CREDIT RISK EVALUATION

By Ahmad Reginald Syahiran

BACKGROUND

- A lending company is faced with the problem that it needs to lend to each customer with efficiency and speed.
- As a Data Science Intern at ID/X Partners, we will process data and create models that are able to predict and evaluate optimal credit applications and predict existing risks.

id/x partners

Rakamin Academy

BUSINESS UNDERSTANDING

1. The concept of credit risk revolves around the possibility of a borrower not repaying a loan

2. Therefore, evaluating the likelihood of repayment by the borrower is vital in managing credit risk

3. Machine learning can be utilized to streamline this evaluation process"

id/x partners

Rakamin Academy

ANALYTICAL APPROACH

1

The concept of credit risk revolves around the possibility of a borrower not repaying a loan

2

Therefore, evaluating the likelihood of repayment by the borrower is vital in managing credit risk

3

Machine learning can be utilized to streamline this evaluation process"

ANALYTICAL APPROACH

1. Analysis through description

2. Analysis using graphical representations

3. Predictive modeling via classification techniques

DATASET REQUIREMENTS & COLLECTION

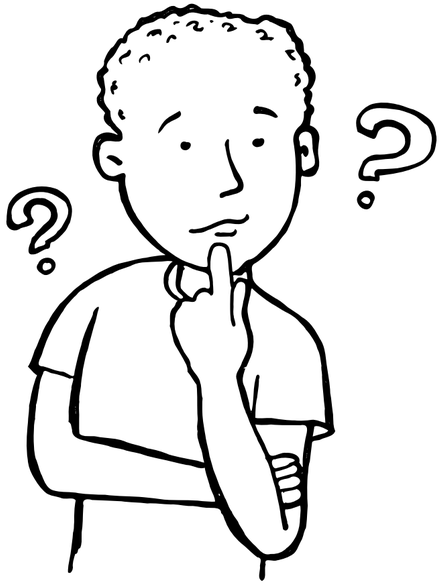
1. Dataset of customer loan from financial company

2. Dataset is collected by ID/X Partners from a company

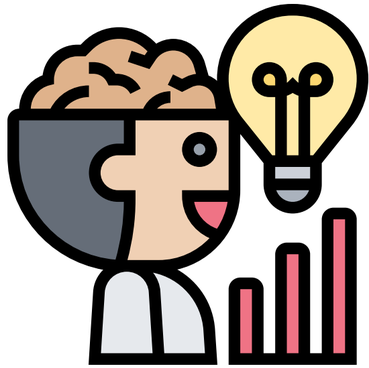
id/x partners

Rakamin Academy

DATA UNDERSTANDING



1. This dataset has 75 columns/ features Consists of 53 numerical & 22 non-numerical features.
2. The percentage of missing values was analysed and sorted.
3. There is no duplicate data in the dataset.
4. Found 17 columns with complete data, suggesting their removal.



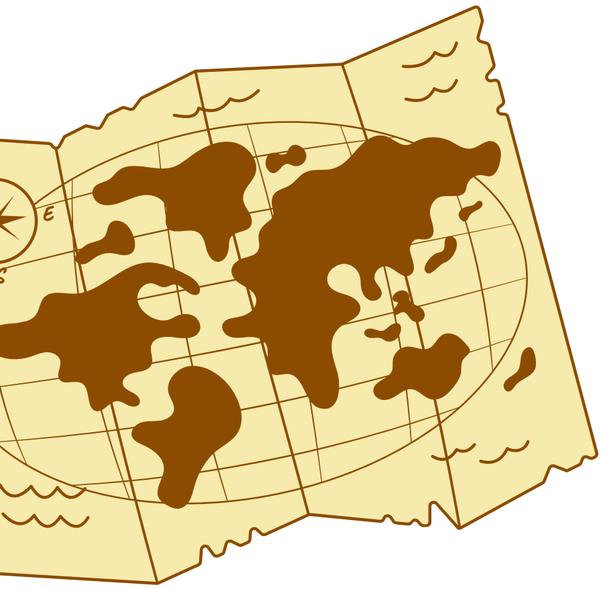
id/x partners

Rakamin Academy

DATA PREPARATION

1. Missing Value: Removing and Imputing
2. Feature Selection using Correlation Analysis
3. Machine learning can be utilized to streamline this evaluation process"
4. Removing outliers using IQR Method

