Volume 12, No.10, October 2025

Journal of Global Research in Mathematical Archives

ISSN 2320 – 5822

UGC Approved Journal

RESEARCH PAPER

Available online at http://www.jgrma.info

CONVERSATIONAL BANKING CHATBOT USING NLP AND SUPERVISED MACHINE LEARNING MODEL

Himanshu Barhaiya[1]

[1] Department of Computer Science and Engineering, Lakshmi Narain College of Technology, Bhopal

himanshub@lnct.ac.in

Abstract: Chatbots are intelligent systems designed by Artificial Intelligence (AI) and are upgraded with Natural Language Processing (NLP) algorithms. In an impressive way, it engages users and interacts with them, answering their questions. Conversation facilitators are mostly used by companies, government departments, and non-profit organisations. Money-related industries such as banks, credit card companies, financial institutions, e-commerce stores, and startups are typical places where we find these chatbots implemented. This research paper depicts the implementation and assessment of a Banking Conversational Chatbot powered by Deep Learning (DL) techniques. The bank chatbot dataset, consisting of real user communication, was preprocessed by cleaning, tokenisation, normalisation, and data balancing using SMOTE to ensure the training data was of the highest quality. The authors proposed a Gated Recurrent Unit (GRU) network to capture the sequential dependencies and contextual patterns of the user query, providing a more efficient and compact solution than the traditional LSTM model. In the conducted comparative experiments with different models, namely SVM, XGBoost, and Naive Bayes, the accuracy recorded was 68%, 79%, and 91%, respectively, while the argued GRU model results showed superiority over the other models with its accuracy of 97%, precision of 97.9%, recall of 96%, and an F1-score of 97%. These figures demonstrate the GRU model's strength and effectiveness in identifying user intent; thus, it can be a significant boost to the performance and reliability of conversational banking applications.

Keywords: Conversational AI, Banking Chatbot, Natural Language Processing (NLP), Bank chatbot dataset, Customer Service, Machine learning, GRU.

1 INTRODUCTION

The financial sector is going through brand new changes due to digital technologies which include machines that automate tasks, technologies with artificial intelligence, and making decisions based on data analysis. Financial systems of the past are struggling to keep up with the demand of customer service driven by real-time, personalized, and efficient banking which requires responsiveness and scalability[1][2][3]. Meanwhile, financial institutions intensify the application of intelligent automation technologies to enhance their service provision as well as their efficiency[4] at the same time. Due to the existence of the conversational interface, the banking industry can now more readily utilize chatbots, which effectively allows the automated financial services to be bridged with human communication.

Chatbots are recognized as a revolution in the banking industry among the significant technological ones[5][6]. These dialogue agents emulate human dialogue —either in text or voice —to assist users with various activities, such as balance inquiries, account inquiries, and loan applications. By doing so, chatbots are beginning to combine automation with human intervention to reduce the number of human labor, increase the accessibility of the services and 24/7 customer support[7][8]. Their application in financial sites is an evidence of how conversational technology can be an alternative to the traditional menu based systems which is stagnant as it offers an interactive and smart conversation[9][10]. However, as the financial sector is correlated with complex and sensitive interfaces, the development of general-purpose chatbots into more special-purpose systems capable of processing financial lexicon and comprehending the contextual background is a precondition of the development of banking chatbots.

In a particular financial environment, a banking chatbot elevates the idea of conversational automation to a new level [11][12]. Users may perform tasks like fund transfers, transaction tracking, and complaint registration since it is both a secure integration partner of the financial databases and a straightforward facilitator of user engagement. Banking chatbots must understand complicated consumer intentions about financial goods and adhere to regulatory standards, unlike generic chatbots.[13]. Hence, the performance of these systems largely depends on their ability to understand human language and provide proper, empathetic, and relevant responses. The increasing demand for contextual accuracy and adaptive learning.

The combination of Natural Language Processing (NLP) and Supervised Machine Learning (ML) models is an essential part of current banking chatbot systems[14][15][16]. NLP is a tool for the chatbot to understand language, recognize user intent, and create a reply[17], where the system learns and makes predictions using annotated conversational datasets with the use of supervised ML techniques like Support Vector Machines (SVM), Random Forests, and Gated Recurrent Units (GRU)[18]. The harmony of NLP and ML transforms standard banking chatbots into smart virtual assistants that can make decisions and interact with users in a personalised way in real time.

1.1 Motivation and Contribution

The increasing demand for efficient, accurate, and user-friendly conversational agents in the banking sector is the main reason for this study. As the number of customer interactions increases and the need for instant query resolution becomes more urgent, traditional rule-based chatbots often struggle to understand complex user intents, resulting in poor user experiences and operational inefficiencies. By using advanced ML and DL techniques, such as the GRU model, chatbots can provide prompt, context-aware, and accurate responses. The present study is motivated by the creation of a banking chatbot system that, in addition to elevating customer satisfaction, would reduce human agents' workload and thus be instrumental in advancing intelligent, responsive banking services. This research makes several key contributions, which are enumerated below:

- Used the Bank Chatbot Dataset with real-world interactions. This ensured practical relevance and genuine evaluation.
- Implemented thorough data preprocessing, including cleaning, tokenisation, normalisation, and data balancing, to improve model performance.
- Proposed a Gated Recurrent Unit (GRU) model that effectively captures the sequence of dependencies and context in user queries.
- Several assessment measures were employed to assess the model's functionality. These included F1-score, recall, accuracy, and precision. These measurements provide a comprehensive assessment of its efficacy.
- Provided a clear method for evaluating chatbot performance with standard metrics. This ensures that results can be repeated and trusted.

1.2 Justification and Novelty

The study is needed to address issues related to rule-based and traditional ML chatbots which are hard to properly comprehend the subtle intentions of the user and keep the context of the conversation in chronological order. The original part of this study is that a Gated Recurrent Unit (GRU) DL model is applied to a banking chatbot, which performs user queries much more effectively; on the one hand, it requires less training time than an LSTM network. In addition, this study employs intensive preprocessing, normalisation, and data balancing to develop a stronger model that outperforms benchmark models such as SVM, XGBoost, and Naive Bayes. The original aspect of this piece of work is the combination of methodological rigor and the application of GRU in the intent recognition in natural banking conversations.

