International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A Trivariate Customer Satisfaction Prediction Framework Using XLNet-Based Text Mining Models

S. Naresh[1], K. Chiranjeevi[2], B. Poojitha[2], K. Balakrishna[2]
[1]Assistant Professor, [1]Department of Computer Science and Engineering (AI & ML), Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

ABSTRACT

Customer satisfaction prediction has become increasingly critical as organizations receive thousands of support tickets daily, with over 80% of customers basing loyalty on timely and accurate service responses; therefore, real-time text analytics plays a vital role in transforming raw customer messages into actionable insights. This research focuses on a trivariate classification framework that simultaneously predicts ticket priority, customer satisfaction categories, and likelihood of issue resolution; however, traditional single-task approaches fail to capture the interconnected nature of these three dimensions, creating the need for an integrated model. The problem arises because many organizations still rely on manual methods for customer satisfaction prediction, including basic text mining and rule-based similarity checks, which are slow, inconsistent, and unable to handle large-scale unstructured data effectively. The objective of this study is to develop an XLNet-driven model combined with machine-learning methods such as Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), Stochastic Gradient Descent (SGD), and Nearest Centroid (NC), but these traditional classifiers produce lower accuracy, whereas the Histogram-Based Gradient Boosting (HGB) approach significantly enhances predictive performance.

Key words: Customer Satisfaction Prediction, Support Ticket Analysis, Trivariate Classification, Text Mining, Natural Language Processing (NLP), XLNet, Machine Learning, Quadratic Discriminant Analysis (QDA).

1. INTRODUCTION

The rapid growth of e-commerce platforms has resulted in millions of customers interacting with online businesses every day through product browsing, purchasing, reviewing, and contacting customer support. These interactions generate a large amount of unstructured text data — including product reviews, feedback messages, chat transcripts, and support requests. Understanding this text data is essential for companies to improve product quality, enhance customer experience, and strengthen customer satisfaction. Customer care interactions also play an important role, and most platforms provide two communication modes: phone calls and text-based messages or chats. While phone calls offer immediate conversation, text messages have become the preferred and more important option because they are quick, non-intrusive, easy to store, and generate rich textual data. This text data can be analyzed using text mining and advanced NLP techniques to understand customer mood, intention, urgency, and overall satisfaction. While phone calls offer immediate conversation, text messages have become the preferred and more important medium because they are quick, non-intrusive, easy to store, and provide a rich set of textual data. This text data can be analyzed through text mining and advanced NLP techniques to understand customer mood, intention, urgency, and satisfaction.

Figure 1 shows the customer analytics dashboard in the second image provides real-time insights through metrics like a 70% CSAT score, detailed star ratings, customer engagement levels, and NPS distribution (55% promoters, 23% passives, 22% detractors). These metrics help identify strengths, weaknesses, and improvement opportunities. Together, the images emphasize that frequent customer experience evaluation combined with data-driven satisfaction analysis leads to higher customer

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

satisfaction, stronger loyalty, and better overall business performance. The two images together highlight the powerful link between continuous customer experience evaluation and strong business performance. The first image clearly shows that companies that frequently review customer touchpoint such as their website, product quality, sales process, and service policies achieve significantly higher success scores compared to those that evaluate only sometimes or not at all.

