Final Project VIX ID/X Partners

CREDIT RISK

BY : Muhammad Fathan Muttaqy

INTRODUCTION

Freshgraduate from agriculture majoring Agrotechnology from Jenderal Soedirman University. Have a basic understanding of data analysts and deeply interested in data related role such as Data Science, Data Analysis, and Machine Learning. Skilled in Statistical Data Analysis, and Research using SQL, Python, Excel tools. Able to quickly adapt to team and individual work, hard worker, problem solver, multitasking, and fast in learning and adapting to new environments.

OUTLINE

BUSINESS UNDERSTANDING

ANALYTICAL APPROACH

DATA REQUIREMENT & DATA COLLECTION

DATA UNDERSTANDING

DATA PREPARATION

EXPLORATORY DATA ANALYSIS

MODELLING & EVALUATION

BUSINESS UNDERSTANDING

Credit risk (credit risk) is the risk of loss associated with the possibility of failure of the counterparty to fulfill its obligations or simply, credit risk is the risk that the Borrower does not pay what is owed. Therefore, it is important to predict the borrower's risk of repaying the loan. The objective of Credit Risk is to maximize the Risk-adjusted Rate of Return from Financial Institutions by maintaining Credit Risk Exposure within acceptable parameters. Therefore, the use of machine learning can help in the prediction process automatically.

ANALYTICAL APPROACH

- Descriptive analysis
- Visualization
- Predictive modelling (classification)

DATA REQUIREMENT & DATA COLLECTION

Dataset of customer loan from Lending Company (Collected by id/x Partners. Collaborating with other departments in this project to provide technology solutions for the company. What is being done is to build a model that can predict credit risk using a dataset provided by the company which consists of data on loans received and rejected.

DATA UNDERSTANDING & DATA PREPARATION

DATA UNDERSTANDING

DATASET Info:

- Credit Risk Dataset has 74 freatures
- Consists of 52 numerical & 22 non-numerical features
- Have 17 columns with full null value

DATA PREPARATION

- Drop & Imputing Missing Value
- Feature Engineering : Categorical Encoding Log Transform, Standardization
- Feature Selection : Correlation Analysis
- Handle Outlier : IQR Method

