Internship Report: Peer-to-Peer Lending Platform Data Analysis

Introduction

Welcome to the README file for my internship report on the analysis of Peer-to-Peer (P2P) lending platform data. In this report, I will provide an overview of the project, its objectives, the approach taken, key steps, and outcomes achieved.

Project Overview

This project focuses on analyzing data from a Peer-to-Peer lending platform. The goal is to gain insights into the dynamics of the lending industry and develop predictive models for Loan Acceptance, Return on Investment (ROI) prediction, and Estimated Monthly Installment (EMI) prediction.

Problem Statement

The main objectives of this project are as follows:

Develop a predictive model for Loan Acceptance to assist investors in making informed lending decisions.

Build a model to predict the ROI for investors, aiding them in portfolio management.

Create a model to estimate borrowers' EMIs, facilitating financial planning.

Approach

The project was carried out using the following structured approach:

Data Exploration: Understand the dataset from the P2P lending platform, identify missing values, and preprocess the data.

Data Visualization: Visualize data patterns and relationships to gain insights.

Feature Engineering: Select relevant features for predictive modeling using mutual information scores.

Modeling: Develop models for Loan Acceptance, ROI prediction, and EMI prediction.

Model Evaluation: Assess model performance using appropriate metrics and cross-validation techniques.

Deployment: Deploy the best-performing models for real-world usage.

Libraries

The following libraries were used for analysis and modeling:

pandas

numpy

matplotlib

seaborn

scipy.stats

plotly.express

pickle

sklearn.preprocessing

sklearn.model\_selection

sklearn.feature\_selection

sklearn.linear\_model

sklearn.decomposition

sklearn.metrics

sklearn.ensemble

sklearn.pipeline

Data Exploration

Loaded the dataset from a CSV file.

Displayed initial rows of the dataset.

Checked data types and information of columns.

Described the statistical summary of the dataset.

Preprocessing

Identified and handled missing values.

Managed outliers using winsorization.

Filled null values in specific columns.

Applied data encoding using label encoding and one-hot encoding.

Exploratory Data Analysis

Conducted various analyses and visualizations to gain insights into the dataset, including credit grades distribution, loan status relationship, borrower rate distribution, correlation analysis, and more.

Feature Engineering

Selected relevant features for predictive models using mutual information scores.

Modeling

Built a Logistic Regression model for Loan Acceptance Prediction.

Developed Linear Regression and Random Forest Regression models for ROI and EMI prediction, respectively.

Model Evaluation & Cross-validation

Evaluated model performance using accuracy, confusion matrix, precision, recall, F1-score, and mean absolute error. Conducted cross-validation to assess generalization.

Deployment

Saved and loaded trained models for Loan Acceptance, ROI, and EMI prediction.

Conclusion

This project successfully achieved the development of predictive models for Loan Acceptance, ROI, and EMI prediction. These models can significantly impact decision-making for investors and borrowers, providing insights and estimates to enhance the lending process.