CREDIT RISK ASSESSMENT

VIRTUAL INTERNSHIP EXPERIENCE ID/X PARTNERS

Created by Yohanes Setiawan







Business Understanding

- Credit risk is known as the risk of borrower's failure to repay a loan
- Assessing borrower's risk to repay the loan is a crucial thing in credit risk assessment
- We can use machine learning to automate the process

Tools:







- Descriptive analysis
- Graph analysis
- Predictive modelling (classification)

Analytical Approach



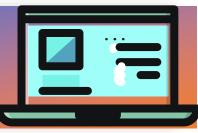


Data Requirements & Collection

- I required a dataset of customer loan from financial company
- The dataset is collected by ID/X Partners from a company

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

Data Understanding



CREDIT RISK ASSESSMENT

VIRTUAL INTERNSHIP EXPERIENCE ID/X PARTNERS

Created by Yohanes Setiawan







Data Preparation

- Missing Value: Removing and Imputing
- Feature Engineering: Categorical Encoding, Log Transform, Standardization
- Feature Selection using Correlation **Analysis**
- Removing outliers using IQR Method

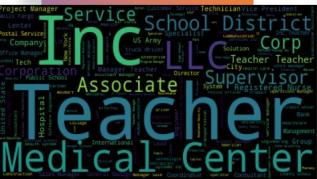
Target Variable

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

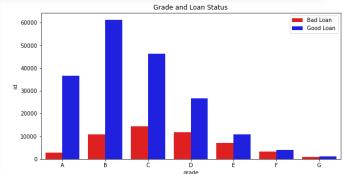
Exploratory Data Analysis

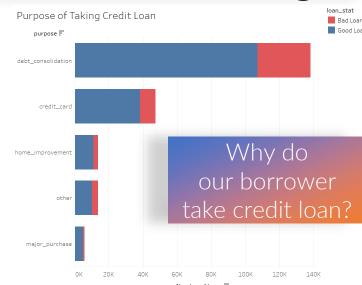


What are the employee titles of our borrowers?



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





CREDIT RISK ASSESSMENT

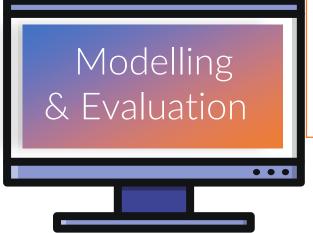
VIRTUAL INTERNSHIP EXPERIENCE ID/X PARTNERS

Created by Yohanes Setiawan



Id/x partners

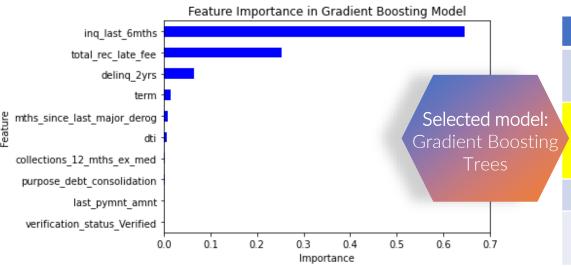




- 70% Training & 30% Testing
- I used SMOTE for handling imbalanced class
- All steps are handled by Pipeline

Evaluation Metrics:

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



	Model	FN	Recall	ROC-AUC	KS
	Random Forest	608	96%	99.41%	94.40%
	Gradient Boosting Trees	386	97%	99.48%	94.28%
	XGBoost	447	97%	99.43%	93.83%
	Voting Classifier	420	97%	99.48%	94.34%