

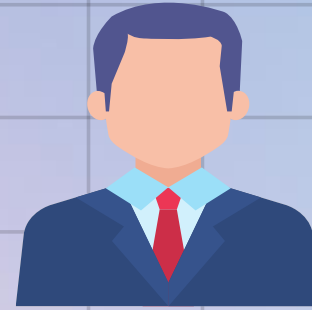


# FINAL PROJECT

## Payment Default Prediction

○ ○ ○ ○

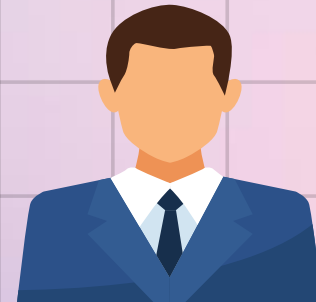
**KELOMPOK  
4-TECH**



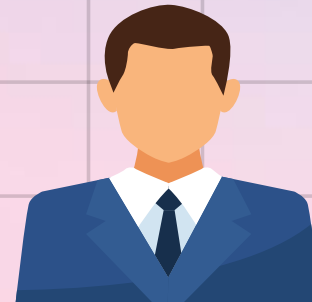
**Azriel Hutagalung**  
(Mentor)



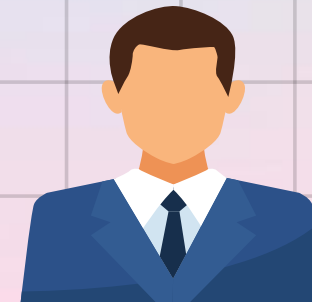
**Nuas Hados Caniago**  
(Student)



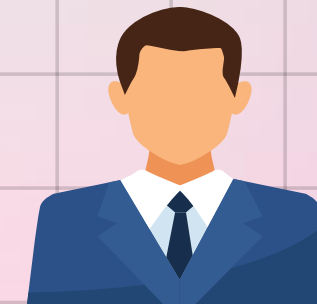
**Fadlan Setiadi**  
(Student)



**Abizar Al Gifari**  
(Student)



**Tanto Hari Wibowo**  
(Student)



**Ferdy Santoso**  
(Student)



**Ivana Joice Chandra**  
(Student)



**Aliyah Nafilah Fachrina**  
(Student)



**Berlian Bidari**  
(Student)



# Table of Content

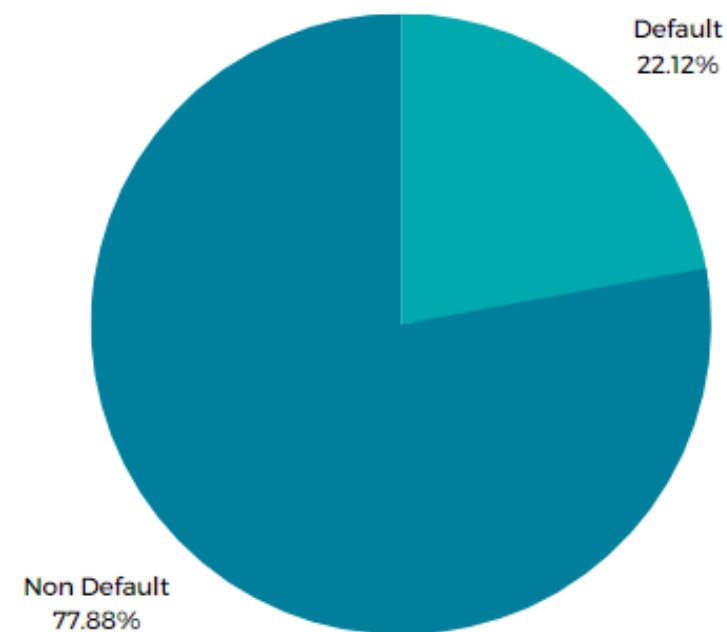
1. Business Understanding
2. Exploratory Data Analysis (EDA)
3. Pre-Processing
4. Modeling
5. Business Recommendation & Simulation

# Business Understanding

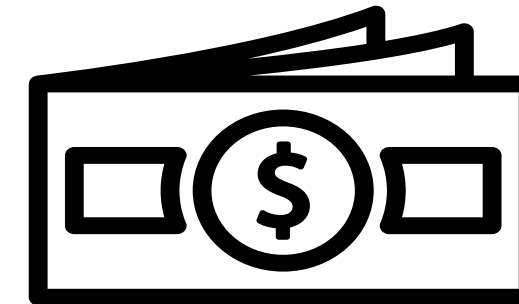


# Problem Statement

Masalah yang dihadapi bank Taiwan pada bulan September 2005:



Default rate (persentase nasabah yang gagal dalam membayar hutang kartu kredit) yang tergolong besar yaitu 22.12%



Potensi kerugian sebesar  
NTD 220.983.433

# ROLE



**4-TECH**

**Data Science Consultant**

## **Project Payment Default Prediction**

- ❑ Membuat model untuk memprediksi nasabah yang akan default pada bulan berikutnya.
- ❑ Memberikan business insight dan recommendation.

## **GOAL**

- Menurunkan kerugian Bank Taiwan
- Menurunkan default rate Bank Taiwan

## **OBJECTIVE**

- Membuat sistem untuk memprediksi nasabah yang mengalami default (gagal bayar) pada bulan berikutnya

## **BUSINESS METRICS**

Default rate → Persentase nasabah yang mengalami default atau gagal dalam membayar hutang kartu kredit

$$\text{default rate} = \frac{\text{total nasabah yang default}}{\text{total nasabah bank}}$$