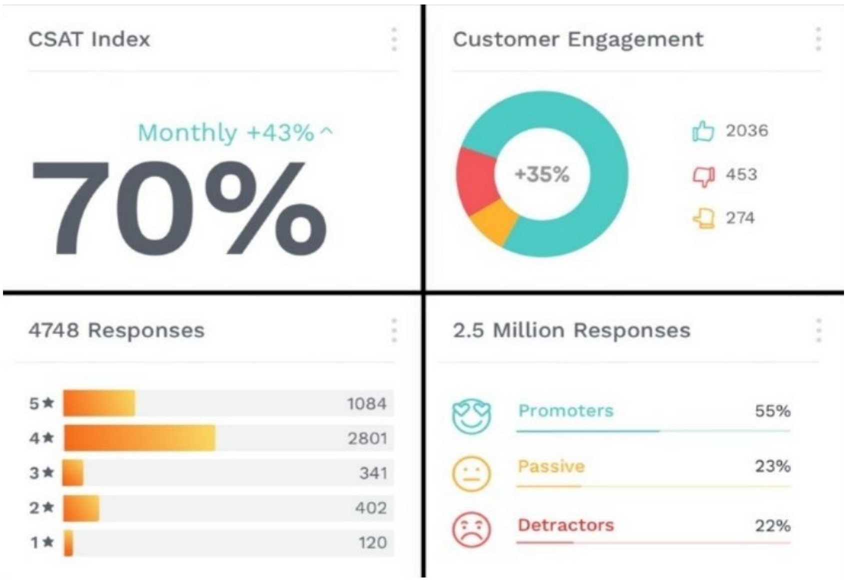

Figure 1: Customer Feedback Analysis

Figure 2 shows that regular customer feedback helps companies quickly identify issues, improve customer journeys, and deliver more satisfying experiences. Both images show that companies become more successful when they regularly check and understand customer experience. The first image shows that businesses that frequently evaluate their website, product, sales process, and services get higher success scores.

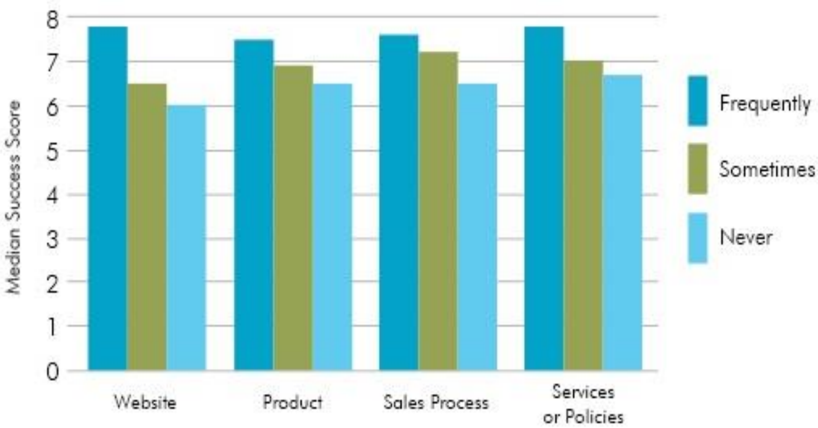

Figure 2: Customer Feedback and Ecommerce Success

Both figures emphasize that frequent evaluation paired with data-driven insights allows companies to make informed decisions, improve products and services, respond to customer needs faster, and build stronger long-term relationships. Ultimately, this leads to higher customer satisfaction, improved loyalty, better brand reputation, and overall business growth. Finally, the two images clearly show that customer experience evaluation is not just a support activity, but a key driver of business success.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

2. LITERATURE SURVEY

Habbat, et al. [1] proposed Sentiment analysis was the task of detecting opinions of people from text using techniques of natural language processing. It was critical in assisting businesses in actively improving their company strategy and better understanding client feedback on their products. Recently, the researchers showed that deep learning models, namely convolutional neural network (CNN), recurrent neural networks (RNNs), and contextualized transformer-based word embeddings, gave hopeful results for extracting sentiment from text.

Filieri, et al. [2] proposed Qualitative analysis had the advantage of generating rich insights from data, but it required intensive manual work. Scholars emphasized the benefits of using algorithms for recognizing and differentiating among emotions. The results showed that the customer experience with service robots was overwhelmingly positive, revealing that interacting with robots triggered emotions of joy, love, surprise, interest, and excitement. Discontent was mainly expressed when customers could not use service robots due to malfunctioning. Service robots triggered more emotions when they moved.

Kufile, et al. [3] explained Brand sentiment analytics played a crucial role in the strategic decision-making processes of contemporary businesses. However, traditional sentiment analysis frameworks often relied heavily on structured data and neglected the rich insights available through unstructured social listening data. Sentiment analytics focused on real-time data ingestion, natural language processing, sentiment classification, and business intelligence interpretation.

