




Seriously Delinquent



April, Endalkachew, Matt, Ryan
March 23, 2021



Motivation

Financial institutions receive an overwhelming amount of credit card and loan applications. Going through each request manually is time consuming and prone to human error. With Machine Learning, we can build a model that can aid decision making for lenders and borrowers to predict potential future financial delinquency.

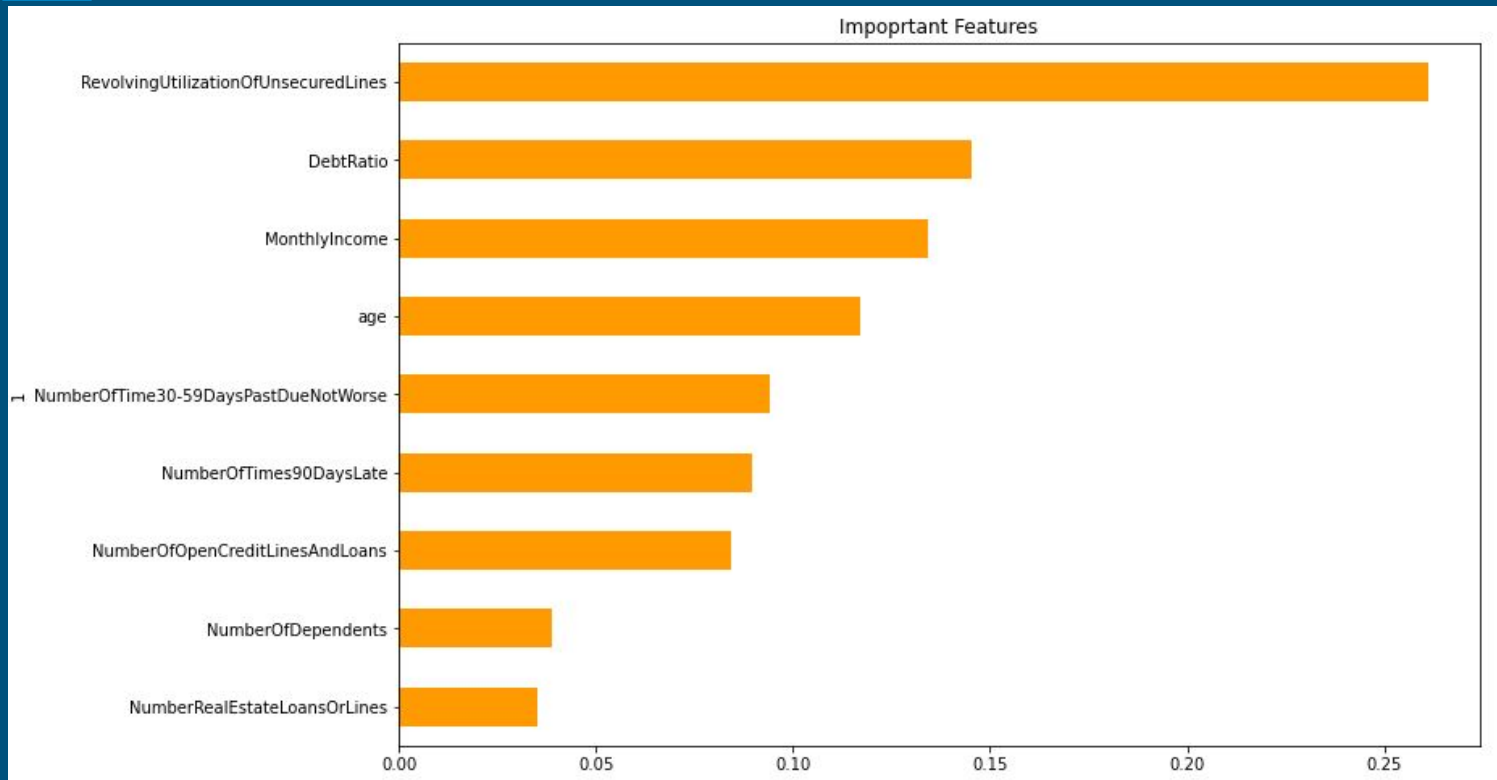
Objective

- Build a ML model that can predict whether borrowers will experience financial distress in the next two years
 - Financial distress: experience ≥ 90 days past due delinquency
- Build a robo adviser