EXPLORATORY DATA ANALYSIS



TARGET VARIABLE

1. Good Loan (1) : Fully Paid, Does not meet the credit policy. Status:Fully Paid

2. Bad Loan (0) : Charged Off, Does not meet the credit policy. Status:Charged Off, Default, Late (31-120 days)

id/x partners

Rakamin Academy

EXPLORATORY DATA ANALYSIS

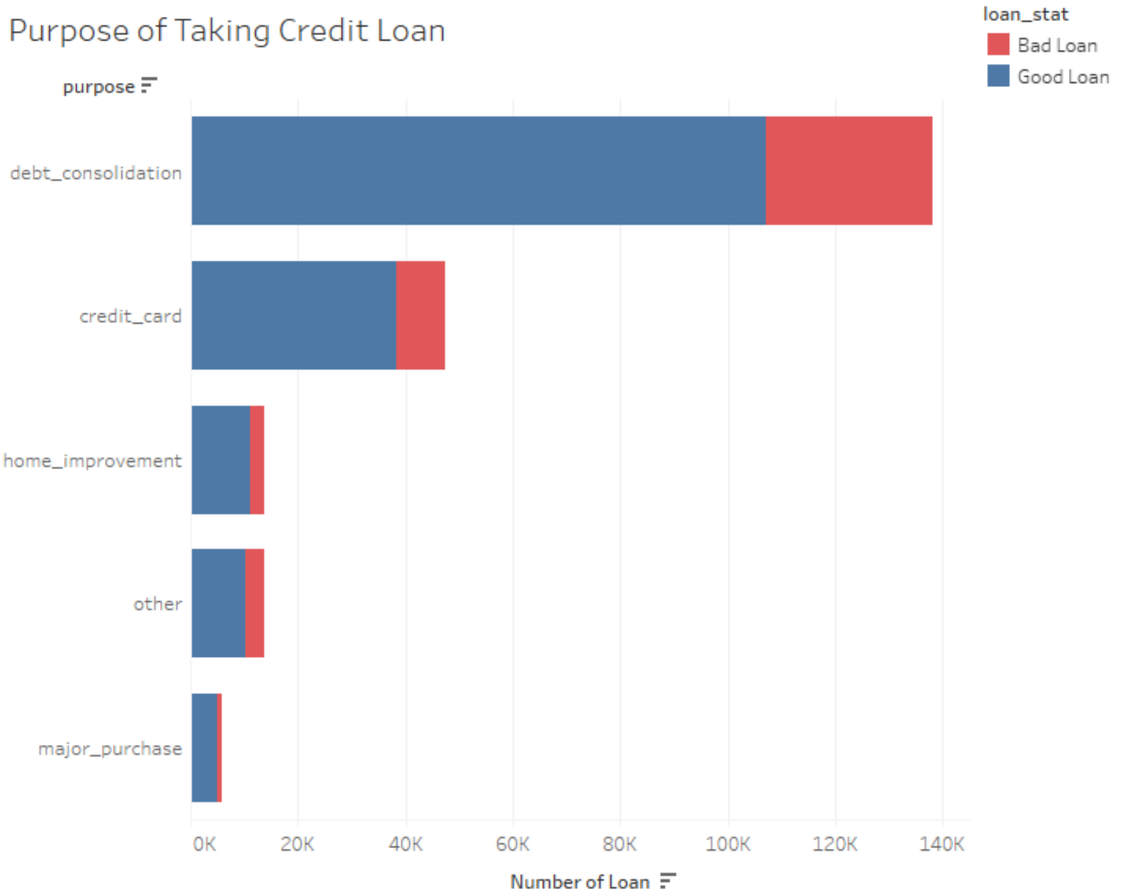
OUR BORROWERS

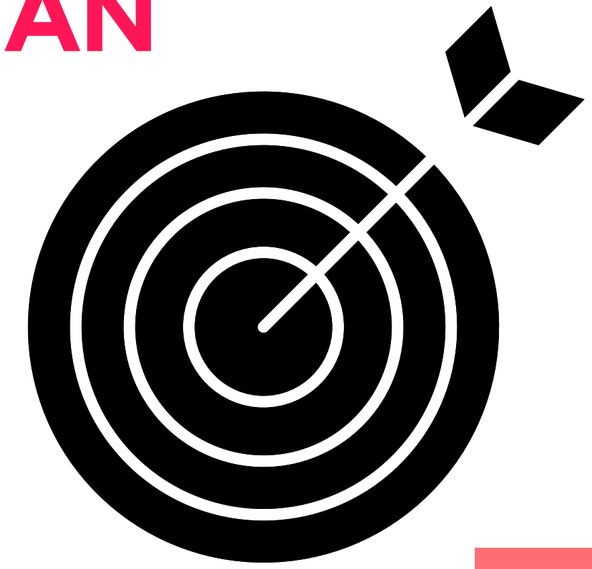