Ferreira, et al. [4] proposed Several approaches such as Cyber-Physical Systems (CPS), Internet of Things (IoT), Internet of Services (IoS), and Data Analytics built a new paradigm called Industry 4.0, which improved manufacturing efficiency and helped industries successfully face economic, social, and environmental challenges. In this scenario, the increased availability of data and processing tools favored, leveraged, and more easily applied approaches such as Zero-Defect Manufacturing (ZDM) and Condition-Based Maintenance (CBM).

Ameer, et al. [5] explained the social distance restrictions across the world, most individuals used social media as their major medium of communication. According to the World Health Organization (WHO), approximately 450 million people were affected. Mental illnesses, such as depression, anxiety, etc., were immensely common and affected individuals' physical health. This study analyzed unstructured user data

on the Reddit platform and classified five common mental illnesses: depression, anxiety, bipolar disorder, ADHD, and PTSD. We trained traditional machine learning, deep learning, and transfer learning multi-class models to detect mental disorders of individuals. This effort benefited the public health system by automating the detection process and informing appropriate authorities about people who required emergency assistance.

Singh, et al. [6] explained Conversational AI intended for machine-human interactions to appear and feel more natural and inclined to communicate in a near-human context. This paper analyzed recent works in the conversational AI domain, examining the exclusive methodologies, existing frameworks or tools, evaluation metrics, and available datasets for building robust conversational agents. Conversational agents used Natural Language Processing (NLP) and other AI techniques to engage in contextual discussion. To begin, an agent had to comprehend the context of the end user. Natural Language Understanding (NLU) drew insights from the users' words, independent of how they were told, including grammar, punctuation, and other faults. The system acquired information and recalled the context of past discussions. The agent then selected the correct response from a set of response variations using NLP or machine learning. As time passed, the response became more diverse, and the accuracy improved regularly. Finally, Natural Language Generation formulated the response naturally and understandably to the user.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Li, Shugang, et al. [7] proposed user-generated content (UGC) provides a valuable source of data to aid in understanding consumers and driving intelligent business. Text mining techniques, such as semantic analysis and sentiment analysis, help to extract meaningful information embedded in UGC. widely used text mining techniques are introduced, including semantic and sentiment analysis. Furthermore, we analyze the development status of semantic analysis in terms of text representation and semantic understanding. In the development of natural language processing (NLP) and deep learning techniques, UGC has gradually replaced consumer surveys as the critical method to understand consumers, consequently contributing to the improvement and innovation of products, services, and brands.

Sirrianni, et al. [8] explained Generative pretrained transformer (GPT) models were one of the latest large pretrained natural language processing models that enabled model training with limited datasets and reduced dependency on large datasets, which were scarce and costly to establish and maintain. There was a rising interest to explore the use of GPT models in health care. For comparison, we also fine-tuned a non-GPT pretrained neural network model, XLNet (large), for next word prediction. We annotated each token in 100 randomly sampled notes by category (e.g., names, abbreviations, clinical terms, punctuation, etc.) and compared the performance of each model by token category.

Guda, et al. [9] proposed this work, we predicted the sentiment of restaurant reviews based on a subset of the Yelp Open Dataset. We utilized the meta features and text available in the dataset and evaluated several machine learning and state-of-the-art deep learning approaches for the prediction task. Through several qualitative experiments, we showed the success of the deep models with attention mechanism in learning a balanced model for reviews across different restaurants. Finally, we proposed a novel multi-tasked joint BERT model that improved the overall classification performance.

Liu, Zhizhong, et al. [10] proposed Multi-access edge computing (MEC) is an emerging computing paradigm that brings services from the centralized cloud to nearby network edge to improve users' Quality of Experience (QoE). As massive services with dynamic Quality of Service (QoS) are available in MEC, it becomes challenging for users to find reliable services that satisfy their needs. To tackle this issue, an accurate and reliable service recommendation (ARSR) approach based on bilateral perception is proposed, which aims to proactively recommend reliable services by perceiving both users' service demands and multi-QoS of candidate services.

Chen, et al. [11] developed this study, we targeted the task of voice chat–based customer response prediction in an emerging online interaction–based commercial mode, the invite-online-and-experience-in-store mode. Prior research showed that satisfaction, which was revealed by the discrepancy between prior expectation and actual experience, was a key factor to disentangle customers' purchase intention, whereas black-box deep learning methods empirically promised us advantageous capabilities in dealing with complex voice data, for example, text and audio information incorporated in voice chat.

