

FINAL PROJECT RAKAMIN BATCH 29



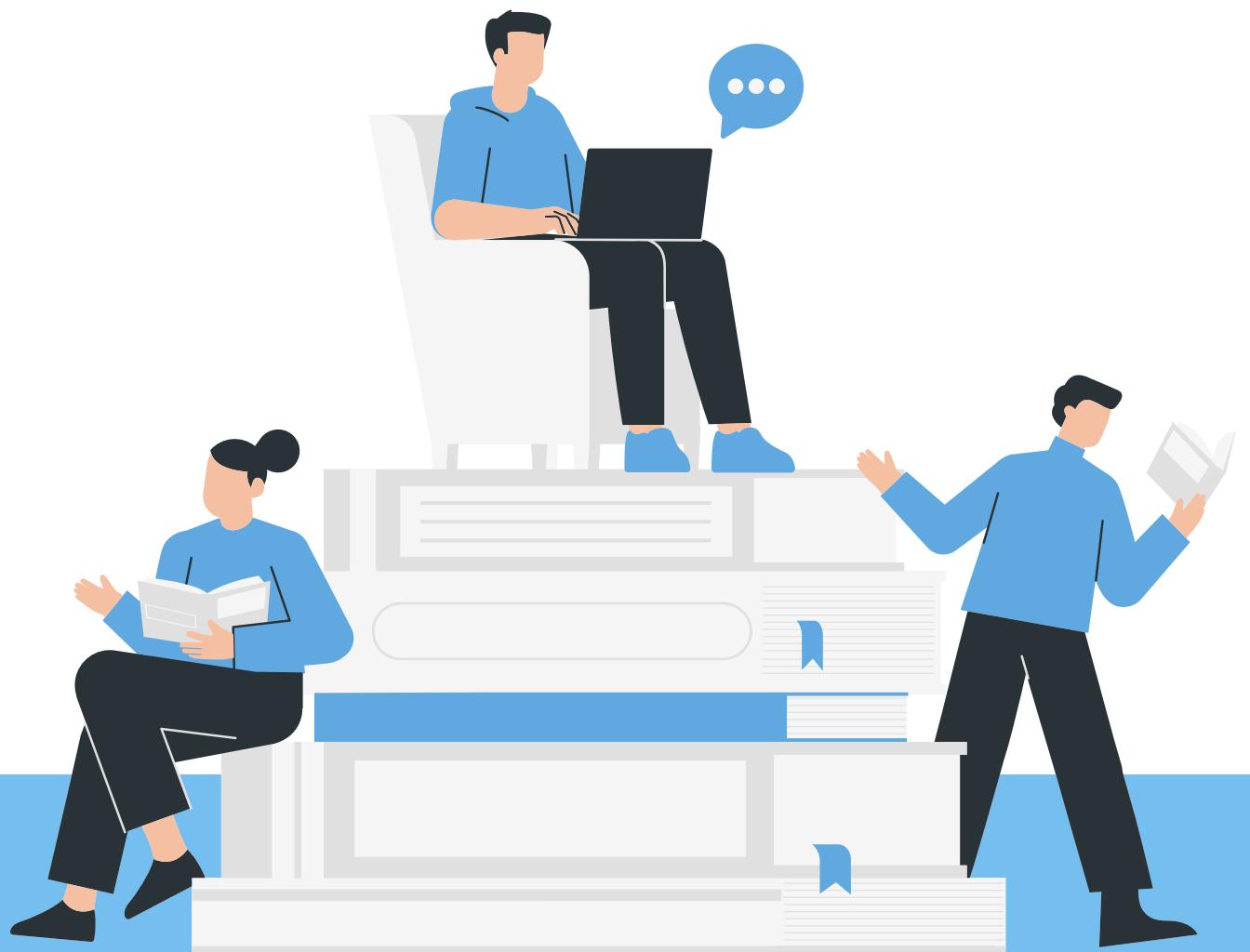
Loan Prediction

By: Pinjol Team

Today's Agenda

1. Who are the people behind the project?
2. What is the problem?
3. How to do exploratory data analysis?
4. What kind of machine learning model to use?
5. What is the solution?





