

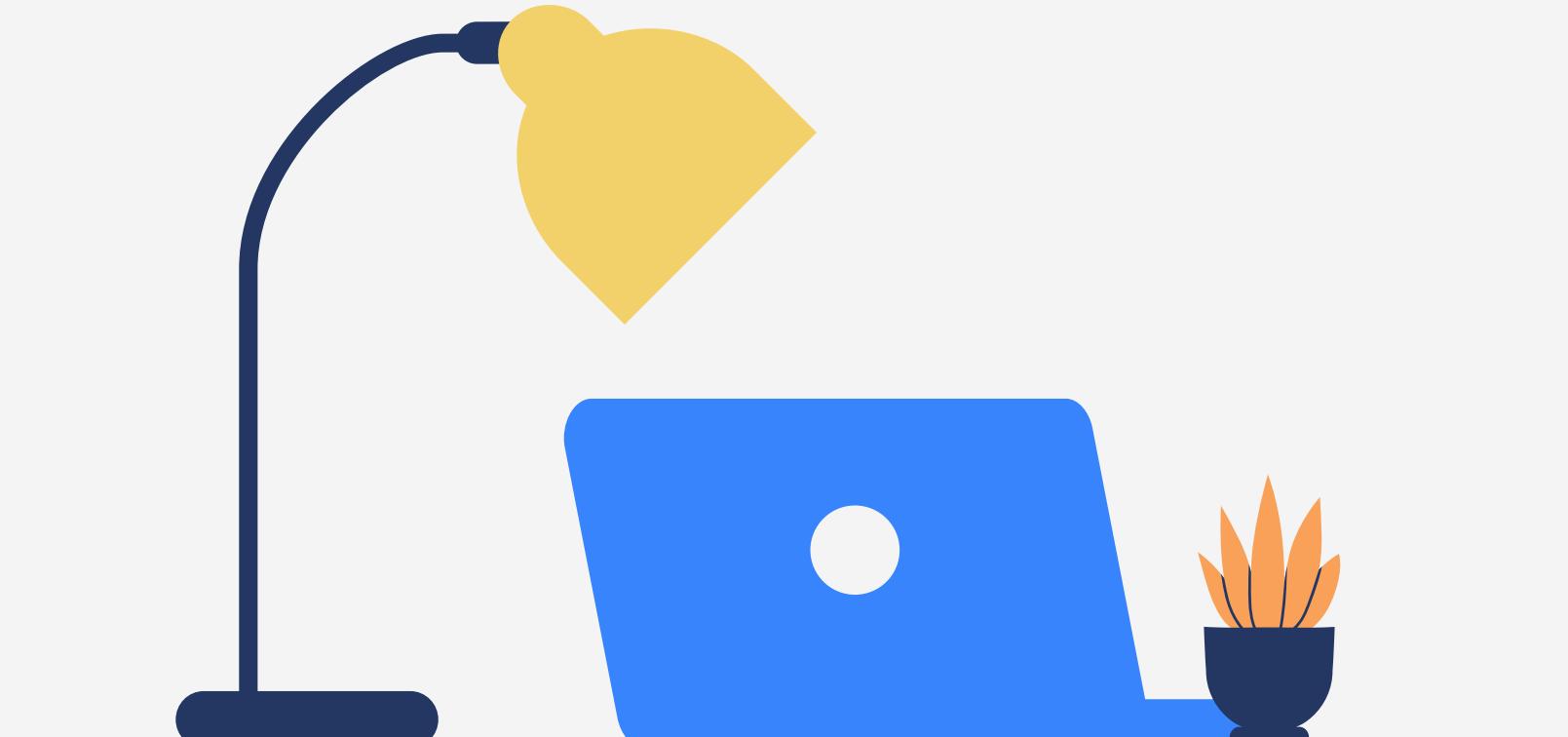
# CREDIT RISK PREDICTION

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  
Business  
Recommendation

Data Analysis  
(EDA)

