

Business Understanding

- Credit risk refers to the potential of a borrower's inability to repay a loan, which can adversely impact the lender's financial well-being.

Analytical Approach

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

Data requirements & collection

- Require a dataset of customer loans from a financial company.
- The dataset was collected by ID/X Partners on behalf of the company.

Data understanding

- Dataset has 74 columns / features
- consist of 52 numerical & 22 non-numerical features
- many features have missing values
- there are 17 null features

Data preparation

- Missing value: remove and imputing
- feature engineering: categorical encoding, log transform, standardization
- feature selection: using correlation analysis

Exploratory data analysis

- 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)

Modeling & evaluation

- 70% Training & 30% testing
- used SMOTE for handling imbalance class
- all steps are handled by pipeline

evaluation metrics:

- addition: ROC-AUC & Kolmogorov Smirnov (KS)