1.3 Organization of the Paper

The paper is structured as follows: Section 2 reviews the related work on Conversational Banking Chatbots using ML models, Section 3 describes the dataset, preprocessing steps, and model implementation, Section 4 presents the experimental results along with a comparative analysis, and Section 5 summarizes the main conclusions and suggests future research directions to wrap up the study.

2 LITERATURE REVIEW

A detailed review of key research studies on conversational banking chatbots that use ML models was done to support the development of this study.

Shrivastava *et al.* (2025) discuss the key NLP methods and tools applied to such systems: ML algorithms, neural networks, and semantic analysis. Consequently, it would also mean that the input of chatbots and virtual assistants is being enhanced by big data and one-on-one learning. Understanding of language, its generation and management of the context are defined in detail, demonstrating how NLP helps to improve the user's journey by enhancing the friendliness and humanness of the interactions. It successfully implemented a multilingual system supporting 10 languages, with a mean intention recognition rate of 89%. Translation Accuracy Machine translation components reached a BLEU score of 85.3, thus delivering translations of a high standard[19].

Sriharsha and Prakash (2024), through integration, provide personalised, efficient customer interactions at any time of day, thereby significantly impacting both customer satisfaction and the bank's operational efficiency. Indeed, this chatbot is going beyond the norms in digital banking, thus reflecting a pledge to innovative, technologically advanced, and customer-friendly solutions. The paper records impressive real-world outcomes (96% accuracy) but chooses to depict them mostly through figures/visuals rather than through detailed numeric tables in the text[20].

Lajčinová, Valábek and Spišiak, (2024) introduced Gemma models (2B-7B parameters) based on Gemini technology, highlighting the emphasis on lightweight open-source alternatives that have strong reasoning capabilities. The paper's experiments demonstrate that the fine-tuned SlovakBERT achieves the highest in-scope accuracy of 77.2% and out-of-scope false positive rate (FPR) of 6.3%, thus it is able to outperform baselines (67.6% accuracy, 22.5% FPR), banking-tailored BERT (68.5%, 4.0%), 8B instruct (75.1%, 7.0%), and even gpt-3.5-turbo fine-tuned (79.5%, 4.3%) to a great extent and hence, SlovakBERT is set as the benchmark for Slovak banking chatbots[21].

Amama (2023) concentrates on developing a banking chatbot system that makes use of ML and a set of NLP tools. This entails compiling and getting ready a dataset of user questions and answers. After evaluating the chatbot's performance, the results show that a working chatbot was successfully developed. Its high F1 score of 0.97 indicates that it understood user inquiries 97% of the time and gave pertinent answers. An existing system demonstrates notable enhancements made possible by the integration of ML and natural language processing technologies. The chatbot's user-friendly and straightforward interface may be accessed via a web application, and administrators can efficiently maintain the knowledge base with separate back-end access[22].

Ananda, Wiharja and Bijaksana (2023) depict a graph where words are nodes and their node characteristics are represented as vectors and edges. These nodes are connected by a dependency parser. The graph-based model used to process the transformed data. This research compares the performance and inference time of graph-based and conventional approaches for identifying a sentence's emotion. The experimental findings show that both the tree-based and graph-based models achieve an accuracy of 0.7173. In contrast, when comparing inference times, the graph-based models are three times faster than the tree-based ones[23].

Abdulkader and Muhammad (2022) is selected as a comparable, popular, no-cost tool for testing its ability to generate structured text from unstructured Arabic. The suggested model trains and validates using the standard Arabic dialect. The knowledge base is a virtual customer support division of Mosul's Al-Rasheed Bank. Wit.ai demonstrated great accuracy and F1 score in the intent classification and entity extraction phases, with respective values of 0.96 and 0.948[24].

Research gaps: Recent studies of banking chatbots through NLP and supervised ML, critical challenges persist in achieving robust, real-world applicability. Existing research often remains limited to single-language or domain-specific datasets, with insufficient exploration of multilingual and low-resource environments. Although advanced models such as GRU, SlovakBERT, and Gemma variants demonstrate strong predictive accuracy, issues like contextual understanding, scalability, and secure data handling continue to hinder practical deployment. Furthermore, limited attention has been given to user experience, privacy, and multimodal interaction. These gaps emphasize the need for more adaptive, ethical, and user-centric chatbot frameworks that balance performance, interpretability, and security in modern conversational banking systems.

Table 1: Recent Studies on Conversational Banking Chatbot Using Machine Learning Model

Author(s) & Year	Methodology	Key Findings	Techniques Used	Limitations	Future Work
Shrivastava et al. (2025)	Analyzed NLP-based chatbot architectures integrating ML algorithms, neural networks, and semantic analysis for multilingual systems.	Achieved 89% intent recognition accuracy across 10 languages; BLEU score of 85.3 for translation quality.	NLP, Neural Networks, Semantic Analysis, Big Data Learning.	Limited exploration of cross-domain adaptability; lacks evaluation on low-resource languages.	Expand to cross-domain and low-resource environments; enhance contextual learning.
Sriharsha & Prakash (2024)	Designed and deployed an AI-driven banking chatbot providing real-time customer interaction.	Delivered 96% accuracy in user intent recognition; demonstrated high satisfaction and efficiency gains.	Supervised ML, NLP Integration, Dialog Management.	Performance metrics primarily presented as visuals; lacks numerical analysis tables.	Provide detailed quantitative evaluation; extend model for multimodal input support.
Lajčinová, Valábek & Spišiak (2024)	Developed lightweight Gemma and SlovakBERT models for Slovak banking chatbot systems.	Fine-tuned SlovakBERT achieved 77.2% in-scope accuracy and 6.3% FPR,	BERT, SlovakBERT, Gemini-based Gemma Models.	Focused only on Slovak language; limited multilingual evaluation.	Generalize approach to multilingual and cross-domain banking datasets.
Amama (2023)	Implemented an NLP- and ML-based smart chatbot using the Natural Language Toolkit (NLTK).	Achieved 0.97 F1 score, indicating 97% accuracy in understanding and response generation; improved usability through web interface.	NLP Toolkit (NLTK), Supervised ML Classification.	Dataset size and domain coverage limited; lacks real-time scalability.	Expand dataset diversity and optimize for real-time financial applications.
Ananda, Wiharja &	Proposed a graph-based NLP model where words are	Graph-based models achieved comparable	Graph Neural Networks (GNN),	Focused only on sentiment detection; not	Extend graph-based modeling to banking chatbots for faster

