OCTAGRAM
IN DATA, WE TRUST

Rakamin Academy
School for Career Acceleration

Final Project Data Science

ANALYSIS PAYMENT DEFAULT ON HOMECREDIT

OCTAGRAM

Meet Our Team

Aldi Vibriani
(Team Leader)

Alfriando C Vean
(Data Scientist)

Andiny Lusy Septiariany
(Business Intelligence)

David Yudha Prasetya
(Data Analyst)

Ramadhian Ekaputra
(Data Analyst)

Reza Yulhansyah
(Data Scientist)

Outline

- Background
- EDA & Insight
- Data Pre-Processing
- Machine Learning
- Business Recomendation

OCTAGRAM

Background

Problem
Goals & Objectives
Business Metrics

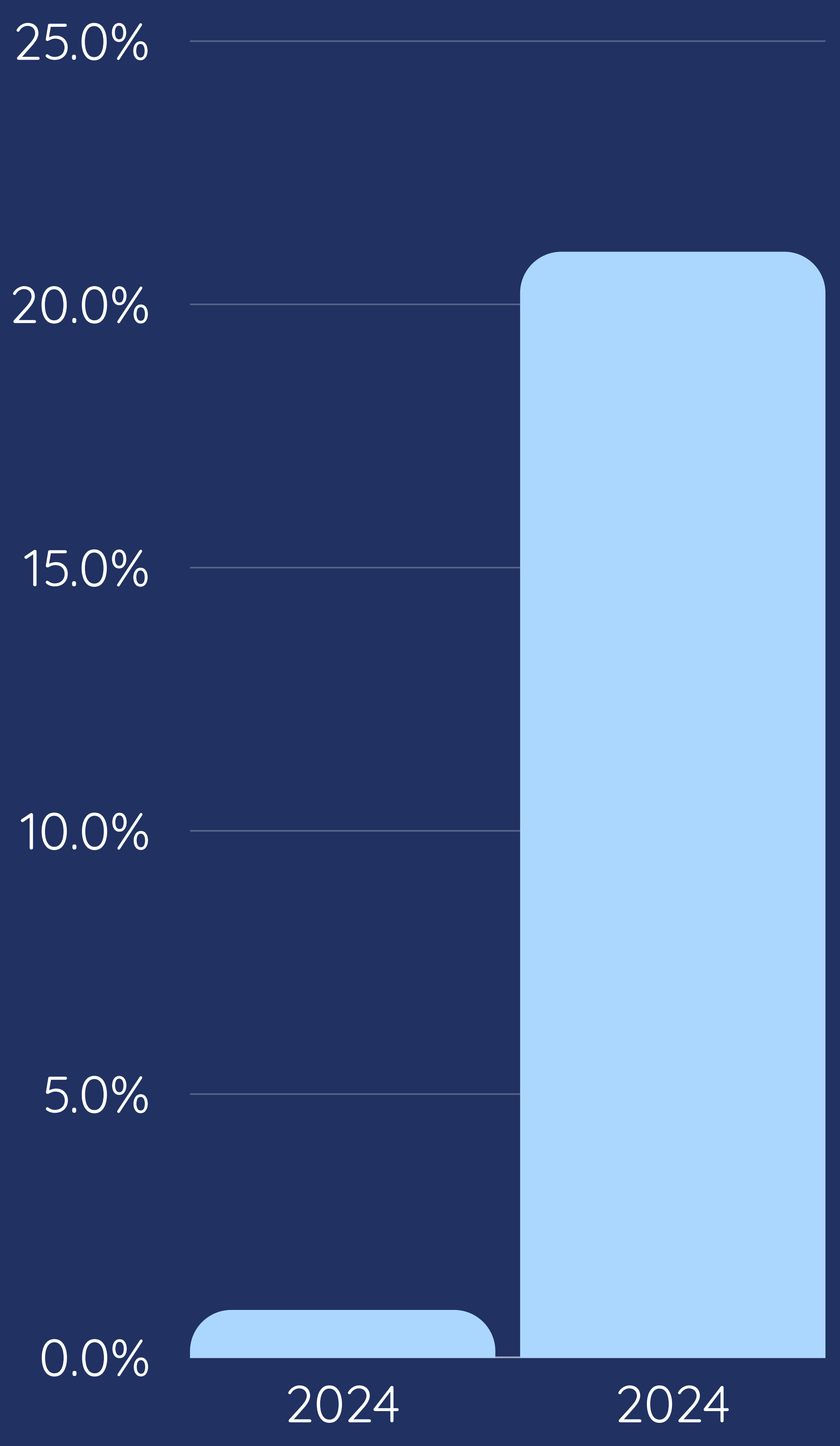
OCTAGRAM

Background Project

Home Credit adalah perusahaan pembiayaan global yang berfokus pada pemberian pinjaman konsumen, terutama kepada pelanggan yang belum terlayani oleh layanan perbankan tradisional.

HOME CREDIT

What is the Problem?



rejected application loan (newyorkfed.org, 2024)

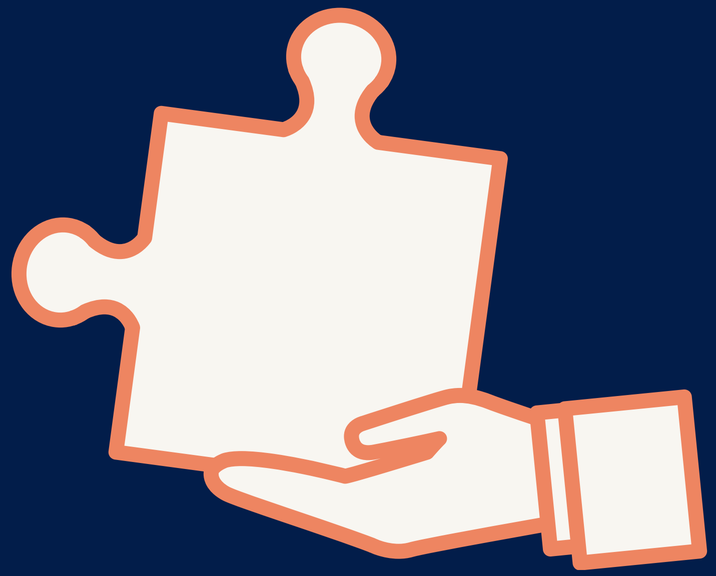
Menurut survey dari Survey of Consumer Expectations (SCE) Credit Access Survey, terjadi **peningkatan penolakan pengajuan pinjaman kredit** dari **0.9%** menjadi **21%** di Tahun 2024 (**newyorkfed.org, 2024**).

Salah satu yang menjadi penyebab pengajuan ditolak adalah **kurangnya credit history customer** dan ketatnya spesifikasi yang harus dipenuhi agar pengajuan pinjaman dapat diterima (**rocketmortgage, 2023., Backendsnews, 2022**).

OCTAGRAM

What is the Problem?

Tantangan bagi HomeCredit adalah banyaknya masyarakat yang **tidak** mempunyai **rekening bank dan credit history** (**rocketmortgage, 2023., Backendsnews, 2022**)



Home Credit berkomitmen untuk layanan pinjaman yang responsible, reliable, dan affordable, termasuk untuk customer yang kurang credit history (**homecredit.co.in, 2024**)

What is the Problem?

OCTAGRAM

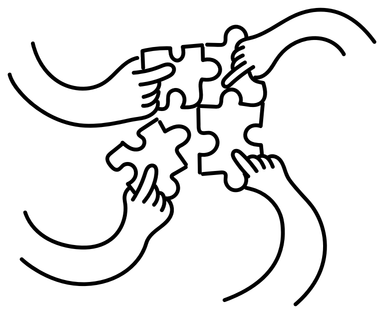
Dengan Machine Learning, HomeCredit dapat menilai kemungkinan gagal bayar dari berbagai peminjam

Kenapa harus Machine Learning?

- 61% pengambilan keputusan dalam bisnis menekankan penerapan AutoML
- Menurut McKinsey, perusahaan yang mengadopsi AI dan ML mencacat peningkatan produktivitas hingga 40%
- Boston Consulting Group melaporkan bahwa perusahaan yang menggunakan AI dapat mengurangi biaya operasional hingga 20-30%

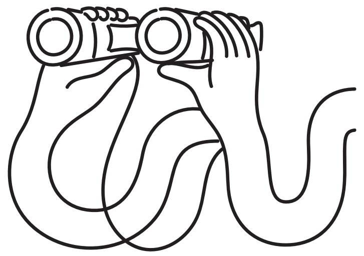
OCTAGRAM

Our Goals & Objective



Goals

Mengidentifikasi nasabah gagal bayar (default rate) dengan lebih akurat.



Objective

Menggunakan Machine Learning untuk menentukan nasabah yang terdeteksi default (karena tidak ada riwayat kredit). Namun, bisa melakukan pinjaman dengan limit menyesuaikan dari total pendapatan

Business Metrics

OCTAGRAM

Credit Default Rate

Mengukur persentase jumlah orang yang gagal bayar setelah pengajuan pinjaman.

EDA & Insight

Dataset

OCTAGRAM

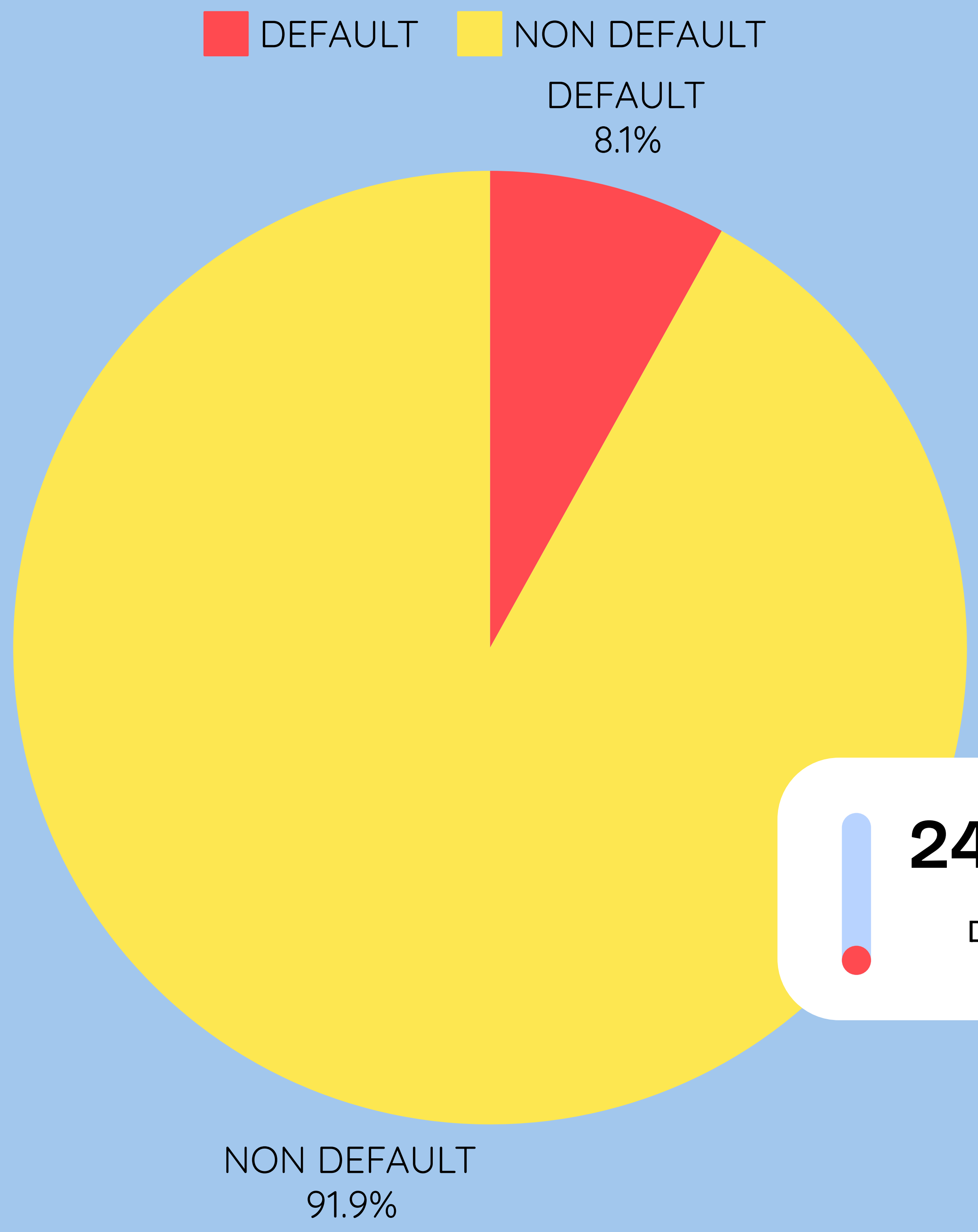
DATASET	Rows	Columns
Application Train	307.511	122
Application Test	48.744	121
Bureau	1.716.428	17
Bureau Balance	27.299.925	3
Previous Application	1.670.214	37
Credit Card Balance	3.840.312	20
Pos Cash Balance	10.001.358	8
Installment	13.605.401	8

