**1. INTRODUCTION & MOTIVATION**

**1.1 INTRODUCTION**

In finance, a loan is the lending of money by one or more individuals, organizations, or other entities to other individuals, organizations etc. The recipient (i.e., the borrower) incurs a debt and is usually liable to pay interest on that debt until it is repaid as well as to repay the principal amount borrowed. The document evidencing the debt (e.g., a promissory note) will normally specify, among other things, the principal amount of money borrowed, the interest rate the lender is charging, and the date of repayment. A loan entails the reallocation of the subject asset for a period of time, between the lender and the borrower.

In India, the number of people applying for the loans gets increased for various reasons in recent years. The bank employees are not able to analyzed or predict whether the customer can payback the amount or not (good customer or bad customer) for the given interest rate. The aim of this project is to find the nature of the client applying for the personal loan. An exploratory data analysis technique is used to deal with this problem. The result of the analysis shows that short term loans are preferred by majority of the clients and the clients majorly apply loans for debt consolidation. The results are shown in graph that helps the bankers to understand the client’s behavior.

It is difficult for bank employees to collect the data manually. Here to make the process simple & easy. The bankers can get the data within few minutes and check whether the person is eligible or Not for loan application. Initially, the customer should apply for home loan after that company validates the customer eligibility for loan. However doing this manually takes a lot of time. Hence it wants to automate the loan eligibility process (real time) based on customer information so the final thing is to identify the factors/ customer segments that are eligible for taking loan. How will the bank benefit if we give the customer segments is the immediate question that arises? The solution is, banks would give loans to only those customers that are eligible so that they can be assured of getting the money back. Hence the more accurate we are in predicting the eligible customers the more beneficial it would be for the banks.

In our banking system, banks have many products to sell but main source of income of any bank is on its credit line. So they can earn from interest of those loans which they credits. A bank’s profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. Distribution of loan is the core business part of almost all banks. By predicting the loan defaulters, the bank can reduce its Non-Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison.[1]

Loan prediction is very helpful for employee of the banks as well as for the applicant also. It can also provide special advantages to bank. The loan prediction system can automatically calculate the weight of each feature taking part in loan processing and on new test data same features are processed with respect of their associated weight. Loan prediction system allows jumping to specific application so that it can be checked on priority basis.

Following very important approaches in predictive analytics are used to study the problem of predicting loan defaulters:

1. Logistic Regression Model

2. Decision Tree Model

3. Random Forest Model

The data is collected from the Kaggle for studying and prediction. The models have been performed and the different measures of performances are computed. The models are compared on the basis of the performance measures such as sensitivity and specificity. The final results have shown that the model produce different results.

An end to end web application has developed to predict the Loan output. The web application must be built with Node-red in JSON format with the machine learning trained models. It deployed on Watson Studio using machine learning service, jupyter notebook, Auto AI service, Cloud storage object.

* 1. **AIM & OBJECTIVE**

**AIM**

With enhancement in the banking sector, lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. The main aim of Loan prediction Model project is to reduce the risk factor behind selecting the safe person so as to save lots of bank efforts and assets. Therefore, by using different approaches, the right customers to be targeted for granting loan can be easily detected by evaluating their likelihood of default on loan. The models concludes that a bank should not only target the rich customers for granting loan but it should assess the other attributes of a customer as well which play a very important part in credit granting decisions and predicting the loan defaulters.

**OBJECTIVE**

**Predictive Analysis:**

The main objective of the project is to create a prediction model for loan approval and to classify that an applicant applying for loan is eligible or not or person is loan defaulter using data collected by bank employee.

**Risk Minimization:**

In present, The number of people applying for the loans are getting increased for various reasons. The bank employees are not able to analyze or predict whether the customer can payback the amount or not (good customer or bad customer) for the given interest rate. Hence risk of borrowers defaulting the loan can be minimized with the help of prediction model.

* 1. **EXISTING SYSTEM**

Loan prediction model of existing system is a powerful tool for a range of possible circumstances. Nevertheless there are limitations to it. The history of prediction model is long and there had been many obstacles in its evolution. Although constantly increasing variations and evolution in technology managed to overcome many obstacles, some of them still remain. The different prediction models use different approaches on dealing with these limitations and therefore their performance and reliability vary.

There are two important observations to make about existing Loan prediction model. For the prediction of loan approval, Logistic regression approach has been used. Preprocessing is major area of model where it consumes more time and then data analysis. Logistic regression is a type of statistical machine learning algorithm which is used to classify the data by considering outcome variables at extreme ends and tries to make algorithmic line that distinguishes between them. By this way the prediction can be made through logistic regression. And no other algorithm for prediction is used.

The existing model for the prediction is taken in account using the sigmoid function in logistic regression as the outcome is targeted binary either 0 or 1. The dataset of bank customers has been divided into training and test data sets. The train dataset contains approximately 600+ rows and 13+ columns whereas the test dataset contains 300+ rows and 12+ columns, the test dataset does not contain the target variable. Both the datasets are having missing values in their rows, and the mean, median or mode is used to fill the missing values but not removing the rows completely because the datasets are already small. Using the Feature Engineering techniques, the existing model is further proceeded and move towards the exploratory data analysis, where the dependent and independent variable is studied through statistics concepts such normal distribution, Probability density function etc. Study of the univariate, bivariate and multivariate analysis will give the view of the inside dependent and independent variable [1].

**1.4 LIMITATIONS ON EXISTING SYSTEM**

**1. Single Algorithm for predictive analysis is used:**

The existing system uses only logistic regression type of predictive analysis. Logistic regression approach is easy to implement and also effective but it can be easily outperformed by other algorithms. Hence there is need to implement more algorithms.

**2. Personal attributes required:**

The Model would be marginally better if it includes variables (personal attributes of customer like age, purpose, credit history, credit amount, credit duration, etc.) other than checking account information (which shows wealth of a customer) that should be taken into account to calculate the probability of default on loan correctly.

**3. Not updated:**

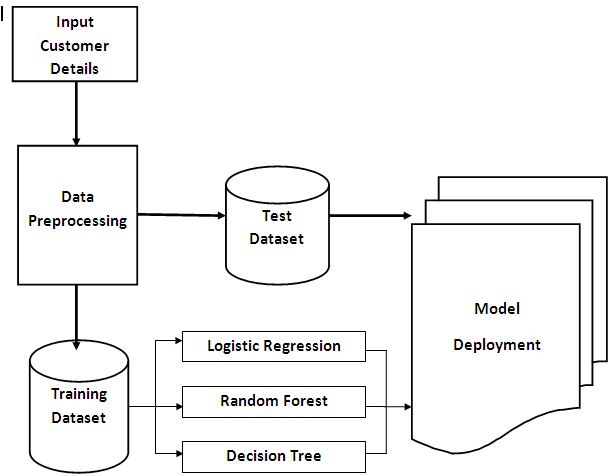
Time also plays a role in how well these techniques work. Though a model may be successful at one point in time, customer behavior changes with time and therefore a model must be updated. The financial crisis in 2020 due to COVID-19 pandemic exemplifies how crucial time consideration is.

**4. Human behavior:**

A Bank that wishes to utilize data-driven decision-making needs to have access to substantial relevant data from a range of activities, and sometimes big data sets are hard to come by. Even if a bank has sufficient data, critics argue that when anticipating human behavior; computers and algorithms fails to consider variables.

* 1. **PROPOSED SYSTEM**

Basically, the proposed system gives the predicted value a customer with details then the person is eligible or not for loan, by just taking some necessary details of the customer in real time, and those details will be collected by bank employee within minutes. Proposed system’s prediction model includes Decision tree model & Random forest model along with Logistic regression model present in existing system.



*Fig.1.5* ***Overview of proposed system***

Using a Decision Tree Regression has improved overall performance; It can be further improved by assembling more trees. Random Forest Regression trains randomly initialized trees with random subsets of data sampled from the training data; this will make the model more robust. Random forests are an ensemble model of machine learning with their roots in Decision Trees. These decision trees individually may over fit the data set and thus they come together to form a much stronger model. Numerous decision trees are first built and based on these by performing random sampling of attributes; a group of decision trees are assembled to form a Random Forest.

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees. After preparing the data, it can be fit in different models on the training data and compare their performance to choose the algorithm with good performance. As this is a regression problem, RMSE (Root Mean Square Error) can be used and $R²$ score as evaluation metrics. And also have created an UI using the Node-red software and Flask for the loan status prediction, this UI will allow the users to predict the loan status very easily and the User interface is user friendly not at least one complication in using the interface, and it can be used just by entering some necessary details into the UI in real time it'll give the predicted value like if the customer is beneficial to take a loan and how often does he pays the loan interest amount to the bank.

**Advantages:**

1. Using Machine learning to predict the loan status will produce less time and more accuracy in predicting the approximately close value can be done easily.
2. It is more trust worthy and cost effective .It also reduces the man power for doing the experiments to find the loan defaulters.

1. **PROJECT OVERVIEW**

**2.1 LITERATURE SURVEY**

The literature survey deals with the topics and the researches that would help to understand the existing systems that are similar to this project.

The objective of this literature survey is to analyze the related work to this project and mechanisms used in previous studies. The different Methods used to implement the similar project are as follows:

**Method 1**

**Paper Name:** “Loan Approval Prediction based on Machine Learning Approach”

**Authors Name:** Kumar Arun, Garg Ishan, Kaur Sanmeet

**Explanation:**

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

**Advantage**: Six machine learning classification models have been used for prediction

**Disadvantage:** The system is trained on old training dataset.

**Method 2**

**Paper Name:** “Loan Prediction by using Machine Learning Models”

**Authors Name:** Pidikiti Supriya, Myneedi pavani, Nagarapu Saisushma, K Vikas

**Explanation:**

The analytical process started from data cleaning and processing, Missing value imputation with mice package, then exploratory analysis and finally model building and evaluation. The best accuracy on public test set is 0.811. This brings some of the following insights about approval. Applicants with Credit history not passing fails to get approved, Probably because that they have a probability of a not paying back. Most of the Time, Applicants with high income sanctioning low amount is to more likely get approved which make sense, more likely to pay back their loans. Some basic characteristic gender and marital status seems not to be taken into consideration by the company.

**Advantage:** The best accuracy on dataset test is 0.811.

**Disadvantage:** In this model only one method used for prediction.

**Method 3**

**Paper Name:** “Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval”

**Authors Name:** Ashlesha Vaidya

**Explanation:**

Various machine learning models exist for predictive analysis like logistic regression, decision trees, artificial neural networks (ANN) and Bayesian Networks. This paper with the logistic regression model. While logistic regression is a statistical model the other three are graphical models. ANNs have a very complex structure with multiple layers of nodes. Logistic regression and ANNs are most widely used because they are easy to develop and provide most accurate predictive analysis. Logistic regression can handle non linear effect and power terms. The independent variable based on which prediction takes place need not be normally distributed.

**Advantage:** Logistic regression is widely used in data analytics where analyzing of the

Pre-existing data within all kinds of organization is required.

