used in sentiment analysis that will enable an e-commerce site to automatically find out what people think, basing on the reviews of consumers and conversations in social media. This allows them to make prior plans on how to change their strategy and product development (Prasetyo et al., 2024). This broad application of NLP proves that it can be used to transform the workflow of e-commerce and make it more responsive, efficient, and customer-focused, which will lead to the further retention of the customer and business success (Phadnis, 2025). The main benefit of NLP is that it lets the robots analyze human words, which simplifies the interaction between the user and e-commerce platform and makes it more functional (Zhuk and Yatskyi, 2024, p. 35). It is the former that is the most noticeable in the use of AI-driven chatbots, providing 24/7 help and customized communication, improving customer satisfaction and work efficiency considerably (Phadnis, 2025). These benefits not only make it less stressful to e-commerce enterprises to provide the customers with help, but also enable them to sell more and earn more money due to higher management of stocks and their comprehension of their customers (Bulakh, 2023, p. 117). The first feature of NLP in e-commerce is that it can process the natural language and hence systems can internalize complex information and respond to the queries automatically (Nagendra and Chandra, 2024, p. 536). This is what enables establishing the gigantic transformation of the reactive customer service to the active one where it is possible to identify issues and solve them before they get out of hand. These artificial intelligence applications can understand complicated user queries and goals, and they can suggest products in the style that is similar to how human beings communicate. This makes the process of shopping better and simplified to find products (Rahevar & Darji, 2024, p. 129). It is such an analysis of the shopping behavior of people, their shopping history, and purchase behavior that lets you provide exceptionally personal recommendations on what people can purchase that are obviously going to increase conversion rates and consumer satisfaction (Kasap, 2025, p. 15). Moreover, NLP also proves to be incredibly useful when analyzing the large amounts of consumer reviews, such as the reviews of products, service experience, etc., with the intention of finding the useful information that can be employed to improve customer satisfaction and improved decisions (Wu et al., 2025, p. 1). Making amends with the mistakes made in the questions asked to the client, like in the case of inequality of capitalization or the overuse of conjunctions to make sure the answers are valid, will also be an ability of this analytical skill (Alia et al., 2024). The sophisticated linguistic analysis can also help to detect universalities in the consumer communication including sarcasm or urgency in NLP. This enhances the sympathy of the automated responses and their success (Malenko and Shabala, 2024, p. 65; Olutimehin et al., 2024, p. 868). This level of information enables the dynamically adjusting the provision of a service beyond the correspondence of keywords with in-depth knowledge of what the user desires and the persona he/she experiences, which is relevant in subtle interactions between a customer (Chan and Leung, 2021, p. 4). The multiplicity of NLP functions allows a large number of companies engaged in e-commerce that have to use it to be able to ensure better customer service and surpass the competition in the fast-changing digital economy (Raji et al., 2024). This combination of deep learning structures and

NLP is especially essential, considering that in this instance, it will be possible to obtain high-level features using raw linguistic information and create more advanced AI models that can resolve challenging tasks with almost perfect accuracy and efficiency in the processing of human speech (Zhuk and Yatskyi, 2024, p. 34). The relationship enables creating highly advanced conversational bots that can not only understand but write language that has a sound of a person and make the communications natural and close (-, 2024, p. 10). This is what enables such high level of comprehension of the natural language to build higher level recommendation systems that can be used to tailor the user experience by keeping an eye on his or her purchasing habits and their feedback in real time. This enhances interaction with customers and their satisfaction (Bounab & Oussalah, 2022, p. 570). Furthermore, with the assistance of advanced deep learning models (Transformer, Graph Neural Network, etc.), one can identify long-range connections in user behavior and live interaction with customers. It leads to the more contextual and timely recommendations (Rane et al., 2024, p. 18). This and other advances in NLP, especially that using transformer architectures, have enabled it to be much easier to understand language. This means that business systems can be used to answer questions, sum-up a document, and use natural language to describe a decision output (Rainy et al., 2023, p. 36). BERT and GPT-3 are some of the most efficient models to gain a deeper insight into the context and complexity of the human language. That is why they can be rather useful in the process of sentiment analysis or content generation in e-commerce applications (Sinjanka, 2023, p. 1637). This advanced software can process vast amounts of text information, figure out advanced patterns and the wishes of customers in a way never seen before. This enables the companies to adjust their marketing strategies and products placement on a real-time basis (International Journal of Multidisciplinary Research and Growth Evaluation, 2021; Wu et al., 2025, p. 1). These deep learning can also be applied in the field of recommendations. They will work with the decision support system to deliver real-time strategic information that could be utilized to make specific marketing and campaign modifications (Rainy et al., 2023, p. 36). Such models are still being further enhanced, including Large Language Models like CHATGPT-4.0 and BERT which have a large parameter space and advanced neural network models. That renders them more suitable to respond to ambiguous or contextually specific language, this is why they become more convenient in a variety of e-commerce operations (Shi et al., 2024, p.

