Springboard DSC–Capstone Project 2: Loan Default Prediction

Problem Statement

By: John Ryan

December, 2021

Summary

A financial institution is interested in estimating the probability a loanee will default on an auto loan from demographic data, loan information and credit history associated with historical loans.

Context

Financial institutions incur significant losses due to default on vehicle loans. In the United States 2021 auto loan debt in 2021 has risen to $1.42 trillion with about 5% annually being defaulted on. Accurate prediction of a future default could save lenders millions of dollars.

Criteria for Success

Classification models will be built, and compared according to appropriate classification performance metrics, which align with the business problem. Connections between variables and the target will be studied to determine how they influence the increase/decrease of the probability of default.

Scope of Solution Space

The variables for which models will be built will be limited to those in the dataset provided.

Constraints within Solution Space

None identified so far.

Stakeholders

Any bank or lender interested in minimizing default rate on auto loans.

Data Sources

https://www.kaggle.com/mamtadhaker/lt-vehicle-loan-default-prediction?select=train.csv

Dataset consists of 41 columns, 233k rows. Dataset has Loanee Information, Loan Information, Bureau data & history.