EXPLORATORY DATA ANALYSIS

PURPOSE OF TAKING CREDIT LOAN

BAD LOAN

GOOD LOAN





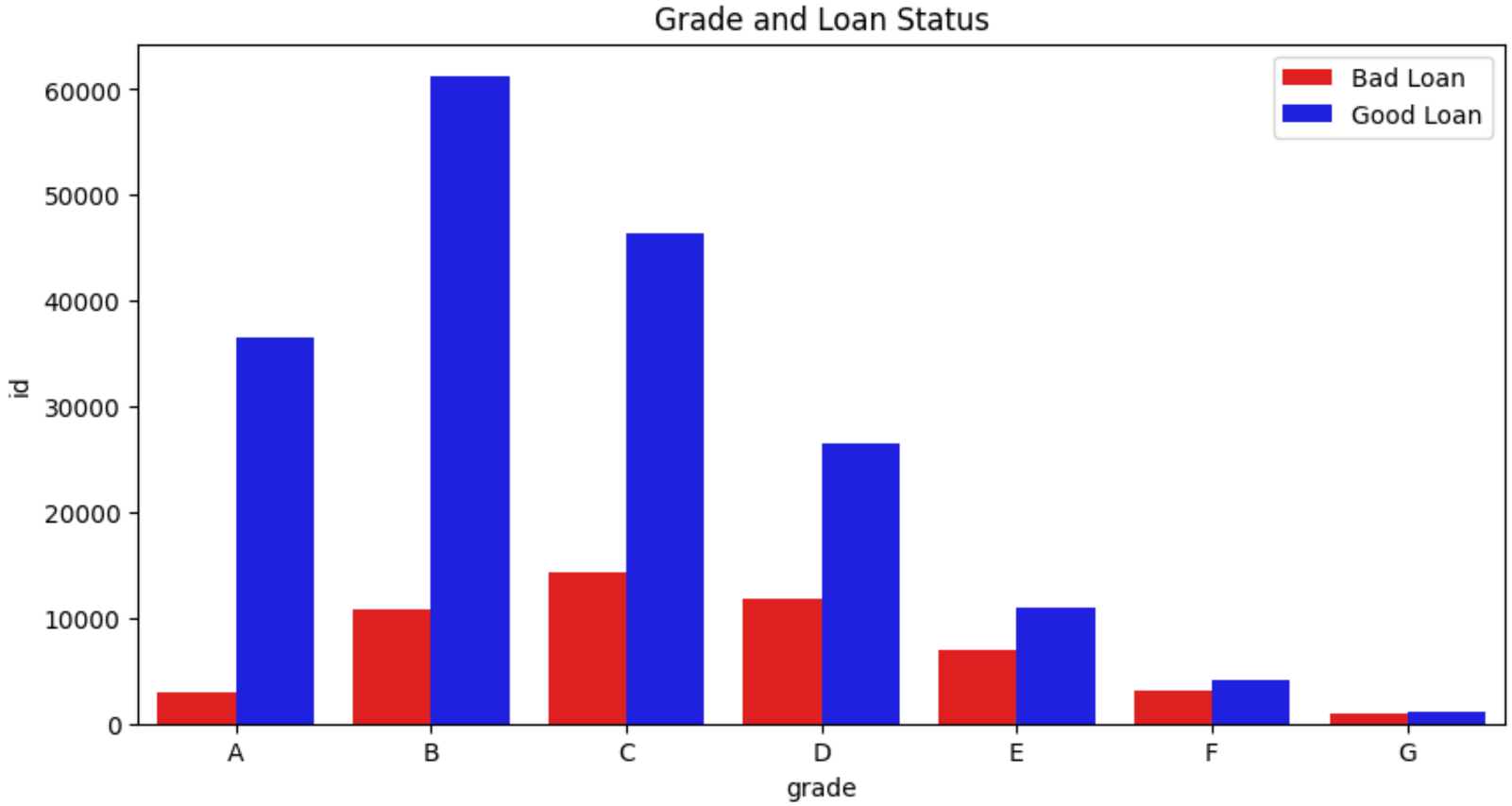
id/x partners

Rakamin Academy

EXPLORATORY DATA ANALYSIS

CLASSIFYING GRADE

classifying grade towards our borrower and the loan status



BAD LOAN

GOOD LOAN

MODELLING & EVALUATION

1. 70% Training & 30% Testing
2. I used SMOTE for handling imbalanced class
3. All steps are handled by Pipeline

