

Credit Risk Prediction - Technical Report

GenAI Capstone Project - NST Sonipat

1. Problem Statement

Access to credit is a fundamental enabler of economic participation, yet lending institutions face the persistent challenge of distinguishing creditworthy borrowers from those likely to default. Inaccurate risk assessments lead to significant financial losses through non-performing assets or, conversely, the denial of credit to deserving applicants.

This project addresses the binary classification problem of predicting loan default risk. Given a set of applicant and loan attributes, the goal is to build a supervised machine learning model that predicts whether a borrower will default on a loan (status=1) or repay successfully (status=0).

2. Data Description

The dataset contains 32,581 records of historical loan applications. Key features include applicant age, income, employment length, home ownership, loan intent, loan grade, loan amount, interest rate, and credit history length.

3. Exploratory Data Analysis (EDA)

- Loan-to-Income Ratio: Borrowers with high ratios show significantly higher default risk.
- Interest Rates: Defaulters are concentrated in higher rate brackets.
- Loan Grades: Default rates increase progressively from Grade A to Grade G.

4. Methodology

We utilized a Decision Tree Classifier with balanced class weights. Preprocessing included median imputation for missing values, ordinal encoding for grades, and StandardScaler normalization.

5. Evaluation

Metric	Decision Tree Result
Accuracy	90.81%
ROC-AUC	0.6817
Recall (Default)	0.77

6. Team Contribution

Member	Contribution
Palak	Data preprocessing, EDA, model training
Samarth	App development, Deployment, UI/UX

Credit Risk Prediction - Technical Report

GenAI Capstone Project - NST Sonipat

7. Deployment

Live at: <https://genaicapstone-a7eipdbqudn2niewt9s2mp.streamlit.app/>