# Exploratory Data Analysis (EDA)

○ ○ ○ ○

# Exploratory Data Analysis (EDA)

○ ○ ○ ○

21.000  
Rows

24  
Features

1  
Target  
Variable

0  
Missing  
Value

0  
Duplicated  
Value

○ ○ ○ ○

## 11 Categorical Columns

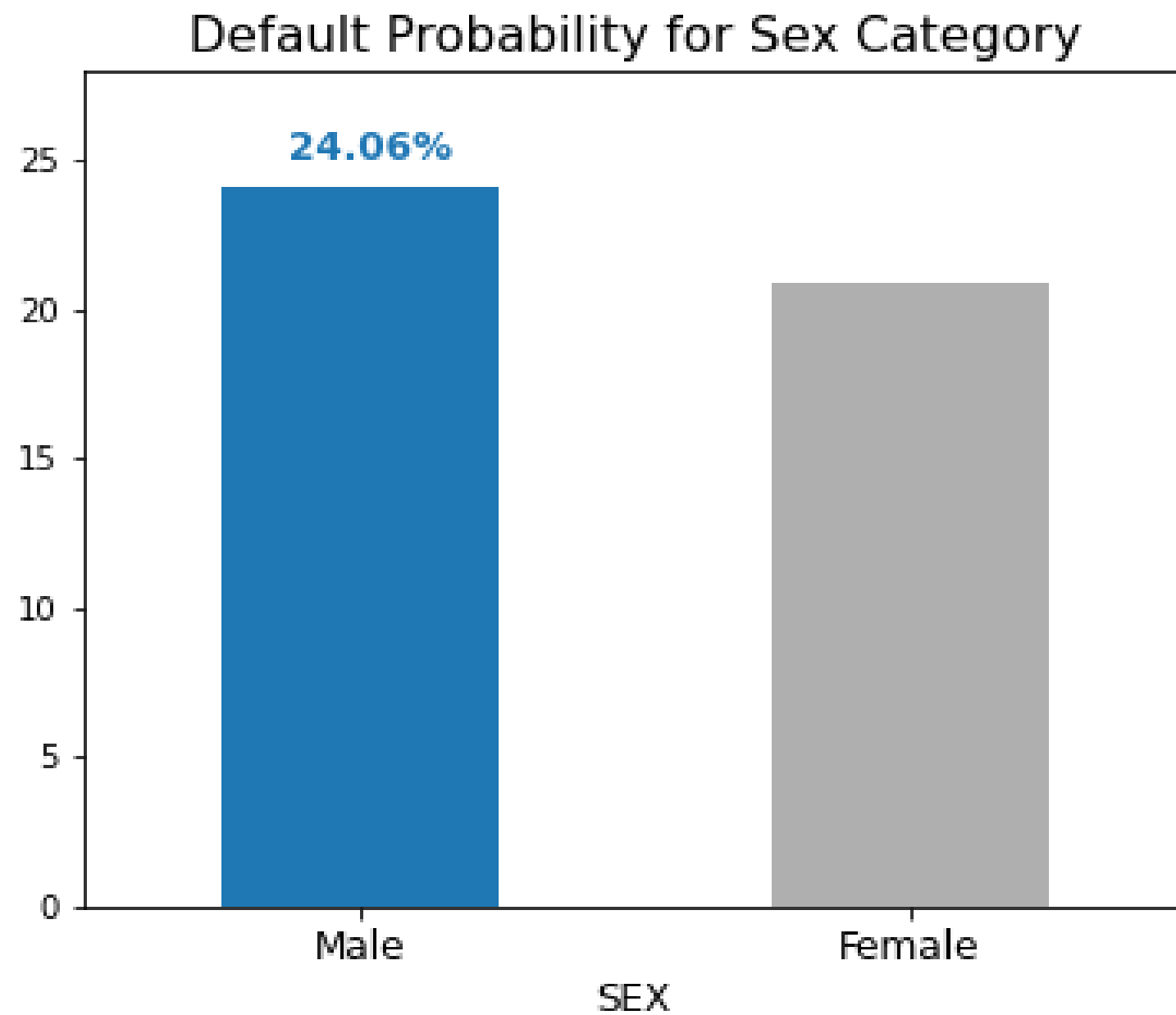
- ID
- SEX
- EDUCATION
- MARRIAGE
- PAY\_0
- PAY\_2
- PAY\_3
- PAY\_4
- PAY\_5
- PAY\_6
- **Default\_Payment\_Next\_Month**

## 14 Numerical Columns

- LIMIT\_BAL
- AGE
- BILL\_AMT1
- BILL\_AMT2
- BILL\_AMT3
- BILL\_AMT4
- BILL\_AMT5
- BILL\_AMT6
- PAY\_AMT1
- PAY\_AMT2
- PAY\_AMT3
- PAY\_AMT4
- PAY\_AMT5
- PAY\_AMT6

TIPE DATA	NAMA KOLOM	DESKRIPSI
KATEGORIKAL	SEX	Jenis kelamin nasabah
	EDUCATION	Pendidikan terakhir nasabah
	MARRIAGE	Status pernikahan nasabah
	PAY_0 - PAY_6	Repayment status nasabah pada bulan September 2005 - April 2005
	ID	ID nasabah
NUMERIKAL	LIMIT_BAL	Limit kartu kredit nasabah (dalam NTD)
	AGE	Umur nasabah
	BILL_AMT1 - BILL_AMT6	Jumlah tagihan kartu kredit nasabah pada bulan September 2005 - April 2005
	PAY_AMT1 - PAY_AMT6	Jumlah pembayaran sebelumnya pada bulan September 2005 - April 2005

