

Loan Default Model

- Yuke Gan, Ying Feng, Shihan Wang

Data

- The data comes from Kaggle:
<https://www.kaggle.com/datasets/hemanthsai7/loandefault/data>
- Total 35 Columns:
 - Except ID and Loan Default target column, 33 Columns left
 - 9 Categorical Columns, 24 Numerical Columns
- Total 67463 Rows

Data Preprocess

1. Load Data

2. Check Data

- a. Drop rows with nan values (No Nan Values)
- b. Drop feature with only one unique value (One column dropped: "Payment Plan")

3. Prepare Train and Test Dataset

- a. Get X, y dataset: X contains 32 features, y is loan default column (target value 0 & 1)
- b. Convert categorical columns to numbers by LabelEncoder()
- c. Check target value imbalance: 90% 0 and 10% 1
 - i. Balance data by imblearn.over_sampling.SMOTE
- d. Split X_train, X_test, y_train, y_test with 80% train data, 20% test data
- e. Standardize the dataset by StandardScaler()

Pick Model

1. Train 5 different models by using Metaflow's foreach method
 - a. RandomForestClassifier(random_state=42),
 - b. XGBClassifier(random_state=42),
 - c. AdaBoostClassifier(random_state=42),
 - d. GradientBoostingClassifier(random_state=42),
 - e. LogisticRegression(random_state=42)
2. Pick the model with highest **precision in target value 1**
 - a. Using classification_report: RandomForestClassifier(random_state=42)

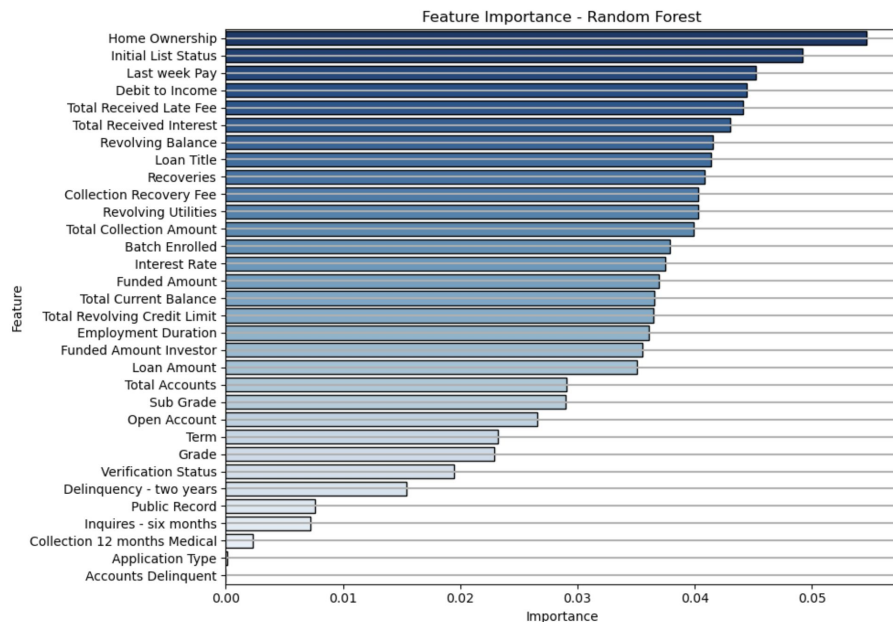
	Model	Precision on 1
0	Random Forest Classifier	0.893130
1	XGBClassifier	0.842205
2	AdaBoost	0.763595
3	Gradient Boosting	0.791810
4	Logistic Regression	0.668601