Meet Pinjol Team

**M. Idhandi
Kurnia**
Project Leader

**Dea
Dahlila**
Data Analyst

**Muhammad
Rafly S.**
Data Analyst

**Ficky
Milando**
Data Science

**Laksaman
Satio**
Data Science

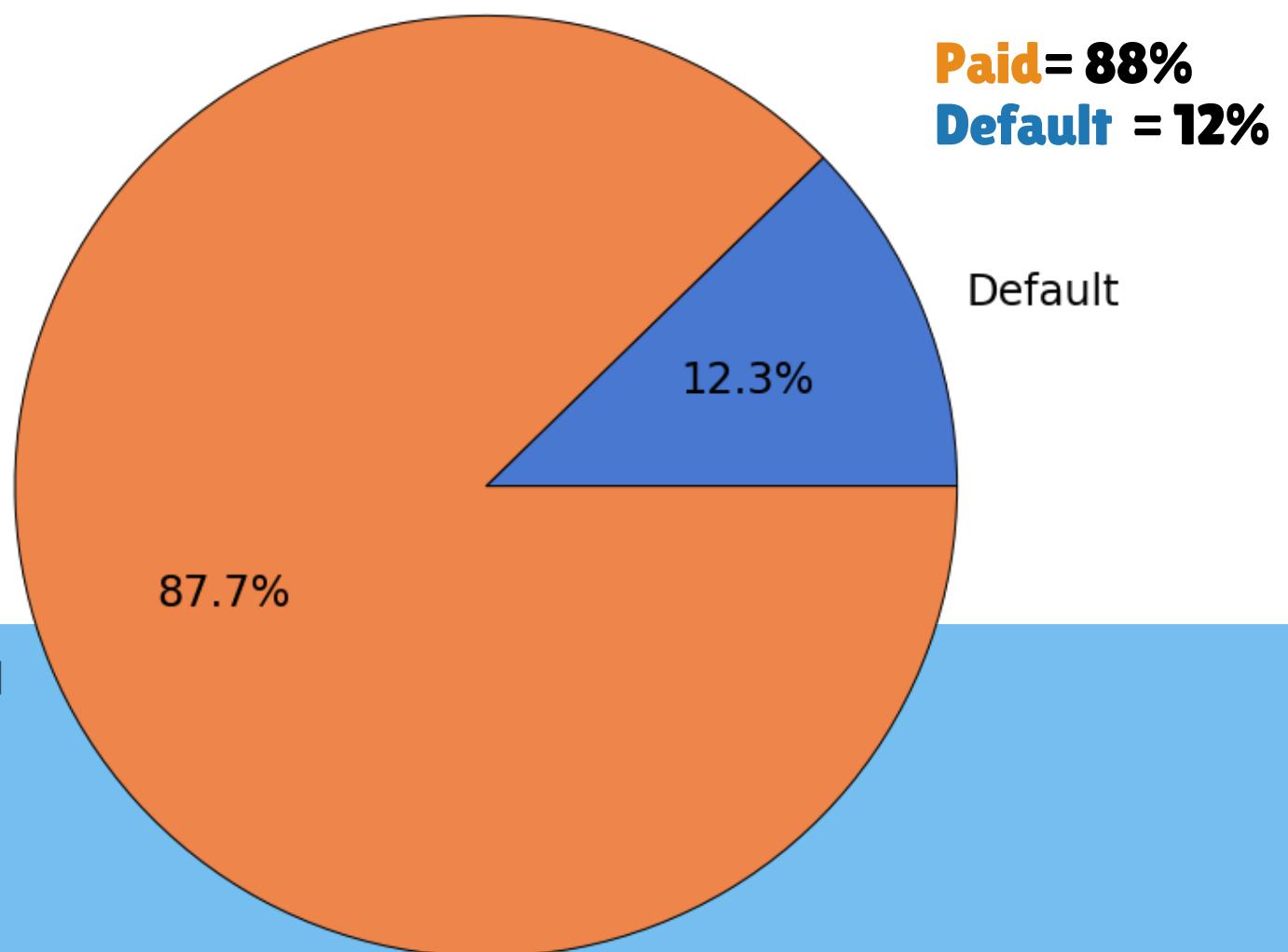
**Drestanta
V. W.**
Data Science

What is the problem?

NoWorries* is a startup company P2P landing who wants to predict **who possible defaulters** are for the consumer loans product. Hence when they acquire new customers they want to predict who is **riskier** and who is **safe**.

NoWorries* has a total of 252,000 customers and 12% of total are Non Performing Loans. Non Performing Loan (NPL) is a loan condition with the debtor failing to make payments scheduled for a certain period of time.

Paid VS Defaulted



PROBLEM STATEMENT

Background

Loss of potential profit caused by customers who cannot repay loans.

Goals

The percentage of defaulted rate decreased

Objectives

Create machine learning models to determine and forecast customers who have the potential to default on loans.

Business Metrics

Default rate



How to do Exploratory Data Analysis?



About Dataset

Column	Description	Type
income	Income of the user	int
age	Age of the user	int
experience	Professional experience of the user in years	int
profession	Profession	string
married	Whether married or single	string
house_ownership	Owned or rented or neither	string
car_ownership	Does the person own a car	string
risk_flag	Defaulted on a loan	string
current_job_years	Years of experience in the current job	int
current_house_years	Number of years in the current residence	int
city	City of residence	string
state	State of residence	string

Description:

- Target Label : Risk Flag
- Feature : 11
- Categorical features:
 - Married/Single
 - House Ownership
 - Car Ownership
 - Profession
 - City
 - State
- Numerical features:
 - Income
 - Age
 - Experience
 - Current Job Years
 - Current House Years