Modeling

# Customer Classification Status

01 **Good borrower**

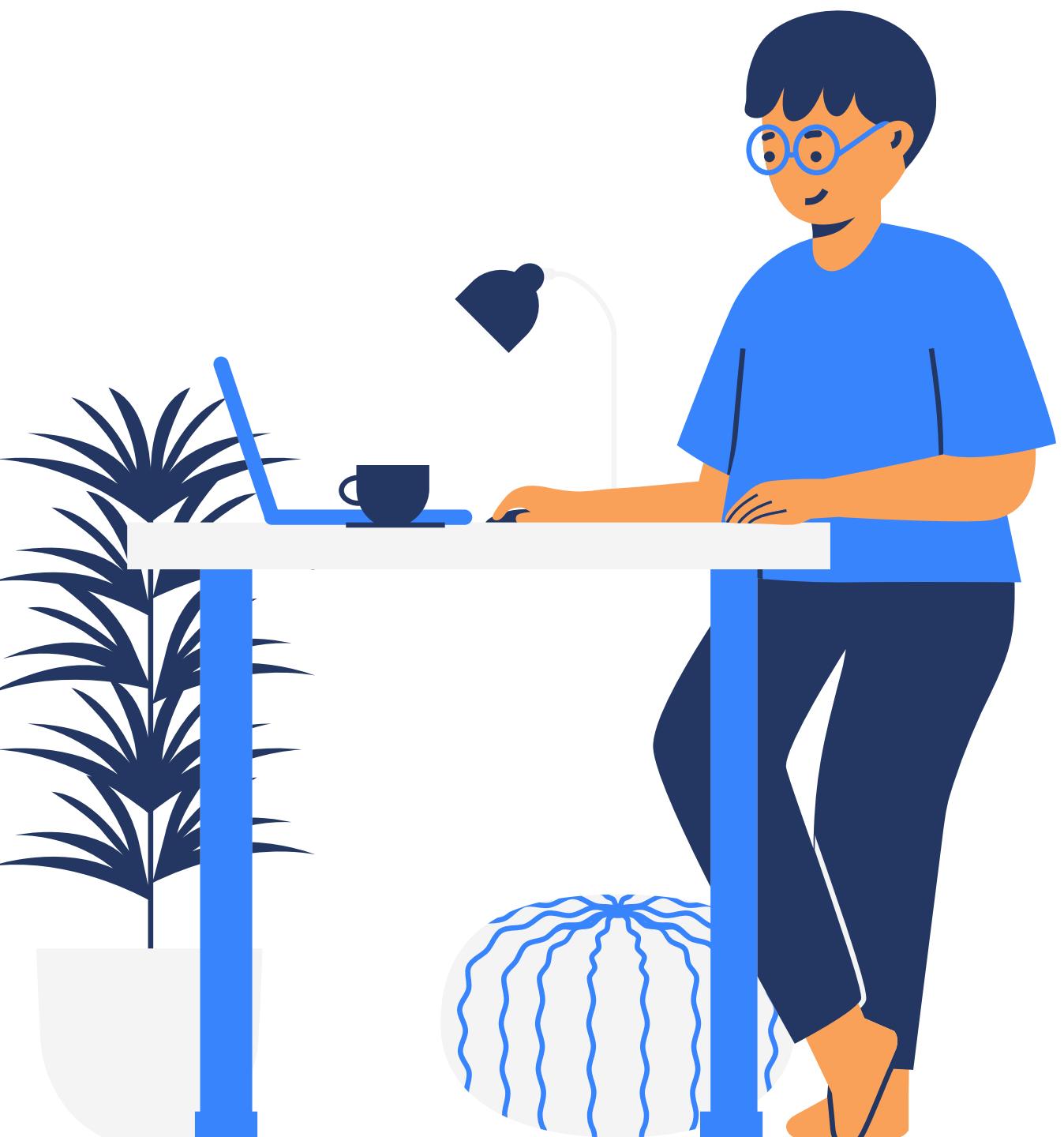
Fully Paid

02 **Bad Borrower**

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