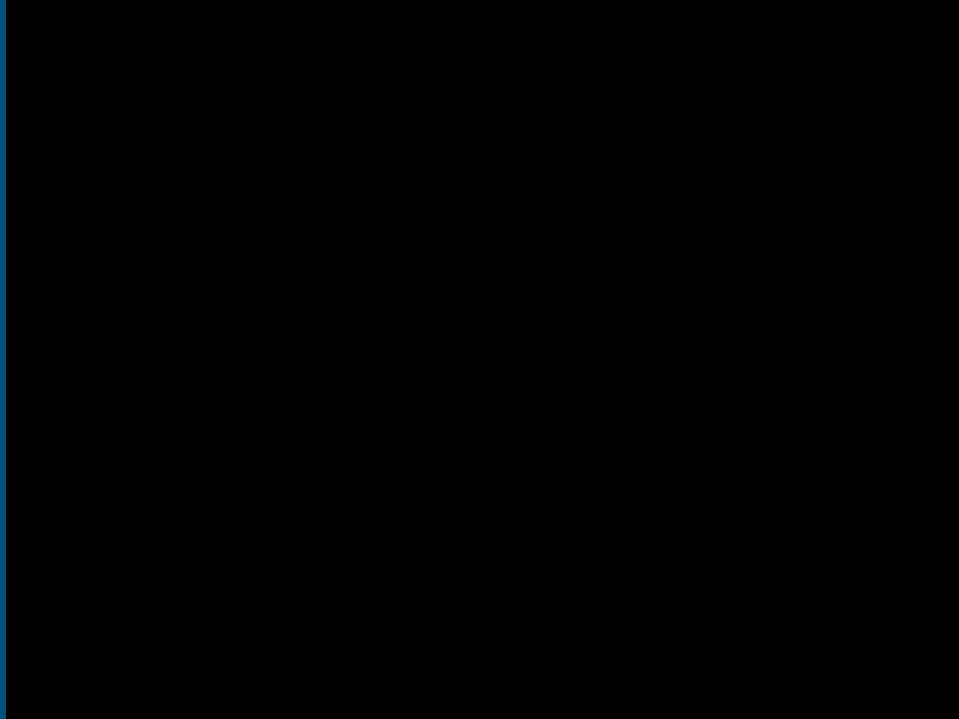
Data Dictionary

Variable Name	Description
SeriousDlqin2yrs (<i>target</i>)	Person experienced 90 days past due delinquency or worse
RevolvingUtilizationOfUnsecuredLines	Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits
age	Age of borrower in years
NumberOfTime30-59DaysPastDueNotWorse	Number of times borrower has been 30-59 days past due but no worse in the last 2 years
DebtRatio	Monthly debt payments, alimony, living costs divided by monthly gross income
MonthlyIncome	Monthly income
NumberOfOpenCreditLinesAndLoans	Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)
NumberOfTimes90DaysLate	Number of times borrower has been 90 days or more past due
NumberRealEstateLoansOrLines	Number of mortgage and real estate loans including home equity lines of credit
NumberOfDependents	Number of dependents in family excluding themselves (spouse, children etc.)

