



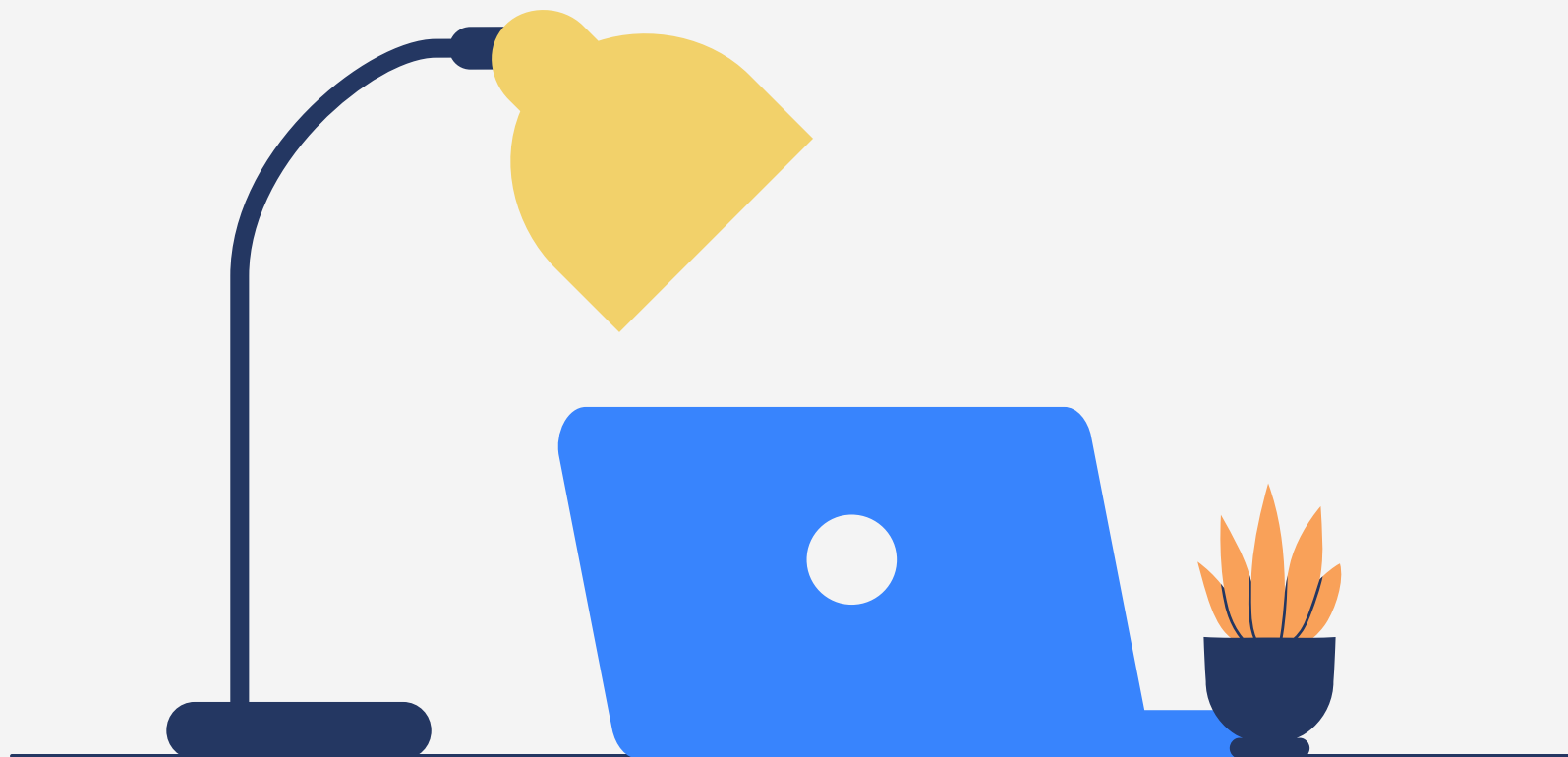
CREDIT RISK PREDICTION

by: Oky Hariawan



Problems

Credit risk refers to the risk of loss that a lender faces due to a borrower's failure to repay any type of loan or debt



Time



It will takes a long time if we do the assessment manually

Human Error



It can be happen if we make wrong decisions & will cause considerable losses

Financial Loss



Credit loss is the amount of money lost by the lender when the applicant refuses to pay or runs away with the money owed.