Dataset

OCTAGRAM

DATASET	Columns with Null	Sum of Duplicates
Application Train	70	0
Application Test	70	0
Bureau	9	0
Bureau Balance	0	0
Previous Application	26	0
Credit Card Balance	9	0
Posh Cash Balance	3	0
Installment	2	0

Proposi jumlah customer default vs non default



24825
DEFAULT

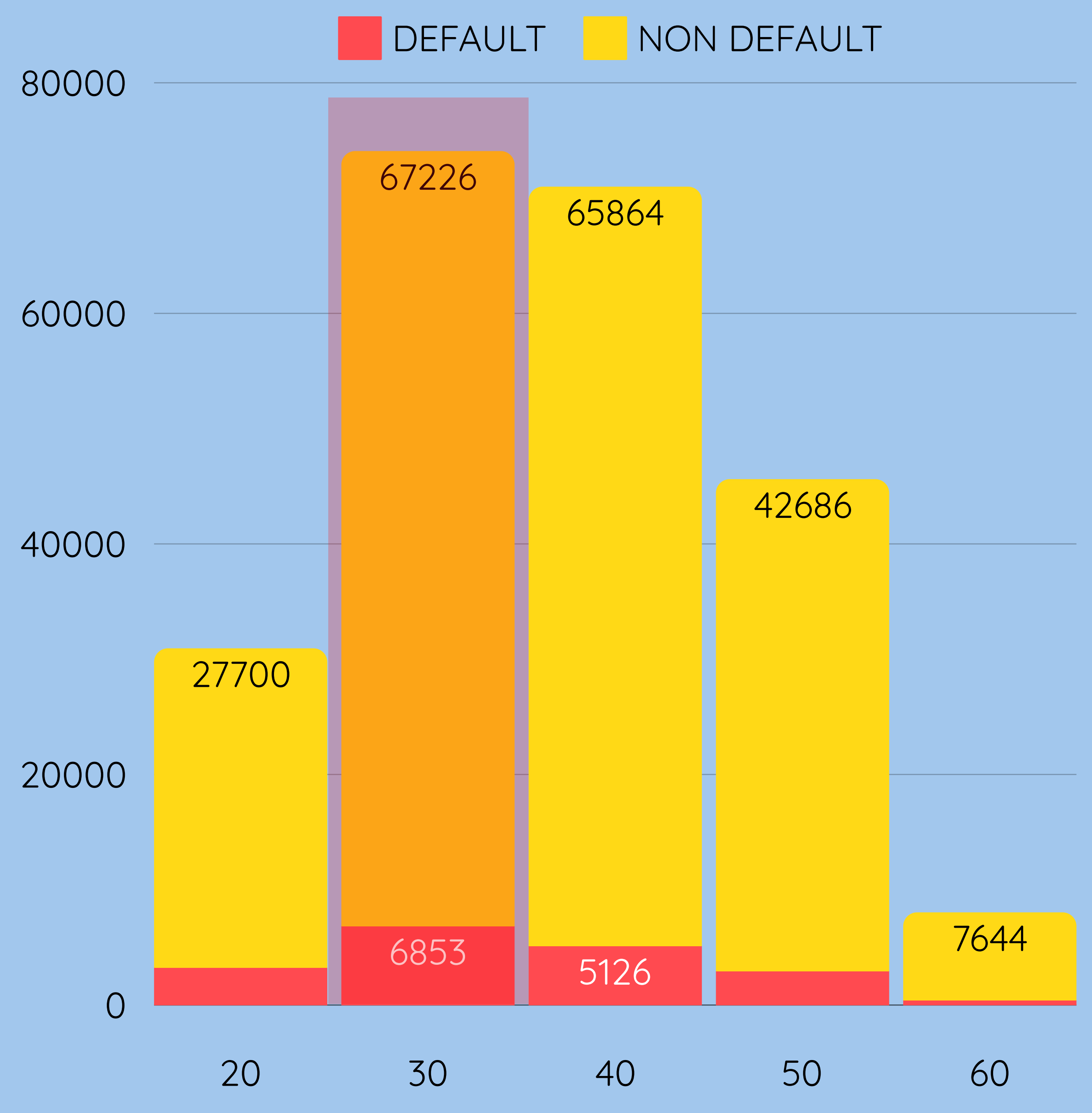
282686
NON DEFAULT

- Label “TARGET” menjelaskan apakah seseorang default atau tidak dalam pembayaran kredit.
- Pada data menunjukan sebanyak 91.9% Default dan 8.1% Not Default.

OCTAGRAM

OCTAGRAM

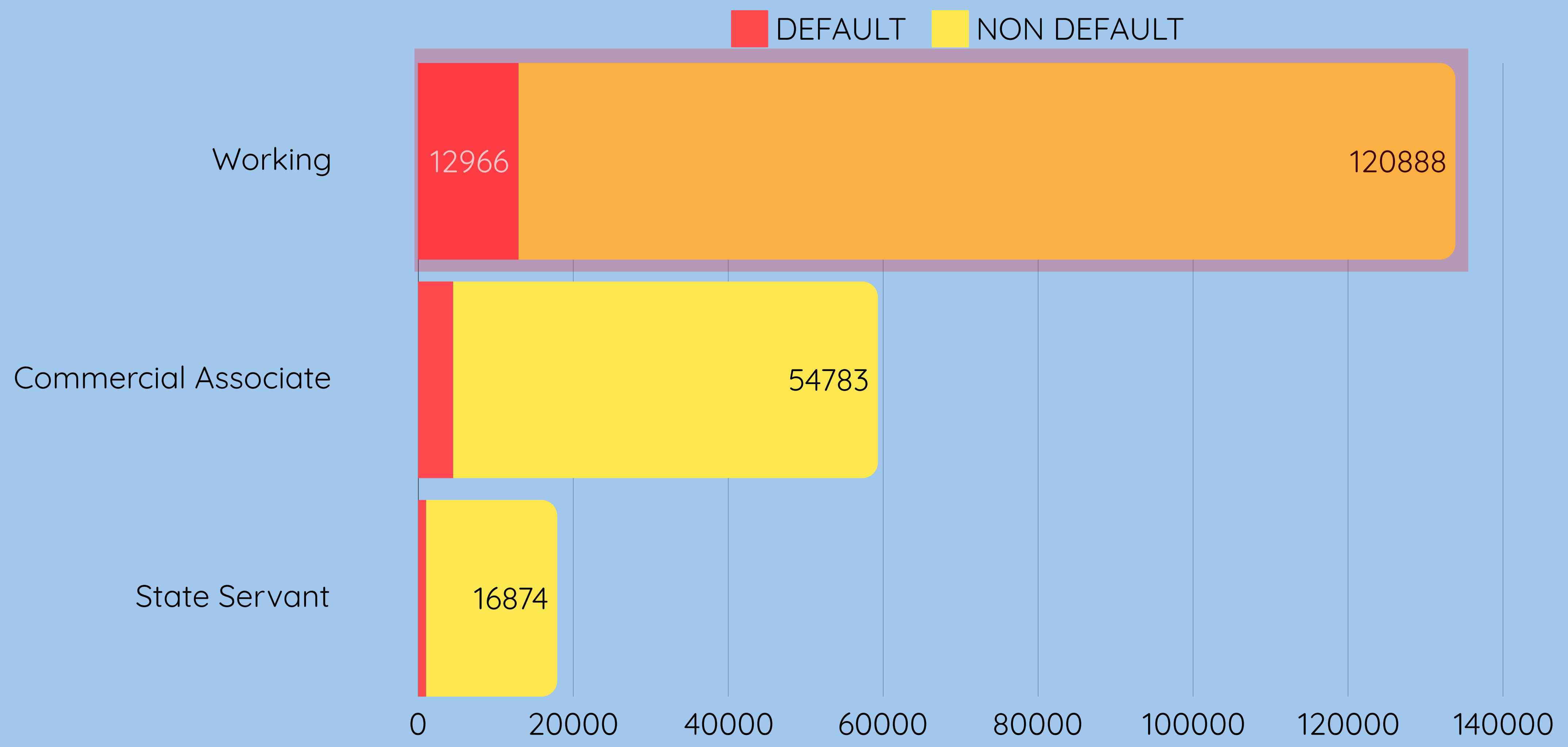
Distribusi umur berdasarkan target



USIA	% TOTAL	% NON DEFAULT	% DEFAULT
20	13.12%	88.29%	11.71%
30	**31.84%**	**89.81%**	**10.19%**
40	31.19%	92.22%	7.78%
50	20.22%	93.13%	6.87%
60	3.62%	94.85%	5.15%

- Mayoritas customer HomeCredit yang gagal untuk membayar cicilan (default) berada di rentang usia 30 tahunan.
- Hal ini perlu menjadi bahan pertimbangan HomeCredit untuk lebih fokus memperhatikan customer dengan usia rentang 30 tahunan karena meskipun memiliki persentase Default yang tinggi (10.19%), tetapi juga menjadi jumlah pelanggan terbanyak yang mengajukan kredit (31.84%).