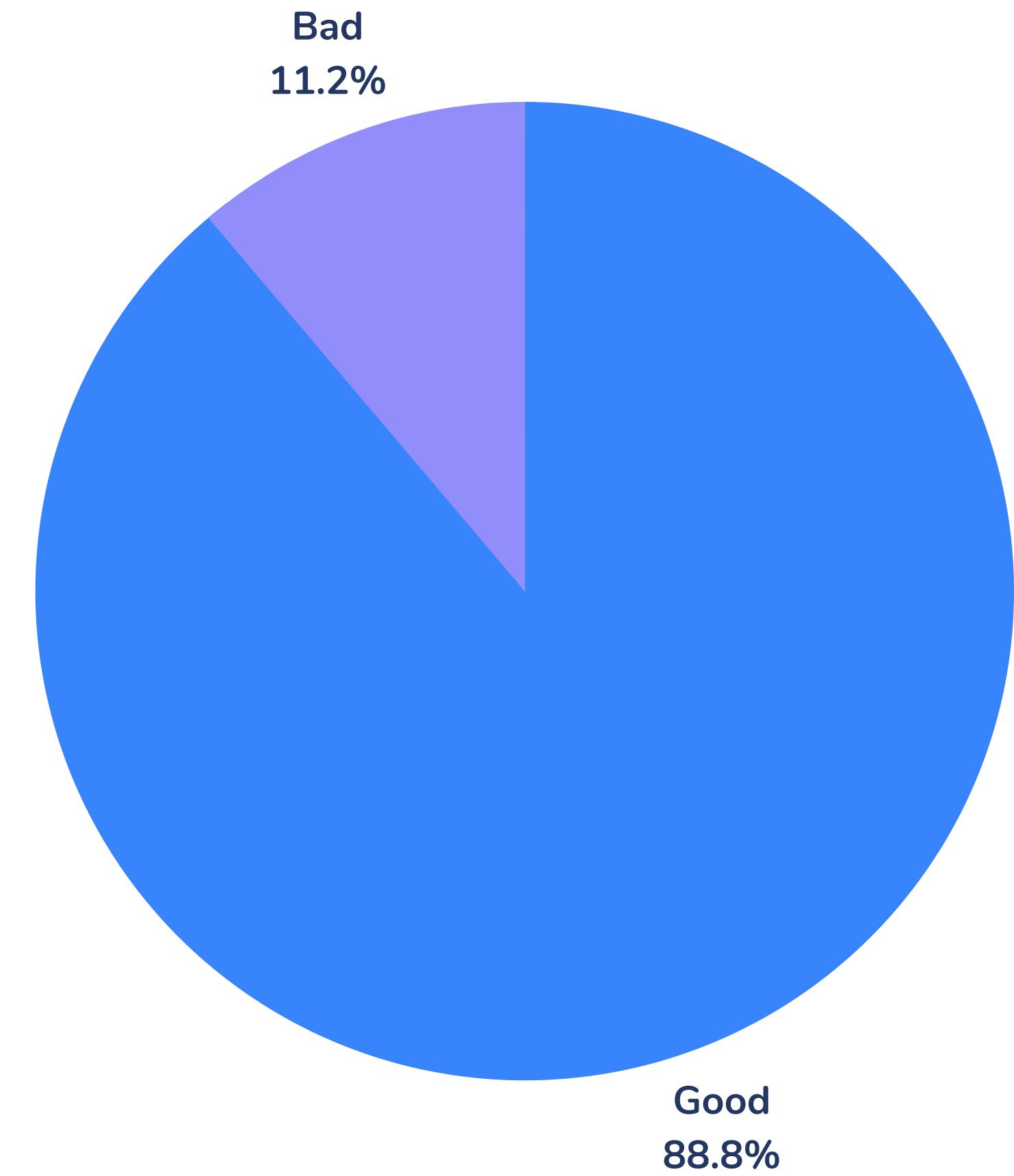
03 **Ambiguous**

Current, In grace period



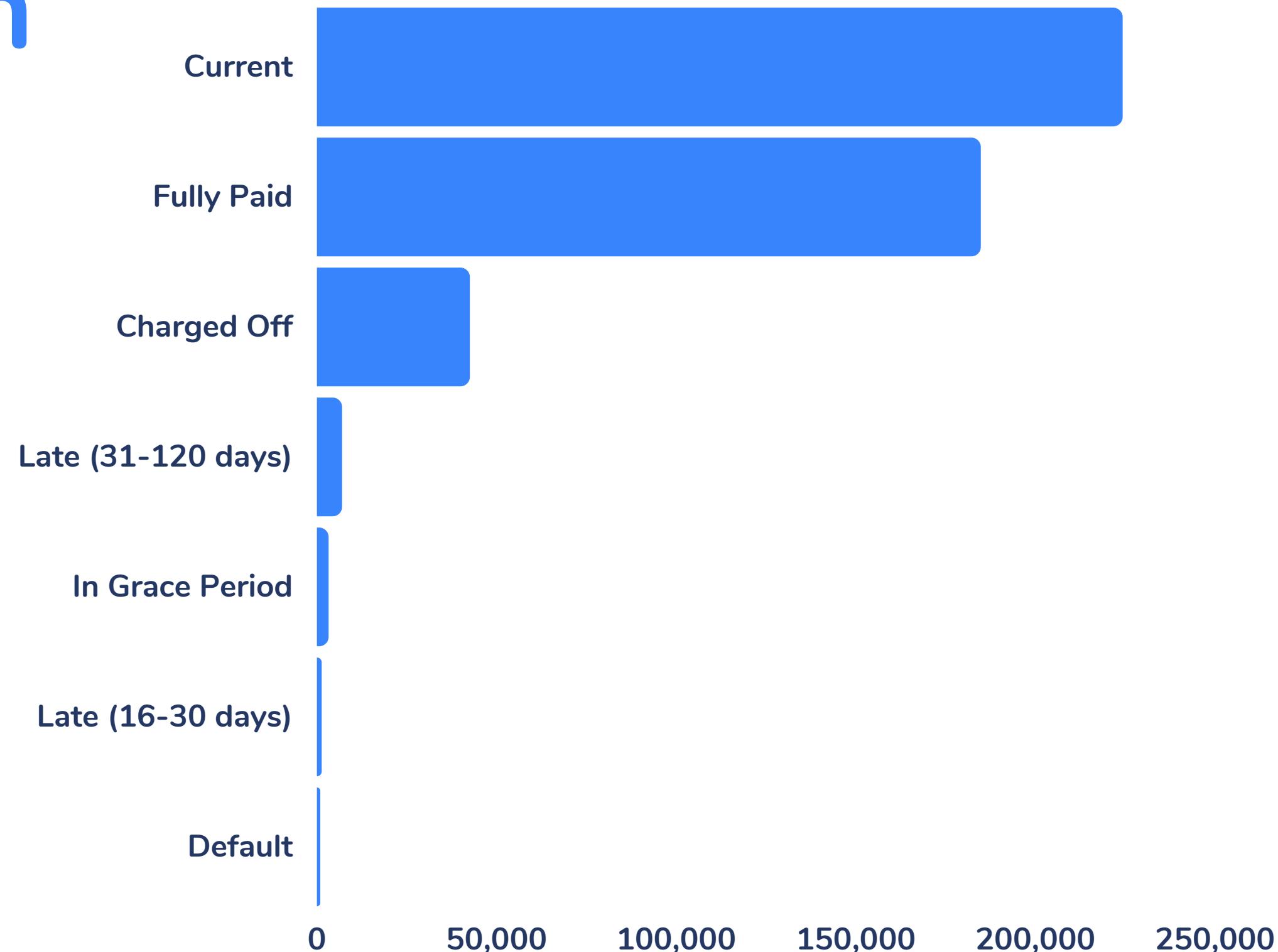
# The loan's problem

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

