Loan Status Prediction

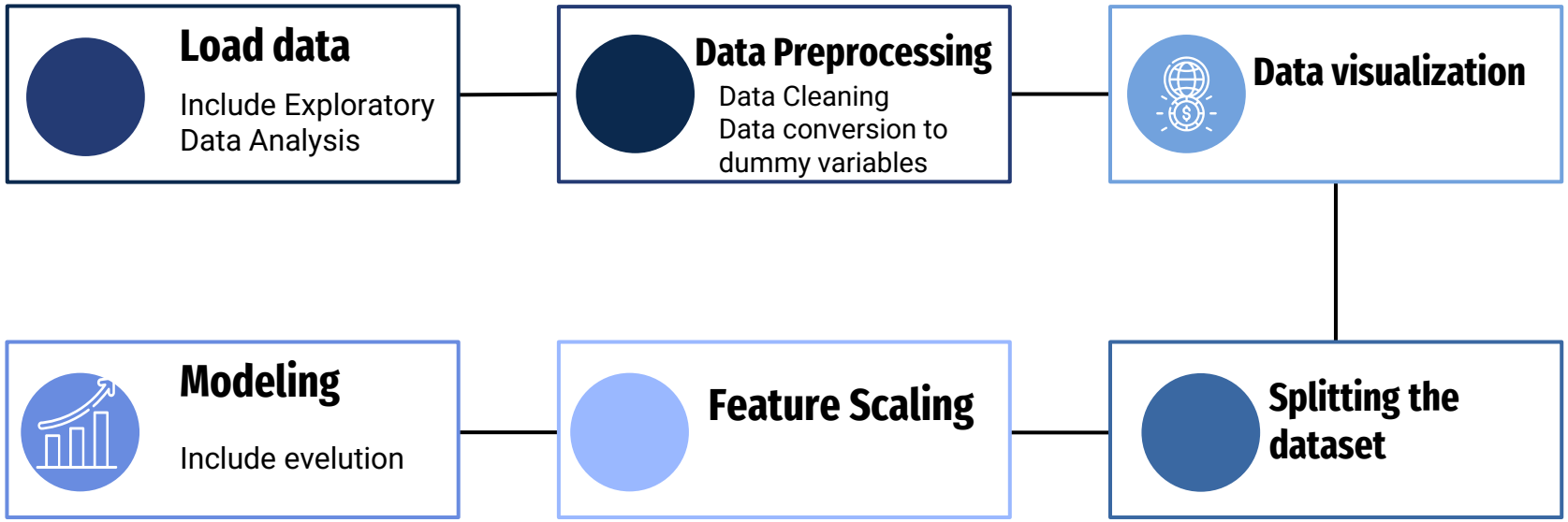
By : Mashael Alsalhi

Introduction

Banks run into losses when a customer doesn't pay their loans on time. Because of this, every year, banks have losses, and this also impacts the country's economic growth to a large extent.

In this Dataset, we look at various attributes such as credit score, annual income, years of employment and try to train the classifier to predict if the loan would be paid or not.

Process modeling



Data

The dataset contains of **100514 observations** and **19 features, After cleaning the observations 89593 and 16 features**

Column name	Column type
0 Loan ID	object
1 Customer ID	object
2 Loan Status	object
3 Current Loan Amount	float64
4 Term	object
5 Credit Score	float64
6 Annual Income	float64
7 Years in current job	object
8 Home Ownership	object
9 Purpose	object
10 Monthly Debt	float64
11 Years of Credit History	float64
12 Months since last delinquent	float64
13 Number of Open Accounts	float64
14 Number of Credit Problems	float64
15 Current Credit Balance	float64
16 Maximum Open Credit	float64
17 Bankruptcies	float64
18 Tax Liens	float64

Models

XGBClassifier

RandomForest Classifier

LogisticRegression

Metrics

01 | Recall

$$\text{Recall} = \frac{\text{True positives}}{\text{Actual positives}}$$

02 | Precision

$$\text{Precision} = \frac{(\text{True positives})}{\text{Predicted positives}}$$

03 | F1 score

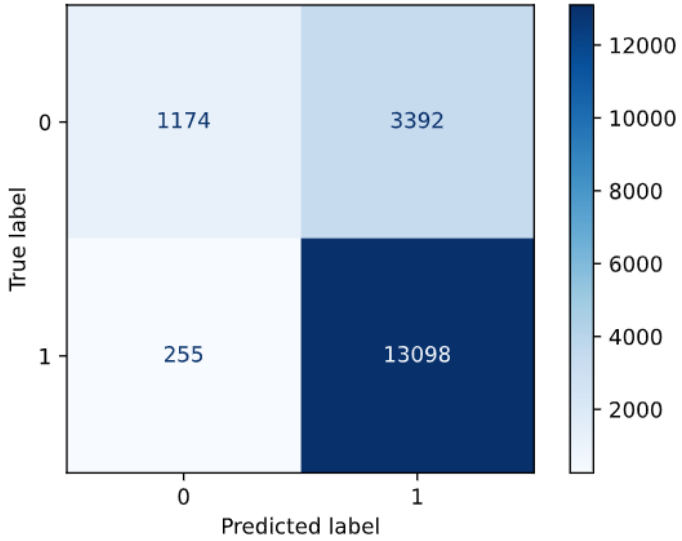
$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

