

Banking Chatbot Using NLP and Machine Learning

B.Sidharth Varma¹, G.Srija², M.Abubakar³, Dr.Preethi Jeevan⁴

^{1,2,3} B.tech Scholars, Dept. of Computer Science and Engineering, Hyderabad-501301, India

⁴Associate Professor, Dept.of Computer Science and Engineering, SNIST, Hyderabad 501301, India

Date of Submission: 01-05-2023

Date of Acceptance: 08-05-2023

ABSTRACT: The increasing demand for efficient and personalized customer service in the banking industry has led to the development of Chatbot technology. In this project, we propose the development of a Bank Chatbot, which acts as an intelligent agent and provides personalized customer experience using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The Bank Chatbot will be designed to handle customer inquiries, such as account balance, transaction history, and account details, in a conversational and intuitive manner. The data for this project is collected from open source like Github or Kaggle and perform the implementation. The Bank Chatbot will be trained on a large dataset of customer queries and responses, using state-of-the-art Natural Language Processing (NLP) techniques for text pre-processing and Machine Learning (ML) models such as Support Vector Machine (SVM) and NavieBaye's (NB) Classifier. The conclusion is based on the present performance of the system can be further improved by using advanced machine learning algorithms such as Naive Bayes having 90.6% accuracy which has shown better accuracy than SVM is having the accuracy of 76.2% and highlights the potential benefits of using chatbots in the banking industry. This will enable the system to understand and respond to customer queries accurately and efficiently, then we will create a front-end application by connecting it to a server and provide answers to the queries. Hence, a convenient customer and bank interface can be developed.

Index Terms— Chat bot, bank, classification, NLP, Machine Learning, Streamlit.

I. INTRODUCTION

BANKING CHATBOT using NLP and Machine Learning is a project aimed at developing a chatbot system for banks that can assist customers with their inquiries, account information, and transactions. This system will utilize Natural Language Processing (NLP) and Machine Learning

(ML) algorithms to provide an intelligent and personalized customer experience. Developing a chatbot system that meets these requirements presents several challenges. The system must be able to accurately understand and interpret customer inquiries, which can vary in complexity and language. The system must also be able to handle a high volume of customer inquiries simultaneously, ensuring that wait times are minimized. Additionally, the system must be able to recognize and personalize responses based on customer history and preferences to provide a more tailored experience. Here we create a customer front-end application which further enhances the customer experience. This hypothesis is supported by previously published literature that highlights the potential benefits of chatbot systems in enhancing customer service in various industries, including banking. To reduce the burden on human customer service representatives and provide customers with immediate assistance at any time. The base paper of this project is a literature survey of previously published works that have explored the use of chatbot systems in customer service. The base paper highlights the potential benefits of using NLP and ML algorithms in developing chatbot systems that can handle complex customer inquiries and provide accurate responses in natural language.

II. LITERATURE

SURVEY (BACKGROUND STUDY)

a. Title- Review on implementation techniques of chatbot:

The methodology employed for this paper is a critical review of chatbots and their current development strategies. The study primarily relies on the analysis of existing literature in the field of chatbots and AI, including academic research papers, industry reports, and online publications. The review is focused on exploring the functionalities and limitations of chatbots, the available development frameworks, and the

underlying technologies that support their implementation. The study reveals that chatbots are intelligent systems developed using AI and NLP algorithms that interface with users and answer inquiries. They are widely used by organizations, government associations, and non-profit organizations, and are deployed by financial institutions, online retail stores, and startups. The review highlights the challenges and limitations of chatbot development, including the handiness and flexibility of real dialogues. Also, popular intelligent personal assistants such as Amazon's Alexa, Microsoft's Cortana, and Google's Google Assistant are identified in the study. The capabilities of these automated assistants are lacking and today's chatbots use rule-based methods, intuitive machine learning algorithms or retrieval techniques that do not produce satisfactory results.

b. Title-Enterprise Crowd Computing for Human Aided Chatbots:

The methodology used in this paper involved a literature review of existing research on chatbots, their limitations, and the concept of Human Aided Chatbots. The paper also discusses the potential benefits and drawbacks of Human Aided Chatbots and how enterprise mesh computing can help overcome the challenges they face. The authors introduce ECrowd, a platform designed for consolidating training data for rational systems, as a potential solution. The paper highlights the limitations of fully automated chatbots and how Human Aided Chatbots can help overcome these limitations by leveraging human intelligence. The authors also discuss the potential benefits of using enterprise crowd computing to assist Human Aided Chatbots, such as scalability and low-latency. However, the paper also acknowledges the potential challenges and concerns related to privacy and security when using chatbots operated by the humans.

c. Title-"Nudge Your Workforce: A Study on the Effectiveness of Task Notification Strategies in Enterprise Mobile Crowd sourcing":

A survey of IBM Benelux's 93 staff members was carried out to determine the factors that might influence engagement with mobile business crowdsourcing. To study the effectiveness of different notification strategies for specific tasks, the results of the survey has been used to develop an experiment. The impact of factors such as time and the context for notification on staff participation and retention was assessed in this study. The study found that break times were the

most suitable for crowd work. Employees were more likely to participate in crowd sourcing tasks during their breaks, as compared to other times of the day. In addition, the study showed that "aggressive disclosure strategies" had been shown to be a deterrent for participation, while simple but periodic nudges were significantly more likely to help retain contributions of proponents. The study also identified several other factors that affected engagement in mobile enterprise crowd sourcing, including perceived usefulness of the task, intrinsic motivation, and job autonomy.

III. METHODOLOGY

Based on our literature review, we found that existing chatbot systems in the banking industry have some limitations, such as limited ability to understand natural language queries and provide accurate responses, lack of personalization, and inability to handle complex requests. The proposed algorithm will rectify these drawbacks by utilizing state-of-the-art NLP and Machine Learning techniques such as Support Vector Machine (SVM) and NavieBaye's (NB) Classifier to understand and appropriately and quickly answer to natural language inquiries. Then we will be creating a front-end application for a more effective customer bank interface. This a very eye-catching feature of our project, where in the time complexity is reduced and better experience is provided to the customer. The system architecture is shown below:

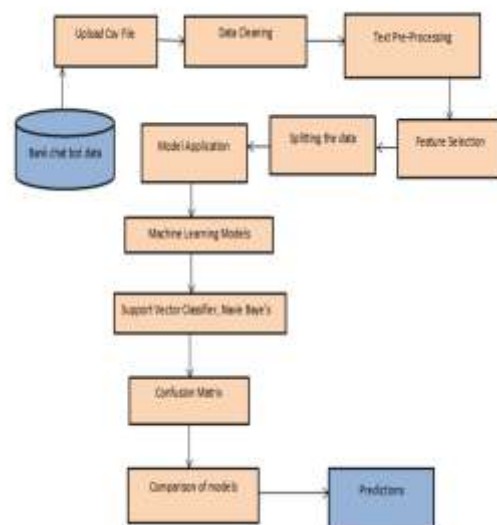


Fig (3.1) System Architecture

a. DATA SET INFORMATION:

The data set is collected from an open-source tool, namely Github. The link for the data set is mention below:

<https://github.com/MrJay10/banking-faq-bot>

b.PYTHON PROGRAMMING:

Python is an advanced programming language designed to make reading and using it easy. It can be used freely, even for commercial purposes, because it is open source. Python, as well as other related virtual machines are available on Mac, Windows or Unix systems.

