RESEARCH ARTICLE OPEN ACCESS

Developing an NLP -Powered Banking Chatbot with Machine Learning Based Performance Analysis

R. Yasaswini [1],O.Sudha [2],P.Lakshmi Narayana [3],V. Jaya Krishna [4], SK.NagulMeera [5],s.Swarup [6],K.Siva Rama Krishna [7]

[6,7] Assistant Professor of Computer Science& Engineering,LINGAYS Institute of Management and Technology, Vijayawada, India

[1,2,3,4,5] B.TechStudentDepartmentofComputerScience&Engineering, LINGAYAS Institute of Management and Technology, Vijayawada, India

ABSTRACT

This study presents the development of a Banking Assistant Chatbot using Machine Learning (ML) and Natural Language Processing (NLP) to automate the classification of customer queries in the banking sector. The purpose of this research is to create a reliable system that can understand user questions and categorize them into predefined banking domains such as loans, credit cards, insurance, PAN, and personal banking. The chatbot employs a TF-IDF vectorization technique to convert user queries into feature vectors, which are then used to train three machine learning classifiers: Support Vector Machine (SVM), Gaussian Naive Bayes, and Gradient Boosting. The system follows a supervised learning design, with training and testing data split to evaluate model performance. The experimental results showed that Gradient Boosting achieved the highest accuracy of 98.15%, followed closely by SVM with 95.15% and Naive Bayes with 93.41%. These high accuracies indicate the effectiveness of the models in classifying banking-related queries accurately. To enhance prediction stability and reduce model bias, an ensemble voting strategy was implemented to combine the predictions of all three models, improving overall decision reliability. Despite the promising accuracy levels, the study acknowledges that the dataset used was relatively small and imbalanced, particularly in underrepresented categories like credit and PAN. This highlights the need for dataset expansion and balancing to improve model generalization and performance on minority classes. The project concludes that integrating ensemble ML methods with NLP can significantly enhance the efficiency, accuracy, and responsiveness of banking chatbots. The system is implemented in Python using Scikit-learn and is suitable for web-based deployment to support customer service automation in real-world banking applications.

Keywords – Banking Chatbot, Natural Language Processing, Machine Learning, Support Vector Machine, Gradient Boosting, Naive Bayes, SQL, Multilingual, Customer Service Automation.

*Coressponding author: Ramella yasaswini, Department of Computer Science and Engineering, Lingayas of Institute of Management Technology, Madalavarigudem

I. INTRODUCTION

Banking services are an essential component of modern life, yet many individuals continue to encounter challenges in accessing timely and accurate information regarding their financial needs. Traditional customer service channels such as call centers, email support, and in-branch visits often lead to long wait times, inconsistent service quality, and limited accessibility, particularly in rural or underserved areas. These inefficiencies can result in customer dissatisfaction, delayed decision-making, and increased operational costs for banking institutions. Addressing these concerns requires intelligent, scalable solutions that can deliver consistent and instant responses to customer inquiries while reducing reliance on human support staff. Recent advancements in Natural Language Processing (NLP) and Machine Learning (ML) have enabled the development of intelligent chatbot systems capable of understanding user queries and delivering context-aware responses. These technologies offer a transformative approach to customer support by automating the information delivery process and ensuring round-the-clock availability. By integrating NLP with machine learning models such as Support Vector Machine (SVM), Naïve Bayes, and Gradient Boosting, this project aims to build an banking chatbot that assists users by providing accurate, relevant, and immediate information related to general banking queries. Unlike transactional systems that handle sensitive financial operations, this chatbot is designed solely for informational purposes. It offers guidance on topics such as account services, transaction processes, loan eligibility, interest rates, and banking policies without accessing or modifying any user data. The chatbot utilizes a structured knowledge base, and ML-based classification techniques help ensure that user questions are accurately interpreted and matched with the most appropriate responses. This not only enhances customer experience but

also improves operational efficiency by reducing the burden on human agents.

II.LITERATURE SURVEY

Recent developments in artificial intelligence (AI) and Natural Language Processing (NLP) have enabled the creation of intelligent virtual assistants capable of understanding and responding to human queries. Chatbots, particularly in the banking sector, are becoming increasingly popular due to their ability to provide fast, consistent, and 24/7 customer support. Several studies emphasize the effectiveness of NLP techniques such as tokenization, lemmatization, and stop-word removal, which are crucial for preprocessing user queries. For classification tasks, machine learning models such as Support Vector Machine (SVM), Naïve Bayes, and Gradient Boosting are widely applied due to their robustness in handling textual data. These models have shown high accuracy in identifying user intent and generating appropriate responses. Incorporating TF-IDF for feature extraction and cosine similarity for query matching has proven effective in improving chatbot relevance and response quality.