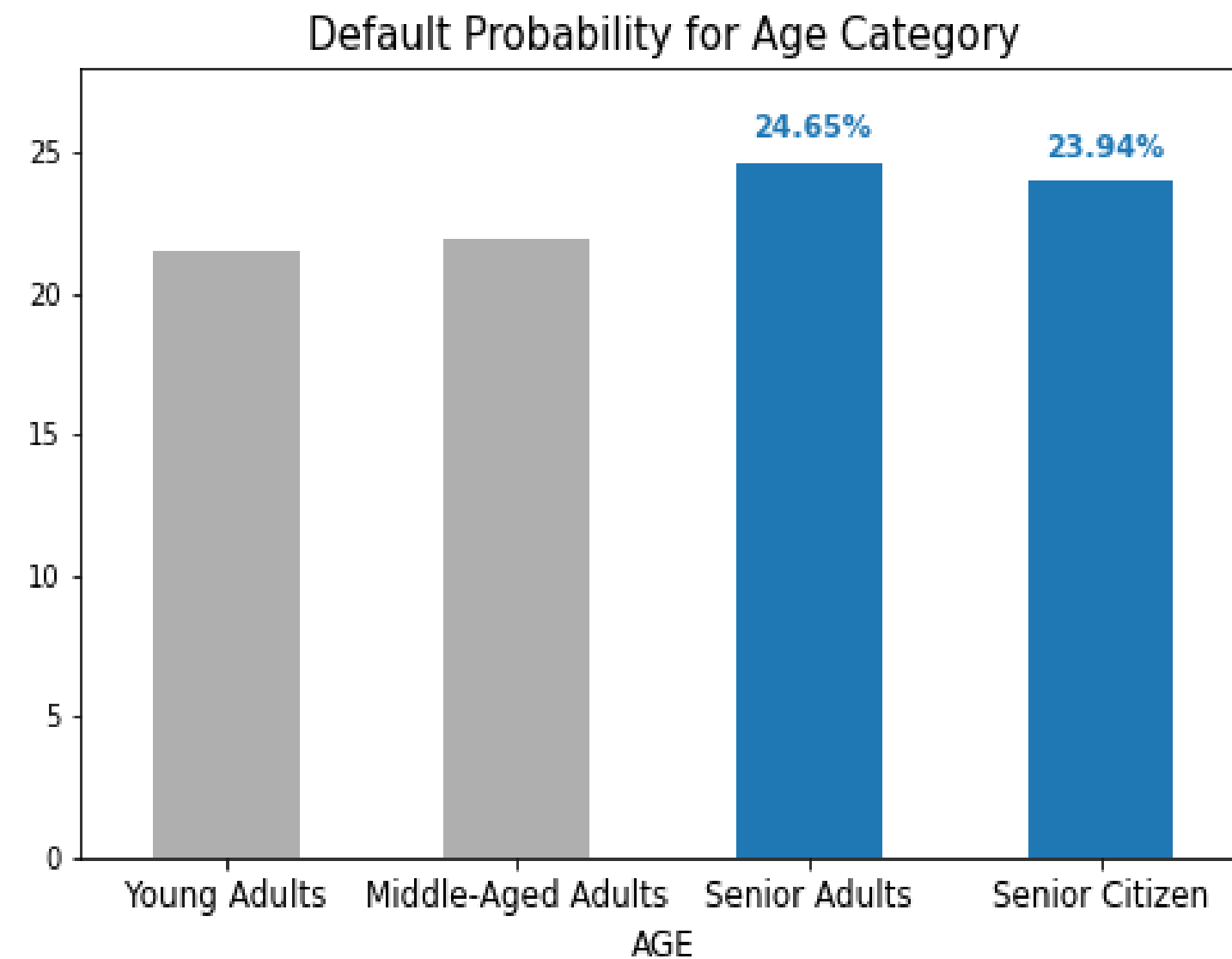
## Potensi Nasabah untuk Default



**Hasil pengamatan:**

**Nasabah laki-laki** lebih cenderung mengalami default

## Potensi Nasabah untuk Default



Young Adult: 20-35 tahun, Middle-Aged Adults: 36-45 tahun,  
Senior Adults: 46- 60 tahun, Senior Citizen: > 60 tahun

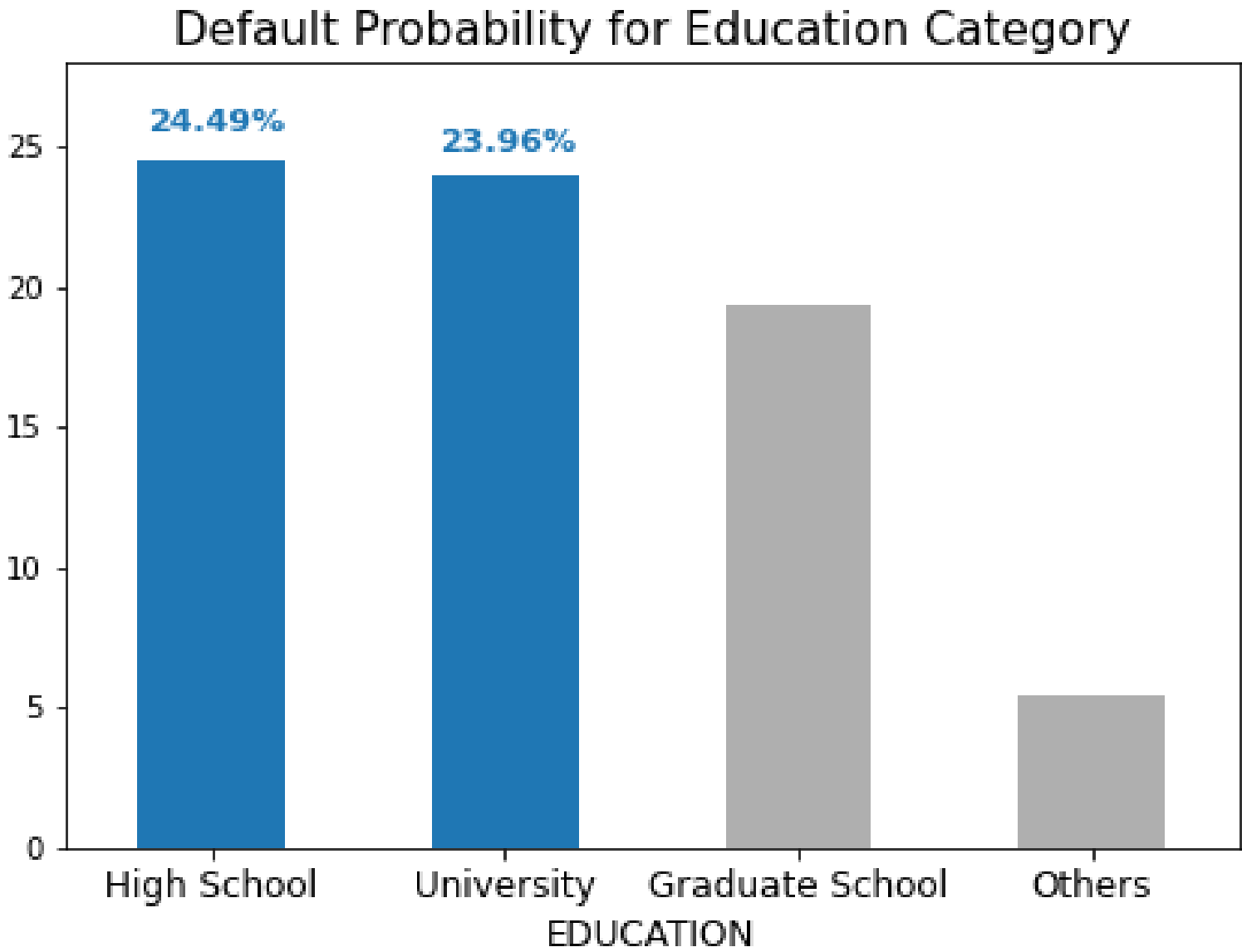
### Hasil pengamatan:

Nasabah yang tergolong **senior adults** dan **senior citizen** memiliki kemungkinan paling besar untuk default

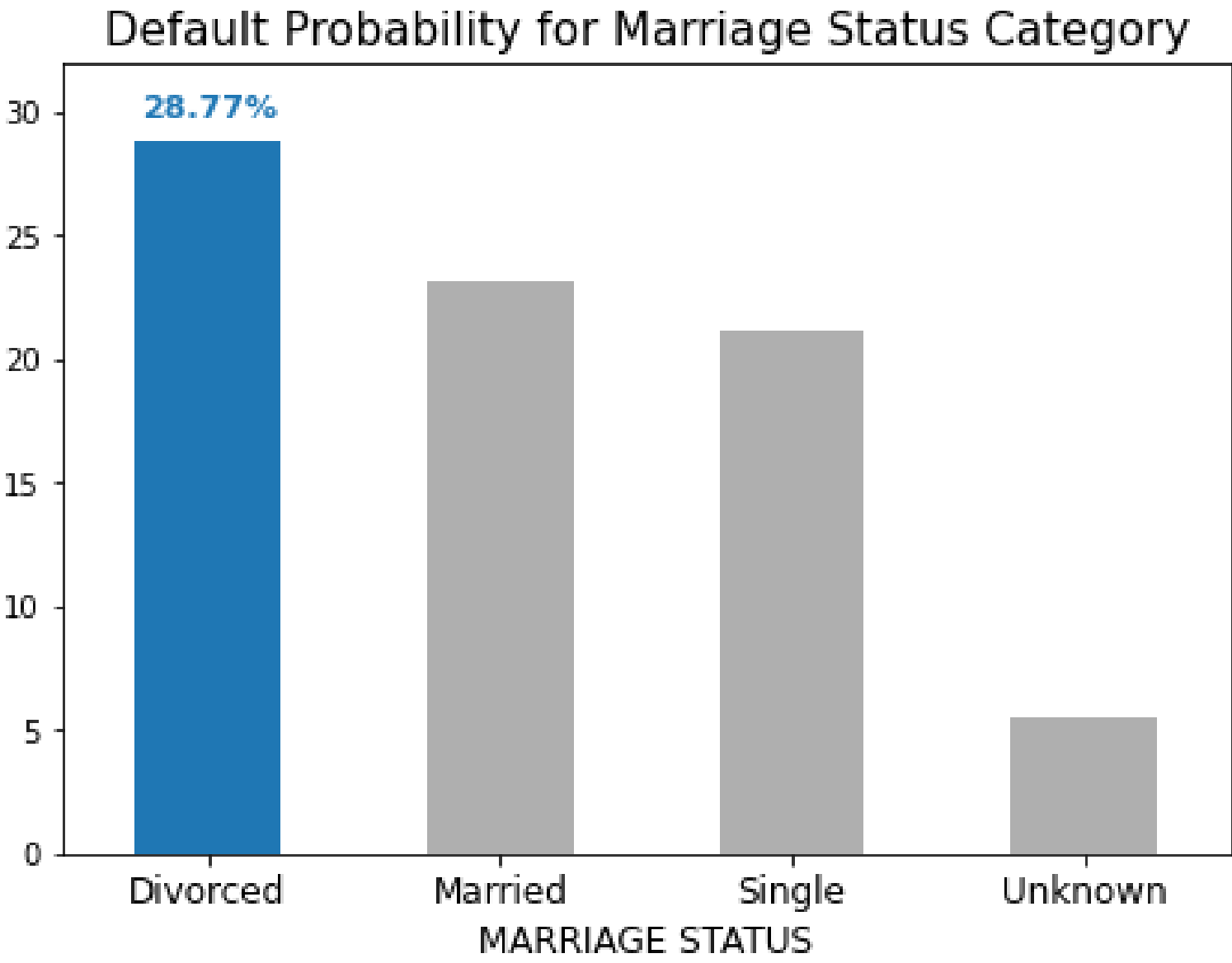
### Treatment yang dapat dilakukan:

Mengirimkan reminder untuk membayar tagihan melalui SMS dan telepon

# Potensi Nasabah untuk Default

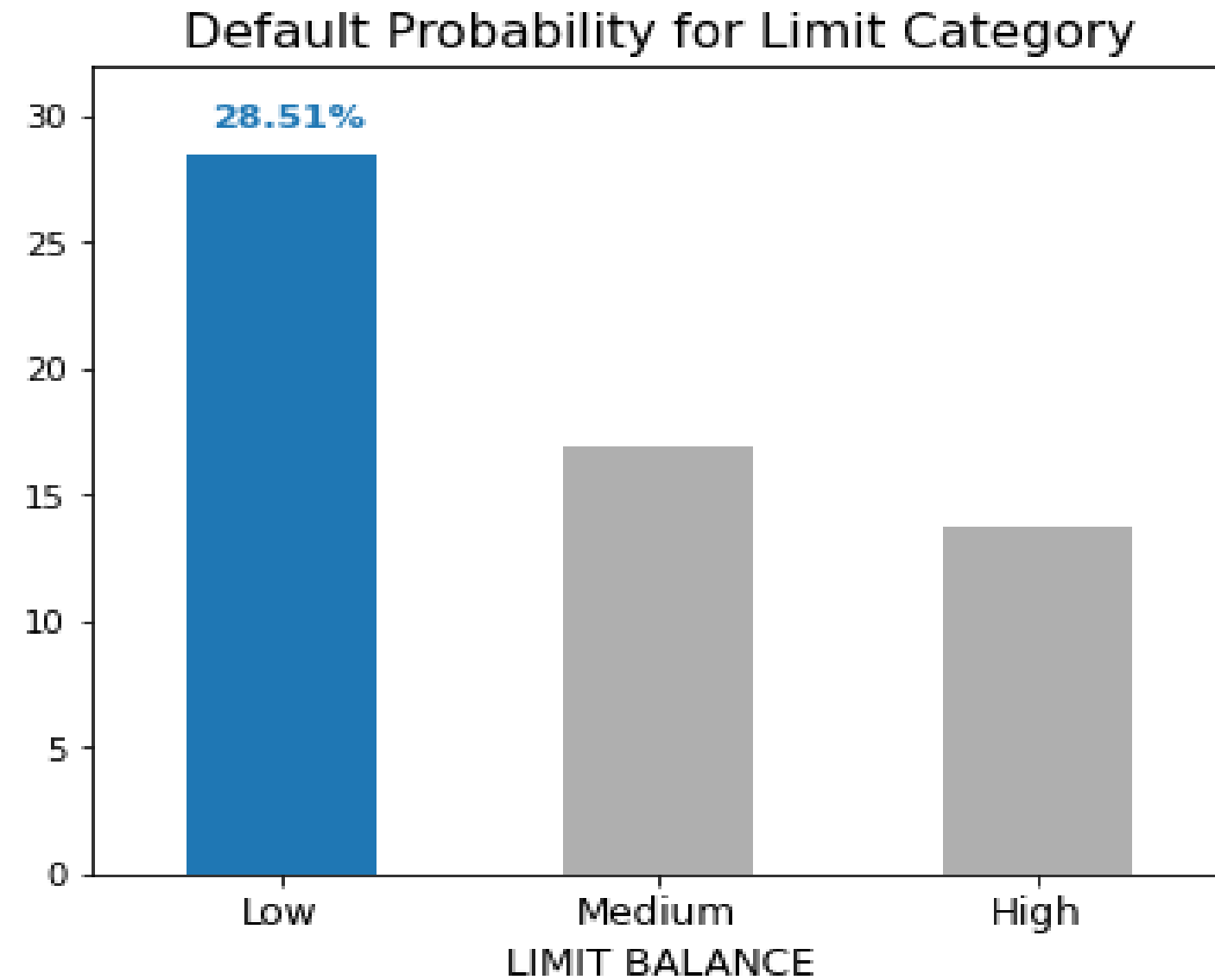


Nasabah dengan lulusan **SMA** dan **Universitas**  
lebih cenderung default



Nasabah yang **statusnya sudah pernah bercerai**  
banyak yang mengalami default

## Potensi Nasabah untuk Default



Low :  $\leq$  NTD 140.000, Medium: NTD 140.001 – NTD 240.000,  
High :  $>$  NTD 240.000

### Hasil pengamatan:

Nasabah dengan peluang default tertinggi adalah nasabah dengan **kategori limit low**

# Pre-Processing







Rename

Mengubah fitur PAY\_0 menjadi PAY\_1

Handling Outliers

Metode Z-score

Feature Encoding

- Label Encoding → Kolom SEX, EDUCATION
- One Hot Encoding → Kolom MARRIAGE, PAY\_1 - PAY\_6