Bijaksana (2023)	represented as nodes connected by dependency parsing.	accuracy (0.7173) but 3× faster inference than tree-based methods.	Dependency Parsing.	applied to banking context.	inference and context retention.
Abdulkader & Muhammad (2022)	Developed Arabic-language chatbot using entity extraction in a virtual bank service.	Attained high precision (0.96) and F1-score (0.948) on Arabic banking dataset.	Wit.ai NLP Framework, Intent Classification, Entity Extraction.	Restricted to Arabic language; lacks multilingual or complex intent handling.	Extend model to handle multiple dialects and integrate contextual emotion detection.

3 RESEARCH METHODOLOGY

The methodology of this work, a conversational banking chatbot based on DL is developed using the Bank Chatbot Dataset. Initially, data gathering and preprocessing were performed, including concatenation, cleaning, handling missing values, removing duplicates and noisy data, correcting spelling errors, tokenisation, and normalisation using z-score to ensure standardised and high-quality input. After applying SMOTE to balance the class distribution, the dataset was split into training (80%) and test (20%) sets to preserve the class distribution. By merging the input and forget gates into a single update gate, the proposed model's Gated Recurrent Unit (GRU) design efficiently simplifies the internal architecture of LSTM networks. The inclusion of a reset gate allows for the recording of both immediate and distant dependencies in the sequential data. By passing input sequences through these gates, the model is trained, memory states are updated, and expected outputs are produced. Lastly, a confusion matrix was used to compute the critical metrics—recall, accuracy, precision, and F1-score —which together provide a thorough evaluation of the model's classification capabilities. Figure 1 illustrates the proposed flowchart for the development of a Conversational Banking Chatbot using a ML model.

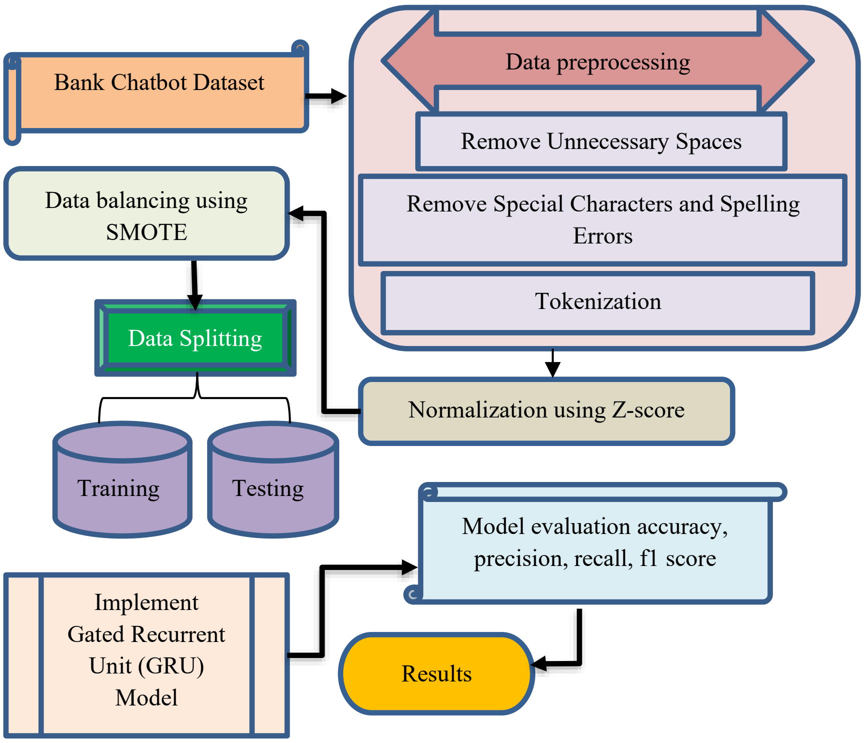

Figure 1: Proposed flowchart for Conversational Banking Chatbot using Machine Learning Model

A thorough explanation of each stage in the suggested technique is given in the section that follows:

3.1 Data Gathering and Analysis

This research is based on the Bank Chatbot Dataset. To authentically depict real user interactions, have employed a test dataset consisting of around 300 (text, intent) pairs randomly selected from the real chatbot deployment. Human annotators have tagged each sentence in this dataset with an intention.

