



**SMART LENDER APPLICATION  
CREDIBILITY DETECTION FOR  
LOAN APPROVAL**



**A PROJECT REPORT**

*Submitted by*

**ADITYA B (19CSR006)**

**GOKUL SARATHY R (19CSR026)**

**KARTHEESWARAN S (19CSR044)**

**MEENACHI S (19CSR060)**

*Inpartial fulfillment for the award of the degree of*

**BACHELOR OF ENGINEERING**

*in*

**COMPUTER SCIENCE AND ENGINEERING**

**VELALAR COLLEGE OF ENGINEERING**

**AND TECHNOLOGY**

**ERODE-638012**

**NOV 2022**

## ACKNOWLEDGEMENT

We are greatly indebted to many people who have contributed to the progress of our project and have encouraged us right from the start to make our ideas a reality.

We are privileged to express our heartfelt thanks to our honourable Secretary and Correspondent **THIRU.S.D.CHANDRASEKAR, B.A.**, who provided all the facilities to build our project.

We express our deep sense of gratitude and sincere thanks to **Dr.M.JAYARAMAN,B.E,M.E.,Ph.D.**,for permitting us to undertake this project.

We are indebted to our Dean **Dr.P.JAYACHANDAR, M.E.,Ph.D.**, for giving us the opportunity and continuous inspiration to carry out this project.

We express our profound gratitude to our beloved Head of the Department **Dr.S.JABEEN BEGUM, M.E., Ph.D.**, for her valuable immensely profited by her comments and reviews.

We express our sincere thanks to our Faculty Mentor **Ms.C.KOTTEESWARI B.E.,(Ph.D.,)** Assistant Professor for their guidance to do our project successfully. We express our sincere thanks to our parents for their support to do our project successfully.

## **ABSTRACT**

Customer loan prediction is usually life time issue so; each and every retail bank faces the issue at the minimum lifetime. If done exactly, it can spare a lot's of man hours at the conclusion of a retail bank. If Company wants to semi automate the loan acceptability process (real time) based on customer detail provided while filling online application form. These subtle elements are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others. To automate this method, they have given an issue to recognize the customers segments; those are allowed for loan amount total so they can clearly target these customers. We need to predict whether or not a loan would be approved. In a classification problem, we need to predict separate values based on a given set of self-sufficient variable(s). What's our objective is to implement machine learning model so as to classify, to the best doable degree of accuracy, and dataset gathered from Kaggle. Random forest classification method shows best accuracy in classifying given on loan candidates using python help on Jupyter notebook.

Keys - Customer loan, Prediction, preprocessing, classification models.

TABLE OF CONTENTS		
CHAPTERNO.	TITLE	PAGE NO
	ABSTRACT	3
	LIST OF FIGURES	5
	LIST OF ABBREVIATIONS	6
1	INTRODUCTION	7
2	LITERATURE SURVEY	9
3	PROPOSED MODEL	12
4	SYSTEM ARCHITECTURE	13
5	PROPOSED ALGORITHM	17
6	SYSTEM FEATURES	18
7	MATHEMATICAL MODEL	19
8	CONCLUSION	20
9	REFERENCE	21

## LISTOFFIGURES

CHAPTERNO	TITLE	PAGENO
4.1	MAJOR ATTRIBUTES	13

## **LIST OF ABBREVIATIONS**

- BN -Bayesian Network
- DL -Deep Learning
- CART -Classification And Regression Tree
- DT -Decision Tree
- ID3 -Iterative Dichotomizer 3
- KNN -kNearest neighbor
- ML -Machine Learning
- NB -Naive Bayes
- RF -Random Forest
- SVM -Support Vector Machine
- XGBoost -eXtreme Gradient Boosting

## **CHAPTER 1**

### **INTRODUCTION**

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. Dream housing Finance Company deals in all loans. They have presence across all urban, semi urban and rural areas. Customer first apply for loan after that company or bank validates the customer eligibility for loan. Company or bank wants to automate the loan eligibility process (real time) based on customer details provided while filling application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and other. This project has taken the data of previous customers of various banks to whom on a set of parameters loan were approved. So the machine learning model is trained on that record to get accurate results. Our main objective of this project is to predict the safety of loan. To predict loan safety, the SVM and Naïve bayes algorithm are used. First the data is cleaned so as to avoid the missing values in the data set.

### **PROBLEM STATEMENT**

Finance companies, banks are deals with different kinds of loans such as education loan, shop loans, home loans, personal loans etc all are part of our country loan types. All the companies and banks are present in villages, towns, cities. After customer apply for loan these banks/companies want to validate the customer details for that candidate eligible for loan or not. The main purpose of the system is applicant loan approved or not based on train models.

## **OBJECTIVE**

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.

## **MOTIVATION**

Loan approval is a very important process for banking organizations. The system approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. Using Machine learning we predict the loan approval.