**Disadvantage:** Logistic regression requires a large sample for parameter estimation.

**Method 4**

**Paper Name:** “An Approach for Prediction of Loan Approval using

Machine Learning Algorithm”

**Authors Name:** Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar

**Explanation:**

The process of prediction starts from cleaning and processing of data, imputation of missing values, experimental analysis of data set and then model building to evaluation of model and testing on test data. On Data set, the best case accuracy obtained on the original data set is 0.811. The following conclusions are reached after analysis that those applicants whose credit score was worst will fail to get loan approval, due to a higher probability of not paying back the loan amount. Most of the time, those applicants who have high income and demands for lower amount of loan are more likely to get approved which makes sense, more likely to pay back their loans. Some other characteristic like gender and marital status seems not to be taken into consideration by the company.

**Advantage:** The advantage of logistic regression is that it is a predictive analysis.

**Disadvantage:** This method is not give accurate result on web application

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr No**  **.** | **Paper Name** | **Author/Publi cation** | **Technology** | **Advantages** | **Disadvant ages** |
| **1.** | An Approach for Prediction of Loan Approval using Machine Learning Algorithm | MohammadAhmad Sheikh/ (ICESC 2020) | Logistic regression,  SVM, KNN | The advantage of logistic regression is that it is a predictive analysis. | This method is not give accurate result on web application. |
| **2.** | Loan Approval Prediction based on Machine Learning Approach | Kumar Arun/ IOSR Journal of Computer Engineering IOSR-JCE(NCRTCSIT-2016) | Decision Trees,RandomForest,Support Vector Machine,LinearModel,Neural network | Six machine learning classification models have been used for prediction | The system is trained on old training dataset. |
| **3.** | Loan Prediction by using Machine Learning Models | PidikitiSupriya/ | Decision tree algorithm | The best accuracy on dataset test is 0.811. | In this model only one method used for prediction. |
| **4.** |  |  |  |  |  |
|  |  |  |  |  |  |
|  | Predictive and probabilistic approach using logistic regression :Application toprediction of loan approval. | Ashlesha Vaidya/ Computer Science Engineering SRM University, Chennai July 3-5, 2017 | logistic regression, decision trees,Artificialneural networks(ANN) and BayesianNetworks. | Logistic regression is widely used in data analytics where analyzing of the pre existing data within all kinds of organization is required. | Logistic regression requires a large sample for parameter estimation. |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  | . |

**2.1.1 Comparative Analysis:**

*Table.2.1.1:* ***comparative analysis***

* + 1. **MATHEMATICAL MODEL**

Basically, Three different machine learning approaches are implemented to predict the loan approval of customer. By comparing all three models, the maximum prediction accuracy can be achieved.

**1. Logistic Regression**

Logistic regression is a mathematical modeling approach used in describing the relationship of several independent variables to a dichotomous dependent variable or a limited dependent variable. The logistic function is employed because the dependent variable default is dichotomous, whereas the proposed covariates were mixture of continuous and categorical random variables. Thus the model was chosen over others due to the data structure and purpose. Also the independent variables need not be interval, normally distributed, nor linearly related, nor equal variance within each group. The logistic model is a derivative of the odds function. The odd of a function is the ratio of the probability of success to that of failure. Thus

(𝑌 = 1) =(𝑌 = 1/ 𝑋 = 𝑥)/𝑃(𝑌 = 0/ 𝑋 = 𝑥)

Where (𝑌 = 1)is the odds of default t; 𝑃(𝑌 = 1) is the probability that default occurs given a set of explanatory variables and 𝑃(𝑌 = 0) is the probability of non-default given set of explanatory variables. If the odds of default is greater than one it means there is a higher probability of default compared to that of non-default. A value less than one indicate a higher probability of non-default than that of default. Given the binary response variable (default or non-default), the probability distribution of the number of defaults in a given loan portfolio size, for given values of explanatory variables is binomial. Thus the probability that the number of default of a given portfolio size n is exactly equal to size x is given by

𝑃(𝑋 = 𝑥) =𝑛!𝑥! \* 𝑝𝑥𝑞𝑛−𝑥 /(𝑛 − 𝑥)!

Where 𝑞 = 𝑃 (0) =; probability of non-default). This means that given a portfolio size n and probability of default P(Y=1) from a financial institution, one can use the theorem to perform risk control analysis.

While selecting the algorithm that gives an accurate prediction we gone through lot of algorithms which gives the results abruptly accurate and from them we selected only one algorithm for the prediction problem that is Logistic regression and it is based on binomial probability theory. The algorithm is selected based on its accuracy. Accuracy is defined as the ratio of the number of samples correctly classified by the classifier to the total number of samples for a given test data set. The formula is as follows

Accuracy=TP+TN/TP+TN+FT+FN

Mathematical equation for sigmoid function used in logistic regression is

**S(x) = 1/(1+e-x)**

**Algorithm:**

Step 1: Data preprocessing

Step 2: Fitting Logistic Regression to our training set

*model import LogisticRegression*

*lr←LogisticRegression(random\_state=0)*

*lr.fit(X\_train, y\_train)*

Step 3: Testing the model

*pred ← lr.predict(X\_test) y\_pred*

Step 4:Measuring Accuracy

*fromsklearn.metrics import accuracy\_score*

*accuracy\_score(y\_test,y\_pred)*

Step 5:import sklearn.metrics as metrics

*fpr,tpr,threshold ← metrics.roc\_curve(y\_test, pred)*

*roc\_auc ← metrics.auc(fpr, tpr)*

*plt.title("Logistic Regression")*

*plt.plot(fpr,tpr,'b',label = 'auc = %0.2f'%roc\_auc)*

*plt.legend(loc = 'lower right')*

*plt.plot([0,1],[0,1],'r--')*

*plt.xlim([0,1])*

*plt.ylim([0,1])*

*plt.ylabel('tpr')*

*plt.xlabel('fpr')*

**2. Decision Tree**

In decision Tree (like CRT, QUAID) build classification model that learn decision rules inferred from data features to make predictions, generating a tree structure with decision nodes corresponding to attributes (input variables).

**Step 1**: use splitting criterion (like Information Gain, Gain Ratio, Gini Index) to select the attribute with the best score that will be chosen to produce the purest node regarding to the target variable (in our case, the attribute that best separates “Granted” from “Not granted”).

**Step 2**: create the root split node with the consequents subsets, then repeat step 1 for each subset by reusing splitting criterion to select the next best attribute to produce the purest sub-nodes regarding to the target variable.

**Step 3**: repeat step 2 until reaching a stopping Criteria, for instance: Purity of the node > pre-specified limit or Depth of the node > pre-specified limit or simply Predictor values for all records are identical (no more rule could be generated)

**Step 4**: apply Pruning to avoid overfitting by using a criterion to remove sections of the tree that provide little power to classify and determine the optimum tree size. To do so, we create distinct dataset “training set” and “validation set”, to evaluate the effect of pruning and use statistical test ( like Chi-square for CHAID) to estimate whether pruning or expanding a given node produce an improvement. We have two types of Pruning:

**Pre-pruning** stop growing the tree earlier, before it perfectly classifies the training set.

**Post-pruning** allow the tree to grow and then prune it back.

**Algorithm:**

Step 1: Data preprocessing

Step 2: Training the Decision Tree Classification with the help of Training set

*from sklearn.tree import DecisionTreeClassifier*

*dt ← DecisionTreeClassifier(criterion = 'entropy', random\_state = 0)*

*dtr.fit(X\_train, y\_train)*

Step 3: testing the prediction results

*pred ← dtr.predict(X\_test) y\_pred*

Step 4: Measuring Accuracy

*from sklearn import metrics*

print('The accuracy of Decision Tree Classifier is: *'metrics.accuracy\_score(y\_test, y\_pred))*

Step 5: import sklearn.metrics as metrics

*fpr,tpr,threshold ← metrics.roc\_curve(y\_test, y\_pred)*

*roc\_auc ← metrics.auc(fpr, tpr)*

*plt.title("Decision Tree Curve")*

*plt.plot(fpr,tpr,'b',label = 'auc = %0.2f'%roc\_auc)*

*plt.legend(loc = 'lower right')*

*plt.plot([0,1],[0,1],'r--')*

*plt.xlim([0,1])*

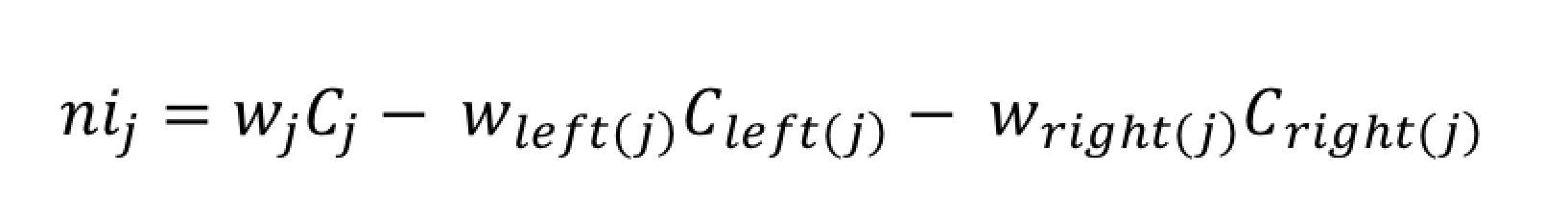
*plt.ylim([0,1])*

*plt.ylabel('tpr')*

*plt.xlabel('fpr')*

**3. Random Forest**

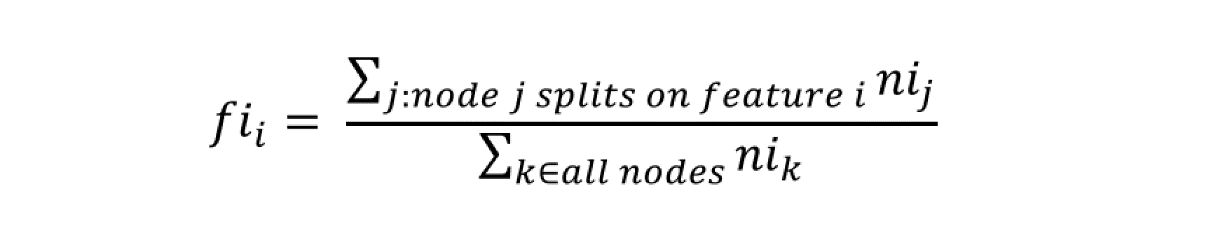
For each Random Forest, Scikit-learn calculates a nodes importance using Gini Importance, assuming only two child nodes (binary tree):

****

* ni sub(j)= the importance of node j
* w sub(j) = weighted number of samples reaching node j
* C sub(j)= the impurity value of node j
* left(j) = child node from left split on node j
* right(j) = child node from right split on node j

