Project Report

Loan Eligibility Prediction

# INTRODUCTION

# Based on the data provided by the applicant, a process known as loan prediction enables you to apply for loans and receive alerts when they are granted. The applicant is informed by the system of the loans that are accessible. In order to solve this classification challenge, we must decide whether to approve or refuse the loan. To develop the loan approval prediction system, we will use a variety of classification methods, including Logistic Regression, SVM, Naive Bayes, Decision Tree, Random Forest, XG boost, Artificial Neural Networks, etc. Each retail bank employee must deal with the very common problem of loan projection at some point in their careers. At the conclusion of a retail bank, it can save a lot of man hours if done effectively.

# METHODOLOGY

1. Jupyter Lab is used to run the code after the data has been loaded using Tensorflow. All libraries and code are implemented using Python.
2. Support vector machine(SVM) algorithm is used for loan prediction.
3. The Silhouette score determines the clusters' quality, with a higher value suggesting better clusters in the characteristics.

**About the Dataset**

We have used the dataset which is downloaded from Kaggle website. It includes the following attributes.

Table

Description automatically generated

**Loading the libraries**

We have loaded the essential libraries, which are required for our model.

Graphical user interface, text, application

Description automatically generated

**Size of the Dataset**

We have 614 rows and 13 columns in our data.

Table

Description automatically generated

# Preprocessing

Data cleaning has been done to our dataset, where we found many missing values in the data and filled the data set with different samples. We have also calculated the percentage of missing values in different columns such as the percentage of missing values in gender column is 2.12%

# Model Description

In this step we have searched for different methods for building our model and tried with the methods which were best for our project. We have used different methods in our model such as Random Forest, K neighbor classifier etc.

**Random Forest**

Random forest is a classification algorithm where it consists of many decision trees and uses techniques such as bagging and feature randomness for building each individual tree and create an uncorrelated forest of trees whose prediction is more accurate than an individual tree.

**K- Nearest Neighbors Algorithm**

KNN (K- Nearest Neighbor) Algorithm is a supervised classifier, non-parametric which use proximity method to make the predictions and classifications about the grouping of an individual point.

**Gradient Boosting**

Gradient boosting method is a machine learning method which relies on the intuition that the best possible model which is combined with all previous models gives the less prediction error. The main idea is to minimize the prediction error.

**Decision Tree**

A decision tree is a type of classifier in machine learning which is used to categorize and make predictions using the previous data or the answers of previously answered questions.

**Support Vector Machine**

Support vector machine (SVM) is a supervised learning model which helps in classification algorithms for 2 group classification problems. When we give a set of labeled data SVM is capable of classifying or categorize that data into new text.

# Experiment and Result

# Training and testing logs

We have trained all the columns as shown below

Text

Description automatically generated

The output of the data after training and testing in a graphical representation is given:

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

For methods having numerical outputs the graphs is shown as:

Chart, histogram

Description automatically generated

**Results**

We got different accuracy for different techniques which we used in our project.

* Accuracy for Gradient Boosting Technique

Text

Description automatically generated with medium confidence

* Accuracy for Random Forest Classifier

A picture containing text

Description automatically generated

* Accuracy for Decision Tree Classifier

A picture containing text

Description automatically generated

* Accuracy for K Neighbor Classifier

Text

Description automatically generated with medium confidence

# Discussion and comparison

As we can see that using different classifier gave different accuracies. Using different classifiers helps in selection of perfect method of model. However, the project can be done with many more methods too, such as, Logistic Regression, but we selected these 4 models because it was more suitable to the project and also the implementation of these models were easy when compared to other models.

# CONCLUSION

# We can predict the eligibility of a person getting loan or not using this model, the accuracies of those models are shown above and according to those accuracies we can say Gradient boosting techniques is the best option for the model because it has the highest accuracy i.e., 78.17%

# REFERENCES

* <https://www.geeksforgeeks.org/loan-approval-prediction-using-machine-learning/>
* <https://www.kaggle.com/code/vikasukani/loan-eligibility-prediction-machine-learning>
* <https://www.analyticsvidhya.com/blog/2022/02/loan-approval-prediction-machine-learning/>
* <https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057>

**By:**

**Vedvikas Reddy Magam**