

# Virtual Assistant System: A Machine Learning Approach

<sup>1</sup>Lokesh Kumar Sharma , <sup>2</sup>Jyoti

<sup>1,2</sup>Department of Mathematics, Chandigarh University Mohali, Punjab, India.

---

**Abstract**— The main focus of this research is chatbots, which are intelligent devices that can comprehend users' natural language inquiries and appropriately answer during a discussion. Using it makes users feel more like they are chatting with a real person, more like a virtual assistant. They can answer questions and communicate in our language. Insufficient staff members usually take a long time to address one request, squandering time and degrading the quality of customer service at banks, customer service centres, and enquiry desks. This chatbot's main goal is to enable clients to converse by asking inquiries in plain English and receiving appropriate answers in return. The recommended solution would aid in emulating the customer care experience while still attending to and addressing the customer's queries, with the exception of the consumer speaking with a bot rather than a human being. By offering solutions for customer care centres, telephone digital assistants, and help desks, it may make life easier. The architecture, methodology, and dataset that we developed from the FAQ sections of bank websites are all described in this post. This research also examines seven categorization techniques that are employed to identify the kind of input.

**Keywords**— Chat bot, Dialog system, Classification, NLP, Vectorization.

## 1. Introduction:

Chatbots that are driven by artificial intelligence are now acknowledged as being essential for organisations since they allow for rapid customer communication from anywhere at any time. One of the crucial components for company success that banks in the financial industry are utilising is the chatbot. An artificially intelligent robot called a chatbot mimics human conversation by responding to commands, text messages, or both. Customers engage in an online conversation with a talking robot in a virtual chat. [2][5] Currently, chatbots may be found via several platforms, such as the web and messaging applications. For instance, Mastercard offers clients enhanced digital services through the use of chatbots on Facebook Messenger, allowing users to see their account balance, spending patterns, and purchase history. [7] Artificial intelligence is being used by banks in their digital strategy, particularly in developed countries like Singapore. Customers are conservative by nature, despite living in one of the technologically sophisticated nations. They

---

also reject change and prefer the conventional ways. This study examines attitudes toward chatbots deployed by banks.

#### Artificial Intelligence in Financial Services:

Technology is used to boost financial processes in a new financial industry known as "fintech" (Schueffel 2016) [1][3]. Artificial intelligence-driven innovation in technology has an impact on how people live. In fintech companies, or organisations employ in recent years, the use of AI to produce better financial services has increased. To gain an advantage over conventional banks and to assure digital progress, major financial institutions are starting to collaborate and engage in fintech. Conversational agents known as chatbots replicate human interaction. The most advanced chatbots utilise machine learning to respond to fresh data or user requests. Chatbots are often used on message-based services including SMS, WeChat (since 2013), Facebook Messenger (since 2016), WhatsApp, and others. Today, B2C marketing, sales, and customer service all use chatbots [2][5][9]. The banks experienced a substantial economic impact as a result of the 2008 financial crisis. The profit margins of the banks have shrunk during the last 10 years. So that they may continue to be profitable even in the case of a future financial catastrophe, banks are working to increase internal operational efficiency and reduce customer service expenses. Chatbot technology can manage a lot of customer calls, increase satisfaction, and change how customers see the bank. As internet behemoths like Alibaba and Amazon raise the bar for customer expectations [6], To meet the needs of banking customers, banks must be able to use and enhance the capabilities of conversational AI-backed systems.

## **2. Methods of objectivity and research:**

The project's objectives and the different methods employed to achieve them are described in the sections that follow. The approaches chosen for the investigation are looked at and backed up.

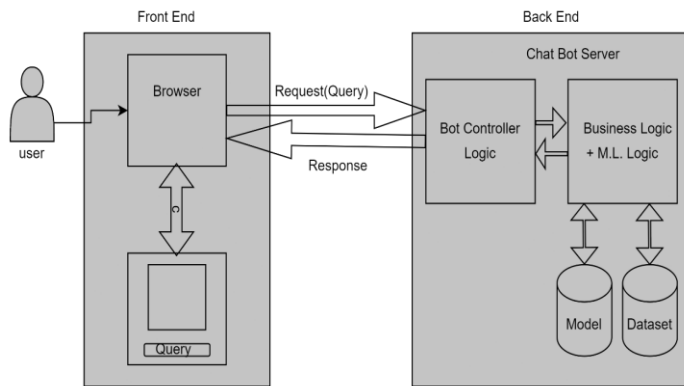
### **2.1 Project's purpose and related work**

Currently, three different approaches are utilised to build chatbots: mandate (hard-coded rules into the code), AI-based, and pattern-based (can only handle patterns specified for answer retrieval).[5] There are frameworks for creating chat bots, yet they also make use of techniques that are based on rules or patterns. Rules like If X then Y else If A then B etc. are used to create rule-based chat bots, which are the simplest to develop. must be written. As a result, the developer must design 100 rules, one for each of the 100 possible outcomes. Because of the amount, diversity, and complexity of data, such tactics are useless. It is practically hard to write rules and/or patterns for extremely large volumes of data. AI-based chatbots are created using NLP and ML. [1] They are predicated on how well humans can learn, but they are more efficient at it. Natural language processing (NLP) is a technique that may be used in place of static or present rules or patterns.

