

Conversational Intelligence via Deep Feature Extraction and Balanced Semantic Representation for Next-Gen Customer Support Automation

B. Vara Lakshmi^{1*}, K. Raveendra Chaitanya¹, Vennapusa Preethi², Shaik Jasmin², Siddu Venkata Sujitha Reddy², Talari Lakshmi Sunayana²

¹Assistant Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering

^{1,2}Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

*Correspondence: B. Vara Lakshmi (varalakshmi@gist.edu.in)

ABSTRACT

The rapid growth of conversational Artificial Intelligence (AI) is transforming customer support systems, with the global market expected to reach USD 32 billion by 2030 and chatbots projected to handle a majority of customer interactions. Despite this advancement, manual intent annotation and query classification remain labor-intensive and inconsistent, posing challenges to scalability and efficiency. This study proposes an advanced Natural Language Processing (NLP) framework built on a Customer Support Bitext dataset annotated with multiple intents and categories. The approach begins with systematic NLP preprocessing and Exploratory Data Analysis (EDA), including text normalization, tokenization, and data distribution analysis to prepare the dataset for modeling. For feature extraction, the framework utilizes a Miniature Language Model (MiniLM), a lightweight yet context-aware language model capable of capturing semantic relationships effectively. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic samples for underrepresented classes, ensuring a more balanced dataset. In contrast to Machine Learning (ML) methods such as Decision Tree Classifier (DTC), K-Nearest Neighbors (KNN), and Naïve Bayes Classifier (NBC), the proposed model incorporates Deep Neural Network (DNN)-based feature selection combined with KNN classification to enhance predictive performance. The system is designed to perform bivariate prediction by identifying both intent and category, thereby improving contextual understanding of customer queries. The trained model is integrated into a chatbot interface to support real-time intent detection and automated response generation, resulting in improved classification accuracy and reduced inconsistencies in annotation.

Keywords: Conversational AI, NLP, Customer Support Chatbot, Text Classification, Multi-Intent Detection.

1. INTRODUCTION

Customer service has become a fundamental pillar for businesses worldwide, especially with the rapid expansion of e-commerce, continuous technological innovation, and rising customer expectations. Organizations are increasingly required to deliver quick, efficient, and personalized support to remain competitive. In India, this sector has experienced significant growth, fueled by the country's expanding digital ecosystem and the need to serve a large and diverse population. With over 1.3 billion people, providing consistent and scalable customer support presents a considerable challenge. The industry itself represents a substantial economic segment, with an estimated value of around \$32 billion, and India continues to be a major global destination for outsourced customer service due to its skilled, English-speaking workforce and cost advantages. Additionally, the country's growing middle class—projected to exceed 600 million by 2030—further increases the demand for efficient service delivery. The rapid shift toward digital platforms has transformed how businesses interact with customers. Communication channels such as emails, social media, and live chat have become dominant, replacing

traditional support methods. However, conventional systems that depend heavily on human agents struggle to keep up with this scale. These systems often suffer from delays, inconsistencies, and limited scalability, leading to reduced customer satisfaction and increased operational costs. As a result, businesses are actively exploring modern solutions that can automate routine processes while maintaining high levels of accuracy and responsiveness.

Artificial Intelligence (AI) and ML technologies have emerged as effective solutions to these challenges. Tools such as chatbots, intent classification systems, and automated response generators enable organizations to manage large volumes of interactions efficiently. These systems provide continuous support, reduce response times, and improve overall service quality. In India, where many customer service operations are still transitioning from manual processes, the adoption of AI-driven systems offers significant potential for improving both efficiency and customer experience. NLP plays a crucial role in enabling these intelligent systems by allowing machines to interpret and understand human language. With advancements in transformer-based models such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT), NLP systems have achieved higher levels of accuracy in understanding context and intent.

