ANALYSIS CREDIT RISK ASSESSMENT

Business Understanding

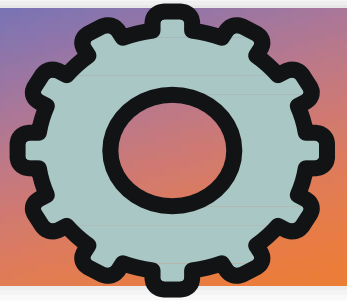
1. Credit risk is known as the risk of borrower's failure to repay a loan
2. Assessing borrower's risk to repay the loan is a crucial thing in credit risk assessment
3. We can use machine learning to automate the process

Tools

python matplotlib scikit learn tableau

1. Descriptive analysis
2. Graph analysis
3. Predictive modelling (classification)

Analytical Approach



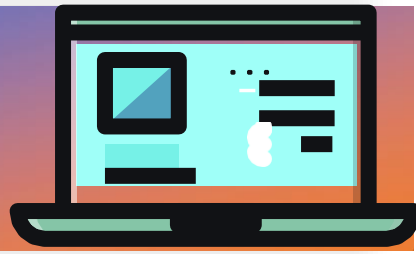
Data Requirements & Collection

1. I required a dataset of customer loan from financial company
2. The dataset is collected from a company

1. This dataset has 74 columns / features
2. Consists of 52 numerical & 22 non-numerical features

1. Many features have missing values
2. There are 17 null features

Data Understanding



ANALYSIS CREDIT RISK ASSESSMENT

Data Preparation

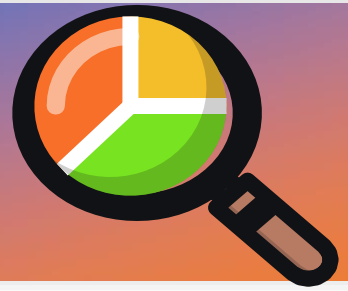
1. Missing Value: Removing and Imputing
2. Feature Engineering: Categorical Encoding, Log Transform, Standardization

1. Feature Selection using Correlation Analysis
2. Removing outliers using IQR Method

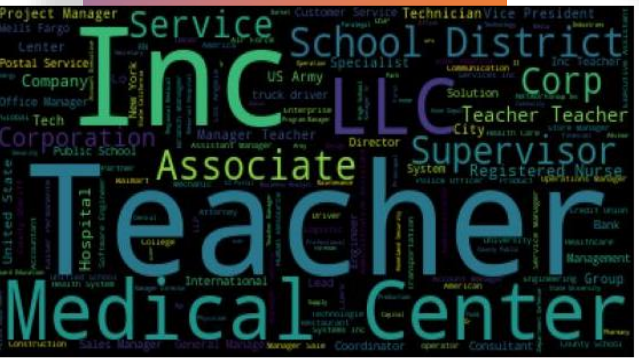
Target Variable

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)

