

Contextual Preference Cognition through XLNet Semantic Encoding and Trivariate Decision Intelligence

Gajavelli Harshitha, Ramesh G*

Vaagdevi Engineering College, Warangal, 506005, Telangana, India.

*Correspondence: Ramesh G (gadde.ramesh@gmail.com)

ABSTRACT

Customer support platforms generate an enormous number of service requests every day, making it difficult for organisations to analyse customer feedback and deliver timely responses. Extracting useful information from these text-based interactions is essential for improving service quality, customer retention, and operational efficiency. This research presents a multi-target prediction model that simultaneously determines the priority of a support ticket, classifies the expected customer satisfaction level, and predicts the possibility of successful issue resolution. Unlike conventional approaches that address each prediction task independently, the proposed framework exploits the relationship among these outputs to produce more reliable and consistent results. Existing customer support systems often depend on manual inspection, keyword-based techniques, or simple text analysis methods, which struggle to process large volumes of unstructured textual data and frequently produce inconsistent outcomes. To overcome these limitations, this work utilises the XLNet transformer model to generate rich contextual representations of customer messages. These features are evaluated using several classification algorithms, including Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Stochastic Gradient Descent (SGD), and Nearest Centroid (NC). Their performance is compared with the Histogram-Based Gradient Boosting (HGB) classifier, which demonstrates the highest prediction accuracy and overall effectiveness. The proposed framework provides a scalable and

intelligent solution for customer support analytics by enabling faster ticket assessment, more accurate satisfaction estimation, and improved resolution forecasting, thereby assisting organisations in making better service decisions and enhancing the overall customer experience.

Keywords: Multi-Class Classification, Text Mining, Natural Language Processing (NLP), XLNet Embeddings, Machine Learning Algorithms, Quadratic Discriminant Analysis (QDA).

1. INTRODUCTION

The rapid expansion of digital commerce has significantly transformed the way businesses interact with customers, resulting in a substantial increase in online transactions and customer engagement across e-commerce platforms. Every customer interaction, including product reviews, service inquiries, complaint descriptions, live chat conversations, support requests, and feedback submissions, generates valuable textual data that reflects customer experiences, expectations, and satisfaction levels. Unlike structured transactional records, these textual interactions contain rich contextual information that provides deeper insights into customer opinions, preferences, emotions, and purchasing behavior. Effectively analysing such unstructured data enables organizations to identify recurring customer concerns, evaluate service quality, improve product offerings, and develop customer-centric business strategies. As online shopping continues to expand, the ability to transform large volumes of textual customer interactions into actionable business intelligence has become increasingly important for maintaining customer satisfaction, improving operational efficiency, and

achieving competitive advantage.

Customer communication is now predominantly conducted through digital channels such as live chat systems, email support, messaging applications, and online customer service portals, where written conversations create a permanent and valuable source of business information. These text-based interactions can be effectively processed using Natural Language Processing (NLP), text mining, and advanced machine learning techniques to automatically identify customer intent, detect sentiment and emotional expressions, classify customer requests, estimate the urgency of issues, and predict overall customer satisfaction. Automated analysis of customer conversations reduces the need for manual review while enabling organizations to provide faster responses, personalize customer support, prioritize critical service requests, and improve decision-making through data-driven insights. Furthermore, continuous monitoring of customer interactions allows businesses to evaluate service performance, identify emerging trends, strengthen customer relationships, and optimize support operations, ultimately contributing to enhanced user experience, improved business performance, and long-term customer loyalty.

2. LITERATURE SURVEY

Khan, et al. [1] proposed a framework for identifying Non-Functional Requirements (NFRs) from software requirement specification documents using advanced transfer learning models, including XLNet, BERT, DistilBERT, DistilRoBERTa, and ELECTRA. The study addressed the challenge of manually identifying NFRs, which is often time-consuming and prone to human error. By leveraging pre-trained transformer-based language models, the proposed approach automatically classified software requirements with improved accuracy and reduced manual effort. The comparative evaluation demonstrated that transfer learning