Improve Model

1. Check Feature Importance

- Model.feature_importances_ :
- Train RFC Again with updated features
 - Original Precision: 0.893130

```
{ '0': { 'precision': 0.8640967260736437,
        'recall': 0.8988158432013067,
        'f1-score': 0.8811144023697062,
        'support': 12245},
  '1': { 'precision': 0.894571136827774,
        'recall': 0.858624632473048,
        'f1-score': 0.8762293715619269,
        'support': 12244},
  'accuracy': 0.8787210584343991,
  'macro avg': { 'precision': 0.8793339314507089,
                  'recall': 0.8787202378371773,
                  'f1-score': 0.8786718869658166,
                  'support': 24489},
  'weighted avg': { 'precision': 0.879333309244601,
                    'recall': 0.8787210584343991,
                    'f1-score': 0.8786719867051038,
                    'support': 24489}}
```



Improve Model

```
{'0': {'precision': 0.8551050469028607,  
      'recall': 0.9007758268681094,  
      'f1-score': 0.8773464842507158,  
      'support': 12245},  
'1': {'precision': 0.8951682484900777,  
      'recall': 0.8473538059457694,  
      'f1-score': 0.8706050180414535,  
      'support': 12244},  
'accuracy': 0.8740659071419821,  
'macro avg': {'precision': 0.8751366476964693,  
              'recall': 0.8740648164069393,  
              'f1-score': 0.8739757511460846,  
              'support': 24489},  
'weighted avg': {'precision': 0.8751358297128524,  
                 'recall': 0.8740659071419821,  
                 'f1-score': 0.8739758887888266,  
                 'support': 24489}}
```

App Website

Using Flask to build the app website.

Loan Default Classifier

Example Applicant's Characteristics:

Loan Amount: 32879
Funded Amount: 1014
Funded Amount Investor: 10989.93948
Term: 59
Batch Enrolled: BAT5629144
Interest Rate: 6.894023179
Grade: B
Sub Grade: A1
Home Ownership: MORTGAGE
Employment Duration: 94327.15399
Verification Status: Not Verified
Loan Title: Debt consolidation
Debt to Income: 34.8737539
Delinquency - two years: 0
Inquires - six months: 0
Open Account: 11
Public Record: 0
Revolving Balance: 2753
Revolving Utilities: 68.81130029
Total Accounts: 19
Initial List Status: w
Total Received Interest: 471.8253429
Total Received Late Fee: 0.092744198
Recoveries: 3.789364581
Collection Recovery Fee: 0.578289675
Collection 12 months Medical: 0
Application Type: INDIVIDUAL

Predict

Predicted label: 0

Loan Amount: loan amount applied

Funded Amount: loan amount funded

Funded Amount Investor: loan amount approved by the investors

Term: term of loan (in months)

Batch Enrolled: batch numbers to representatives

Interest Rate: interest rate (%) on loan

Grade: grade by the bank

Sub Grade: sub-grade by the bank

Home Ownership: Type of home ownership

Employment Duration: Employment duration in hours

Verification Status:

Loan Title: Purpose for loan

Debt to Income: ratio of representative's total monthly debt repayment divided by self reported monthly income excluding mortgage

Delinquency - two years: number of 30+ days delinquency in past 2 years

Inquires - six months: total number of inquiries in last 6 months

Open Account: number of open credit line in representative's - credit line

Public Record: number of derogatory public records

Revolving Balance: total credit revolving balance

Revolving Utilities: amount of credit a representative is using - relative to revolving_balance

Total Accounts: total number of credit lines available in - representatives credit line

Initial List Status: unique listing status of the loan - - W(Waiting), F(Forwarded)

Total Received Interest: total interest received till date

Total Received Late Fee: total late fee received till date

Recoveries: post charge off gross recovery

Collection Recovery Fee: post charge off collection fee

Collection 12 months Medical: total collections in last 12 months - excluding medical collections

Application Type: indicates when the representative is an individual or joint

Last week Pay: indicates how long (in weeks) a representative has paid EMI after batch enrolled

Accounts Delinquent: number of accounts on which the representative is delinquent

Total Collection Amount: total collection amount ever owed

Total Current Balance: total current balance from all accounts

Total Revolving Credit Limit: total revolving credit limit

Predict