Web applications and dynamic Web content is largely created on the python platform or other frameworks like Ruby, Perl etc. A range of multi-dimensional imaging applications, making it possible for users to develop custom plugins and extensions, also support python. The Python script.PY files can be interpreted and executed very easily; this increases the efficiency of the code. The compiled programs.pyC files, commonly used as programming modules that other Python programs may refer to, are also available for saving.Python comes with a number of interesting features that distinguish it from the rest of the programming languages. Object-Oriented Programming, Procedural Programming and Dynamic Memory Allocations are supported.

c.JUPYTER NOTEBOOK:

This is an open-source web application that allows you to create and develop major projects which include code implementation, driver codes front the application development and exploratory data analysis and also the test narrative. Its applications include the cleaning and conversion of data, numerical simulation, statistical modelling, data visualization, machine learning or more.This application was once known as IPython Notebook is a web-based interactive computational environment for creating documents for interactive web applications. Project Jupyter is a project dedicated to developing opensource software in various programming languages for interactive computing, according to its official website.

d.NUMPY:

You have a Python package named NumPy. It's called 'Numerical Python'. This library is composed of multidimensional array objects, as well as a collection of arrays processing routines. The new Numarray has also been developed, which adds a few other features. Travis Oliphant's Numpy package was developed in 2005 and incorporated features from Numarray into the Numeric Package. The Open- Source project is being supported by a number of contributors.NumPy's main feature is its ability to perform mathematical computations on entire

arrays, which makes it significantly faster than traditional Python code that relies on loops. This library is widely used in the scientific and research communities because of its efficiency and versatility.

e.PANDAS:

Pandas is a well-liked open-source Python toolkit for handling and analysing data. It provides fast, efficient, and easy-to-use data structures for working with tabular, structured, and time-series data. The two primary data structures in Pandas are Series and Data Frame, which allow for easy indexing, filtering, aggregation, and transformation of data. With its rich set of tools for data manipulation and analysis, Pandas has become a standard tool for data scientists and analysts working with data in Python.

f.SCIKIT-LEARN:

The most effective and robust machine learning package available for Python is Scikitlearn. A set of rapid tools for machine learning and statistical modeling, such as classification, regression, clustering or dimensionality reduction has been developed using the Python Consistency Interface. The basis of this package, which is mostly based on the Python language, are NumPy, Scipy and Matplotlib.SciKit features a number of functions that help you learn everything from data manipulation to metrics processing One of the greatest things about Scikit is its built learning algorithms for machine learning, which can be only tried with minimal alterations such as segmentation, regression, mode, model.

g.NATURAL LANGUAGE PROCESSING:

Natural Language Processing (NLP) plays a crucial role in the development of the proposed Banking Chatbot using NLP and Machine Learning.The system will utilize various NLP techniques to analyze and understand customer queries in natural language and provide accurate responses in real-time. These techniques include text pre-processing techniques such as tokenization, lemmatization, and cosine similarity, as well as sentiment analysis and topic modeling. The system will be trained on a large dataset of customer interactions, allowing it to learn and enhance its response efficiency over time. We will be using Count Vectorizer class for performing natural language processing. By leveraging NLP, the system will be able to provide a personalized and efficient customer service experience, improving customer satisfaction and loyalty.

```

In [7]: from sklearn.feature_extraction.text import CountVecorizer
vectorizer = CountVecorizer()

In [8]: from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

Splitting Data into Train and Test

In [9]: train_data, test_data = train_test_split(X, test_size=0.1, random_state=42)

Applying Word embedding Techniques on Features

In [10]: # Transform the test data into feature vectors
X_train = vectorizer.fit_transform(train_data['question'])
X_test = vectorizer.transform(test_data['question'])
  
```

Fig (3.2) Text preprocessing using NLP

h.SUPPORT VECTOR CLASSIFIER:

In the Bank Chat Bot project, Support Vector Machine (SVM) algorithm is one of the techniques in machine learning used to build an intelligent and wise system that can provide personalized and efficient customer service in the banking industry. SVM is a powerful algorithm for classification and regression tasks, widely used in various domains, including Natural Language Processing (NLP). In the project, the SVM algorithm is trained on a large dataset of customer interactions to classify the customer's query into different categories, such as account information, transaction history, and bill payments, and provide accurate responses.