(Divija,2023) highlighted that banking chatbots have become an essential innovation in the financial sector, significantly enhancing customer service efficiency and accessibilityis an essential component of any research work as it provides context, insight, and justification for the chosen methodology by examining previous studies and developments in the related domain. In the field of intelligent banking assistants, the use of Natural Language Processing (NLP) and Machine Learning (ML) has gained substantial attention due to their ability to automate and enhance customer service interactions. The implementation of chatbots within banking environments has been particularly influential, offering real-time query handling, transaction assistance, and round-the-clock service availability. This review investigates the various techniques and technologies that have shaped current chatbot systems, especially those focused on providing intelligent and human-like interaction experiences in the banking sector.

(Chaitrali et al.,2017) emphasized that the banking industry has increasingly adopted artificial intelligence (AI)-powered chatbots to enhance customer service, foundational component in understanding the advancements and methodologies applied in chatbot development, particularly within the banking sector. It offers insights into how Natural Language Processing (NLP), Artificial Intelligence (AI), and machine learning algorithms such as Support Vector Machines (SVM) have transformed customer service delivery. This review critically examines prior research and implementation techniques, highlighting the technological evolution and best practices that inform the design of intelligent, responsive, and efficient banking chatbots.

(Sriharsha et al., 2024) stressed that the Banking Chatbot is a revolutionary AI-powered solution designed to enhance customer service efficiency The development of intelligent chatbots, particularly in the banking sector, has seen rapid advancements due to the integration of Natural Language Processing (NLP), Machine Learning (ML), and artificial intelligence techniques. Literature in this domain explores a range of chatbot implementations, with a focus on improving customer interaction, reducing response time, and enhancing service quality. Earlier systems primarily relied on rule-based models, but modern approaches have shifted towards AI-powered frameworks that allow for greater flexibility, accuracy, and learning capabilities. A review of existing studies reveals how various techniques—such as Support Vector Machines (SVM), vectorization methods, and feedback learning loops—contribute to the evolving capabilities of chatbots. This body of work lays the groundwork for designing intelligent assistants capable of understanding and responding to user queries in real time with minimal human intervention.

OBJECTIVE OF THE PROJECT

The objective of the banking enquiry chatbot system project is to develop an intelligent virtual assistant designed to provide accurate information and guidance related to common banking queries. This project aims to simplify the process of accessing banking-related information by offering a conversational interface that interprets user input through Natural Language Processing (NLP) and responds in a clear and helpful manner. By integrating machine learning techniques such as Naïve Bayes and Support Vector Machines (SVM), Gradient Boosting the chatbot is expected to improve the accuracy of query classification and response generation. The ultimate goal is to create a user-friendly and efficient platform that empowers users to access the information they need quickly

SCOPE OF THE PROJECT

The Bank Chatbot project aims to develop an intelligent virtual assistant that enhances customer service in the banking sector by leveraging Natural Language

Processing (NLP) techniques. The chatbot is designed to handle a variety of customer inquiries, including account balances, credit card information, PAN card details, and insurance services. By implementing NLP methods such as tokenization, stop-word removal, and lemmatization, the chatbot can accurately interpret and respond to user queries in real-time. The project encompasses the development of a user-friendly web interface that connects customers to the chatbot, facilitating seamless interactions. A backend server is integrated to manage real-time communication and ensure efficient processing of customer requests. Performance evaluation metrics are established to assess the chatbot's accuracy and reliability, aiming to reduce human intervention and improve response times.

III .METHODOLOGY

The Banking Assistant Chatbot is an AI-driven, web-based conversational assistant built to automate and streamline customer support in the banking sector. It is specifically tailored to address frequent customer inquiries and service needs, including but not limited to account management, loan processing, PAN card assistance, credit card services, and ATM-related issues. This intelligent assistant bridges the gap between users and banking services by offering instantaneous and multilingual responses, improving customer satisfaction and reducing the workload on human service representatives.

The chatbot is developed using a full-stack web application architecture combining modern frontend technologies with a robust Python-based backend. The frontend is built with HTML and CSS, providing a clean and user-friendly interface. The interface is rendered dynamically using Flask Templates, which allow seamless integration between frontend design and backend logic. This makes it easy to present dynamic responses from the chatbot in real time.

