CHATBOT IN BANKING DOMAIN USING NLP AND MACHINE LEARNING

B. Sidharth Varma[1], G. Srija[2], M. Abubakar[3], Dr. Preeti Jeevan[4]

[1] B.Tech Scholar, Dept. Computer Science & Engineering,SNIST,Hyderabad-501301,India
[2] B.Tech Scholar, Dept. Computer Science & Engineering, SNIST,Hyderabad-501301,India
[3] B.Tech Scholar, Dept. Computer Science & Engineering,SNIST,Hyderabad-501301,India
[4]Associate Professor, Dept. Computer Science & Engineering, SNIST, Hyderabad-501301, India

---

Abstract - There are a lot of applications in the banking domain such as YONO SBI, Phonepe bot etc. There are a lot of queries which are being asked by the customers at every stage of transaction classes such as money transfers, loan updates and other issues. Now the problem of this applications is that they are facing several interrupt issues where the responses are very slow and the sites keep on loading, this reduces the number of customers using the applications, so we need an agent which provides a very good customer interaction and is very appealing to the customers. 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 model developed in the project works with a very large dataset and a huge load applied on the model to estimate the predictions. The data for this project is collected from open source like Github or Kaggle and perform the implementation. In this project the dataset is first preprocessed using language processing tools, the data is also unstructured in the csv file. The conclusion is based on the efficiency of the model on application of the machine learning algorithms which are classification algorithms such as Naïve Bayes and Support Vector Classifier and the predictions are estimated accordingly using a front-end local host application which is dedicated to work for the customer queries. This is a very good initiate which will help various banks to provide a very good interface where in the customers can solve their queries.

***Key Words*: Bank chatbot, Customer queries, Classification, Text processing, Machine Learning.**

1.INTRODUCTION

This project **Chatbot in banking domain using NLP and machine learning** mainly revolves around developing an intelligent agent which deals with quick query responses and provides all the information regarding the bank account. The data set has several classes such as finance, security, account etc. This is the most important issue that is to be dealt with, because the model is trained with the question and the respective and the answer is to be displayed for that query. This metric of quick responses sums up the efficiency of the model. This project uses language processing tools such as NLTK which has CountVectorizer and other modules and machine learning algorithms to provide an intelligent and personalized agent with a good customer and bank environment. There are several challenges faced during developing a chatbot. The system must be able to accurately understand and interpret customer inquiries, which can vary in complexity and language. The system must be able to provide and interpret personalized responses based on customer history and preferences to provide a more tailored experience. To provide such environment a local host application is developed which only works on the customer queries and solves them, this application is built using the streamlit library. 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.

2. LITERARY SURVEY (Background study)

2.1 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.

2.2 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. Where as the opponents stated that a very good application interface can attract the customers as there is a huge ECrowd in several work spaces. This cluster of work spaces provide a very good sphere of customers that can be targeted.

3. METHODOLOGY

Based on the above studies we can see that the chatbots have a lot of 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 Navie Baye'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:

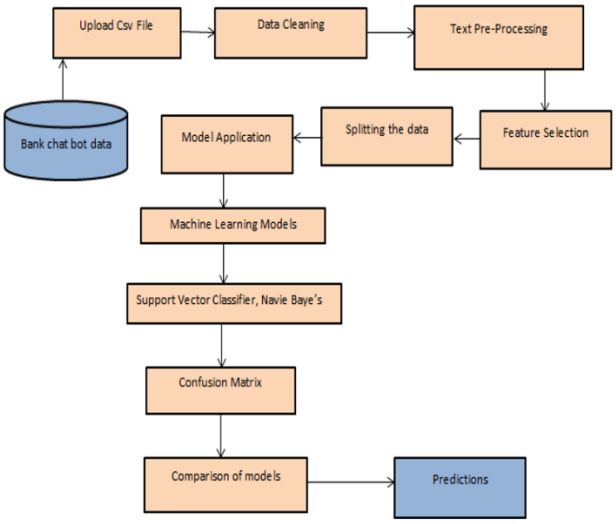

Fig (3.1) System Architecture

The diagram shows the iterative application development of the bank chatbot in stages.