significantly enhanced the identification of software quality attributes, thereby supporting more reliable software development and requirement engineering processes. Filieri, et al. [2] investigated the application of Artificial Intelligence (AI) for emotion detection in customer experience analysis. The study highlighted that traditional qualitative analysis provides detailed customer insights but requires extensive manual interpretation of customer feedback. AI-based emotion recognition techniques reduced this workload by automatically identifying customer emotions from textual data. The research further revealed that customers generally expressed positive experiences with service robots, particularly when the robots demonstrated natural movements and interactive behaviors, thereby improving customer satisfaction. Li, Shugang, et al. [3] emphasized the importance of User-Generated Content (UGC), including online reviews, ratings, and customer comments, as valuable sources of business intelligence. The study utilized semantic analysis and sentiment analysis techniques to extract customer opinions and identify hidden patterns from textual feedback. These insights enabled organizations to improve product quality, enhance service delivery, strengthen brand reputation, and make more informed business decisions. Liu, Zhizhong, et al. [4] proposed a reliable service recommendation framework for Multi-access Edge Computing (MEC) environments by considering both user preferences and service quality attributes. The recommendation model evaluated service reliability, user requirements, and Quality of Service (QoS) parameters to identify the most appropriate services for users. The proposed framework improved recommendation accuracy, enhanced service reliability, and provided better user experiences in edge computing environments. Guda, et al. [5] developed a sentiment analysis model for restaurant reviews using the Yelp dataset by combining textual reviews with additional

contextual features. The study employed a multi-task BERT architecture to classify customer opinions more accurately than conventional machine learning methods. Experimental results demonstrated improved sentiment classification performance, enabling restaurants to better understand customer preferences and improve service quality.

Ameer, et al. [6] examined the growing importance of social media platforms as communication channels during the COVID-19 pandemic and proposed AI-based techniques for detecting mental health conditions through Reddit discussions. The study utilized Natural Language Processing (NLP) and deep learning models to identify symptoms related to depression, anxiety, Attention Deficit Hyperactivity Disorder (ADHD), bipolar disorder, and Post-Traumatic Stress Disorder (PTSD). The findings demonstrated that social media analysis can serve as an effective tool for early mental health assessment and public health monitoring. Chen, et al. [7] developed a customer response prediction framework for online shopping environments using multimodal deep learning techniques. The proposed system simultaneously analysed customer speech and textual conversations during voice-based interactions to predict customer satisfaction levels and purchasing intentions. By combining speech recognition with sentiment analysis, the framework improved customer behavior prediction and assisted businesses in enhancing customer engagement and sales performance. AbdElminaam, et al. [8] investigated employee turnover prediction using several Artificial Intelligence and machine learning algorithms, including Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), AdaBoost, Naïve Bayes, and Gradient Boosting Machine (GBM). The study highlighted that accurate employee churn prediction enables organizations to reduce recruitment costs, improve workforce retention strategies, and

support effective human resource management. Habbat, et al. [9] reviewed recent advances in sentiment analysis using Natural Language Processing and deep learning techniques. The study emphasized that modern transformer-based architectures, along with Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), significantly improve the accuracy of opinion mining and customer sentiment classification across diverse application domains. Ferreira, et al. [10] investigated the role of Industry 4.0 technologies, including the Internet of Things (IoT), Cyber-Physical Systems (CPS), Internet of Services (IoS), and advanced data analytics in improving industrial productivity. The study demonstrated that integrating these technologies supports Zero-Defect Manufacturing (ZDM), Condition-Based Maintenance (CBM), predictive maintenance, and intelligent manufacturing processes, thereby enhancing production efficiency and reducing operational costs.

Kufile, et al. [11] proposed a comprehensive brand sentiment analysis framework that continuously collected customer opinions from multiple social media platforms using real-time data acquisition techniques. The framework integrated Natural Language Processing, sentiment classification algorithms, and business intelligence tools to monitor brand reputation and identify customer perceptions. The generated insights enabled organizations to improve marketing strategies, customer engagement, and product development. Shah, et al. [12] examined the adoption of Artificial Intelligence technologies within modern contact centers, highlighting the increasing use of machine learning, Natural Language Processing, intelligent chatbots, and automated voice systems for enhancing customer support services. The study reported that these technologies became particularly important during the COVID-19 pandemic by enabling organizations to maintain efficient customer service while reducing operational costs and

response times. Bhuiyan, et al. [13] explored the application of Artificial Intelligence, predictive analytics, and real-time customer data in modern revenue management systems. The research demonstrated that incorporating customer sentiment analysis from online reviews allows organizations to optimize pricing strategies, improve demand forecasting, and enhance overall business profitability beyond traditional historical data analysis methods. Sirrianni, et al. [14] investigated the effectiveness of Generative Pre-trained Transformer (GPT) models for Natural Language Processing tasks involving limited training data. The study compared GPT with XLNet for next-word prediction in healthcare records and evaluated their predictive performance. Experimental findings showed that transformer-based language models effectively capture contextual information and generate accurate predictions, demonstrating their suitability for healthcare text analysis and other domain-specific NLP applications.