## 2.2 ARCHITECTURE

### A. Bank Chat Bot:

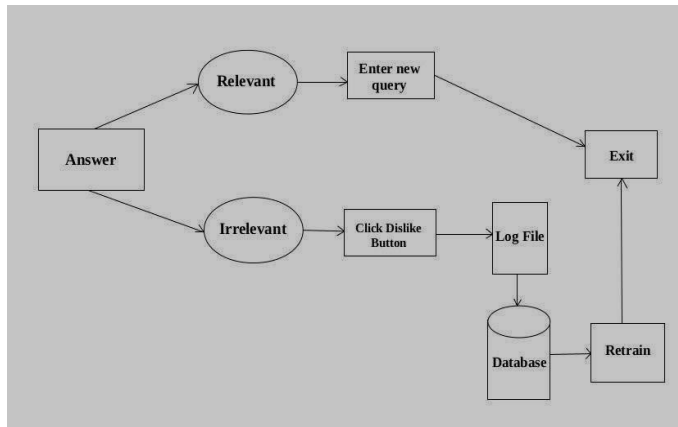
Web applications will be used by users to communicate with the system. He will type his inquiry into the text field on the web application's front end. After he pushes the Enter key or submits his request, the logic of the bot controller will take care of it. [4]



**Figure 1** Framework of chat bot functionality

The functionality of the bot controller uses the Flask framework to process user requests and deliver replies in response. After that, business and machine learning logic will get the query. Natural language processing (NLTK library) [9][4] as well as its vectorization used in business logic to pre-process user input queries. NLP will tokenize the question before extracting lemmas for each token, as well as removing any excess spaces or stop words. Then, using vectorization, this text-format question will be transformed into a vectorized representation. This modified query will now be subjected to the classification procedure in order to establish the class to which it belongs. The classification technique will be used in line with the stored model using train data from the previous run. All input questions whose class matches the determined class will be subject to the cosine similarity function. [7] According to the similarity values we determine, the most comparable response will be delivered to the user as a response.

### B. Feedback System:

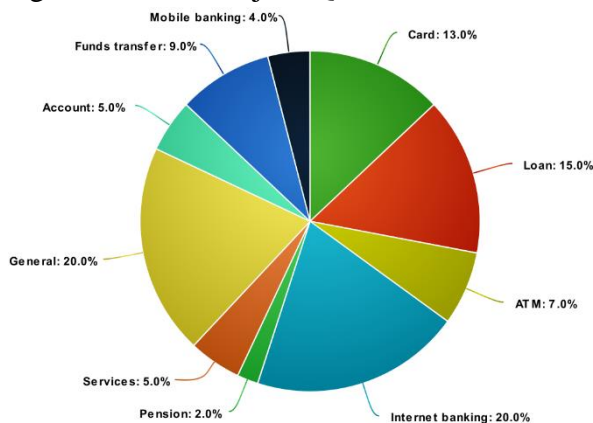


**Figure 2** Feedback system of chat bot

A chatbot might not be able to respond or give an answer when a question is given that isn't covered by the database. For our chatbot's feedback system, we developed a set of scenarios that it can handle. A web application will provide a hate button in addition to the submit button. If the user is unhappy with the system's response, he can click the Dislike button. After that, the query will either be added to an already-existing log file or a new log file created just for it. A developer now arrives on the scene to address these issues. [11][12] He will look up the answers in the classes, enter the appropriate information, and retrain the classification model. The user types the identical query in order to get the right answer the following time. As a result, the chatbot's accuracy and dataset will grow.

## 2.3 IMPLEMENTATION

**2.3.1 Creating Dataset:** Our data collection consists of the queries and responses that customers frequently ask officials at help desks or customer care centres. As a source of information, we used FAQs that we found on the websites of different banks. [3] We've used a number of web scraping programmes for this job. Question distribution throughout the Dataset:



**Figure 3** Various types of created datasets

**2.3.2 Pre-processing:** We utilised the NLTK library to parse natural language. To assist the machine, understand the English phrases that the user will be submitting, we apply natural language processing. [9] This pre-processing was done to cut down on further processing and get rid of the ambiguity that came from using the same phrase in different ways. The actions in this task are:

- I. **Tokenization** - The user's input query was tokenized to produce a word sequence. Eliminating stop words - To improve system performance, the majority of frequent words that do not need to be taken into consideration during processing are deleted, such as "desire," "are," and "can."
- II. **Lemmatization** - We used the WordNet Lemmatizer to get the lemma (root form). [4] Our data collection consists of the queries and responses that customers frequently ask bank staff at help desks or customer care centres. We used FAQs from the websites of many banks as a source of information. For this task, we employed a variety of web scraping tools. The distribution of each token's questions in the dataset, for instance, the equal treatment of the terms "processing" and "process" throughout processing. Lemmatization is used to distinguish between "process" and "processing."