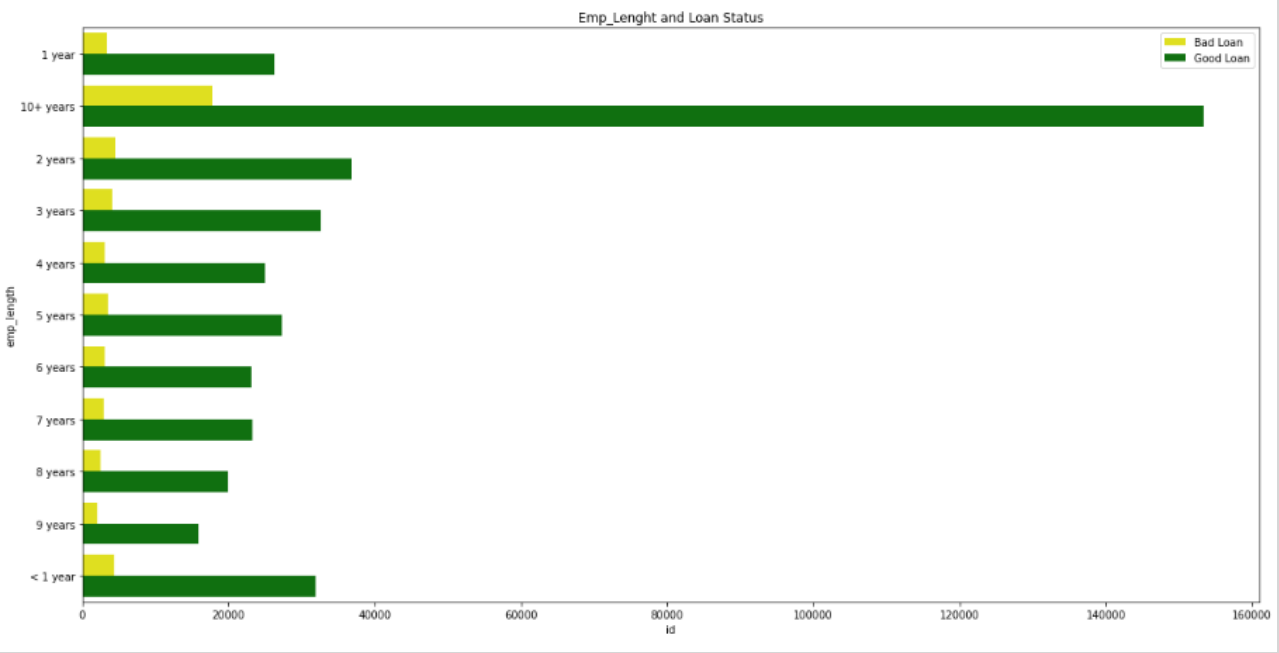
EXPLORATORY DATA ANALYSIS

Target Variables

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

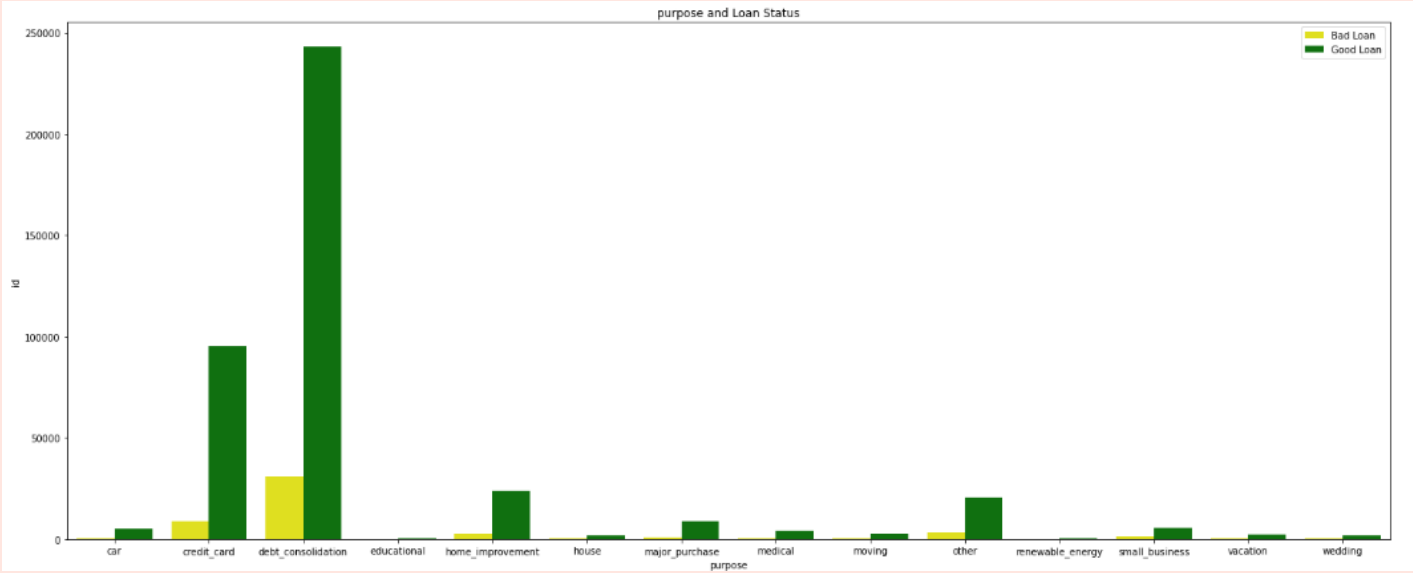
EXPLORATORY DATA ANALYSIS

Does the employment length have an impact to good or bad loan?

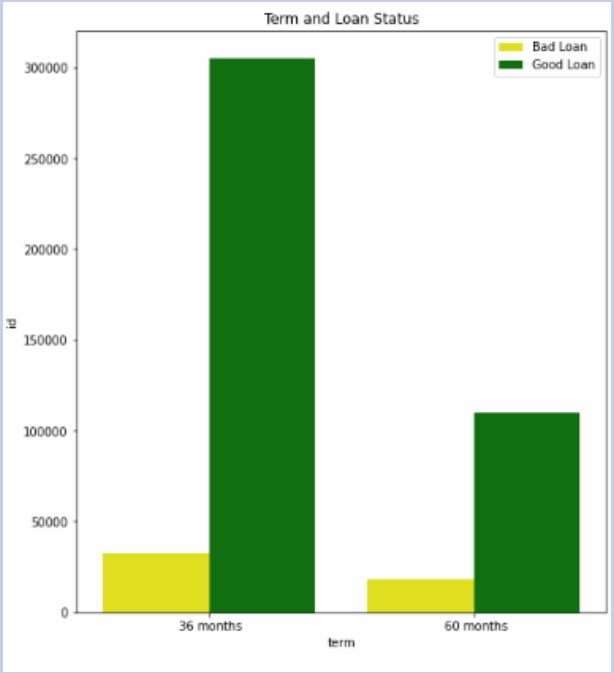


EXPLORATORY DATA ANALYSIS

Why do our borrower take credit loan?



