

00. Credit Risk Scoring

Problem Statement:

We want to build a credit risk model for loan approvals. How would you approach this?

Clarification of the problem:

- Whether to **approve or reject** a loan
- What **interest rate** to offer
- How much **risk reserve** to hold

1. Data Collection and Preparation:

a. Customer Details or Demographics:

customer_id, age, education, employment_type, income, credit_history, gender, martial_status, address

b. Loan (Credit) Details:

loan_amount, loan_term, interest_rate, previous_default_loans

c . Financial Behavior:

credit_score, active_credit_accounts, debt_to_income, loan_issue_date

d. Loan Payment Transaction:

transaction_id, customer_id, loan_id, issue_date, due_date, actual_amount, payment_amount, payment_status

e. Target Dependent Variable:

default 1 and no default 0

2. Exploratory Data Analysis

- Checking for **Class Imbalance**

- Checking for **Missing** and **Outliers**
- Making some Visualizations for **Distribution** and **Correlation**
- Applying Feature **Correlation** **Technics**

3. Feature Engineering

New Features:

1. `remaining_amount`
2. `days_delays`
3. `is_late_payment` → (if it is more than **5 working days** it is late **(1)** else no **(0)**)
4. `debt_burden_ratio` → (debt/income)

Categorical Data Encoding:

1. One Hot Encoding
2. Label Encoding

Data Normalization:

1. income, loan_amount, age, remaining_amount and so on