3. PROPOSED SYSTEM

This research is designed to predict customer satisfaction and ticket priority from customer responses using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The process begins with uploading the customer response dataset, followed by text preprocessing to clean and prepare the data for analysis. As shown in the figure 1, the processed text is converted into numerical features using deep learning and transformer-based embedding techniques. These features are then passed to multiple classification models to predict customer satisfaction, ticket priority, and resolution category. Finally, the results are displayed through a user-friendly graphical interface, allowing users to perform real-time predictions and generate reports efficiently. This architecture ensures accurate text analysis, faster processing, and effective decision-making for customer support applications.

Step 1: Data Upload and Inspection

The system starts by uploading the customer

response dataset in CSV format. The uploaded data is checked for missing values, incorrect entries, and overall structure. This ensures that only clean and valid data is used for further processing.

Step 2: NLP Data Preprocessing

The uploaded text is cleaned using NLP techniques such as lowercasing, tokenization, stopwords removal, and lemmatization. These preprocessing steps remove unnecessary words and convert the text into a meaningful format that improves model performance.

Step 3: Feature Extraction

After preprocessing, the cleaned text is transformed into numerical features. The system uses both deep learning methods (Keras text padding) and transformer-based embeddings (XLNet) to capture the semantic meaning and contextual information of customer responses.

Step 4: Predictive Classification

The extracted features are given to different machine learning and deep learning models, including Histogram Gradient Boosting (HGB), LDA, SGD, and Deep Neural Networks. These models classify the customer responses and predict the most suitable output with high accuracy.

Step 5: GUI Interaction and Prediction

The user enters customer ticket information through a graphical user interface (GUI). The trained model processes the input instantly and predicts the ticket priority, customer satisfaction score, and possible resolution category. The predicted results are displayed in an easy-to-understand format, helping customer support teams make faster decisions.

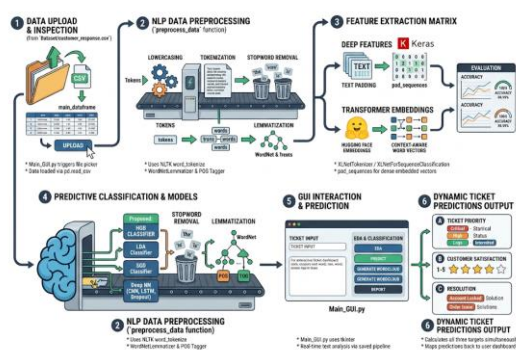


Figure 1: Proposed system architecture for customer satisfaction prediction using xlnet-based text mining

4. RESULTS ANALYSIS

Figure 2 shows the graphical user interface of the proposed XLNet-driven trivariate classification system developed using Tkinter. The interface presents a structured and user-friendly layout that allows users to execute the complete workflow, including dataset loading, preprocessing, exploratory data analysis, XLNet feature extraction, classifier execution (QDA, LDA, SGD, NC, and HGB), prediction, and comparison of performance metrics. The title banner clearly highlights the objective of high-precision customer satisfaction prediction, while the left panel contains function-specific control buttons for stepwise execution. The central display area is used to visualize dataset previews and intermediate outputs, and the integrated text console provides real-time feedback during processing. Figure 4 illustrates how the front-end interface effectively integrates user interaction with backend NLP and machine learning operations in a cohesive manner.

