SMALL BUSINESS LOAN APPROVAL FORECAST

AGENDA

Overview of Problem

Overview of Datasets

Important findings from EDA

Baseline models and evaluation metrics

Next Steps

Overview of Problem

- High-interest rate environment
- Potential recession threat slower the rate cuts
- Lack of access to traditional commercial banks
- Proposed solutions and hope to build a model to help small businesses to have easier access to financing

Overview of Datasets

Lending club data 2007 to 2018

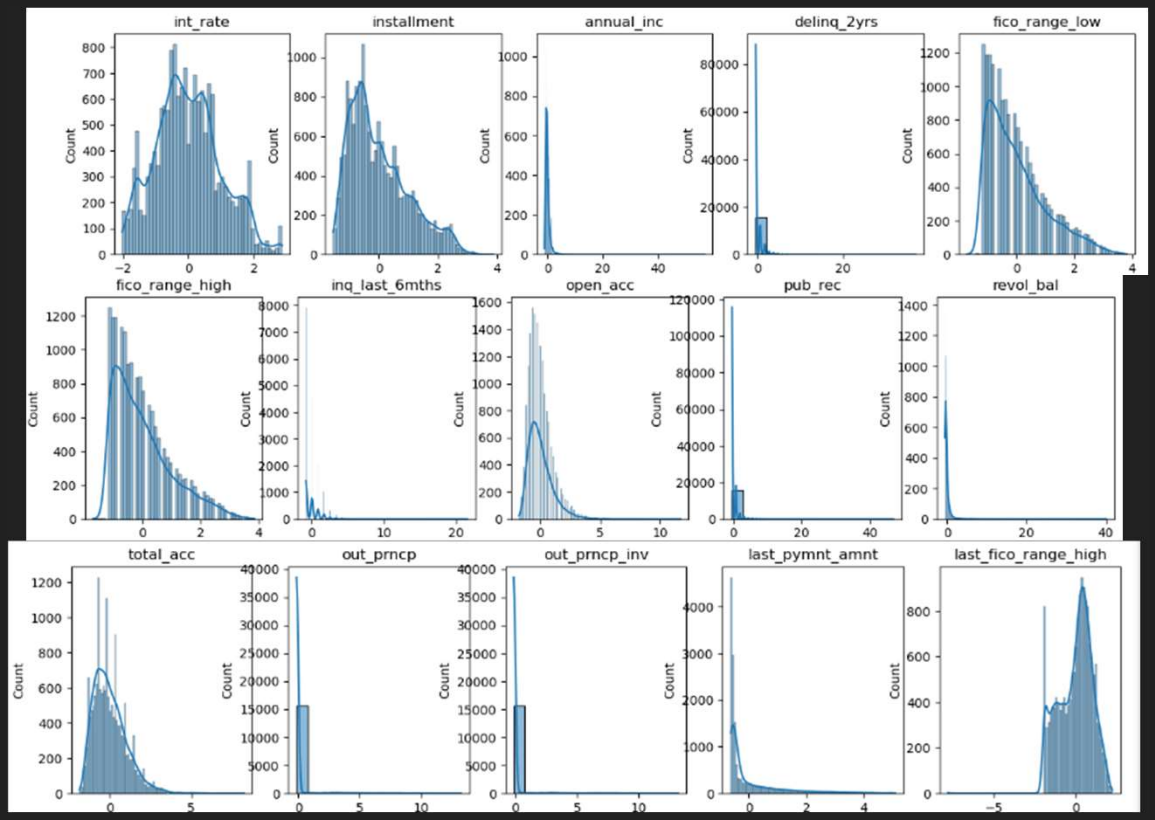
Only use the data related to Business use

Reduced rows from 2M to 20,000 rows.

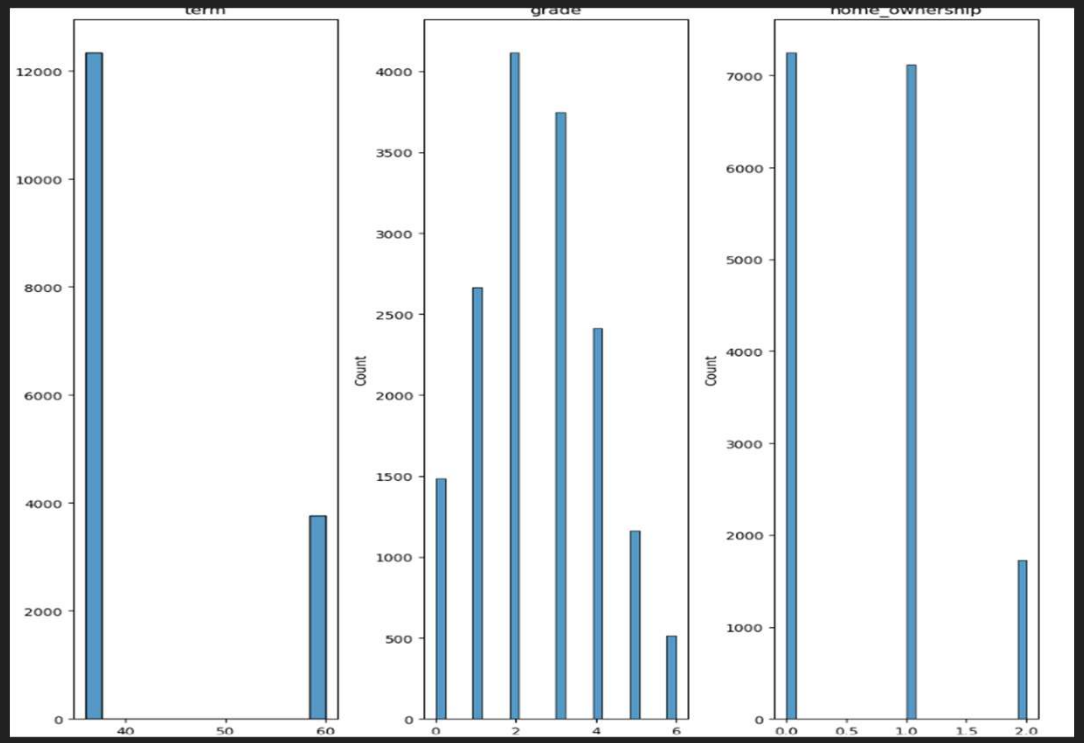
Reduced columns from 150 columns to 21 columns

Convert categorical columns to numerical columns

Overview of Datasets



Overview of Datasets



Scaling and Building Models

Random Forecast Result

Accuracy Score: 0.9142857142857143

Confusion Matrix:

[[894, 137], # Class 0

[139, 2050]] # Class 1

For loan status equal to 1(fully paid), the model accuracy is 94%, and for loan status as default/ not fully paid, the model accuracy is 87%.

ANN

Train Accuracy: 0.6784

Test Accuracy: 0.6798

Next Steps

Inbalance class distributions(Class 1: 10925 vs. Class 0 : 5273)

Add industry specific risk for interest rates

Product Demo, i.e. a web tool to let lender to decide if the loan will default.

THANK YOU

Abigail Deng