Data Cleansing



Missing Values

There is **no missing values**



Duplicated Data

There is **no duplicated data**



Outliers

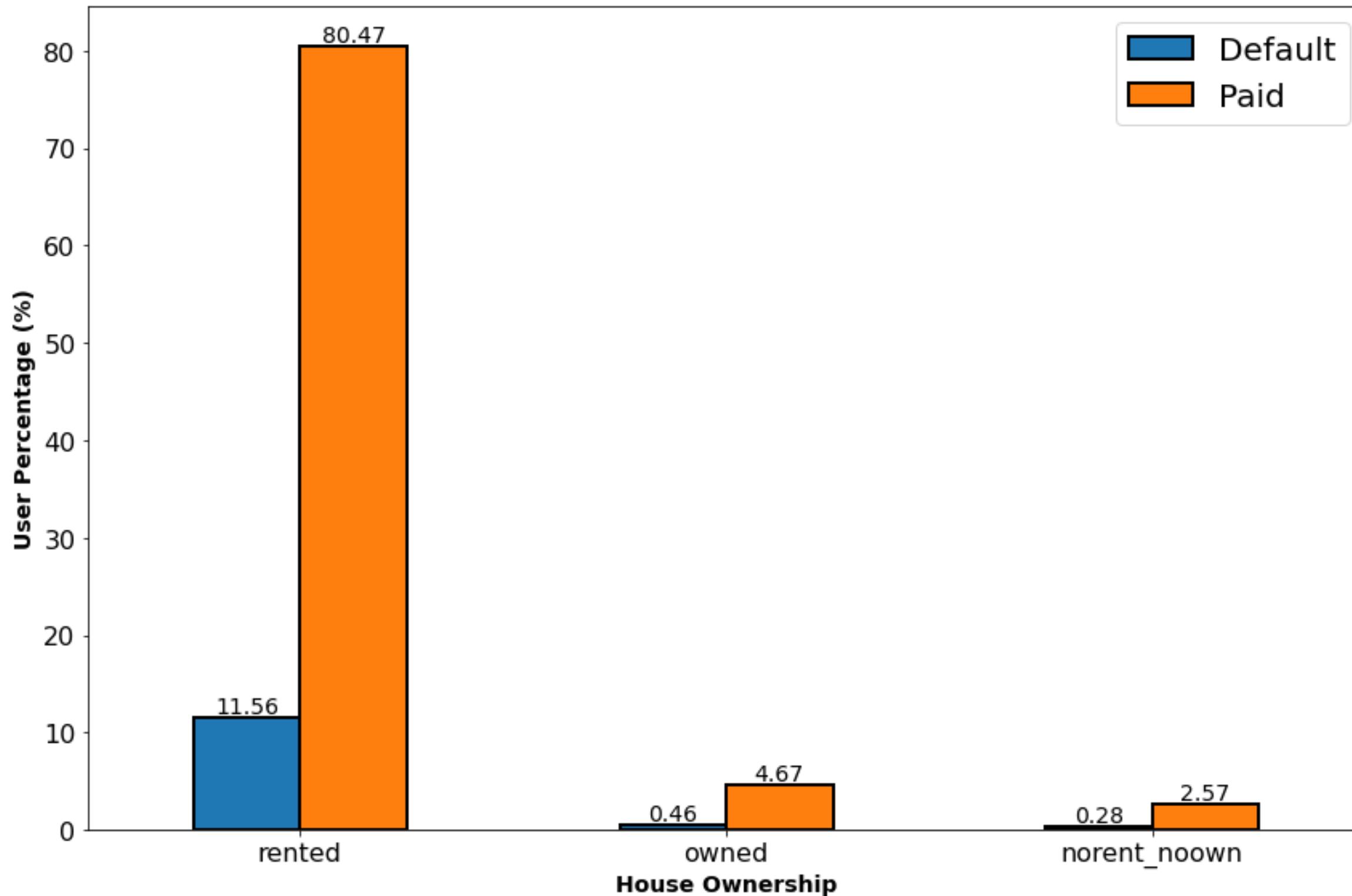
There is **no outliers**



Edit Columns

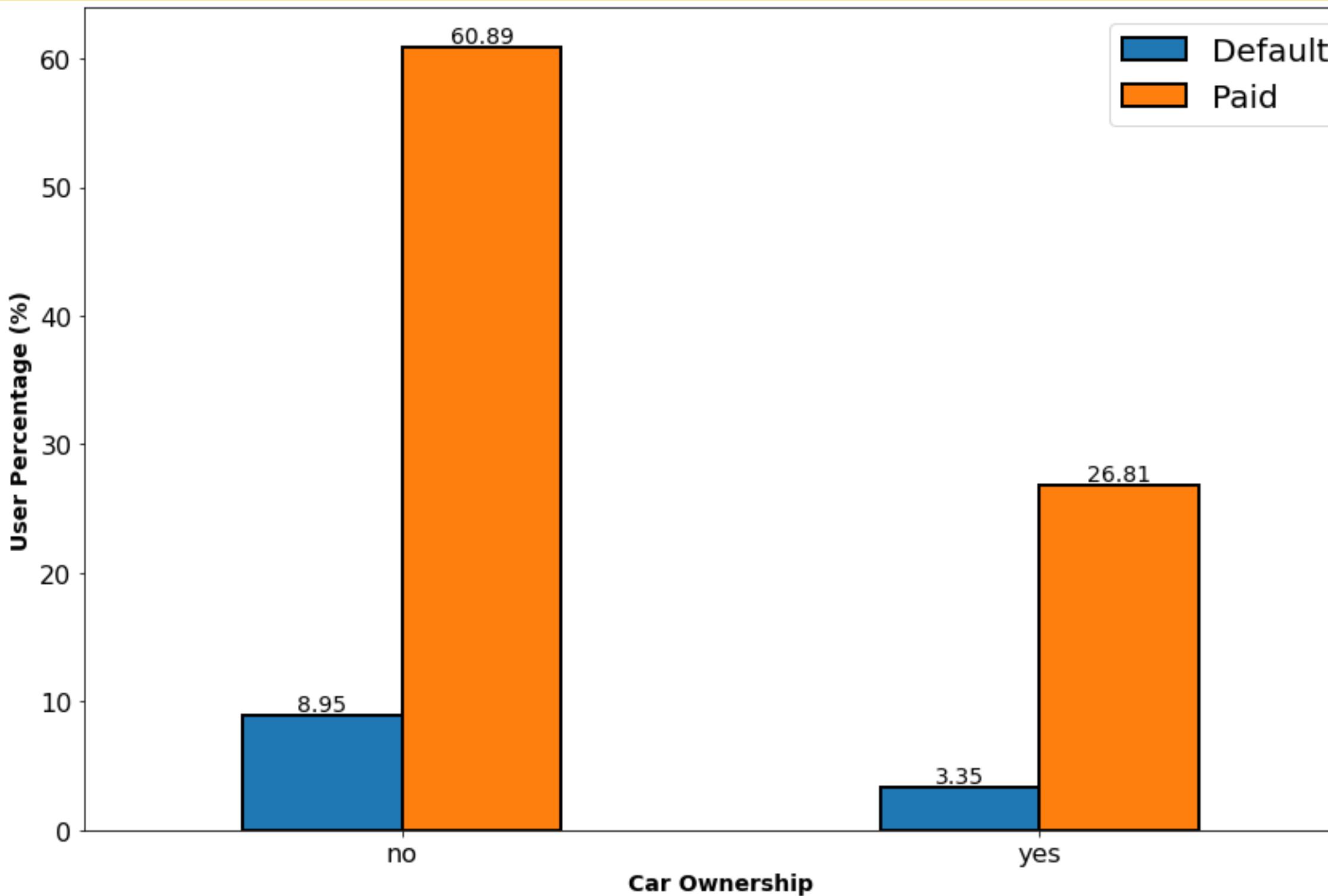
remove symbols and
arrange letters in certain
columns.