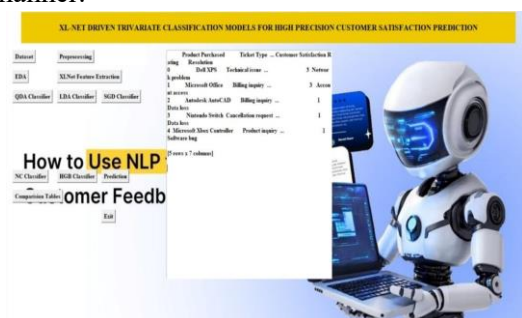


Figure 2: GUI of research work

Figure 3 shows the confusion matrix obtained

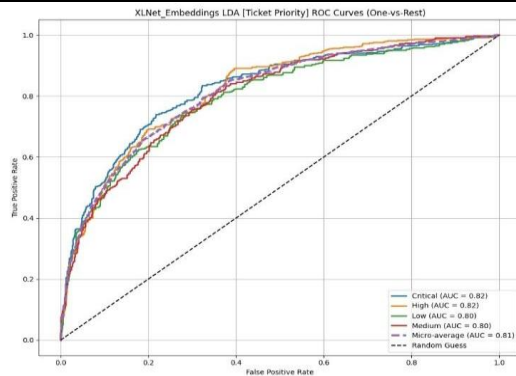
for Ticket Priority classification using XLNet embeddings with the QDA classifier. The matrix illustrates the distribution of correctly and incorrectly classified instances across four priority classes, namely Low, Medium, High, and Critical. It can be observed that a significant number of samples are correctly predicted along the diagonal, indicating the model's ability to capture contextual priority-related patterns from textual data. However, some misclassifications occur between adjacent priority levels, such as High and Medium or Low and Medium, which is expected due to semantic overlap in customer issue descriptions. Figure 3 demonstrates the effectiveness of combining XLNet-based contextual representations with QDA for multi-class ticket priority prediction while also highlighting areas where class boundary confusion exists.

XLNet_Embeddings QDA [Ticket Priority] Confusion Matrix

	Critical	High	Low	Medium
True Class				
Medium	20	33	210	284
Low	24	40	250	198
High	21	47	259	241
Critical	24	41	254	275
	Critical	High	Low	Medium
	Predicted Class			

Figure 3: Confusion matrix of XLNet_Embeddings QDA

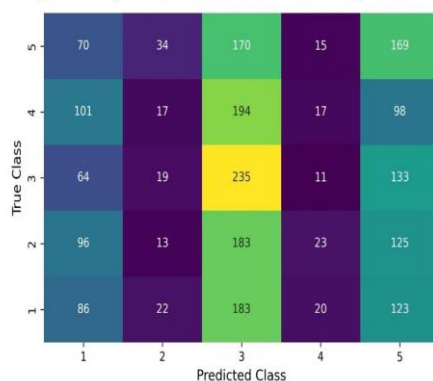
Figure 4 shows the one-vs-rest ROC curves for ticket priority classification using XLNet embeddings with the LDA classifier. All priority classes exhibit ROC curves well above the random guess line, indicating effective discrimination among Critical, High, Low, and Medium tickets. The AUC values range from approximately 0.80 to 0.82, with the micro-average AUC of 0.81 reflecting stable overall performance. This result demonstrates the model's ability to reasonably distinguish ticket priorities in a multi-class setting.



**Figure 4: ROC Curve of
XLNet_Embeddings LDA**

Figure 5 shows the confusion matrix for customer satisfaction rating prediction using XLNet embeddings with the NC classifier. The model demonstrates the highest correct predictions along the diagonal, particularly for the mid-level rating class (rating 3), indicating better recognition of neutral customer feedback. Some confusion is observed between adjacent rating levels, such as ratings 2, 3, and 4, reflecting their close semantic nature. The matrix suggests that the model captures general satisfaction trends effectively while facing challenges in fine-grained rating separation.

XLNet_Embeddings NC [Customer Satisfaction Rating] Confusion Matrix

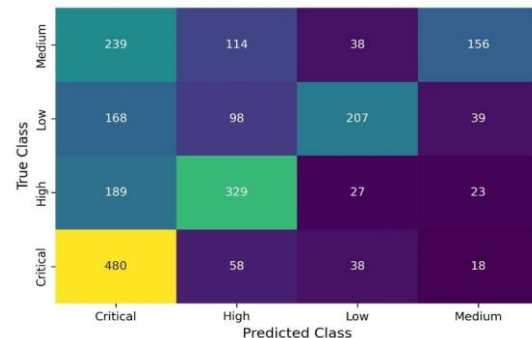


**Figure 5: Confusion Matrix of
XLNet_Embeddings NC**

Figure 6 shows the confusion matrix for the XLNet embeddings combined with the Stochastic Gradient Descent (SGD) classifier for ticket priority prediction. The figure indicates that the model performs relatively better in identifying High priority tickets, as reflected by a higher number of correct classifications along the diagonal. However,

notable misclassifications are observed between Critical and Medium priorities, where several critical tickets are incorrectly predicted as lower priority classes. This highlights the challenge faced by the SGD classifier in clearly separating closely related priority levels when using high-dimensional XLNet feature representations.