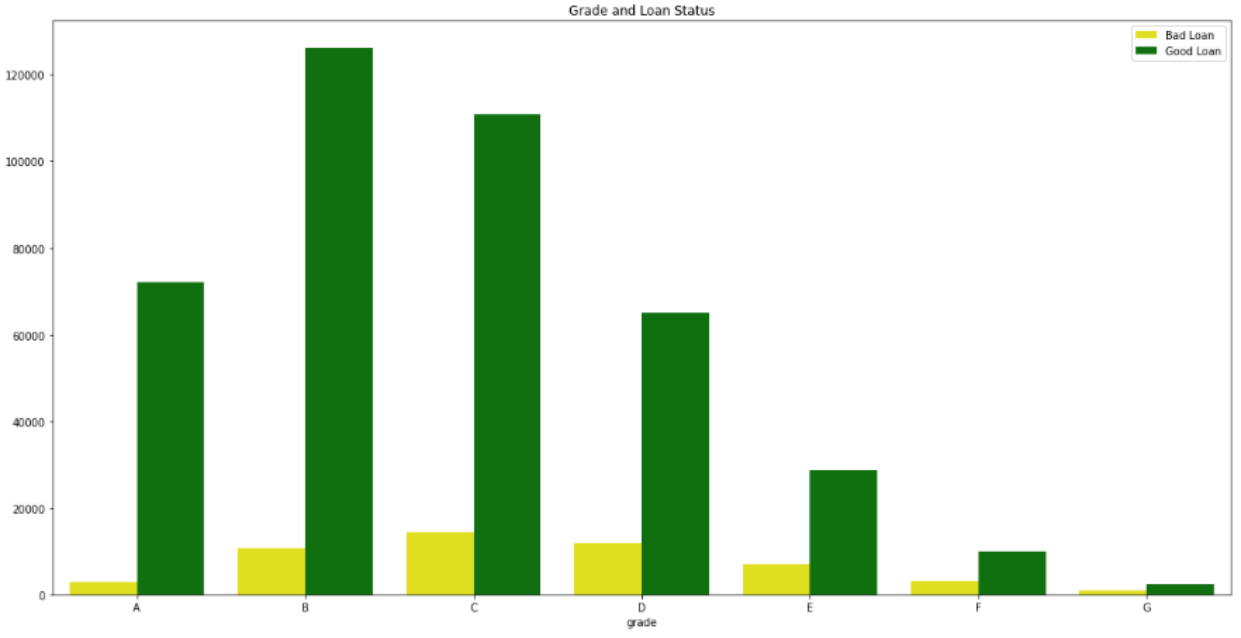
EXPLORATORY DATA ANALYSIS



What can be concluded from good or bad loans based on their number of payments on the loan (term)?

EXPLORATORY DATA ANALYSIS

How about classifying grade towards our borrower and the loan status?



MODELLING & EVALUATION

SPLIT DATA

70% Training & 30% Testing

IMBALANCE CLASS

Handling Imbalance Class Using SMOTE

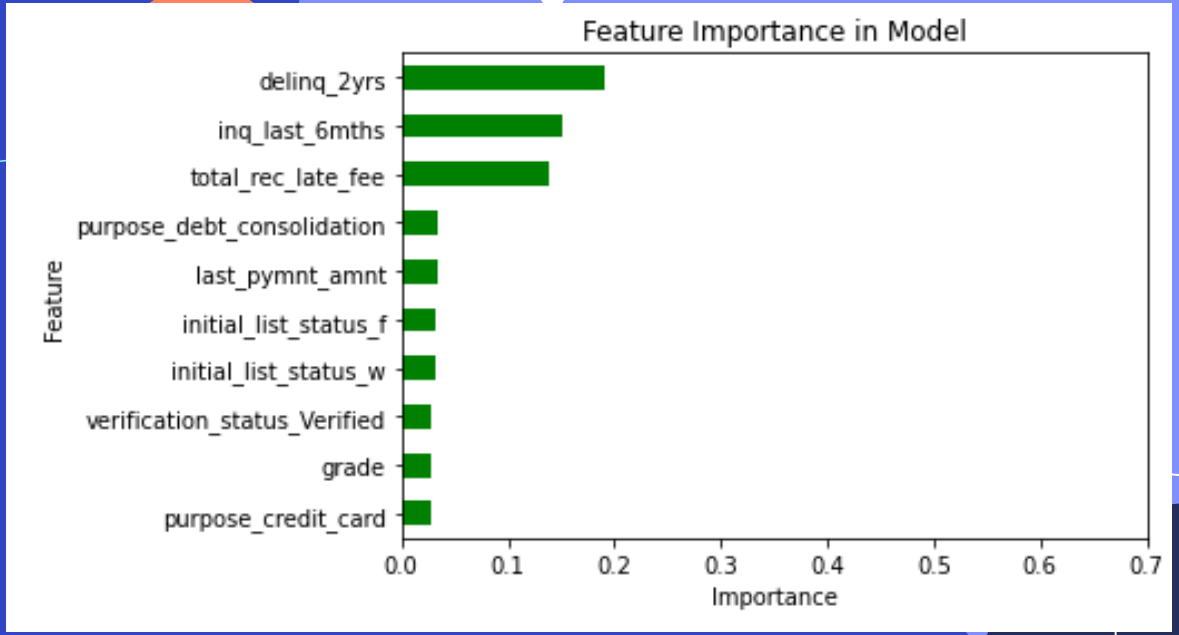
EVALUATION MATRICS

To minimized wrong predicted, main metrics : Recall Additional : ROC-AUC & Kolmogorov-Smirnov (KS)

MODEL	RECALL	ROC-AUC	KS
Logistics Regression	0.87	0.9368	0.7369
Random Forest	0.84	0.9505	0.7606
XGBoost Classifier	0.84	0.9421	0.7226
Vooting Classifier	0.85	0.9493	0.7493
Gradient Boosting Classifier	0.84	0.9421	0.7226

FEATURE IMPORTANCE

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



Deling_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