**2.3.3 Vectorization:** Our text data was vectorized utilising the Bag of Words (BOG) idea. BOG is a technique for getting text ready for our machine learning algorithm's input. [2] From each text, the BOG model initially creates a vocabulary, and then it models each document by figuring out how often each word appears.

**2.3.4 Classification:** Finding connections between the user's query and the queries from the big data collection and delivering a response takes longer as the size of the data set rises. In By speeding up response and response times, efficiency may be increased, classification has been used. [10][12] With the aid of the Scikit-learn package, these classifiers were put into action. Data mining and machine learning tools may be found in Python's Scikit-learn library. In order to choose the final classifier for the chat bot that performs the best, we chose the following subset of classifiers as part of our literature study and initial training. [9]

- a) Decision Tree classifier
- b) Bernoulli Naive Bayes Classifier
- c) Gaussian Naive Bayes Classifier
- d) K-nearest neighbour classifier
- e) Multinomial Naive Bayes classifier
- f) Random Forest classifier
- g) Support vector machine

We have also made parameter changes to enhance the algorithm's performance in the context of our data set. There are two methods for carrying out parameter optimization.: i. Grid search, which is merely a form of exhaustive searching is necessary to explicitly identify a subset of a

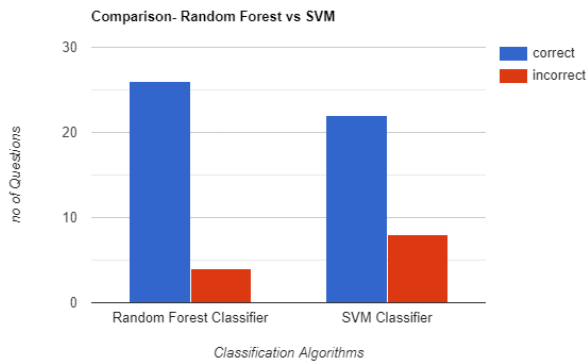
learning model's hyper-parameter space before using grid search. [4] Hyper-parameters are variables that estimators automatically learn. Grid search is expensive and exhaustive, hence randomised search is utilised in its place. As a result, randomised search samples better parameter values more frequently. In our case, we used a random search approach.

**2.3.5 Creating the learning model:** We have combined the classification, vectorization, and natural language processing algorithms into a single model at this point that will be applied moving forward. We will just obtain it whenever a new query enters the system in order to verify the query on the stored model and identify its class. [8] We may reduce the processing time by doing away require the model to be trained for each new query.

**2.3.6 Testing model:** In order to determine which classification method will work best in practice, we compare each algorithm's cross-validation score, precision score, and recall score. [6] The following table displays the outcomes of each algorithm:

**2.3.7 Choosing the best approach:** According to the findings of the above table, the Random Forest classifier and the Support Vector Machine classifier [4] are the two most accurate algorithms.

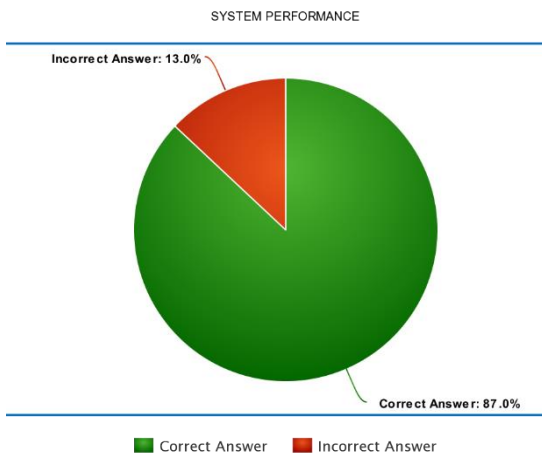
	Cross Validation Score	Accuracy Score	Precision Score	Recall Score
<b>BernouliNB Classifier</b>	0.6027	0.9252	0.9252	0.9252
<b>GaussianNB Classifier</b>	0.3893	0.8262	0.8262	0.8262
<b>Multinomial Classifier</b>	0.5966	0.9185	0.9185	0.9185
<b>Decision Tree Classifier</b>	0.5769	0.9845	0.9845	0.9845
<b>Random forest Classifier</b>	0.6187	0.9845	0.9845	0.9845
<b>SVM Classifier</b>	0.6524	0.9582	0.9582	0.9582
<b>K Neighbour Classifier</b>	0.3388	0.9845	0.9845	0.9845



**2.3.8 mapping queries to responses (Using Cosine similarity):** After receiving it from the classifier, we extract all of the questions from our data collection that fall under this classification. We check for cosine similarity between the user's query and these extracted questions. [8] The user is then presented with the most pertinent response to his question. This bot has a locked domain and is only allowed in banks. We have set a threshold on the cosine similarity measure values for handling queries that are beyond the domain.

