Predicting Loan Defaulting using Machine Learning

Ashhar Aziz
IIITD
ashhar21137@iiitd.ac.in

Lakshya Goel
IIITD
lakshya21469@iiitd.ac.in

Sanmay Sood
IIITD
sanmay21095@iiitd@ac.in

Srimant Mohanty
IIITD
srimant21207@iiitd.ac.in

Abstract
This project aims to predict the likelihood of people defaulting on loans. In this project, we will implement various algorithms on our data such as Random Forest, Logistic Regression, Decision Trees, Support Vector Machines and Multi-Layer Perceptron Classifiers.

1. Motivation
The rapid growth of India's banking sector has led to an upsurge in loan applicants, amplifying the challenge of accurately predicting loan defaults. Recent data from the Reserve Bank of India (RBI) reveals a concerning trend. Loan write-offs by banks surged to an alarming Rs 209,144 crore in the fiscal year ending March 2023 which is a 19.53% increase from FY22. Although existing strategies encompass collateral evaluation, income validation, and scrutiny of borrowing records, a transformative approach is imperative. Through harnessing the potential of machine learning (ML), we aim to formulate an advanced predictive model to bolster the precision of loan default prediction.

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

This innovation holds the potential to equip Indian banks with proactive measures against potential defaults, mitigating financial losses and fortifying the lending landscape.

2. Related Work
2.1 **Loan Default Prediction Techniques using Machine Learning**
2.2 **A Study in predicting Loan Default Risk using Random Forest Algorithm**
2.3 **Fairness Constraints in Loan Default Prediction**
2.4 **Learning Adversarially Fair and Transferable Representations**

3. Timeline
3.1 Pre-Mid Sem
Week 3-4: Data Visualisation, Data Preprocessing, Literature Review
Week 5-7: Logistic Regression, Decision Trees
3.2 Post Mid Sem
Week 9-10: MLP(ANN), SVM
Week 11-12: Random Forest (+XGBoost)
Week 13: Compiling the results, Conclusion and Final Report

4. Tasks (Intermediate + Final)
- Ashhar Aziz : Literature Review, Logistic Regression + MLP (ANN) + SVM
- Lakshya Goel : Data Visualization, Data Preprocessing, Exploratory Data Analysis, Outlier Detection and Removing + Random Forest (+XGBoost)
- Sanmay Sood : Literature Review, Decision Trees + Random Forest (+XGBoost)
- Srimant Mohanty : Data Visualization, Data Preprocessing, Exploratory Data Analysis, Outlier Detection and Removing + SVM

5. Final Outcome/Conclusion
Summing up the project concerning the prediction of loan defaults, the successful completion of this study will offer insights into the practical application of exploratory data analysis and machine learning to address genuine business challenges. Throughout this case study, we will also cultivate a foundational comprehension of risk analytics within the realms of banking and financial services. This endeavor will provide us with a clear understanding of how data is harnessed to mitigate the potential financial losses associated with lending to customers.