Feature Transformation

Standardization

Feature Selection

Membuang kolom ID

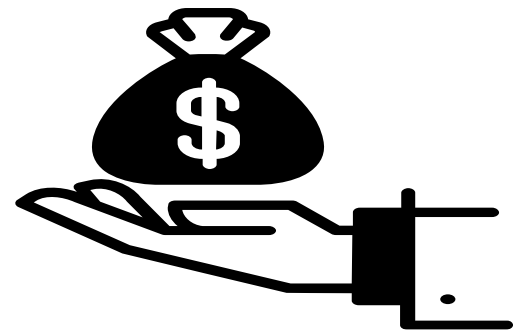
Feature Extraction

Tidak ada

Splitting Data

Train : Test = 70 : 30

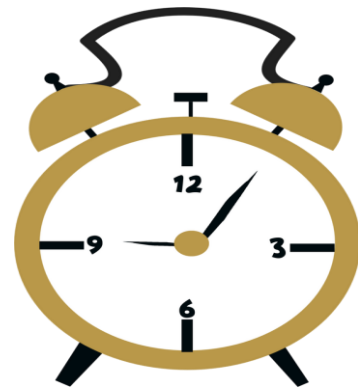
## Rekomendasi Feature Tambahan



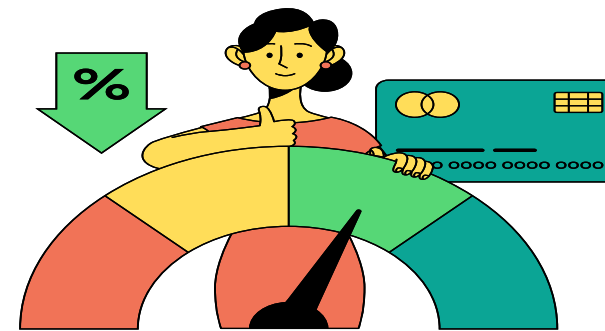
**Income**



**Job**



**Masa Kerja**



**Credit Score**

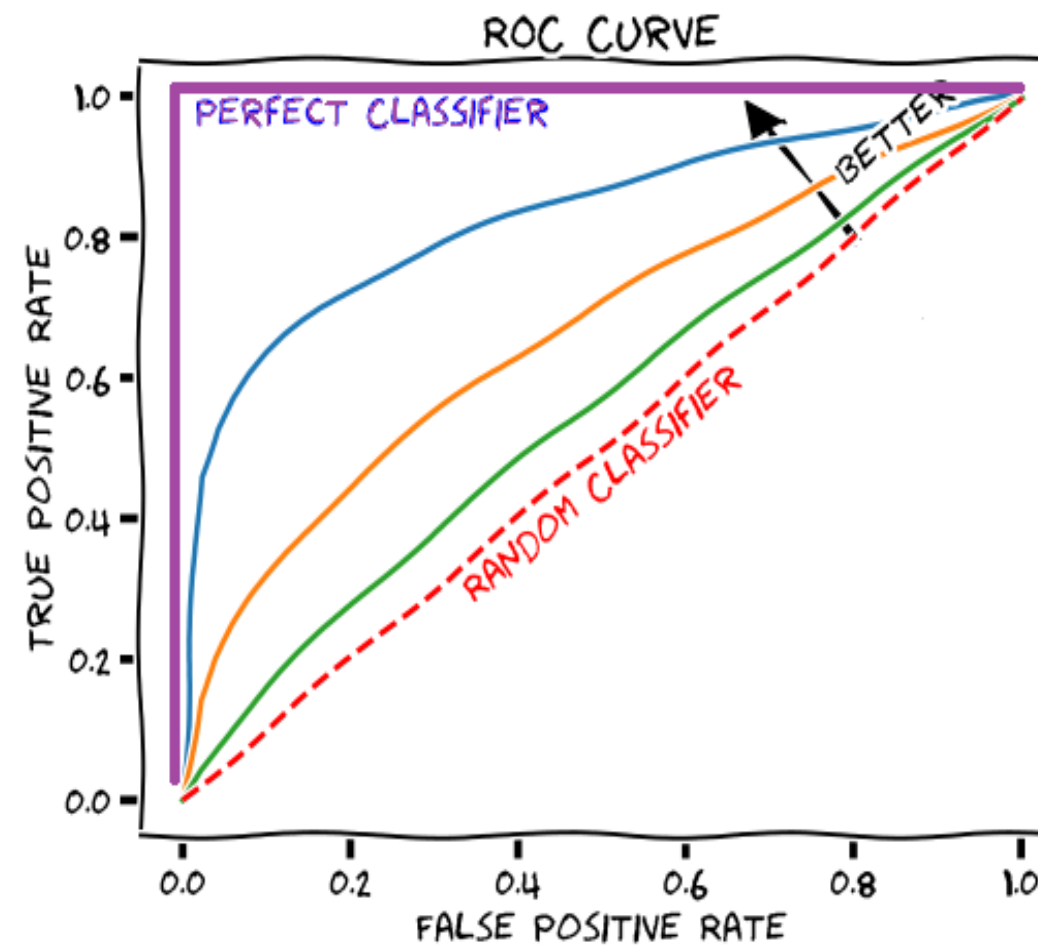
# Modeling

○ ○ ○ ○

## Algoritma

- ❖ SVM (Support Vector Machine)
- ❖ KNN (K-Nearest Neighbor)
- ❖ Logistic Regression
- ❖ Decision Tree
- ❖ Random Forest
- ❖ XGBOOST

## Model Evaluation Metrics



Metrics ROC-AUC dipilih karena robust terhadap data yang imbalance

Pada dataset Payment Default Prediction terdapat 22.12% nasabah yang default dan 77.88% nasabah yang tidak default

# Model Evaluation

ROC-AUC						
Before Hyperparameter Tuning						
Method	SVM	KNN	LOGREG	DECISION TREE	RANDOM FOREST	XGBOOST
Train	0.83	0.88	0.78	1.00	1.00	0.81
Test	0.68	0.69	0.77	0.63	0.74	0.75
False Positive	-	351	256	1039	242	164
False Negative	-	938	906	750	930	631

ROC-AUC						
After Hyperparameter Tuning						
Method	SVM	KNN	LOGREG	DECISION TREE	RANDOM FOREST	XGBOOST
Train	0.79	0.80	0.78	0.74	0.78	0.79
Test	0.68	0.74	0.77	0.74	0.76	0.75
False Positive	-	239	247	325	178	232
False Negative	-	989	912	871	1022	939

# XGBOOST - Confusion Matrix

Confusion Matrix		Predicted Label	
Actual Label		Positive	Negative
		True Positive 469	False Negative 939
		False Positive 232	True Negative 4660

