

# LOAN APPROVAL PREDICTION USING SUPERVISED MACHINE LEARNING ALGORITHM

Dr.E.Neelima<sup>1</sup>, Venkata Ayyappa Reddy Maruprolu<sup>2</sup>, Sai Prasad Penta<sup>3</sup>,  
Karthikeya Mallareddy<sup>4</sup>

#1 Assistant Professor, Department of CSE, GITAM (Deemed to be University),

Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, India

#2,#3,#4 Student, Department of CSE, GITAM(Deemed to be University),

Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, India

**Abstract\_** As we all are experiencing, this Covid-19 Pandemic extended the demand for a center class/lower center classification candidate to collect mortgage thru authorities banks or personal banks. Even although authorities attempting to increasing the cash-flow in the banks has its personal threshold candiation which it has to given to few candidates only, so predicting or classifying to whom the mortgage be sanctioned which will be a impervious desire for the financial institution is an essential aspect.

Therefore, In our assignment we strive to limit the uncertainty in the back of opting the authentic individual so that we can minimize the Bank's Human Resource, can tightly closed Banks assets, can decrease the length of mortgage get sanction. The important significance of this undertaking is to classify whether or not sanctioning the mortgage to a unique candidate will be impervious or not. If not, then we won't proceed the similarly method with that candidate.

Classifying a mannequin such that it is educated by using Loan applicant's/client's required datasets and determining whether or not the new consumer is in all likelihood to repay the mortgage quantity which they are obtaining .

Our Project includes: Data necessities Gathering, Developing of laptop mastering model, Training of machine on our model, Testing, UI acquiring Decision.

## 1.INTRODUCTION

Since April 2020, our united states of america is shifting into monetary recession. Our Country's GDP shifted to terrible scale due to terrible money drift all over the Banks, Industries, and many different fields in which cash-flow is must. Bank performs an essential position in market economy. To resume and to overcome from this monetary recession authorities agreed to furnish loans and assist middle/lower center

classification families, small scale businesses, medium scale organizations etc.

As we all understand to correctly aquire Loan via a financial institution there are many steps to be observed by using the candidate as nicely as the financial institution mortgage officials.

Candidate ought to post application, Application must be verified, Candidate's financial institution account associated small print like Credits, Debits, Bank

Balance, Credit History and so forth need to be verified, Candidate's Physical Assets/properties write-up must be updated, all the different associated files to be validated, etc.

To function all these technique the bank desires a couple of human assets and all these methods takes almost 1 week to get his/her mortgage quantity to sanction. And primarily based on single individual's data, financial institution would possibly now not take ideal selection to furnish the mortgage or not. So these variety of conditions can also drag the financial institution to uncertainty about their decision.

Though this looks to be a big technique there is a want for Banks to become aware of or to predict the candidates who are greater probably to the mortgage quantity pay lower back in which they acquire at this immediately emergency situation. So that they can proceed with the similarly procedures except any insecurities .

By this notion of our assignment we can clearly classify or predict inside few minutes whether or not the candidate is in all likelihood to pay lower back the mortgage quantity as soon as sanctioned.

This will really minimise the length of loan techniques so that candidate as nicely as the financial institution can shop their time. Bank can limit their human useful resource as we at once furnish the facts to mannequin (though it is skilled with large series of data). Bank can securely furnish loans solely for these who are in all likelihood to the quantity pay lower back as soon as they obtained.

## 2.LITERATURE SURVEY

[1] Raj, J. S., & Ananthi, J. V., "Recurrent neural networks and nonlinear prediction in support vector machine" *Journal of Soft Computing Paradigm (JSCP)*, 1(01), 33-40, 2019.

The detection of edges is the one of the important stage in the application, associated with the machine vision, computer vision and the image processing. It is most commonly and highly preferred in the area were the extraction or the detection of the attribute are necessary. As the manual methods of diagnosis in the medical images acquired from the CT (computed tomography) and the MRI (magnetic resonance images) are very tedious and as well as time consuming, the paper puts forth the methodology to detect the edges in the CT and the MRI by employing Gabor Transform as well as the soft and the hard clustering. This proposed method is highly preferred among the image with dynamic variations. The technique used in the paper is evaluated using 4500 instance of the MRI and 3000 instance of CT. The results on the basis of the figure of merit (FOM) and Misclassification rate (MCR) are compared with other standard approaches and the performance was evinced.

[2] X.Frencis Jency, V.P.Sumathi,Janani Shiva Shri, "An exploratory Data Analysis for Loan Prediction based on nature of clients", *International Journal of Recent Technology and Engineering (IJRTE)*,Volume-7 Issue-4S, November 2018.