i.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

Question	Answer	Class
Do I need to enter splcharacters after keying in my Card number/ Card expiry date/ CVV nu	Please listen to the recorded message and follow the instructions while entering your card deta	security
What details are required when I want to perform a secure IVR transaction	To perform a secure IVR transaction, you will need your 16-digit Card number, Card expiry date	security
How should I get the IVR Password if I hold an add-on card	An IVR password can be requested only from the registered mobile number and will be sent to	security
How do I register my Mobile number for IVR Password	Please call our Customer Service Centre and ensure that your mobile number is updated in our	security
How can I obtain an IVR Password	By Sending SMS request: Send an SMS 'PWD<space>1234' to 9717455555 or to 5676712 from y	security
Can I use the same IVR Password to perform multiple transactions	No, each IVR password can be used only for a maximum of 3 attempts (including decline attem	security
Can I generate multiple IVR Passwords	No, only one IVR password can be generated at a time. Only when the first one is used / expire	security
How do I register for IVR password	There is no registration process. However you will have to obtain a 3D Secure IVR password to	security
What is IVR Password and how does it work	To make telephonic (IVR) transactions more secure, all purchases/payments to merchants that	security
In how much time will the IVR Password be delivered to my mobile phone/email ID	IVR 3D Secure password is delivered instantly. It takes less than a minute. However, network cc	security
Can the OTP be generated prior to the transaction	No, OTP can be generated only during the transactions.	security
Can I continue using the Verified by Visa/MasterCard SecureCode/Protectbuy password fo	Yes, you can continue using your secure password, OTP is an additional option provided for che	security
On which mobile number will I receive the OTP	OTP will be sent to the registered mobile number (updated in Credit card system).	security
Is it necessary to register the card for an OTP transaction	Prior registration is not necessary for an OTP transaction.	security
Is this option available on all the online transactions	This option is available on all websites which are participating in the VBV/MSC/Portectbuy secu	security
Can the OTP be used for multiple online transactions	OTP can be used only for a single online transaction.	security

Fig (3.2) Data set

ii.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.

iii.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.

iv.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.

v.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.

vi.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. 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.

vii.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. By leveraging NLP, the system will be able to provide a personalized and efficient customer service experience, improving customer satisfaction and loyalty.

Fig (3.3) Text preprocessing mechanism

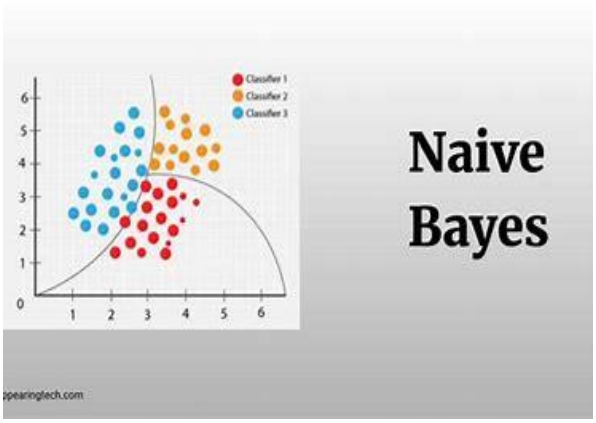

Fig (3.4) Naïve Bayes mechanism

viii.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.

ix.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.

xi.UML DIAGRAMS:

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. 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.