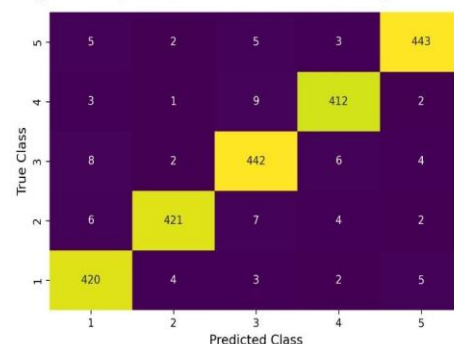
XLNet_Embeddings SGD [Ticket Priority] Confusion Matrix



**Figure 6: Confusion matrix of
XLNet_Embeddings SGD**

Figure 7 shows the confusion matrix of the XLNet embeddings combined with the Histogram-Based Gradient Boosting (HGB) classifier for customer satisfaction rating prediction. The matrix exhibits strong diagonal dominance, with a high number of correct classifications for ratings 1 (420), 2 (421), 3 (442), 4 (412), and 5 (443). Only a small number of misclassifications occur between adjacent rating levels, indicating the model's high accuracy and robustness in predicting customer satisfaction scores using XLNet-based textual representations.

XLNet_Embeddings HGB [Customer Satisfaction Rating] Confusion Matrix



**Figure 7: Confusion Matrix of
XLNet_Embeddings HGB**

Figure 8 shows the confusion matrix of the

XLNet embeddings combined with the Histogram-Based Gradient Boosting (HGB) classifier for ticket priority classification. The matrix is strongly dominated by diagonal values, indicating a very high number of correct predictions across all classes, with 577 Critical, 551 High, 483 Low, and 525 Medium tickets accurately classified. Only a small number of misclassifications are observed between adjacent priority levels, demonstrating the model's strong discriminative capability and its effectiveness in minimizing confusion among ticket priority categories.

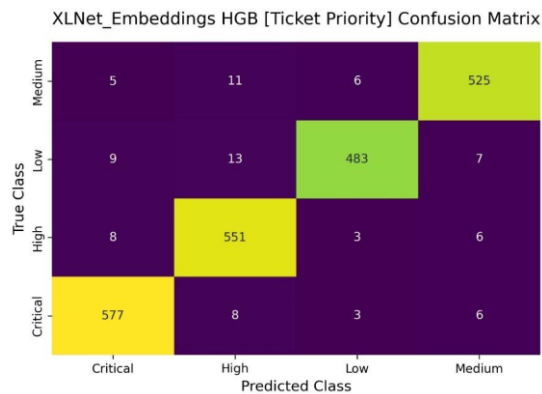


Figure 8: Confusion matrix of
XLNet_Embeddings HGB

Figure 9 shows the ROC curves for XLNet embeddings combined with the Histogram-Based Gradient Boosting (HGB) classifier for ticket priority classification using a one-vs-rest strategy. All priority classes—Critical, High, Medium, and Low—exhibit ROC curves that closely follow the top-left boundary, achieving an AUC value of 1.00 for each class as well as for the micro-average curve. This indicates near-perfect class separability and an exceptional ability of the HGB model to

distinguish between different ticket priorities when leveraging contextual XLNet representations, significantly outperforming the random guess baseline.