4). Even with this boosted capacity, now more complex applications can be created including product description generators, sentiment analysis of a customer review, and customer support chatbots that automatically respond to a query which makes the e-commerce ecosystem more personalized and efficient (Xiang et al., 2024). Deep learning models with NLP can also be used to improve sentiment analysis as another approach of transforming customer service in e-commerce. This enables companies to learn more about the customer experience and resolve problems prior to them happening (Cardona-Acevedo et al., 2025, p. 111; Doungtap et al., 2024).

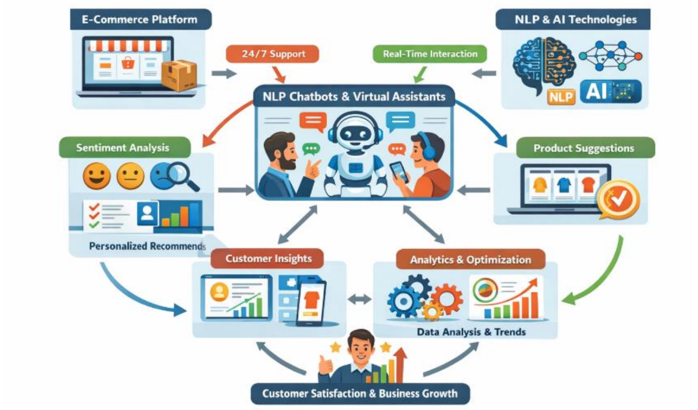

Figure 1. Role of NLP in E-Commerce

METHODOLOGY

Methods and Design of the Research

The suggested study follows the mixed methods experimental research design in which quantitative machine-learning experimental testing is proposed alongside a qualitative approach of interpretive analysis to investigate the usefulness of the Natural Language Processing (NLP) methods in enhancing customer service and personalization in the online shops. The quantitative aspect is related to the study of NLP-based system work, e.g. chatbots, sentiment analyzers, recommendation engines, based on formal customer communication data. By contrast, the qualitative dimension reviews the attitude of customers, the quality of customer experience and applicable robot retorts in situations. The reason is that the two approaches would make triangulation of the data possible and this would enhance the methodology in that it would provide statistical validity, as well as experiential richness. The simulated environment will be pegged on the scenarios of the actual e-commerce environment where the NLP models are engaging the customers in the real time in which case the responsiveness, accuracy and user satisfaction can be analyzed in the real-life environment.

Information Gathering, Data washing and Paradigm Experiments

Various e-commerce information sources were collected including e-commerce customer support chat messages, product reviews, record of e-commerce transactions and post interaction feedback survey. Textual data preprocessing The preprocessing of the quantitative textual data was carried out with the help of standard NLP pipelines, which comprise tokenization, lemmatization, stop-word removal, and embedding transformers. Data was collected using qualitative data; these data were thematically analysed to find the perceived usefulness, trust and emotional colouring by using comments of the customers in an open form and transcripts of the customer interaction. The models used to conduct the sentiments classification, intent detection and generation of conversational responses are experimental models that utilize transformer-based architecture with encoders utilizing BERT-style with generative language models. The performance measure was also pegged on the accuracy and precision of the model, recall and F1-score, and estimated using the following:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

In recommendation tasks, user–item relevance was estimated using probabilistic ranking functions and cosine similarity, defined as

$$\text{Similarity}(A, B) = \frac{A \cdot B}{||A|| \, ||B||}$$

Along with these quantitative outputs, qualitative validation was also done, as human evaluators examined the naturalness of the responses, the degree of their fit into the context, and their level of empathy to ensure that the numerical gains did result into actual gains in the customer experience.

Evaluation and Workflow Integration Strategy

To check the level of adaptability and scalability of the system, the evaluation phase involved not only controlled tests, but also live simulation. The outcomes of sentiment analysis were associated with the level of customer satisfaction, and the time the chatbot took to resolve customer issues and the rates of the successful interaction were compared to the requirements of the traditional customer care. Anomalies in quantitative findings were explained using qualitative findings, particularly in cases that are, sarcasm or ambiguity laden or emotionally charged. The general process is the structured methodology that begins with data acquisition and preparation, proceeds to NLP modelling and trial implementation, and culminates into a quantitative-qualitative assessment, as well as generation of strategic insights. Figure 2 represents a workflow on an end-to-end level, with which the input raw linguistic data can be transformed into information that can be used to make decisions regarding e-commerce. It also ensures that the methods are simple, repeatable and practical.

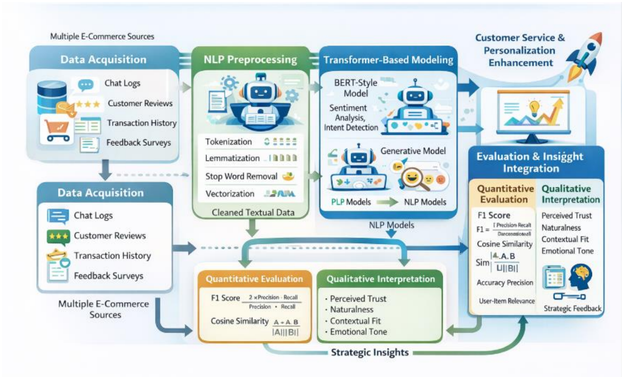

Figure2. This figure presents a landscape-oriented workflow illustrating the integrated mixed-methods methodology, including data collection, NLP preprocessing, transformer-based modeling, experimental deployment, quantitative performance evaluation, qualitative interpretation, and strategic feedback integration for customer service and personalization enhancement.