Visualization of House Ownership



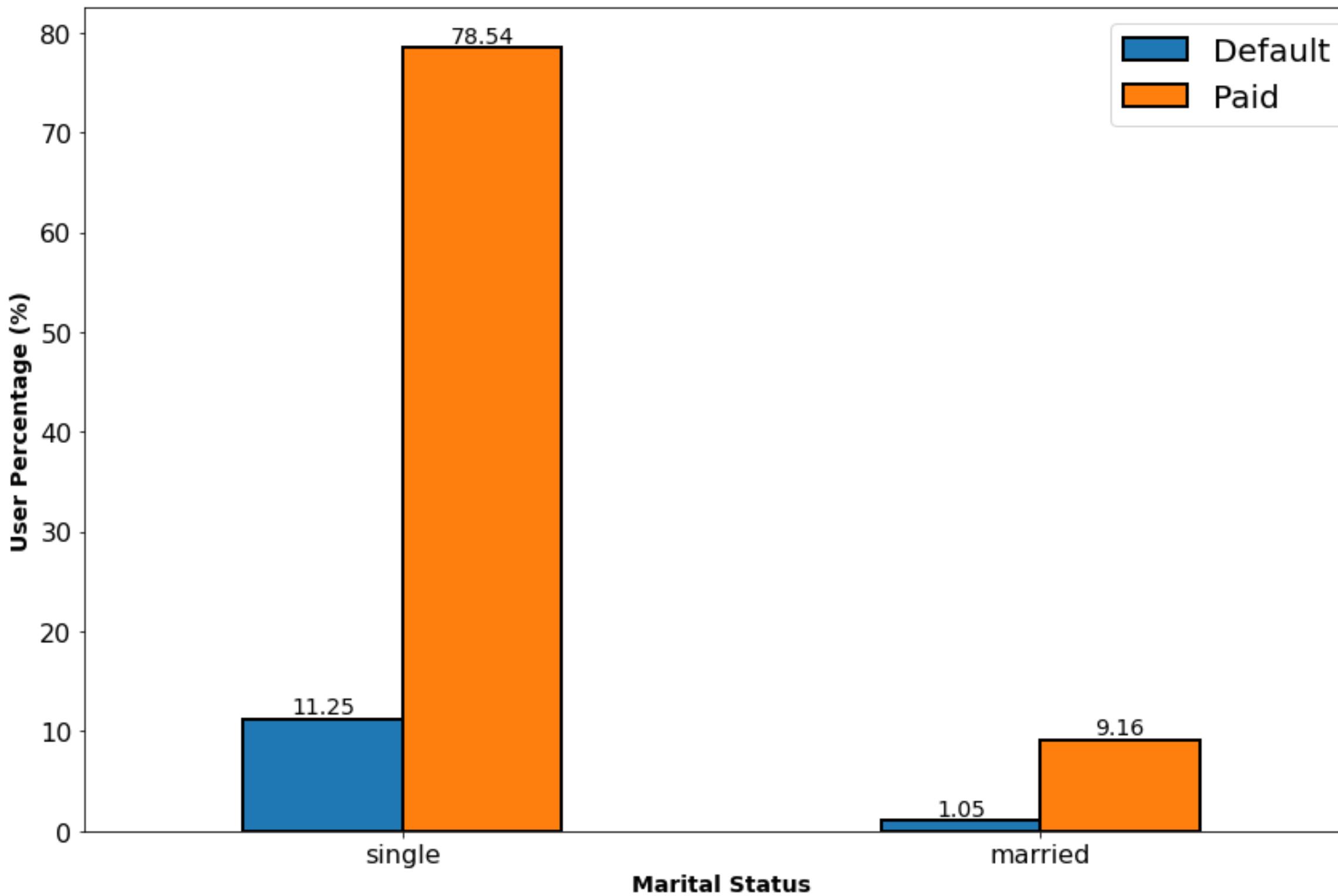
From the results of the visualization, it was found that users with rental residence status tend to make the most loans and have the highest risk flags.

Visualization of Car Ownership



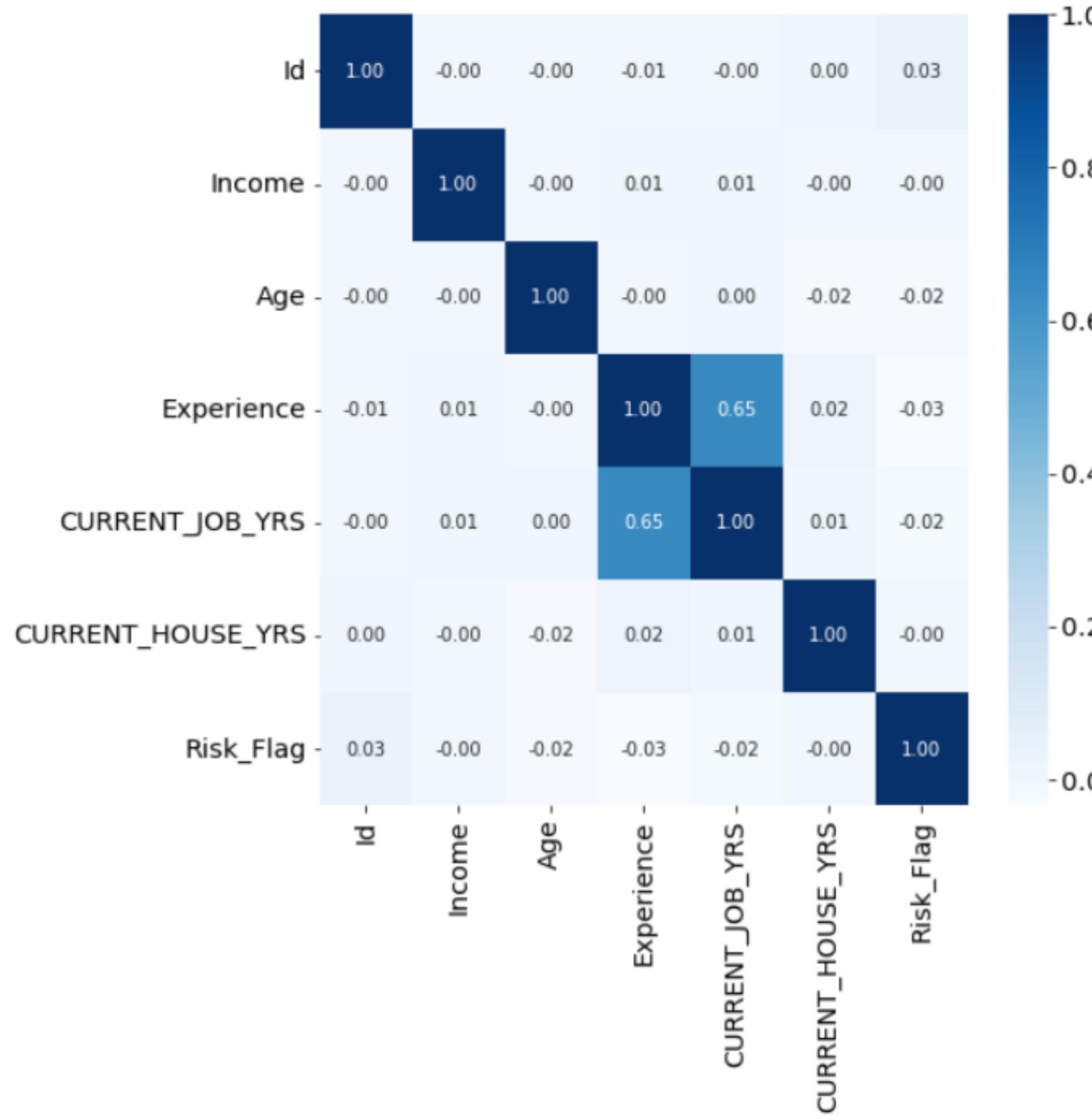
From the results of the visualization, it was found that users with no car ownership tend to make the most loans and have the highest risk flags.