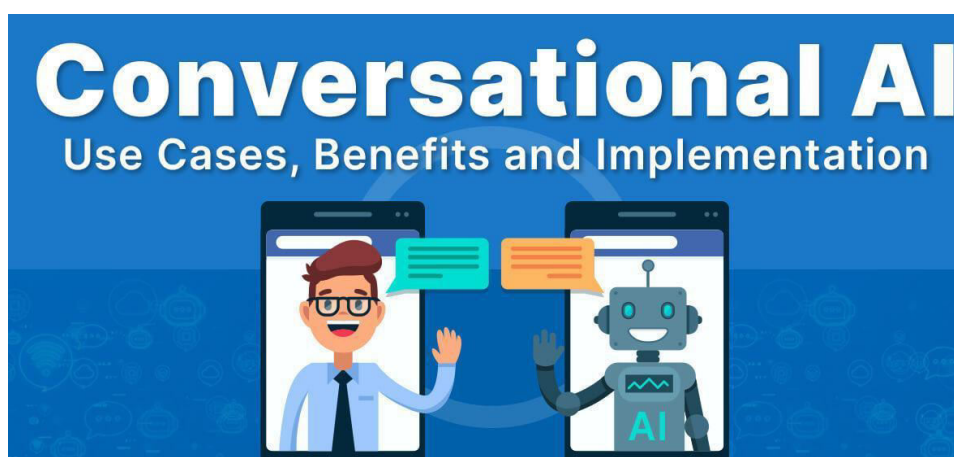


Fig. 1. Customer Support Dataset with Multi-Intent Annotations for Conversational AI

This is particularly important in a linguistically diverse country like India, where multiple languages and dialects are used in daily communication. The ability to process such varied linguistic data is essential for delivering effective automated support. As customer expectations continue to evolve toward faster and more personalized interactions, businesses in India are increasingly investing in AI and NLP-based solutions. These technologies not only automate processes but also enable more intelligent, context-aware interactions that enhance customer engagement and satisfaction.

2. LITERATURE SURVEY

Huang et al. [1] proposed approach uses different bidirectional gated recurrent unit (BiGRU) combined with attention mechanisms to encode the contextual semantic information of different types of conversational texts. For modeling affective interactions, we use directed graph structures to portray the affective interactions between speakers and encode them with affective interaction features using graph convolutional neural networks (GCN). Finally, the two features are fused to recognize customer sentiment. The experimental results on the JDDC dataset show that our model can more accurately recognize customer sentiment than other baseline models in customer service conversation. Wagner et al. [2] evaluated metrics tailored for multi-user contexts. We conduct a comprehensive analysis of relevant literature, employing both quantitative and qualitative methodologies to identify common patterns and challenges in multi-user interactions. The findings underscore the importance of

developing robust interfaces that can effectively manage overlapping dialogues, ensure collaborative group work, and enhance overall conversational quality. This work contributes to the understanding of Multi-User Conversational Interfaces and may help as a basis for future research aiming to develop more natural, user-friendly, and effective conversational interfaces. Lin et al. [3] provided an instant and automated response, which can be leveraged in many application areas. Chatbots can handle a wide range of inquiries and tasks, such as answering frequently asked questions, booking appointments, or making recommendations. Modern conversational chatbots use artificial intelligence (AI) techniques, such as natural language processing (NLP) and artificial neural networks, to understand and respond to users' input. In this study, we will explore the objectives of why chatbot systems were built and what key methodologies and datasets were leveraged to build a chatbot. Bercaru et al. [4] explored the influence of various feature vectors on the task of intent classification using RASA's text classification capabilities. The second part of this work consists of a generic method for efficiently augmenting textual corpora using large datasets of unlabeled data. The proposed method is able to efficiently mine for examples similar to the ones that are already present in standard, natural language corpora. The experimental results show that using our corpus augmentation methods enables an increase in text classification accuracy in few-shot settings. Particularly, the gains in accuracy raise up to 16% when the number of labeled examples is very low (e.g., two examples). We believe that our method is important for any Natural Language Processing (NLP) or NLU task in which labeled training data are scarce or expensive to obtain.

Chuang et al. [5] focused on developing conversational systems based on the Chinese corpus over military scenarios. The soldier will need information regarding their surroundings and orders to carry out their mission in an unfamiliar environment. Additionally, using a conversational military agent will help soldiers obtain immediate and relevant responses while reducing labor and cost requirements when performing repetitive tasks. This paper proposes a system architecture for conversational military agents based on natural language understanding (NLU) and natural language generation (NLG). The NLU phase comprises two tasks: intent detection and slot filling. Detecting intent and filling slots involves predicting the user's intent and extracting related entities. Nicolescu et al. [6] purposed of is to analyze the overall customer experience with customer service chatbots in order to identify the main influencing factors for customer experience with customer service chatbots and to identify the resulting dimensions of customer experience (such as perceptions/attitudes and feelings and also responses and behaviors). The analysis uses the systematic literature review (SLR) method and includes a sample of 40 publications that present empirical studies. The results illustrate that the main influencing factors of customer experience with chatbots are grouped in three categories: chatbot-related, customer-related, and context-related factors, where the chatbot-related factors are further categorized in: functional features of chatbots, system features of chatbots and anthropomorphic features of chatbots. Merdivan et al. [7] develop a benchmark dataset with human annotations and diverse replies that can be used to develop such a metric for conversational agents. The paper introduces a high-quality human annotated movie dialogue dataset, HUMOD, that is developed from the Cornell movie dialogues dataset. This new dataset comprises 28,500 human responses from 9500 multi-turn dialogue history-reply pairs. Human responses include: (i) ratings of the dialogue reply in relevance to the dialogue history; and (ii) unique dialogue replies for each dialogue history from the users. Izadi et al. [8] provided an overview of chatbots, conducts an analysis of errors they encounter, and examines different approaches to rectifying these errors. These approaches include using data-driven feedback loops, involving humans in the learning process, and adjusting through learning methods like reinforcement learning, supervised learning, unsupervised learning, semi-supervised learning, and meta-learning. Through real life examples and case studies in different fields, we explore how these strategies are implemented. Looking