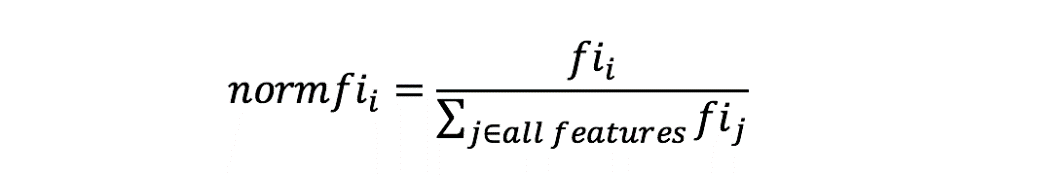
*sub() is being used as subscript isn’t available in Medium*

The importance for each feature on a decision tree is then calculated as:

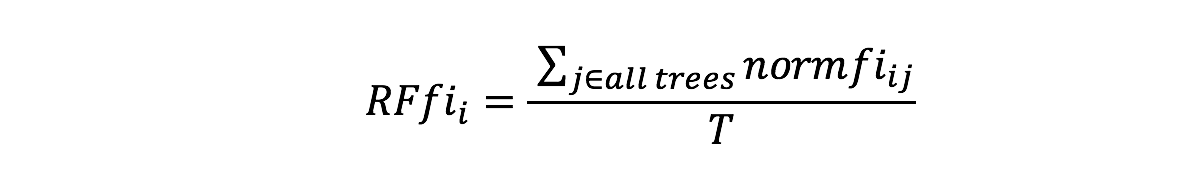


* fi sub(i)= the importance of feature i
* ni sub(j)= the importance of node j

These can then be normalized to a value between 0 and 1 by dividing by the sum of all feature importance values:



The final feature importance, at the Random Forest level, is it’s average over all the trees. The sum of the feature’s importance value on each trees is calculated and divided by the total number of trees:



* RFfi sub(i)= the importance of feature i calculated from all trees in the Random Forest model
* normfi sub(ij)= the normalized feature importance for i in tree j

T = total number of trees

**Algorithm**

Step 1: Data preprocessing

Step 2: Training the Random forest model with the help of Training set *from sklearn.ensemble import RandomForestClassifier*

*rf ← RandomForestClassifier(n\_estimators = 10, criterion = 'entropy', random\_state = 0)*

*rf.fit(X\_train, y\_train)*

Step 3:Predicting the Test set results

*pred ← rf.predict(X\_test) y\_pred*

Step 4: Measuring Accuracy

*metrics.accuracy\_score(pred, y\_test))*

Step 5:*import sklearn.metrics as metrics*

*fpr,tpr,threshold ← metrics.roc\_curve(y\_test, y\_pred)*

*roc\_auc ← metrics.auc(fpr, tpr)*

*plt.plot([0,1],[0,1],'r--')*

*plt.xlim([0,1])*

*plt.ylim([0,1])*

*plt.ylabel('tpr')*

*plt.xlabel('fpr')*

* 1. **PROBLEM STATEMENT**

Problem of existing is the less accuracy for predictive analysis. Only one prediction model is available to predict the loan approval. Also prediction credibility is a challenge not only in the field of prediction. All aspects of reality can never be implemented in prediction models and therefore compromises have to be made. Such compromises can be absence of certain personal attributes of applicants. A prediction model can be useful for banks if the probability is gives and result it delivers is applicable to real world facts. The prediction model developer can never be one-hundred percent sure if this is the case. Hence a certain level of trust is needed by banks when working with model.

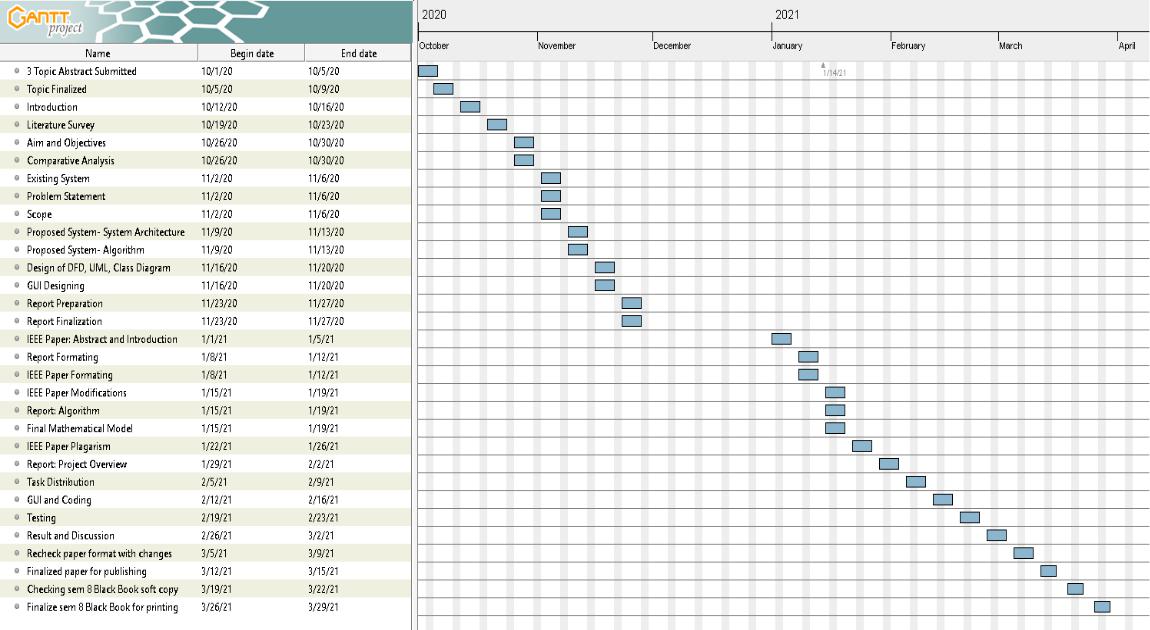
* 1. **SYSTEM OVERVIEW**

The Software is user-friendly and easy to use as the user is naive and hence the GUI is very less complex. The main purpose of this software is to provide prediction about loan approval i.e. the eligibility of applicant. Hence, for this purpose machine learning algorithms are required. The GUI is so simple that we can understand easily.

The scope of this research is to provide bank employee the prediction about whether the applicant is eligible for loan or not. The Project consists of three prediction models Logistic regression, Decision tree and Random forest. In order to achieve the high accuracy for prediction; all the models are initially trained with quantitative and qualitative dataset. Comparison is done between all three models to know, which one is giving the highest accuracy.

* 1. **PROJECT TIMELINE CHART**

To complete the project successfully, it must be needed to control or manage the number of activities, and ensure that they are completed on schedule provided by guide. If any task miss a deadline or finish out of sequence, there could be chances of knock on effects on the rest of project. Hence Gantt charts convey this information visually. They outline all of the tasks involved and their order, shown against a timescale. It gives you an instant overview of project, its associated tasks and when these needed to be finished. In following Gantt chart, the sequences of task with the deadline are given and also Gantt chart is present.



*Fig. 9.1* ***Gantt chart***

**2.5 TASK DISTRIBUTION**

**2.5.1 DESIGN PHASE**

**Table No 2.5.1*: Design Phase***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Name of Student** |  | **Task Performed** |  | **Result** | | |
|  | |  |  |  |  |  |  | |
|  | | **Rahul P. Bhere** |  | Design of the front End |  | Completed Successfully | | |
|  | |  |  |  |  |  |  | |
|  | | **Dhiraj R. Manje** |  | Suggested Appropriate contents |  | Completed Successfully | | |
|  | |  |  | for User interface |  |  |  | |
|  | |  |  |  |  |  |  | |
|  | | **Rahul K. Pawade** |  | Suggested orientation and |  | Completed Successfully | | |
|  | |  |  | appropriate looks of User |  |  | |  |
|  | |  |  | interface |  |  | |  |
|  | |  |  |  |  |  | |  |
| **2.5.2 IMPLEMENTATION PHASE** | | | | | |  | |  |
|  |  | |  | **Table No 2.5.2*: Implementation Phase*** | |  | |  |
|  |  | |  |  | |  | |  |
|  | **Name of Student** | |  | **Implementation Task** | | **Results** | |  |
|  |  | |  |  | |  | |  |
|  | **Rahul P. Bhere** | |  | Implemented the web appication of the project | | Completed Successfully | |  |
|  |  | |  |  | |  | |  |
|  | **Dhiraj R. Manje** | |  | Implemented project Algorithm | |  | |  |
|  |  | |  | Completed Successfully | |  |
|  |  | |  |  |  |  |
|  |  | |  |  | |  | |  |
|  | **Rahul K. Pawade** | |  | Implemented back end part of the project | | Completed Successfully | |  |
|  |  | |  |  | |  | |  |

1. **SOFTWARE REQUIREMENT SPECIFICATION**
   1. **HARDWARE REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| System | : | Pentium IV 2.4 GHz. |
| Hard Disk | : | 40 GB. |
| Floppy Drive | : | 1.44 Mb. |
| Monitor | : | 14’ Color Monitor. |
| Mouse | : | Optical Mouse. |
| Ram | : | 4 GB. |
| Keyboard | : | 101 Keyboard Keys. |

* 1. **SOFTWARE REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| Operating system | : | Windows 7 or 10. |
| Browser | : | Any Browser. For E.g. Google Chrome |
| Coding Language | : | JSON, Python 3.8. |
| Software’s used | : | Node-red, Jupyter Notebook, Watson Studio. |

1. **SYSTEM DESIGN**
   1. **DESIGN SPECIFICATION**

**4.1.1 ALGORITHM**

**Algorithm of prediction model using Logistic regression, Random forest, Decision tree**

**Step 1: Start**

Importing libraries;

*import numpy as np*

*import pandas as pd*

*import matplotlib.pyplot as plt*

**Step 2: loading dataset**

*data←pd.read\_csv("loan\_data.csv")*

**Step 3**: Taking care of null values or All null values removed

*data.apply(lambda x: sum(x.isnull()),axis=0)*

**Step 4: Data visualization**

*sns.pairplot(data)sns.heatmap(data.corr(), annot = True)*

*a←data["Gender"].value\_counts().to\_numpy()*

*b←data["Married"].value\_counts().to\_numpy()*

*c←data["Dependents"].value\_counts().to\_numpy()*

**Step 5: Analyzing the data**

*X ← data.iloc[:, 1: 11].values*

*y ← data.iloc[:, 11].values*

**Step 6: Label Encoding**

*From sklearn.preprocessing*

*import LabelEncoderle = LabelEncoder()*

*fori in range(0, 5):*

*X[:,i] ← le.fit\_transform(X[:,i])*

*X[:,9] ← le.fit\_transform(X[:,9])*

*y ← le.fit\_transform(y)*

OneHotEncoding *fromsklearn.preprocessing import OneHotEncoder*

*one ← OneHotEncoder()*

*z ← one.fit\_transform(X[:,9:11]).toarray()*

*X ← np.delete(X, 9, axis ← 1)*

*X ← np.concatenate((z,X), axis = 1)*

**Step 7: Splitting into train and test**

From sklearn.model\_selection

import train\_test\_split

*X\_train, X\_test, y\_train, y\_test ← train\_test\_split(X, y, test\_size = 1/3, random\_state =0)*

