PROJECT BASED INTERNSHIP PROGRAM

ID/X Partners
Data Scientist

Business Understanding

Definition

- Credit risk refers to the potential risk that a borrower may fail to meet their financial obligations, such as repaying a loan or meeting contractual commitments.

Objective

- Assess the creditworthiness of loan applicants to mitigate the risk of default.

Solution

- Use machine learning techniques to automate the process

Analysis Approach

- Descriptive Analysis
- Graph Analysis
- Predictive Modelling (Binary Classification)

Data Requirements

- Required a dataset of a customer loan from a company

Data Collection

- The dataset is collected by ID/X Partners from a company

id/x partners

Rakamin Academy

Data Understanding

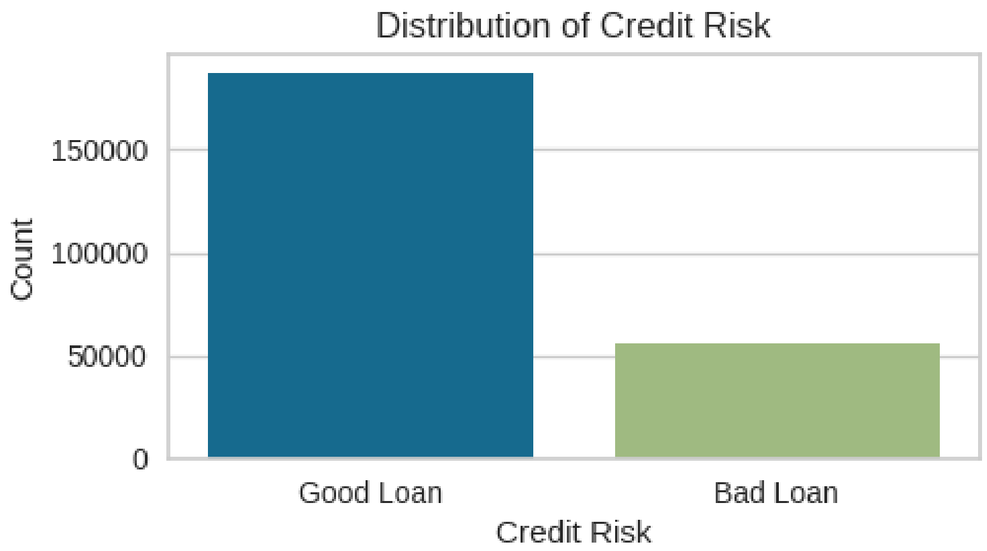
- This dataset has 74 features
- Consists of 52 numerical features & 22 non-numerical features
- Many features have missing values
- There are 17 null features

Data Preparation

- Missing Value: Removing and Imputing (Mode & Median)
- Feature Engineering: Categorical Encoding, Log Transform, Standardization
- Feature Selection using Correlation Analysis
- Removing outliers using IQR method