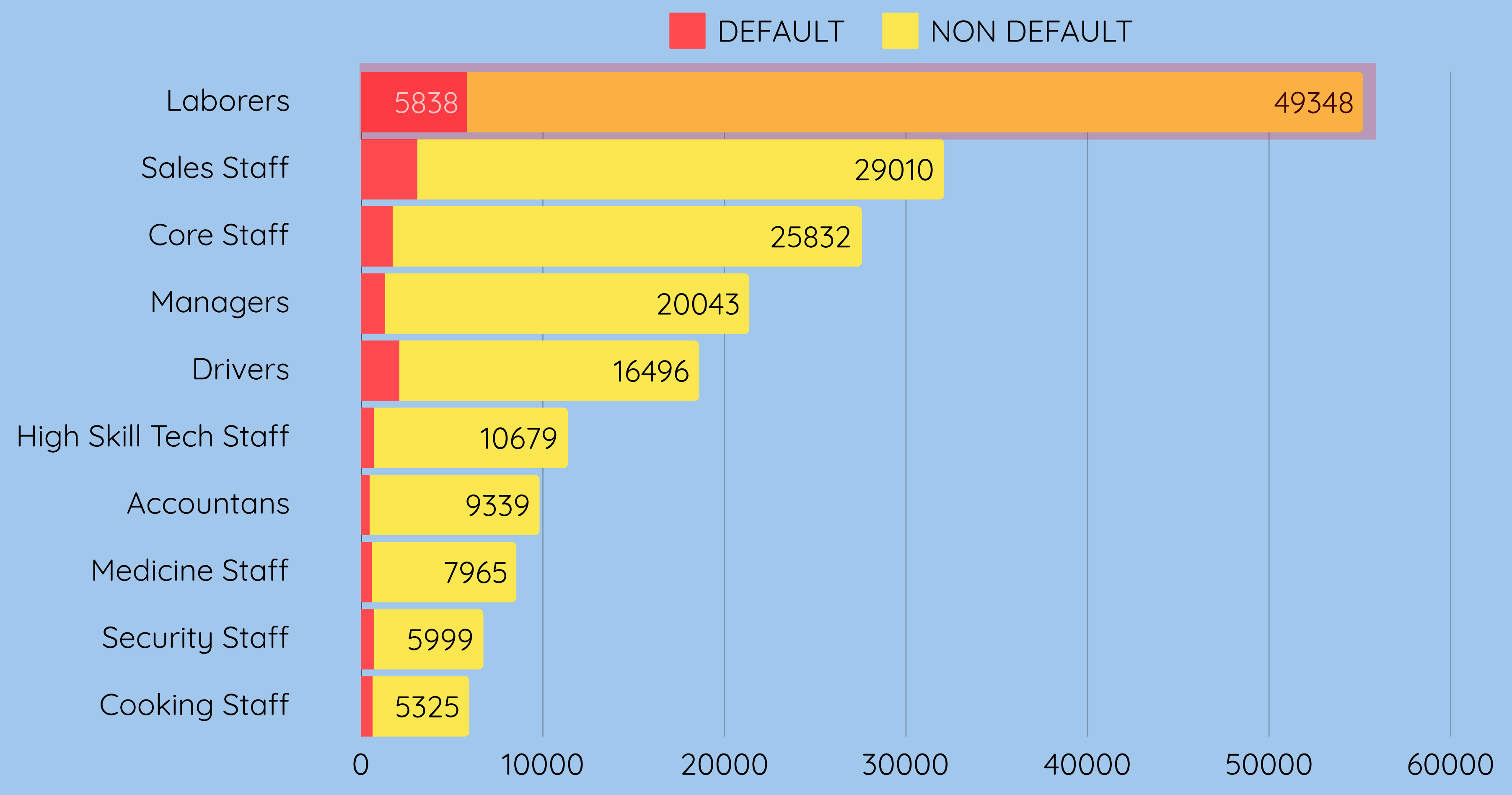
Distribusi Income Type berdasarkan target



Income Type	% TOTAL	% NON DEFAULT	% DEFAULT
Working	**62.78%**	**89.27%**	**10.73%**
Saless Staff	28.45%	91.72%	8.28%
Core Staff	8.76%	93.82%	6.18%

- **Sebanyak 62.78% data mayoritas memiliki pendapatan dari Working (karyawan).**
- **Sebanyak 10.73% dari data mayoritas adalah orang - orang yang tidak capable untuk membayar kredit.**
- **Hal ini menunjukan bahwa customer dengan sumber pendapatan dari Working / karyawan perusahaan adalah orang - orang yang paling banyak melakukan cicilan di HomeCredit.**

Distribusi Occupation Type



Occupation Type	% TOTAL	% NON DEFAULT	% DEFAULT
Laborers	**27.41%**	**88.17%**	**11.83%**
Saless Staff	16.11%	89.34%	10.66%
Core Staff	14.35%	93.27%	6.73%

- **Top 3 level pekerjaan customer adalah Laborers, Sales Staff, dan Core Staff.**
- **Laborers menjadi tipe pekerjaan terbanyak sebesar 27.41% dari total data.**
- **Terdapat 11.83% customer yang tidak capable pada tipe pekerjaan sebagai Laborers.**

Data Preprocessing

OCTAGRAM

Data Preprocessing

Null Values

- Null value over 40%: Drop features atau isi dengan 0.
- under 40%: isi dengan median dan modus.

Duplicated Data

- Tidak ada data duplikat.

Outliers

- Tidak handling outliers karena masih acceptable.
- Agar model mampu belajar dan memprediksi untuk data baru yang kemungkinan outliers.

Feature Encoding & Extraction

- Tipe data distinct dan ordinal dilakukan Label Encoding.
- Application_{train | test}: DEBT_TO_INCOME
- Credit Card Balance: LIMIT_RATIO

Feature Transform

- Handling skewness dengan metode Yeo-Johnson.
- Menyamakan skala dengan Standardization.

Merging Data

- Aggregate dataset berdasarkan SK_ID_CURR dan digabungkan ke dataset application_{train|test}

Train Test Split

- Membagi proporsi 80% untuk data training dan 20% untuk data testing.
- Label "TARGET" menjelaskan seseorang default atau not default.

Imbalance Data

- Menerapkan teknik oversampling: "Synthetic Minority Over-sampling Technique".

Feature Selection

- Menggunakan SelectKBest untuk menyeleksi jumlah feature terbaik.
- Namun, hasil score menunjukan bahwa perlu menggunakan seluruh jumlah features.

OCTAGRAM

OCTAGRAM

Machine Learning

MACHINE LEARNING

OCTAGRAM

BEST FIT

Logistic Regression		
	Train	Test
Recall	0.72	0.66
ROC AUC	0.77	0.74

Random Forest		
	Train	Test
Recall	0.99	0.72
ROC AUC	0.99	0.69

Decision Tree		
	Train	Test
Recall	1.00	0.74
ROC AUC	1.00	0.50

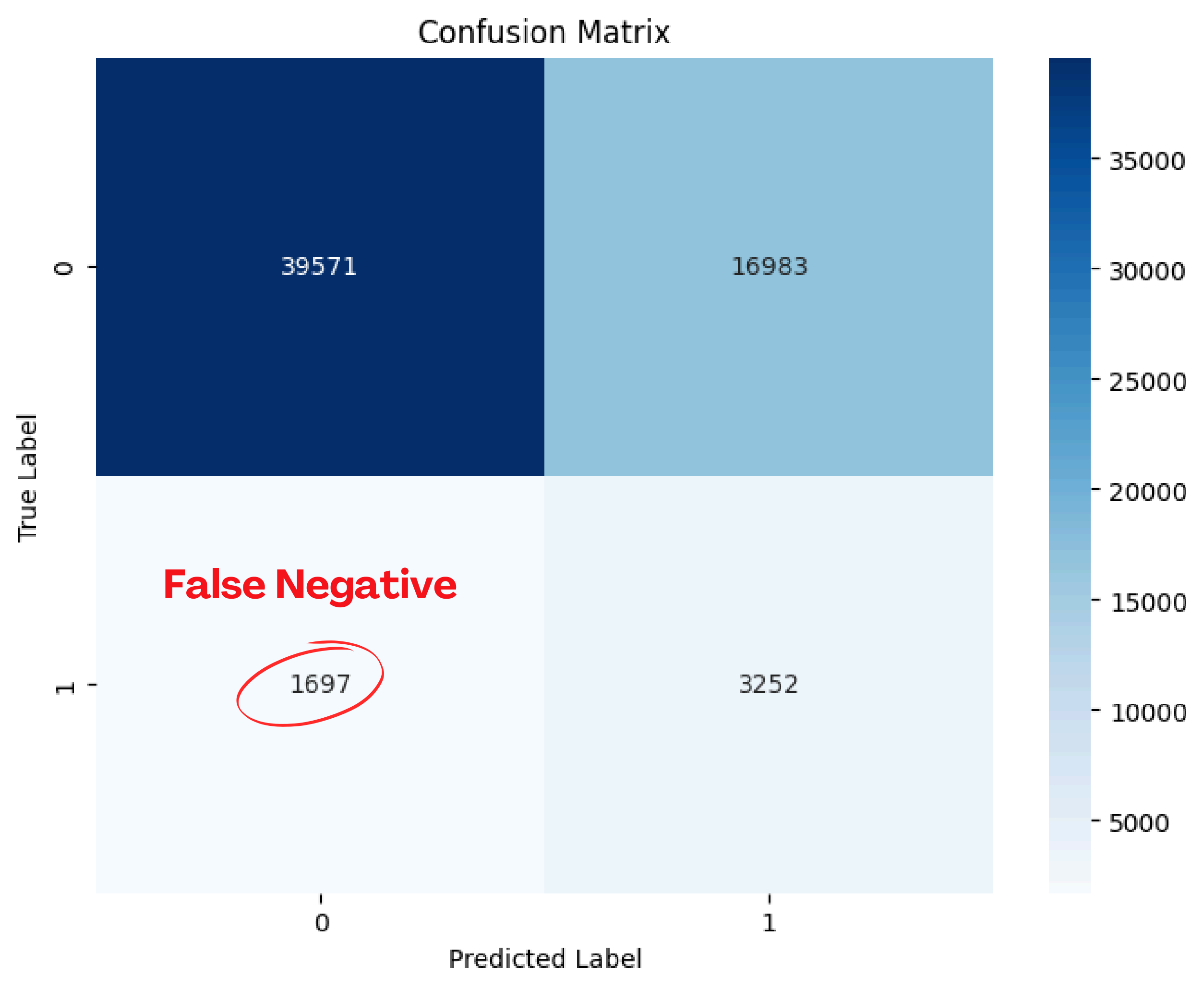
Ada Boost		
	Train	Test
Recall	0.88	0.98
ROC AUC	0.95	0.62

XGBoost		
	Train	Test
Recall	0.92	0.93
ROC AUC	0.99	0.38

Summarizing Results

Logistic Regression		
	Train	Test
Recall	0.72	0.66
ROC AUC	0.77	0.74

“Model ini menunjukkan kemampuan yang kuat dalam meminimalkan False Negatives, yang sangat penting untuk mengidentifikasi potensi default dengan akurat.”



Dengan kemampuan prediksi ini, Home Credit dapat **mengidentifikasi** pelanggan yang **berpotensi menimbulkan financial risk** sebelum masalah terjadi.