Khan, et al. [12] proposed this research study, we addressed the critical task of identifying and classifying non-functional requirements (NFRs) in software development. NFRs, described in the software requirements specification (SRS) document, offered a comprehensive system view and were closely aligned with software design and architecture. We evaluated multiple state-of-the-art transfer learning models, including XLNet, BERT, Distil BERT, Distil Roberta, Electra-base, and Electra-small, for this purpose. Using XLNet, we aimed to make software development easier, optimize resource usage, and improve the overall quality of software systems.

Bhuiyan, et al. [13] explained Traditional revenue management relied heavily on historical booking patterns, seasonal trends, and manual adjustments, which often lacked the flexibility to respond to sudden market fluctuations and evolving customer preferences. In contrast, modern revenue management incorporated real-time data processing, predictive analytics, and AI-powered decision-

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper
making to dynamically optimize pricing, demand forecasting, and customer segmentation. NLP-powered sentiment analysis played a crucial role in refining revenue strategies by analyzing customer feedback and online reviews.
Shah, et al. [14] proposed Contact centres had been highly valued by organizations for a long time. However, the COVID-19 pandemic highlighted their critical importance in ensuring business continuity, economic activity, and quality customer support. Next-generation platforms that incorporated machine learning techniques and natural language processing, such as self-service voice portals and chatbots, were implemented to enhance customer service. These platforms offered robust features that equipped customer agents with the necessary tools to provide exceptional customer support.
AbdElminaam, et al. [15] explained the loss of employees significantly impacted the company's growth and success, from decreased productivity to increased costs, including recruiting and training new employees. To mitigate this problem, many organizations adopted artificial intelligence (AI) techniques to detect employee churn. In this approach, AI algorithms like KNN, Decision Tree, Logistic Regression, Random Forest, SVM (Support Vector Machines), ADA boost, Naïve Bayes, and GBM (Gradient Boosted Machine Tree) were trained on historical data.

3. PROPOSED SYSTEM

The proposed system presents an XLNet-driven trivariate classification framework for high-precision customer satisfaction prediction using text mining. The system processes customer support ticket data and transforms unstructured textual information into meaningful contextual representations using XLNet. These representations are then used to build and compare multiple machine learning classifiers, with a focus on a proposed XLNet–Histogram Gradient Boosting model for enhanced predictive performance. As shown as figure 3 the framework simultaneously predicts ticket priority, customer satisfaction level, and resolution status, enabling efficient ticket handling and improved customer service decision-making. The system further integrates a graphical user interface using Tkinter to allow real-time prediction from raw text input in a user-friendly environment.

Step 1: Dataset Collection: The dataset consists of structured and unstructured attributes including product purchased, ticket type, ticket description, subject, ticket priority, customer satisfaction rating, and resolution status. These attributes collectively represent customer interactions and service outcomes, forming the foundation for trivariate classification.

Step 2: NLP Preprocessing: Textual fields such as ticket description and subject are cleaned and standardized through NLP preprocessing techniques. This step includes lowercasing, removal of noise, punctuation, and stopwords, as well as tokenization and normalization to ensure consistency and improve model learning.

Step 3: Exploratory Data Analysis: Exploratory data analysis is performed to understand data distribution, class balance, text length variations, and correlations among target variables. This step provides insights into data quality and helps identify patterns that influence model performance.

Step 4: XLNet Feature Extraction: Preprocessed text is passed through the XLNet model to generate deep contextual embeddings. XLNet captures semantic meaning, contextual dependencies, and sentiment information from customer feedback, producing high-dimensional feature vectors suitable for downstream classification tasks.

Step 5: Machine Learning Model Building: Multiple machine learning classifiers, including Nearest Centroid, Stochastic Gradient Descent, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Histogram Gradient Boosting, are implemented using the XLNet-generated feature vectors. This step establishes baseline and comparative models.

