
CONSUMER LENDING IN INDIA: PREDICTING LOAN DEFAULTS



Project Proposal

Prepared for: Metis Staff Instructors(Rita Biagioli, Brian McGarry)

Prepared by: Rahul Raju

August 25, 2021

EXECUTIVE SUMMARY

Opportunity

Over the past 20 years the economy of India has experienced unprecedented growth, increasing at an average annual rate of 6.50%. This has been largely led by the information technology boom that has created a shift in demography, an affluent middle class and growth in rural economies. Millennials and Gen-Z now have access to better education and employment opportunities, resulting in rising incomes and growing consumption habits. This in turn has led to a consumer credit boom with growth in credit averaging 15.00% per year over the past 2 decades. With this rise in credit comes great opportunity as well as great risk, particularly as the evolution in technology leads to greater financial inclusion in lower socio-economic classes. As such, India will need enhanced risk management techniques to ensure the continued growth of credit and well as the overall stability of its financial system.

Impact

Historically, legacy banks provided personal loans based on close business relationships as well as high value collateral. This lent itself to a very archaic and crude credit approval process. Recently, technological innovation in the finance industry has led to many smaller institutions offering unsecured loans. Data science methodologies could benefit both legacy and new firms alike in providing a more scientific approach to the credit approval process. This would allow for improved risk management, greater financial stability and more extensive access to credit markets across all socio-economic classes.

Solution

A classification model could provide for the ability to better predict consumer loan defaults based on sound statistical techniques. An array of features can be used to target whether an applicant will default on a loan or not. Some of the features included would be age, income, occupation etc. A variety of algorithms can be evaluated to determine which best serves the

respective firms's needs. One potential algorithm is logistic regression, which will provide a probability of a default from 0 to 1 based on a pre-defined threshold.

Data

The data set used in this analysis was found on Kaggle and is maintained by the University of Alabama. Below are the major characteristics of the data set:

- Number of Observations: 252,000
 - Loan Defaults(Positives): 30,996
 - No Loan Defaults(Negative): 221,004
- Number of Columns: 13
- Number of features: 11
 - Quantitative: 5
 - Categorical: 6

Tools

- 1). Data Cleaning & EDA: Pandas
 - 2). Visualizations: Seaborn, Matplotlib
 - 3). Statistical Analysis: Sci_kit Learn
-

Minimum Viable Product Vision

An MVP for this problem would consist of:

- 1). A list of all features included in the analysis
- 2). Insight into specific features
- 3). A baseline classification model
- 4). Appropriate evaluation metrics based on needs of the firm