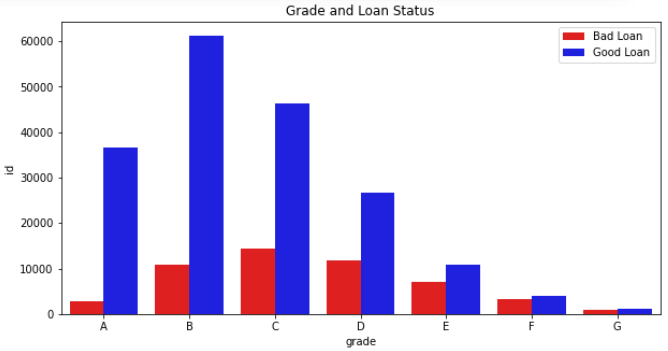
Exploratory Data Analysis

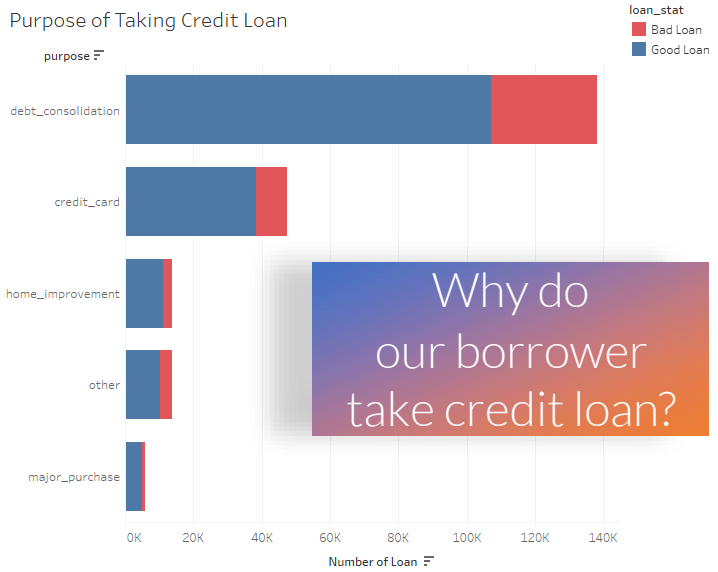


What are the employee titles of our borrowers?



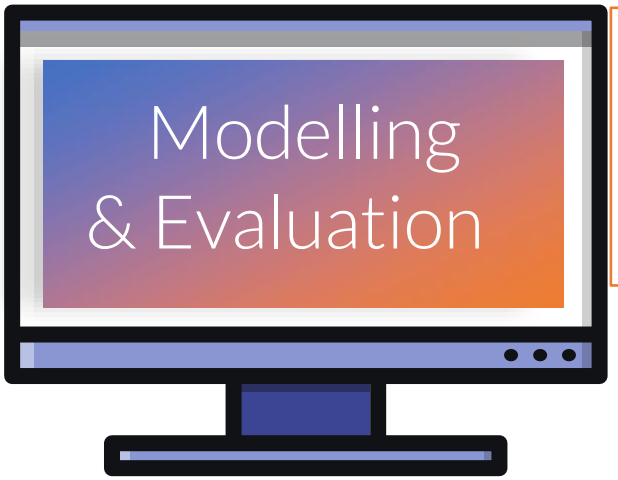
How about classifying grade towards our borrower and the loan status?





Why do our borrower take credit loan?

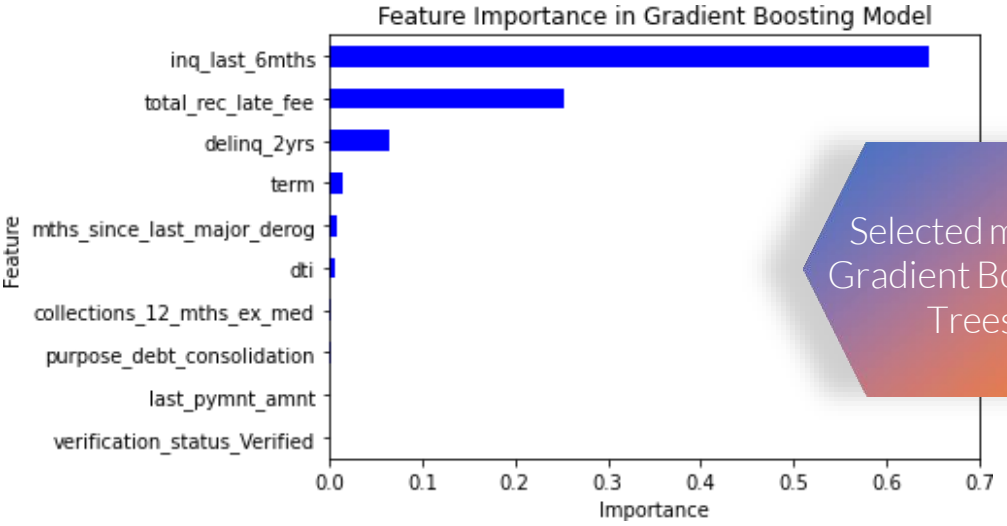
ANALYSIS CREDIT RISK ASSESSMENT



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 & KolmogorovSmirnov (KS)



Selected model: Gradient Boosting Trees

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%