The term banking can be defined as receiving and protecting money that is deposited by the individual or the entities. This also includes lending money to the people which will be repaid within the

given time. Banking sector is regulated in most of the countries as it is the important factor in determining the financial stability of the country. The provision of banking regulation act allows public to obtain loans. Loans are good sum of money borrowed for a period and expected to be paid back at given interest rate. The purpose of the loan can be anything based on the customer requirements. Loans are broadly divided as open-ended and close-ended loans. Open-ended loans are the loans for which the client has approval for a specific amount. Examples of open-end loans are credit cards and a home equity line of credit (HELOC). Close-ended loans decrease with each payment. In other words, it is a legal term that cannot be modified by the borrower. Personal loans, mortgages, auto payments, instalment loan and student loans are the most common examples of close-ended loans. Secured or collateral loan are those loans that are protected by an asset. Houses, Vehicles, Savings accounts are the personal properties used to secure the loan. Unsecured loans are also known as personal or signature loans. Here the lender believes that the borrower can repay the loan based on financial resources possessed by the borrower. Liquidity risk is the risk that arises from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. The interest rate risk is the risk in which the interest rates priced on loans will be too low to earn the bank money. Revised Version Manuscript Received on 25 November, 2018. Ms.X.Francis Jency, CSE Department, Kumaraguru College of Technology, Coimbatore, India Ms.V.P.Sumathi, CSE Department, Kumaraguru College of Technology, Coimbatore, India Janani

Shiva Sri, C S Department, Kumaraguru College of Technology, Coimbatore, India The primary objective of the bank is to provide their wealth in the safer hands. In recent times, banks approve loan after verifying and validating the documents provided by the customer. Yet there is no guarantee whether the applicant is deserving or not. This paper classifies the customers based on certain criteria. The classification is done using Exploratory Data Analysis. Exploratory Data Analysis (EDA) is an approach to analyse the datasets that summarizes the main characteristics with visual methods. The purpose of using EDA is to uncover the underlying structure of a relatively larger set of variables using visualizing techniques.

**[3] Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, k Vikash, "Loan Prediction by using Machine Learning Models", International Journal of Engineering and Techniques. Volume 5 Issue 2, Mar-Apr 2019**

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 project 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 project 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. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

### 3. PROPOSED SYSTEM

In our proposed model, we train the model with huge collection of datasets called training datasets (80% of total datasets) and after completing this task we then test our model with testing datasets which is 20% of total datasets and perform any remaining preparations to get ready with our model to work on real applicants.

Client's Account details in the respective bank is provided to the model along with that we also provide client's other banking details, physical assets details and all required details to our model.

Based on the algorithms and data provided to the model, the model provides the decisions as how likely the client is about to repay his loan amount. These outcomes are validated by the Loan Administrator and takes his final decision whether to provide the loan or not.

### 3.1 IMPLEMENTATION

To move into methodologies we need to understand what is Machine Learnings. Machine Learning is an art of developing Models such that will provides the decisions as outcomes based on the past data in which it is being trained (training and testing datasets) without any Human

Intervention. The shown some model methodology involved step wise represented in it.

#### 3.1.1 DATA SELECTION

Machine Learning Model we need to train our model and most of the developed process is totally based on datasets. So we need to get access for a huge collection of datasets hence we can move further in our model development. Many big firms providing free access to the Machine Learning

Datasets as an encouragement to the developers and provide security for the data. Example: Google's KAGGLE is open source model development platform and dataset management for the machine Learning Enthusiast.

#### 3.1.2 DATA PRE-PROCESSING

The change we perform to our data before delivering it to the algorithm is referred to as pre-processing. Data Preprocessing is a technique for transforming unclean or unintelligible data into a clean or understandable data collection. When data is obtained from various sources in raw format, it is analyzed to see if it is practical.. The provided data must be formatted in a sensible manner in order for our model to produce better results. Some input in a specific format is required by a Machine Learning model.

For example, because the Random Forest method does not tolerate null values, the original raw data set must be handled in order to run the algorithm.

So to perform Missing Value/Null Value

Data Pre-Processing on the raw data, we have few solutions i.e, we can insert either max value or min value or Average value or Mode Value in the missing fields of a respective attribute in the given raw dataset.

### 3.1.3 FEATURES SELECTION

Feature selection is the process of automatically or manually picking the features that contribute the most to the prediction variable or output that you are interested in.

Irrelevant features in your data might impair model accuracy and lead your model to train on irrelevant features.

The first and most crucial phase in model design should be feature selection and data cleansing, since these assist the model to:

**Reduces Overfitting:** When there is less duplicated data, there is a lower probability of reaching judgements or conclusions based on noise.

**Improves Modeling Accuracy:** Less misleading data implies better modelling accuracy.

**Reduces Training Time:** Data points reduce algorithm complexity, which allows algorithms to learn quicker.

### 3.1.4 BUILDING CLASSIFICATION MODEL

There are two types of machine learning models: supervised and unsupervised. Supervised learning is a machine learning activity that entails learning a function that translates an input to an output by utilising input-output pairs as examples.

Unsupervised learning is a type of machine learning that looks for previously overlooked patterns in a data collection with no pre-existing labels and very no human supervision.

Because we intend to incorporate Supervised Learning Model in our model, we will go more into Supervised Learning Model.

If the model is supervised, it is classified as either a regression or a classification model.

A regression model is a set of machine learning techniques for predicting a continuous outcome variable (y) based on one or more predictor variables' values (x). A regression model's main goal is to generate a mathematical equation that describes y as a function of x variables.

