

Credit Risk Modeling

ID/X Partners - Data Scientist

Presented by
M.Idris Mangku Jagat



Lampung, Indonesia



idrismangkujagat11@gmail.com



M.Idris Mangku Jagat

Hi, I'm Idris

Bachelor of Mathematics from the Sumatra Institute of Technology with a GPA of 3.48. Have an understanding of data analysis, statistics, and modeling in solving problems. Have experience working in the data field, and the ability to carry out analysis, visualize data, and present data.

Table of Contents

1 Business Understanding

2 Data Understanding

**3 Exploratory Data Analysis
(EDA)**

4 Data Preparation

**5 Machine Learning Models
& Evaluation**

6 Conclusion

Business Understanding

Background of the Problem

Carrying out a project from a lending company (multi-finance), where the client wants to increase accuracy in assessing and managing credit risk, so that they can optimize their business decisions and reduce potential losses.

Scope and Goals

- **The dataset** used is loan data from **2007-2014**.
- The machine learning models used are **Logistic Regression**, **Decision Tree**, and **Random Forest**.
- Developing a machine learning model that can predict credit risk.

Tools



Python



Google
Colaboratory

Data Understanding

1

- **Dataset**

Loan data consists of **466,285 rows** and **75 columns or features**.

2

- **Checking for Missing Value**

There are datasets that **have missing values**.

3

- **Checking for Duplicate Data**

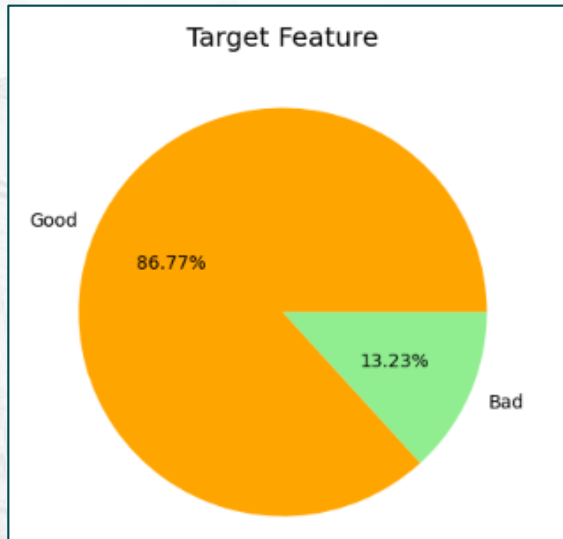
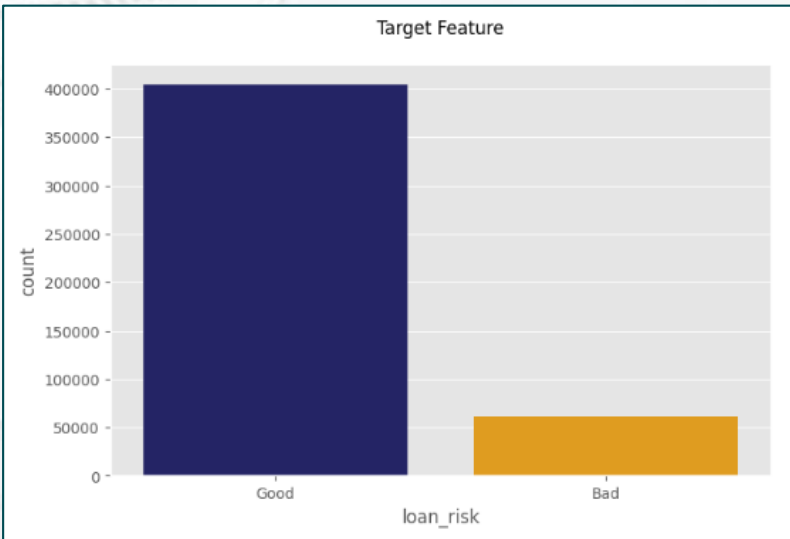
The dataset does **not have duplicate values**.