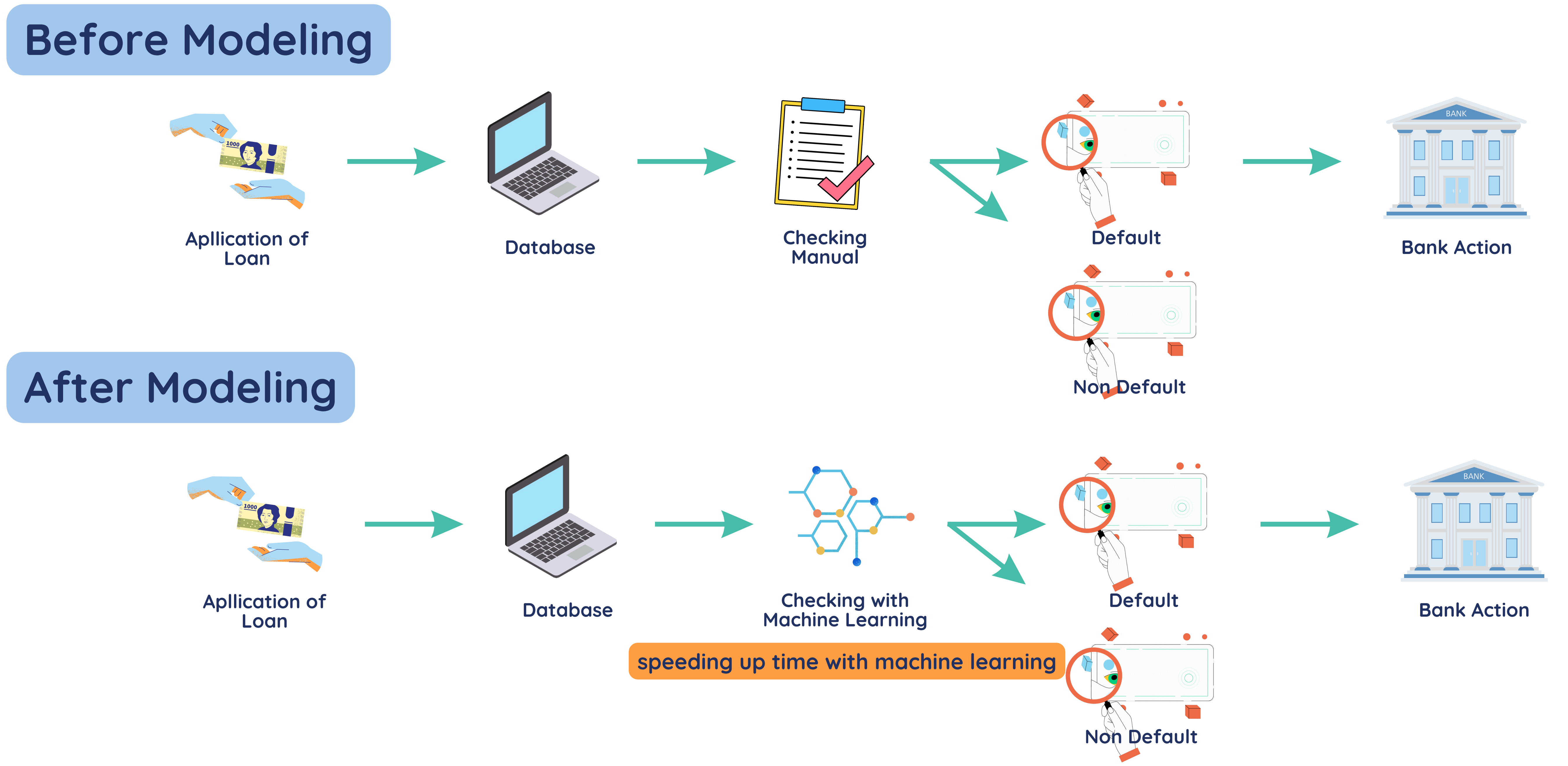
OCTAGRAM

Business Recomendation

OCTAGRAM

Business Recomendation

WORKFLOW



OCTAGRAM

Simulasi Perhitungan Kerugian dalam Penerapan Machine Learning

Asumsi terdapat **61.503** data pengajuan kredit yang diterima dan rata - rata kreditnya adalah **$596.564.**

Jumlah orang yang default **sebelum** ada model: **8.1% x 61.503 = 4.949 orang**

Jumlah orang yang default **setelah** ada model: **5.3% x 61.503 = 2.252 orang**

Kerugian per total data

$$\textit{Total Kerugian} = \frac{\textit{Jumlah orang yang gagal bayar (default)}}{\textit{Total Kreditur}}$$

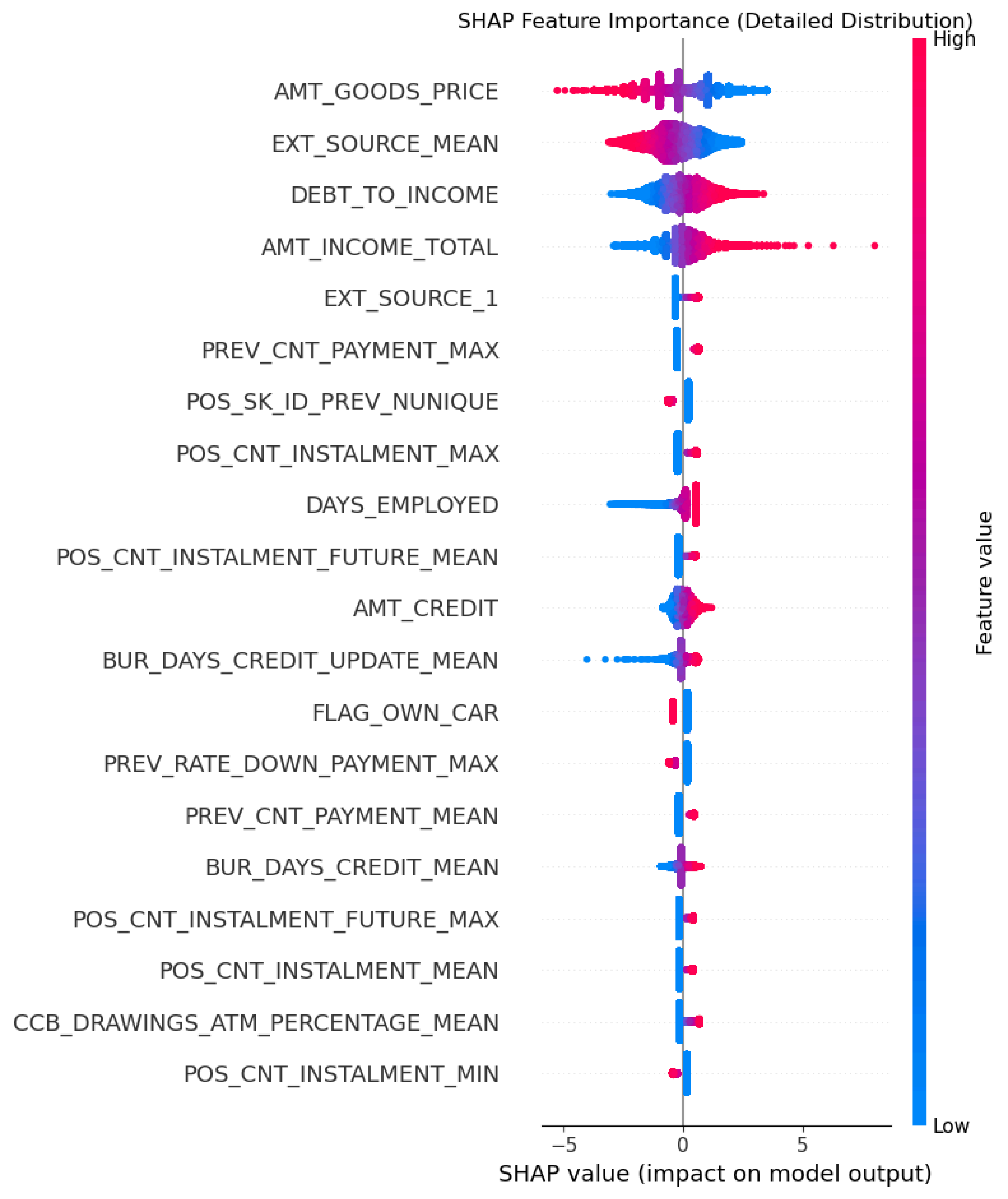
Total Kerugian : **4.949 x $ 596.564 = $ 2.964.464.725**
Sebelum ada model

Total Kerugian : **3.252 x $ 596.564 = $ 1.948.029.300**
Setelah ada model

Persentase penurunan kerugian

34.29%

OCTAGRAM



Feature Importance

- Terdapat **empat features** yang memiliki **pengaruh yang tinggi** dalam memperediksi seseorang default atau tidak.
- AMT_GOODS_PRICE dan EXT_SOURCE_MEAN semakin **tinggi** nilainya maka diprediksi **tidak default**, vice versa.
- DEBT_TO_INCOME dan AMT_INCOME_TOTAL semakin **tinggi** nilainya maka diprediksi **semakin default**, vice versa.

Business Recomendation

OCTAGRAM

Default:

- Penilaian score perlu memberikan bobot lebih pada **AMT_GOODS_PRICE, EXT_SOURCE_MEAN,** dan **DEBT_TO_INCOME** karena hal ini **penting** dalam **memprediksi seseorang gagal bayar atau tidak**.
- Mengembangkan **risk-based pricing** untuk menentukan **suku bunga** berdasarkan customer risk.
- Melakukan **monitoring** untuk melihat default rate dan profit margins untuk setiap **4 bulan**.

saran saat proses monitoring:

- Menerapkan sistem auto reminder melalui email, WhatsApp, dan notifikasi di smartphone.
- Early detection dan warning untuk customer yg telat bayar 3 kali

Rekomendasi setelah 4 bulan pemantauan:

Jika **Tingkat Default Rate Meningkat**:

- Tetapkan batas kredit sesuai dengan kemampuan pelanggan.
- Kurangi aktivitas pemberian pinjaman kepada pelanggan dengan risiko tinggi.

Jika **Tingkat Default Rate Menurun**:

- Tawarkan produk keuangan lain kepada pelanggan yang selalu membayar tepat waktu.
- Berikan promosi, seperti pengajuan pinjaman jangka panjang dengan manfaat tambahan, seperti pembebasan biaya tahunan.

THANK YOU

You can reach us to:

- **Aldi Vibriani:**
  - **https://www.linkedin.com/in/aldi-vibriani/**
- **Andiny Lusy:**
  - **https://www.linkedin.com/in/andinylusy/**
- **Alfriando C Vean**
  - **https://www.linkedin.com/in/alfriandocvean/**
- **David Yudha Prasetya**
  - **https://www.linkedin.com/in/davidyudhaprasetya/**
- **Reza Yulhansyah**
  - **https://www.linkedin.com/in/reza-yulhansyah/**
- **Ramadhian Ekaputra**
  - **https://www.linkedin.com/in/ramadhian-ekaputra/**

APPENDIX

- Back End News. (2022). Home Credit develops programs that promote financial inclusion. Diakses dari https://backendnews.net/home-credit-develops-programs-that-promote-financial-inclusion/
- Home Credit India. (2019). Home Credit launches a multi-city drive to deepen its offering and broaden financial inclusion in India. Diakses dari https://www.homecredit.co.in/en/home-credit-launches-a-multi-city-drive-to-deepen-its-offering-and-broaden-financial-inclusion-in-india
- Home Credit Indonesia. (2024). Home Credit Indonesia completes change in shareholder. Diakses dari https://www.homecredit.co.id/tentang-perusahaan/siaran-pers/Home-Credit-Indonesia-completes-change-in-shareholder
- Federal Reserve Bank of New York. (2024). Consumers Report Higher Credit Rejection Rates, Expect Fewer Credit Applications. Diakses dari https://www.newyorkfed.org/newsevents/news/research/2024/20241118
- Rocket Mortgage. (2023). What to do if your mortgage loan application is denied. Diakses dari https://www.rocketmortgage.com/learn/what-to-do-if-your-mortgage-loan-application-is-denied
- Back End News. (2022). Home Credit develops programs that promote financial inclusion. Diakses dari https://backendnews.net/home-credit-develops-programs-that-promote-financial-inclusion/
- Springer. (2023). Machine learning techniques for default prediction: an application to small Italian companies. Diakses dari https://link.springer.com/article/10.1057/s41283-023-00132-2
- DataRobot. (n.d.). Loan default. Diakses dari https://docs.datarobot.com/en/docs/api/guide/common-case/loan-default/index.html
- ScienceDirect. (2021). Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. Diakses dari https://www.sciencedirect.com/science/article/abs/pii/S1057521921002878
- FICO. (n.d.). Explainable Machine Learning for Probability of Default Calculations. Diakses dari https://www.fico.com/blogs/Explainable-Machine-Learning-for-Probability-of-Default-Calculations