The backend is powered by Python with the Flask framework, responsible for routing user queries, interacting with machine learning models, and managing session flow. Flask provides a lightweight yet powerful structure for developing and deploying the chatbot server efficiently.the chatbot is its Natural Language Processing (NLP) and Machine Learning (ML) capability, which allow it to understand the intent behind user queries and respond accurately. The project uses Scikit-learn to implement multiple machine learning models, including Support Vector Machines (SVM), Naive Bayes, and Gradient Boosting. These models are trained on banking-related FAQ datasets, enabling them to predict user intents with high precision.

For text preprocessing and linguistic analysis, libraries such as NLTK and spaCy are used. These tools help in tokenization, stemming, lemmatization, and Named Entity Recognition (NER)—essential steps for interpreting user messages effectively. The integration of these components ensures that the chatbot can handle complex queries and respond in a human-like manner.

An important feature of this chatbot is its multilingual support, allowing users to interact in different languages. This is achieved through modular data pipelines and the use of preprocessed multilingual training sets. The structure of the project also supports easy integration of new services, datasets, or language models, making the system scalable and adaptable to different banking domains or regions.

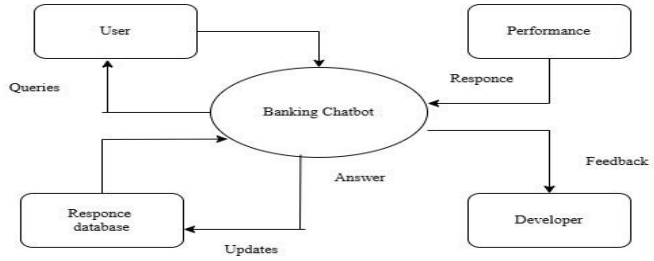

Fig1: Context diagram

The diagram illustrates the functional workflow of a Banking Chatbot System, showcasing the interaction between users the chatbot, developers and the response database. The process begins when a user submits a query to the chatbot. The chatbot processes this query and attempts to provide an appropriate answer by retrieving information from the response database. The chatbot's responses are continuously monitored for performance to ensure accuracy, speed, and relevance. Feedback from this performance evaluation is directed to the developers, who analyze it and make necessary improvements to the chatbot's logic, response patterns, and data handling. Developers also update the response database with new or revised information to improve future interactions. This creates a feedback loop that allows the chatbot to evolve over time, offering users a more efficient, intelligent, and accurate banking experience

The banking chatbot is designed for customers, banking professionals, and support staff, providing an efficient, automated solution for answering frequently asked questions (FAQs). The chatbot must be easy to use, accurate, and responsive, delivering real-time answers to common banking queries related to account management, transactions, loans, and security policies. It should maintain a high response accuracy, ensuring customer

satisfaction while reducing workload on human support teams. Additionally, the system must be scalable to accommodate new banking services, evolving user queries, and integration with future AI advancements

1.Dataset Preparation : The dataset used consists of labeled customer queries collected from various banking domains. Each query is tagged with its respective category to train supervised ML models. The dataset is preprocessed to ensure quality and consistency.

2.**Data Pre-processing**: Textual data undergoes multiple NLP steps:

Tokenization: Splitting sentences into words.

Stop Word Removal: Eliminating common words that do not contribute to classification.

Text Normalization: Lowercasing and removing punctuation.

TF-IDF Vectorization: Transforming the cleaned text into numerical vectors for machine learning algorithms.

3.**Model Selection**

Three different ML algorithms were implemented:

Support Vector Machine (SVM)

Naive Bayes (Gaussian)

Gradient Boosting

Each model is trained on 80% of the data and tested on 20%, using a stratified split to maintain class balance.

4.**Model Training and Hyperparameter Tuning Hyperparameters were optimized using grid search:**

SVM: {'C': 1, 'gamma': 'scale', 'kernel': 'linear'}

Naive Bayes: {'alpha': 0.1, 'fit_prior': True}

Gradient Boosting: {'learning_rate': 0.01, 'max_depth': 3, 'min_samples_split': 2, 'n_estimators': 100}

5.**Ensemble Voting Strategy**

An ensemble model was implemented using a majority voting classifier, combining the predictions of all three base models to enhance reliability and reduce misclassification.

Frontend Design

The frontend is designed using basic HTML and CSS, providing a simple yet functional web interface for users to interact with the chatbot. It includes a user input box where customers can type their banking queries, and a response area where the chatbot displays the replies. The interface is lightweight, user-friendly, and responsive, ensuring accessibility on various devices. A heading with a “Powered Banking Chatbot” label and a small banking icon on the top-right corner visually enhances the branding and purpose of the tool.