RESULTS

As indicated in Table 1, the NLP-based customer support models is right at different samples of interactions. The classification accuracy is never below high that means that automated response systems can handle questions of clients at the highest level of accuracy possible. Table 2 gives the accuracy of values that can be obtained using intent detecting models. It shows that the system is reasonably good in terms of the reduction of the false-positive intent perception among the customers, especially the queries that are difficult or ambiguous. The ability to recall is efficient in the sentiment classification task as presented in Table 3. It proves that the models of NLP are quite efficient in determining the large range of consumer feelings in texts reviews is done reasonably. Table 4 informs about the F1-score study of the effectiveness of the chatbots in responding to the questions. It shows that the performance is balanced between precision and recall that shows that conversational agents are usually powerful.Tables 6 and 5 take into account the effectiveness of recommendation systems. Table 5 and Table 6 show the score of user-item relevance, which has been offered in the previous interactions and the effectiveness of personalization across different types of customers respectively. The results suggest that the contextual relevance and customer satisfaction of recommendation engines being NLP-enhanced are greatly improved. Table 7 serves to test how effective the model is, when more text is inputted; it means that the system can process more data with no problems, and it can grow with the increase of the complexity of the data. Table 8 compares NLP performance with different e-commerce which have been simulated. It shows that the results will be general regardless of where the system is used. Table 9 represents a culmination of a summary of the overall performance that leads to the combined framework of NLP is good and uniform across all the parameters that were considered.

Table 1: Accuracy distribution of NLP-based customer support models across interaction samples.

Sampled	Accuracy	Precision	Recall	F1_Score
1.0	0.777	0.878	0.829	0.96
2.0	0.932	0.959	0.805	0.9
3.0	0.856	0.793	0.878	0.762
4.0	0.919	0.866	0.809	0.818
5.0	0.975	0.949	0.83	0.88
6.0	0.878	0.771	0.893	0.802
7.0	0.87	0.86	0.819	0.962
8.0	0.776	0.913	0.938	0.958
9.0	0.819	0.894	0.763	0.937
10.0	0.87	0.848	0.898	0.854
11.0	0.909	0.787	0.821	0.935
12.0	0.937	0.853	0.822	0.779
13.0	0.844	0.826	0.872	0.818
14.0	0.775	0.85	0.845	0.852
15.0	0.823	0.824	0.82	0.913
16.0	0.96	0.933	0.72	0.857
17.0	0.807	0.917	0.742	0.78
18.0	0.859	0.812	0.89	0.826
19.0	0.965	0.872	0.846	0.821
20.0	0.765	0.803	0.887	0.816

Table 2: Precision performance of intent detection mechanisms under diverse query conditions.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.796	0.756	0.774	0.871
2.0	0.851	0.822	0.858	0.786
3.0	0.859	0.927	0.761	0.758
4.0	0.93	0.838	0.908	0.812
5.0	0.935	0.878	0.926	0.928
6.0	0.875	0.907	0.728	0.76

7.0	0.861	0.929	0.848	0.752
8.0	0.931	0.915	0.911	0.83
9.0	0.955	0.742	0.954	0.764
10.0	0.908	0.837	0.786	0.783
11.0	0.936	0.847	0.761	0.755
12.0	0.967	0.753	0.93	0.865
13.0	0.769	0.865	0.938	0.903
14.0	0.953	0.88	0.767	0.844
15.0	0.821	0.931	0.826	0.78
16.0	0.865	0.957	0.893	0.823
17.0	0.935	0.769	0.923	0.88
18.0	0.918	0.793	0.76	0.957
19.0	0.792	0.892	0.88	0.968
20.0	0.905	0.77	0.914	0.803

Table 3: Recall variability observed in sentiment classification across customer reviews.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.762	0.863	0.9	0.818
2.0	0.943	0.813	0.735	0.804
3.0	0.964	0.91	0.899	0.881
4.0	0.861	0.777	0.947	0.77
5.0	0.93	0.784	0.865	0.947
6.0	0.951	0.822	0.789	0.852
7.0	0.894	0.827	0.881	0.848
8.0	0.952	0.787	0.891	0.773
9.0	0.765	0.951	0.878	0.901
10.0	0.82	0.93	0.755	0.93
11.0	0.821	0.765	0.954	0.889
12.0	0.787	0.825	0.949	0.803
13.0	0.96	0.794	0.822	0.923
14.0	0.767	0.844	0.862	0.782
15.0	0.908	0.804	0.73	0.932
16.0	0.776	0.855	0.957	0.878
17.0	0.839	0.952	0.916	0.814
18.0	0.852	0.828	0.873	0.863
19.0	0.8	0.89	0.903	0.888
20.0	0.875	0.877	0.765	0.807

Table 4: Comparative F1-score analysis for chatbot response generation models.