Detail Metrics Model

Final (Agregat data setiap dataset)

Decision Tree

	Train	Test
Accuracy	1.00	0.30
Precision	1.00	0.08
Recall	1.00	0.74
F1-Score	1.00	0.14
F2-Score	1.00	0.28
ROC AUC	1.00	0.50

Logistic Regression

	Train	Test
Accuracy	0.71	0.70
Precision	0.70	0.16
Recall	0.72	0.66
F1-Score	0.71	0.26
F2-Score	0.71	0.41
ROC AUC	0.77	0.74

Random Forest

	Train	Test
Accuracy	0.99	0.57
Precision	0.99	0.13
Recall	0.99	0.72
F1-Score	0.99	0.22
F2-Score	0.99	0.37
ROC AUC	0.99	0.69

Ada Boost

	Train	Test
Accuracy	0.89	0.13
Precision	0.90	0.09
Recall	0.88	0.98
F1-Score	0.89	0.15
F2-Score	0.88	0.31
ROC AUC	0.95	0.62

XGBoost

	Train	Test
Accuracy	0.96	0.23
Precision	0.99	0.09
Recall	0.92	0.93
F1-Score	0.96	0.16
F2-Score	0.93	0.32
ROC AUC	0.99	0.38

LightGBM

	Train	Test
Accuracy	0,95	0,08
Precision	0,99	0,08
Recall	0,91	1,00
F1-Score	0,95	0,14
F2-Score	0,92	0,30
ROC AUC	0,98	0,61

Detail Metrics Model

Tidak agregat data & drop fitur korelasi tinggi

Decision Tree

	Train	Test
Accuracy	1,00	0,81
Precision	1,00	0,14
Recall	1,00	0,20
F1-Score	1,00	0,16
F2-Score	1,00	0,18
ROC AUC	1,00	0,54

Logistic Regression

	Train	Test
Accuracy	0,71	0,69
Precision	0,70	0,17
Recall	0,72	0,64
F1-Score	0,71	0,27
F2-Score	0,72	0,41
ROC AUC	0,77	0,73

Random Forest

	Train	Test
Accuracy	1,00	0,90
Precision	1,00	0,38
Recall	1,00	0,03
F1-Score	1,00	0,06
F2-Score	1,00	0,04
ROC AUC	1,00	0,71

Ada Boost

	Train	Test
Accuracy	0,90	0,86
Precision	0,93	0,22
Recall	0,87	0,18
F1-Score	0,90	0,20
F2-Score	0,88	0,18
ROC AUC	0,95	0,26

XGBoost

	Train	Test
Accuracy	0,96	0,90
Precision	0,99	0,41
Recall	0,92	0,06
F1-Score	0,96	0,10
F2-Score	0,94	0,07
ROC AUC	0,99	0,72

Detail Metrics Model

Tidak agregat data, SMOTETomek, SelectKBest 140 Fitur

Decision Tree

	Train	Test
Accuracy	0,77	0,33
Precision	0,86	0,09
Recall	0,65	0,74
F1-Score	0,74	0,16
F2-Score	0,68	0,31
ROC AUC	0,85	0,53

Logistic Regression

	Train	Test
Accuracy	0,71	0,35
Precision	0,70	0,10
Recall	0,72	0,79
F1-Score	0,71	0,18
F2-Score	0,72	0,33
ROC AUC	0,77	0,56

Random Forest

	Train	Test
Accuracy	0,77	0,33
Precision	0,86	0,09
Recall	0,65	0,74
F1-Score	0,74	0,16
F2-Score	0,68	0,31
ROC AUC	0,85	0,53

Ada Boost

	Train	Test
Accuracy	0,90	0,09
Precision	0,92	0,08
Recall	0,87	0,99
F1-Score	0,90	0,16
F2-Score	0,88	0,32
ROC AUC	0,95	0,63

XGBoost

	Train	Test
Accuracy	0,96	0,08
Precision	0,99	0,08
Recall	0,92	1.00
F1-Score	0,96	0,16
F2-Score	0,94	0,32
ROC AUC	0,99	0,55

Detail Metrics Model

Tidak agregat data, Drop fitur VIF >10%, SelectKBest 125 Fitur

Decision Tree

	Train	Test
Accuracy	1,00	0,12
Precision	1,00	0,08
Recall	1,00	0,91
F1-Score	1,00	0,15
F2-Score	1,00	0,31
ROC AUC	1,00	0,48

Logistic Regression

	Train	Test
Accuracy	0,70	0,37
Precision	0,69	0,10
Recall	0,72	0,78
F1-Score	0,70	0,18
F2-Score	0,71	0,33
ROC AUC	0,77	0,57

Random Forest

	Train	Test
Accuracy	1,00	0,09
Precision	1,00	0,08
Recall	1,00	0,99
F1-Score	1,00	0,16
F2-Score	1,00	0,33
ROC AUC	1,00	0,58

Ada Boost

	Train	Test
Accuracy	0,88	0,09
Precision	0,89	0,08
Recall	0,87	1,00
F1-Score	0,88	0,16
F2-Score	0,88	0,32
ROC AUC	0,95	0,62

XGBoost

	Train	Test
Accuracy	0,96	0,08
Precision	0,99	0,08
Recall	0,92	1.00
F1-Score	0,96	0,16
F2-Score	0,94	0,32
ROC AUC	0,99	0,49

Detail Metrics Model

Feature Extraction (VIF >10%), Drop fitur VIF >10%, SelectKBest 122 Fitur

Decision Tree

	Train	Test
Accuracy	0,81	0,16
Precision	0,86	0,09
Recall	0,74	0,93
F1-Score	0,80	0,16
F2-Score	0,76	0,32
ROC AUC	0,89	0,51

Logistic Regression

	Train	Test
Accuracy	0,70	0,90
Precision	0,69	0,21
Recall	0,72	0,00
F1-Score	0,70	0,01
F2-Score	0,71	0,00
ROC AUC	0,77	0,61

Random Forest

	Train	Test
Accuracy	0,99	0,09
Precision	0,99	0,09
Recall	1,00	1,00
F1-Score	0,99	0,16
F2-Score	1,00	0,33
ROC AUC	1,00	0,58

Ada Boost

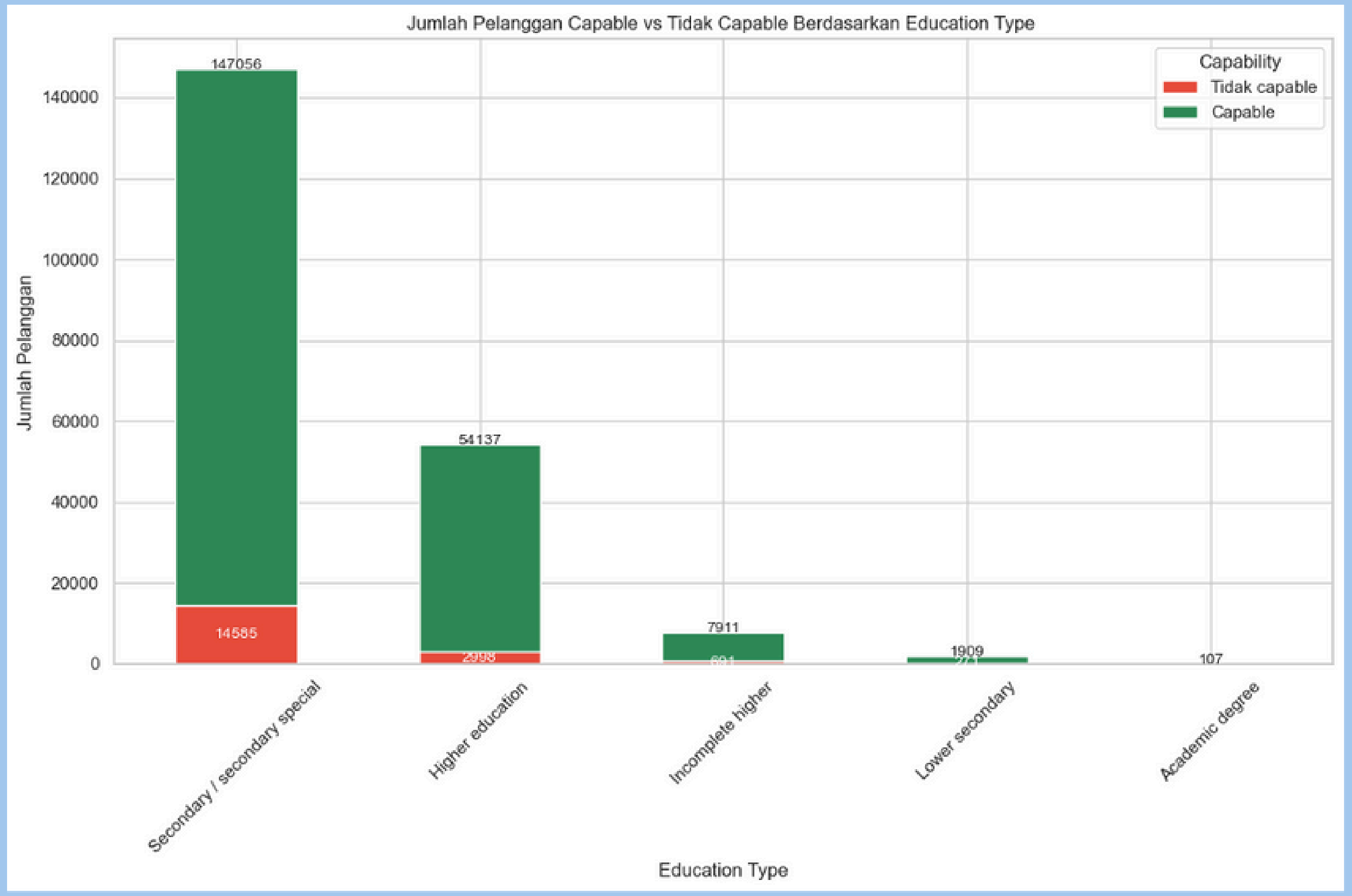
	Train	Test
Accuracy	0,88	0,09
Precision	0,89	0,08
Recall	0,87	1,00
F1-Score	0,88	0,16
F2-Score	0,88	0,32
ROC AUC	0,95	0,62

