Bank Loan Term Prediction

Ali
Ahmed
Abdulrahman

Table of contents

01

Introduction

Backstory

- Loan is one of the most important schemes of banks.
- Short Term Loan or Long Term Loan.
- Buying a house → Long term.
- Take a trip → short term.
- Help bankers to determine the type of loan.

Data set

- **Bank Loan Status Dataset**
- Kaggle.
- 110867 rows.
- 19 column.
- 16 feature columns.
- 1 binary class target column
- Target column:
  - Short term
  - Long term

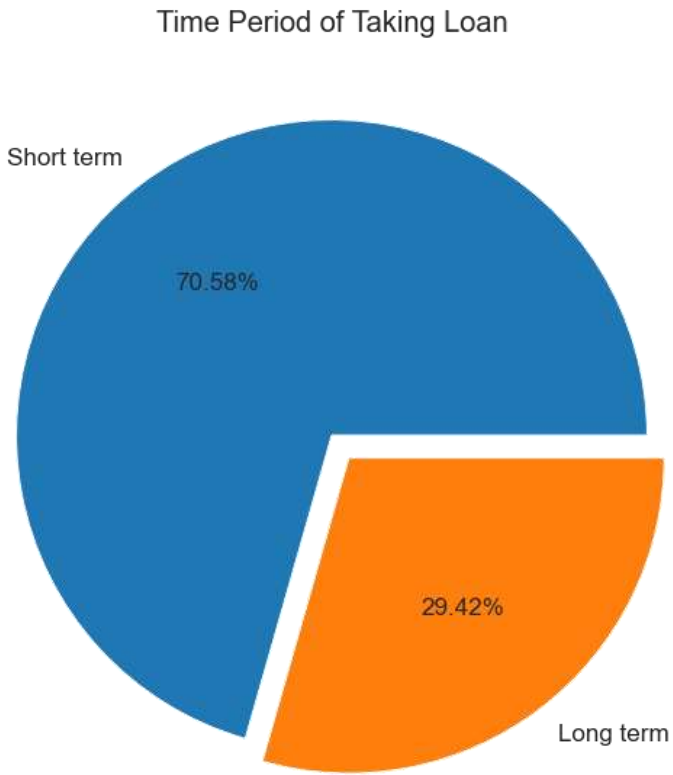
Tools

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Sklearn
- XGBoost
- Pickle

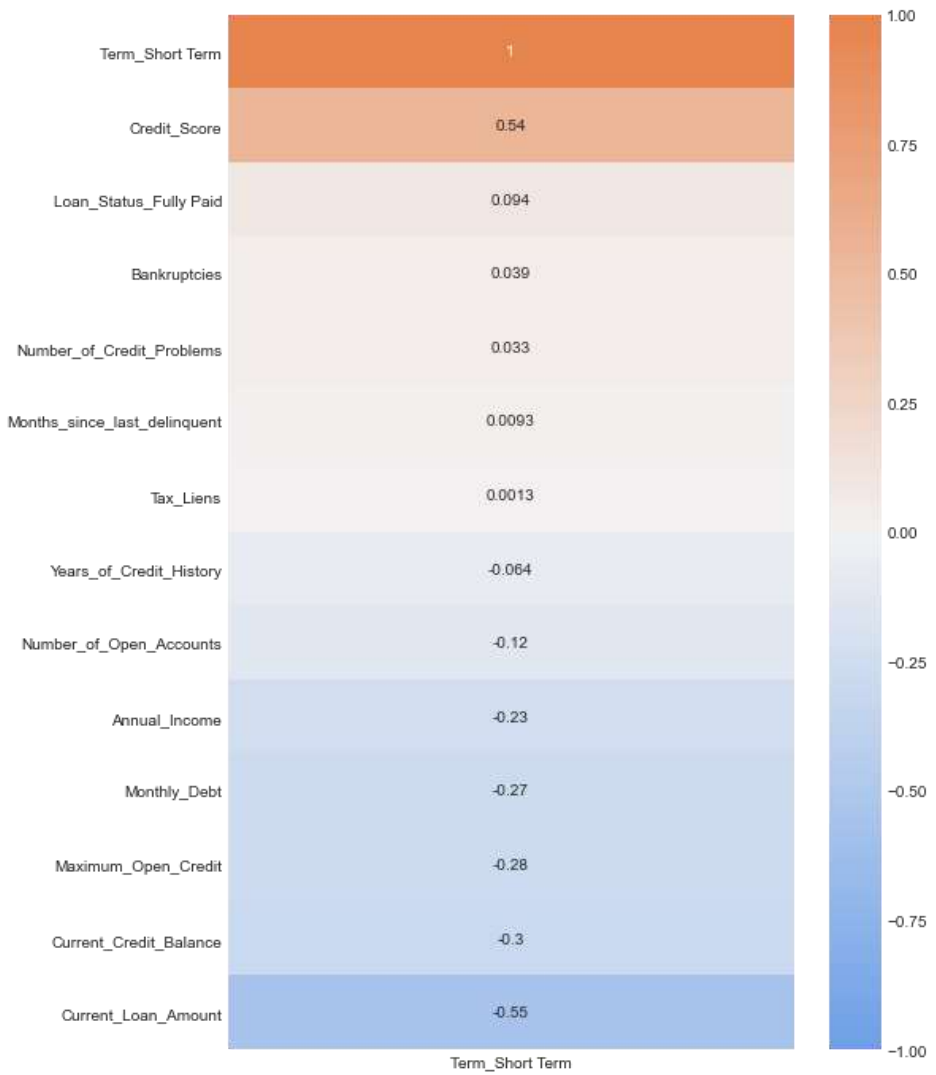
02

Data Analysis

Type of Terms Plot

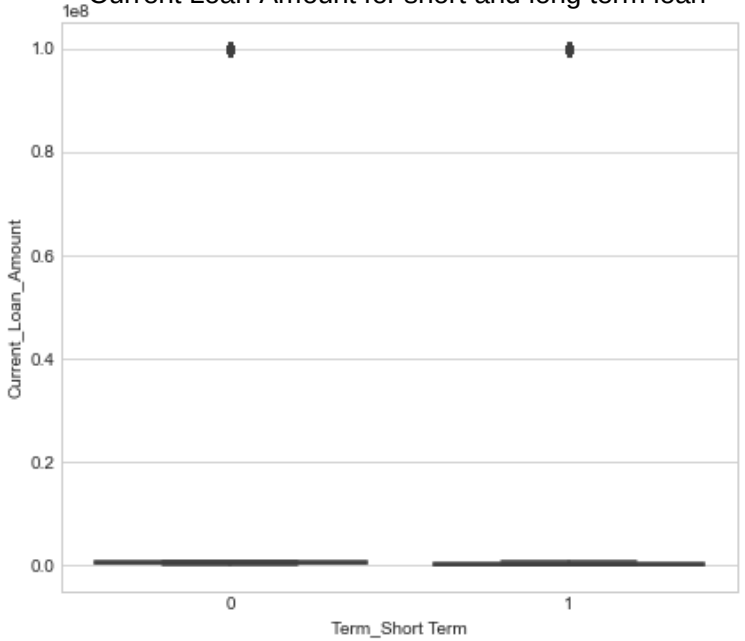


Features and Target Correlation

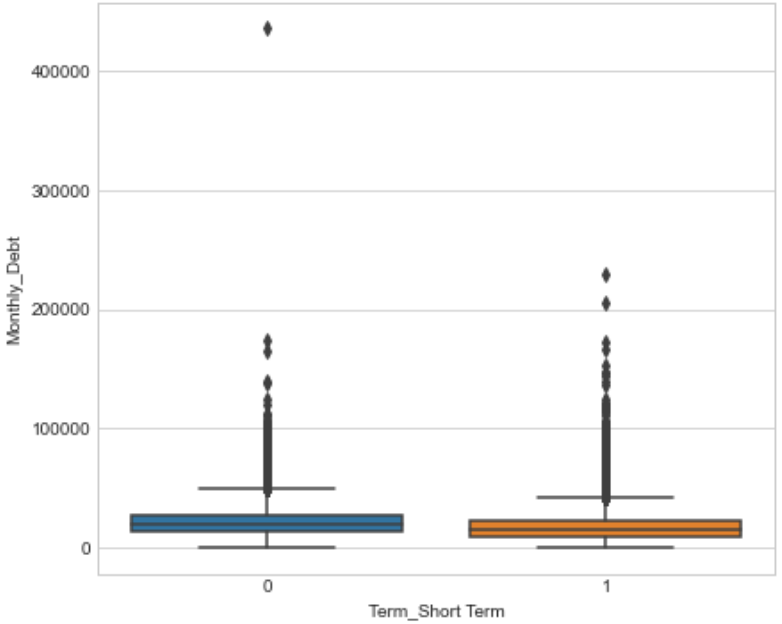


Outliers Boxplot

Current Loan Amount for short and long term loan



Monthly Dept for short and long term loan



03

Data Cleaning and Feature Engineering