ahead, we explore the different challenges faced by AI-powered chatbots, including ethical considerations and biases during implementation.

Al-Mutawa et al. [9] proposed artificial intelligence and data augmentation to predict customer satisfaction ratings from conversations by analyzing the responses of customers and service providers. For the study, the authors obtained actual conversations between customers and real agents from the call center database of Jeddah Municipality that were rated by customers on a scale of 1–5. They trained and tested five prediction models with approaches based on logistic regression, random forest, and ensemble-based deep learning, and fine-tuned two pre-trained recent models: ArabicT5 and SaudiBERT. Then, they repeated training and testing models after applying a data augmentation technique using the generative artificial intelligence, GPT-4, to improve the unbalance in customer conversation data. They found that the ensemble-based deep learning approach best predicts the five-, three-, and two-class classifications. Sultana et al. [10] explored content within their repositories. However, the keyword-based search cannot identify the users' search intent accurately. Integrating a query-understandable framework into keyword search engines has the potential to enhance their performance, bridging the gap in interpreting the user's search intent more effectively. In this study, we have proposed a novel approach that focuses on spatial and temporal information, phrase detection, and semantic similarity recognition to detect the user's intent from the search query. We have used the n-gram probabilistic language model for phrase detection. Coppola et al. [11] proposed to provide aid to the researchers and practitioners of the field. We came up with a final pool of 118 contributions, including grey (35) and white literature (83). We categorized 123 different quality attributes and metrics under ten different categories and four macro-categories: Relational, Conversational, User-Centered and Quantitative attributes. While Relational and Conversational attributes are most commonly explored by the scientific literature, we testified a predominance of User-Centered Attributes in industrial literature. We also identified five different academic frameworks/tools to automatically compute sets of metrics, and 28 datasets (subdivided into seven different categories based on the type of data contained) that can produce conversations for the evaluation of conversational interfaces. Allouch et al. [12] proposed technology has finally ripened to advance the use of CAs in various domains, including commercial, healthcare, educational, political, industrial, and personal domains. In this study, the main areas in which CAs are successful are described along with the main technologies that enable the creation of CAs. Capable of conducting ongoing communication with humans, CAs are encountered in natural-language processing, deep learning, and technologies that integrate emotional aspects. The technologies used for the evaluation of CAs and publicly available datasets are outlined. In addition, several areas for future research are identified to address moral and security issues, given the current state of CA-related technological developments.

3. PROPOSED SYSTEM

The proposed system aims to revolutionize customer service operations by integrating ML and NLP to automate the classification of customer intents and the generation of appropriate responses. The system involves multiple stages, starting with the preprocessing of text data to clean and normalize customer queries. Next, the features are extracted using advanced transformer-based models, such as MiniLM, to capture the semantic meaning of each query. The system then uses machine learning models, specifically DTC, KNN, and NBC, for intent classification. Additionally, DNN are used to further refine feature extraction and enhance prediction accuracy. The system includes a multi-output classification approach to simultaneously predict both customer intent and category, leveraging SMOTE for class balancing as shown in Fig. 2. Finally, the system provides automated, real-time responses to customers through a trained model, significantly improving response times and service quality. By employing

these techniques, the system is designed to scale with business growth, reduce operational costs, and provide a more efficient and personalized customer service experience.

