

SDAIA Academy T5C04 Bootcamps: Data Science
Classification Module Project Presentation



Predicting the Risk of Loan Default

Presented by Nada Rambu

01

Problem and Solution

Project background and objectives

Objective

Predict whether a customer's loan would be approved or denied depending on their risk of defaulting

02

Data Processing

Getting the data from the its source to be processed

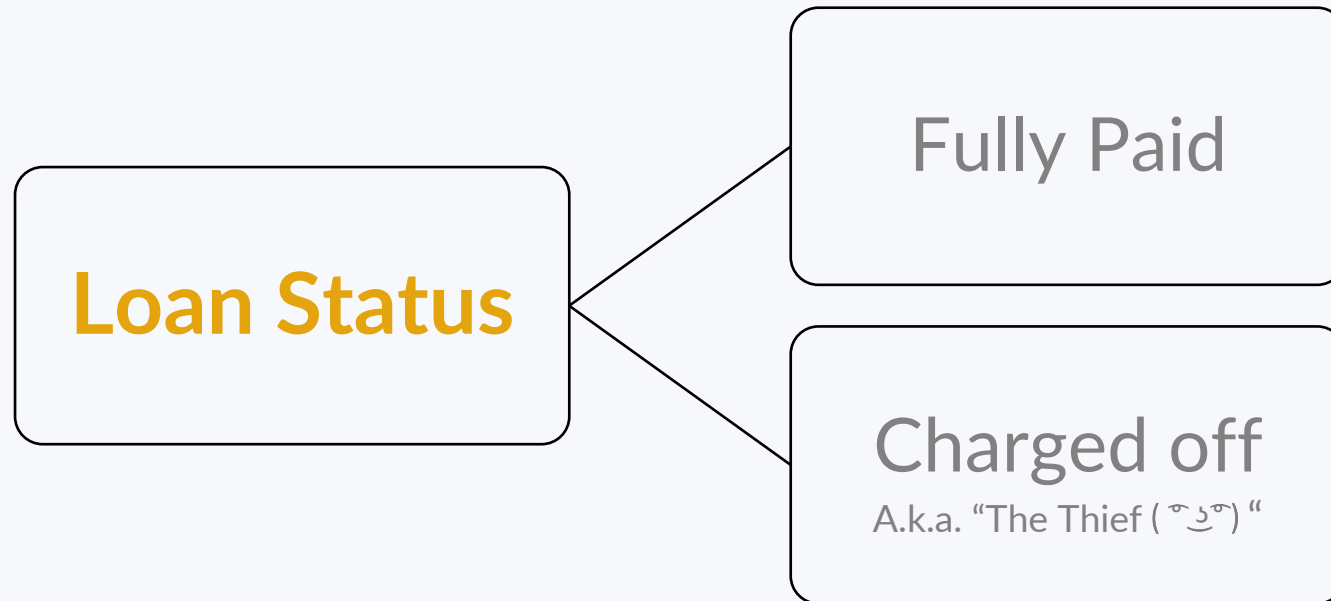