Goal

We aim to predict the secure loan before application is approved by implement machine learning algorithms. In addition, this model can be considered by lending companies to provide the amount of interest, the loan term and the loan period.

TABLE OF CONTENTS

Data Preparation

Data Analysis
(EDA)

Modeling

Business
Recommendation

Customer Classification Status

01 Good borrower

Fully Paid

02 Bad Borrower

Charge Off, Default, Late (16-30 days), Late (31-120 days)

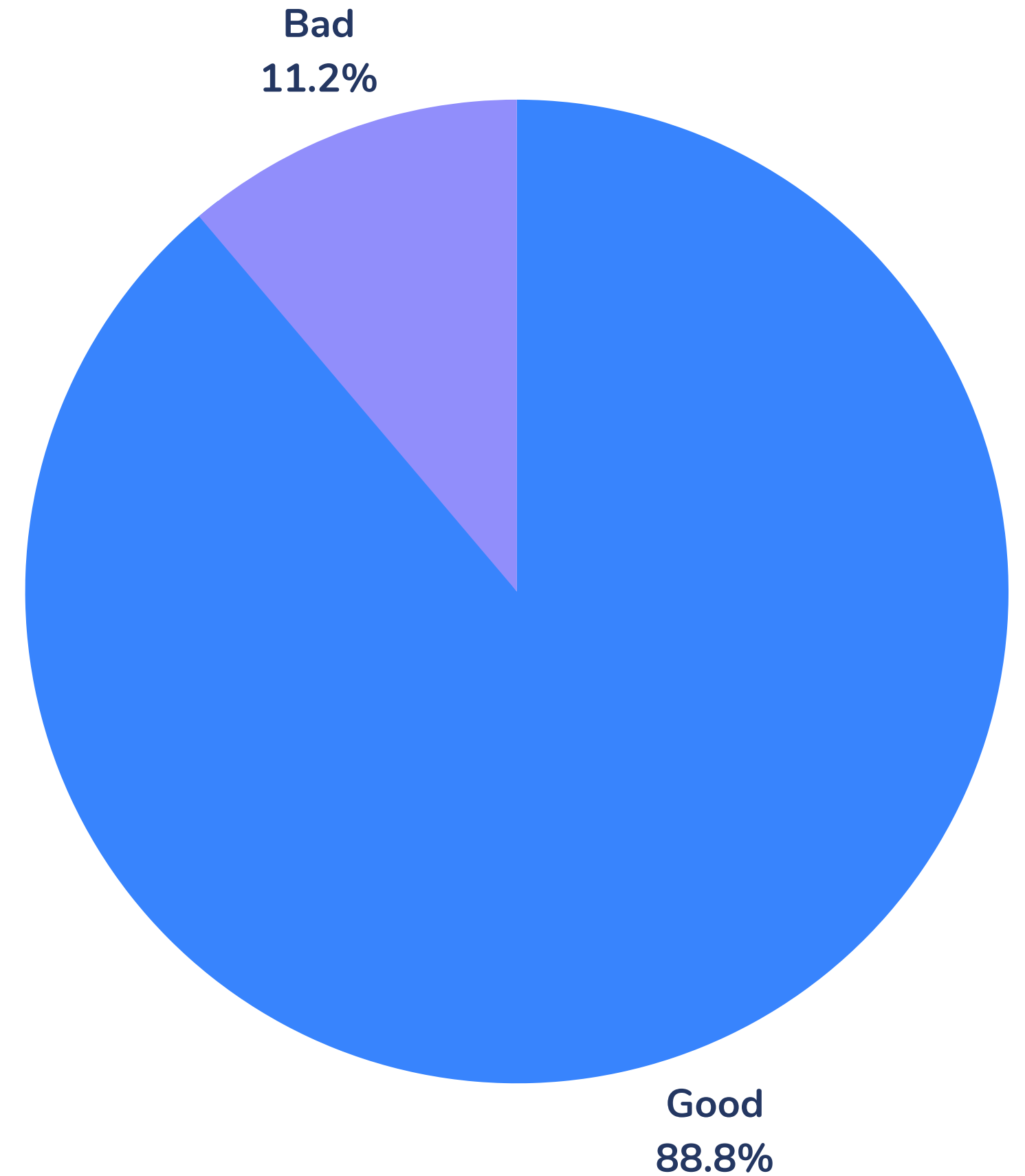
03 Ambiguous

Current, In grace period



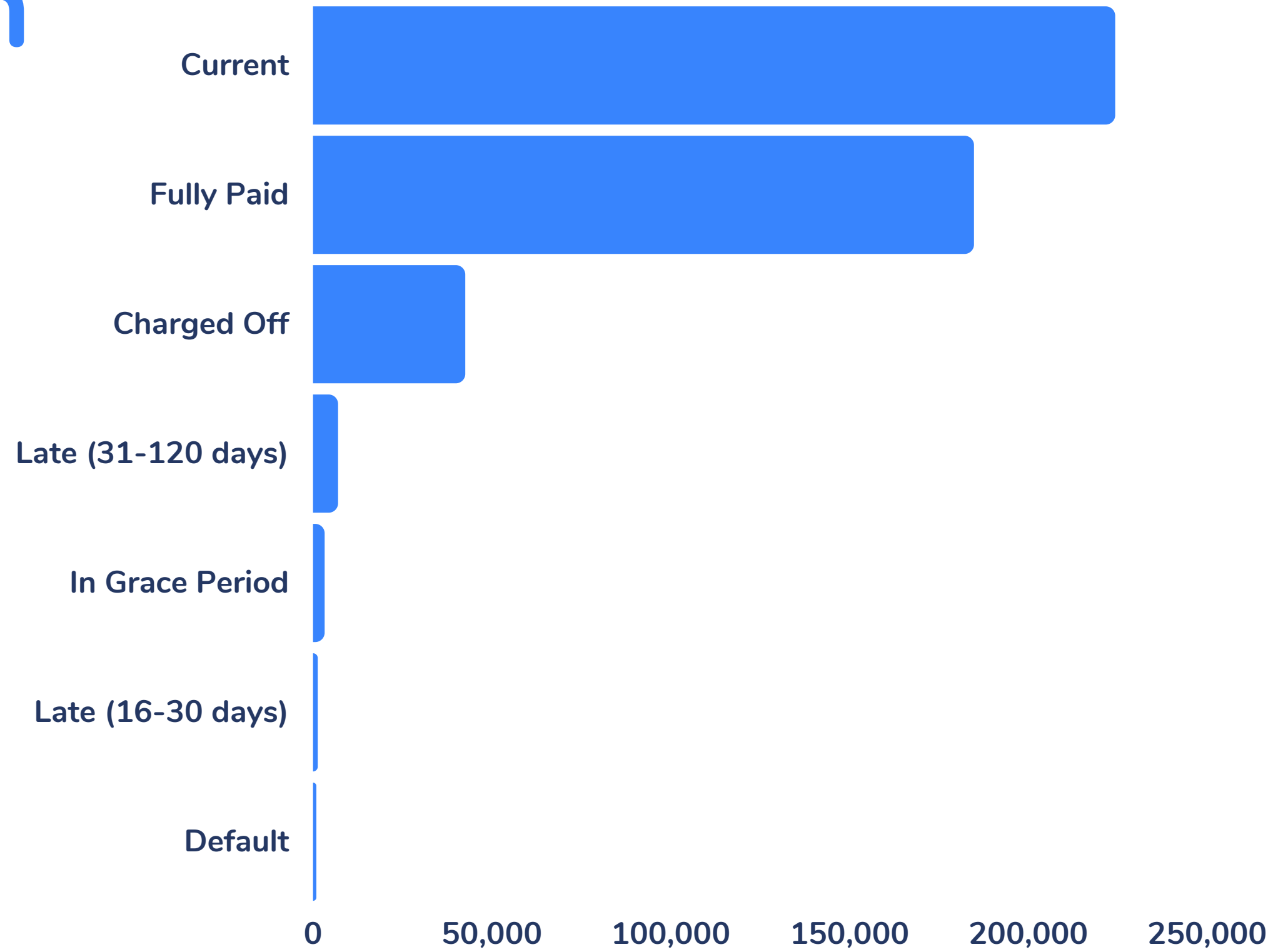
The loan's problem

We get the imbalance data which have 88,8% good borrower and 11,2% bad borrower from 2007 until 2014