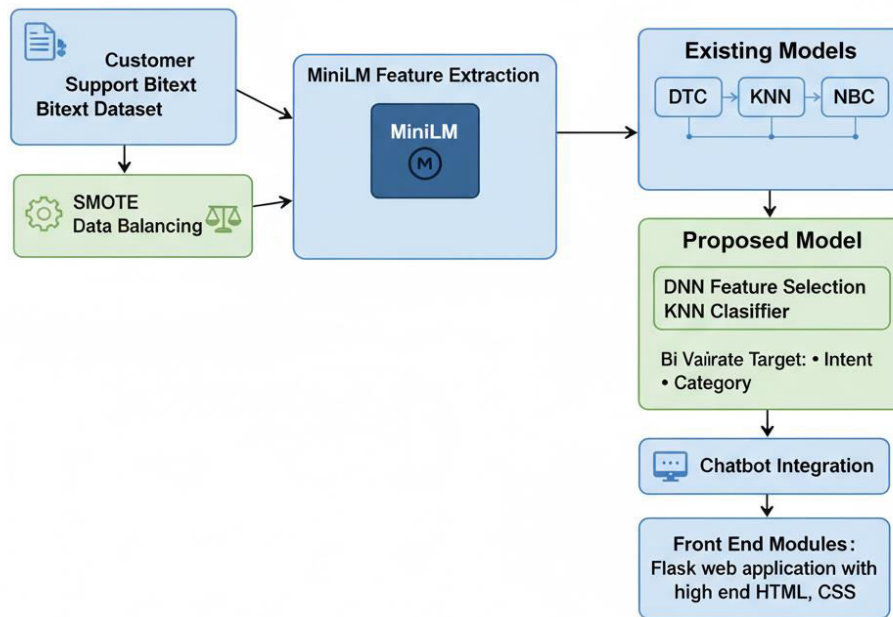


Fig. 2: Proposed System Architecture.

1. Customer Support Bitext Dataset: This step involves collecting paired text data (input–response) used for training the chatbot. The dataset typically contains customer queries and their corresponding support replies. It serves as the foundational data source for learning semantic relationships and intent patterns.

2. SMOTE Data Balancing: SMOTE is applied to handle class imbalance in the dataset. It generates synthetic samples for underrepresented classes to ensure the model does not become biased toward majority classes. This improves classification performance, especially for rare intents.

3. MiniLM Feature Extraction: MiniLM converts text into dense vector embeddings that capture semantic meaning. These embeddings represent sentences numerically while preserving contextual relationships. This step transforms raw text into machine-understandable features for model training.

4. Existing Models: machine learning models like DTC, KNN, NBC are trained on extracted features. These models serve as baseline approaches for comparison. Their performance helps evaluate the effectiveness of the proposed model.

5. Proposed Model: A hybrid model is used where a DNN performs feature selection and KNN handles classification. The DNN identifies the most relevant features, improving accuracy and reducing noise. This combination enhances prediction performance compared to standalone models.

6. Bivariate Target: The system predicts two outputs simultaneously: user intent and category. This multi-output approach allows more detailed understanding of user queries. It improves chatbot response precision by capturing both purpose and context.

7. Chatbot Integration: The trained model is integrated into a chatbot system for real-time interaction. It processes user inputs, predicts intent/category, and generates appropriate responses. This enables automated customer support functionality.

8. Front-End Modules: A Flask-based web interface is developed for user interaction with the chatbot. HTML and CSS are used to design a responsive and user-friendly interface. This layer connects users with the backend model seamlessly.

4. RESULTS AND DISCUSSION

Fig. 3 illustrates the Home Page Screen of the ConvoSense system, which serves as the central interface for initiating AI-driven customer support interactions. It depicts the integration of intelligent intent detection and real-time conversational capabilities within a unified platform. The figure represents how users can access core functionalities such as chat initiation and analytical dashboards through a structured interface. It also highlights multiple system features including AI intent detection, real-time responses, multi-intent handling, and continuous availability, which collectively enhance user interaction.

Fig. 4 illustrates the Real-Time AI Support Dashboard for Conversation Monitoring and Analytics, providing a quantitative overview of ongoing system activities and performance. It depicts key numerical indicators such as active conversations (1), average response time (713 ms), resolved issues (2), and user satisfaction score (96%), which collectively reflect system efficiency and user experience. The figure represents temporal analysis through a 24-hour conversation timeline, showing fluctuations in interaction volume over time. It also highlights intent distribution across categories such as technical, billing, account, product, and general, enabling insight into user query patterns. Additionally, the dashboard includes recent conversation logs with attributes like user ID, intent classification, resolution status, and interaction duration (e.g., 8 minutes), demonstrating real-time tracking and evaluation.