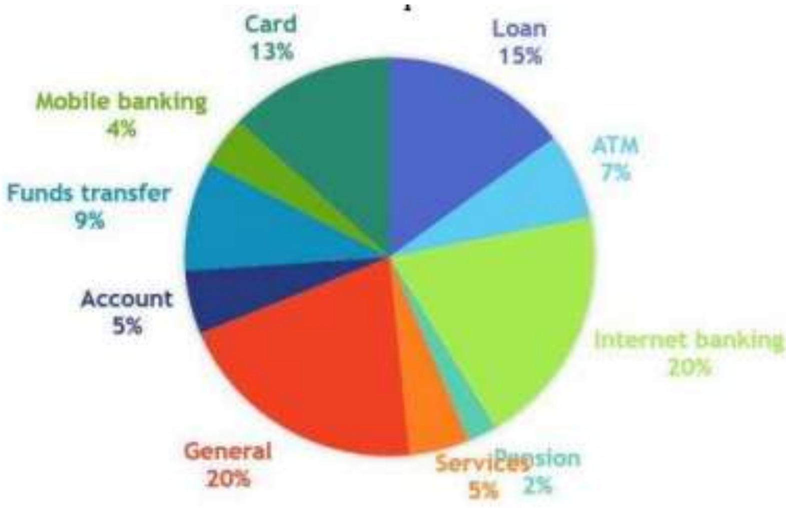

Figure 2: Distribution of User Queries in Banking Chatbot System

The pie chart illustrates the proportions of user intents in a banking chatbot dataset as in figure 2. Based on the figures, the two major classes are Internet banking (20%) and General queries (20%), which are the ones with the highest number of occurrences. These two categories are then followed by Loan inquiries (15%) and Card-related requests (13%) respectively. Mid-range categories include Funds transfer (9%) and ATM services (7%), while smaller segments comprise Account management (5%), Service requests (5%), Mobile banking (4%), and Pension-related queries (2%). This distribution reflects typical customer interaction patterns with banking services, providing insights for chatbot training and optimization.

3.2 Usage of chatbot

Chatbots start with the design, much to the process used to create web pages and mobile apps. The interaction between the user and the bot is described in this design [25]. The pattern also covers the development of a bot that incorporates input analysis through the use of a NLP engine. Following the early phases, the bots are studied and maintained.

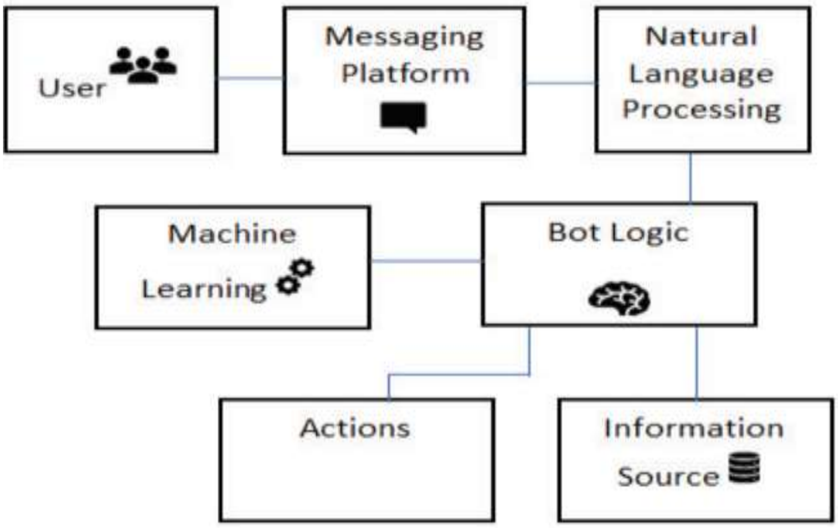

Figure 3: Users interaction with chatbot

Figure 3 displays platforms that Platform-as-a-Service vendors provide in which chatbot development can take place. These are SnatchBot, Oracle cloud platform and IBM Watson. Perspectives that consumers are using messaging applications more frequently than social networking seem evident due to the recent surveys. In many ways, conversational bots can help businesses in various industries to automate and simplify their processes, enhance productivity, and enhance employee and customer interactions. Chatbots are computer software, the AI of which resembles human speech. The design is meant to be a full-fledged virtual assistant and entertainment device.

3.3 Data Pre-processing

To prepare data, the Bank Chatbot Dataset was used, and it involved concatenation, cleaning, and feature engineering. The preprocessing procedure included addressing missing data, eliminating duplicates and erroneous data, erasing redundant information, and normalizing and tokenizing data to ensure high-quality data to feed on the models:

- **Remove Unnecessary Spaces:** Removing unnecessary spaces is the process of cleaning up text by eliminating extra whitespace, such as multiple spaces between words, leading/trailing spaces, or excessive line breaks. This is a routine task in data processing and text editing. It ensures that the text is well-formatted and polished.
- **Remove Special Characters and Spelling Errors:** The process of removing and standardizing text data includes the elimination of special characters and spelling errors. It is a step that requires data preprocessing that will be used. Such applications as data analysis and NLP demand the data to be of a high-quality, consistent, and accurate.
- **Tokenization:** The process of replacing sensitive data by non-sensitive information, called a token, which is one single code that is analyzed and protected, is referred to as tokenization. The main stage of the data cleaning process of ML and NLP is tokenization, which specifies how to divide the text into smaller units, known as tokens.

3.4 Normalization using Z-score

The practice of standardizing or scaling data to make sure it has a uniform distribution is known as data normalization. The two most often used techniques among the others are min-max normalizing and z-score normalization. In this investigation, z-score normalization was used, which ensures that values are centered around the average with one standard deviation, calculated by transforming the data such that the mean is 0 and the standard deviation is 1. The z-score normalization is mathematically represented in Equation (1).

$$E' = \frac{E-\bar{M}}{\sigma_M} \quad (1)$$

Where,

For every data input, $\bar{M}$ represents the mean, σ_M is the standard deviation, and $E'$and E are new and old.

3.5 Data balancing using SMOTE

In machine learning, data balancing refers to techniques used to correct a skewed class distribution in a dataset, which is a common problem known as "imbalanced data". Figure 4 SMOTE bar chart shows how to handle class imbalance in a banking chatbot dataset. The graph compares sample distributions across five banking intent categories before and after applying SMOTE.