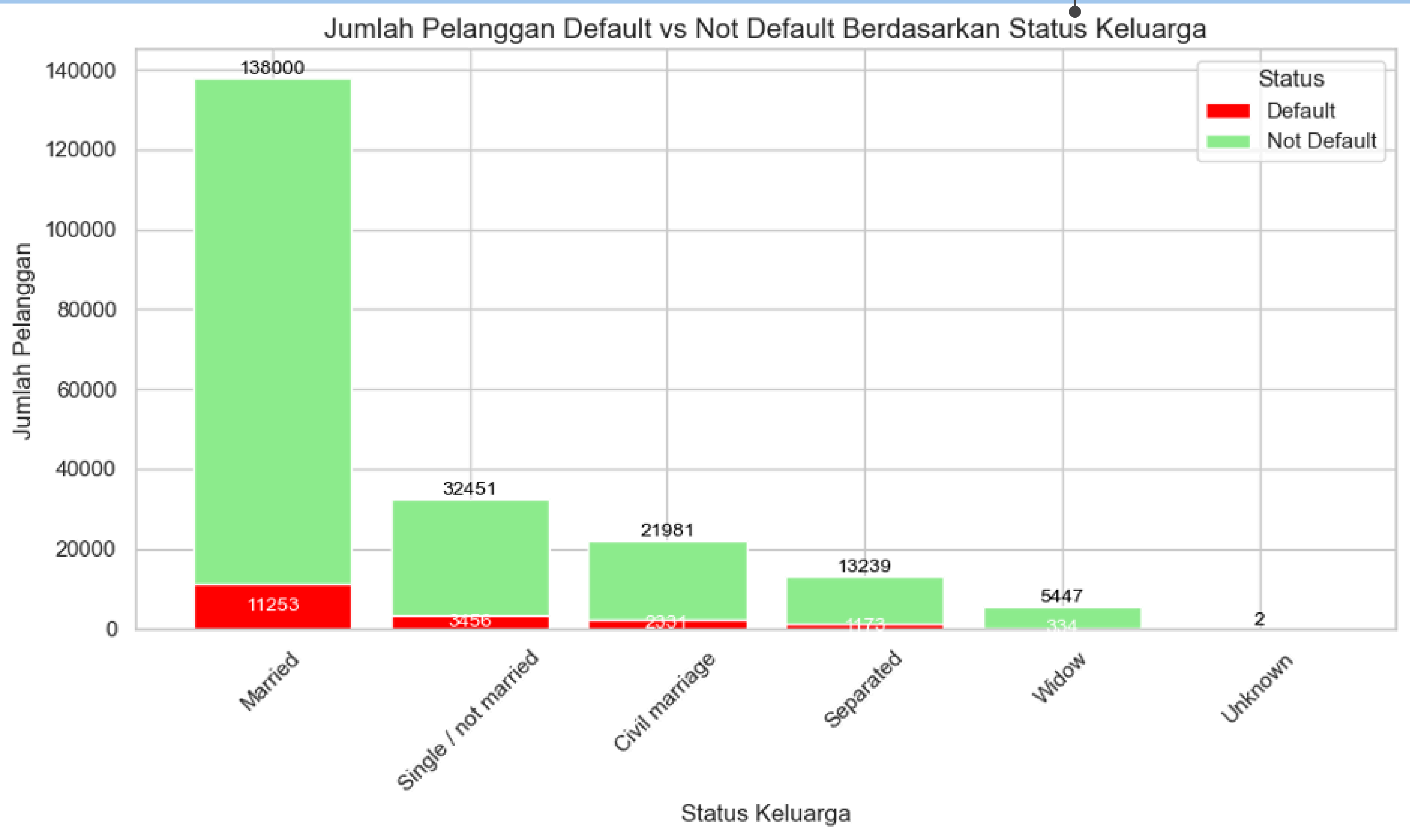
XGBoost

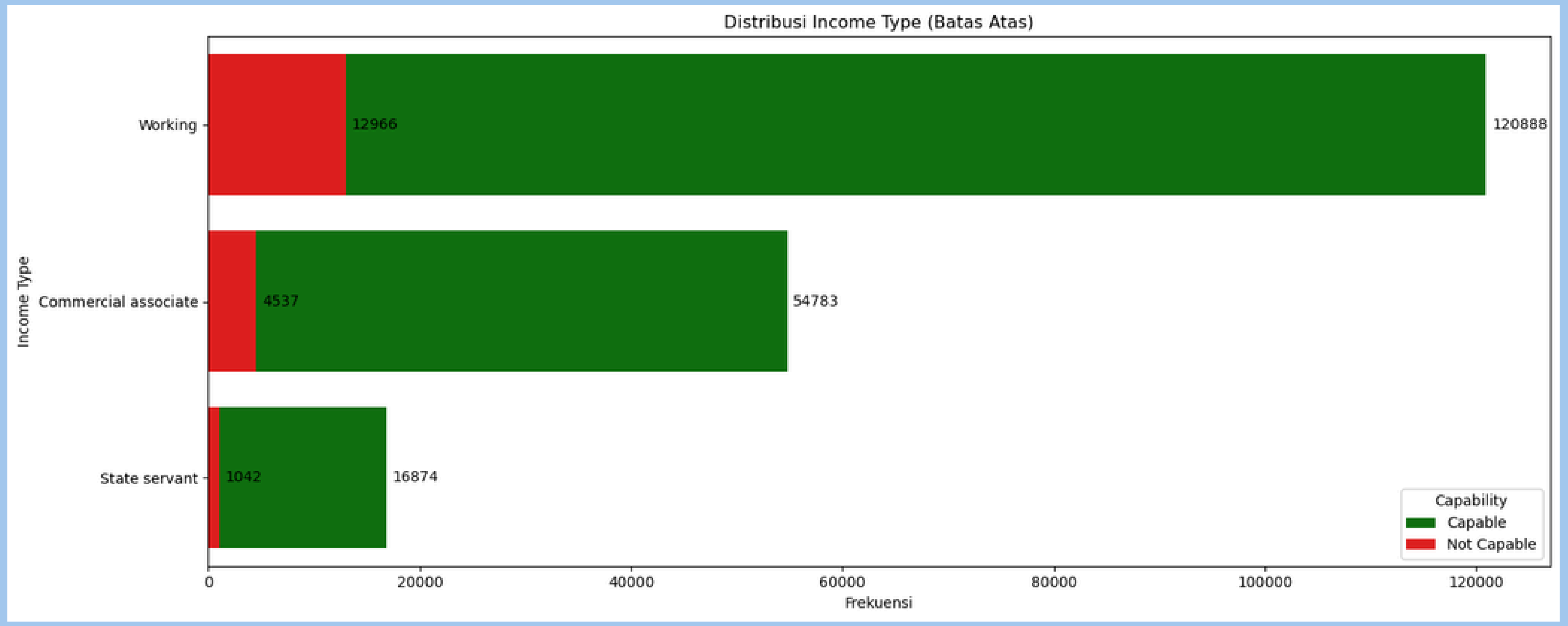
	Train	Test
Accuracy	0,96	0,08
Precision	0,99	0,08
Recall	0,92	1.00
F1-Score	0,96	0,16
F2-Score	0,94	0,32
ROC AUC	0,99	0,49