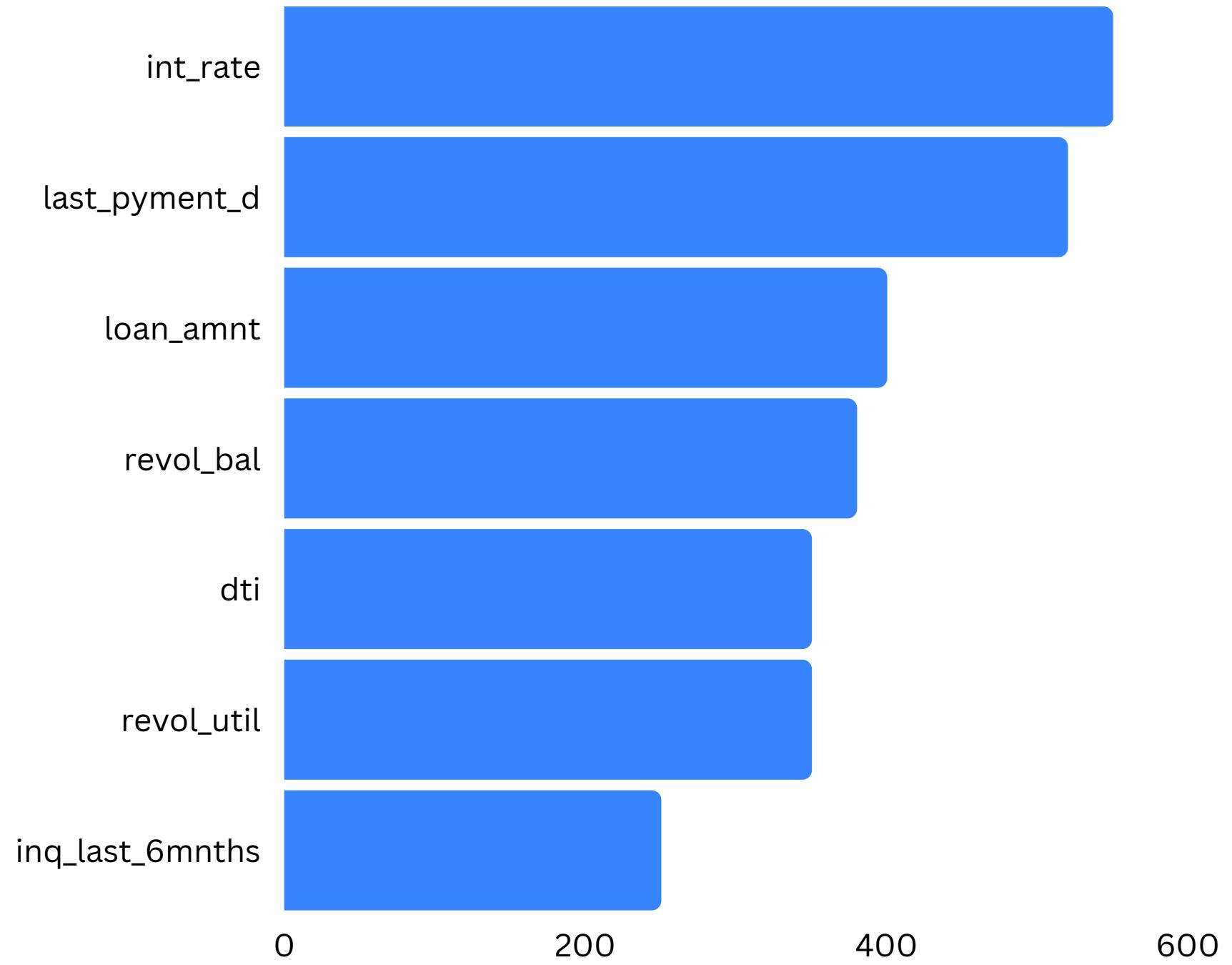
# Modelling & Evaluation

Model	Accuracy	Recall	AUC
LightGBM	0.889	0.992	0.859
XgBoost	0.888	0.993	0.855
Random Forest	0.887	0.986	0.847
Gradient Boosting	0.888	0.992	0.855
K-Nearest Neighbor	0.873	0.983	0.551

- 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\_payment\_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