## **CHAPTER 2**

### **LITERATURE SURVEY**

1." Loan Approval Prediction based on Machine Learning Approach"  
Author- Kumar Arun, Garg Ishan, Kaur Sanmeet Year- 2018  
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

2.“Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process” Author- E. Chandra Blessie, R. Rekha -  
Year- 2019  
Extending credits to corporates and individuals for the smooth functioning of growing economies like India is inevitable. As increasing number of customers apply for loans in the banks and non-banking financial companies (NBFC), it is really challenging for banks and NBFCs with limited capital to devise a standard resolution and safe procedure to lend money to its borrowers for their financial needs. In addition, in recent times NBFC inventories have suffered a significant downfall in terms of the stock price. It has contributed to a contagion that has also spread to other financial stocks, adversely affecting the benchmark in recent times. In this paper, an attempt is made to condense the risk involved in selecting the suitable person who could repay the loan on time thereby keeping the bank's nonperforming assets (NPA) on the hold. This is achieved by feeding the past records of the customer who acquired loans from the bank into a trained machine learning model which could yield an accurate

result. The prime focus of the paper is to determine whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting.

3. “Loan Prediction using machine learning model” Year2019whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting. 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.

4. “Loan Prediction using Decision Tree and Random Forest” Author- Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar Year-2020. In India the number of people or organization applying for loan gets increased every year. The bank has to put in a lot of work to analyse or predict whether the customer can pay back the loan amount or not (defaulter or non-defaulter) in the given time. The aim of this paper is to find the nature or background or credibility of client that is applying for the loan. We use exploratory data analysis technique to deal with problem of approving or rejecting the loan request or in short loan prediction. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.

## **CHAPTER 3**

### **PROPOSED MODEL**

This system predict whether the loan is approve or reject . This System refers the following things or ways.

Data Collection Data Pre-processing (Data Cleaning)

Model Selection

Model Evaluation

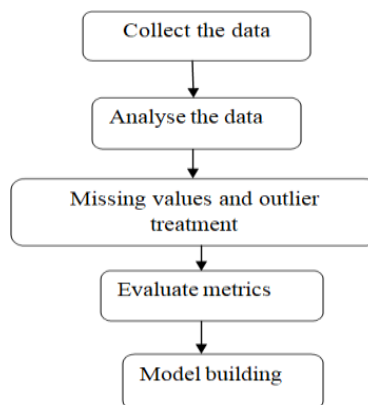
Classification Result (output)

## CHAPTER 4

### SYSTEM ARCHITECTURE

**A. EXPERIMENT OVERVIEW:** In this experiment firstly collect the data and understand the data with help of (.describe()) and then analyses of data then search for any missing/null/nosy data present in the dataset and then evaluate the confusion matrices(accuracy, precision, recall, f1-score) and finally model building i.e., used methods Procedures are designed to detect errors in data at a lower level of detail. Data validations have been included in the system in almost every area where there is a possibility for the user to commit errors. The system won't accept invalid information. Whenever invalid information is keyed in, the system like a shot prompts the user and also the user should once more key within the information and also the system will accept for the info provided that the info is correct. Validations are enclosed wherever necessary. The system is designed to be a user friendly one. In alternative words the system has been designed to speak effectively with the user. The system has been designed with popup menus.

#### OVERVIEW OF EXPERIMENT



#### *4.1 Major Attributes*

## **IMPLEMENTATION DETAILS (MODULES):**

**4.1. Loan Dataset :** Loan Dataset is very useful in our system for prediction of more accurate result. Using the loan Dataset the system will automatically predict which costumer's loan it should approve and which to reject. System will accept loan application form as an input. Justified format of application form should be given as an input to get processed.

**4.2. Determine the training and testing data:** Typically , Here the system separate a dataset into a training set and testing set ,most of the data use for training ,and a smaller portions of data is use for testing. after a system has been processed by using the training set, it makes the prediction against the test set.

**4.3. Data cleaning and processing:**InData cleaning the system detect and correct corrupt or inaccurate records from database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing , modifying or detecting the dirty or coarse data. In Data processing the system convert data from a given form to a much more usable and desired form i.e. make it more meaningful and informative.

### **4.4. Models used :**

#### **1.RANDOM FOREST :**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

## **2.DECISION TREE:**

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.

In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.

A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees. The feature in the dataset include

- 1. Application\_Id
- 2. Gender
- 3. Marital Status

- 4. Number of dependents
- 5. Educational Profile
- 6. Employment Status
- 7. Applicant's Income
- 8. Co-Applicant's Income
- 9. Loan Amount
- 10. Credit History
- 11. Loan Status

## **EXPLORATORY DATA ANALYSIS SYSTEM**

verify the documents and forward the details to loan evaluator for approval or rejection. System approve the loan if documents are cleared and reject the loan if documents are not cleared Report is delivered to the applicant according to their status.