### 3. EXPERIMENTAL RESULTS

A. Exercise 1 We evaluated a number of inquiries to confirm the Bank Chat Bot's installation. We have included similar queries from our data gathering in this experiment. An analysis of the result is given below:



B. Exercise 2. In this test, we tried out the same question in several forms. There are several techniques to find out whether a bank account may be opened, for instance:

- 1) how to open an account
- 2) account opening procedure
- 3) opening a bank account
- 4) I desire to register an account.

- 5) opening account
- 6) the steps involved in creating a bank account

The chatbot can provide the same answer accurately to each form even if we input the identical inquiry in multiple forms.

#### **4. Conclusion:**

As banks transition to hiring staff with greater digital literacy, By providing user-friendly features, a few of them have had success implementing IT. For the continued competition, Institutions of finance must automate their traditional services in order to appeal to customers who are growing more tech-savvy. More features beyond merely automated savings are predicted for chatbots in the future. It is projected that they would provide wealth management for the general public via an automated virtual assistant while considering each customer's risk profile, evaluating loans and insurance, offering data analysis and advanced analytics, and spotting and notifying fraudulent conduct.

Banking services are now accessible at any time and from any location thanks to chatbot technology. As a result, the institutions' running costs are significantly reduced. [12] They are still in the early phases of adoption and cannot yet handle sophisticated financial services. There is room for improvement when it comes to responding to customers' questions and providing them with the specific information they want. Since customers prioritise quality information while using the chatbot, this will enhance the customer experience.

#### **Acknowledgement:**

I want to sincerely thank my academic department, and my supervisor Dr. Jyoti for their patient supervision, passionate support, and helpful criticism of this effort. Last but not least, I want to thank my parents for their inspiration and support during my studies.

#### **References:**

1. Yuhua Li, David McLean, Zuhair A. Bandar, James D. O'Shea, Keeley Crockett, "Sentence Similarity Based on Semantic Nets and Corpus Statistics", IEEE Transactions on Knowledge and Data Engineering, Volume 18 - No. 8, August 2006.
2. Emanuela Haller, Traian Rebedea, "Designing a Chat-bot that Simulates an Historical Figure", IEEE Conference Publications, July 2013.
3. Pratik Slave, Vishruta Patil, Vyankatesh Gaikwad, Girish Wadhwa, "College Enquiry Chat Bot", International Journal on Recent and Innovation Trends in Computing and Communication, Volume 5, Issue 3, March 2015.
4. "AIML Based Voice Enabled Artificial Intelligent Chatterbot", International Journal of u- and e- Service, Science and Technology Volume 8 - No. 2, 2015.



5. Amey Tiwari, Rahul Talekar, Prof. S. M. Patil, “College Information Chatbot System”, International Journal of Engineering Research and General Science, Volume 2, Issue 2, April 2017.
6. Rachit Kulkarni, Ankit Methwani, Nakul Pawar, Charmi Valecha, Pooja Shetty, “College Chat-bot”, International Journal of Advanced Research in Computer Engineering & Technology, Volume 6, Issue 4, April 2017.
7. Chaitrali S. Kulkarni, Amruta U. Bhavsar, Savita R. Pingale, Prof. Satish S. Kumbhar, “BANK CHATBOT - An Intelligent Assistant System Using NLP and Machine Learning”, International Research Journal of Engineering and Technology, Volume 4, Issue 5, May 2017.
8. Yash Mehta, Shreya Sawkar, “The college chatbot”, International Journal of Computer Applications, Volume 173 - No. 7, September 2017.
9. Prof. K. Bala, Mukesh Kumar, Sayali Hulawale, Sahil Pandita, “Chat-Bot For College Management System Using A.I”, International Research Journal of Engineering and Technology, Volume 4, Issue 11, Nov 2017.
10. Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., Das, T.: A study of today’s A.I. through chatbots and rediscovery of machine intelligence. Int. J. u- e- Serv. Sci. Technol. 8(7), 277–284 (2015)
11. Shawar, B.A.: Different measurements metrics to evaluate a chatbot system. Association for Computational Linguistics, pp. 89–96 (2007)
12. Kuligowska, K.: Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. Prof. Center Bus. Res. 2, 1–16 (2015)