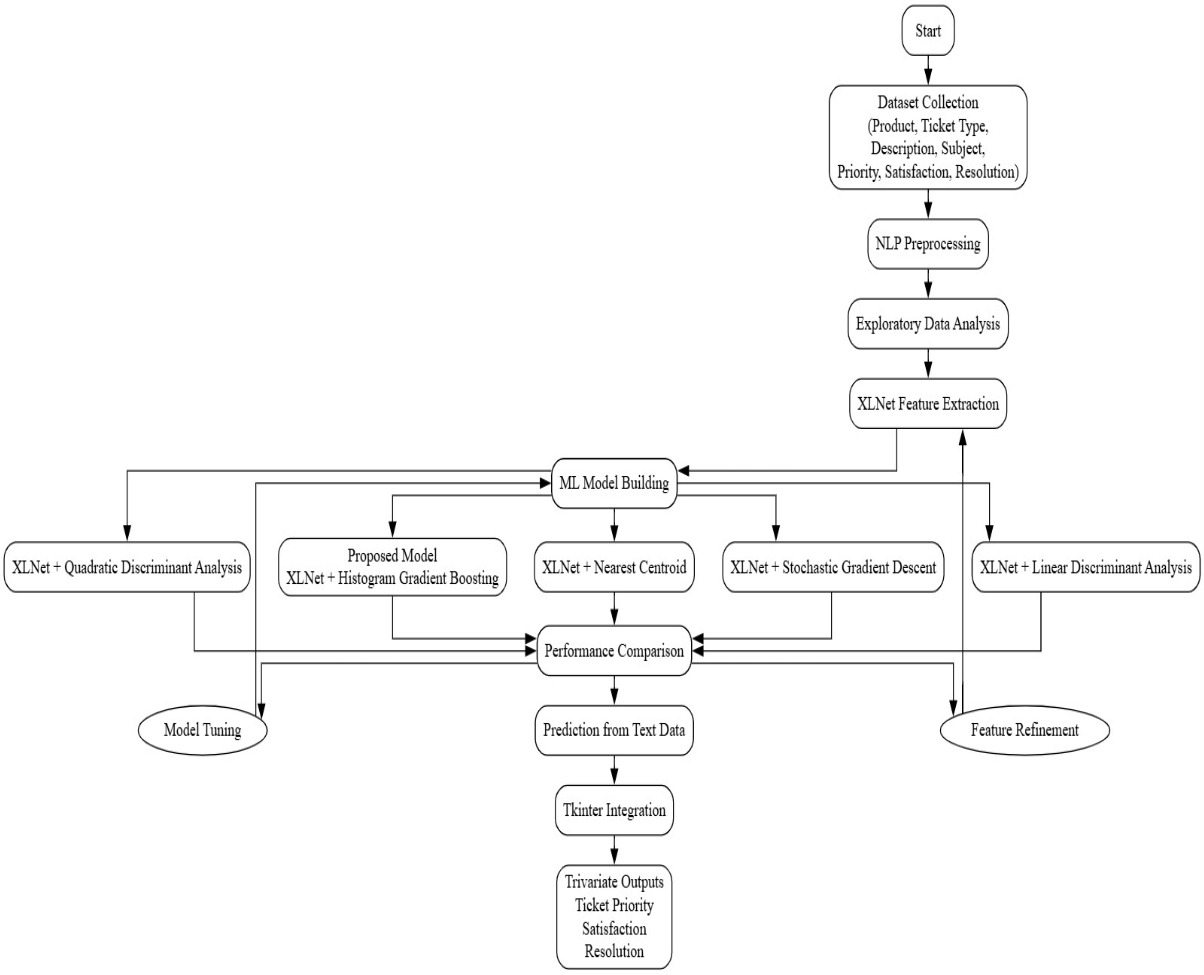

Figure 3: Proposed system architecture

Step 6: XLNet with Nearest Centroid: In this step, XLNet feature representations are classified using the Nearest Centroid algorithm. The model assigns classes based on proximity to class centroids, providing a simple and computationally efficient baseline for trivariate prediction.

Step 7: XLNet with Stochastic Gradient Descent: XLNet embeddings are combined with a Stochastic Gradient Descent classifier to perform large-scale linear classification. This approach enables fast learning and efficient optimization for high-dimensional text features.

