

ChatBot using Machine Learning and AI

Pushkar Sathe¹, A J Haravindan², Akshayprasad Arunkumar³, Sakshi S Gupte⁴, Mrinalini Murugan⁵

¹Pushkar Sathe, Assistant Professor, EXTC Department, S.I.E.S.G.S.T, Nerul

²A J Haravindan, Student, EXTC Department, S.I.E.S.G.S.T, Nerul

³Akshayprasad Arunkumar, Student, EXTC Department, S.I.E.S.G.S.T, Nerul

⁴Sakshi S Gupte, Student, EXTC Department, S.I.E.S.G.S.T, Nerul

⁵Mrinalini Murugan, Student, EXTC Department, S.I.E.S.G.S.T, Nerul

Abstract -Banks are places where people stand in line for hours and still don't manage to pull off a constructive outcome of their purpose of visit to the bank. Also, the loan approval system of the modernized banks are slow now-a-days. Thus, there arises a need of an automated system which can be used to predict if a particular user is eligible to get a loan or not based on some factors like single/married, no of children, annual income, annual expenditure, income of spouse if necessary, etc. Sometimes the basic FAQ's of a particular bank are not clear to the account holders. A basic work of money deposit or money withdrawal or application for a loan could become a tedious job without the right knowledge of working principles of the bank. Thus, it becomes a very important to help the account bearer to understand the principles and abide to the limits of the bank. This is where a FAQ ChatBot comes in handy. The account bearer now can get his doubts cleared and come prepared for their work so that their time could be saved and the energy of the manpower of the bank could be saved.

The loan prediction system will use **Machine Learning Algorithm** to predict whether the customer is eligible to get the loan or not.

The ChatBot will use **Natural Language Processing (NLP)** and various techniques in NLP to answer the questions of the account bearer and the responsive nature of the ChatBot will help user to understand better.

Keywords—Loan Prediction System, ChatBot, Machine Learning Algorithm, Natural Language Processing.

1. INTRODUCTION

A ChatBot is an artificial intelligence (AI) software that can simulate a conversation (or a chat) with a user in natural language through messaging applications, websites, and mobile apps or through the telephone. Why are ChatBots important? A ChatBot is often described as one of the most advanced and promising expressions of interaction between humans and machines. However, from a technological point of view, a ChatBot only represents the natural evolution of a Question Answering system leveraging Natural Language Processing (NLP).

Formulating responses to questions in natural language is one of the most typical Examples of Natural Language Processing applied in various enterprises' end-use applications.

A loan prediction system uses Machine Learning Algorithm (Classification) to predict if a particular user is eligible for a loan or not. Classification is used as there are only two possible classes of outputs (that is "YES" or "NO"). Various loss functions are calculated and global minima of error is found out using Adam optimizer. The model uses **Random Forest Classifier** from the **SciKit-Learn** library of Python.

The complete project has two stages of development.

1. Loan Prediction System
2. FAQ ChatBot.

For the sake of conceptualizing the steps to be taken to predict if an individual is eligible to get a loan or not, various parameters are to be identified over which the

decisions are made. Also, the influence of one parameter over the other should also be kept as a parameter. Machine learning models reiterate and learn from examples that were pondered upon by humans. Thus, the machine learning algorithm more or less thinks in the same way as humans think and the decisions are pretty close and the margin of error is very less.

We make use of this property of **Artificial Intelligence** and **Machine Learning** to make our model learn from previous examples where the decisions were made by humans and every parameter was reiterated before arriving at the conclusion. Thus, the margin of error for error is very less in the model. The same task could be done using various machine learning models and **Deep Learning** could also be used to bring in the concept of artificial neurons and make the model learn as a human brain learns.