**Step 8: Feature Scaling**

from sklearn.preprocessing import StandardScaler

*sc ← StandardScaler()*

*X\_train ← sc.fit\_transform(X\_train)*

*X\_test ← sc.fit\_transform(X\_test)*

**Step 9: Use all of three machine learning models to train models:**

#Fitting all ML models to the Training set

*From sklearn.tree import DecisionTreeClassifier*

*From sklearn.linear\_model import LogisticRegression*

*from sklearn.ensemble import RandomForestClassifier*

*dt ← DecisionTreeClassifier(criterion ← 'entropy',random\_state ← 0)*

*lr←LogisticRegression(random\_state=0)*

*rf ← RandomForestClassifier(n\_estimators = 10, criterion = 'entropy', random\_state = 0)*

*dt.fit(X\_train, y\_train)*

*rf.fit(X\_train, y\_train)*

*lr.fit(X\_train, y\_train)*

# Predicting the Test set results

*y\_pred ← classifier.predict(X\_test)y\_pred*

# Measuring Accuracy

*fromsklearn import metrics*

*print('The accuracy of model is: ', metrics.accuracy\_score(y\_test, y\_pred))*

# Making confusion matrix

*fromsklearn.metrics import confusion\_matrixprint(confusion\_matrix(y\_test, y\_pred))*

**Step 9:Evaluation of models**

*fromsklearn.metrics import r2\_score, mean\_absolute\_error*

*foralgo,model in fit\_models.items():*

*yhat ← model.predict(X\_test)*

**Step 10: launch the web application.**

**Step 11: Enter the values for prediction.**

**Step 12: Result display you are eligible or not for loan approval prediction & as well as you get some tips regarding bank loan.**

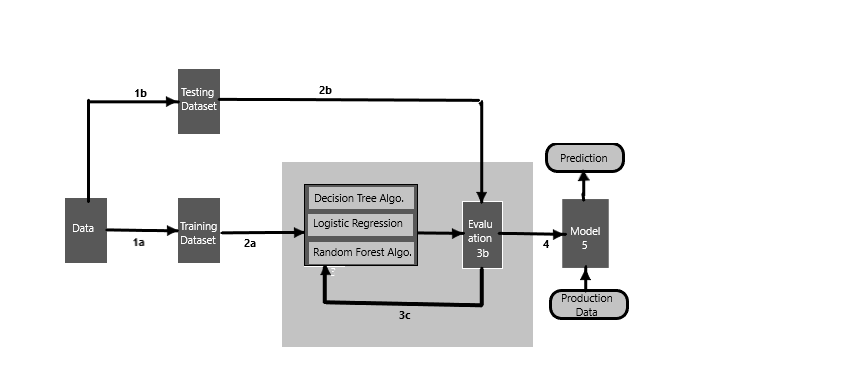
**Step 13: Stop**

* + 1. **WORKING OF ALGORITHM**

The general idea of working of algorithm is given as below:

* The algorithms imports the libraries going to use in program to run functions
* The loading of dataset is done by using function read\_csv( ) and named as data
* Null values present in dataset are removed by applying conditions.
* The categorical values are transferred into numerical values using OneHotEncoder.
* After the dataset will be divided into Training and testing dataset.
* Scaling of the dataset is done by StandardScaler( ).
* The acquired Training dataset will be used to train all of three Machine Learning models.
* All of the required models are imported from sklearn library.
* model.fit( ) function is used to train the models with the dataset.
* For testing purpose, some value string is given as input to get the prediction output.
* By using accuracy\_score( ) functions, the accuracy of each model is measured. The model having highest accuracy will give more accurate prediction.
* By using any web browser eg. Google Chrome, the web application will be launched.
* Now user can enter the required details for loan approval prediction.
* The application will provide prediction whether the applicant is eligible for loan or not.
  1. **SYSTEM ARCHITECTURE**

In present generation everyone have bank accounts, the rarely find anyone without an account. In lifetime peoples may at least take one loan from the bank for any of the need. Due to lot of transactions have been occurring every day huge data volumes are available which represent the customers behavior and the risks around loan are increased.

Data Mining is one of the most motivating and vital area of research with the aim of extracting information from tremendous amount of accumulated data sets. Here a new model for classifying loan risk in banking sector by using Machine Learning concepts. The model has been built using data from banking sector to predict the status of loans. Three algorithms have been used to build the proposed model: Random Forest, Logistic Regression and Decision Tree. By using the algorithm a Node-red and Flask model has been implemented and tested. The result has been discussed and a full comparison between algorithms was conducted [2].

*Fig 4.2* ***System architecture of proposed system***

**1] Data Collection**

The quantity & quality of your data dictate how accurate our model is The outcome of this step is generally a representation of data; which will be useed for training Using pre-collected data, by way of datasets from Kaggle, UCI, etc., still fits into this step.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Gender | Male/Female |
| Married | Applicant married(Y/N) |
| Dependents | No of dependents |
| Education | Applicant Education(Graduate/Under Graduate) |
| Self\_Employed | Self employed(Y/N) |
| Applicantincome | Applicant Income |
| LoanAmount | Loan amount |
| Loan\_Amount\_term | Term of loan in month |
| Credit\_History | Credit history meets guidelines |
| Property\_area | Urban/Semi urban/Rural |

***Table 4.2 Example of Dataset***

**2] Data Preprocessing**

Wrangle data and prepare it for training Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis Split into training and evaluation sets.

**3] Choose a Model**

Different algorithms are for different tasks; choose the right one. In this project Logistic regression, Decision tree, Random forest models are used.

**4] Training the Model**

The goal of training is to answer a question or make a prediction correctly as often as possible Linear regression example: algorithm would need to learn values for m (or W) and b (x is input, y is output)Each iteration of process is a training step.

**5] Evaluate the Model**

Uses some metric or combination of metrics to "measure" objective performance of model Test the model against previously unseen data This unseen data is meant to be somewhat representative of model performance in the real world, but still helps tune the model (as opposed to test data, which does not)Good train/evaluate split? 80/20, 70/30, or similar, depending on domain, data availability, dataset particulars, etc.

**6] Parameter Tuning**

This step refers to hyperparameter tuning, which is an "art form" as opposed to a science Tune model parameters for improved performance Simple model hyperparameters may include: number of training steps, learning rate, initialization values and distribution, etc.

**7] Make Predictions**

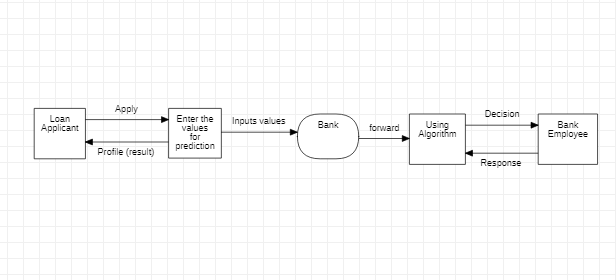
Using further (test set) data which have, until this point, been withheld from the model (and for which class labels are known), are used to test the model; a better approximation of how the model will perform in the real world.

**8] Display Prediction result (Web Application)**

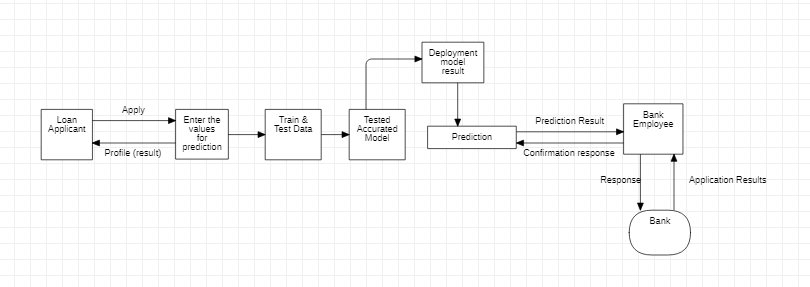
An end to end web application has developed to predict the Loan output. The web application must be built with Node-red in JSON or Flask (python) in HTML format with the machine learning trained models. It deployed on Watson Studio using machine learning service, jupyter notebook, Auto AI service, Cloud storage object

.