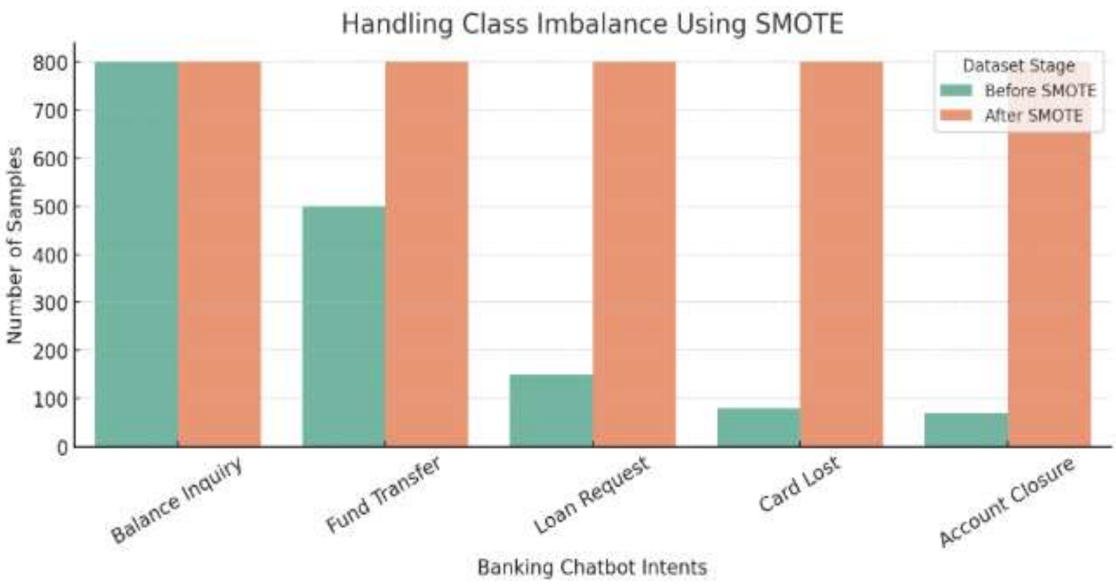

Figure 4: Handling class imbalance using SMOTE

Initially, "Balance Inquiry" had around 800 samples while minority classes like "Loan Request," "Card Lost," and "Account Closure" had only 150-80 samples. After SMOTE application, all classes were balanced to approximately 800 samples each, ensuring the chatbot model receives equal training representation across different banking intents for improved performance.

3.6 Data Splitting

It is important to note that the dataset was partitioned according to feature mapping. Using stratification and randomization, to preserve the original class distribution, the data were separated into training (80%) and testing (20%) groups.

3.7 Proposed Gated Recurrent Unit (GRU) Model

A suggested GRU Model deep learning-based conversational banking chatbot is presented in this paper. An LSTM neural network's more intricate internal structure and difficult parameter tweaking lead to longer model training times. GRU is an LSTM in a simple form. The GRU model offers comparable prediction accuracy to the LSTM model while requiring less training time. There are just two gating components in the memory module: GRU combines the LSTM's input and forget gates into a single update gate, as well as the reset gate. The update gate controls the amount of data that is transferred from the previous data state to the current data state. It is denoted by Z_t. A higher value for the update gate shows that the present neuron remembers more past data than the prior neuron, whereas the latter remembers far less. The update gate's main roles are clearing out memory and detecting trends in the water-quality data series over the long period. As demonstrated in equation (2), the update gate is able to capture information:

$$Z_t = \sigma(W_Z * [h_{t-1}, X_t]) \tag{2}$$

The reset gate or R_t is very important in deciding how much of the past history should be kept. A low reset gate indicates that more historical data is preserved, which is ideal for identifying short-term patterns in the parametric data on water quality. To get the reset gate value from information, the formula is given in eqn (3):

$$R_t = \sigma(W_r * [h_{t-1}, X_t]) \tag{3}$$

where h_t represents the output state of the unit at time t and $\bar{h}_t$ is the inferred state of h_t . In this condition, the data from the current unit is saved and passed on to the next one, while the anticipated value for the output from the prior time is determined using equation (4):

$$\bar{h}_t = tanh(W_{\bar{h}} * [r_t * h_{t-1}, X_t]) \tag{4}$$

It is possible to represent the anticipated results of the data from the water quality parameters using equation (5):

$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \bar{h}_t \tag{5}$$

where W_Z, W_r, and $W_{\bar{h}}$ indicate the cell's weight matrix, "[]" denotes the link between two matrices, "$*$" denotes the matrix product, σ represents the activation function, tanh represents the activation function's bisecting curve, and X_t represents the input data value at the current time, t_t. Additionally, the memory cell's output value for the water quality parameter data at time $t-1$ is h_{t-1}.

3.8 Evaluation metrics

The effectiveness of the suggested model was measured through various essential metrics. To illustrate classification results, a confusion matrix was created, showing the count of correct and incorrect predictions for each class. The figures in the matrix indicated four different types of results: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The main performance assessment parameters that were utilized to ascertain accuracy, precision, recall, and F1-score are detailed below:

Accuracy: The proportion of instances (input samples) correctly anticipated in the dataset. It is given as (6)-

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \tag{6}$$

Precision: The proportion of accurately predicted positive instances among all of a model's positive predictions is known as precision. It identifies the accuracy with which the classifier selects positive cases, and the formula for this is shown in Equation (7)-

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

Recall: This metric shows the ratio of correctly predicted positive events to all positive events that actually occurred in the dataset. It is mathematically expressed in Equation (8)-

$$Recall = \frac{\text{TP}}{TP+FN} \tag{8}$$

F1 score: It stands for the precision and recall harmonic mean. This provides a fair assessment of both. The value ranges from 0 to 1, and its mathematical formulation is shown in Equation (9)-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

4 RESULTS AND DISCUSSION

This section describes the experimental setup and provides performance analysis of the proposed model at both the training and testing stages, demonstrating its efficiency and computational efficiency. The experiments were designed and analyzed using Python in Jupyter Notebook, with help from NumPy, Pandas, and Scikit-learn. Table 2 shows the main performance measures used to develop the proposed model. It was trained on the Bank Chatbot Dataset and its results are displayed there. Results of the classification of the proposed Gated Recurrent Unit (GRU) model applied to the conversational banking chatbot with the Bank Chatbot dataset. This model achieves 97% accuracy, indicating strong generalisation in predicting user intents. It was also shown to have a precision of 97.9%, meaning it can produce relevant answers with few false positives. The model catches the majority of cases of interest, as indicated by the 96% recall, while the 97% F1-score shows that precision and recall are well-balanced. These findings verify that GRU model is most effective in interpreting and replying to the user queries in the banking chatbot scenario.