## **CHAPTER 5**

### **PROPOSED ALGORITHM**

The following shows the pseudo code for the proposed loan prediction method

1. Load the data
2. Determine the training and testing data
3. Data cleaning and pre-processing.
  - a) Fill the missing values with mean values regarding numerical values.
  - b) Fill the missing values with mode values regarding categorical variables.
  - c) Outlier treatment.
4. Apply the modelling for prediction
  - a) Removing the load identifier
  - b) Create the target variable (based on the requirement). In this approach, target variable is loan-status
  - c) Create a dummy variable for categorical variable (if required) and split the training and testing data for validation.
  - d) Apply the model: Random forest method, Decision tree method
5. Determine the accuracy followed by confusion Matrix

## **CHAPTER 6**

### **SYSTEM FEATURES**

- Data collection.
- Data cleaning and preprocessing
- Model selection
- Data verification
- Classification.
- Report deliver.

## **CHAPTER 7**

### **MATHEMATICAL MODEL**

Consider any decision problem, where for given number of inputs, decision oriented solution is available so our project is NP complete but some cases like not proper input format provided or if dataset not trained proper it's NP hard.

Let  $s$  be System :

$S=I, P, O$

$S$ : is a System

$I=I_1, I_2$

$P= DC, DP, DV, NBA, CL$

$O=RD$

$I_1$ : Loan Dataset

$I_2$ : Trained Dataset.

$DC$ : Data Cleaning

$D DP$ : Data Processing

$DV$ : Data Verification

$RFA$ : Random Forest Algorithm

$CL$ : Classification

$RD$ : Report Deliver Success

Condition : Proper features trained Dataset will give proper output

Failure Condition No Trained Dataset.

## **CHAPTER 8**

### **CONCLUSION**

In this paper, we have proposed customer loan prediction using supervised learning techniques for loan candidate as a valid or fail to pay customer. In this paper, various algorithms were implemented to predict customer loan. Optimum results were obtained using Logistic Regression, Random Forest, KNN, and SVM, decision Tree Classifier. Compare these five algorithms random forest is the high accuracy. From a correct analysis of positive points and constraints on the part, it can be safely ended that the merchandise could be an extremely efficient part. This application is functioning properly and meeting to all or any Banker necessities. This part is often simply obstructed in several different systems. There are numbers cases of computer glitches, errors in content and most significant weight of option is mounted in machine-driven prediction system, therefore within the close of future the therefore called software system might be created more secure, reliable and dynamic weight adjustment. In close to future this module of prediction can be integrated with the module of machine-driven processing system.

### **FUTURE SCOPE**

The system is trained on old training dataset in future software can be made such that new testing data should also take part in training data after some fix time.

## CHAPTER 9

### REFERENCE

- 1 Yu Jin and Yudan Zhu, “A data-driven approach to predict default risk of loan for online Peer-to-Peer (P2P) lending,” School of Information, Zhejiang University of Finance and Economics, 310018 Hangzhou, China.
- 2 <https://www.kaggle.com/telco-churn>
- 3 Bhoomi Patel, Harshal Patil, Jovita Hembram, Shree Jaswal “Loan default forecasting using data mining” Department of Information Technology, St. Francis Institute of Technology, Mumbai, India (2020)
- 4 Octave Iradukunda, Haiying Che, Josiane Uwineza, Jean Yves Bayingana, Muhammad S Bin-Imam, Ibrahim Niyonzima ”Malaria Disease Prediction Based on Machine Learning” School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China (2019).
- 5 G. Arutjothi, Dr. C. Senthamarai “Prediction of Loan Status in Commercial Bank using Machine Learning Classifier” department of computer applications government arts college (Autonomous) Salem, India (2017.)
- 6 Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar “An Approach for Prediction of Loan Approval using Machine Learning Algorithm” School Of Computer Science And Engineering Galgotias University Greater Noida, India (2019).
- 7 Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li “Overdue Prediction of Bank Loans Based on LSTM-SVM”Jiangsu Key Lab of Big Data and Security and Intelligent Processing Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.
- 8 Aakanksha, Tamara Denning, Vivek Srikumar, Sneha Kumar Kesera “secrets in source code: reducing false positives using ML” software engineering (Microsoft) school of computing, USA (2020)
- 9 Arutjothi .G, Dr. C. Senthamarai. "Credit Risk Evaluation using Hybrid

Feature Selection Method. Software engineering and technology (2017)

- 10 Ch. Balayesu and S Narayana, "An Improved Algorithm for Efficient Mining of Frequent Item Sets on Large Uncertain Databases" in International Journal of Computer Applications, Volume 73, No. 12 July 2013, Page No. 8-15.
- 11 Bala brahmeswara kadam et al."A novel ensemble decision tree classifier using hybrid feature selection measures for parkinson's disease prediction", Int. J. Data science (IJDS), ISSN: 2053-082X, Vol.3, No.4,2018.
- 12 Mrunal Surve, Pooja Thitme, Priya Shinde, Swati Sonawane, and Sandip Pandit. "Data mining techniques to analyze risk giving loan (bank)" International Journal of Advance Research and Innovative Ideas in Education Volume 2 Issue 1 2016 Page 485-490