Prediction of Loan Approval by Machine Learning with Detailed Analysis Graphs/Figures and Reports

Lovely Professional University

Requirement Specification

Version 1.0

Developed By Anurag Mittal(12209102)
Sukhdeep Kaur(12209123)
Ekhlakh Ahmad(12209166)

Table of Contents

1. Introduction

This project aims to develop a machine-learning model to predict the approval of loan applications. The model will be trained on a historical dataset of loan applications, including information about the applicant's credit history, income, employment status, and other relevant factors. Once the model is trained, it can be used to predict the likelihood of approval for new loan applications.

2. Requirements

- The model must be able to predict loan approval with an accuracy of at least 80%.
- The model must be able to run on a standard laptop or desktop computer.
- The model must be able to be easily updated with new data as it becomes available.

3. Design

The model will be implemented using a random forest algorithm. Random forests are a type of ensemble learning algorithm that combines the predictions of multiple decision trees to produce a more accurate prediction.

The model will be trained on a historical dataset of loan applications. The dataset will include information about the applicant's credit history, income, employment status, and other relevant factors. The target variable will be whether the loan was approved.

Once the model is trained, it can be used to predict the likelihood of approval for new loan applications. The model will take the applicant's information as input and produce a probability score between 0 and 1, where 1 indicates that the loan is likely to be approved and 0 indicates that the loan is unlikely to be approved.

4. Testing

The model will be tested on a held-out test set of loan applications. The test set will not be used to train the model, so it will provide an unbiased evaluation of the model's performance.

The accuracy of the model will be measured by the percentage of loan applications that the model correctly predicts. The model will also be evaluated using other metrics, such as precision, recall, and F1 score.

5. Deployment

The model will be deployed as a web service. The web service will allow users to submit loan applications and receive a prediction of the likelihood of approval.

6. Maintenance

The model will be updated regularly with new data as it becomes available. This will help to ensure that the model remains accurate and up to date.

Detailed Analysis

The following is a detailed analysis of the prediction of loan approval using machine learning:

Data Collection

The first step in developing a machine learning model is to collect data. The data should be representative of the population that the model will be used to predict. In this case, the data should include information about loan applicants, such as their credit history, income, employment status, and other relevant factors.

Data Preparation

Once the data has been collected, it needs to be prepared for machine learning. This may involve cleaning the data, removing outliers, and converting the data to a format that is compatible with the machine learning algorithm.

Feature Selection

Not all features in the data will be equally important for predicting loan approval. Therefore, it is important to select the features that are most relevant to the prediction task. This can be done using a variety of methods, such as correlation analysis or information gain.

Model Training

Once the features have been selected, the next step is to train the machine learning model. This involves feeding the model the training data and allowing it to learn the relationship between the features and the target variable.

Model Evaluation

Once the model has been trained, it needs to be evaluated on a held-out test set. This will give an unbiased estimate of the model's performance on new data.

Model Deployment

Once the model has been evaluated and found to be satisfactory, it can be deployed to production. This may involve integrating the model into a web service or mobile app.

Graphs/Figures and Reports

The following are some graphs/figures and reports that can be used to analyze the prediction of loan approval using machine learning:

- Confusion matrix: A confusion matrix is a table that shows the number of correctly and incorrectly predicted loan applications.
- ROC curve: An ROC curve shows the trade-off between sensitivity and specificity. Sensitivity is the percentage of loan approvals correctly predicted, and specificity is the percentage of loan denials correctly predicted.

- PR curve: A PR curve shows the trade-off between precision and recall. Precision is the percentage of predicted loan approvals that are approved, and recall is the percentage of actual loan approvals that are predicted.
- Feature importance: A feature importance plot shows the relative importance of each feature for predicting loan approval.
- Model performance report: A model performance report should include accuracy, precision, recall, and F1 score metrics.

Conclusion

Machine learning can be used to develop accurate and reliable models for predicting loan approval.