



# Artificial Intelligence Powered Banking Chatbot

K.Satheesh Kumar<sup>1</sup>, S.Tamilselvan<sup>2</sup>, B.Ibrahim Sha<sup>3</sup>, S.Harish<sup>4</sup>  
Assistant Professor<sup>1</sup>, BE Student<sup>2,3,4</sup>

Department of Computer Science and Engineering  
University College of Engineering, Thirukkuvallai, India

## Abstract:

Chat robot, a computer program that simulates human conversation, or chat, through artificial intelligence an intelligence chat bot will be used to give information or answers to any question asked by user related to bank. It is more like a virtual assistant, people feel like they are talking with real person. They speak the same language we do, can answer questions. In banks, at user care centres and enquiry desks, human is insufficient and usually takes long time to process the single request which results in wastage of time and also reduce quality of user service. The primary goal of this chat bot is user can interact with mentioning their queries in plain English and the chat bot can resolve their queries with appropriate response in return The proposed system would help duplicate the user utility experience with one difference that employee and yet get the queries attended and resolved. It can extend daily life, by providing solutions to help desks, telephone answering systems, user care centers. This paper defines the dataset that we have prepared from FAQs of bank websites, architecture and methodology used for developing such chatbot. Also this paper discusses the comparison of seven ML classification algorithm used for getting the class of input to chat bot.

**Keywords:** ChatBot, Artificial Intelligence, Web Scrapping.

## I. INTRODUCTION

Banks play an important role in every country's economic development. In day-to-day life, everybody needs banks. But most of the people, especially the first-timers, struggle to know various procedures and processes required to get their work done at the bank and avail of its different versed with technology or in some cases where the information is too scattered to search for simply. There are, many types of platforms provided by different banks but users are facing problems accessing them (different User Interface much navigation). Although User Care centers are available, there are lot of weight times and redirection in some cases, leaving the user with no choice but to experience considerable delays getting a simple informational query resolved. People have queries about many bank policies, ATM, fixed deposits. This results in unnecessary crowd in banks for inquiry. Banks also face problems solving repeated queries of users. This is time consuming and banking staffs gets frustrated. Human work load and wealth gets wasted for separate inquiry counter unnecessarily.

### A. Basics of Chat Bot

A chat bot is a conversational agent that interacts with clients in a certain field on certain topic with NLTK sentences. commonly a chat bot reply by a user asking a question or initiating a new topic. These are the software with artificial intelligence which allows them to understand users query and deliver meaningful answer using predefined knowledge base.

### B. Chat Bot for Banks

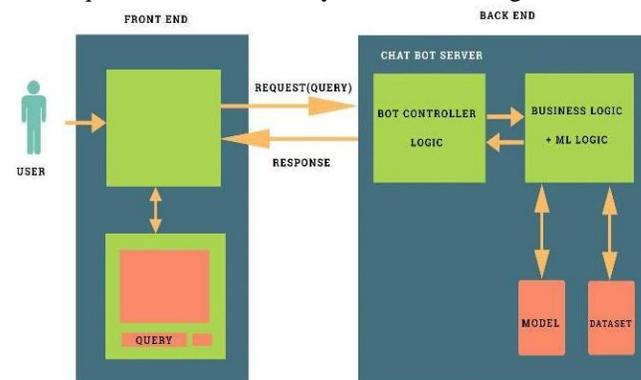
Developing a chat bot could provide a smart solution to solve these bank queries, provide details as and when required, improve service and increase number of users. It removes human factors included in organization and can give 24/7

hours service to increase productivity. We intend to provide a chat bot interface for users which could be available on the web and on any hand-held devices. Users can mention their queries in natural language and the chat bot can respond to them with correct answer. Proposed chat bot application is simply reachable to user thereby solving redundant queries anywhere anytime. As there will be fast response for inquiry, this will be time saving for both bank and users. The planned system would be a stepping stone in having to place an intelligent query handling software which could in next stages not just respond but self-learn to develop itself thereby increasing not just the quality of user service but also reducing human work load, increase in productivity and of course increasing number of fulfilled users.