EVALUATION METRICS

1. Main: False Negative (FN) & Recall from “0” (I minimized wrong predicted bad loan)
2. Additional: ROC-AUC & Kolmogorov-Smirnov (KS)

id/x partners

Rakamin Academy

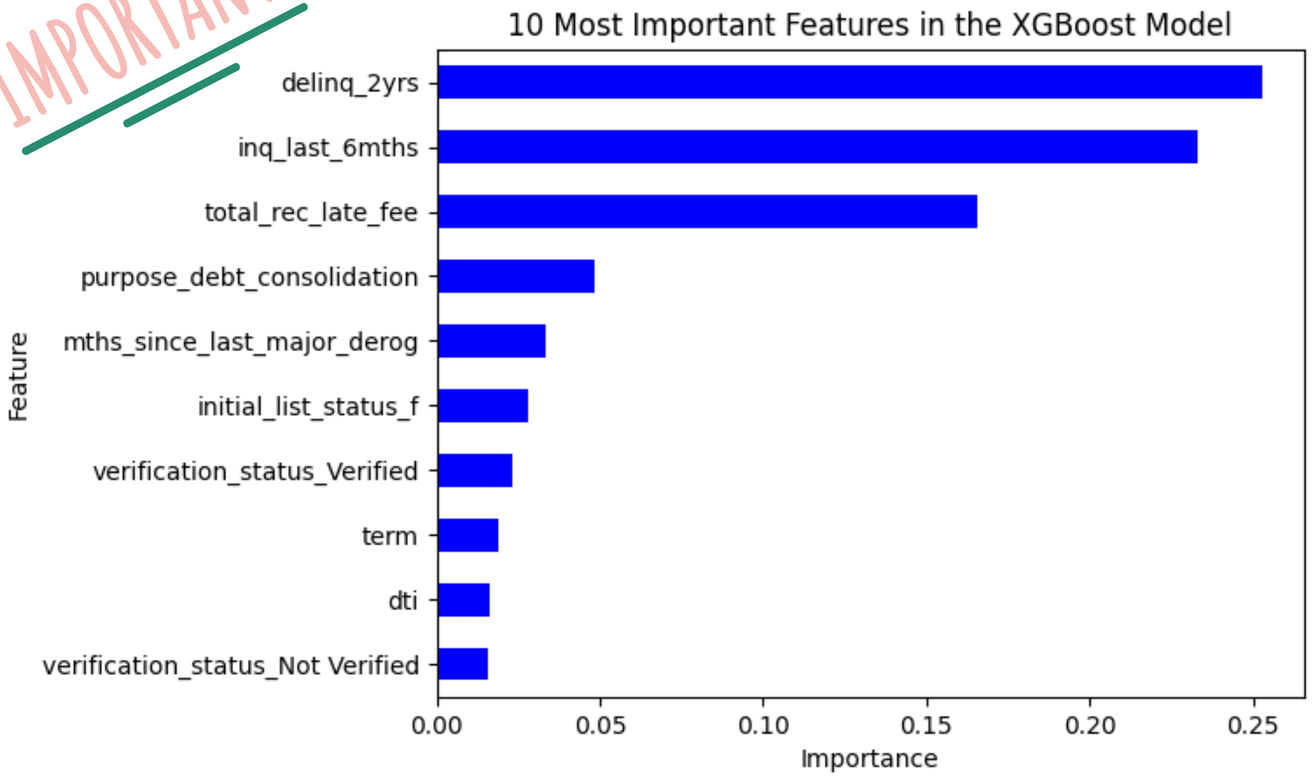
MODELLING & EVALUATION

Model	FN	Recall	ROC-AUC	KS
Random Forest	625	96%	99.43%	94.43%
Gradient Boosting Tress	386	97%	99.48%	94.28%
XGBoost	420	97%	99.76%	96.27%
Voting Classifier	380	97%	99.66%	95.08%

SELECTED MODEL

TOP FEATURES IN XGBOOST MODEL

IMPORTANT!



Top 3 feature importances in predicting credit risk is good or bad:

- delinq_2yrs: The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
- inq_last_6mths: Number of credit inquiries in past 12 months
- total_rec_late_fee: Late fees received to date

THANK YOU

Ahmad Reginald Syahiran