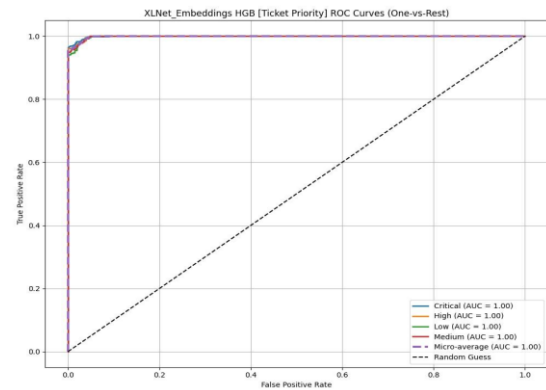


Figure 9: ROC Curve
XLNet_Embeddings HGB

Table 1 shows the overall performance comparison of different classifiers using XLNet embeddings in terms of accuracy, precision, recall, and F1-score. The QDA and Nearest Centroid (NC) models exhibit relatively low performance, achieving an accuracy of 27.24% and an F1-score of 22.59%, indicating their limited ability to model high-dimensional contextual embeddings. The SGD classifier demonstrates moderate improvement with 52.77% accuracy and 51.02% F1-score, while LDA further enhances classification reliability by achieving 57.63% accuracy and 57.45% F1-score. In contrast, the Histogram-Based Gradient Boosting (HGB) model significantly outperforms all other classifiers, attaining 96.17% accuracy, 96.23% precision, 96.11% recall, and 96.16% F1-score, demonstrating its superior capability in capturing complex non-linear patterns in XLNet-based feature space.

Table 1: Performance Comparison (Accuracy, Precision, Recall, F1-Score)

Method	Accuracy(%)	Precision (%)	Recall (%)	F1-Score (%)
QDA	27.24	27.24	27.58	22.59
LDA	57.63	57.63	57.46	57.45
SGD	52.77	52.77	58.10	51.02
NC	27.24	27.24	27.58	22.59
HGB	96.17	96.23	96.11	96.16

Table 2 shows the F1-score comparison, which

represents the harmonic balance between

precision and recall for each classifier. The QDA and NC models produce low F1-scores, particularly for the Critical class with a value of 0.07, highlighting weak classification capability. The SGD classifier achieves moderate F1-scores, reaching 0.57 for Critical and 0.56 for High, but performance decreases

for the Medium class. LDA demonstrates consistent F1-scores ranging from 0.54 to 0.61 across all ticket priorities. The HGB model achieves near-perfect F1-scores between 0.96 and 0.97, emphasizing its robust and balanced performance.

Table 2: F1-Score Comparison (Ticket Priority Classes)

Class	QDA	LDA	SGD	NC	HGB
Critical	0.07	0.61	0.57	0.07	0.97
High	0.13	0.60	0.56	0.13	0.96
Low	0.34	0.55	0.50	0.34	0.96
Medium	0.37	0.54	0.40	0.37	0.96

Table 3 shows the macro-average comparison of precision, recall, and F1-score across all classes, giving equal importance to each category. The QDA and NC models record low macro-F1 values of 0.23, reflecting poor generalization across multiple classes. The SGD classifier improves macro-F1 to 0.51,

while LDA further enhances balanced performance with a macro-F1 of 0.57. The HGB classifier achieves the highest macro-average values of 0.96 for precision, recall, and F1-score, indicating uniform and highly reliable classification performance across all classes.

Table 3: Macro-Average Comparison (All Tasks)

Method	Macro Precision	Macro Recall	Macro F1
QDA	0.28	0.28	0.23
LDA	0.57	0.57	0.57
SGD	0.58	0.52	0.51
NC	0.28	0.28	0.23
HGB	0.96	0.96	0.96

Table 4 shows the micro-average performance comparison, which aggregates contributions from all classes and closely aligns with overall accuracy. The QDA and NC models achieve low micro-average scores of 0.27, while the SGD and LDA classifiers demonstrate

moderate performance with values of 0.53 and 0.58, respectively. The HGB model again demonstrates outstanding results, achieving a micro-average precision, recall, and F1-score of 0.96, confirming its dominance in large-scale multi-class classification scenarios.