Visualization of Marital Status



From the results of the visualization, it was found that users with single status tend to make the most loans and have the highest risk flags.

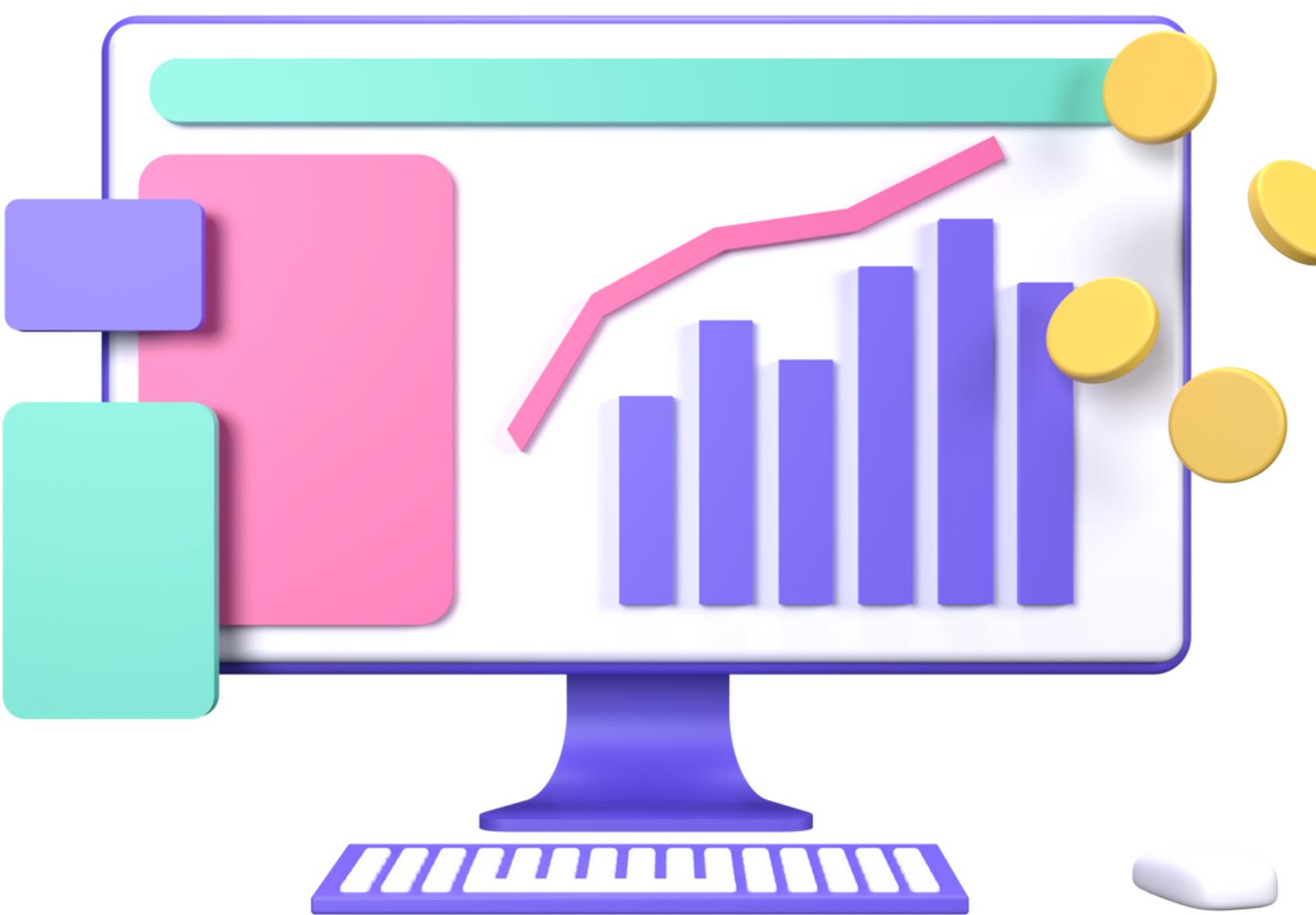
Multivariate Analysis



Explanation:

Feature and Target Label is not linear.
Because the correlation value is very small (< 0.1)