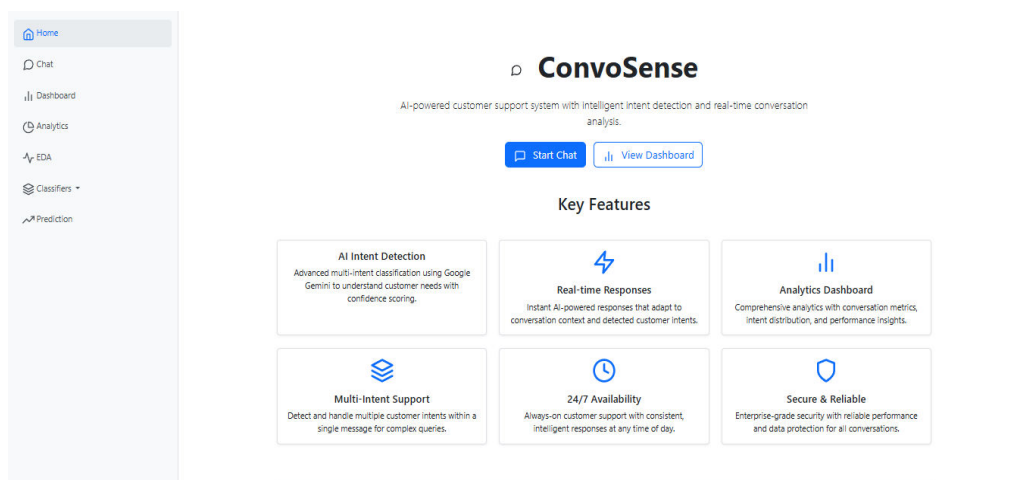


Fig. 3: Home page Screen.

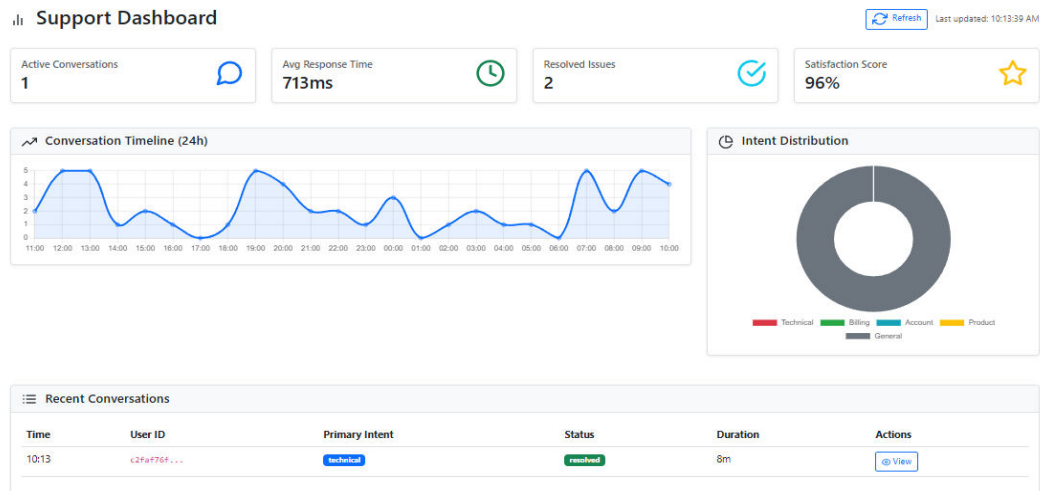


Fig. 4: Real-Time AI Support Dashboard for Conversation Monitoring and Analytics.

Fig. 5 illustrates the Advanced Analytics Dashboard for AI Support System Performance Monitoring, providing a comprehensive quantitative assessment of system behavior. It depicts key numerical performance indicators such as average response time (651.598 ms), throughput rate (19 requests per minute), active connections (1), and system error rate (3%), which collectively evaluate system responsiveness and reliability. The figure represents resource utilization metrics, including memory usage (59%) and CPU usage (27%), highlighting computational efficiency during real-time operations. It also presents statistical visualizations such as response time trends, confidence score distributions (high >70%, medium 40–70%, low <40%), and intent-wise accuracy comparisons across categories like technical, billing, account, product, and general.

Fig. 6 illustrates the Exploratory Data Analysis of the Customer Support Intent Dataset, emphasizing statistical insights and feature-level understanding of textual data. It depicts numerical distributions such as document length frequencies, word occurrence counts, and class-wise sample distributions, providing a quantitative overview of dataset characteristics. The figure represents the top 100 words through word cloud visualization and quantifies the top 20 most frequent words and bigrams using frequency counts, revealing dominant linguistic patterns. It also highlights part-of-speech (POS) tag frequency distributions, indicating syntactic structure and language usage trends within the dataset. Additionally, the intent class distribution graph presents the balance of categories across multiple classes, ensuring data suitability for training.