Exploratory Data Analysis

OCTAGRAM



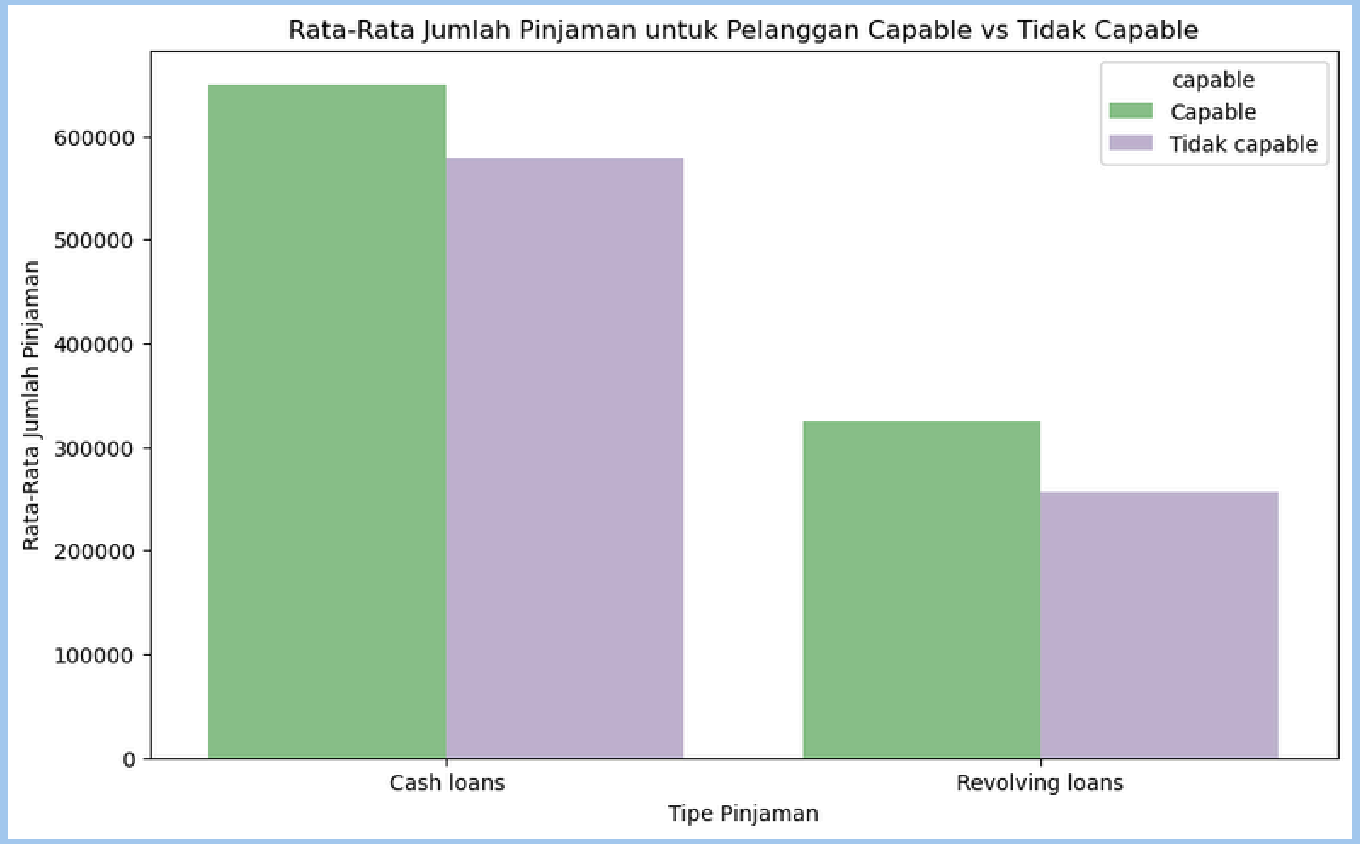


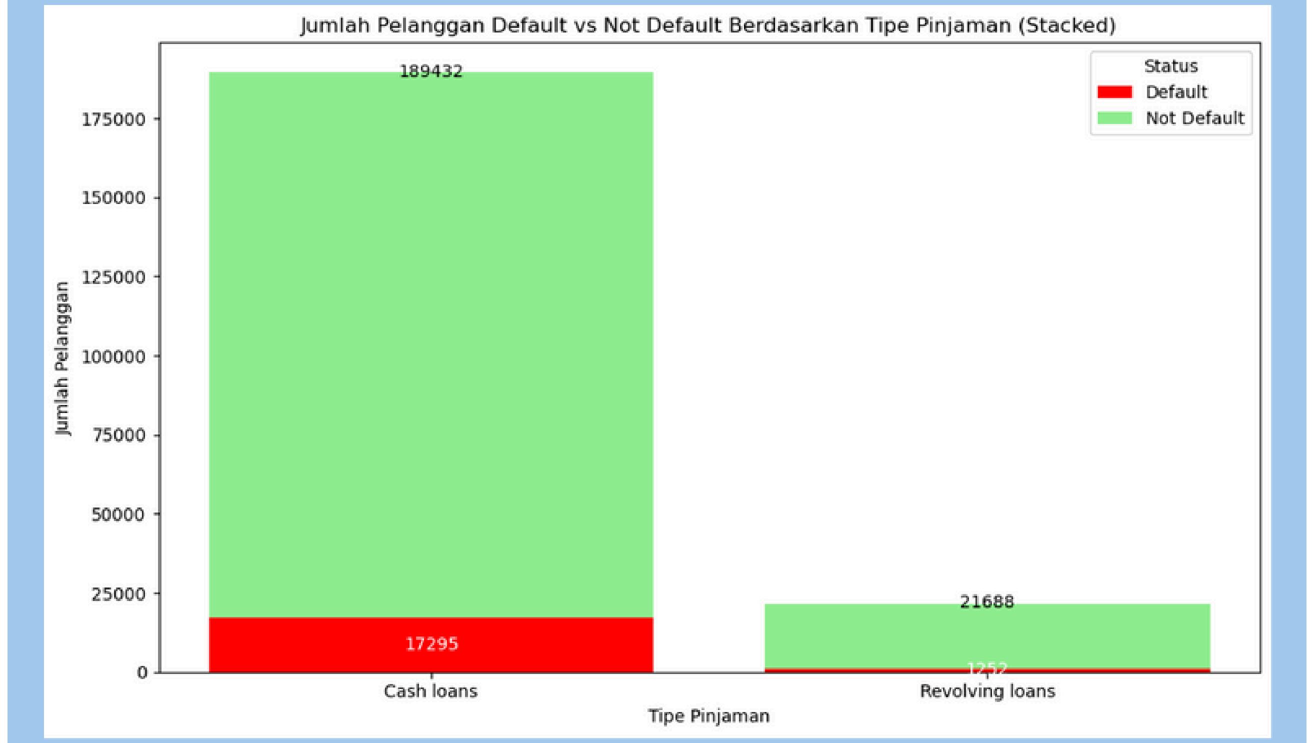


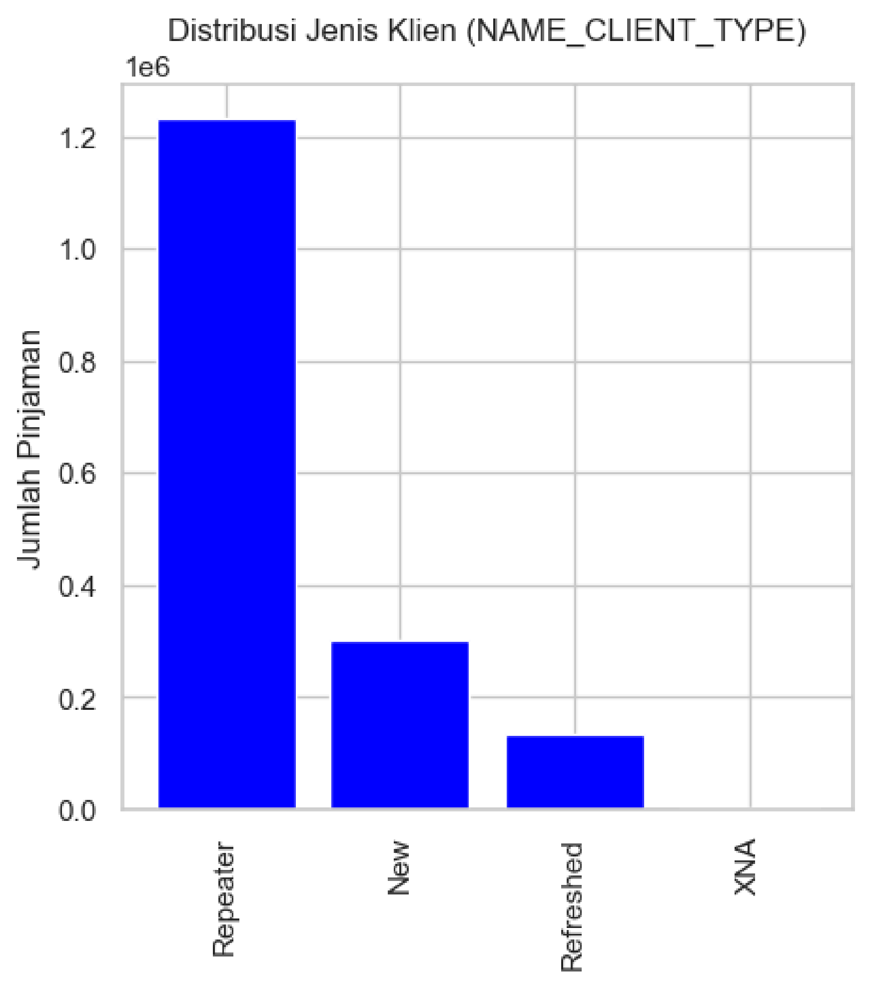
- The majority of customers are married.
- Customers with secondary or secondary special education are more likely to default on their loans.
- The predominant income type among customers is "Working" and more likely to default on their loans.