The Random Forest algorithm is a tool for supervised learning. Because of its simplicity and versatility, it is also one of the most extensively used algorithms (it can be used for both classification and regression tasks). Random Forest generates a large number of decision trees and then combines them to provide a more precise and reliable forecast. Random forests are also difficult to beat in terms of performance. Of course, a better model, like as a neural network, may always be found, but they take longer to create and can handle a wide range of feature types, including binary, category, and numerical data. Random Forest Classifier is a classification algorithm that uses a random forest

The following uses pseudo code to forecast using the learned random forest method.

1. Takes the test characteristics and uses the rules of each randomly created decision tree to predict the outcome, then saves the predicted outcome (target)

2. Determine each projected target.

3. Use the random forest algorithm's highest-voted projected target as the final prediction

## 4.RESULTS AND DISCUSSION

### 4.1 Models Comparison

```
compare = pd.DataFrame({'Model': ['Logistic Regression', 'K Neighbors', 'SVM', 'Decision Tree', 'Random Forest'],
                        'Accuracy': [LRAcc*100, KNAcc*100, SVCAcc*100, DTAcc*100, RFAcc*100,]})
compare.sort_values(by='Accuracy', ascending=False)
```

	Model	Accuracy
6	Random Forest	93.333333
1	K Neighbors	86.666667
2	SVM	86.666667
0	Logistic Regression	84.444444
5	Decision Tree	82.222222

In general, it can be seen that **all models can achieve up to 80% accuracy**.The highest accuracy is **93%**

### Loan Approval Prediction UI

Step1: Open command prompt install streamlit type: pip install streamlit

Step2: Run the file,streamlit run sample.py

Step3:Streamlit appear in a new tab in your browser

### User Interface

#### Bank Loan Prediction using Machine Learning

Full Name: \_\_\_\_\_

Gender: **Female** -

Marital Status: **Mar** -

Age: \_\_\_\_\_ -

Dependents: **0** -

Education: **Not Graduate** -

Employment Status: **Job** -

Property Area: **Rural** -

Credit Score: **Between 300 to 500** -

Applicant's Monthly Income(in thous(k)) **0** - +

Co-Applicant's Monthly Income(in thous(k)) **0** - +

Loan Amount: **0** - +

Loan Duration: **2 Month** -

### Approval case

Property Area: **Urban** -

Credit Score: **Between 300 to 500** -

Applicant's Monthly Income(in thous(k)) **100** - +

Co-Applicant's Monthly Income(in thous(k)) **20** - +

Loan Amount: **200** - +

Loan Duration: **2 Month** -

Hello: sai !! Congratulations!! you will get the loan from Bank.

## Not Approval case

The screenshot shows a web form for a loan application. The fields are as follows:

- Property Area: Urban
- Credit Score: Between 300 to 500
- Applicant's Monthly Income(in thous(k)): 100
- Co-Applicant's Monthly income(in thous (k)): 20
- Loan Amount: 500
- Loan Duration: 2 Month

There is a 'Submit' button at the bottom of the form. Below the form, a red message box displays the text: "Hello: sai || According to our Calculations, you will not get the loan from Bank".

## 5. CONCLUSION

In this Project, we have proposed a device for Loan Uncertainty Prediction that makes use of Random Forest Algorithm which belongs to Machine Learning. During the COVID-19 pandemic times, many are financially degraded as they hit with the aid of a number of factors throughout lockdown. So to limit the Loan Approval duration, to decrease the Human Resource in Bank Organization, and to enhance the sure bet to supply loans to true clients, We designed this Model.

This Model is skilled with the aid of the information of current clients Financial status, stability, financial institution associated details and so on and can supply binary selection as 0's and 1's relying on the statistics in which we provide. Where zero represents Repayment will be

accomplished and 1 represents no Repay. With 93.3 as the F1\_score which is the high-quality overall performance evaluated for this approach.

## REFERENCES

- [1] A. Goyal and R. Kaur, "A survey on Ensemble Model for Loan Prediction", International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.
- [2] A. J. Hamid and T. M. Ahmed, "Developing Prediction Model of Loan Risk in Banks using Data Mining".
- [3] G. Shaath, "Credit Risk Analysis and Prediction Modelling of Bank Loans Using R".
- [4] A. Goyal and R. Kaur, "Accuracy Prediction for Loan Risk Using Machine Learning Models".
- [5] M. Sudhakar, and C.V.K. Reddy, "Two Step Credit Risk Assessment Model for

Retail Bank Loan Applications Using Decision Tree Data Mining Technique”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5(3), pp. 705718, 2016.

[6] Gerritsen, R. (1999). Assessing loan risks: a data mining case study. IT professional, 1(6), 16-21.

[7] X. Francis Jency, V.P. Sumathi, Janani Shiva Sri: An Exploratory Data Analysis 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.

[8] Mohit Lakhani, Bhavesh Dhotre, Saurabh Giri: Prediction of Credit Risks in Lending Bank Loans.J.

[9] 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).

[10] Vikash V, Mohammad Amir Ahmed: Loan Default Prediction using Machine Learning Technique.