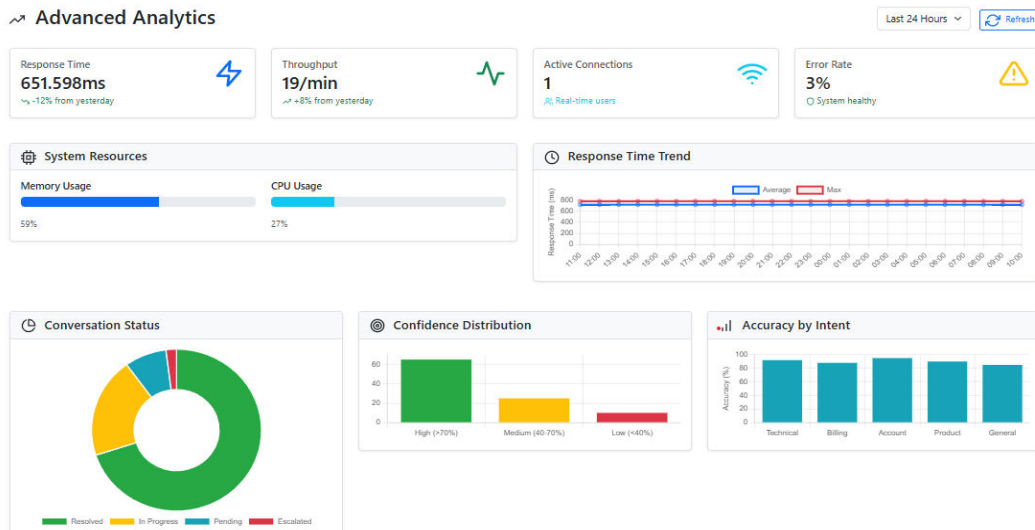


Fig. 5: Advanced Analytics Dashboard for AI Support System Performance Monitoring.

Exploratory Data Analysis (EDA)

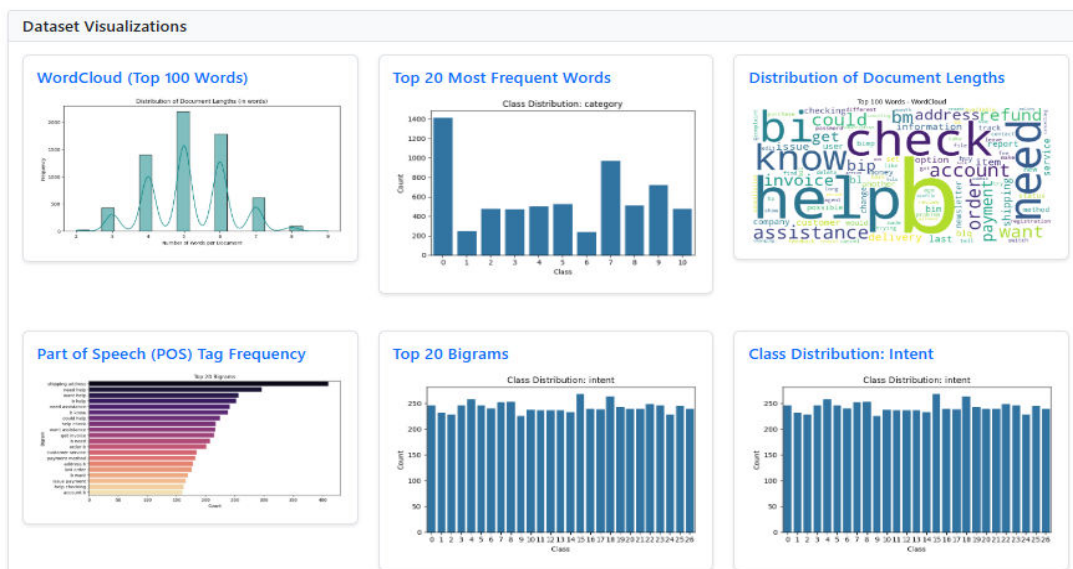


Fig. 6: Exploratory Data Analysis of Customer Support Intent Dataset.

Fig. 7 illustrates the performance evaluation of the proposed MiniLM-WE DNN KNN model for intent classification using both confusion matrix and ROC curve analysis. In Fig. 9.5(a), the confusion matrix presents class-wise prediction results, where the majority of values are concentrated along the diagonal, indicating high true positive rates across multiple intent classes. The numerical values, predominantly around 52–54 samples per class, demonstrate consistent classification accuracy with minimal misclassification. In Fig. 9.5(b), the ROC curves for each intent class show Area Under Curve (AUC) values close to 1.00, indicating near-perfect separability between classes. The curves closely follow the top-left boundary, confirming high sensitivity and specificity.