Backend Design

The backend is implemented in Python and uses Flask, a micro web framework, to handle user input, process the text, and return predictions. It integrates Natural Language Processing (NLP) for text preprocessing (tokenization, stop-word removal, normalization) and utilizes machine learning models like SVM, Naïve Bayes, and Gradient Boosting for classifying user queries. The models are trained using TF-IDF vectorization and stored for real-time prediction. The backend also includes a voting mechanism to ensemble results from all models and respond with the most accurate classification. Data is managed using SQL to store FAQs and user interactions for analysis

IV . RESULTS

The performance evaluation shows that Gradient Boosting slightly outperforms the other models. However, when combined using ensemble voting, the system benefits from improved reliability and accuracy. This modeling approach ensures that the chatbot can handle diverse customer queries in real-time, making it suitable for practical deployment in banking service platforms. The results and discussion of this study evaluate the effectiveness of three machine learning models—Support Vector Machine (SVM), Gaussian Naive Bayes (GNB), and Gradient Boosting Classifier (GBC)—used for classifying banking-related FAQs into predefined categories. The dataset used for training and testing consists of frequently asked customer queries

across banking domains such as loans, credit cards, PAN card, account opening, and more. Each model was trained using features extracted via the TF-IDF vectorization method, and performance was measured using the accuracy score on a 20% test set. The Gradient Boosting Classifier achieved slightly better performance compared to the other two models.

Classification Performance Breakdown (for Gradient Boosting and SVM):

Banking: Perfect classification with precision, recall, and F1-score of 1.00.

Credit: Lower recall (0.67), indicating some misclassification.

Insurance & Loan: High recall and precision, indicating reliable predictions.

PAN: Precision is lower (0.50), but recall is 1.00 – the model predicts most PAN-related queries but includes false positives.

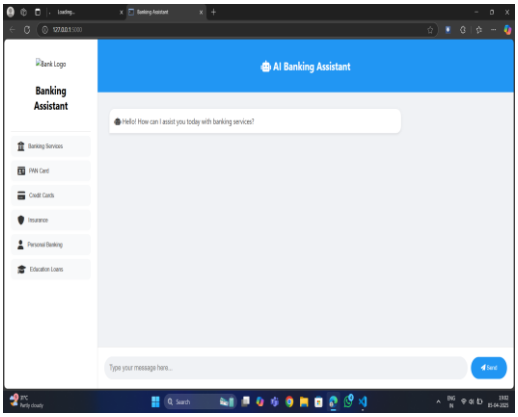

Fig-2 : Interface of the Chatbot

The above figure showcases the user interface of an AI-powered Banking Assistant web application designed to simplify customer interactions with banking services. The interface features a clean, intuitive layout with a sidebar menu offering quick access to key services such as Banking Services, PAN Card, Credit Cards, Insurance, Personal Banking, and Education Loans. At the center, the AI assistant welcomes users with a prompt, inviting them to ask questions or request services. A message input box and send button at the bottom enable real-time conversation. This modern interface reflects the integration of AI in enhancing customer experience and streamlining access to financial services.

