Metis Classification Project – Sept 2021
Sam Reiff

Abstract

Household debt is growing at alarming rates in India and abroad. Using historical consumer credit data, the Indian government can proactively target citizens at high risk of loan default to offer debt relief programs, with the end goal of stabilizing population financial health. Due to limited government outreach resources, acute identification of the highest-risk individuals is a priority. Additionally, government agencies seek to optimize relief funds, and outreach to individuals not in need of relief should be minimized. Finally, given the staggering amount of consumer credit-related data on-hand, a computationally efficient model is preferable for incremental test observations for prediction.

Design

This analysis seeks to establish a meaningfully predictive classification model for identifying high-risk debtors (i.e., at higher risk of default). A model with acceptable accuracy, recall, precision, and F1/F-beta may be utilized for more informed outreach for debt relief for Indian consumers. Accurate classification could allow the Indian government (or any sovereignty assuming similar credit profiles/regimes) to appropriately dispense assistance to consumers in need of financial help or adice.

Data

This analysis leverages a public data set of >250k observations and 11 features, including income, car/house ownership, profession, and location for Indian citizens to classify each observation as defaulted or not defaulted (the target classes). Default rates for the entire sample approximate 12%, requiring class balancing to enhance model precision and recall. Feature engineering was limited to regularization of numerical values, and categorical features were dummified. I suspect that models employed in this analysis can be supplemented with additional information, such as monthly non-discretionary expenditures (e.g., total credit payments, number of children etc.).

Algorithms/tools

- Python Pandas and Numpy for data pre-processing
- Python Scikit Learn for model application
- Seaborn and Matplotlib for visualization

Communication

The findings of this exploratory analysis are principally communicated in the presentation associated with this document.