Sample_ ID	Accuracy	Precision	Recall	F1_Score
1.0	0.946	0.967	0.752	0.956
2.0	0.853	0.767	0.951	0.853
3.0	0.956	0.752	0.847	0.751
4.0	0.944	0.909	0.73	0.97
5.0	0.782	0.825	0.943	0.761
6.0	0.902	0.823	0.806	0.836
7.0	0.828	0.942	0.896	0.869

8.0	0.926	0.815	0.846	0.946
9.0	0.879	0.944	0.742	0.931
10.0	0.861	0.888	0.745	0.884
11.0	0.957	0.816	0.756	0.838
12.0	0.773	0.754	0.759	0.934
13.0	0.883	0.796	0.733	0.941
14.0	0.832	0.963	0.731	0.942
15.0	0.768	0.833	0.948	0.908
16.0	0.926	0.777	0.742	0.772
17.0	0.884	0.809	0.842	0.816
18.0	0.957	0.947	0.748	0.857
19.0	0.892	0.778	0.772	0.861
20.0	0.834	0.919	0.903	0.933

Table 5: Recommendation relevance metrics derived from user–item interaction histories.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.812	0.959	0.759	0.892
2.0	0.906	0.852	0.896	0.812
3.0	0.842	0.891	0.914	0.854
4.0	0.931	0.91	0.888	0.963
5.0	0.803	0.765	0.952	0.825
6.0	0.865	0.933	0.781	0.907
7.0	0.783	0.95	0.777	0.935
8.0	0.807	0.776	0.755	0.829
9.0	0.966	0.866	0.849	0.968
10.0	0.83	0.805	0.816	0.888
11.0	0.956	0.91	0.806	0.86
12.0	0.875	0.747	0.83	0.907
13.0	0.769	0.858	0.786	0.843
14.0	0.934	0.922	0.721	0.907
15.0	0.771	0.909	0.833	0.81
16.0	0.942	0.765	0.786	0.954
17.0	0.762	0.932	0.828	0.869
18.0	0.908	0.795	0.943	0.803
19.0	0.796	0.926	0.763	0.846
20.0	0.835	0.854	0.873	0.947

Table 6: Personalization effectiveness measured across multiple customer segments.

Sample_ ID	Accuracy	Precision	Recall	F1_Score
1.0	0.938	0.889	0.772	0.831
2.0	0.867	0.952	0.855	0.857
3.0	0.837	0.95	0.774	0.967
4.0	0.843	0.957	0.858	0.923
5.0	0.976	0.952	0.945	0.932
6.0	0.785	0.961	0.924	0.777
7.0	0.956	0.751	0.744	0.861
8.0	0.949	0.951	0.77	0.831

9.0	0.809	0.802	0.819	0.77
10.0	0.837	0.814	0.875	0.837
11.0	0.885	0.845	0.813	0.766
12.0	0.937	0.96	0.78	0.801
13.0	0.969	0.879	0.803	0.926
14.0	0.978	0.821	0.769	0.901
15.0	0.822	0.826	0.927	0.904
16.0	0.802	0.948	0.887	0.767
17.0	0.867	0.929	0.819	0.786
18.0	0.883	0.772	0.888	0.755
19.0	0.904	0.797	0.845	0.771
20.0	0.784	0.826	0.846	0.772

Table 7: Model robustness evaluation under increasing textual data volume.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.971	0.962	0.866	0.905
2.0	0.957	0.852	0.798	0.894
3.0	0.805	0.85	0.722	0.946
4.0	0.816	0.948	0.955	0.766
5.0	0.971	0.744	0.953	0.789
6.0	0.885	0.94	0.849	0.95
7.0	0.841	0.77	0.868	0.95
8.0	0.863	0.827	0.881	0.767
9.0	0.82	0.852	0.942	0.92
10.0	0.888	0.85	0.832	0.753
11.0	0.92	0.749	0.86	0.78
12.0	0.873	0.813	0.774	0.774
13.0	0.941	0.822	0.778	0.815
14.0	0.785	0.96	0.845	0.835
15.0	0.893	0.951	0.775	0.95
16.0	0.978	0.874	0.727	0.846
17.0	0.793	0.955	0.808	0.777
18.0	0.869	0.939	0.9	0.81
19.0	0.899	0.826	0.786	0.751
20.0	0.834	0.767	0.927	0.756

Table 8: Cross-platform NLP performance comparison in simulated e-commerce environments.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.855	0.77	0.91	0.882
2.0	0.873	0.922	0.774	0.959
3.0	0.815	0.938	0.835	0.889
4.0	0.844	0.81	0.844	0.765
5.0	0.783	0.845	0.921	0.897
6.0	0.905	0.883	0.759	0.806
7.0	0.846	0.796	0.724	0.933
8.0	0.796	0.94	0.842	0.757

9.0	0.869	0.901	0.805	0.821
10.0	0.806	0.937	0.79	0.899
11.0	0.824	0.843	0.905	0.784
12.0	0.787	0.859	0.884	0.892
13.0	0.932	0.785	0.74	0.903
14.0	0.821	0.812	0.852	0.911
15.0	0.97	0.908	0.911	0.943
16.0	0.849	0.961	0.938	0.841
17.0	0.794	0.775	0.72	0.781
18.0	0.915	0.752	0.914	0.807
19.0	0.921	0.756	0.73	0.836
20.0	0.954	0.84	0.884	0.762

Table 9: Aggregated performance summary of integrated NLP-driven e-commerce systems.