Data Cleaning

01 **Check for NaN and deal with them.**

02 **Drop unwanted columns.**
Loan ID , customer ID

03 **Check and drop duplicate.**

04 **Check and drop outliers.**

Feature Engineering

New columns

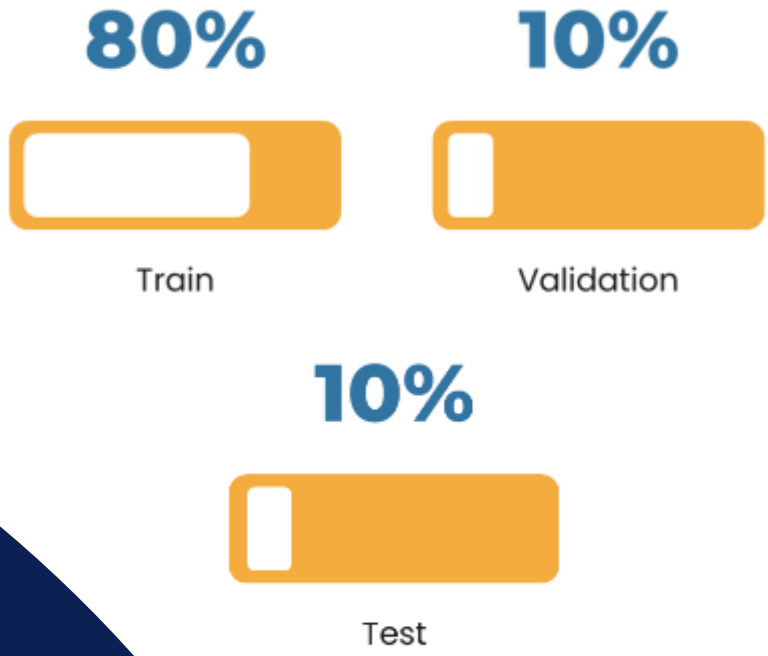
$(Credit\ Score)^3$

$(Current\ Loan\ Amount) * (Credit\ Score)$

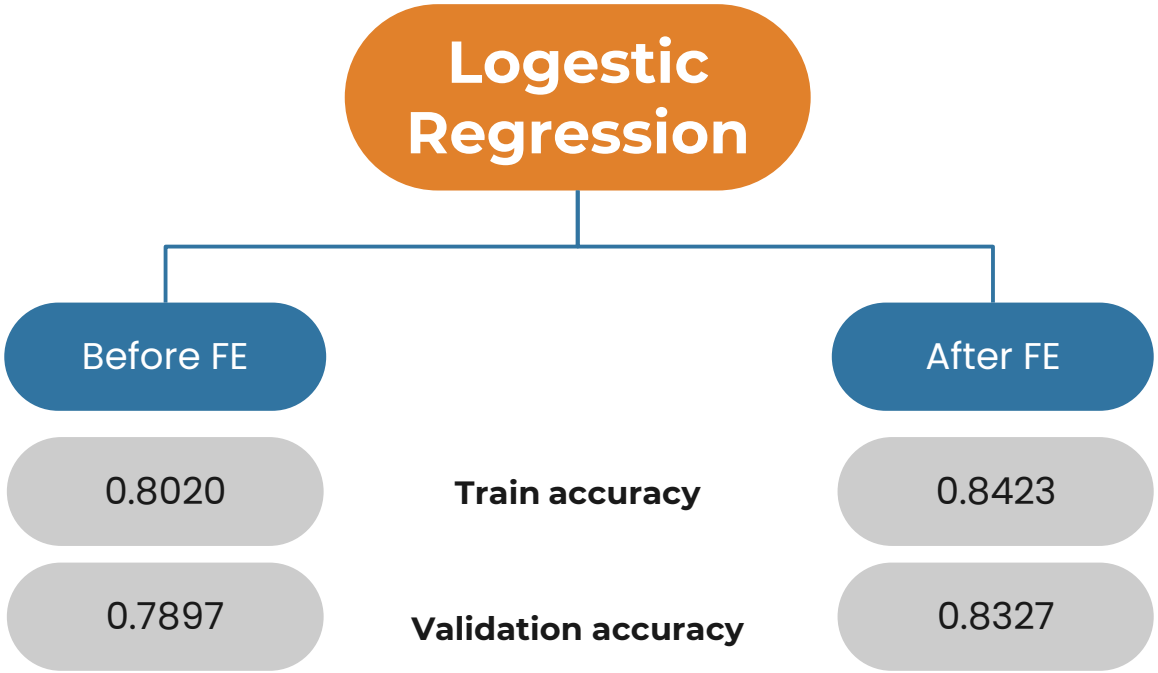
$(Annual\ Income)^{0.05} * (Current\ Loan\ Amount)$

$\left(\sqrt{Current\ Credit\ Balanc} * (Credit\ Score)\right)^2$

04 Classification Models



Baseline Model



Logistic Regression

Model	Accuracy	
	Train	Validation
Logistic Regression	0.8423	0.8327
Logistic Regression Scaled	0.8613	0.8670
LogisticRegression class weight {Long Term : 2 , Short term : 1}	0.8435	0.8466
LogisticRegression class weight : balanced	0.8382	0.8363

Naive Bayes

Model	Accuracy	
	Train	Validation
Gaussian NB	0.8311	0.8308
Bernoulli NB	0.6871	0.6888
Multinomial NB	0.7807	0.7771