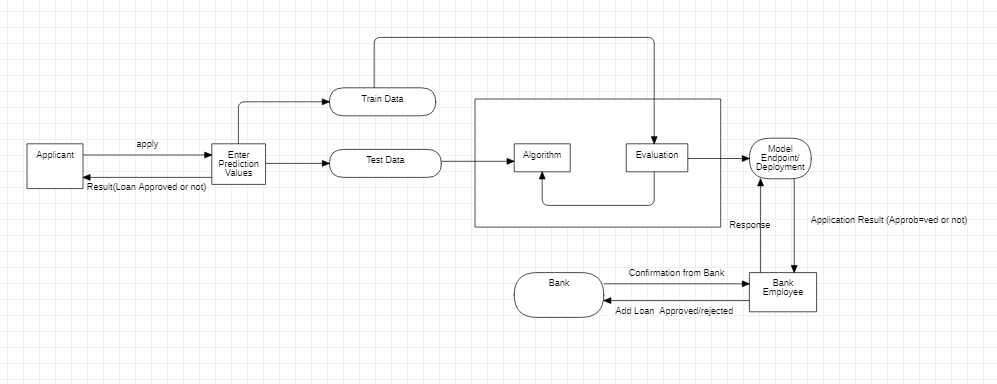
* 1. **DATA FLOW DIAGRAMS**



***Fig. 4.3.1 DFD level 0 diagram***

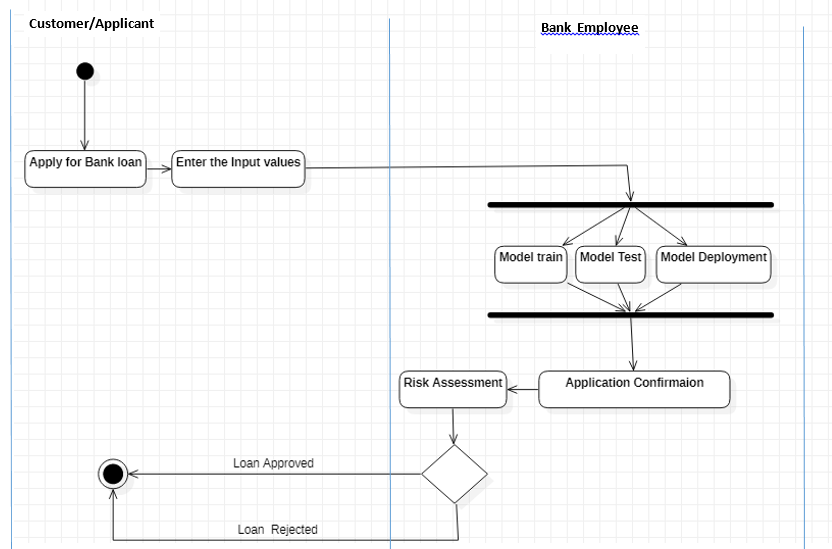


***Fig.4.3.2 DFD Level 1 Diagram***



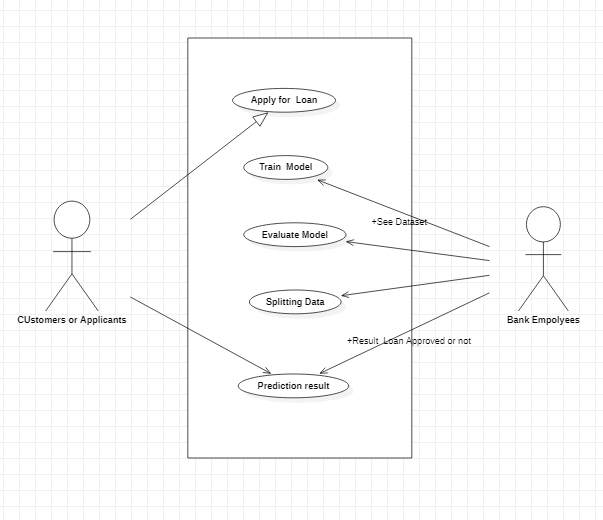
***Fig: 4.3.3 DFD Level 2 Diagram***

* 1. **UML DIAGRAMS**
     1. **ACTIVITY DIAGRAMS**



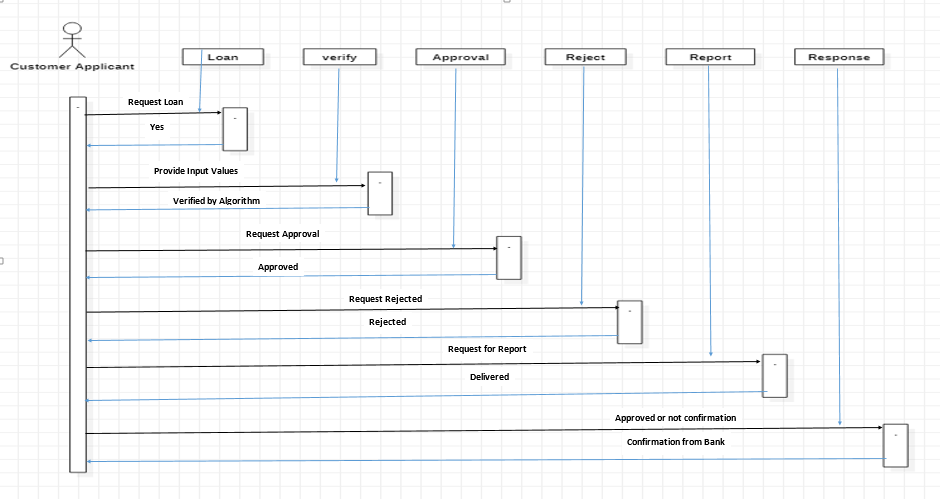
***Fig: 4.4.1 Activity Diagram***

* + 1. **USERCASE DIAGRAM**

****

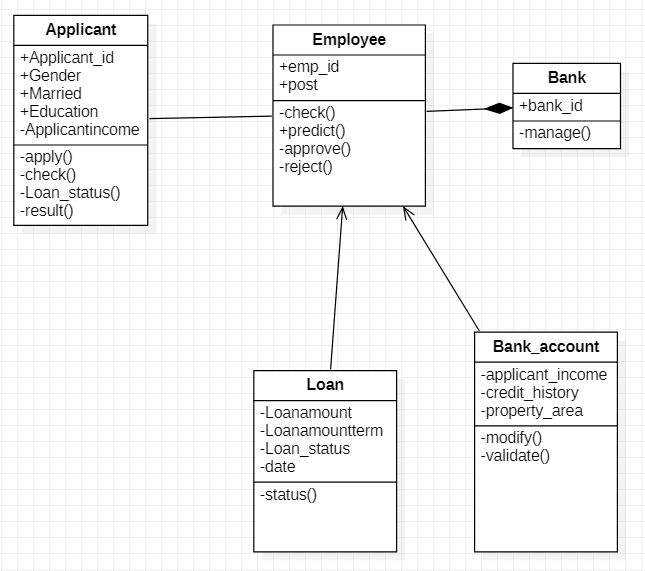
***Fig: 4.4.2 Use Case Diagram***

* + 1. **SEQUENCE DIAGRAM**

****

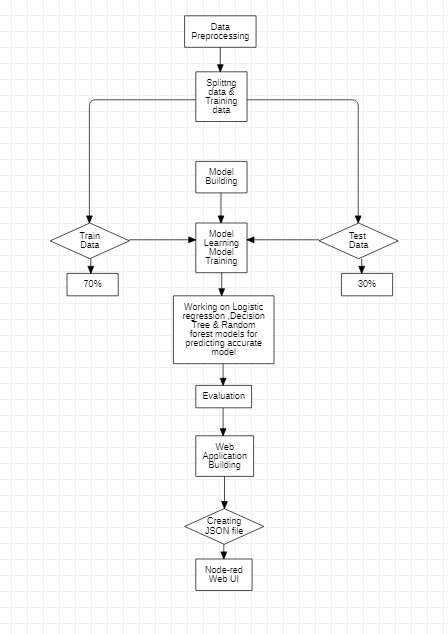
***Fig: 4.4.3 – Sequence Diagram***

* + 1. **CLASS DIAGRAM**

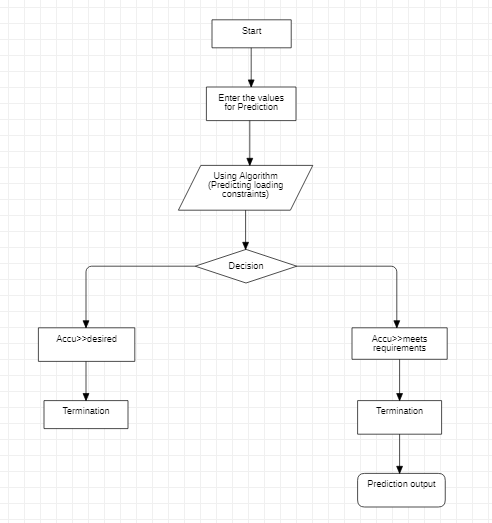


***Fig: 4.4.4 – Class Diagram***

**4.5 FLOWCHARTS**

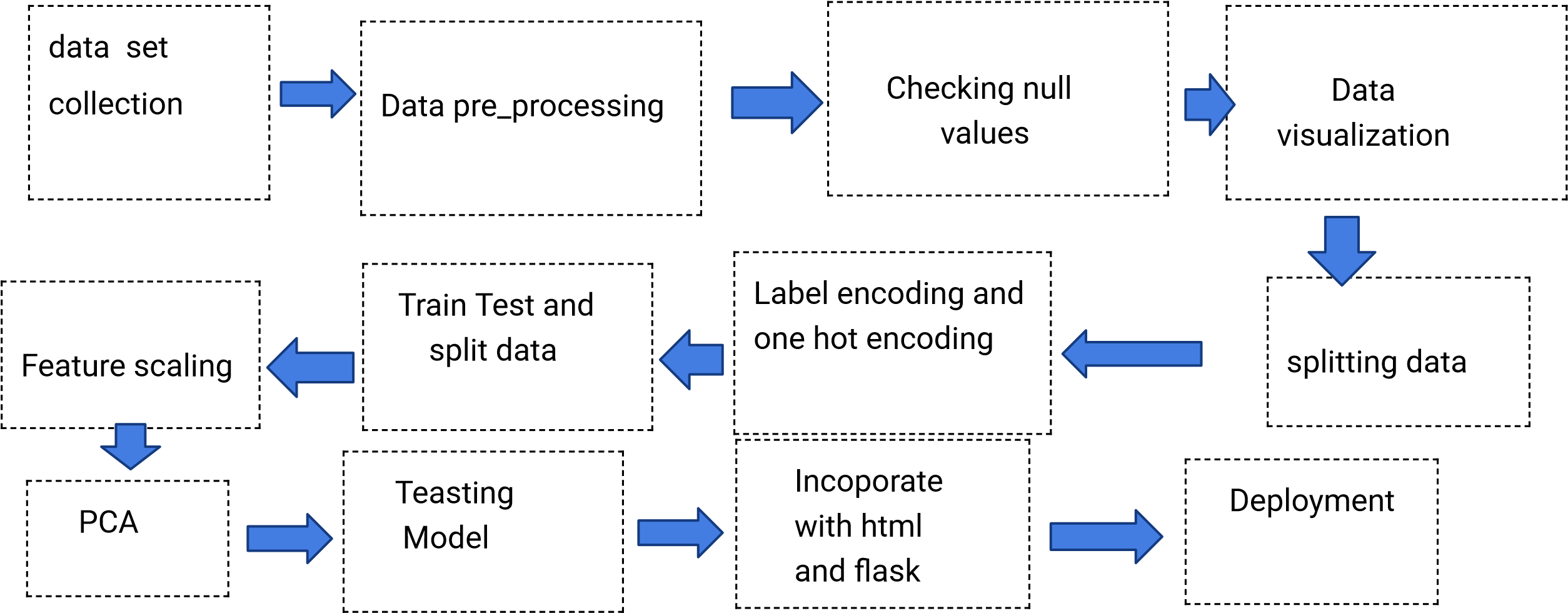
****

***Fig: 4.5.1 Flowchart I***



***Fig: 4.5.2 Flowchart II***

* 1. **BLOCK DIAGRAM**



***Fig: 4.6 Block Diagram***

1. **PROJECT IMPLEMENTATION**

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it’s constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

**5.1 TECHNOLOGY OVERVIEW**

**MACHINE LEARNING**

Nowadays, there are many risks related to bank loans, for the bank and for those who get the loans. The analysis of risk in bank loans need to understand what is the meaning of risk. In addition, the number of transactions in banking sector is rapidly growing and huge data volumes are available which represent the customers behavior and the risks around loan are increased. Here a new model for classifying loan risk in banking sector by using Machine Learning concepts. The model has been built using data from banking sector to predict the status of loans. Three algorithms have been used to build the proposed model: Random Forest, Logistic Regression and Decision Tree. By using a this algorithm, on Watson studio software the web application, model have been implemented and tested. The web application must be built with JSON or HTML5 languages with the machine learning model trained & deployed on IBM Watson Studio. The model needs to be deployed in the IBM cloud to get scoring endpoint which can be used as API in web app building. The model prediction needs to be showcased on User Interface.

**Technologies Used:**

1. IBM Watson Studio
2. Node-RED
3. Jupyter Notebook
4. Anaconda 3 prompt

**IBM WATSON STUDIO**

IBM Watson Studio provides tools to work more easily and collaboratively with data to build and train models at scale. It gives the flexibility to build models where your data resides and to deploy anywhere in a hybrid environment so you can operationalize data faster. Use The model needs to be deployed in the IBM cloud to get scoring endpoint which can be used as API in web app building. The model prediction needs to be showcased on User Interface. IBM Watson Machine Learning (WML) Service enables to create, train, and deploy manual-learning models using an Notebook, collaborative workflow. Create, train, and deploy manual-learning models. Use own data to create, train, and deploy machine learning and deep learning models.