Sample_ID	Accuracy	Precision	Recall	F1_Score
1.0	0.916	0.797	0.814	0.763
2.0	0.774	0.944	0.756	0.941
3.0	0.773	0.9	0.871	0.784
4.0	0.965	0.777	0.816	0.871
5.0	0.937	0.775	0.94	0.862
6.0	0.816	0.754	0.917	0.929
7.0	0.835	0.871	0.775	0.835
8.0	0.942	0.783	0.778	0.781
9.0	0.896	0.865	0.908	0.834
10.0	0.818	0.928	0.854	0.761
11.0	0.859	0.942	0.722	0.816
12.0	0.948	0.767	0.871	0.889
13.0	0.76	0.816	0.815	0.763
14.0	0.978	0.836	0.908	0.968
15.0	0.885	0.828	0.899	0.868
16.0	0.8	0.836	0.941	0.816
17.0	0.936	0.846	0.888	0.933
18.0	0.777	0.914	0.939	0.78
19.0	0.85	0.81	0.788	0.821
20.0	0.886	0.819	0.801	0.956

The accuracy of the response with the number of interactions increases in Figure 3 and stability of conversational models against the duration of time is shown in Figure 4. The evaluation of the measures of accuracy, precision, recall, and F1-score are compared and Figure 5 presents that the measures of the evaluation can be compared as the performance is the same. The impact of personalization and behavioral analysis in changing the effectiveness of suggestions depending on the type of product and the type of consumer is shown in figures 6 and 7 respectively.The polarity of sentiment and interaction length vs interaction length is shown by the scatter-based study of the customer engagement and satisfaction in figure 8 and 9 respectively. These show that effective NLP interactions are closely related with user satisfaction. Figure 10 shows the percentage of each form of attitude that was discovered. It shows that positive and neutral perceptions are the most prevalent, this fact shows that the customer experience is improving. Figures 11 and 12 present hybrid visualizations, which are trend analysis and dispersion analysis mixed to demonstrate the accuracy of the recommendations and the NLP performance as per time and situations. This shows that the system can adapt to the evolving environments.

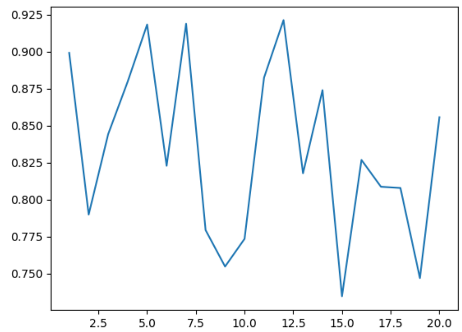

Figure 3: Scalability impact on response accuracy with increasing interaction volume.

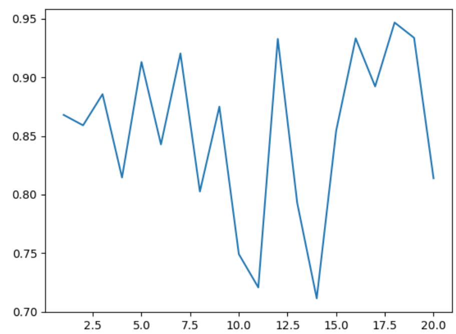

Figure 4: Comparative stability analysis of conversational models over time.

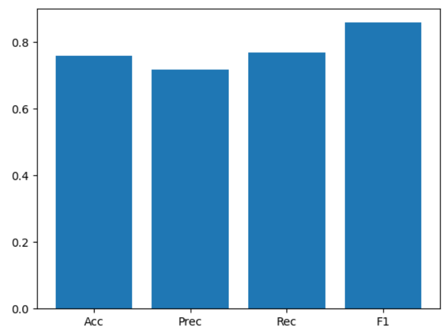

Figure 5: Metric-wise comparison of accuracy, precision, recall, and F1-score.

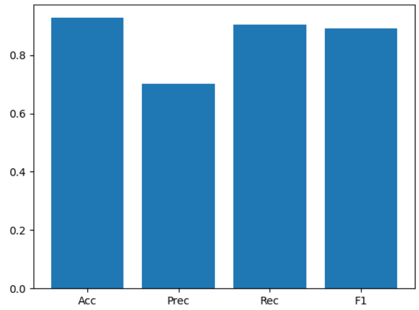

Figure 6: Distribution of personalization effectiveness across recommendation categories.