2. Need of Project

In today's generation with the increase of mobile technology and resources we can easily obtain information, authenticate it and perform transactions easily without need of human resource. But the only missing factor occurred with the help of machines is the interaction between the human and the machines like in human to human interaction where a personal conversation occurs. Hence, we need to create a platform where we can improve the interaction between the man and the machine. This is where the evolution of ChatBot took place. ChatBot is widely used by most of the industries to bring out more advance in technology by creating a bot which solves most of the queries and problems. These ChatBots not only helps us save our time and money but also is trained in such a way that it gives us more accurate analysis and results than the human resource. So using this technology we are creating a loan prediction ChatBot which will be used by the banks to enable the customers of the bank to save their time just by clicking few buttons and getting the information they needed immediately. The main aim of loan prediction system is to replace the traditional method that is human resource used for approving loan. There are various parameters on basis of which loan approval is predicted. To take all parameters into consideration, it is difficult for a human to decide whether loan should be granted or not. This can be easily done by our loan prediction system. FAQ ChatBot solves all the bank related doubts of the user. It is one to

one conversation ChatBot. The Loan Prediction System and FAQ ChatBot can be useful for the society as you can get to know whether u can get loan or not and can also solve any banking related doubts without visiting the bank.

3. Literature Survey

There are loan prediction systems and few FAQ ChatBots present today. But the domain and method of current systems differs from ours. There are various FAQ ChatBots in working on the domains such as food delivery, online shopping, share markets, etc. FAQ ChatBot and Loan prediction system is rarely found. The loan prediction system is fully automated system and it considers all factors to predict loan unlike through direct bank access which may sometimes look over just one strong factor and approve loan to the individual. This will eliminate the work load on bank to make the right choice on to whom the loan should be approved. There are few research papers talking about loan prediction system and NLP which is the base of FAQ ChatBot.

In paper [1], researchers analyse dataset for loan prediction based on nature of the client. The research paper tells that whenever the bank makes decision to give loan to any customer then it automatically exposes itself to several financial risks. It is necessary for the bank to be aware of the clients applying for the loan. Knowing the client means getting to know his criminal history, mental ailment, annual income, purpose of loan, property he/she owns, etc. This problem motivates to do an EDA on the given data set and thus analysing the nature of the customer. The data set using EDA undergoes the process of normalization, missing value treatment, choosing essential columns using filtering, deriving new columns, identifying the target variables and visualizing data in graphical form. Python is a very efficient way of processing the data. This paper used the panda's library available in python in order to process and extract information from the data set. For better visualization of the results, processed data is converted into graphs. This is done by Matplotlib library.

In paper [2], machine learning classifier is used. It comprises of supervised learning for known datasets. Semi supervised learning and unsupervised learning for unknown datasets. Machine learning uses the logistic regression, decision tree support vector machine, neural

network as different classifiers. In this, classifier is used as the combination of K nearest neighbour (K-NN) classifier and Min-Max normalization is used to predict the loan applicant as a valid customer or default customer. K-NN is used to predict the credit score of the customer and min-max normalization is used to normalize, sample and compress the data to achieve linear transformation. Using these two techniques the data is split into 70% training data and 30% testing data sets before doing this, first the data is normalized. This results in 75.08% accuracy.

Paper [3] explains the various Machine Learning methods of making loan prediction system. The methods in this paper includes Decision tree, Random forest, Support Vector Machine, Adaboost (Adaptive boosting), Neural Networks, Linear Models. The paper also talks about types of parameters on the basis of which loan approval is predicted.

Our Loan prediction system is inspired from this paper and we made 2 models. One by using Neural networks and other one by using Random forest. The accuracy got by using Neural Network was 75% and accuracy got by using Random Forest is 85%.

In paper [4], ChatBot is made using Artificial Intelligence (AI) for the users to gather information about the loans and insurance schemes provided by the government aided banks such as RBI,SBI,NABARD etc. This ChatBot used machine learning with python and also used Recurrent Neural Network(RNN) along with Neural Machine Translation(NMT).These concepts have been used to get accurate results which satisfies the required queries of the users.First dataset is created to test and train it so that our bot starts learning from itself and improve the quality of the results. Two text files are created one which contains the questions asked frequently and another text file which contains all the possible answers corresponding to the questions. RNN and NMT methods are used to take a sentence question as input, process it and find a target sentence answer which is matching the question asked and this way the ChatBot starts to learn the deep meaning itself and give corresponding result in Graphical User Interface (GUI).GUI is the interface between the user and the backend which is not visible to the user. Backend plays an important role in processing the data and provide result.