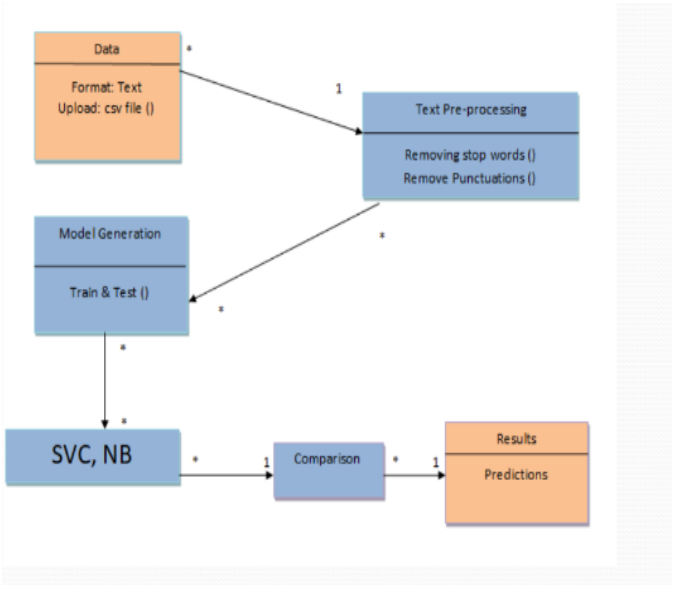

Fig (3.5) Class diagram

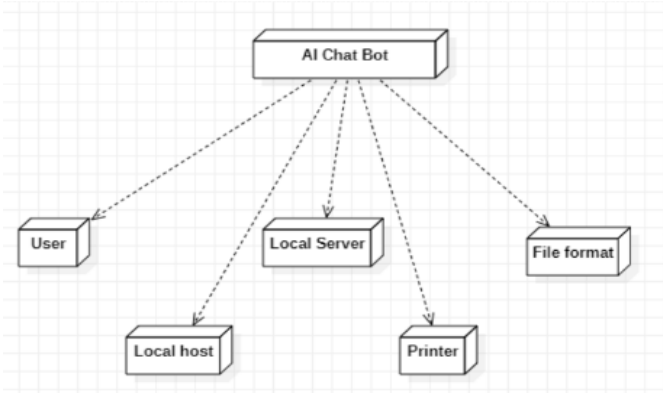

Fig (3.6) Deployment diagram

xii.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.

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

Fig (3.7) Accuracy metric

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

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

Fig (3.8) Recall metric

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

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

Fig (3.9) Precision metric

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

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

Fig (3.10) Specificity for the cases

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

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

Fig (3.11) F1 score of the model

NOTE: The notations mentioned above stand for the following:

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.

4.CONCLUSIONS

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. Upon boosting the two algorithms using NB and SVM the resultant performance was recorded as 91%. The predictions are recorded and shown accurately and the confusion matrices and classification reports are developed, 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. Then a local host front-end application will be

developed using a streamlit library and this medium is used to directly interact with the customer. This provides a special attention to the customer and solves the personalized customer queries.

	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 (3.12) Resultant classification report

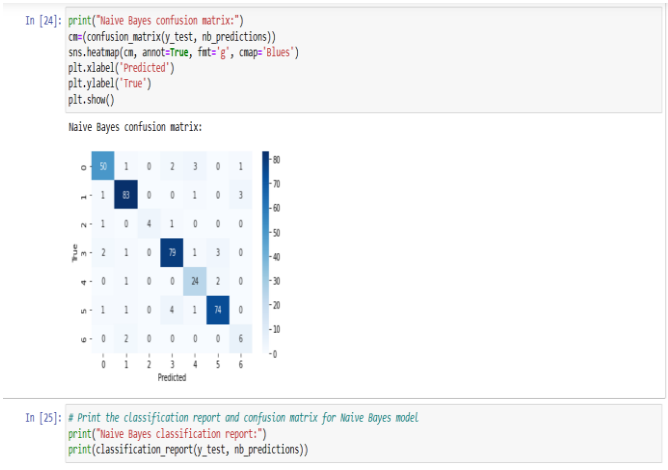

Fig (3.13) Resultant confusion matrix

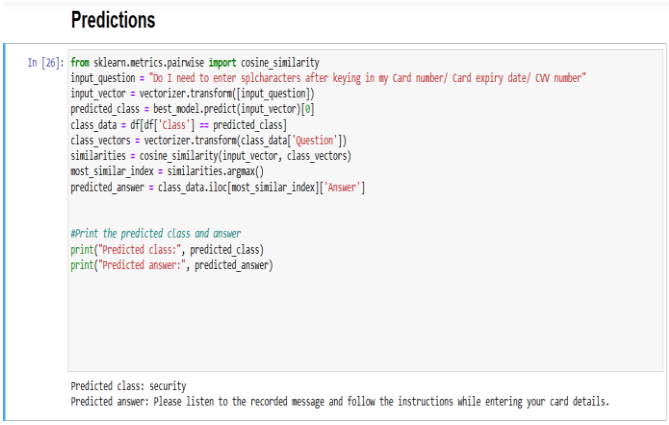

Fig (3.14) The predictions for output

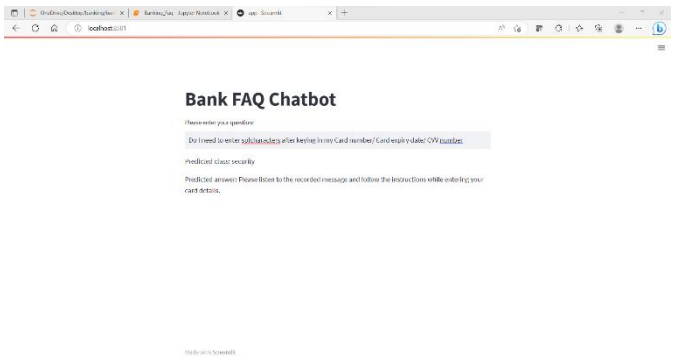

Fig (3.15) Chatbot local host application

5.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.Nikita Hatwar, 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