Step 8: XLNet with Linear Discriminant Analysis: Linear Discriminant Analysis is applied to XLNet features to maximize class separability. This model assumes linear decision boundaries and helps evaluate discriminative performance under linear constraints.

Step 9: XLNet with Quadratic Discriminant Analysis: Quadratic Discriminant Analysis is used with XLNet embeddings to capture non-linear class boundaries. This step enhances classification flexibility by modelling class-specific covariance structures.

Step 10: Proposed Model XLNet with Histogram Gradient Boosting: The proposed model integrates XLNet feature extraction with Histogram Gradient Boosting for trivariate classification. This ensemble approach efficiently handles complex, non-linear relationships in high-dimensional feature space, resulting in improved prediction accuracy and robustness.

Step 11: Performance Comparison: All implemented models are evaluated and compared using standard performance metrics such as accuracy, precision, recall, and F1-score. This comparison highlights the effectiveness of the proposed model over baseline approaches.

Step 12: Prediction from Text Data: The trained models are used to generate predictions directly from

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

new customer ticket text. The system outputs predicted ticket priority, customer satisfaction level, and resolution status based on learned patterns.

Step 13: Integration with Tkinter: The final system is integrated with a Tkinter-based graphical user interface. This interface allows users to input ticket text and receive real-time classification results, making the system practical for real-world customer support applications.

Classification Outputs:

- Ticket Priority
- Customer Satisfaction
- Resolution

4. RESULTS ANALYSIS

Figure 4 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.

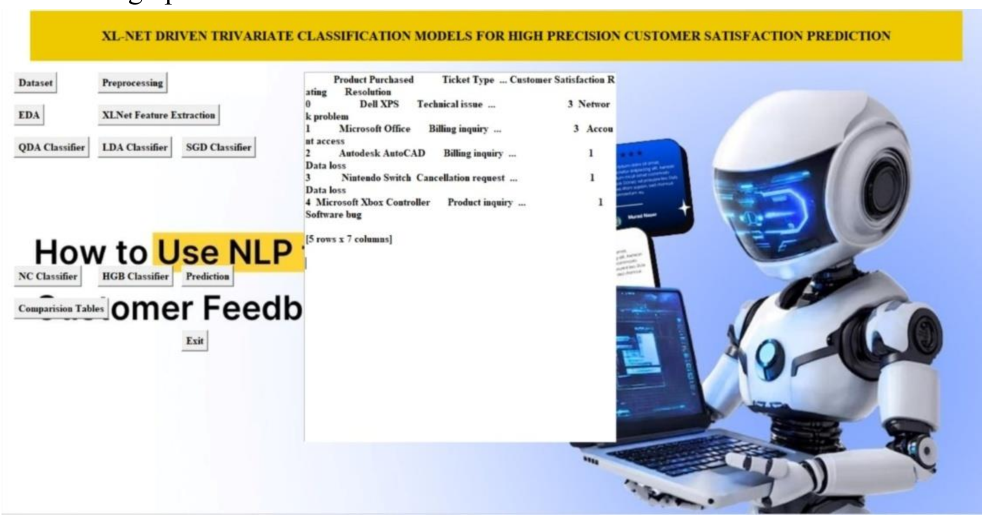

Figure 4: GUI of research work

Figure 5 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 5 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.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

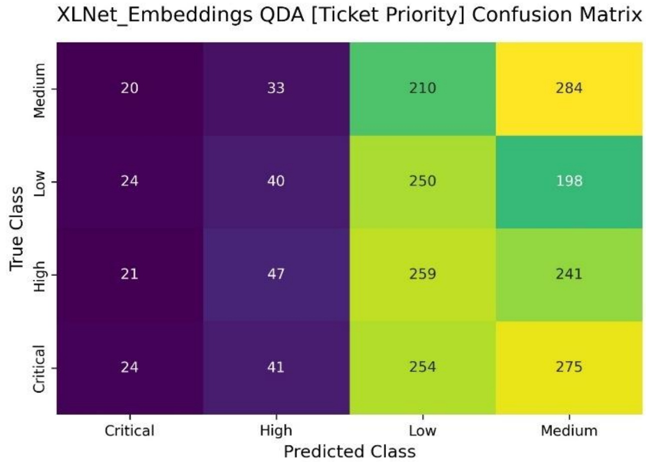

Figure 5: Confusion matrix of XLNet_Embeddings QDA

Figure 6 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.