In paper [5],the analysis of the Indian Bank ChatBots that is on how much the technology has reached in India

in terms of online banking sector. The different ChatBots of various banks are SIA by SBI bank, EVA by HDFC bank, iPal by ICICI bank, YES ROBOT by YES bank, Keya by Kotak bank, Aha by Axis bank, Indusassist by IndusInd bank and many more banks have ChatBots which solves a lot of queries of users and works 24/7 and hence bank employees have reduced queries at their desk. Now the banks have started to work on improvising the ChatBots by including voice, video and text recognition in all possible languages so that users can easily access the bots. Still many customers are unaware of the ChatBots provided by the banks and hence measures are to be taken to improvise this situation as India is a place with vast diversity in population. The ChatBots are in process of improvisation of their security standards.

4. Proposed System

1) Loan Prediction System

The developed code makes use of Random Forest classifier. Random Forest or Random Decision Forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode if the classes (classification) or mean prediction (regression) of the individual tree. Random decision forests correct for decision trees' habit of over fitting to their training set.

The beginning stages of the project includes finding an apt data set, as none of the banks will be willing to give their data for a research and development purpose as the data may be misused if it reaches in the wrong hands.

The second step involves cleaning the data. This step is very important to understand the aesthetics of the data and to see whether all the parameters of each and every input vector is present or not.

Working of Random Forest Classifier

Suppose there is matrix of Training Data set S that we are going to give to the algorithm to create a classification model.

$$S = \begin{bmatrix} f_{A1} & f_{B1} & f_{C1} & C_1 \\ \vdots & & \vdots & \vdots \\ f_{AN} & f_{BN} & f_{CN} & C_N \end{bmatrix}$$

Here we are having 3 features A, B, and C and there are N samples for each feature. The last column contains the Training Class. So, the aim is to create a random forest to classify the sample set. We create the sample set and from the sample set a lot of subsets are created with random values.

$$S_1 = \begin{bmatrix} f_{A12} & f_{B12} & f_{C12} & C_{12} \\ f_{A15} & f_{B15} & f_{C15} & C_{15} \\ \vdots & & \vdots & \\ f_{A35} & f_{B35} & f_{C35} & C_{35} \end{bmatrix} \quad S_2 = \begin{bmatrix} f_{A2} & f_{B2} & f_{C2} & C_2 \\ f_{A6} & f_{B6} & f_{C6} & C_6 \\ \vdots & & \vdots & \\ f_{A20} & f_{B20} & f_{C20} & C_{20} \end{bmatrix}$$

$$S_M = \begin{bmatrix} f_{A4} & f_{B4} & f_{C4} & C_4 \\ f_{A9} & f_{B9} & f_{C9} & C_9 \\ \vdots & & \vdots & \\ f_{A12} & f_{B12} & f_{C12} & C_{12} \end{bmatrix}$$

For example, in S1, randomly line number 12, 15, 35 and some other rows are used to create the subset from the sample set. Similarly, S2 and SM are created randomly. Each subset matrix gives rise to a decision tree. This is why we call it a forest because we have a lot of decision trees. Then with all these decision tree we have different variations of the main classification. Then using all these classifications, we are going to create ranking of the classifiers.

Prediction of Class for a new element

If we have a new element to classify, we are going to ask to each tree about the prediction of class. Let us suppose we have 4 trees and we are going to predict the class of the new element. The number of votes for each class given by the decision trees is calculated and the class which is having the highest votes is selected as the class of the classification. Creation of a lot of decision tress belongs to the algorithm of decision tree. So, this is the idea behind Random Forest Algorithm.

2) FAQ ChatBot

Natural Language Processing for ChatBot

The later part of the project deals with natural language processing (NLP). This is the method of converting the given text to vectors using various techniques such as tokenization, stemming, lemmatizing, vectorization etc. Nltk(Natural Language ToolKit) module of python was used for the same.