Data



Features



Features



Preprocessing



Handling Missing values



Data Type Conversion



Dataset Resampling



Encoding Categorical Features



Removing Outliers

Preprocessing

38560 rows × **17** columns

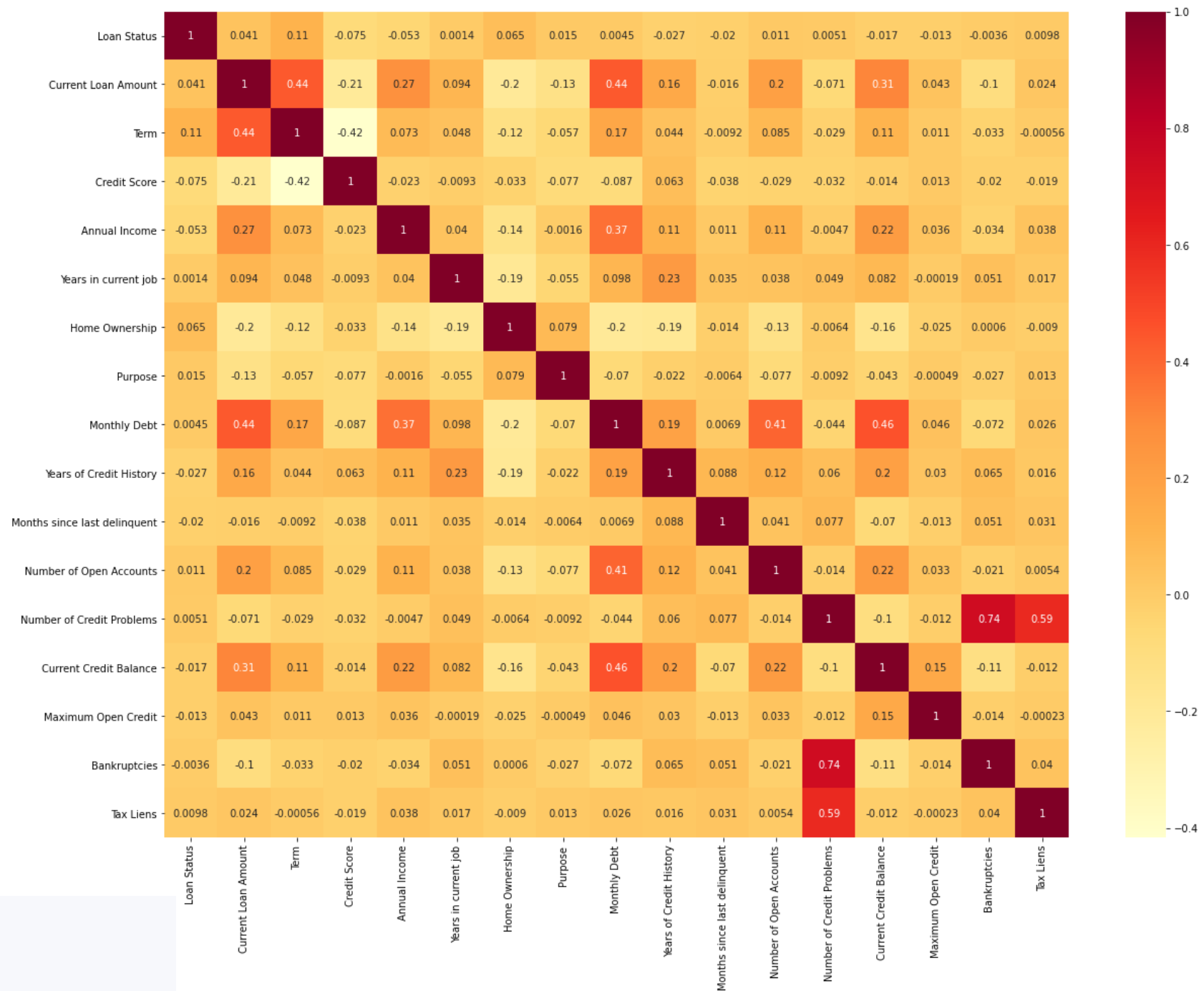
Preprocessing

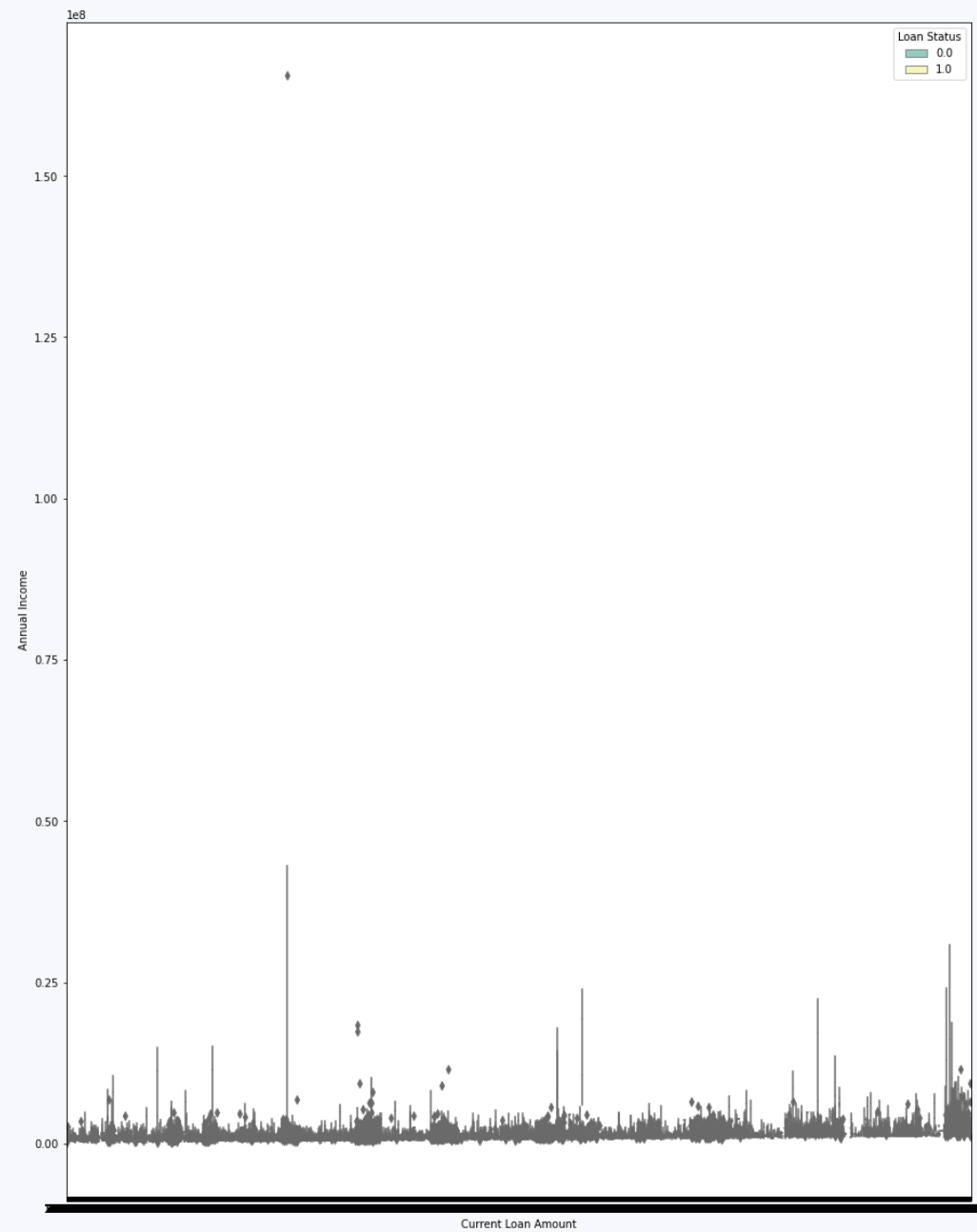
38560 rows × **17** columns

03

Data Visualization

Gaining insights from the data





04

Model Development