Results

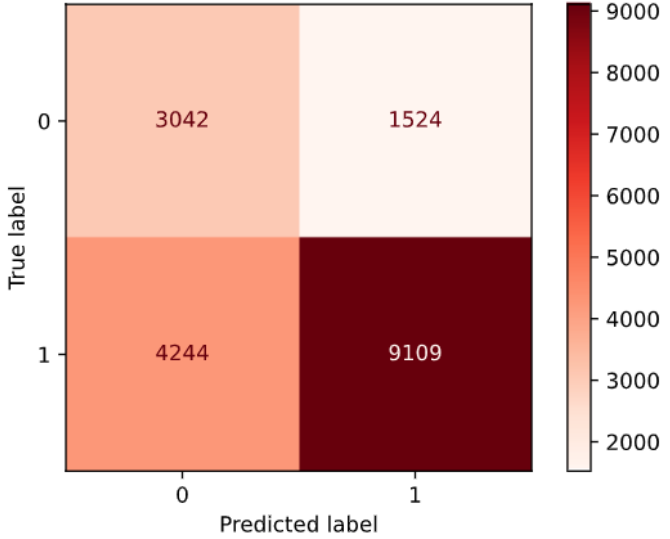
Summary

Model	Recall	Precision	F1 score
XGB	**98.1%**	**79.4%**	**87.8%**
XGB with under-sampling	**67.6%**	**85.6%**	**75.9%**
Random Forest	**99%**	**79.1%**	**87.9%**
LogisticRegression	**71.7%**	**84.4%**	**77.6%**

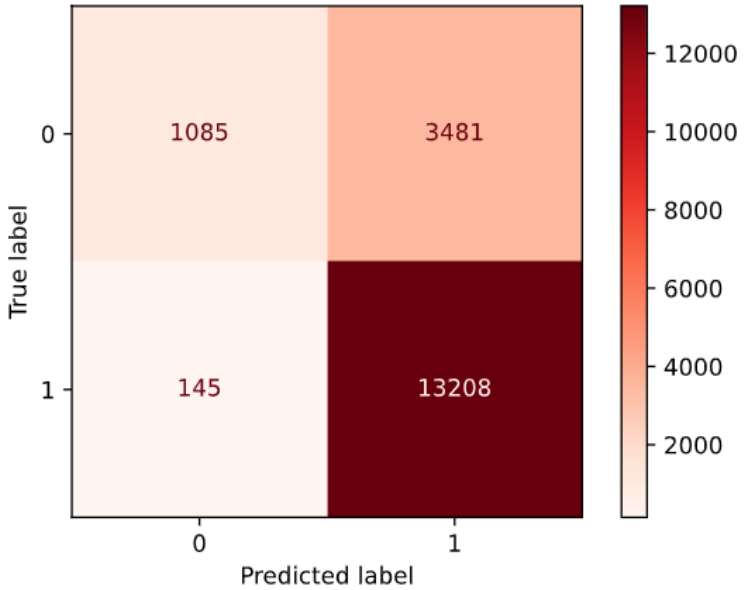
XGB



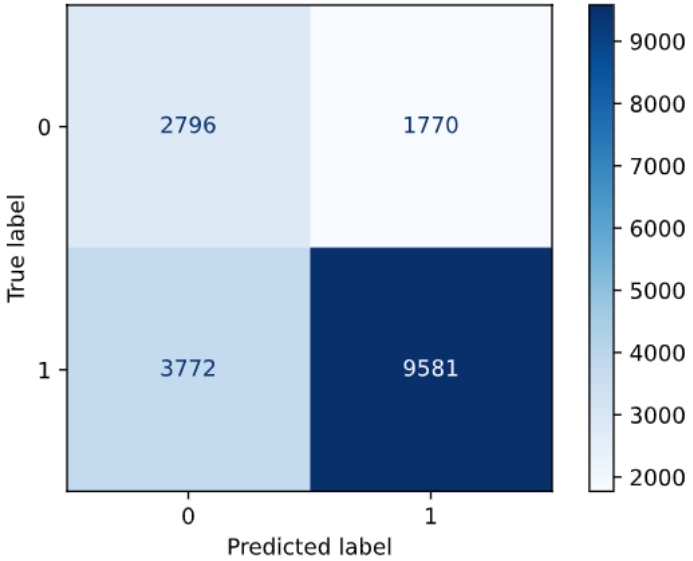
XGB with under sampling



Random Forest



LogisticRegression



Demo

Conclusions and Future Work

Appendix

Appendix

- Dataset : Bank Loan Status Dataset | Kaggle
- GitHub : MashaelAlsalhi/T5_Loan-Status_Prediction (github.com)

Thank You