Important Features via RandomForestClassifier



Bot Demo



ML Model Evaluation

- Sklearn libraries

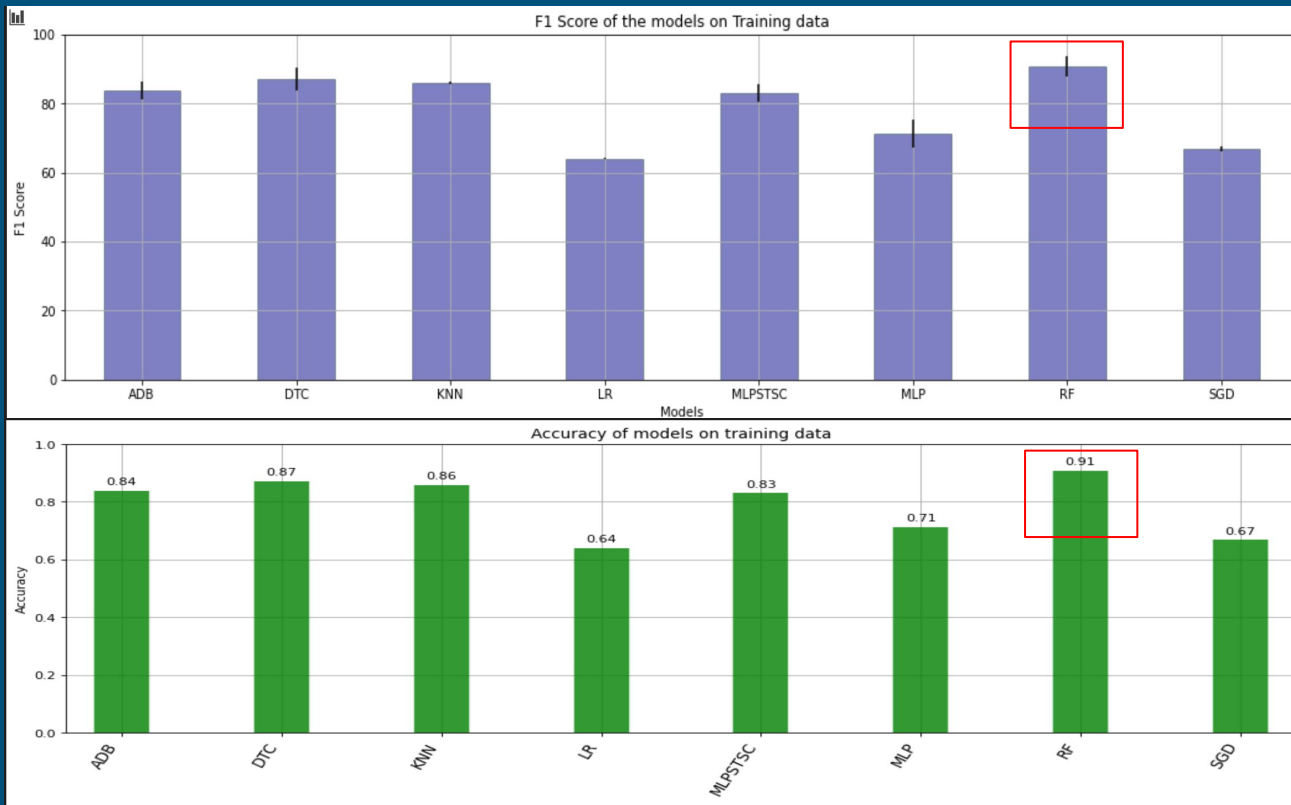
Data preparation:

- Imputed the data (for NA, NAN, NULL values)
- Imbalanced data - synthesized the minority class data using SMOTE technique and adaptive synthetic sampling technique on the training data
- Removed the highly correlated columns

Parameter tuning:

- Applied Grid search cross validation technique to search for model selection and parameter tuning

Summary



Summary ...

- Based on the grid search result, selected models with f1 and accuracy scores > 80 (Random forest, Decision tree, AdaBoost, KNN, Multilayer perception and XGBoost models)
- Tested selected models and evaluated models based on classification report - Weighted f1 avg ≥ 85 (Random forest, Decision tree, AdaBoost, Multilayer perception and XGBoost models)
- Applied ensemble classifiers (stacking, Voting soft and voting hard)
- Run the ensemble models and found Weighted f1 avg of 91 and accuracy of 90.45 using the Voting Hard ensemble classifier
- Selected Voting Hard classifier as the final model.

Postmortem

- Better data set that contain additional features:
 - Credit score, prior default, education level, years employed, race and ethnicity, zip code etc.
- Polish bot design and execution
 - Get Lambda Running
 - Get API connected
- Difference between Binary outcome vs probability



Q & A