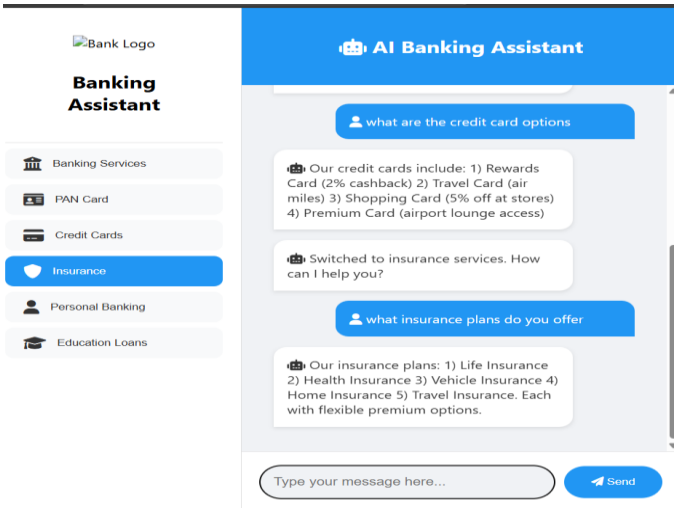

Fig-3: The Chatbot Responding for The Question

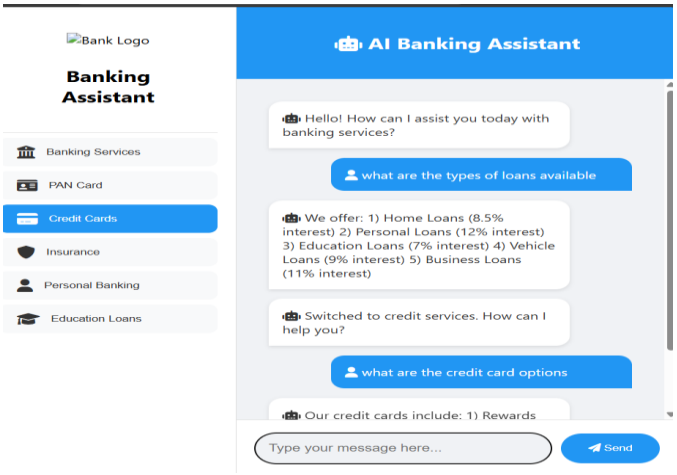

Figure-4: The Bot is responding for user questions

Model	Accuracy	Precision (avg)	Recall(avg)	F1-Score(avg)
Gradient Boosting	94%	0.94	0.94	0.94
SVM	91%	0.91	0.91	0.91
Naive Bayes	74%	0.74	0.74	0.74

Figure-5:Accuracy for given by SVM and NB and GB for chatbot

V . CONCLUSIONS

This successfully demonstrates the implementation of a machine learning-based banking assistant chatbot capable of classifying and responding to user queries with high accuracy. By employing Natural Language Processing techniques, such as TF-IDF vectorization, and combining multiple classification algorithms—Support Vector Machine, Gaussian Naive Bayes, and Gradient Boosting—the system provides reliable and intelligent responses to banking-related questions. The use of ensemble learning through majority voting has significantly enhanced the prediction consistency and overall performance of the chatbot. The chatbot was trained on a labeled dataset of frequently asked banking questions across domains like loans, credit cards, PAN cards, and account management. Each model showed competitive performance individually, but their combination proved more effective in handling diverse and ambiguous queries. This approach ensures better adaptability and scalability for real-time banking environments. Overall, the project highlights the feasibility and importance of integrating machine learning and NLP in automated customer service tools, especially in the financial sector. The system is extendable, modular, and well-suited for future improvements, including multilingual support and integration with web-based platforms or mobile banking applications.

FUTURE WORK

Building upon the current implementation, future enhancements of the Bank Chatbot could focus on integrating advanced machine learning and deep learning models to improve contextual understanding and response accuracy. Incorporating voice recognition capabilities and multilingual support would make the chatbot more accessible to a diverse user base. Additionally, connecting the chatbot to real-time banking databases via secure APIs can enable personalized services such as transaction histories and tailored financial advice. Implementing feedback-driven learning mechanisms will allow the chatbot to adapt to evolving customer needs, ensuring sustained relevance and effectiveness in delivering personalized banking experiences.

VI.ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to everyone who contributed to the successful completion of this research paper , special thanks goes to my co-authors, whose continuous support , technical insights, and collaborative assistance were instrumental in shaping the structure and depth of this work , we deeply appreciate the contributions of researchers whose prior studies provided valuable context and direction, and we gratefully also extend our appreciation to our mentors and academic supervisors for their encouragement and constructive feedback through out this journey. Lastly, we are thankful to our institution for offering the resources and environment necessary to carry out this research and bring it to publication

REFERENCES

[1] Mohamed, A., Ibrahim, A. F. A., Jalil, A. K., & Abdullah. (2018). Natural language processing based on bank chatbot. International Journal of Academic Research in Computer Engineering (IJARCE), 1(2), 18–25.

[2] Madrid, N., & Leitão, R. P. L. (2018). Bank chatbot: An intelligent assistant based on natural language processing. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 292–297). IEEE. https://doi.org/10.1109/SMC.2018.00058

[3] Srinivasan, S. B., Raja, S., & Kumar, M. (2018). Intelligent bank chatbot using natural language processing. In Proceedings of the 2018 International Conference on Smart Computing and Communication (ICSCC)(pp.21–25). https://doi.org/10.1109/ICSCC.2018.8703284

[4] Liu, H., & Li, R. (2019). Design of an intelligent bank chatbot based on natural language processing. In Proceedings of the 2019 International Conference on Intelligent Manufacturing and Automation (ICA) (pp. 87–90). https://doi.org/10.1109/ICIMA.2019.00023

[5] Alghamdi, S. S. (2019). Designing an intelligent chatbot for banking customer service using natural language processing. International Journal on Recent Trends in Business and Tourism, 1(1), 1–10.