Train and test the model to estimate results

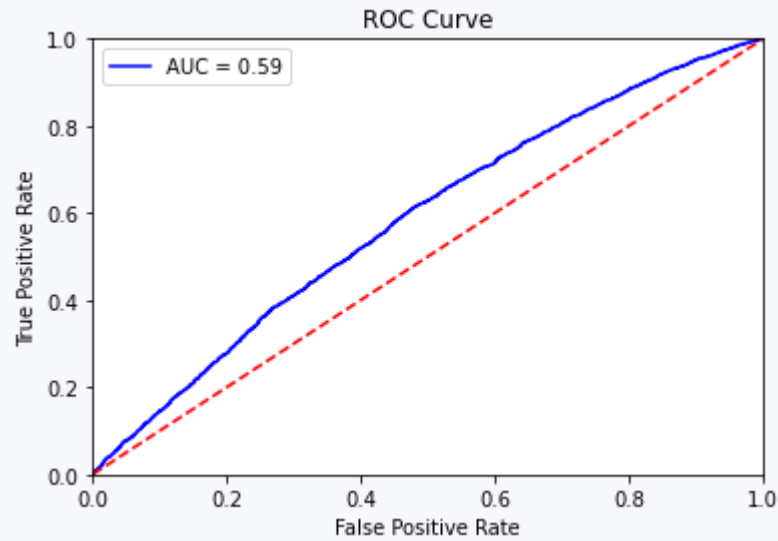
Models

Algorithm	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.57	0.58	0.57	0.56
Random Forest	0.60	0.64	0.62	0.60
Gradient Boosting	0.56	0.60	0.58	0.56

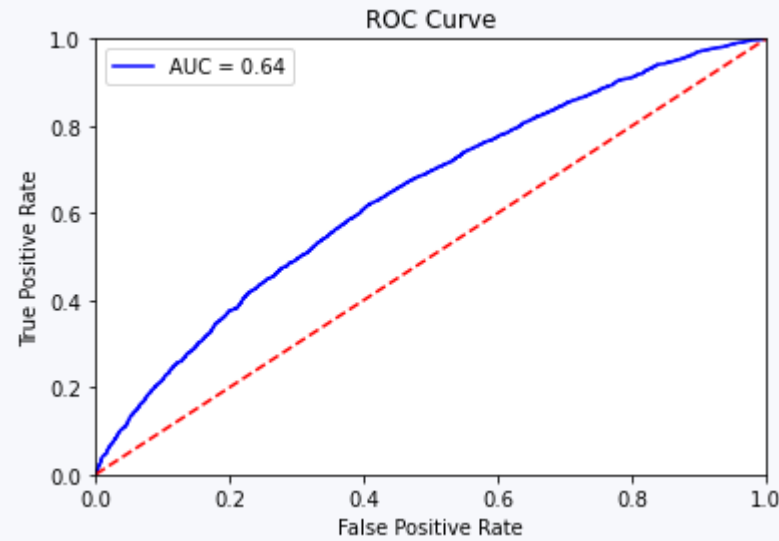
Models

Algorithm	Precision	Recall	F1-score	Accuracy
Logistic Regression	0.57	0.58	0.57	0.56
Random Forest	0.60	0.64	0.62	0.60
Gradient Boosting	0.56	0.60	0.58	0.56