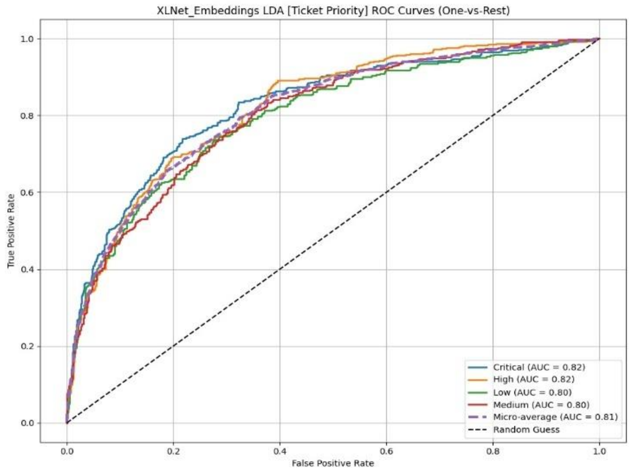

Figure 6: ROC Curve of XLNet_Embeddings LDA

Figure 7 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.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Figure 7: Confusion Matrix of XLNet_Embeddings NC

Figure 8 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.

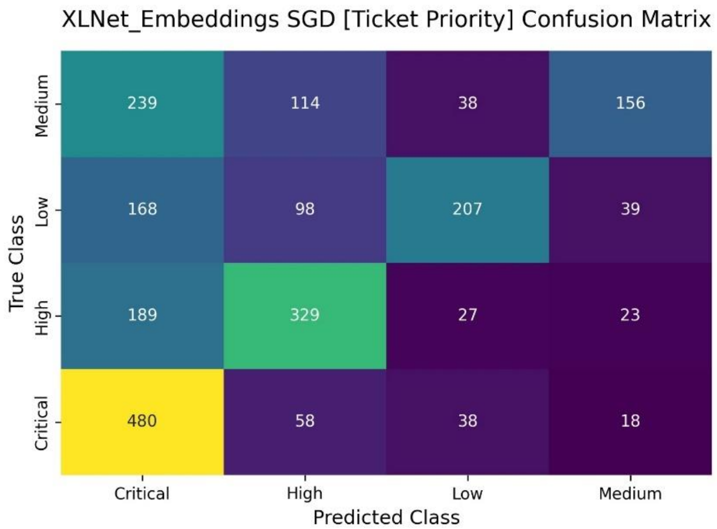

Figure 8: Confusion matrix of XLNet_Embeddings SGD

Figure 9 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.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Figure 9: Confusion Matrix of XLNet_Embeddings HGB

Figure 10 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.

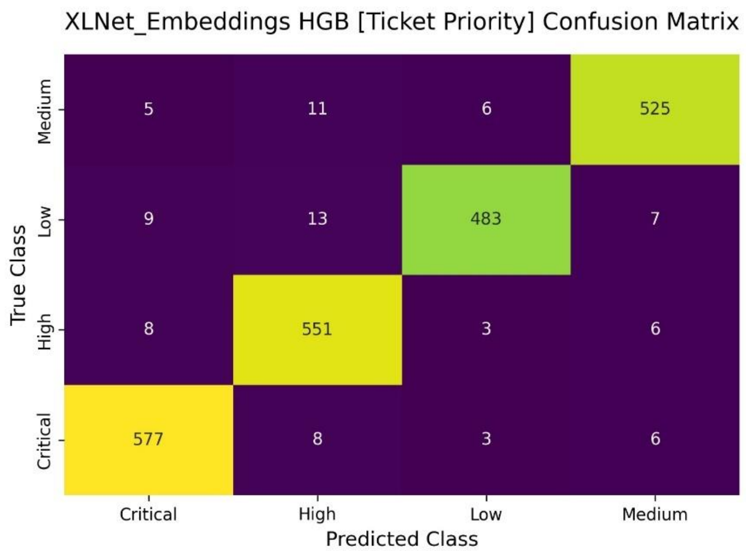

Figure 10: Confusion matrix of XLNet_Embeddings HGB

Figure 11 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.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