Table 4: Micro-Average Comparison (All Tasks)

Method	Micro Precision	Micro Recall	Micro F1
QDA	0.27	0.27	0.27
LDA	0.58	0.58	0.58
SGD	0.53	0.53	0.53
NC	0.27	0.27	0.27
HGB	0.96	0.96	0.96

5. CONCLUSION

The results of this study show that the proposed XLNet-based trivariate classification model is highly effective for customer support ticket analysis. Among all the machine learning models tested, the Histogram-Based Gradient Boosting (HGB) classifier achieved the best performance when used with XLNet embeddings. It obtained an accuracy of 96.17%, precision of 96.23%, recall of 96.11%, and an F1-score of 96.16%. These results were much better than baseline models such as QDA and Nearest Centroid, which achieved only 27.24% accuracy and a 0.23 F1-score. Other models like LDA and SGD also performed lower, with accuracies of 57.63% and 52.77%, showing that they were less effective in understanding complex text features. The proposed HGB model also performed well in classifying ticket priorities. It achieved 97% recall for Critical tickets, 97% for High, 94% for Low, and 96% for Medium priority tickets. The macro-average and micro-average scores were both 0.96, showing balanced performance across all classes. The confusion matrix showed that most tickets were classified correctly, while the ROC curves achieved an AUC value of 1.00 for all three prediction tasks, indicating excellent classification performance. Overall, the results prove that combining XLNet contextual embeddings with the HGB classifier provides an accurate, reliable, and scalable solution for customer satisfaction and support ticket prediction.

REFERENCES

- [1] M. A. Khan et al., "Non-functional requirements identification and classification using transfer learning model," *IEEE Access*, vol. 11, pp. 74997–75005, 2023.
- [2] R. Filieri, Z. Lin, Y. Li, X. Lu, and X. Yang, "Customer emotions in service robot encounters: A hybrid machine-human intelligence approach," *Journal of Service Research*, vol. 25, no. 4, pp. 614–629, 2022.
- [3] S. Li et al., "Text mining of user-generated content (UGC) for business applications in e-commerce: A systematic review," *Mathematics*, vol. 10, no. 19, p. 3554, 2022.
- [4] Z. Liu et al., "Accurate and reliable service recommendation based on bilateral perception in multi-access edge computing," *IEEE Transactions on Services Computing*, vol. 16, no. 2, pp. 886–899, 2022.
- [5] B. P. R. Guda, M. Srivastava, and D. Karkhanis, "Sentiment analysis: Predicting Yelp scores," *arXiv preprint arXiv:2201.07999*, 2022.
- [6] I. Ameer, M. A. Arif, G. Sidorov, H. Gomez-Adorno, and A. Gelbukh, "Mental illness classification on social media texts using deep learning and transfer learning," *arXiv preprint arXiv:2207.01012*, 2022.
- [7] G. Chen, S. Xiao, C. Zhang, and H. Zhao, "A theory-driven deep learning method for voice chat-based customer response prediction," *Information Systems Research*, vol. 34, no. 4, pp. 1513–1532, 2023.
- [8] D. S. Abdelminaam et al., "EmpTurnoverML: An efficient model for employee turnover and customer churn prediction using machine learning algorithms," in *Proc. MIUCC*, 2023.
- [9] N. Habbat, H. Anoun, and L. Hassouni, "Combination of GRU and CNN deep learning models for sentiment analysis on French customer reviews using XLNet model," *IEEE Engineering Management Review*, vol. 51, no. 1, pp. 41–51, 2022.
- [10] C. Ferreira and G. Gonçalves, "Remaining useful life prediction and challenges: A literature review on the use of machine learning methods," *Journal of Manufacturing Systems*, vol. 63, pp. 550–562, 2022.
- [11] O. T. Kufire, B. O. Otokiti, A. Y. Onifade, B. Ogunwale, and C. H. Harriet, "A framework for integrating social listening data into brand sentiment analytics," *Journal of Frontiers in Multidisciplinary Research*, vol. 3, no. 1, pp. 393–402, 2022.
- [12] S. Shah et al., "A review of natural language processing in contact centre

automation,” Pattern Analysis and Applications, vol. 26, no. 3, pp. 823–846, 2023.

[13] M. Bhuiyan et al., “Advanced analytics and machine learning for revenue optimization in the hospitality industry: A comprehensive review of frameworks,” American Journal of Scholarly Research and Innovation, vol. 2, no. 2, pp. 52–74, 2023.

[14] J. Sirrianni, E. Sezgin, D. Claman, and S. L. Linwood, “Medical text prediction and suggestion using generative pretrained transformer models with dental medical notes,” Methods of Information in Medicine, vol. 61, no. 5–6, pp. 195–200, 2022.