

---

# Prediction for Loan Approval using Machine Learning Algorithm

---

**Abstract** - *In our banking system, banks have many products to sell but main source of income of any banks 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. 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. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters (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.*

**Key Words:** Big data, Machine Learning, SVM, Naïve Bayes, Prediction.

## 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.

---

## 1.1 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.

## 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" Year-2019 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 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 have 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.

### 3. PROBLEM DEFINITION

Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries. These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validates the eligibility of customers to get the loan or not. This project provides a solution to automate this process by employing machine learning algorithm. So the customer will fill an online loan application form. This form consist details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others.

### 3. IV. 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)

### 4. SYSTEM ARCHITECTURE

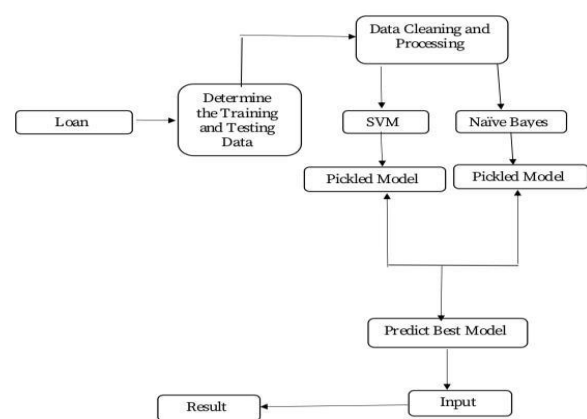


Fig -1: Loan Prediction Architecture

### 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:** In Data 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) SVM:

In this approach, each data item is plotted in a n-dimensional space, where n represents the number of features with each feature represented in a corresponding co-ordinates. A hyper plane is determined to distinguish the classes (possibly two) based on their features.

##### 2) Naïve Bayes (NB) Model :

The basis for NB model is Bayes Theorem (BT), where events are mutually exclusive similar to rolling a die. Moreover, the BT presumes that the input features also referred as predictors are independent in nature. Similarly, NB also presumes that the input features are independent in nature. But, this is impossible in the realistic procedures. Since this assumption leads to naïve, this algorithm is termed as Naïve Bayes algorithm. Thus, NB is a probabilistic algorithm, where the conditional probability is determined regarding the input features. On the other hand, during the dependent input features scenario, conditional probability is calculated twice resulting in improper results. Hence, for better prediction results with respect to NB model, independent input features are selected and processed. dataset collected from Kaggle source. 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

#### 4.5. 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.

#### 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 loan 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: NB method, SVM method

5. Determine the accuracy followed by confusion Matrix.

#### 5.1. SYSTEM FEATURES

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

#### 6. 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

$I = I1, I2$

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

$O = RD$

I1: Loan Dataset

I2: Trained Dataset.

DC: Data Cleaning

D DP: Data Processing

DV: Data Verification

NBA: Naïve Bayes Algorithm

CL: Classification

RD: Report Deliver Success

Condition : Proper features trained Dataset will give proper output

Failure Condition No Trained Dataset.

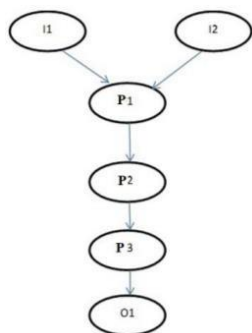


Fig -2: Mathematical model

## 7. USE CASE DIAGRAM

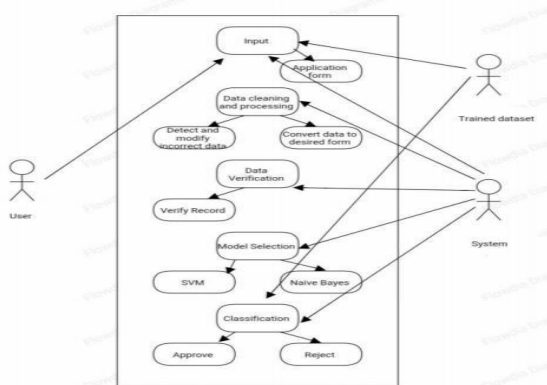


Fig -3: use case diagram

## CONCLUSION

So here, it can be concluded with confidence that the Naïve Bayes model is extremely efficient and gives a better result when compared to other models. It works correctly and fulfills all requirements of bankers. This system properly and accurately calculate the result. It predicts the loan is approve or reject to loan applicant or customer very accuratly.

## REFERENCES

1) Kumar Arun, Garg Ishan, Kaur Sanmeet, —Loan Approval Prediction based on Machine Learning Approach||, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).

2) Aboobyda Jafar Hamid and Tarig Mohammed Ahmed, —Developing Prediction Model of Loan Risk in Banks using Data Mining||, Machine Learning and Applications: An International Journal (MLAIJ), Vol.3, No.1, pp. 1-9, March 2016.

3) S. Vimala, K.C. Sharmili, —Prediction of Loan Risk using NB and Support Vector Machine||, International Conference on Advancements in Computing Technologies (ICACT 2018), vol. 4, no. 2, pp. 110-113, 2018.

4) Pidikiti Supriya, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, kVikash,“Loan Prediction by using Machine Learning Models”, InternationalJournalofEngineering andTechniques.Volume 5 Issue 2, Mar-Apr 2019

5) Nikhil Madane, Siddharth Nanda,“Loan Prediction using Decision tree”, Journal of the Gujrat Research History, Volume 21 Issue 14s, December 2019.