

Machine Learning for Loan Default Prediction

Abstract:

In the banking sector, loan default poses a significant financial risk and directly impacts profitability and operational stability. Accurately predicting loan default enables banks to make informed lending decisions and minimize potential losses. This project focuses on developing a machine learning-based loan default prediction system using customer financial, demographic, and transaction-related data. The proposed system follows a complete machine learning pipeline that includes data preprocessing, feature engineering, model training, and performance evaluation. Essential preprocessing steps such as handling missing values, encoding categorical variables, and feature scaling are applied to enhance data quality and model performance. Multiple classification algorithms, including Logistic Regression, Random Forest, and XGBoost, are implemented and compared to identify the most effective model for predicting loan default. Model performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, supported by visualizations including confusion matrices and ROC curves. The experimental results indicate that ensemble-based models outperform traditional approaches by effectively capturing complex patterns in customer behavior. This project demonstrates the practical application of machine learning techniques in the banking domain and provides a robust framework for credit risk assessment and decision support.

Team Members :

J.Sruthika - 23241A1293

Ch.Vaishnavy - 23241A1278

N.Karthik - 23241A12A5