Table 2: Classification results of the proposed model, FOR Conversational Banking Chatbot using Bank Chatbot Dataset

Matrix	Gated Recurrent Unit (GRU) Model
Accuracy	97
Precision	97.9
Recall	96
F1-score	97

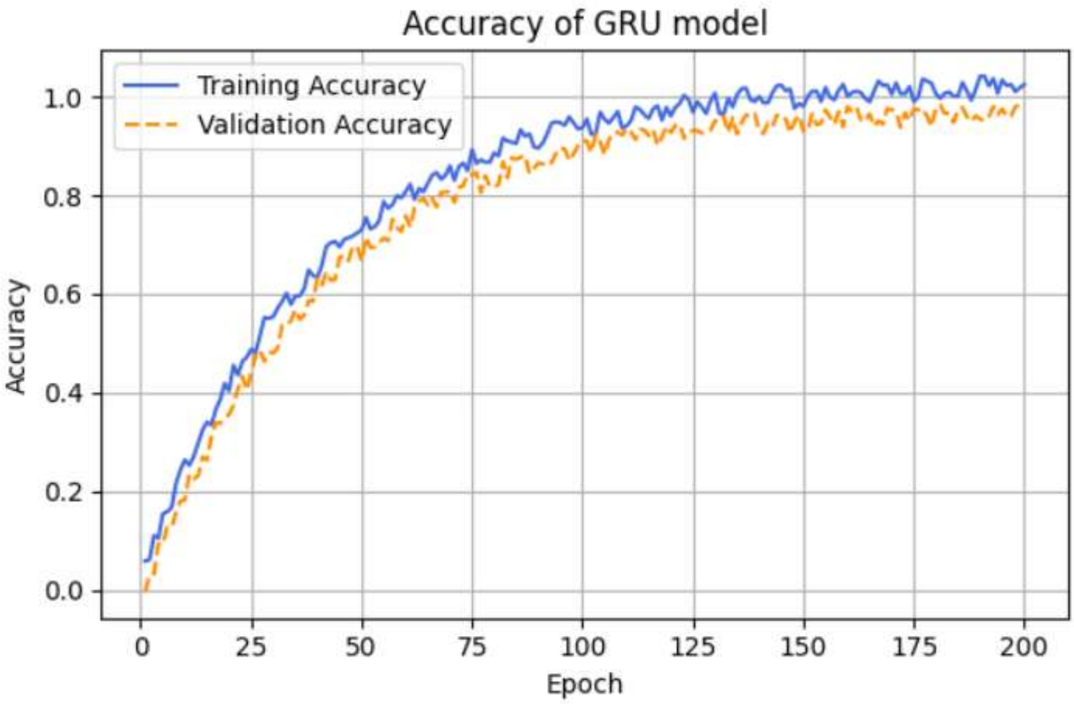

Figure 5: Accuracy curve for the GRU Model

Figure 5 shows the accuracy development of a GRU (Gated Recurrent Unit) model across 200 training epochs. Training and validation accuracy curves increase rapidly in the first 50 epochs, starting at almost zero and reaching around 80%. The curves then level off slowly, and there is approximately 100 per cent accuracy, with a small gap between the curves, indicating good generalisation and reasonable overfitting.

Figure 6. The graphs on the loss of a GRU model in 200 epochs of training. Training (blue) and validation (orange) loss are both high at the beginning, at around 4.0, and sharply diminishing in the initial 50 epochs. They keep reducing gradually approaching 0.2-0.4 at epoch 200 and exhibit effective learning with little overfitting since the curves approach each other in their behavior.