Applicants by Loan Status

There are about 48% applicants with loan status of Current, followed by loan status of Fully Paid with 39.6%.



Data Preparation



Data Cleansing

- Drop feature with null value $>50\%$
- Impute feature with missing value $<20\%$ with mode & median
- Drop unnecessary feature & feature with 1 unique value
- Drop high correlation feature (>0.7) with target



Feature Engineering

- Change datatype for some feature with incorrect datatype
- Feature selection using weight of evidence and information value
- Encode all features for modeling with label encoding and one hot encoding

Before & After Data Preparation



75 Feature
466285 Rows



23 Feature
466285 Rows

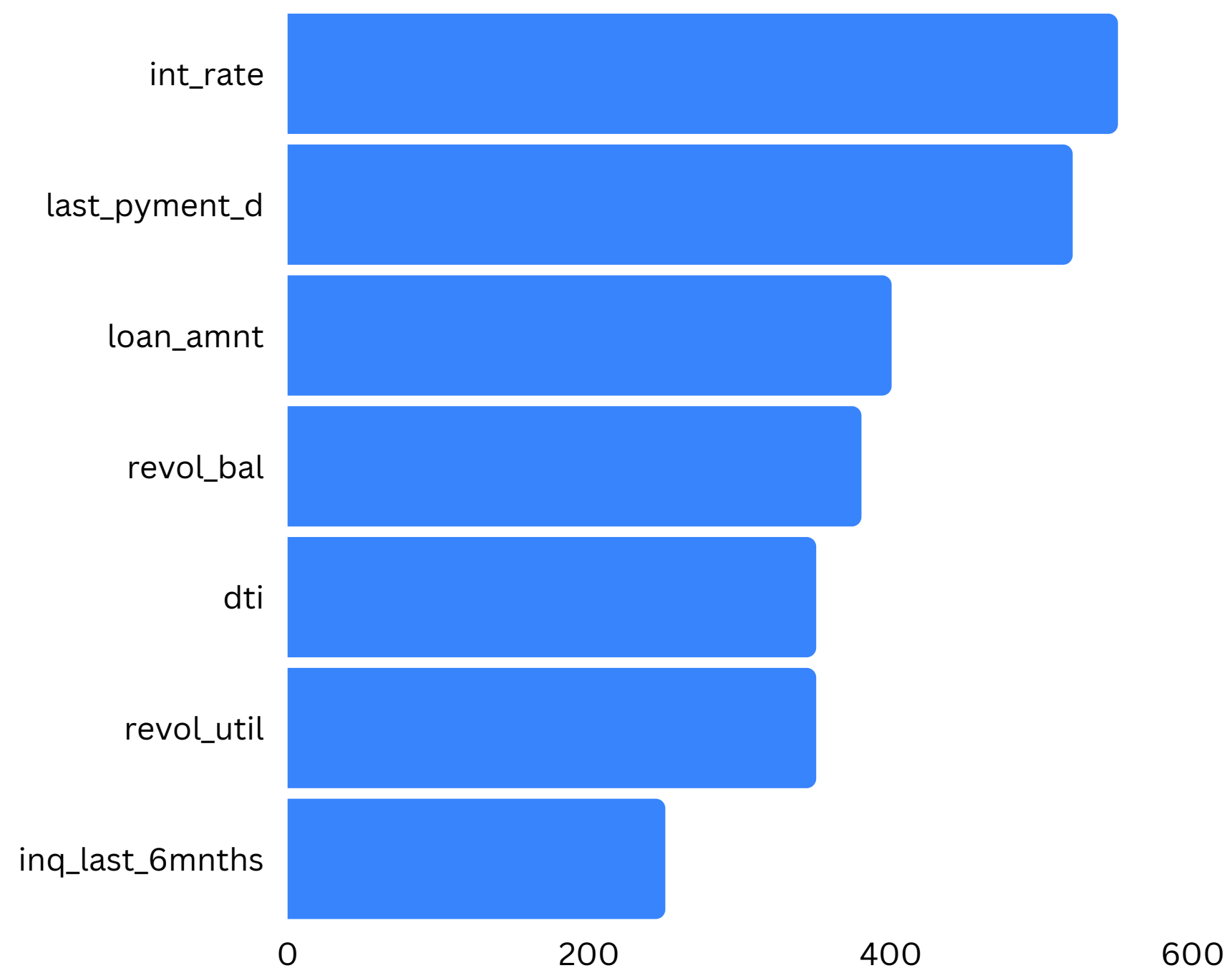
Modelling & Evaluation

Model	Accuracy	AUC	Duration
K-Nearest Neighbor	87.3%	55.1%	0.22
XgBoost	88.8%	85.5%	26.3
Random Forest	88.7%	84.6%	64.2
Gradient Boosting	88.8%	85.5%	69.7
LightGBM	88.9%	85.9%	4.1

- Metrics evaluation that important for this model is **AUC** Score
- Best model is **LightGBM** with the best score & short duration

Actual	Predict	
	Bad	Good
Bad	1758	14494
Good	964	122670

Feature Importance



int_rate

Interest Rate on the loan

last_pymnt_d

Last month payment was received

loan_amnt

The listed amount of the loan applied for by the borrower

revol_bal

Total credit revolving balance

revol_bal

Total credit revolving balance

dti

A ratio calculated using the borrower's total monthly debt payments on the total debt obligations

revol_util

Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.

inq_last_6mnths

The number of inquiries in past 6 months (excluding auto and mortgage inquiries)