Data Pre-Processing



FEATURE

ENCODING

Change categorical data to numerical data

FEATURE

SELECTION

Drop Feature Id, CITY, Risk_Flag, Profession, STATE

SPLIT

DATA

Data Train 80%
Data Test 20%

SCALING

FEATURE

Use StandarScaler

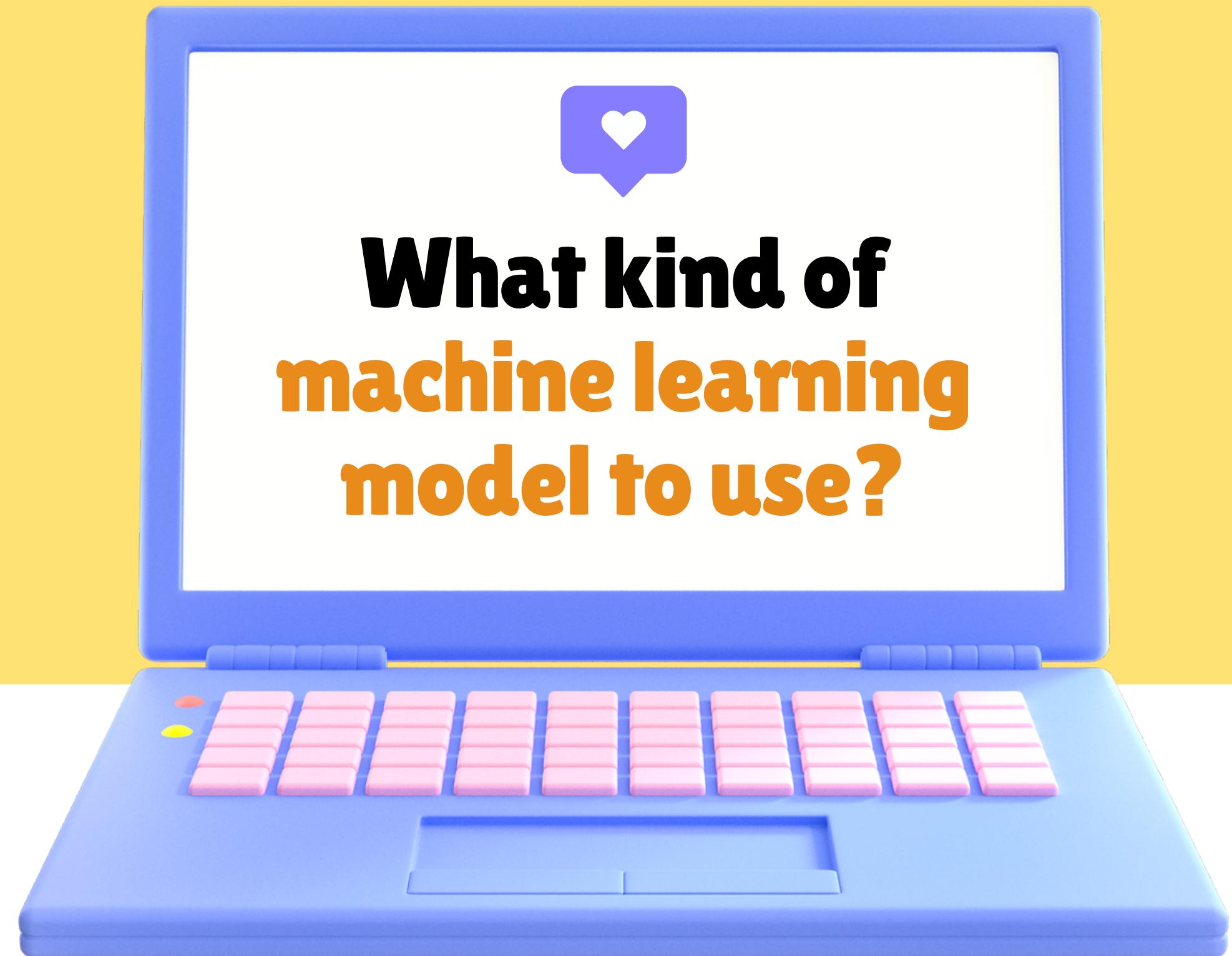
CLASS

IMBALANCE

Using `class_weight = 'balanced'`



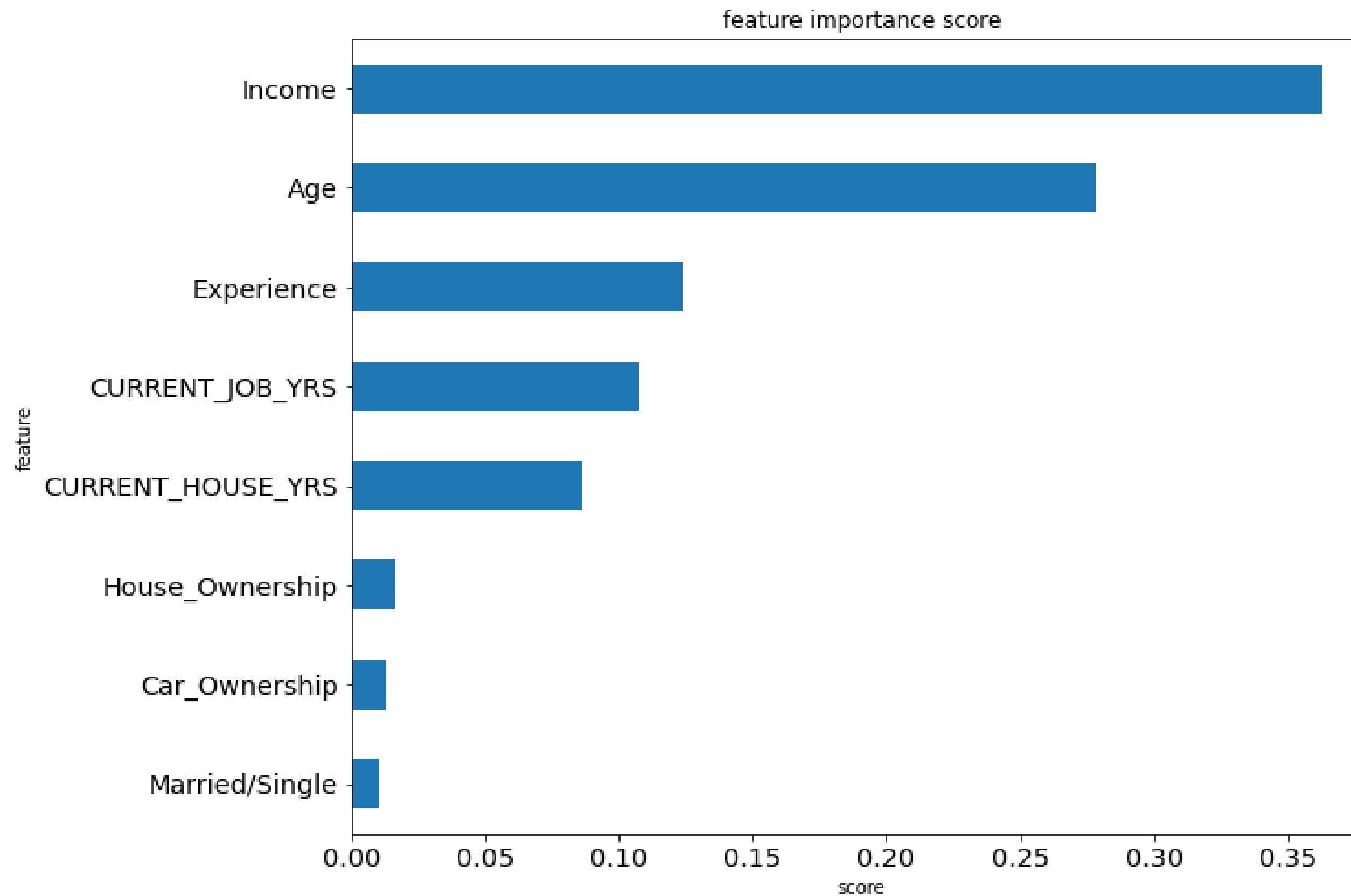
**What kind of
machine learning
model to use?**