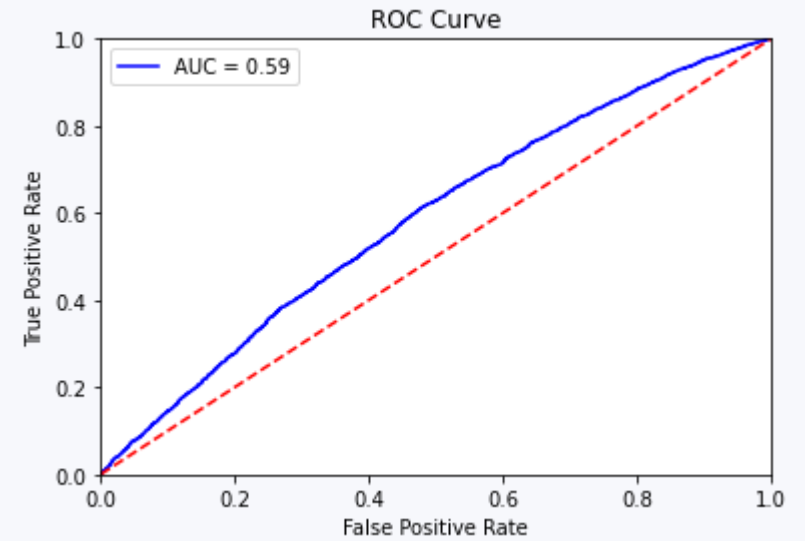
Models



**Logistic
Regression**



**Random
Forest**



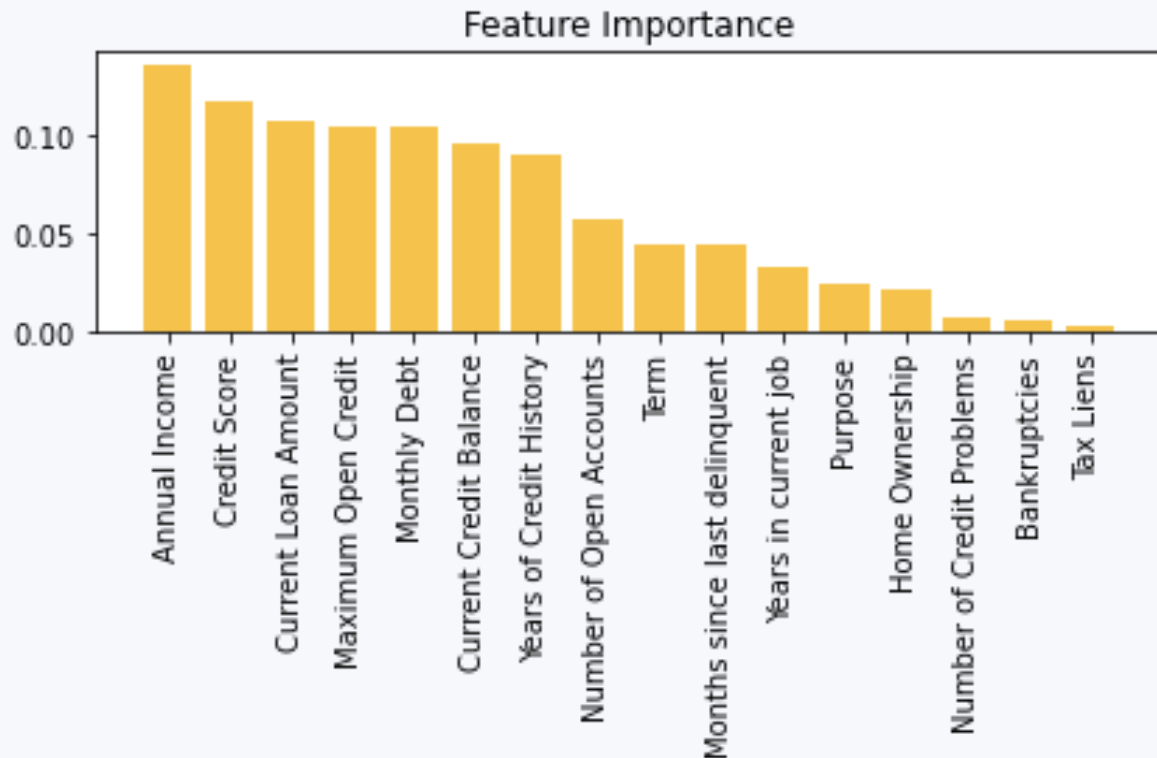
**Gradient
Boosting**

05

Conclusion

Some recommendation to solve the problem

Insights



According to the bar chart, these features:

1. Annual Income
2. Credit Score
3. Current Loan Amount
4. Monthly Debt
5. Current Credit History

Are the most important features in explaining the risk of defaulting.



Thank You!