## II. ARCHITECTURE

### A. Bank Chat Bot

End user will access with the system using web application. He will enter his query in text box provided on front end of this web application. Once he press Enter button or submit the doubt, this request will be handled by bot controller logic.

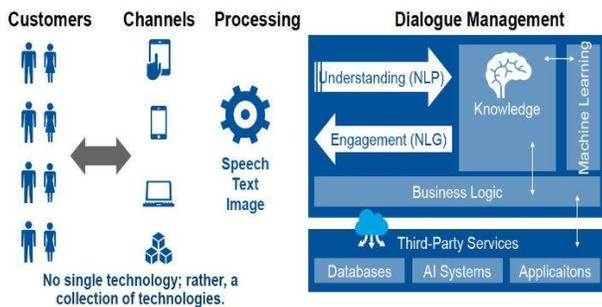


The bot controller logic which contains the implementation of Flask framework, which has been used for handling user requests and then to sending answers to those queries as a reply response. Then, the query will be forward to the implemented Business logic and Machine learning logic. The Business logic contains basic and advanced pre-processing techniques of the user input query using Natural Language Processing (NLTK library) and its vectorization method. NLP will tokenize the query, remove unnecessary spaces, stop-words and then extract lemmas for each token. Then this text-format query will be converted to vectorized format using vectorization. Now, using the Machine Learning logic, the classification algorithm will be applied to this reformed query to find the class in which it belongs. Classification algorithm will be applied based on the previous saved model executed on train data. All the requested queries from the user input data which having class equal to the retrieved class will be fetched and the cosine similarity will be applied to these class. According to the respected similarity values we get, the most similar answer will be returned to the user as a response to the query.

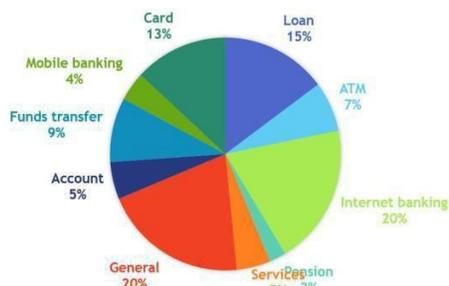
### III. IMPLEMENTATION

**1) Preparing Data Set:** We have started to prepare our own data set as questions and answers that banking customer's used to ask the bank staffs, at customer care centers or enquiry desks. In this we have referred a number of banking websites and collected FAQs as our data. We have used different web scrapping tools for this task. The following diagram shows the Distribution of questions in the Data-set format.

**Data-set format:** The Queries that customers requested were entered, the entered queries will get the approximate desired answer from the model by using Natural language



Processing whenever user enters the same query next time he will get the correct answer. by using the Machine learning the chat bot will improve its accuracy and gain knowledge from the Dataset.



**2) Web Scrapping:** To get the dataset we used a python

library called Beautiful Soup. By using this library web pages can be downloaded using get requests method. Then the downloaded page passed to the soup method. Soup element can be used to access the elements of the webpage. Web scraping that will automatically then extracts the data and in an easy format that you can make the matters easily.

**3) Pre-processing:** By using the NLTK library, the library which has been used for the purpose of Natural Language Processing. On because of the user input, will be in English, we have to let the machine to understand the query language that we used for the Natural Language Processing. then To decrease further processing and also to removing the ambiguity caused due to use of same word of different forms, we are using this pre-processing technique.

**4) Those steps included in this task are:**

Removing punctuation marks and extra spaces

**Tokenization** - The method Tokenization that we have used in the process is used to generate a sequence of words from user's input query. Removing stop words - Most of the common words like 'want', 'are', 'can', which we don't need to be considered while processing are re-moved for improving the performance of system.

**Lemmatization** - By using the Word Net Lemmatizer method for getting all the lemma (which means root form of the word) of the each token. e.g. 'processing' and 'process' should be considered as equal while processing. So for getting 'process' from 'processing', the method lemmatization is used.