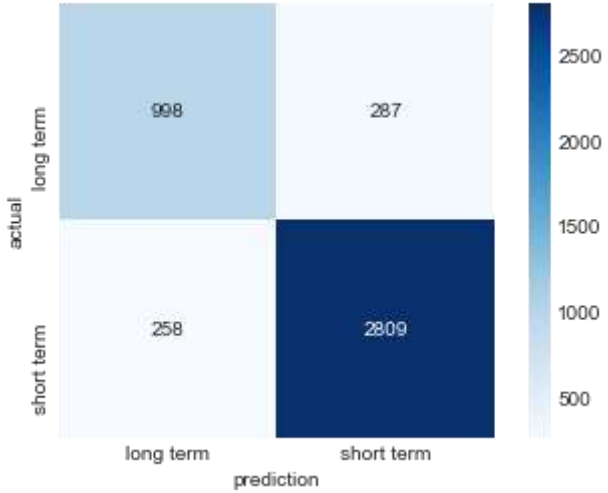
More models

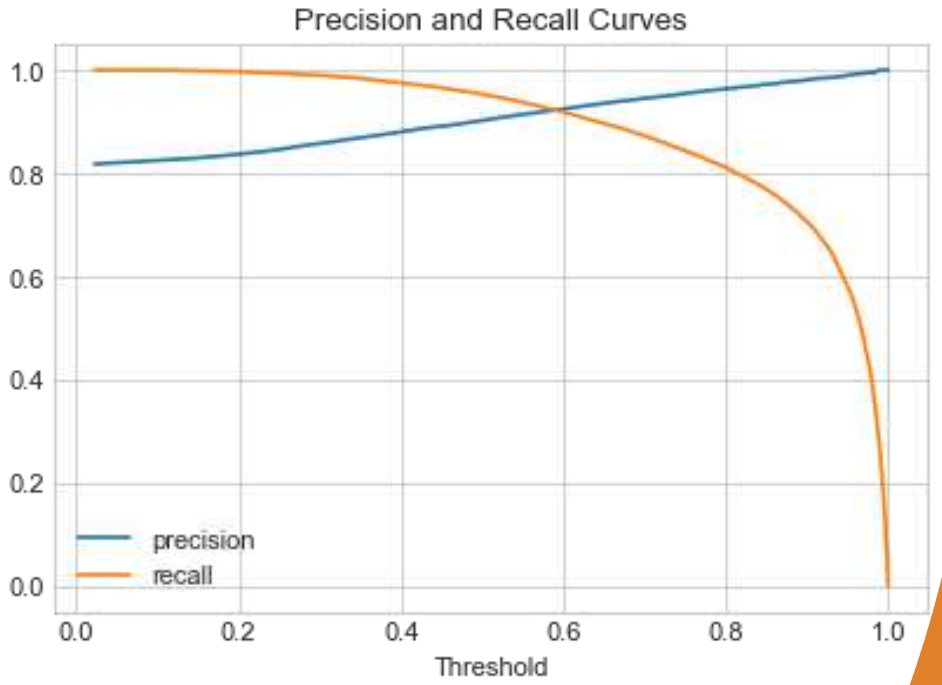
Model	Accuracy		F1 Score	
	Train	Validation	Train	Validation
Logistic Regression Scaled	0.8613	0.8670	0.9045	0.9044
K-Neatest Neighbors (3)	0.9051	0.8332	0.9339	0.8849
Decision Tree	0.8741	0.8683	0.9148	0.9073
Random Forest	0.9999	0.8732	1.0	0.9141
Extra Tree	1.0	0.8699	1.0	0.9111
Ada Boost	0.8738	0.8758	0.9131	0.9155
Stochastic Gradient Descent	0.8580	0.8616	0.9035	0.9067
XGBoost	0.8916	0.8856	0.9266	0.9179

Best Classification model

XGBoost classifier

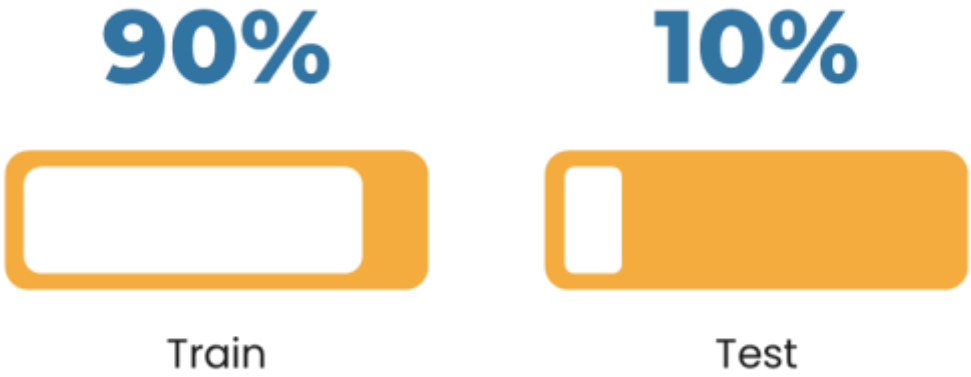
Threshold:	0.59
Precision:	0.9225
Recall:	0.9220





05

XGBoost Conclusion



Accuracy

Train: 0.8859

Test: 0.8812

F1 Score

Train: 0.9215

Test: 0.9181