Data set was in json format as for the ChatBot to differentiate particular type of question, each question required a unique tag.

Keras on Tensorflow module was used for fast experimentation of deep neural network. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation function, optimizers and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code.

For hosting the ChatBot, Flask web framework was used. Flask is a micro-framework written in python. It is called as micro-framework as it doesn't require a particular tool or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Extensions are updated far more frequently than the core Flask program.

5. Flowchart

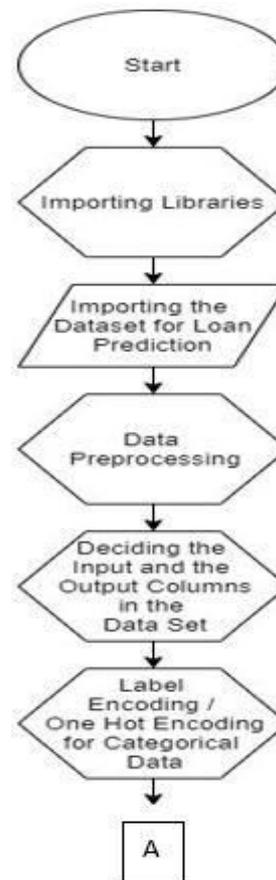


Fig 1: Loan Prediction Flowchart

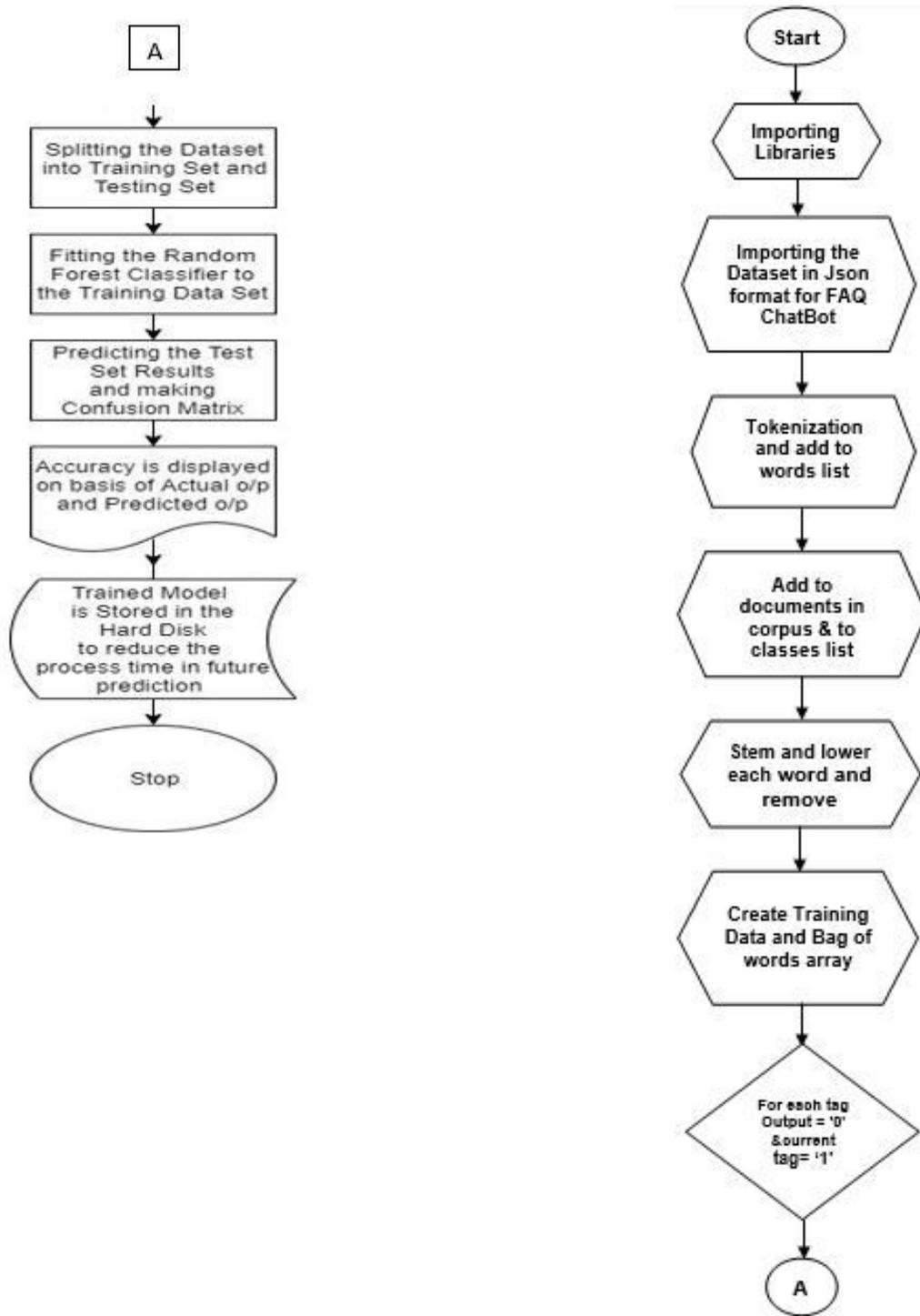
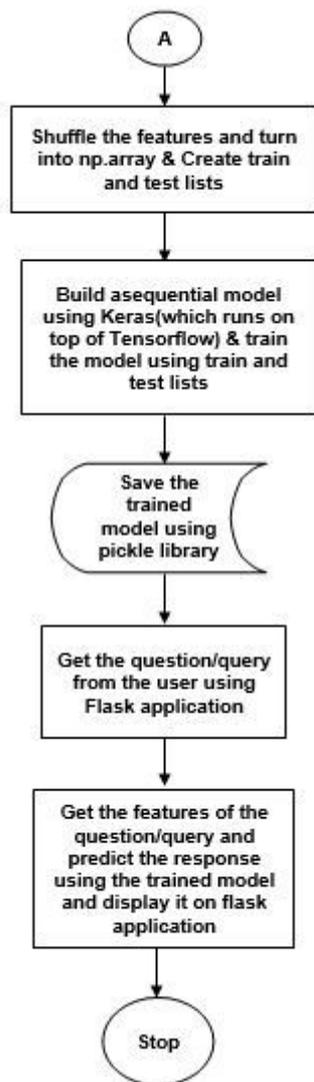


