

Loan Amount Prediction - Problem Statement

A financial services company wants to improve its loan disbursement strategy by predicting the expected loan amount for new applicants based on their financial and employment profiles. The goal is to build a machine learning model that can estimate loan amounts accurately and fairly, helping the company:

Assess applicant eligibility Optimize loan offerings Reduce manual underwriting effort

You are provided with a synthetic dataset containing applicant details such as income, employment length, credit history, and loan purpose. Your task is to:

Clean and preprocess the data Engineer meaningful features Analyze distribution and multicollinearity Train and compare regression models (Linear, Ridge, Lasso)

Select the best-performing model based on R^2 and RMSE Build a prediction pipeline that generalizes well to new applicants

The final model should be interpretable, scalable, and suitable for integration into a client-facing loan recommendation system.