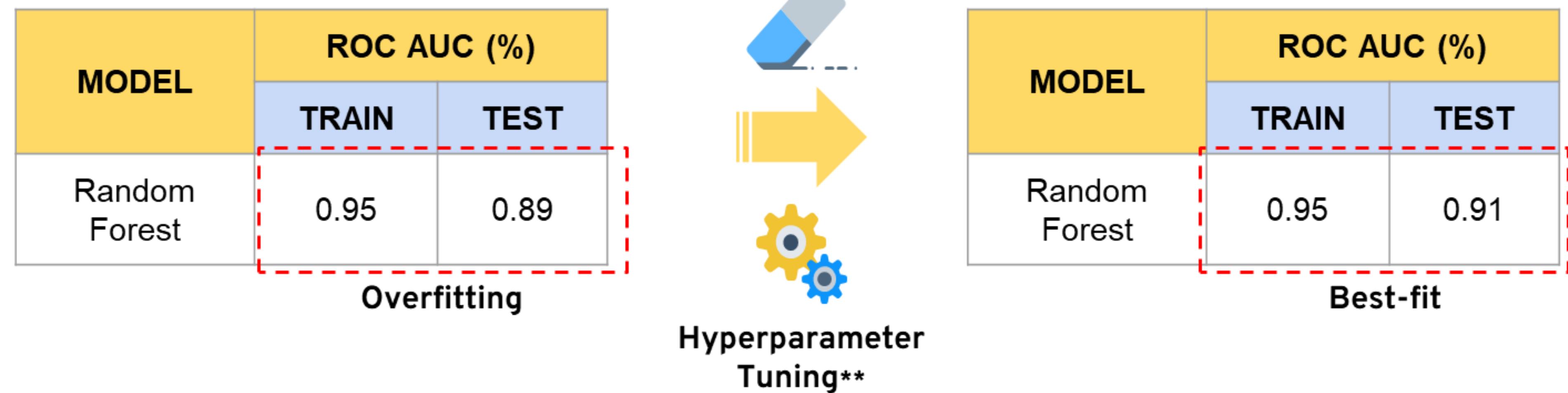
Modeling Machine Learning

Eval Metrics \ Model	Logistic Regression	Decision Tree	Random Forest	XGBoost	KNN
Accuracy	0.51	0.88	0.89	0.88	0.89
Precision	0.14	0.50	0.54	0.66	0.56
Recall	0.56	0.81	0.78	0.12	0.50
F1-Score	0.22	0.62	0.64	0.20	0.53
ROC-AUC (Test)	0.55	0.85	0.94	0.86	0.87
ROC-AUC (Train)	0.55	0.98	0.96	0.90	0.94

Feature Importance



Model Evaluation



What is the solution?