4

- **Checking for Imbalance Data**

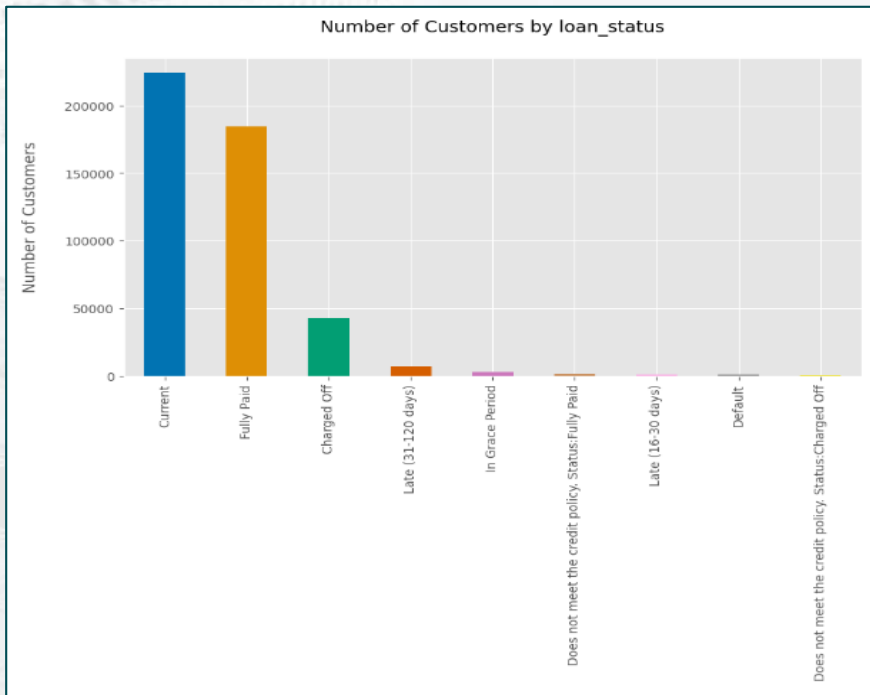
The dataset is **imbalanced**.

Exploratory Data Analysis (EDA)



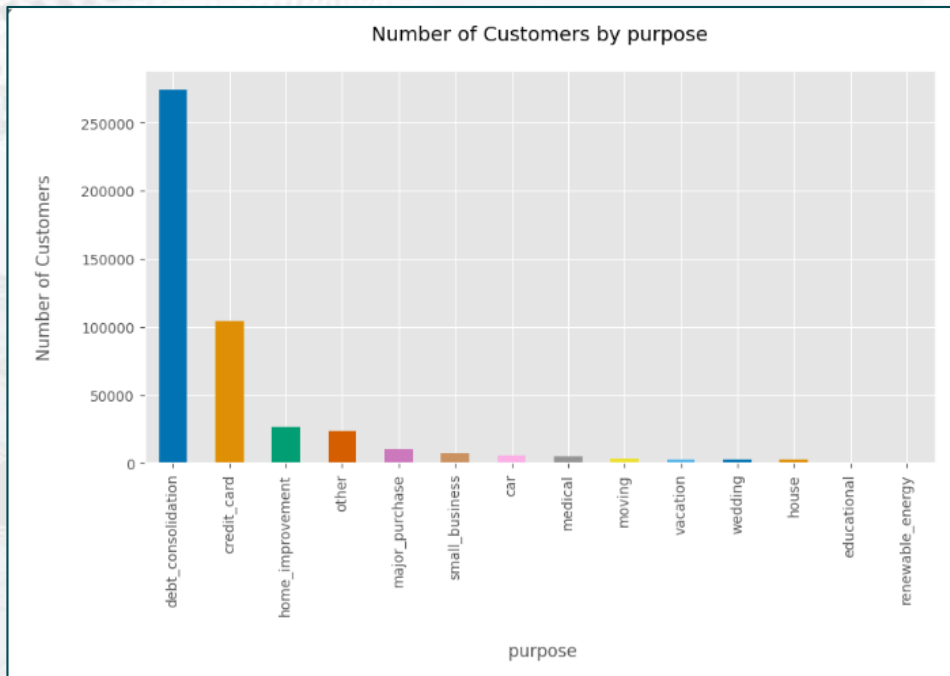
The comparison of loan risks tends to be towards good loans, with a risk percentage for **good loans** of **86.77%** and **bad loans** of **13.23%**.

Exploratory Data Analysis (EDA)