**False Negative:** Hasil prediksi nasabah yang tidak default sedangkan data aktualnya default

**False Positive:** Hasil prediksi nasabah yang default sedangkan data aktualnya tidak default

Positive = Default

Negative = Tidak Default

## Dampak False Positive dan False Negative untuk Bank Taiwan

### False Positive :

Berpotensi menaikkan biaya operasional

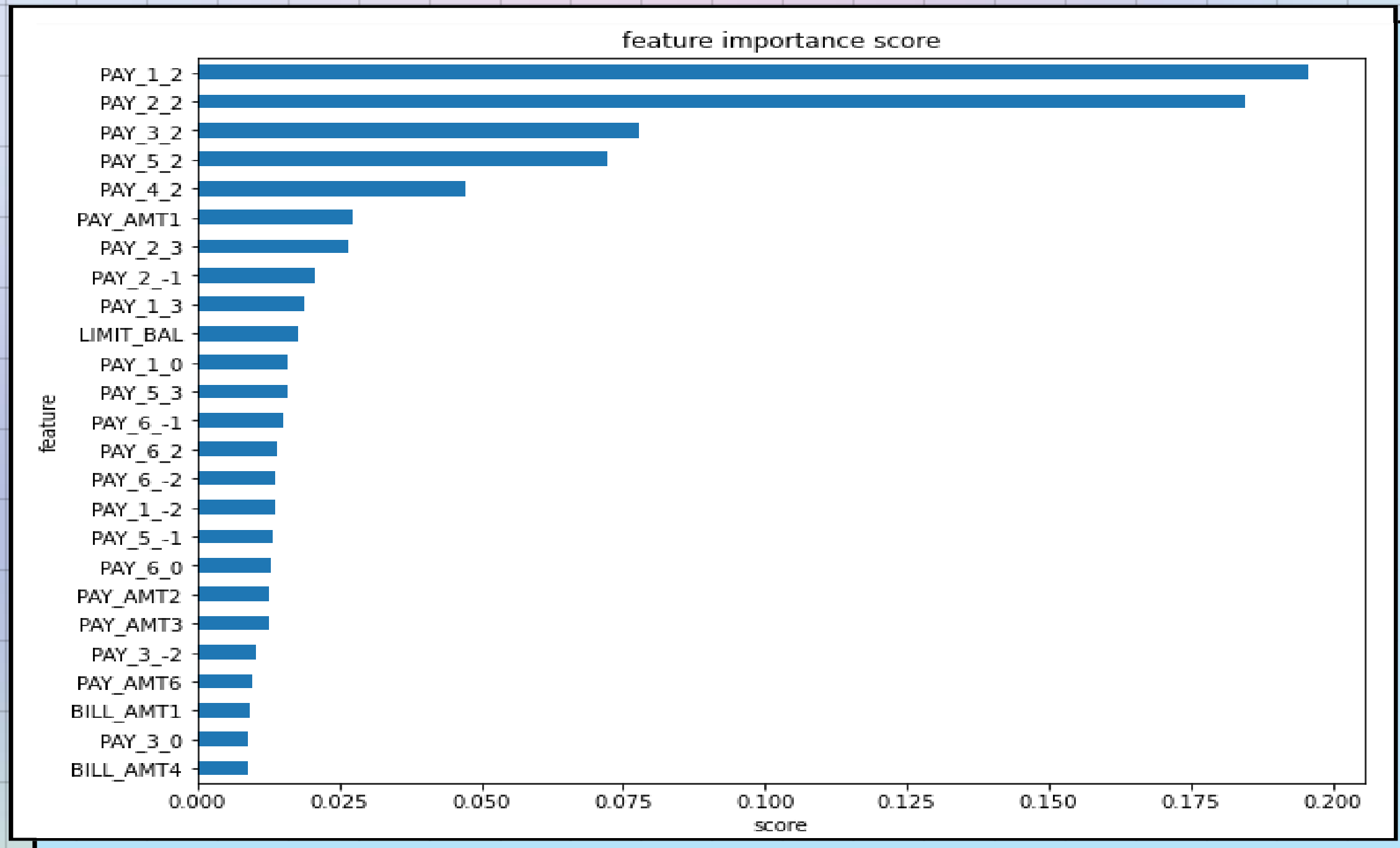
### False Negative :

Berpotensi meningkatkan kerugian bank karena pemberian treatment salah sasaran

### Score Evaluasi XGBOOST

- Accuracy: 0.81
- Precision: 0.67
- Recall: 0.33
- F1-Score: 0.44
- AUC (test-proba): 0.75
- AUC (train-proba): 0.79

# XGBOOST - Feature Importance



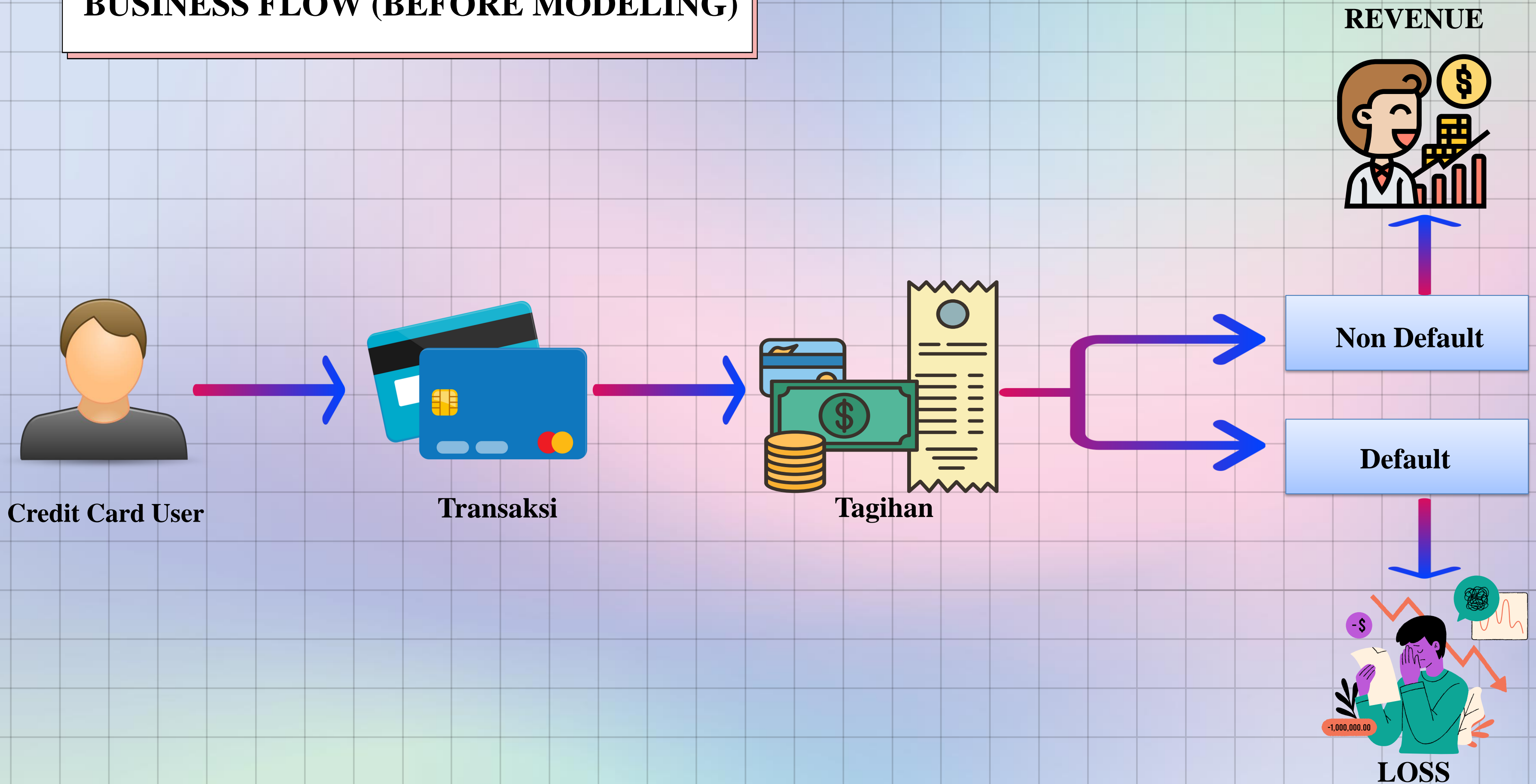
**PAY\_1\_2** merupakan feature yang paling berpengaruh untuk melakukan klasifikasi nasabah