**5)Vectorization:** On using the Bag Of Words (BOG) We have converted our text data to vectorized format concept. BOG is a method, which has been used for preparing text for input to our machine learning algorithm. This BOG model develops a vocabulary from all of the documents which has been used and then model each document by counting number of time

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

**6).Classification:** When the size of the data set getting increases, then it takes more time to find the similarity between user's query and the questions from large data set and return the answer. So we have used classification to boost the efficiency by lowering the response time need to get the answer. We have used "Scikit-learn" library from python for implementing these classifiers. Scikit - learn is a tool for data mining and machine learning in Python. As a part of article survey and initial

training we have selected following subset of classifiers to choose the best performing one as the final classifier for the chat bot.

- Decision Tree classifier
- Bernoulli Naive Bayes Classifier
- Gaussian Naive Bayes Classifier
- K-nearest neighbor classifier
- Multinomial Naive Bayes classifier
- Random Forest classifier
- Support vector machine

Also, for enhance the algorithm’s performance by using our data set, we have developed parameter optimization. There are two approaches for implementing parameter optimization -

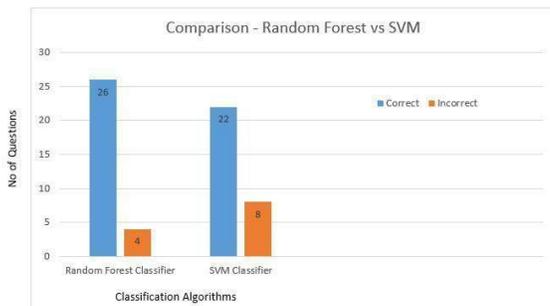
**i. Grid search** - Grid search is simply exhaustive searching method. In grid search, “it is needed to manually specify subset of hyper-parameter space of a learning model”. Hyper-parameters are not directly learned within estimators.

**ii. Randomized Search** – “As grid search is exhaustive and therefore expensive”. So, in randomized search, it samples parameter settings a static number of times that are more effective. We have used Randomized Search tactic in our case.

**7) Develop learning model:** In this stage, we have joined NLTK, Vectorization and Machine learning algorithms all together and save this model for future use. So, whenever the new query comes to model, we will just fetch this saved system, test this query on that system and get its class. In this way, we don’t need to train new model every time for each new query, thereby lowering the treating time.

**8) Testing model:** Testing for cross-validation score and accuracy and recall score of each classification algorithm, so that we can choose best for final use. Following is the table holding scores of each algorithm

**9) Selecting best approach:** According to scores of above table, 2 most correct algorithms are - Random Forest classifier and Support Vector Machine classifier.



**10) “Enquiry mapping” and receiving reply (Using Cosine similarity) :** After classifier provides us the class, we extract all queries that have this class from our data set. We check for cosine likeness of user’s query with these extracted queries. Then answer of the most similar question is

chosen as response to user’s query and is reverted to him. This bot is closed domain i.e. restricted to bank only. We have set a threshold on values of cosine similarity measure for handling queries that are out of domain.

#### IV. EXPERIMENTATION RESULT

##### A. Experiment(i)

“We have tried many queries to approve the execution of Bank Chat Bot”. In this experiment, we have entered questions which are alike to the questions existing in our data set. The analysis of the result is shown below:

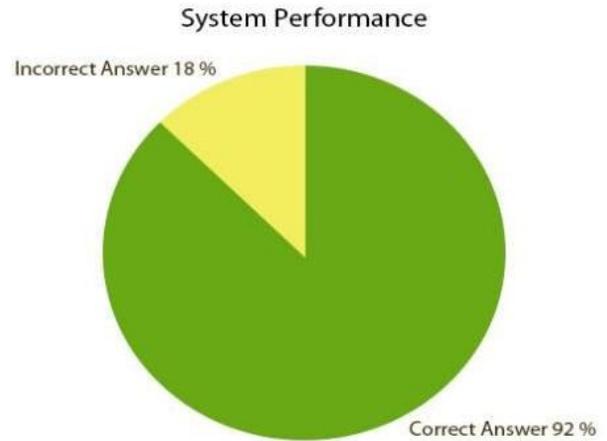


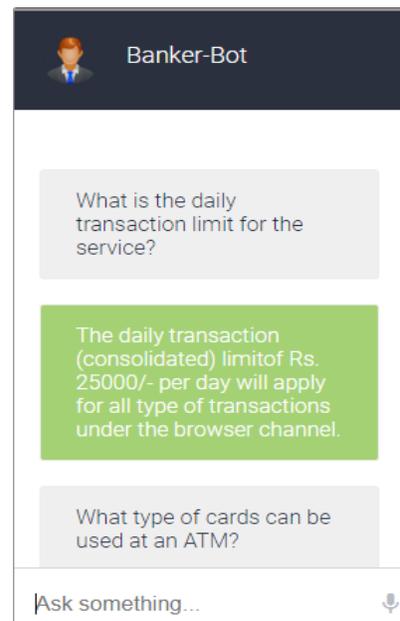
Figure.1. Gate-way

##### B. Experiment (ii)

In this research, we have tried similar query in altered forms. e.g. Query to open an account in bank can be asked in changed ways like :

- 1) steps for opening account
- 2) process for opening account
- 3) how to open an account
- 4) i want to open account
- 5) opening account
- 6) procedure for opening an account in bank

Even if we enter same question in similar and different forms, hat bot can reply same answer exactly to each form.



#### IV. CONCLUSION

The future system would be a stepping stone in consuming in place an intelligent question handling program which could in next stages not just respond but self-learn to improve itself thereby increasing not just the quality of user service but also reducing human load, increase in productivity and of course increasing number of satisfied users.

#### V. FUTURE SCOPE

1. Broadening the domain
2. Intelligent responses constructed by joining not just the existing list of FAQs but also from numerous other sources like internet, databases and other sources of data
3. Providing close suggestions
4. Intelligent demonstration of response images, links
5. Merging semantic similarity along with cosine similarity
6. Presentation account related info using Bank's

#### VI. REFERENCES

[1]. Automatic Extraction of Chatbot Training Data from Natural Dialogue Corpora, Bayan AbuShawar, Eric Atwell

[2]. Chatbot Evaluation and Database Expansion via Crowdsourcing, Zhou Yu, Ziyu Xu, Alan W Black, Alexander I. Rudnicky

[3]. AI BASED CHATBOT, Prof.Nikita Hatwar, Ashwini Patil, Diksha Gondane

[4] Data Mining: Concepts and Techniques Jiawei Han and Micheline Kamber

[5]. H. Kopka and P. W. Daly, Data Mining Practical Machine Learning Tools and Techniques Ian H. Witten Eibe Frank Mark A. Hall

[6]. <https://www.cse.iitb.ac.in/~bibek/WriteUP2016.pdf> <http://stanford.edu> [7] <https://scikit-learn.org>

[8]. <http://www.nltk.org>

[9]. <http://www.wikipedia.org>

[10] <https://chatbotsmagazine.com/the-complete-beginners-guide-to-chatbots-8280b7b906ca12zgql2op>

[11] <https://www.quora.com/What-is-the-best-way-to-learn-and-write-a-AI-Chat-bot>

[12]. <https://chatbotlife.com/ultimate-guide-to-leveraging-nlp-machine-learning-for-your-chatbot-531ff2dd870c5mcveo57b>

[13] <https://apps.worldwritable.com/tutorials/chatbot/>

[14] <http://machinelearningmastery.com/>

[15] <http://flask.pocoo.org/>

[16] [pandas.pydata.org](https://pandas.pydata.org)

[17] <https://www.pandorabots.com/>

[18]. <http://www.wildml.com/2016/04/deep-learning-for-chatbots-part-1-introduction/>

[19]. <https://stanfy.com/blog/advanced-natural-language-processing-tools-for-bot-makers/>