**Services Used:**

1. IBM Cloud Object Storage
2. AI Service
3. IBM Machine Learning Service

**5.2 CODING**

**Frontend Source Code ( .JSON):**

[{"id":"6b689a25.9f21a4","type":"tab","label":" LOAN PREDICTION MODEL","disabled":false,"info":""},{"id":"f43c106a.9ef2b","type":"ui\_form","z":"6b689a25.9f21a4","name":"","label":"Applicant Details : Please enter the Applicant details to predict the Loan Application ","group":"3ea44d2.b3e7db2","order":2,"width":0,"height":0,"options":[{"label":"Enter your gender","value":"Gender","type":"text","required":true,"rows":null},{"label":"Enter your marital status","value":"Married","type":"text","required":true,"rows":null},{"label":"Enter no.of dependents","value":"Dependents","type":"number","required":true,"rows":null},{"label":"Enter your education status","value":"Education","type":"text","required":true,"rows":null},{"label":"Are you self-employed ?","value":"Self\_Employed","type":"text","required":true,"rows":null},{"label":"Enter your income","value":"ApplicantIncome","type":"number","required":true,"rows":null},{"label":"Enter your co-applicant income","value":"CoapplicantIncome","type":"number","required":true,"rows":null},{"label":"Enter your loan amount:","value":"LoanAmount","type":"number","required":true,"rows":null},{"label":"Choose your loan amount term","value":"Loan\_Amount\_Term","type":"number","required":true,"rows":null},{"label":"Select your credit history","value":"Credit\_History","type":"number","required":true,"rows":null},{"label":"Choose your property area","value":"Property\_Area","type":"text","required":true,"rows":null}],"formValue":{"Gender":"","Married":"","Dependents":"","Education":"","Self\_Employed":"","ApplicantIncome":"","CoapplicantIncome":"","LoanAmount":"","Loan\_Amount\_Term":"","Credit\_History":"","Property\_Area":""},"payload":"","submit":"Submit","cancel":"Reset","topic":"","x":390,"y":440,"wires":[["1558e048.e1374","5e24657f.23a75c"]]},{"id":"1558e048.e1374","type":"debug","z":"6b689a25.9f21a4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":590,"y":580,"wires":[]},{"id":"5e24657f.23a75c","type":"function","z":"6b689a25.9f21a4","name":"pre-token","func":"global.set(\"Gender\",msg.payload.Gender)\nglobal.set(\"Married\",msg.payload.Married)\nglobal.set(\"Dependents\",msg.payload.Dependents)\nglobal.set(\"Education\",msg.payload.Education)\nglobal.set(\"Self\_Employed\",msg.payload.Self\_Employed)\nglobal.set(\"ApplicantIncome\",msg.payload.ApplicantIncome)\nglobal.set(\"CoapplicantIncome\",msg.payload.CoapplicantIncome)\nglobal.set(\"LoanAmount\",msg.payload.LoanAmount)\nglobal.set(\"Loan\_Amount\_Term\",msg.payload.Loan\_Amount\_Term)\nglobal.set(\"Credit\_History\",msg.payload.Credit\_History)\nglobal.set(\"Property\_Area\",msg.payload.Property\_Area)\nvar apikey=\"6uAKe\_DH0OKqvkTR\_Y6oIsCf5nR7thrmkbZGLasq-kqA\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant\_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;","outputs":1,"noerr":0,"x":600,"y":360,"wires":[["b9b9e5e5.d0d0f8"]]},{"id":"b9b9e5e5.d0d0f8","type":"http request","z":"6b689a25.9f21a4","name":"","method":"POST","ret":"obj","paytoqs":true,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":752.0000038146973,"y":404.00000286102295,"wires":[["3f70f1b2.9a949e","3913c180.7c18be"]]},{"id":"3f70f1b2.9a949e","type":"debug","z":"6b689a25.9f21a4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":940,"y":460,"wires":[]},{"id":"3913c180.7c18be","type":"function","z":"6b689a25.9f21a4","name":"Pre Prediction","func":"var Gender = global.get(\"Gender\")\nvar Married = global.get(\"Married\")\nvar Dependents = global.get(\"Dependents\")\nvar Education = global.get(\"Education\")\nvarSelf\_Employed = global.get(\"Self\_Employed\")\nvarApplicantIncome = global.get(\"ApplicantIncome\")\nvarCoapplicantIncome = global.get(\"CoapplicantIncome\")\nvarLoanAmount = global.get(\"LoanAmount\")\nvarLoan\_Amount\_Term = global.get(\"Loan\_Amount\_Term\")\nvarCredit\_History = global.get(\"Credit\_History\")\nvarProperty\_Area = global.get(\"Property\_Area\")\nvar token=msg.payload.access\_token\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"input\_data\":[{\"fields\":[\"Gender\",\"Married\",\"Dependents\",\"Education\",\"Self\_Employed\",\"ApplicantIncome\",\"CoapplicantIncome\",\"LoanAmount\",\"Loan\_Amount\_Term\",\"Credit\_History\",\"Property\_Area\"],\"values\":[[Gender,Married,Dependents,Education,Self\_Employed,ApplicantIncome,CoapplicantIncome,LoanAmount,Loan\_Amount\_Term,Credit\_History,Property\_Area]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":944.0000076293945,"y":356.0000021457672,"wires":[["a4bc2dc3.0ba1b"]]},{"id":"9a6f2338.7db27","type":"debug","z":"6b689a25.9f21a4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":1322.9999914169312,"y":464.00000381469727,"wires":[]},{"id":"d633e6ea.c87678","type":"inject","z":"6b689a25.9f21a4","name":"","topic":"","payload":"","payloadType":"date","repeat":"","crontab":"","once":false,"onceDelay":0.1,"x":470,"y":260,"wires":[["5e24657f.23a75c"]]},{"id":"1717e8b7.7eff07","type":"debug","z":"6b689a25.9f21a4","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":1210,"y":220,"wires":[]},{"id":"75fa7f6a.4628c","type":"function","z":"6b689a25.9f21a4","name":"Parsing","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;","outputs":1,"noerr":0,"x":1020,"y":280,"wires":[["1717e8b7.7eff07","301e5ea6.b7d1c2","4fd2c9de.45bb28"]]},{"id":"301e5ea6.b7d1c2","type":"ui\_text","z":"6b689a25.9f21a4","group":"3ea44d2.b3e7db2","order":3,"width":0,"height":0,"name":"","label":"YOU ARE ELIGIBLE OR NOT(Y-Yes, N-No): ","format":"{{msg.payload}}","layout":"row-left","x":1227.6000366210938,"y":521.2000122070312,"wires":[]},{"id":"a4bc2dc3.0ba1b","type":"http request","z":"6b689a25.9f21a4","name":"","method":"POST","ret":"obj","paytoqs":true,"url":"https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/200340e6-676d-43e6-a406-525145ab6d0d/predictions?version=2020-11-22","tls":"","persist":false,"proxy":"","authType":"","x":1170,"y":380,"wires":[["9a6f2338.7db27","75fa7f6a.4628c"]]},{"id":"5b78eda0.5a3014","type":"ui\_template","z":"6b689a25.9f21a4","d":true,"group":"3ea44d2.b3e7db2","name":"Footer","order":5,"width":0,"height":0,"format":"\n<!DOCTYPE html>\n\n<html lang=\"en\" dir=\"ltr\">\n <head>\n <meta charset=\"utf-8\">\n\t\t<meta name=\"viewport\" content=\"width=device-width, initial-scale=1.0\">\n\t\t<link rel=\"shortcut icon\" href=\"https://dl.dropbox.com/s/yofbigw3fu8xx7g/styles.css\">\n <link rel=\"stylesheet\" type=\"text/css\" href=\"\">\n <script src=\"https://kit.fontawesome.com/5f3f547070.js\" crossorigin=\"anonymous\"></script>\n \n\t</head>\n\n <body>\n\n <!-- Footer -->\n <div class='footer'>\n <p class='footer-description'>Developed by K N Pavani</p>\n </div>\n\n </body>\n</html>\n","storeOutMessages":true,"fwdInMessages":true,"resendOnRefresh":true,"templateScope":"local","x":530,"y":760,"wires":[[]]},{"id":"fa60b20b.4dba9","type":"ui\_template","z":"6b689a25.9f21a4","group":"3ea44d2.b3e7db2","name":"Header and Background","order":1,"width":"10","height":"1","format":"\n<!DOCTYPE html>\n\n<html lang=\"en\" dir=\"ltr\">\n <head>\n <meta charset=\"utf-8\">\n <meta name=\"viewport\" content=\"width=device-width, initial-scale=1.0\">\n\t\t<title>Predict Loan Approved or not using IBM Auto AI service</title>\n\t\t<link rel=\"shortcut icon\" href=\"{{ url\_for('static', filename='logo.ico') }}\">\n <link rel=\"stylesheet\" type=\"text/css\" href=\"https://dl.dropbox.com/s/yofbigw3fu8xx7g/styles.css\">\n <script src=\"https://kit.fontawesome.com/5f3f547070.js\" crossorigin=\"anonymous\"></script>\n \n\t \n\t</head>\n\n <body>\n\n <!-- Website Title -->\n \t<div class=\"container\" >\n <h2 class='container-heading'><span class=\"heading\_font\">YOUR ARE ELIGIBLE FOR LOAN?</span></h2>\n <div class='description'>\n \t\t\t<p>Check you are eligible for loan or Not</p>\n \t\t</div>\n \t</div>\n \t\n<style>\n body {\n background-image: url(\"\");\n \n background-size: auto;\n \n }\n</style>\n\n\n </body>\n</html>\n","storeOutMessages":true,"fwdInMessages":true,"resendOnRefresh":true,"templateScope":"local","x":590,"y":720,"wires":[[]]},{"id":"4fd2c9de.45bb28","type":"ui\_template","z":"6b689a25.9f21a4","group":"3ea44d2.b3e7db2","name":"Prediction Results","order":4,"width":0,"height":0,"format":"<!DOCTYPE html>\n<html lang=\"en\">\n <head>\n\t\t<meta charset=\"utf-8\">\n\t\t<meta name=\"viewport\" content=\"width=device-width, initial-scale=1.0\">\n\t\t<title>BANK LOAN APPROVAL PREDICTION</title>\n\t\t<link rel=\"shortcut icon\" href=\"https://image.freepik.com/free-vector/glossy-red-heart-with-white-heartbeat-pulse-blue-molecules-background-medical-concept\_1302-5653.jpg\">\n\t\t<link rel=\"stylesheet\" href=\"{{ url\_for('static', filename='styles.css') }}\">\n\t\t<script src=\"https://kit.fontawesome.com/5f3f547070.js\" crossorigin=\"anonymous\"></script>\n\t\t\n\t</head>\n<body><style>\nbody {\n background-image: url('https://miro.medium.com/max/700/0\*eC1EUwoo6rMOypir');\n background-repeat: no-repeat;\n background-attachment: fixed;\n background-size: cover;\n}\n\n</style>\n\t<!-- Result -->\n\t\t<div class=\"results\">\n <div ng-if=\"msg.payload == 'Y' \">\n \n <h1>Prediction: <span class='safe'>Great! You are eligible for loan.</span></h1>\n\t\t\t\t<div class=\"gyan\">\n\t\t\t\t\t<h1 style=\"color: green;\"><p>Tips for Loan Prediction:</p><br><p>Dear customer, Coonragulations!!<br> You can now apply to the Bank loan approval application by clicking on the Link below.</p><br><p>Please Click here:</p><br<p><a href=\"https://www.google.com\">Visit google.com!</a></p></h1>\n\t\t\t\t\t<ul>\n\t\t\t\t\t\n\t\t\t\t\t</ul>\t\t\t\t\n\t\t\t\t</div>\n \n </div>\n <div ng-if=\"msg.payload == 'N' \">\n \n <h2>Prediction: <span class='danger'>Sorry,You are not eligible for loan.</span></h2>\n \t <h2 style=\"color: red;\">Tips for Loan Prediction:<br><p>Thank you for visiting our bank loan eligiblity prediction.</p><br><p>Please,you will try in others bamk for loan.</p></h2>\n </div>\n <div ng-if=\"msg.payload == 'Null' \">\n \n \n \n </div>\n\n </div>\n \n</body>\n</html>","storeOutMessages":true,"fwdInMessages":true,"resendOnRefresh":true,"templateScope":"local","x":1210,"y":680,"wires":[[]]},{"id":"3ea44d2.b3e7db2","type":"ui\_group","z":"","name":"Home","tab":"15ba0e23.8c4ec2","order":1,"disp":false,"width":"10","collapse":false},{"id":"15ba0e23.8c4ec2","type":"ui\_tab","z":"","name":"Loan prediction","icon":"dashboard","order":13,"disabled":false,"hidden":false}]