Business Recommendation

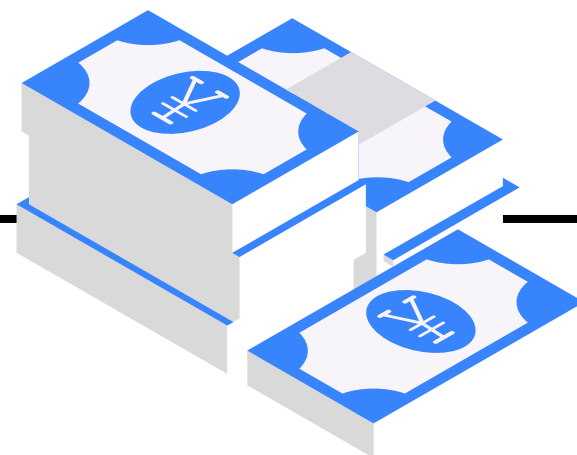


Simulation

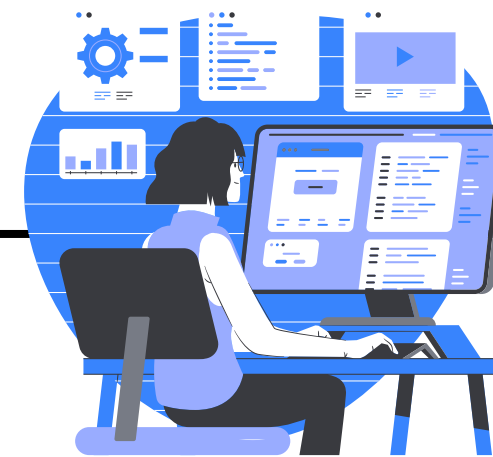
With Machine Learning



Bank Partners



Credit Loan



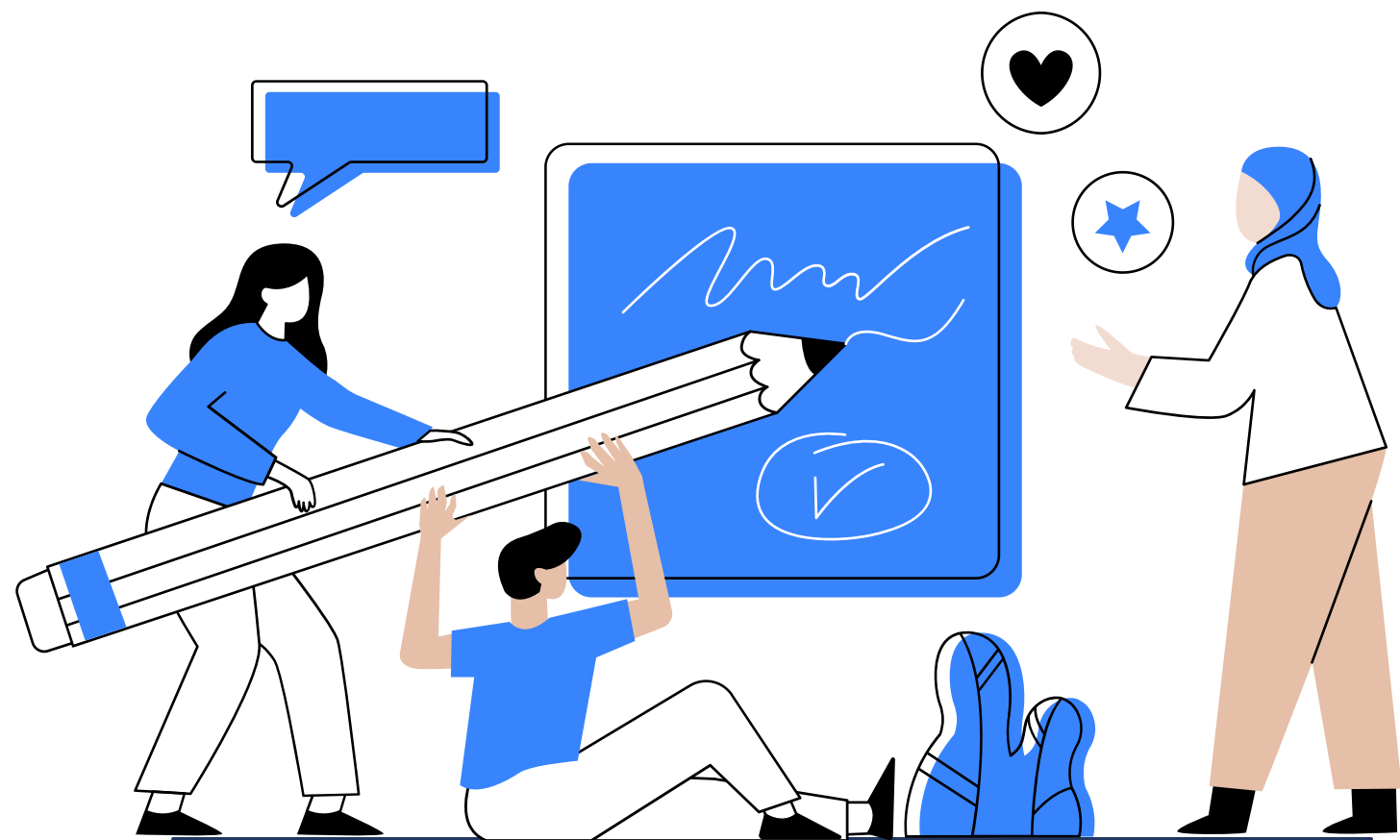
**Automate Using
Machine Learning**



Result:

- Bad Loan decrease 21.8% >> 2.4%
- Suppress the loss due to error determining credit risk

Business Recommendation



Feature Importance

Needs to monitor feature importance because its important features to identify whether the applicant has the possibility of not repaying the loan & to reduce the risk of loss.

Interest Rate

More interest rate that customer take, increasing more bad loan rate. Ideally if lending companies want to keep bad loan low, they can offer interest rate below 13%.

Loan Amount

The amount of the loan given is related to the interest rate that must be paid. The larger the loan amount, the higher the interest rate that must be paid.

Thank you



Github Repositories