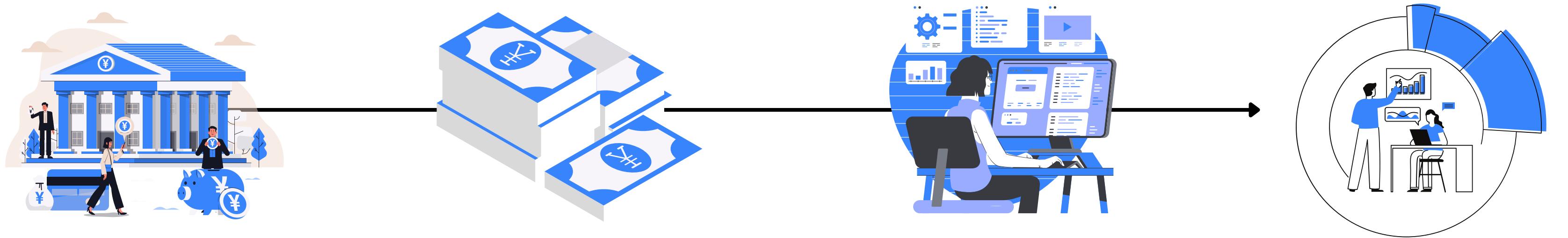
## **revol\_util**

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)

# Simulation With Machine Learning



**Bank Partners**

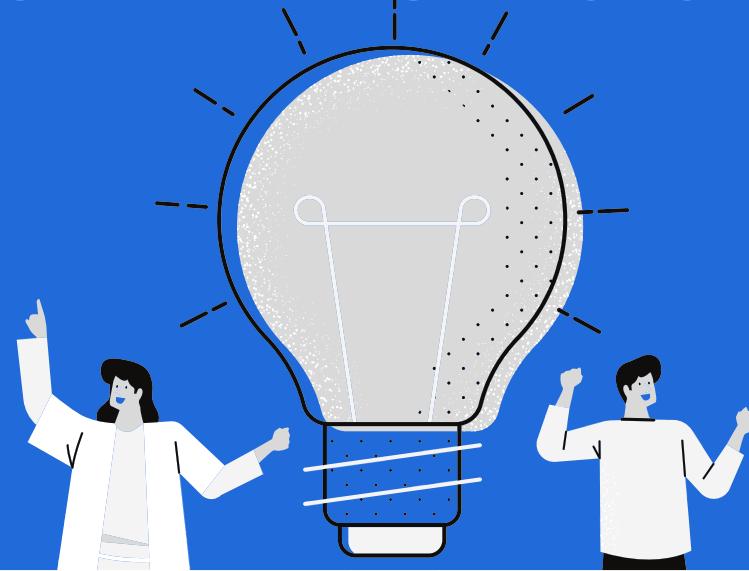
**Credit Loan**

**Automate Using  
Machine Learning**

**Result:**

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

# Business Recommendation



## Loan Amount

It is important to carefully evaluate the borrower's creditworthiness and the potential risks associated with the loan amount to ensure that the loan is viable and sustainable for both the lender and the borrower. This may include conducting a thorough credit risk assessment, implementing a risk-based pricing strategy.



