**Project Requirement and Specification**

on

**Loan prediction using machine learning**

(CSE 6th Semester Mini project )

2022



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

# GRAPHIC ERA UNIVERSITY, DEHRADUN

University Roll no.: Submitted by:

2014407 JITENDRA SINGH

Section:

G

**---------------------------CONTENTS-----------------------**

**1.1ABOUT PROJECT--------------------------------------**

**1.2 MOTIVATION-----------------------------------------------------------**

**1.3 SYSTEM REQUIREMENT -------------------------**

**1.4 Model Used --------------------------------------**

# 1.5 Algorithm used ------------------------------------------

## 1.6 System Architecture -------------------------------

## 1.7 Library Used----------------------------------------------

**1.8 Use Case Diagram-----------------------------------------**

**1.9 Accuracy of different types of algorithm----------------**

**1.10 Conclusion -------------------------------------------**

**1.11 References -------------------------------------------**

----------------------------------------------------------------------

### 1.1About Project

### 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.

### 1.2 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.

### 1.3 SYSTEM REQUIREMENT

**1.3.1Hardware Requirement:**

* **PROCESSOR: INTEL® PENTIUM® CPU @ 2.0 GHZ**
* **RAM: 2 GB(MINIMUM)**
* **STORAGE: 50MB(MAXIMUM)**
* **OS: WINDOWS 10,8,7**

**1.3.2Software Requirement:**

* **PYCHARM**
* **PYTHON**
* **Jupyter Notebook**

**1.4 Models used**

**1)** **SVM**: In this approach, each data item is plotted in a ndimensional 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 aspredictors 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. Aplication\_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

# 1.5 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: NB method, SVM method 5. Determine the accuracy followed by confusion Matrix.

**1.6 SYSTEM ARCHITECTURE**

**Diagram

Description automatically generated**

**1.7 Libraries Used:**

1). Pandas

1). Numpy

3). Sklearn

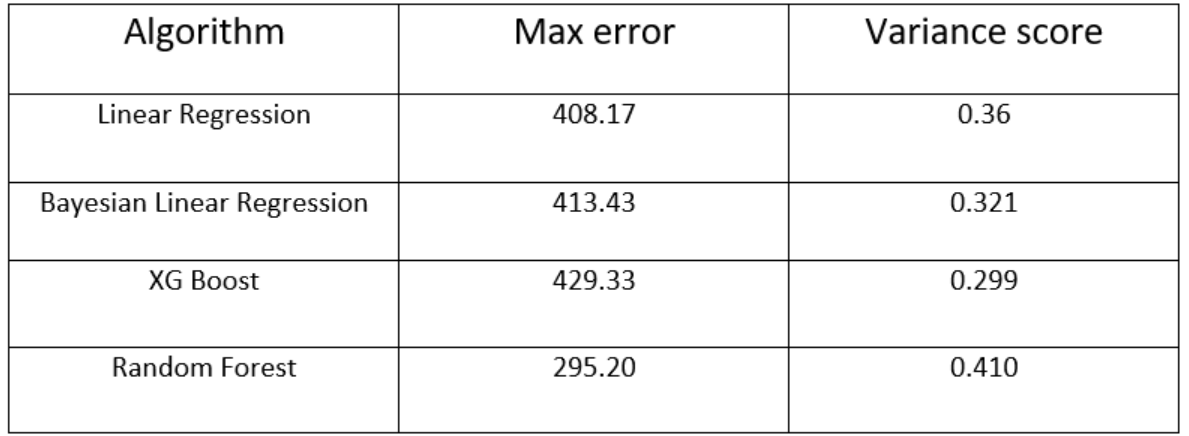
4). Seaborn

**1.8 USE CASE DIAGRAM**

**Diagram

Description automatically generated**

**1.9 Accuracy of different types of algorithm**

****

**1.10 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

# 1.11 REFERENCE

1. GEEKS FOR GEEKS.
2. W3 SCHOOL.
3. STACK OVERFLOW.

4. Machine learing coursera

**Source code is available at my Github Account:**

<https://github.com/Jitendra-singh-123/All-projects/tree/main/Loan_prediction>