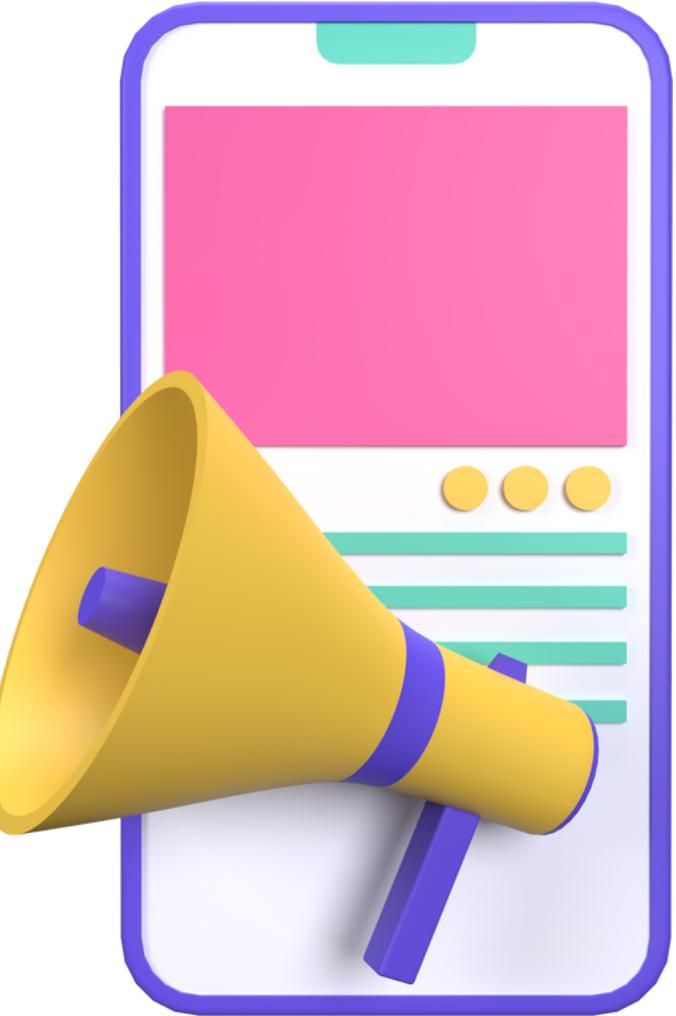
**Business Insight and
Recommendation**



Business Impact



Business Insight and Recommendation



Risk Flag Checking

Mengidentifikasi dari awal nasabah yang mempunyai risk flag 1 agar bisa dipertimbangkan lagi untuk mengambil pinjaman atau memberikan pinjaman khusus

Extra Limit

Memberikan extra limit kepada user yang membayar pinjaman sebelum jatuh tempo.

Business Insight and Recommendation



Partner with house or car dealership

Melakukan partnership dengan dealer mobil atau property, dikarenakan mayoritas user yang tidak memiliki mobil atau rumah

Automatic Reminder

Pengingat otomatis yang dikirimkan kepada nasabah melalui email atau SMS sebelum jatuh tempo

Business Impact

		Predicted Label	
		Tidak Gagal Bayar	Gagal Bayar
Actual Label	Tidak Gagal Bayar	40081	4120
	Gagal Bayar	1388	4811

Sample Test:
50.400

Predicted Label

Default Rate

Sample = 50.400

Tidak Gagal Bayar = 44.201

Gagal Bayar = 6.199

12.3 %



2.75 %

Business Impact

Rupee	Tanpa Machine Learning	Dengan Machine Learning
Revenue (Interest)	48.621.100	44.089.100
Cost (Loan Loss)	6.199.000	1.388.000
Profit	42.422.100	42.701.100



279.000 Rupee

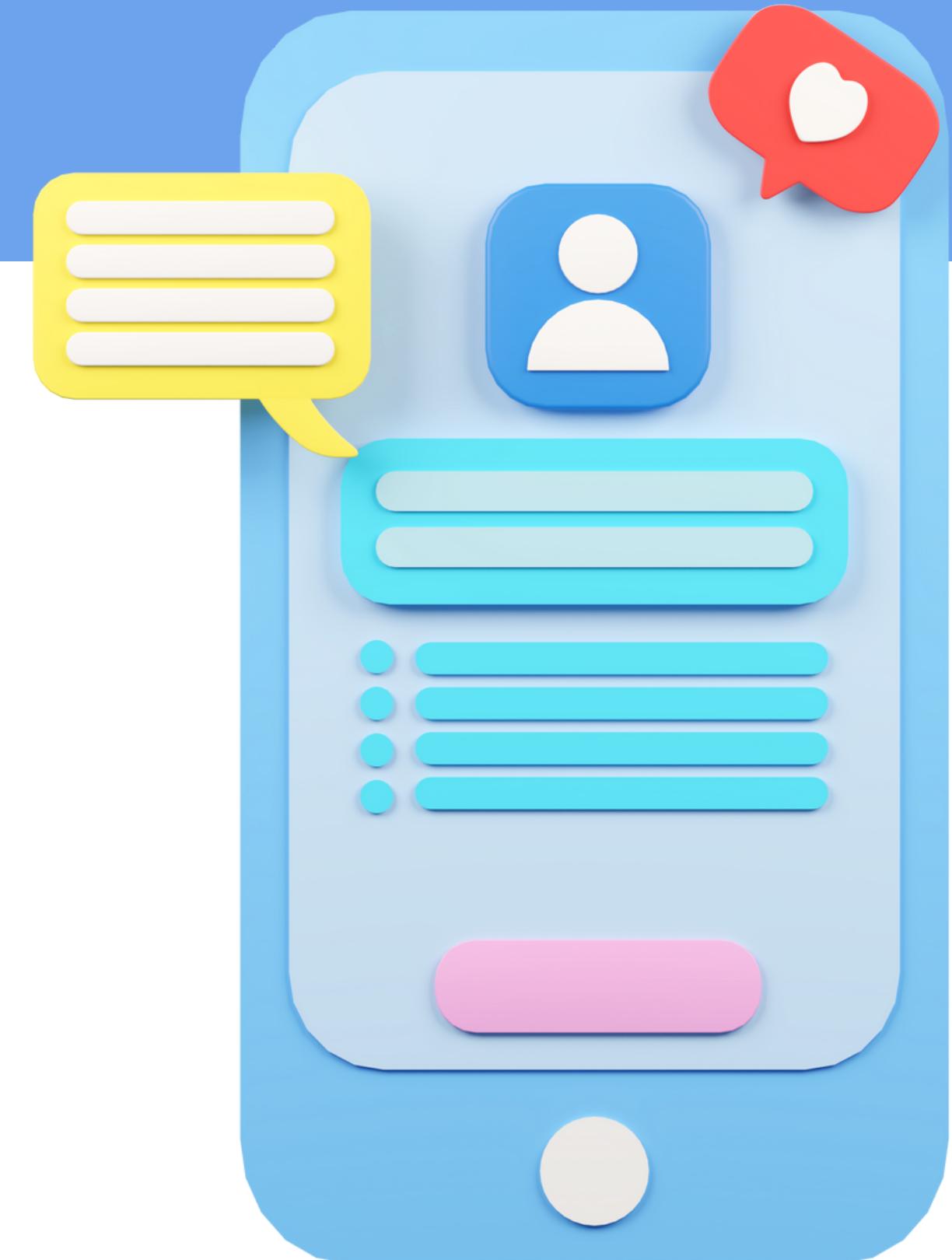
Skema Loan

Interest Rate = 10% p.a

Loan = 1000 rupee

Sample = 50.400 orang

FINAL PROJECT RAKAMIN BATCH 29



Thank You!
Any Question?