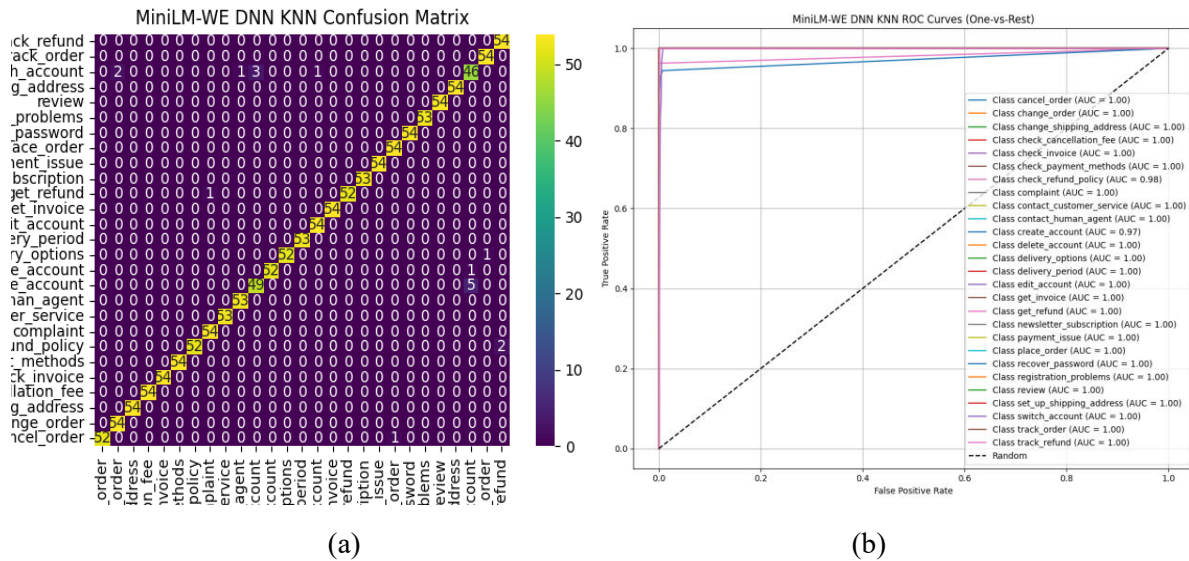


Fig. 7: Proposed Mini LM-WE KNN model (a) Confusion matrix and (b) ROC Curve for Intent Classification column.

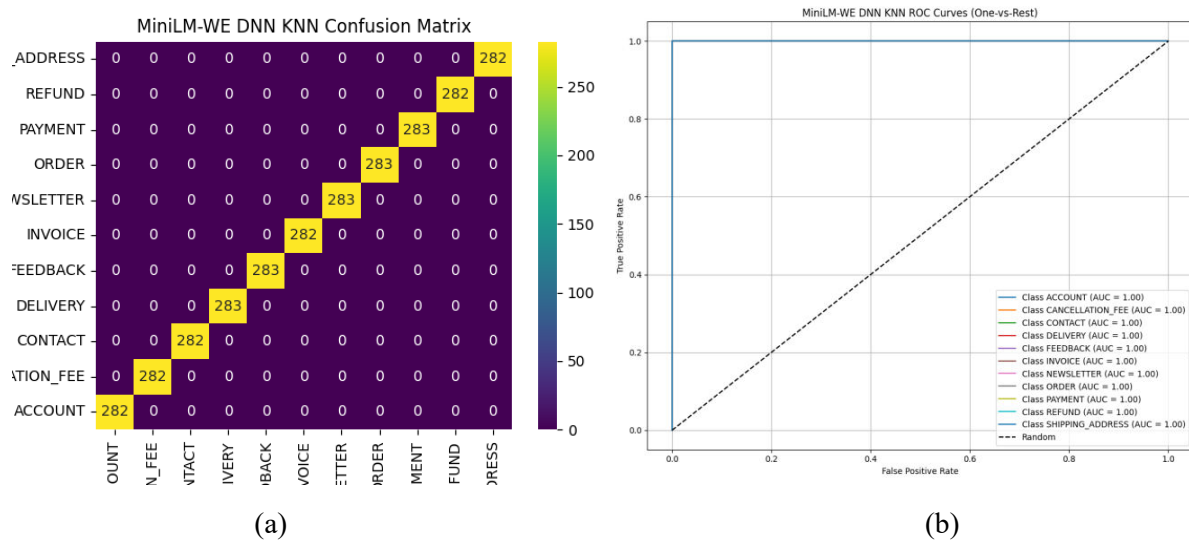


Fig. 8: Proposed Mini LM-WE KNN model (a) Confusion matrix and (b) ROC Curve for Category Classification.

