PPP - application rejection analysis:

Manikanta Koneru
mkoneru@umd.edu
Jayasree Karthik Nandula
snandul1@umd.edu
Team052 - IC2022 - UMCP

ABSTRACT:

Paycheck protection program is an SBA backed plan that helps keep their workforce employed during the covid 19 crisis. SBA has released data on more than 11.5M approved applications. Of these, some are removed from database. Through this project, analysis will be made on why those loans were removed.

Initial exploratory data analysis has rendered variables of considerable importance. Majority of the records removed belonged to Salons with number of employees retained as 1 (solo-proprietorship), belonging to 4 predominant counties in Georgia. Economic breakdown to small business occurred during March to June of 2021 evident with the number of PPP applications. Number of jobs retained are less than 100 where loan amount approved is less than 1M as per the removed applications. The lender Prestamos CDFI LLC has majority of the applications with status of disbursed but not paid.

Two machine learning models – decision tree and random forest were tested. Top 3 features are loan status, undisbursed loan amount, loan amount. Inclusion of these variables produces accuracy of 99.37 and 99.44 respectively. Exclusion of these variables (due to possible bias) produced an ROC curve that is just as good as the line of random guessing. There is scope for further analysis.

Demographic information from the PPP application form, not currently available in the dataset, if captured, could give better insights retaining accuracy and reducing bias.

Tools: Excel, Tableau, Python.