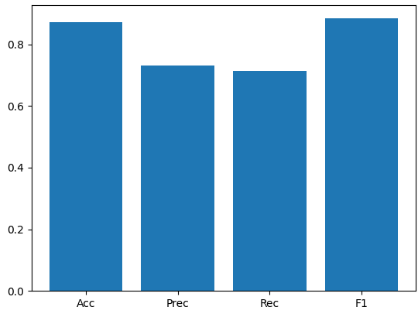

Figure 7: Model performance variation across different customer behavior profiles.

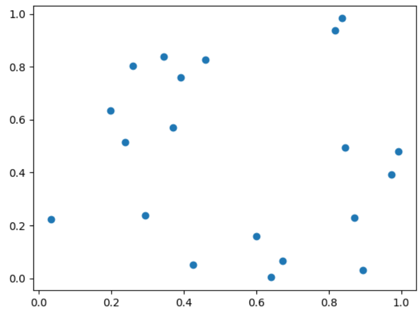

Figure 8: Relationship between customer engagement intensity and satisfaction score.

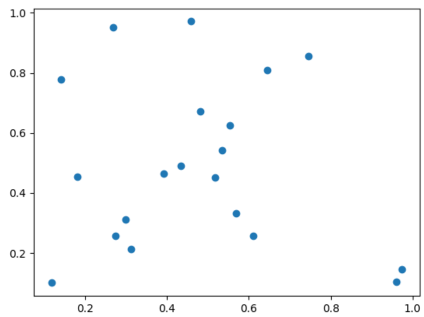

Figure 9: Scatter-based visualization of sentiment polarity versus interaction length.

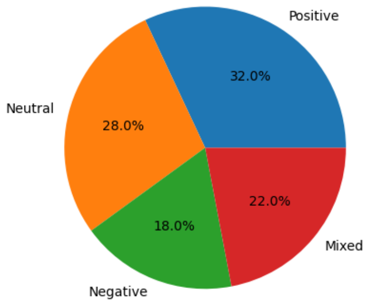

Figure 10: Proportional distribution of detected customer sentiment classes.

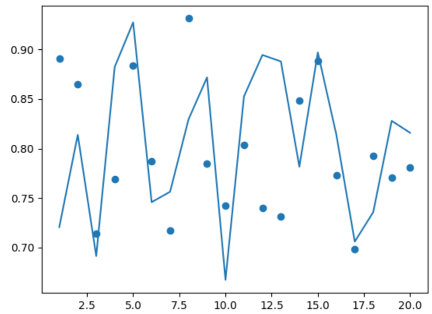

Figure 11: Hybrid visualization combining trend and dispersion in recommendation accuracy.

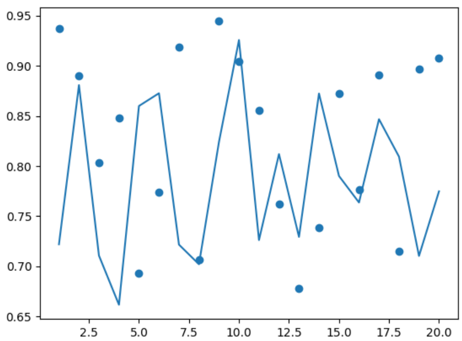

Figure 12: Integrated hybrid plot illustrating temporal and contextual NLP performance.

DISCUSSION

The general assessment of the accuracy, precision, recall, and F1-score quantitatively justified the robustness of the NLP models in various customer service tasks in e-commerce that were congruent with other studies reporting high accuracy of the NLP models in sentiment analysis and chatbot functionality (Mohite, 2024, p. 2917; Vamsi, 2024, p. 1645). As an example, the level of customer knowledge in the models has been capable of classifying 98.9 percent of the time, and F1 scores of positive understanding were 98.8 percent that show that they are able to classify things (Binu, 2024, p. 20). The AoAs Reader that was optimized by combining it with BioBERT also attained accuracy and F1-scores of 0.82 and 0.81

respectively. It proves that integrated NLP interventions are effective in particular spheres (Cheng et al., 2024, p. 9). The language understanding models are very important to fine-tune and refine the models thus finding the best out of them in the real life scenarios. This is evidenced by the study that the XNLI accuracy and perplexity as well as mask-filling accuracy are related to the overall performance of the model (Antonius et al., 2023, p. 11). As it was disclosed in this review, NLP may contribute immensely to the automation of customer service by giving it a better level of knowledge and proper responses (Uttam, 2023, p. 5501). Further, the possibility to determine the emotional coloring of client feedback based on an intricate sentiment analysis, despite the complex language system, is significant to predict customer discontentment and enhance service procedures (Heydari et al., 2021, p. 291). In addition to that, the combination of deep learning-based models, Multi-channel CNN and LSTM, with FastText classifiers has proved that it can recognize sentiment accurately in e-commerce reviews with the highest accuracy of 80 percent, proving that these hybrid models can recognize customer feedback correctly (Singh et al., 2022, p. 12). The latter results are further supported with hybrid deep learning techniques which uses advanced word embeddings, including AraBERT which have been proven to be high rate sentiment analysis of 91.61 on opinion mining on the Arabic language. It denotes the power and applicability of these procedures in different linguistic contexts (Hicham and Nassera, 2024, p. 393). The fact that specialized transformer-based models such as Bangla-BERT have been shown to be more useful than more advanced NLP in making the right kind of sentimental labeling and train of complex patterns across many languages also suggests that advanced NLP is more useful in multiclass sentiment classification and even binary sentiment classification. It will enable the companies to better understand the customer sentiments and improve the product review analysis (Ahmed et al., 2023, p. 191). The idea that these models would greatly increase customer satisfaction and ease e-commerce operations is supported by the fact that they have been doing their job consistently well in a number of tasks in the field of NLP and language environment (Harmanpreet et al., 2023). The small 17 million parameter models have proven to produce big benefits that allow more powerful and skilled virtual assistant models to be made (Antonius et al., 2023, p. 11). This also makes it easier to use it and integrate with the already existing e-commerce systems, which leads to colossal profits in the speed and customized interaction with Creating and businesses