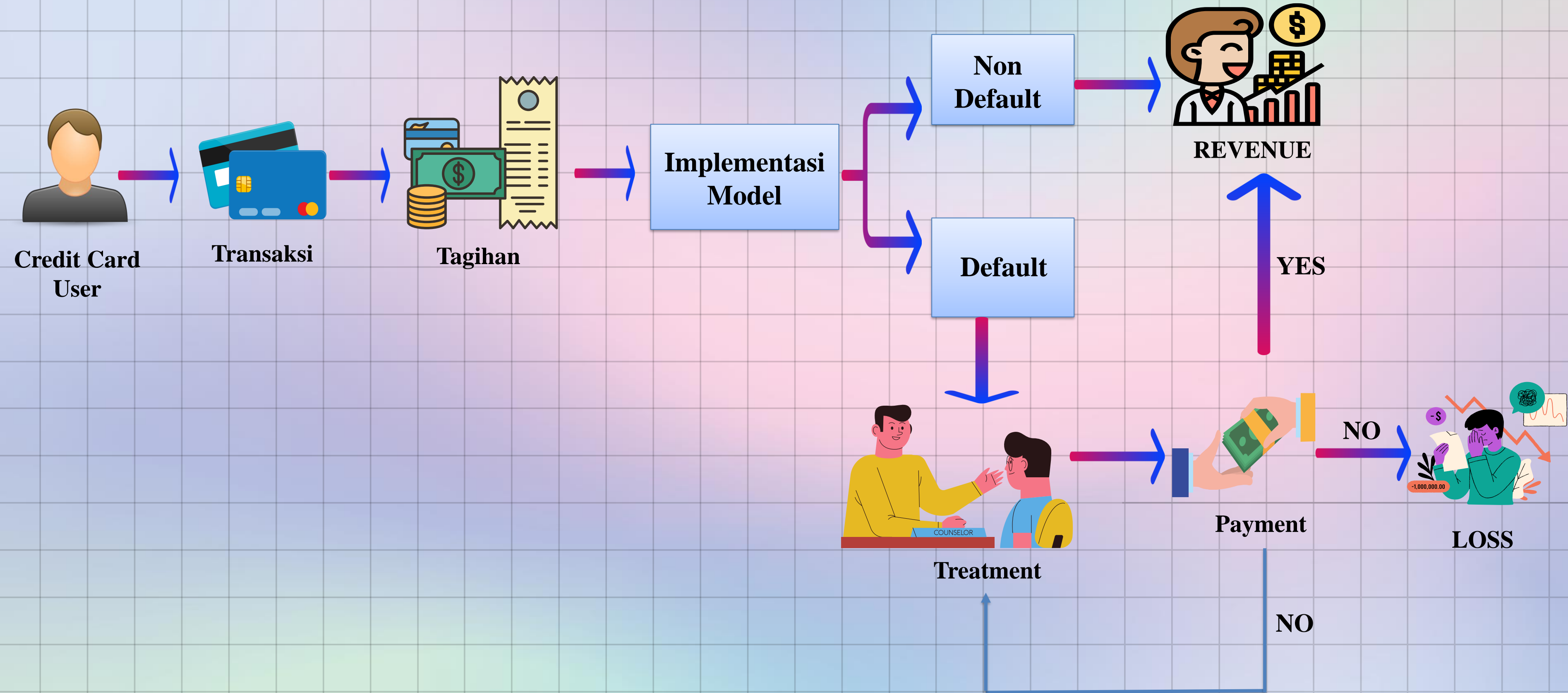
# Business Recommendation & Simulation



# BUSINESS FLOW (BEFORE MODELING)



# BUSINESS FLOW (AFTER MODELING)



## **BUSINESS RECOMMENDATION**

**Treatment untuk nasabah yang berstatus memiliki tunggakan selama 2 bulan di bulan September dan/atau yang terprediksi default :**

1. Mengirim reminder melalui SMS dan telepon
2. Melakukan pemblokiran kartu kredit sementara
3. Menghubungi nasabah untuk menanyakan kendala dan menawarkan keringanan dalam pembayaran tagihan:
  - Mengurangi biaya denda keterlambatan cicilan kartu kredit
  - Menurunkan suku bunga kartu kredit
  - Memberikan kesempatan kepada nasabah untuk mencicil hutang dalam periode waktu yang disepakati bersama
  - Memberikan diskon jika nasabah dapat membayar seluruh hutang dalam 1 kali bayar
  - Melakukan pembebasan bunga dan denda seluruhnya, sehingga nasabah hanya membayar sisa hutang pokok
4. Melakukan penagihan langsung

## **BUSINESS RECOMMENDATION**

### **Treatment untuk nasabah non default**

- Memberikan kemudahan dalam pengajuan peningkatan limit kartu kredit

	BEFORE MODELING	AFTER MODELING
<b>Total Nasabah</b>	21000	21000
<b>(i) Default</b>	4645	3129
<b>(ii) Non Default</b>	16355	17871
<b>Default Rate</b>	22,12%	14,90%
<b>Total Tagihan September 2005</b>	NTD 1.081.532.390	NTD 1.081.532.390
<b>Total Tagihan Gagal Dibayar (Loss)</b>	NTD 220.983.433	NTD 148.860.530
<b>Persentase Kerugian</b>	20,43%	13,76%
<div>Kerugian per Nasabah → 47.574</div>		

Dengan menggunakan model yang dibuat, **default rate** Bank Taiwan berhasil turun sebanyak **7.22%** sehingga **persentase kerugiannya** juga turun sebanyak **6.67%** dengan total kerugian yang mungkin dialami berkurang sebanyak **NTD 72.122.903**.



# THANK YOU



**KELOMPOK  
4-TECH**