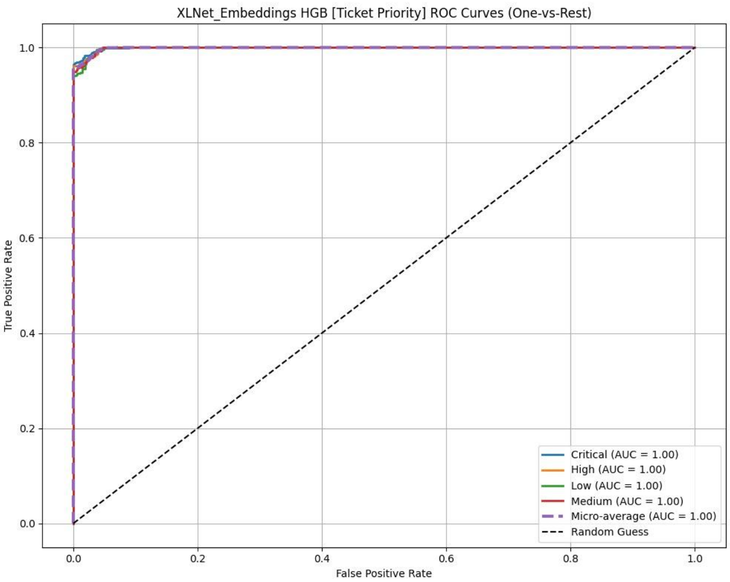

Figure 11: ROC Curve XLNet_Embeddings HGB

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

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

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

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	Marco 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 obtained in this study clearly validate the effectiveness of the proposed XLNet-driven trivariate classification framework for customer support ticket analysis. Among all evaluated classifiers, the Histogram-Based Gradient Boosting (HGB) model demonstrated outstanding performance when combined with XLNet contextual embeddings. The proposed model achieved an overall accuracy of 96.17%, precision of 96.23%, recall of 96.11%, and an F1-score of 96.16%, significantly outperforming baseline models such as QDA and Nearest Centroid, which recorded only 27.24% accuracy and 0.23 macro-F1 score. Even comparatively stronger baselines like LDA and SGD achieved moderate performance, with accuracies of 57.63% and 52.77%, respectively, highlighting their limitations in capturing complex semantic relationships within high-dimensional XLNet embeddings.

Detailed class-wise evaluation further confirms the robustness of the proposed approach. For ticket priority classification, the HGB model achieved recall values of 0.97 for Critical, 0.97 for High, 0.94 for Low, and 0.96 for Medium, ensuring reliable identification of urgent tickets. Macro-average and micro-average metrics across all tasks reached 0.96, indicating both balanced class-wise performance and strong scalability. Confusion matrices and ROC curves revealed near-perfect diagonal dominance and AUC values of 1.00 for all three prediction tasks, demonstrating excellent class separability. The experimental findings establish that integrating transformer-based contextual representations with non-linear ensemble learning provides a highly accurate, stable, and scalable solution for multi-output customer satisfaction prediction.

REFERENCES

[1] 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.

[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] O. T. Kufile, 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.

[4] 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.

[5] 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.

[6] S. Singh and H. Beniwal, “A survey on near-human conversational agents,” Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, pp. 8852–8866, 2022.

[7] 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.

[8] 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.

[9] B. P. R. Guda, M. Srivastava, and D. Karkhanis, “Sentiment analysis: Predicting Yelp scores,” arXiv preprint arXiv:2201.07999, 2022.

[10] 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.

[11] 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.

[12] M. A. Khan et al., “Non-functional requirements identification and classification using transfer learning model,” IEEE Access, vol. 11, pp. 74997–75005, 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] 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.

[15] D. S. AbdElminaam et al., “EmpTurnoverML: An efficient model for employee turnover and customer churn prediction using machine learning algorithms,” in Proc. MIUCC, 2023.