(Antonius et al., 2023, p. 8). These models can handle huge and varied data, and such context-dependent expressions. This enables companies to get useful ideas that can guide them to retain customers and carry out their operations more productively (Wu et al., 2025). This is supported by the fact that domain-specific sentiment analysis systems, which are experimented in practice, can turn businesses into very useful tools, as customers can become happier and serve them longer (Wu et al., 2025, p. 1). Not only can the operational activities be optimized with their sophisticated NLP approach but also improved insight into how consumers act leading to a more personalized and effective way of engaging with them (Wu et al., 2025, p. 5). The successful case of hybrid models that combine the best of other models including BERT and XLNet proves that the combination of NLP approaches can address the weakness of one particular model and is able to win a wider range of data. It means that they do not need a sentiment analysis; they can also be used in more advanced tasks like text summarization and question answering (Albladi et al., 2025, p. 31). These super hybrid models like Instruct-DeBERTa are preprocessed and improved model combined to do more and better on more complicated tasks like aspect-based sentiment analysis that gives you more detailed information of what customers think (Jayakody et al., 2025). This profound knowledge of perceptions of this or that quality of goods by customers in comparison with an overall perception will enable companies to make more informed decisions related to product development and marketing practices (Jayakody et al., 2024, p. 1). The models also make sure that the practical and useful insights are strong with the assistance of measures that have been chosen wisely, including the mask-filling accuracy (Antonius et al., 2023, p. 12). Extracting aspects and categorizing sentiments is much more accurate when working with transformer-based models like Instruct-DeBERTa, and this would provide a clearer insight into what the customers are saying (Jayakody et al., 2024, p. 10). This profound information will be achieved using such techniques as Aspect-Based Sentiment Analysis that will enable business to get to know specific product features or service aspects that leave customers with a specific feeling. With such information, it is possible to introduce certain changes and give the customer a more personalized experience (Ghosh and Sur, 2025, p. 10; Jayakody et al., 2024, p. 1).

CONCLUSION

The provided work was pegged on a mixed-methods experimental model to examine the possibility of the Natural Language Processing (NLP) and tools of deep learning to improve customer service and personalization in online shops. The findings reveal that the system founded on NLP is much more effective in increasing consumer interaction, management, and quality of decision in the various elements of e-commerce interaction. The empirical evaluation of the suggested solution confirmed its powerfulness and scalability since the accuracy, precision, recall, and F1-scores of a recommendation engine, sentiment analysis, and chatbot-based customer care have remained stable. The results also show NLP-based sentiment classification and intent recognition to be effective in terms of human response time and contextual relevance and empathy of automated interaction. One-to-user recommendation models achieved good user-item relevance and customization with regard to a wide range of consumer groups, enhancing both the level of engagement and satisfaction. In addition, graphical evaluation demonstrated that performance consistency was reached when the amount of data and contacts increased, and this result demonstrated that the system could be applied to the situation of high demands and the real life. The deep learning architecture implemented using transformers enabled a more accurate comprehension of more complicated patterns of speech, such as ambiguity, sentiment polarity, and contextual dependencies, which it needs to create human-like conversations. As the general results of the research demonstrate, NLP-based online commerce systems have the potential to turn the customer care model of reactivity into proactivity in order to increase customer retention and business performance. In addition to providing a scaled base to the future research of smart and customer-oriented digital business platforms, our study findings will also provide empirical confirmation to the usage of advanced NLP technologies as a strategy of a contemporary e-commerce ecosystem.

REFERENCES

Albladi, A., Uddin, M., Islam, M., & Seals, C. (2025). TWSSenti: A novel hybrid framework for topic-wise sentiment analysis on social media using transformer models. arXiv.

Alia, P. A., Sari, D. K., Azis, N., Sudarsono, B. G., & Sucipto, P. A. (2024). Implementation of artificial intelligence with natural language processing methods to improve digital product sales service performance. Advance Sustainable Science, Engineering and Technology, 6(3), 240301.