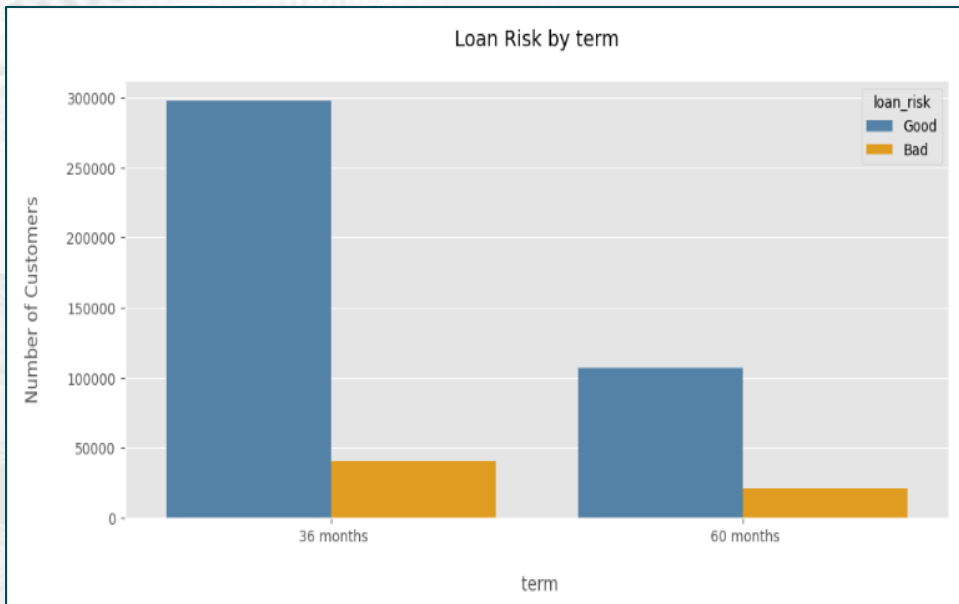
Based on the graph, the **majority** of customers are in **Current loan status** followed by Fully Paid. This shows that the majority of customers can be categorized as Good.

Exploratory Data Analysis (EDA)



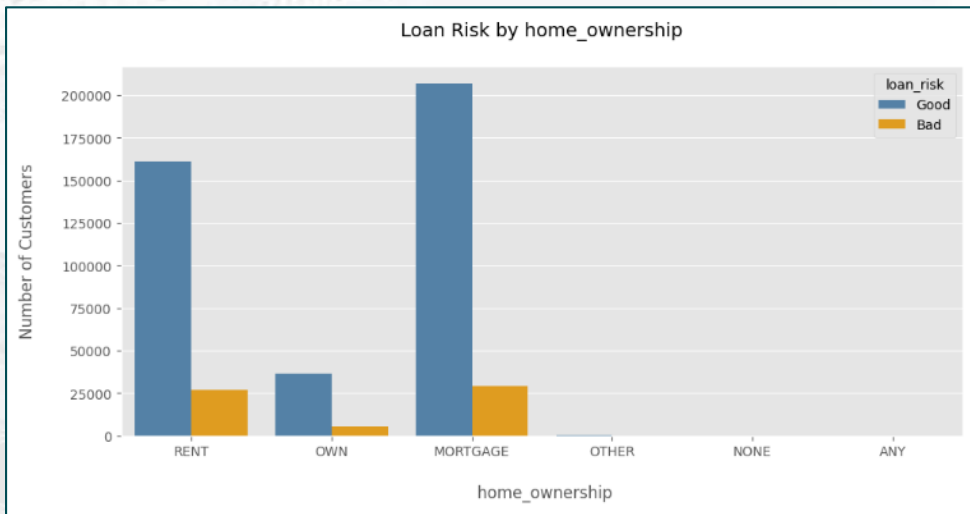
Based on their purpose, the **majority** of customers apply for loans for **debt_consolidation**, followed by credit_card.

Exploratory Data Analysis (EDA)



Most loans have shorter terms, namely 36 months. Loans with a term of **36 months** are likely to experience a **greater risk of bad loans** than those with a term of 60 months.

Exploratory Data Analysis (EDA)



Most customers who take out loans have **mortgages** or are currently **renting a house**. In addition, these customers have a higher risk of bad loans.

Data Preparation

1

- Handling Missing Value

2

- Feature Engineering

3

- Label Encoding

4

- Split Data (Train and Test)

5

- Handling Imbalance Data

6

- Correlation Data

7

- Data Standardization

Machine Learning Models & Evaluation

Machine Learning Models	Training Accuracy (%)	Testing Accuracy (%)
Random Forest	92.37	92.35
Decision Tree	92.33	92.32
Logistic Regression	77.81	77.82

Based on the accuracy of the prediction results from training and testing data, the **Random Forest model is better** than Decision Tree and Logistic Regression.

Github link (to see the full project) : <https://github.com/midrismj/Credit-Risk-Modeling.git>

Conclusion

- Comparison of loan risks tends towards good loans. The majority of customers have Current loan status, and can be categorized as Good.
- Loans with a term of 36 months will likely experience a greater risk of bad loans than those with a term of 60 months.
- The Random Forest model is better at analyzing customer loans, with training and testing accuracy of 92.37% and 92.35%.

Thank You

