

CSE/ECE 343: Machine Learning Project Proposal
Title: Loan Default Prediction for Loss minimization

Ayush Garg

ayush19030@iiitd.ac.in

Pranansh Goyal

pranansh19073@iiitd.ac.in

Vasudev Singhal

vasudev19126@iiitd.ac.in

Vikrant

vikrant19398@iiitd.ac.in

1 Motivation

With the improving banking sector in recent times and the increasing trend of taking loans, a large population applies for bank loans. But one of the major problems banking sectors face in this ever-changing economy is the increasing rate of loan defaults, and the banking authorities are finding it more difficult to correctly assess loan requests and tackle the risks of people defaulting on loans. The two most critical questions in the banking industry are: (i) How risky is the borrower? and (ii) Given the borrower's risk, should we lend him/her?.

2 Related Work

1. A study on predicting loan default based on the random forest algorithm [1] by Lin Zhua , Dafeng Quia used the method of Random forest supervised learning algorithm to predict the loan defaults on P2P lending based on the data of 3 million and investment of over 50 million USD.
2. Prediction of Loan Approval using Machine Learning [2] by Rajiv Kumar, Vinod Jain used the various algorithms and the many classification learning techniques, comparing, analyzing, and improving the results and the accuracy of the model. Various methods were also applied to increase the quality of the available lending data.
3. Loan default prediction using decision trees and random forest: A comparative study [3] by Mehul Madaan, Aniket Kumar used the algorithms like J48, Bayes Net , Naive Bayes computing the accuracy and the mean absolute error value for each and employed other data mining techniques to classify the loan risk .

3 Timeline

- Week 1 : Topic & Data Exploration.
- Week 2 : Data collection and scraping.
- Week 3: Pre-processing and Data Visualization.
- Week 4 : EDA, Feature Extraction and Analysis
- Week 5 :Feature Selection, Correlation, HeatMaps.
- Week 6 :K- Nearest Neighbours ,Linear Regression model.
- Week 7 :Logistic Regression , Gaussian Naive Bayes
- Week 8 : Decision Trees , Random Forest.
- Week 9 : Linear SVM , XG boost
- Week 10 : Neural networks, Analysis of Model
- Week 11 : Hyperparameter Tuning , Model fit
- Week 12 : Report Writing.

4 Individual Tasks

Tasks	Team Member/s
Topic & Data Exploration.	Ayush, Vasudev
Pre-processing and Data Visualization.	Ayush, Vasudev, Vikrant, Pranansh
EDA, Feature Extraction and Analysis	Ayush , Pranansh
Feature Selection, Correlation, HeatMaps.	Vasudev, Pranansh
K- Nearest Neighbours ,Linear Regression model.	Vasudev, Vikrant
Logistic Regression , Gaussian Naive Bayes	Ayush, Vasudev
Decision Trees , Random Forest.	Pranansh, Vikrant, Vasudev
Linear SVM , XG boost	Ayush , Pranansh
Convolutional Neural networks, Analysis of Model	Vasudev ,Vikrant
Hyperparameter Tuning , Model fit	Ayush, Vikrant, Vasudev , Pranansh

5 Final Outcome

This project aimed to explore, analyze, and build a machine-learning algorithm to correctly identify whether a person, given certain attributes, has a high probability to default on a loan.Lending Club must be careful when identifying potential borrowers who fit certain criteria. In light of the above problems, we aim to propose machine learning models to predict whether an individual should be given a loan by assessing certain attributes and therefore help the banking authorities by easing their process of selecting suitable people from a given list of candidates who applied for a loan. We look forward to increasing the accuracy of the ML model by analysing various features provided through the dataset and try to refine the available datasets to include the main features affecting the output the most.

References

[1] A study on predicting loan default based on the random forest algorithm by Lin Zhua , Dafeng Quia.

[2] Prediction of Loan Approval using Machine Learning by Rajiv Kumar, Vinod Jain

[3] Loan default prediction using decision trees and random forest: A comparative study by Mehul Madaan, Aniket Kumar