## Feature Importance

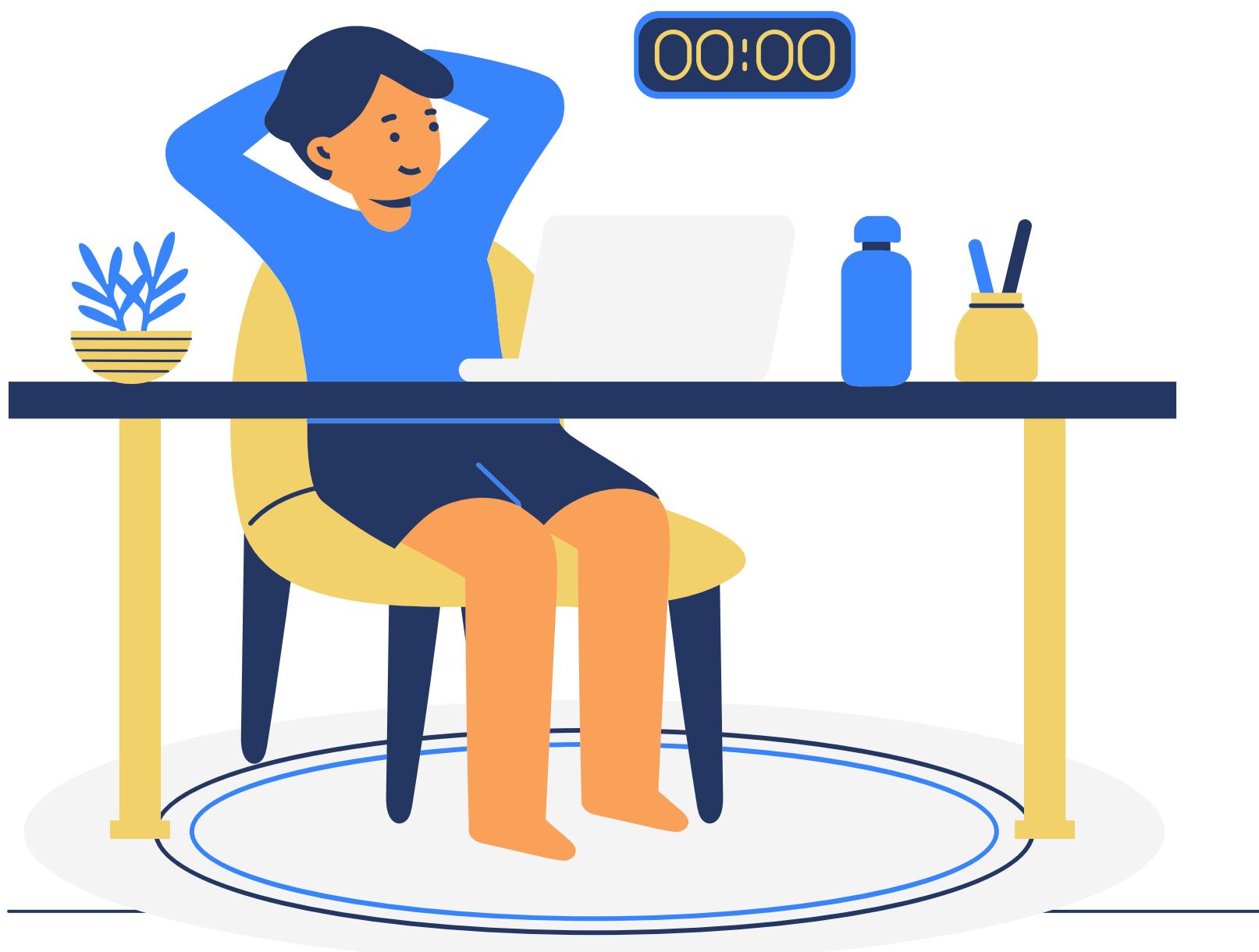
Focusing on feature importance could be a valuable business opportunity. By identifying the most important features and building a focused and effective model, this could help to reduce credit risk and make more informed lending decisions.



## Interest Rate

It would be beneficial to offer flexible interest rates that can be adjusted based on the borrower's creditworthiness and payment history. This allows for a more personalized approach to lending and can help the lender manage their credit risk more effectively.

# Thank you



GitHub Repositories

Credential: 28782IAPMGII30102022



## Virtual Internship

### Competencies Certification

has been presented to

# Muhammad Oky Hariawan

for able to finish all data scientist Virtual Internship Task from **ID/X Partners**

Completed Tasks: Beginner Level (2), Intermediate Level (2), Advance Level (1)

Softskill   Knowledge   Programming Language Scripting   Big Data   Machine Learning  
Statistics and Data Analytics   Data Warehouse   Programming Language Scripting

CEO Rakamin Academy

A handwritten signature in black ink.

Andika Deni Prasetya

VP Risk & Decision,  
ID/X Partners

A handwritten signature in black ink.

Iwan Setiawan