Exploratory Data Analysis

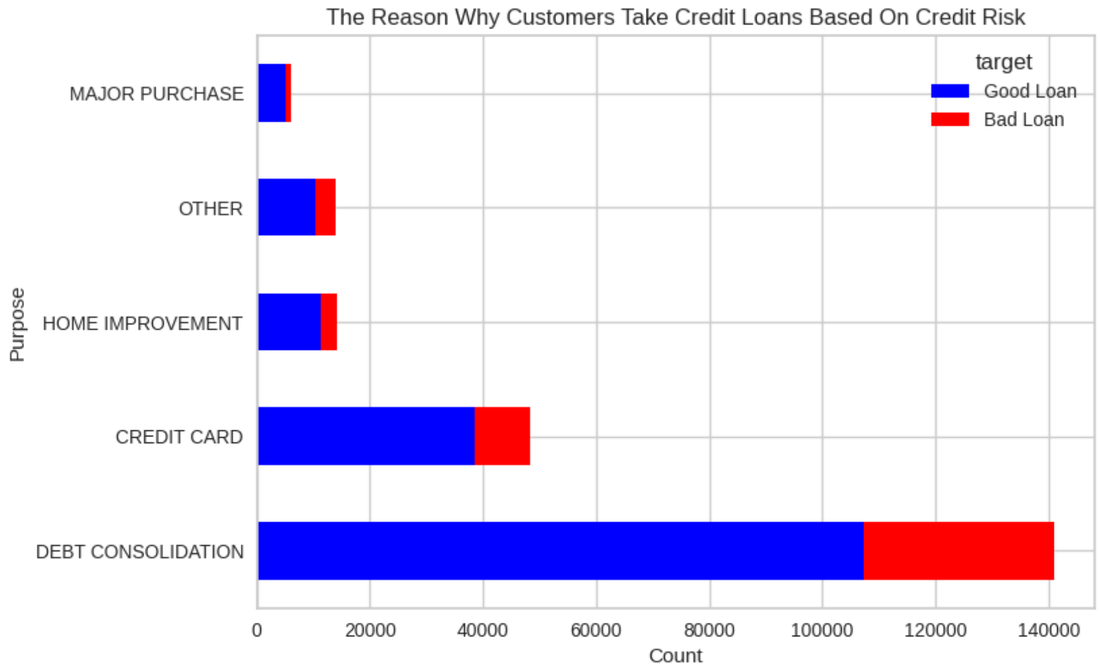
Target Features



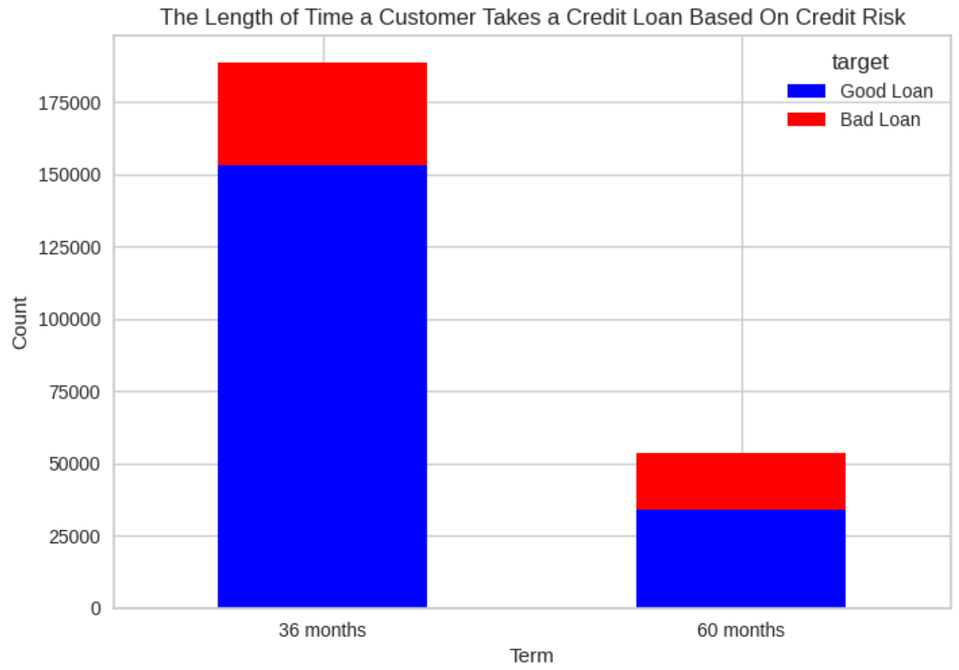
- Good Loan (1): Fully Paid, Does not meet the credit policy. Status:Fully Paid
- Bad Loan (0): Charged Off, Default, Late (31-120 days), In Grace Period, Late (16-30 days), Does not meet the credit policy. Status:Charged Off

Exploratory Data Analysis

Why do customers take out credit loans?



How long does the customer take a credit loan?



Modelling

- 80% Training 20% Testing
- Using SMOTE techniques for handling imbalanced data
- All steps are handled by pipeline

Evaluation

- Main: False Negative (FN) & Recall (minimized wrong predicted bad loan)
- Additional: ROC-AUC & Kolmogorov-Smirnov (KS)

Model	FN	Recall	ROC-AUC	KS
XGBClassifier	794	0.9763	0.9892	0.8835
LGBMClassifier	1063	0.9683	0.9877	0.8778
RandomForestClassifier	1371	0.9591	0.9846	0.8735

Feature Importance