Fig. 8 illustrates the evaluation of the proposed MiniLM-WE DNN KNN model for category classification through confusion matrix and ROC curve visualization. In Fig. 9.6(a), the confusion matrix shows strong diagonal dominance with numerical values approximately ranging from 282 to 283 samples per category, indicating highly accurate predictions with negligible misclassification. The off-diagonal elements are nearly zero, reflecting excellent class separation and model precision. In Fig. 9.6(b), the ROC curves demonstrate AUC values equal to 1.00 for all categories, signifying perfect classification capability. The curves align closely with the ideal classification boundary, confirming optimal true positive and false positive trade-offs.

5. CONCLUSION

The development of the Customer Support Dataset with Multi-Intent Annotations system demonstrates the effectiveness of combining natural language processing, machine learning, and web-based frameworks for intelligent conversational AI. By integrating multiple classifiers including DTC, NBC, KNN, and the proposed hybrid MiniLM-WE DNN-KNN model. The system can accurately detect and

respond to diverse customer intents. The Flask-based implementation ensures a seamless user experience, providing real-time predictions, visualization of model performance, and analytics dashboards that track conversation metrics such as response time, intent distribution, and user satisfaction. Through modular design and session management, the system maintains reliability, scalability, and ease of deployment, demonstrating a significant improvement over traditional single-intent classification system. The research validates the practicality of multi-intent recognition in enhancing customer support efficiency, reducing response times, and improving overall user satisfaction.

REFERENCES

- [1] Huang Z, Liu H, Zhu J, Min J. Customer Sentiment Recognition in Conversation Based on Contextual Semantic and Affective Interaction Information. *Applied Sciences*. 2023; 13(13):7807. <https://doi.org/10.3390/app13137807>
- [2] Wagner N, Kraus M, Minker W, Griol D, Callejas Z. A Survey on Multi-User Conversational Interfaces. *Applied Sciences*. 2025; 15(13):7267. <https://doi.org/10.3390/app15137267>
- [3] Lin C-C, Huang AYQ, Yang SJH. A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022). *Sustainability*. 2023; 15(5):4012. <https://doi.org/10.3390/su15054012>
- [4] Bercaru G, Truică C-O, Chiru C-G, Rebedea T. Improving Intent Classification Using Unlabeled Data from Large Corpora. *Mathematics*. 2023; 11(3):769. <https://doi.org/10.3390/math11030769>
- [5] Chuang H-M, Cheng D-W. Conversational AI over Military Scenarios Using Intent Detection and Response Generation. *Applied Sciences*. 2022; 12(5):2494. <https://doi.org/10.3390/app12052494>
- [6] Nicolescu L, Tudorache MT. Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review. *Electronics*. 2022; 11(10):1579. <https://doi.org/10.3390/electronics11101579>
- [7] Merdivan E, Singh D, Hanke S, Kropf J, Holzinger A, Geist M. Human Annotated Dialogues Dataset for Natural Conversational Agents. *Applied Sciences*. 2020; 10(3):762. <https://doi.org/10.3390/app10030762>
- [8] Izadi S, Forouzanfar M. Error Correction and Adaptation in Conversational AI: A Review of Techniques and Applications in Chatbots. *AI*. 2024; 5(2):803-841. <https://doi.org/10.3390/ai5020041>
- [9] Al-Mutawa RF, Al-Aama AY. Arabic Opinion Classification of Customer Service Conversations Using Data Augmentation and Artificial Intelligence. *Big Data and Cognitive Computing*. 2024; 8(12):196. <https://doi.org/10.3390/bdcc8120196>
- [10] Sultana T, Mandal AK, Saha H, Sultan MN, Hossain MD. Intent Identification by Semantically Analyzing the Search Query. *Modelling*. 2024; 5(1):292-314. <https://doi.org/10.3390/modelling5010016>
- [11] Coppola R, Ardito L. Quality Assessment Methods for Textual Conversational Interfaces: A Multivocal Literature Review. *Information*. 2021; 12(11):437. <https://doi.org/10.3390/info12110437>
- [12] Allouch M, Azaria A, Azoulay R. Conversational Agents: Goals, Technologies, Vision and Challenges. *Sensors*. 2021; 21(24):8448. <https://doi.org/10.3390/s21248448>