Fig 2: FAQ ChatBot Flowchart



	A	B	C	D	E	F	G	H	I	J	K	L
1	Gender	Married	Dependents	Education	Self_Employ	ApplicantInc	CoapplicantInc	LoanAmount	Loan_Amount	Credit_Histo	Property_Ari	Loan_Status
2	Male	No	0	Graduate	No	5849	0	360	1	Urban	Y	
3	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
4	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
5	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	Y
6	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
7	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
8	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	Y
9	Male	Yes	3	Graduate	No	3036	2504	158	360	0	Semiurban	N
10	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
11	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
12	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
13	Male	Yes	2	Graduate	No	2500	1840	109	360	1	Urban	Y
14	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
15	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
16	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
17	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y
18	Male	No	1	Not Graduate	No	3596	0	100	240		Urban	Y
19	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	N
20	Male	Yes	0	Not Graduate	No	4887	0	133	360	1	Rural	N
21	Male	Yes	0	Graduate	No	2600	3500	115	360	1	Urban	Y
22	Male	Yes	0	Not Graduate	No	7660	0	104	360	0	Urban	N
23	Male	Yes	1	Graduate	No	5955	5625	315	360	1	Urban	Y

Fig 4: Loan Prediction Data Set

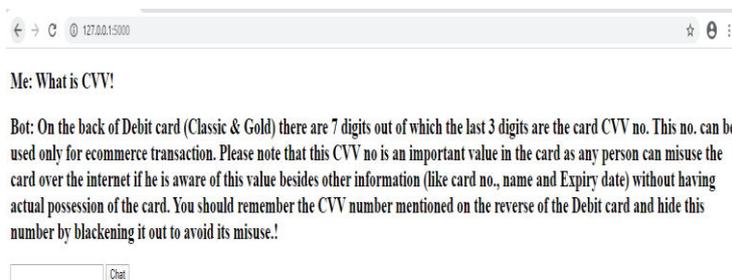
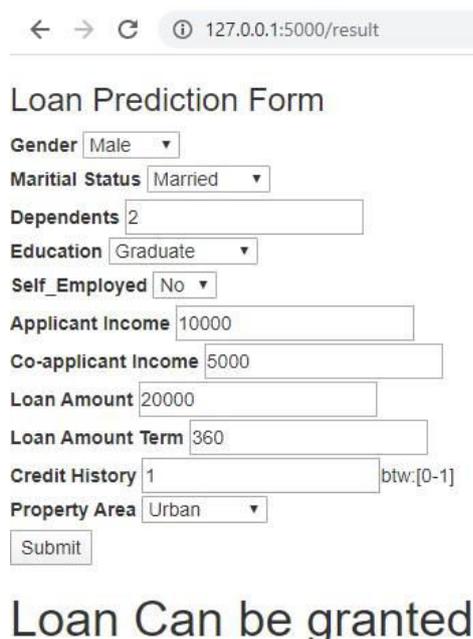


Fig 5: FAQ ChatBot output

7. Results



Loan Prediction Form

Gender: Male

Marital Status: Married

Dependents: 2

Education: Graduate

Self_Employed: No

Applicant Income: 10000

Co-applicant Income: 5000

Loan Amount: 20000

Loan Amount Term: 360

Credit History: 1 (btw:[0-1])

Property Area: Urban

Submit

Loan Can be granted

Fig 3: Loan Prediction Output

8. Conclusion

The project is completed successfully in two phases as our project has 2 parts i.e Loan Prediction System and FAQ ChatBot. The Loan Prediction System was done by making use of various machine learning algorithms to predict if a particular person is eligible for a loan or not based on the parameters mentioned above. The loan prediction system uses these parameters, converts all the non-numerical data to numerical format and learns on the examples provided. The model makes prediction with an accuracy of 85%(approx).

The development of FAQ ChatBot is successfully completed with the working of ChatBot in accordance to the data procured. The FAQ ChatBot makes use of Natural Language Processing(NLP) where all the words are brought to their root form and then vectorized. The vectors were then used to train the neural network model made with the help of Keras which runs on Tensorflow backend to make the bot successfully

answer to all the questions procured in the dataset. The bot may not be able to answer the questions if asked with a completely different figure of speech. Else the model works efficiently well and works with 100% accuracy on the dataset.

REFERENCES

1. X. Fransis Jency, V.P Sumathi, Janani Shiva Sri, An Exploratory Data Analytics for Loan Prediction Based on Nature of the Clients, International Journal of Recent Technology and Engineering, (IJRTE) ISSN:2277-3878, Volume-7 Issue-4S, November 2018.
2. G. Arutjothi, Dr. C. Senthamarai, Prediction of Loan Status in Commercial Bank using Machine Learning Classifier, Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017) IEEE Xplore Compliant – Part Number: CFP17M 19-ART, ISBN:978-1-5386-1959-9.
3. Kumar Arun, Garg Ishan, Kaur Sanmeet, Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN : 2278-0661, p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. I (May-Jun. 2016), PP 79-81.
4. Dr. Jaba Sheela L1 ,Safrin , Shanmugapriyaa , Sindhu Professor, Dept. of CSE, LOAN PAL – A CHATBOT TO PROFFERSPECIFICS ON LOAN SCHEMES Panimalar Engineering College, Chennai ,India.
5. NetraPal Singh, Devendra Singh, ChatBots and Virtual Assistants in Indian Banks, JEL: O3 DOI:10.5937/industrija47-24578 UDC: 004.738.5:336.71(540) 004.5:159.9.