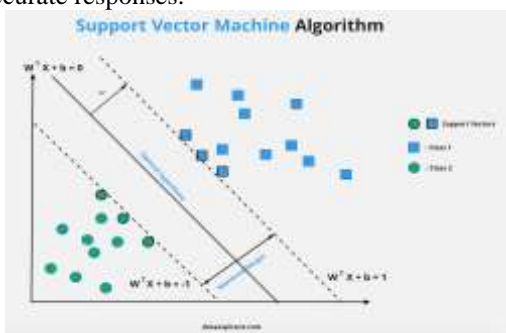


Fig (3.3) Support Vector Machine mechanism

i.NAIVE BAYES:

Another supervised machine learning algorithm that is commonly used for the class separation is the Naïve Bayes algorithm. Naïve Bayes is based on Bayes' theorem, which is a conditional probability theory that explains the likelihood of an event occurring based on the previous knowledge of conditions that might be related to the event. Naive Bayes works by

assuming that all features are independent of each other, which is why it's called "naive". In spite of this presumption, the Nave Bayes can often be very good in situations where there are a large number of variables and limited data. In spite its simplifying assumptions, Naïve Bayes has shown to be effective in many applications, including text classification. In this project, it can be used in conjunction with other machine learning algorithms such as SVM to improve the chatbot's overall accuracy and performance.

j.STREAMLIT:

The scope of data science and analytics is increasing rapidly day by day and the most important issue to be dealt with is the model deployment. Streamlit is a python library which helps us to create applications to deploy our model. This adds more definition to our problem statement. In our project we will be creating an application where in the customer can ask any queries which will be answered by our model. The main purpose of this application is to create a very sustainable environment for the customer to access the chat bot.

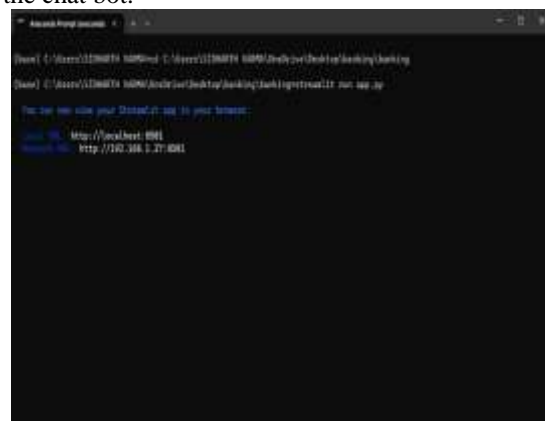


Fig (3.4) Command prompt implementation for application

k.UML DIAGRAMS:

To specify, visualize, build and document software systems' artifacts, UML is a standard language. It stands for Unified Modelling Language. The development of the UML requires a number of objectives, but defining some General Model Language that can be used by every modeler is essential in order to facilitate their understanding and application. The system may be a computer, or not a computer. It must therefore be made clear that UML is not a development technique, rather it forms part of the process in order to create an effective system. Lastly, the UML goal can be defined as an easy modelable tool to build any type

of useful system we might encounter from today's complicated environment.



Fig (3.5) Use case diagram

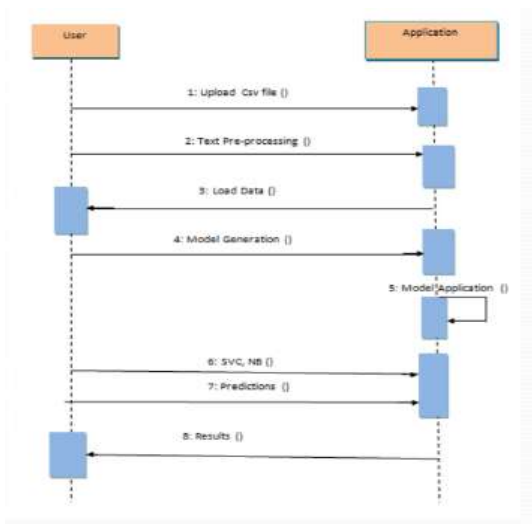


Fig (3.6) Sequence diagram

1.EVALUATION METRICS:

Apart from our new measurement, the corona score, it is worth noting that a number of statistical measures are being used to evaluate the performance of the planned architecture.