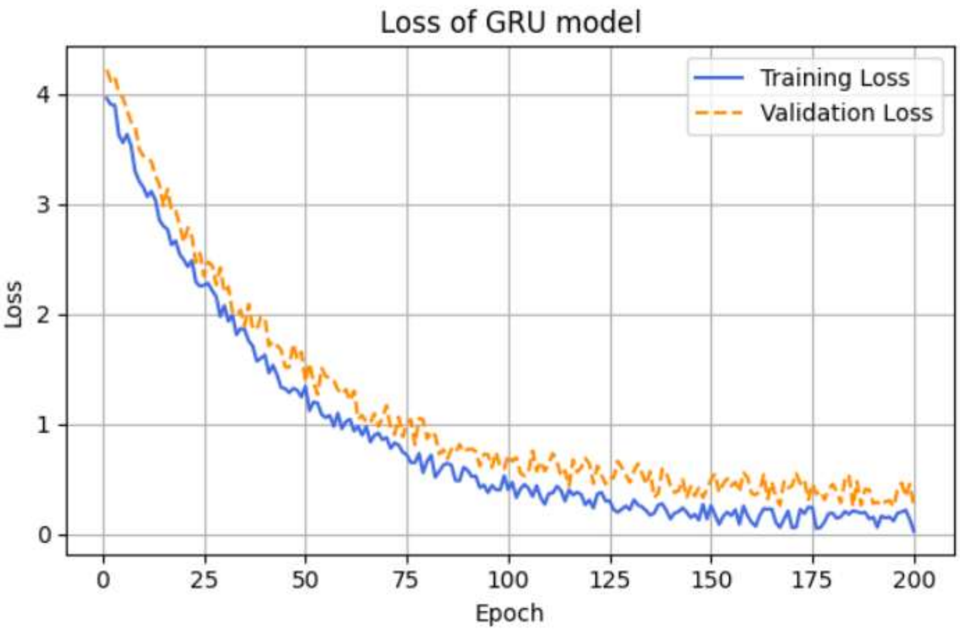

Figure 6: Loss curve for the GRU Model

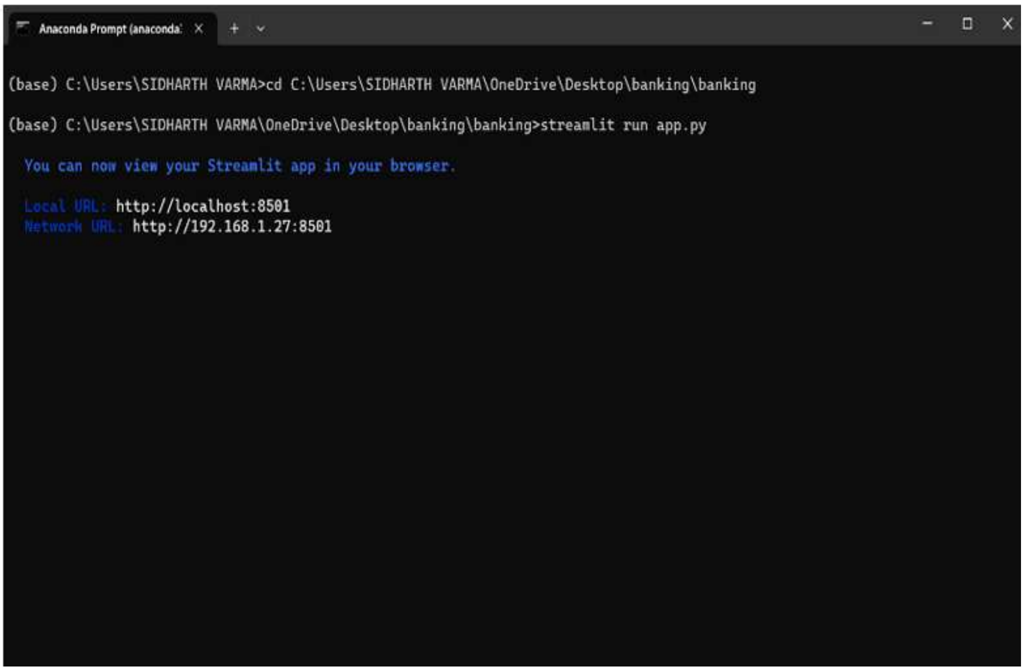

Figure 7: Command prompt implementation for application

This screenshot demonstrates an Anaconda Prompt terminal in which a chatbot banking application is being started with Streamlit in Figure 7. The executed command is a Python file (app.py) in a directory of a bank project, which opens the web interface of the chatbot. The application is made available as soon as it is initiated, and the users have access to it using local and network address, and can use their web browser to interact with the banking chatbot, asking it to answer different banking-related queries and perform different transactions.

4.1 Comparative analysis

The efficacy of the suggested GRU model was tested with the comparison of its accuracy to that of other existing models. Table 3 provides a comparative analysis of different ML models used on the banking chatbot dataset. Accuracy, precision, recall, and F1-score are these assessment metrics. Although the F1-rate was not shown, the SVM model achieved 95% in the accuracy, precision, and recall measures. XGBoost gave a high degree of sensitivity with 79% accuracy, 70.2% precision, a recall of 99% and an F1-score of 82.1%, which suggests a high level of sensitivity and weak precision. Naive Bayes (NB) model demonstrated better results at 91% accuracy, 91% precision, 88% recall and 89% F1-score, which is a balance in predictive performance. Impressively, the suggested GRU model outperformed all others with a 97% accuracy rate, 97.9% precision, 96% recall, and 97% F1-score. These findings confirm the GRU model's greater effectiveness in modelling sequential dependencies in conversational data and the stability and accuracy of responses in the banking chatbot project.

Table 3: Comparison of Different Machine learning AND DEEP LEARNING Models for Conversational Banking Chatbot

Model	Accuracy	Precision	Recall	F1-score
SVM[26]	95	95	95	-
XGBoost[27]	79	70.2	99	82.1
NB[28]	91	91	88	89
GRU	97	97.9	96	97

The suggested GRU model has several strengths, the first being its high predictive accuracy, reaching 97, which is much higher than that of ML models such as SVM, XGBoost, and Naive Bayes. The GRU architecture is simpler than LSTM networks, since instead of having separate input and forget gates, the update gate is used, resulting in shorter training time and lower computational complexity. Moreover, the GRU is very efficient at retaining both long and short sequences in sequential data; hence, it is especially appropriate for intent recognition in conversational banking chatbots. This is its effectiveness and resilience, the model is able to provide valid, context-sensitive responses which result in high levels of user satisfaction and system reliability.

5 CONCLUSION AND FUTURE STUDY

Banking chatbots are mostly applied to enhance the customer experience. They are beneficial to personnel, though, and help prevent awkward situations that may arise from direct client work. Automation of tedious, repetitive tasks is the most common application of chatbots in bank customer service. According to the experiment's outcomes, the comparative analysis of various models of the Conversational Banking Chatbot reveals that the proposed GRU model is far more successful than the others, achieving an accuracy of 97%. Although the traditional ML models, including SVM, XGBoost, and Naive Bayes, reported 68%, 79%, and 91% accuracy, respectively, it is clear that the GRU model performed better than the traditional models, which are unable to capture the sequential dependencies and contextual patterns during the interaction between a user and an agent. This shows that deep learning-based architectures such as GRUs are well-suited for intent recognition in banking chatbots and provide more reliable, accurate responses than traditional approaches. Further development of multilingual, real-time, scalable, and flexible banking chatbots should be pursued in the future. The combination of graph-based models and transformer models can increase the speed and comprehension of inferences and the multimodal and language expansion on low-resource will improve accessibility. Also, data privacy and security, as well as ethical issues, must be discussed as key factors for reliable, user-centric conversational banking systems.

REFERENCES

[1] E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications,” *Mach. Learn. with Appl.*, 2020, doi: 10.1016/j.mlwa.2020.100006.

[2] A. Przegalinska, L. Ciechanowski, A. Stroz, P. Gloor, and G. Mazurek, “In bot we trust: A new methodology of chatbot performance measures,” *Bus. Horiz.*, 2019, doi: 10.1016/j.bushor.2019.08.005.

[3] X. Sánchez-Díaz, G. Ayala-Bastidas, P. Fonseca-Ortiz, and L. Garrido, “A knowledge-based methodology for building a conversational chatbot as an intelligent tutor,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. doi: 10.1007/978-3-030-04497-8_14.