1. **TESTING**

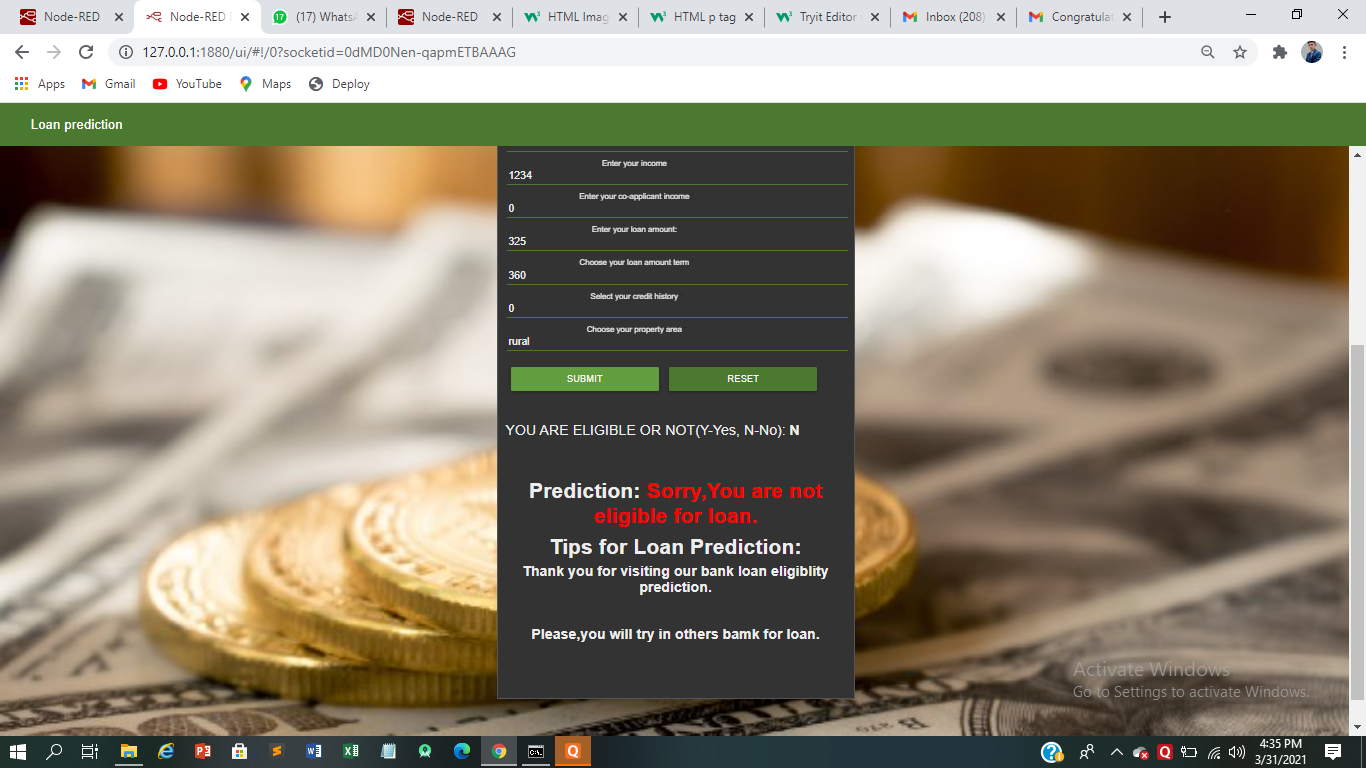
**6.1 SYSTEM TESTING**

**6.1.1 MANUAL TEST**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Table no 6.1: *Test Case 1*** | |
|  |  |  |  |
| **Test Case 01** | |  |  |
|  |  |  |  |
| **Purpose** | | Data Formalization | |
|  |  |  |  |
| **Pre-requisite** | | Gender, married status, no. of dependents, income, loan amount | |
|  |  |  |  |
| **Test Data** | | Validation of applicants data | |
|  |  |  |  |
| **Steps** | | Take Information from loan applicant | |
|  |  |
|  |  |  |  |
| **Expected result** | | Applicant data input | |
|  |  |  |  |
| **Actual Result** | | Applicant data accepted successfully | |
|  |  |  |  |
|  |  | **Table no 6.2: *Test Case 2*** | |
|  |  |  |  |
| **Test Case 02** |  |  |  |
|  |  |  |  |
| **Purpose** |  | Testing Machine Learning models |  |
|  |  |  |  |
| **Pre-requisite** |  | Prediction, Testing dataset |  |
|  |  |  |  |
| **Test Data** |  | Testing dataset |  |
|  |  |  |  |
|  |  | Prediction |  |
| **Steps** |  | Accuracy score |  |
|  |  |  |
|  |  |  |  |
| **Expected result** |  | Prediction and accuracy results |  |
|  |  |  |  |
| **Actual Result** |  | Prediction and accuracy results achieved |  |
|  |  |  |  |