- **ACCURACY:** The measurement of the methods ability to identify the right expected cases.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

1. **True Positive:**The count of cases that have been accurately predicted are equal to the True Positive.

2. **False Positive:**The count of false positive cases is the same as the count of wrong expected true cases.
3. **True Negative:** The count of correct negative forecasted cases is equal to the true negative.
4. **False Negative:** False negatives are equivalent to the number of incorrectly forecast adverse cases.

- **RECALL:** It is the metric which gives the sensitivity of the method.

$$Recall = \frac{TP}{TP+FN}$$

- **PRECISION:** Precise is the measure of accuracy which only considers the true positives and false positives.

$$Precision = \frac{TP}{TP+FP}$$

- **SPECIFICITY:** Specificity is the model's ability to predict the ratio of true negative to the sum of true negative and false positives.

$$Specificity = \frac{TN}{TN+FP}$$

- **F1- SCORE:** The F1 score shall be the measure of the complete model's accuracy, that we get as the output of the project.

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

IV.CONCLUSION

In conclusion, the Bank Chatbot using NLP and Machine Learning is a promising solution for the banking industry to improve customer service and automate various banking processes. The performance of the system can be furtherimproved by using advanced machine learning algorithms such as Naive Bayes having 90.6% accuracy which has shown better accuracy than SVM is having the accuracy of 76.2%. With

the advancement of technology, it can be expected that the chatbot system will become fully integrated into banks operations, providing more efficient and personalized services to customers. Using streamlit we have created the application for the chat bot for establishing a more effective interface between the bank and the customer.

	precision	recall	f1-score	support
accounts	0.91	0.88	0.89	57
cards	0.93	0.94	0.94	88
fundstransfer	1.00	0.67	0.80	6
insurance	0.92	0.92	0.92	86
investments	0.80	0.89	0.84	27
loans	0.94	0.91	0.92	81
security	0.60	0.75	0.67	8
accuracy			0.91	353
macro avg	0.87	0.85	0.85	353
weighted avg	0.91	0.91	0.91	353

Fig (3.7) Confusion Matrix



Fig (3.7) Front end application for chat bot

REFERENCES

- [1]. Mohamed, A. Ibrahim, A. F. A. Jalil, and A. K. Abdullah, "Natural language processing based on bank chatbot." International Journal of Academic Research in Computer Engineering (IJARCE), vol. 1, no. 2, pp. 18-25, 2018.
- [2]. N. Madrid and R. P. L. Leitão, "Bank chatbot: An intelligent assistant based on natural language processing." Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 292-297, 2018.
- [3]. S. B. Srinivasan, S. Raja, and M. Kumar, "Intelligent bank chatbot using natural language processing." Proceedings of the 2018 International Conference on Smart Computing and Communication (ICSCC), pp. 21-25, 2018.
- [4]. H. Liu and R. Li, "Design of an intelligent bank chatbot based on natural language processing." Proceedings of the 2019 International Conference on Intelligent Manufacturing and Automation (ICA), pp. 87-90, 2019.
- [5]. S. S. Alghamdi, "Designing an intelligent chatbot for banking customer service using natural language processing." International Journal on Recent Trends in Business and Tourism, vol. 1, no. 1, pp. 1-10, 2019.
- [6]. Automatic Extraction of Chatbot Training Data from Natural Dialogue Corpora, Bayan AbuShawar, Eric Atwell
- [7]. Chatbot Evaluation and Database Expansion via Crowd sourcing, Zhou Yu, Ziyu Xu, Alan W Black, Alexander I. Rudnicky
- [8]. AI BASED CHATBOT, Prof.NikitaHatwar, Ashwini Patil, Diksha Gondane
- [9]. Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber
- [10]. H. Kopka and P. W. Daly, Data Mining Practical Machine Learning Tools and Techniques Ian H. Witten Eibe Frank Mark A. Hall