Details

Title: Alternative Credit Scoring Prediction Model Using Machine Learning
Team Name: Pioneers (Team 75)
Team Members: 23BAI1576 (Parun Sethupathy), 23BAI1544 (Aayush Sundaresan), 23BAI1535 (Jonnalagadda Aditya)

Product Explanation

A production pipeline ML model trained to predict loan approval in a binary format. It merges three datasets, removes leakage, creates simple ratios and trains a tuned Random Forest. The final product is a refined prediction model that chooses to either accept or reject a loan applicant based on non traditional requirements,

Data and Labels

Federal Loan Data: https://www.federalreserve.gov/data.htm

Household Data:
https://www.census.gov/#:~:text=Get%20in%20the%20weeds%20with,Bureau's%20premiere%20data%20dissemination%20platform.

Income Data:
https://www.statista.com/statistics/203183/percentage-distribution-of-household-income-in-the-us/

Labels:
[ApplicantIncome, A15, Married, Property_Area, Education, A7, TotalIncome, DTI, LoanAmount, CoapplicantIncome]

Test-Training Split

- Stratified 75/25 train/test (keeps class balance).
- Train: 1,111 rows | Test: 371 rows
- Approved rate: ~83.5% in both splits.

Formulae

Accuracy - Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision_Approved = TP / (TP + FP)
Recall_Approved = TP / (TP + FN)
F1_Approved = 2 * Precision_Approved * Recall_Approved
/ (Precision_Approved + Recall_Approved)

Precision_NotApproved = TN / (TN + FN)
Recall_NotApproved = TN / (TN + FP)
F1_NotApproved = 2 * Precision_NotApproved * Recall_NotApproved
/ (Precision_NotApproved + Recall_NotApproved)

Threshold Value - 0.62