1. **RESULTS AND DISCUSSIONS**

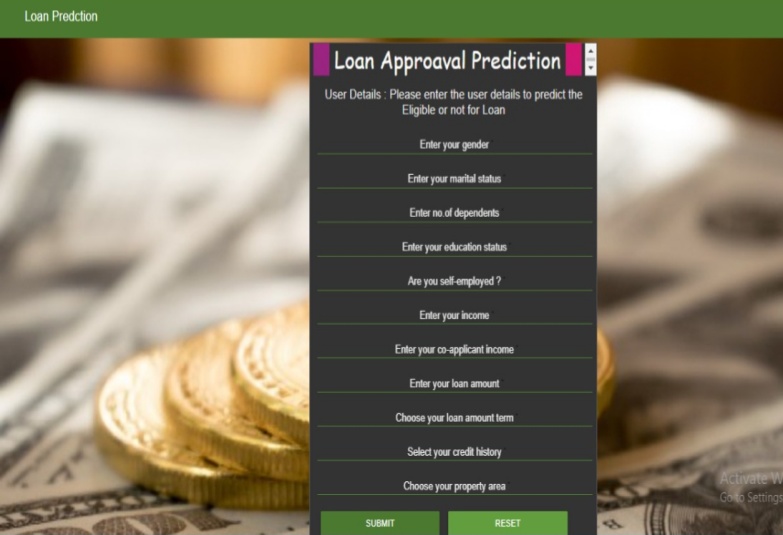
**7.1 RESULT SETS**

****

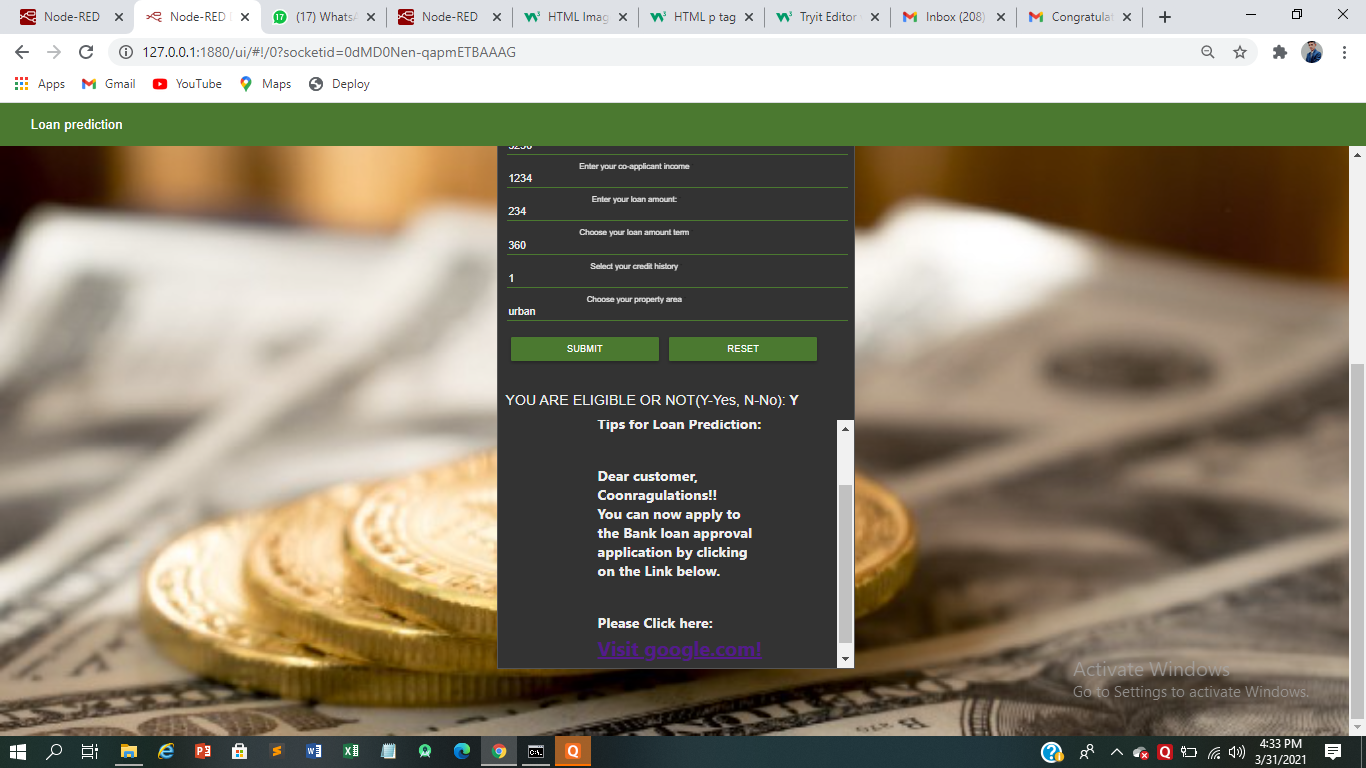
*Fig 7.1.1* ***Prediction Result***

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Male | No | 0 | Not Graduate | No | 3189 | 2598 | 120 |  | 1 | Rural | Y |
| Female | No | 0 | Graduate | No | 8333 | 0 | 280 | 360 | 1 | Semiurban | Y |
| Male | Yes | 1 | Graduate | No | 3155 | 1779 | 140 | 360 | 1 | Semiurban | Y |
| Male | Yes | 1 | Graduate | No | 5500 | 1260 | 170 | 360 | 1 | Rural | Y |
| Male | Yes | 0 | Graduate |  | 5746 | 0 | 255 | 360 |  | Urban | N |
| Female | No | 0 | Graduate | Yes | 3463 | 0 | 122 | 360 |  | Urban | Y |
| Female | No | 1 | Graduate | No | 3812 | 0 | 112 | 360 | 1 | Rural | Y |
| Male | Yes | 1 | Graduate | No | 3315 | 0 | 96 | 360 | 1 | Semiurban | Y |
| Male | Yes | 2 | Graduate | No | 5819 | 5000 | 120 | 360 | 1 | Rural | Y |
| Male | Yes | 1 | Not Graduate | No | 2510 | 1983 | 140 | 180 | 1 | Urban | N |
| Male | No | 0 | Graduate | No | 2965 | 5701 | 155 | 60 | 1 | Urban | Y |

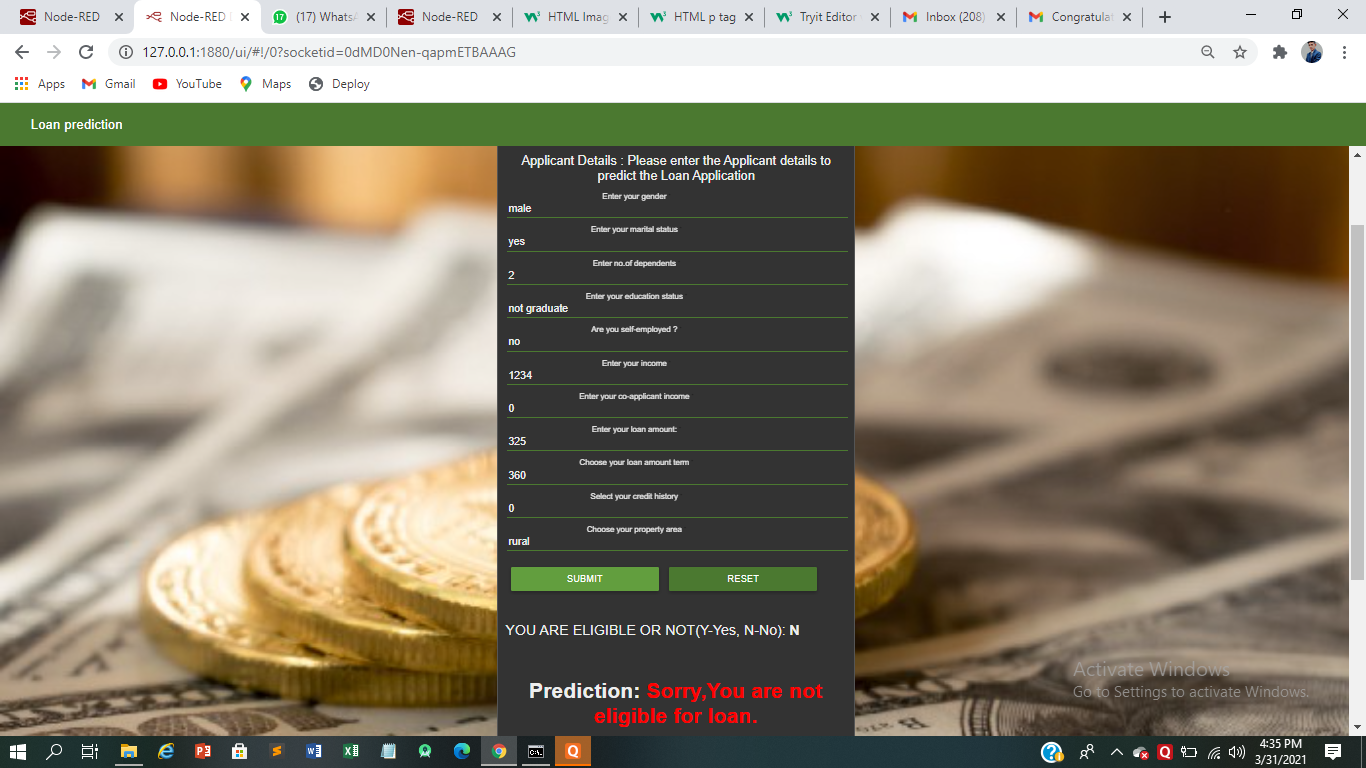
*Table 7.1.1* ***Prediction Result set***

**7.2 SCREENSHOTS**

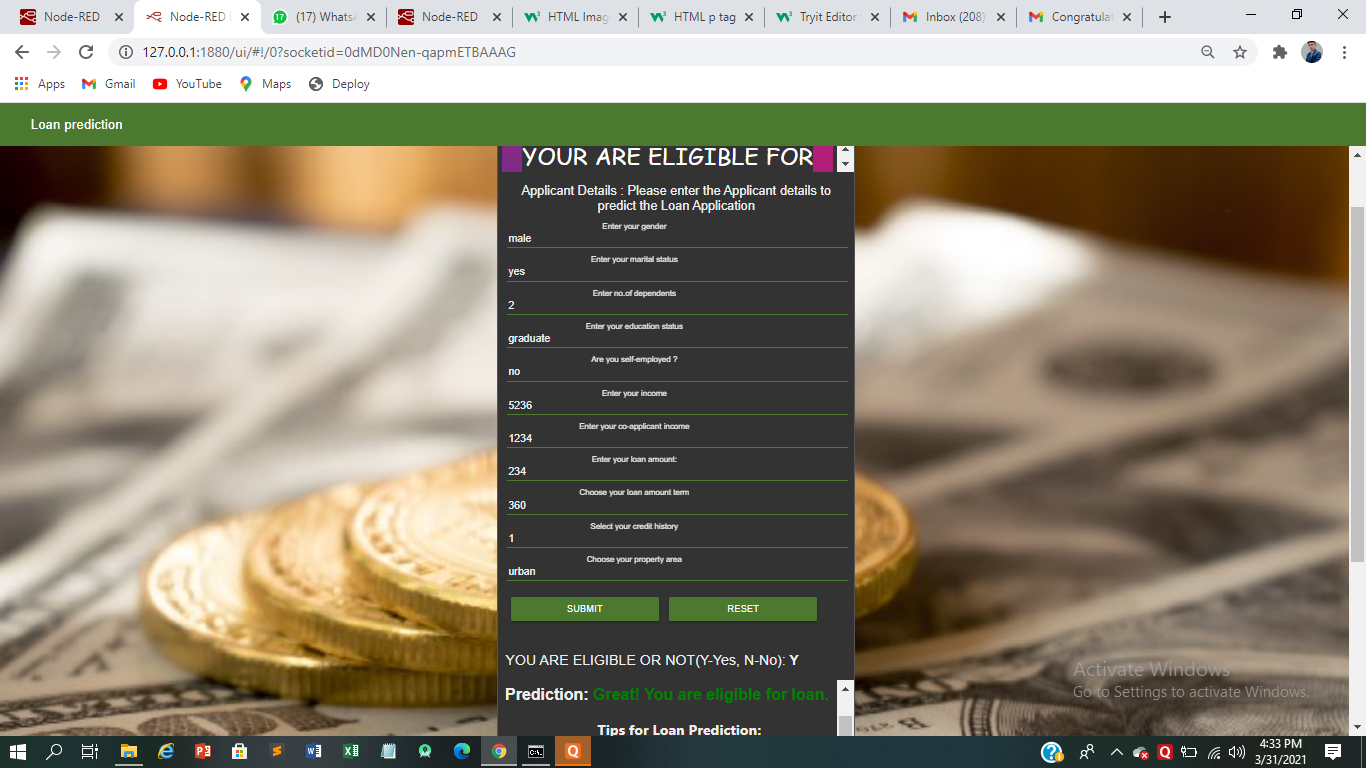
*Fig 7.2.1* ***Main page 1***

****

*Fig 7.2.2* ***Main page 02***

****

*Fig 7.2.3* ***not eligible result***

****

*Fig 7.2.4* ***eligible Result***

1. **ADVANTAGES**

* Easy and simple User Interface for the bank people who is going to evaluate the customer loan status.
* Logistic Regression give the accurate result of the prediction up to 83% which is the algorithm we used for prediction.
* It is composed using the JSON and Python for the web usage in real time.
* It can work in real time and predict as soon as the necessary details for prediction are given to the model.

**CONCLUSION**

Thus we have tried to implement the paper present by **Mohammad Ahmad Sheikh, “An Approach for Prediction of Loan Approval using Machine Learning Algorithm,” Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)** and according to implementation; the conclusion is that the logistic regression algorithm is most accurate in three of them hence adopted to build a user interface for predicting loan eligibility. The accuracy is compared with other algorithms i.e. random forest and decision tree. It can be seen that the logistic regression algorithm gives high accuracy than the other algorithms in the prediction of loan default and has strong ability of generalization; But still there is no definitive standards that which algorithm should be used. Therefore, cross validation is done to get more accurate result.

**FUTUREWORK**

This project is for certain banks as well as loan applicants as they can predict the eligibility level of the applicant for the loan. This application is working properly and meeting to all bankers requirements. This component can be easily plugged in many other systems. In near future, so-called software could be made more secure, reliable, accurate and dynamic weight adjustment. In future, this module of prediction can be integrate with module of automated processing system. Also the algorithms implemented in models can be upgraded to have efficient and less time complexity and high accuracy. This model concludes that a bank should not only target rich customers to be targeted for granting the loan but it can assess the other attributes of applicants as well.