[4] Hemish Prakashchandra Kapadia, “Generative AI for Real Time Conversational Agents,” *Int. J. Curr. Sci.*, vol. 13, no. 3, pp. 201–208, 2023.

[5] R. Chanda and S. Prabhu, “Secured Framework for Banking Chatbots using AI, ML and NLP,” in *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 2023. doi: 10.1109/ICICCS56967.2023.10142289.

[6] L. Steinhoff, D. Arli, S. Weaven, and I. V. Kozlenkova, “Online relationship marketing,” *J. Acad. Mark. Sci.*, 2019, doi: 10.1007/s11747-018-0621-6.

[7] N. Malali, “AI Ethics in Financial Services: A Global Perspective,” *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 2, pp. 2456–2165, 2025, doi: /10.5281/zenodo.14881349.

[8] Nihar Malali, “THE ROLE OF DEVSECOPS IN FINANCIAL AI MODELS : INTEGRATING SECURITY AT EVERY STAGE OF AI / ML MODEL DEVELOPMENT IN BANKING AND INSURANCE,” *Int. J. Eng. Technol. Res. Manag.*, vol. 6, no. 11, pp. 218–225, 2022.

[9] H. Kali, “Optimizing Credit Card Fraud Transactions Identification And Classification In Banking Industry Using Machine Learning Algorithms,” *Int. J. Recent Technol. Sci. Manag.*, vol. 9, no. 11, pp. 1–12, 2024.

[10] K. C. C. Hemish Prakashchandra Kapadia, “AI Chatbots for Financial Customer Service: Challenges & Solutions,” *J. Adv. Futur. Res.*, vol. 2, no. 2, p. 7, 2024.

[11] R. Q. Majumder, “Machine Learning for Predictive Analytics: Trends and Future Directions,” *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 04, pp. 3557–3564, 2025.

[12] D. Patel, “Enhancing Banking Security: A Blockchain and Machine Learning- Based Fraud Prevention Model,” *Int. J. Curr. Eng. Technol.*, vol. 13, no. 06, Dec. 2023, doi: 10.14741/ijcet/v.13.6.10.

[13] S. B. Shah, S. A. Pahune, V. Prajapati, and M. Menghnani, “Efficient NLP-Based Framework for Tweets Sentiment Analysis

via Advanced Machine Learning Model," in *2025 Global Conference in Emerging Technology (GINOTECH)*, IEEE, May 2025, pp. 1–6. doi: 10.1109/GINOTECH63460.2025.11076911.

[14] H. K. Lee, J. S. Lee, and M. Keil, "Using Perspective-Taking to De-escalate Launch Date Commitment for Products with Known Software Defects," *J. Manag. Inf. Syst.*, 2018, doi: 10.1080/07421222.2018.1523604.

[15] A. S. Lokman and J. M. Zain, "One-match and all-match categories for keywords matching in chatbot," *Am. J. Appl. Sci.*, 2010, doi: 10.3844/ajassp.2010.1406.1411.

[16] Y. Wu and Z. Huang, "Application of a case-based reasoning method in estimating the power grid project cost," in *2008 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008*, 2008. doi: 10.1109/WiCom.2008.1743.

[17] T. Shah, "Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing," *Int. J. Sci. Res. Arch.*, vol. 15, no. 03, pp. 618–627, 2025.

[18] S. Ashwini, N. R. Rajalakshmi, P. Victer Paul, and L. Jayakumar, "Dynamic NLP Enabled Chatbot for Rural Health Care in India," in *2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022*, 2022. doi: 10.1109/ICCSEA54677.2022.9936389.

[19] N. Shrivastava, P. Tewari, S. Sujatha, S. R. Bogireddy, N. Varshney, and V. Sharma, "Natural Language Processing for Conversational AI: Chatbots and Virtual Assistants," in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, 2025, pp. 1–6. doi: 10.1109/IATMSI64286.2025.10984818.

[20] T. Sriharsha and D. R. V. Prakash, "Banking Chatbot Using Nlp And Support Vector Machine," *Educ. Adm. Theory Pract.*, vol. 30, no. 5, pp. 1958–1964, 2024, doi: 10.53555/kuey.v30i5.3207.

[21] B. Lajčinová, P. Valábek, and M. Spišiak, "Intent Classification for Bank Chatbots through LLM Fine-Tuning," pp. 1–7, 2024.

[22] C. Amama, "Smart Chatbot System for Banking using Natural Language Processing Tools," 2023, doi: 10.1109/IATMSI64286.2025.10984818.

[23] N. R. Ananda, K. R. S. Wiharja, and M. A. Bijaksana, "Sentiment Analysis on Banking Chatbot using Graph-based Machine Learning Model," in *2023 International Conference on Data Science and Its Applications, ICoDSA 2023*, 2023. doi: 10.1109/ICoDSA58501.2023.10276448.

[24] Z. N. Abdulkader and Y. F. Muhammad, "A Comparison of Natural Language Understanding Platforms to Build Arabic Chatbots," in *ac*, 2022. doi: 10.1109/AEST55805.2022.10412985.

[25] S. F. Suhel, "Conversation to Automation in Banking Through Chatbot Using Artificial Machine Intelligence Language," pp. 611–618, 2020.

[26] C. S. Kulkarni, A. U. Bhavsar, S. R. Pingale, and P. S. S. Kumbhar, "BANK CHAT BOT – An Intelligent Assistant System Using NLP and Machine Learning," 2017.

[27] M. S. Mujeeb and U. Hassan, "an Intelligent Ml Chatbot for Banking Systems," *Dogo Rangsang Res. J. UGC Care Gr. I J.*, vol. 13, no. 5, pp. 2347–7180, 2023.

[28] B. Divija, M. P. Pavani, S. A. Reddy, and A. Kumari, "Banking Chatbot Using NLP and Machine Learning Algorithms," *Int. Res. J. Eng. Technol.*, vol. 10, no. 5, pp. 475–480, 2023.