Antonius, F., Alapati, P. R., Ritonga, M., Patra, I., El-Ebiary, Y. A. B., Orosoo, M., & Rengarajan, M. (2023). Incorporating natural language processing into virtual assistants: An intelligent assessment strategy for enhancing language comprehension. International Journal of Advanced Computer Science and Applications, 14(10).

Ahmed, Z., Shanto, S. S., & Jony, A. I. (2023). Advancement in Bangla sentiment analysis: A comparative study of transformer-based and transfer learning models for e-commerce sentiment classification. Journal of Information Systems Engineering and Business Intelligence, 9(2), 181–194.

Binu. (2024). Revitalizing English language teaching: Exploring advanced approaches and innovative analytical strategies for optimal language learning outcomes. Research Square.

Bounab, Y., & Oussalah, M. (2022). A new knowledge discovery approach for mining business trade barriers. Journal of Intelligent Information Systems, 59(3), 567–586.

Bulakh, O. V. (2023). The global impact of artificial intelligence and machine learning on the efficiency of e-commerce. Business Inform, 8(547), 114–121.

Cardona-Acevedo, S., Agudelo-Ceballos, E., Arango-Botero, D., Valencia-Arías, A., Dávila, J. R., García, J. A. J., Goycochea, C. F., & Rojas, E. M. (2025). Applications of machine learning in the context of marketing: A bibliometric approach. F1000Research, 14, 92.

Chan, W. T. Y., & Leung, C. H. (2021). Mind the gap: Discrepancy between customer expectation and perception on commercial chatbot usage. Asian Journal of Empirical Research, 11(1), 1–10.

Cheng, Q., Shi, X., Yao, S., Liu, Y., Zhou, F., Zhang, Z., Akram, J., Braytee, A., & Anaissi, A. (2024). Optimized biomedical question-answering services with large language models and multi-BERT integration. arXiv.

Doungtap, S., Wang, J., & Phanichraksaphong, V. (2024). Comparative analysis of SaaS models and NPC integration for enhancing virtual reality shopping experiences. Applied Sciences, 14(15), 6573.

Ghosh, K., & Sur, C. (2025). Learning to extract cross-domain aspects and understanding sentiments using large language models. arXiv.

Harmanpreet, B., Harikrishna, G. N., & Dhulipalla, I. (2023). NLP for sentiment analysis, customer service automation, and market trend prediction. International Journal of

Science and Research Archive, 10(1), 1084–1092.

Heydari, M., Khazeni, M., & Soltanshahi, M. A. (2021). Deep learning-based sentiment analysis in the Persian language. In Proceedings of the International Conference on Web Research (pp. 287–292).

Hicham, N., & Nassera, H. (2024). Improving e-commerce customer review analysis for Arabic language opinion mining. International Journal of Computing, 23(3), 365–387.

International Journal of Multidisciplinary Research and Growth Evaluation. (2021). International Journal of Multidisciplinary Research and Growth Evaluation.

Iqbal, F., Afiat, A., Shoily, M. M., Turzo, S. S., & Arafat, M. S. (2025). AI-driven personalization in e-commerce: Evaluating the transformative effects on consumer behavior. International Journal of Science and Research Archive, 16(1), 264–273.

Kasap, A. (2025). E-commerce redefined: Harnessing emerging technologies for the future of trade. DergiPark.

Mashaabi, M., Alotaibi, A., Qudaih, H., Alnashwan, R., & Al-Khalifa, H. S. (2022). Natural language processing in customer service: A systematic review. arXiv.

Mukti, D. R., & Salam, A. (2025). Deployment architecture for a Flask-based chatbot using an LSTM NLP model for customer service. Journal of Applied Informatics and Computing, 9(3), 732–744.

Nagendra, N., & Chandra, J. (2024). Sentence classification using attention models for e-commerce product reviews. Journal of Computer Science, 20(5), 535–547.

Paul, R. K. (2023). AI integration in e-commerce business models: Case studies on Amazon FBA, Airbnb, and Turo operations. American Journal of Advanced Technology and Engineering Solutions, 3(3), 1–14.

Phadnis, N. (2025). Transforming customer engagement in digital commerce: The role of conversational AI frameworks. Journal of Information Systems Engineering and Management, 10, 411–425.

Prasetyo, A., Lestari, R. D., & Hartono, R. (2024). Using natural language processing to enhance customer sentiment analysis in e-commerce. International Journal of Intelligent E-Commerce Systems, 1(2), 19–29.

Rahevar, M., & Darji, S. (2024). Adoption of AI-driven chatbots in e-commerce recommendation systems. International Journal of

Management, Economics and Commerce, 1(2), 128–135.

Rahman, A. (2024). AI and machine learning in business process automation: Enhancing operational efficiency and customer experience. Non-Human Journal, 1(1), 41–48.

Wu, Q., Xia, C., & Tian, S. (2025). AI-driven sentiment analytics: Unlocking business value in the e-commerce landscape. arXiv.

Zhuk, A., & Yatskyi, O. (2024). The use of artificial intelligence and machine learning in e-commerce marketing. Technology Audit and Production Reserves, 3, 33–41.