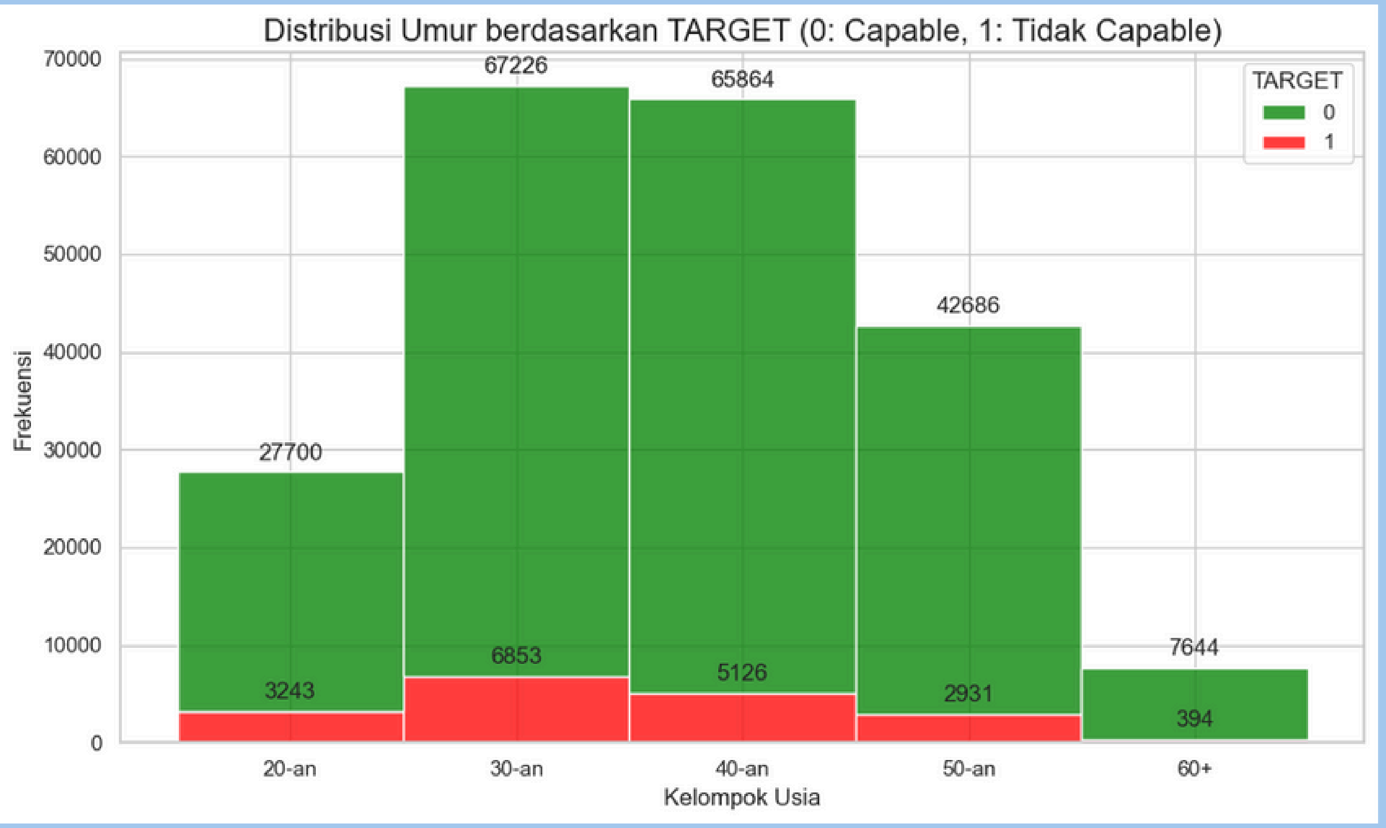
Exploratory Data Analysis

OCTAGRAM





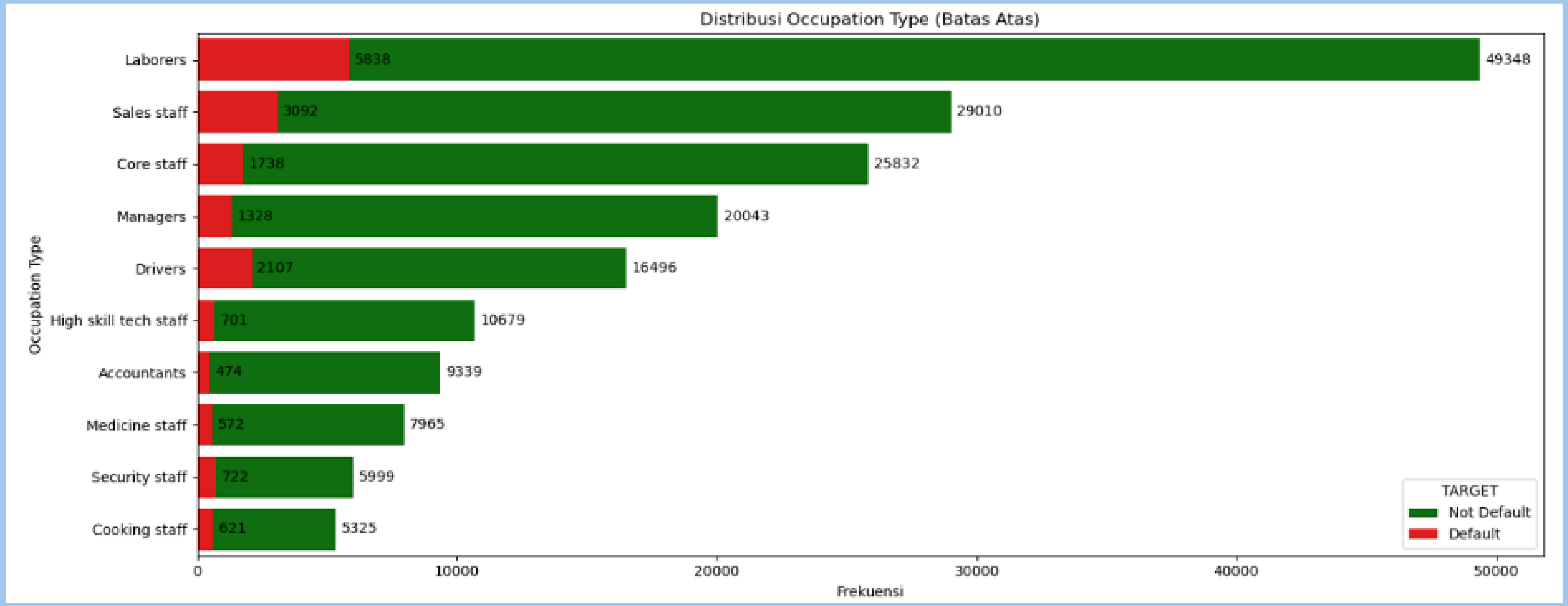


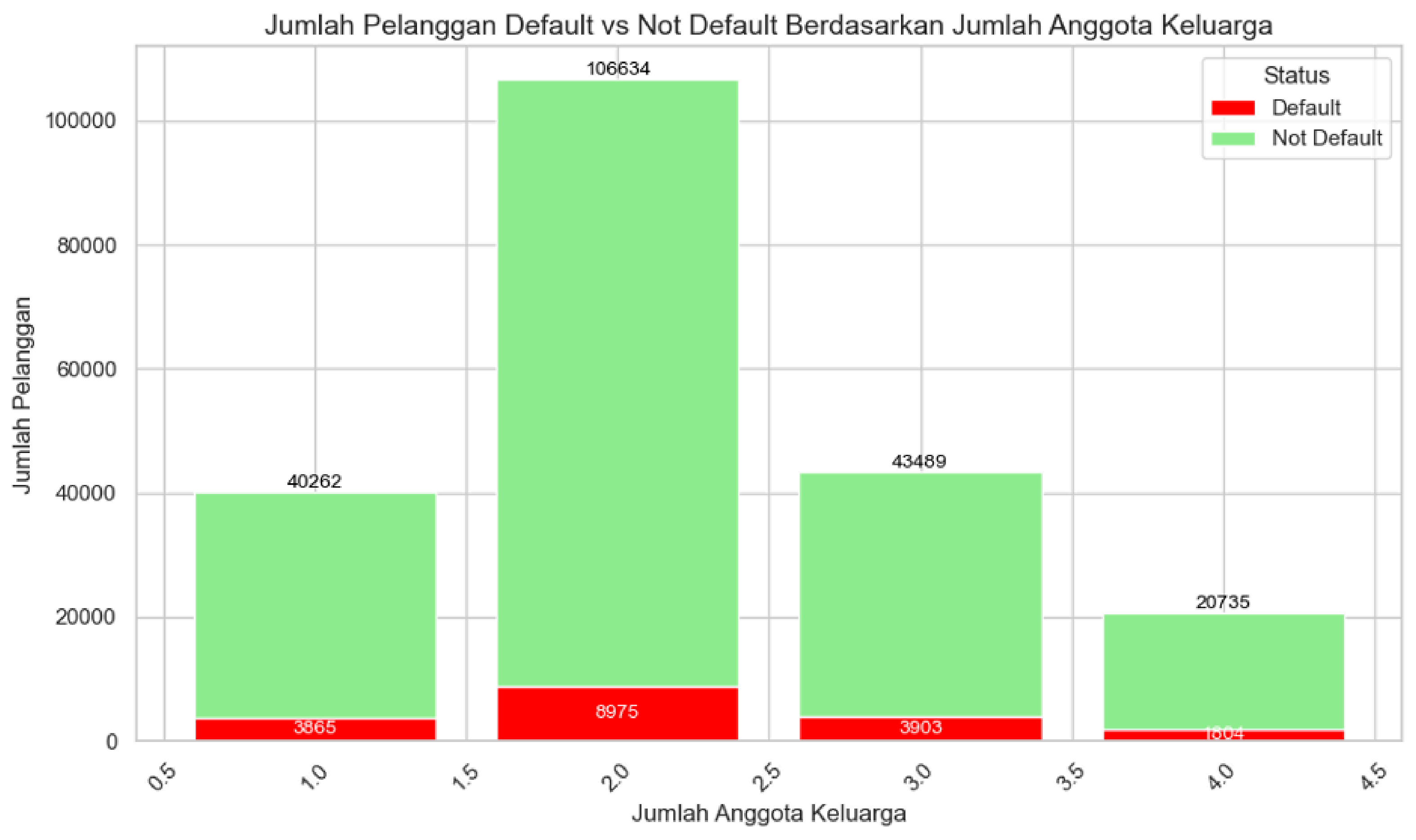


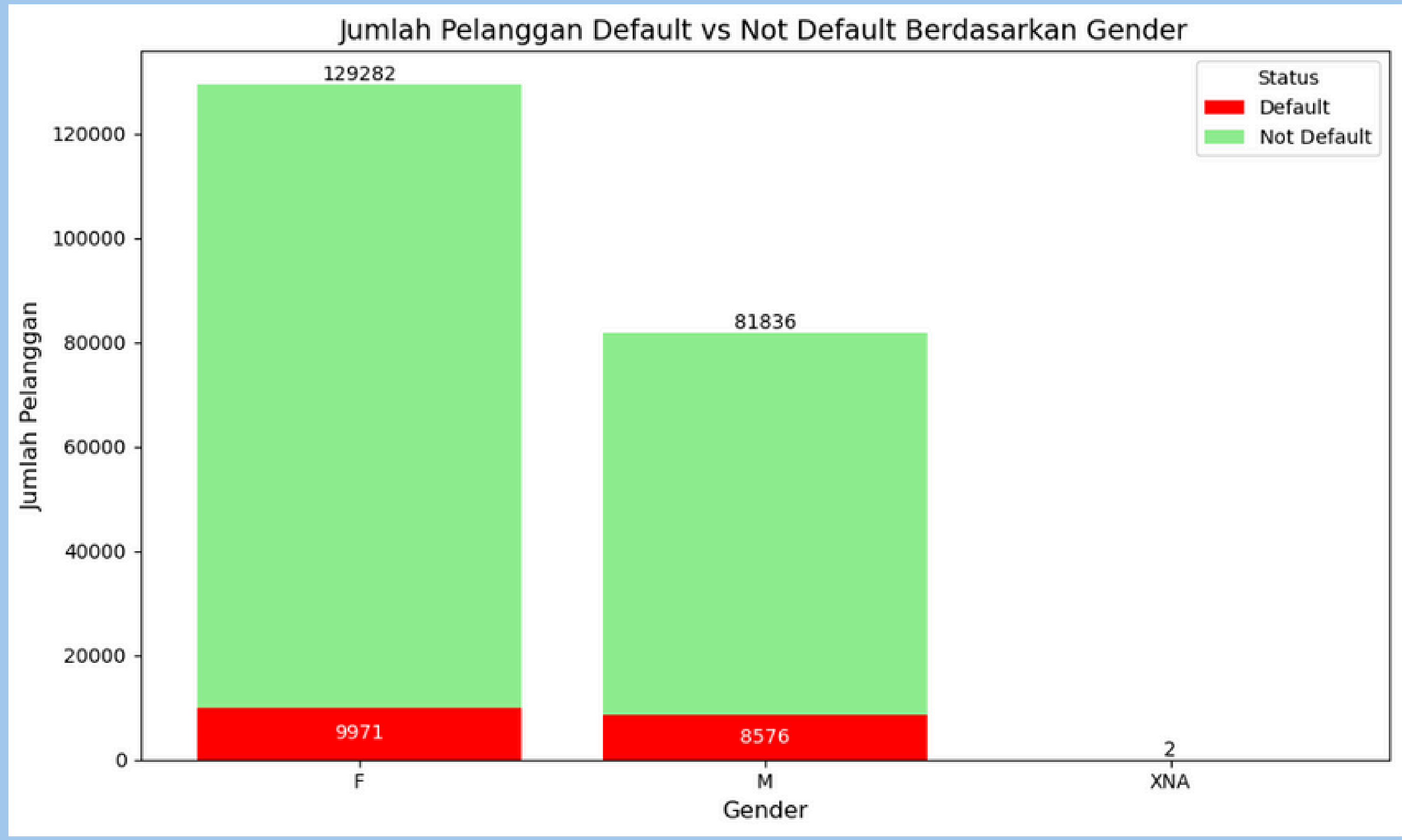
- Customers are more likely to utilize cash loans rather than revolving loans, and on average, both types are susceptible to default.
- Customers aged 20 to 39 years are particularly vulnerable to default.
- The majority of customers who take loans from Home Credit own a house or apartment.
- Repeat customers tend to borrow larger amounts.

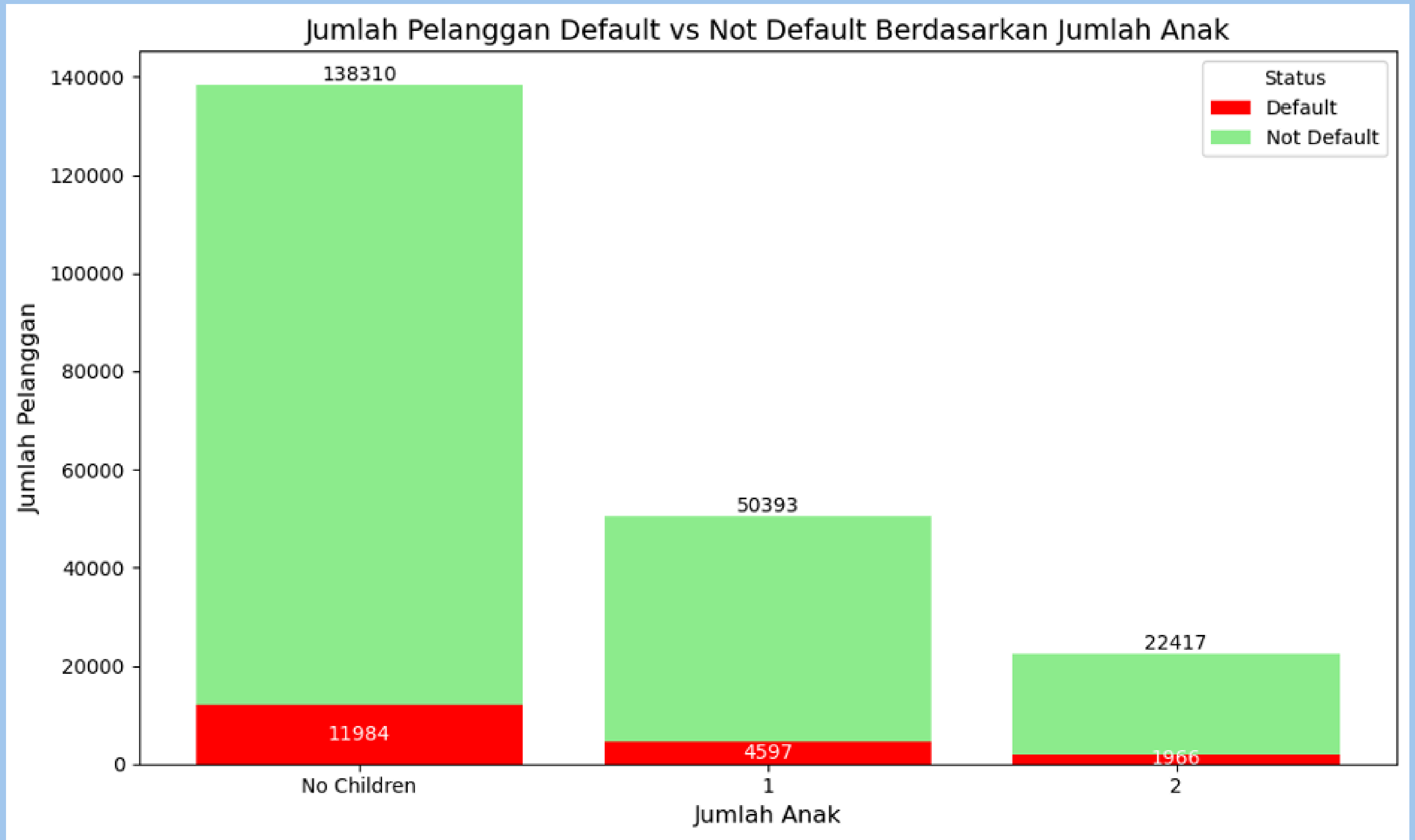
Exploratory Data Analysis

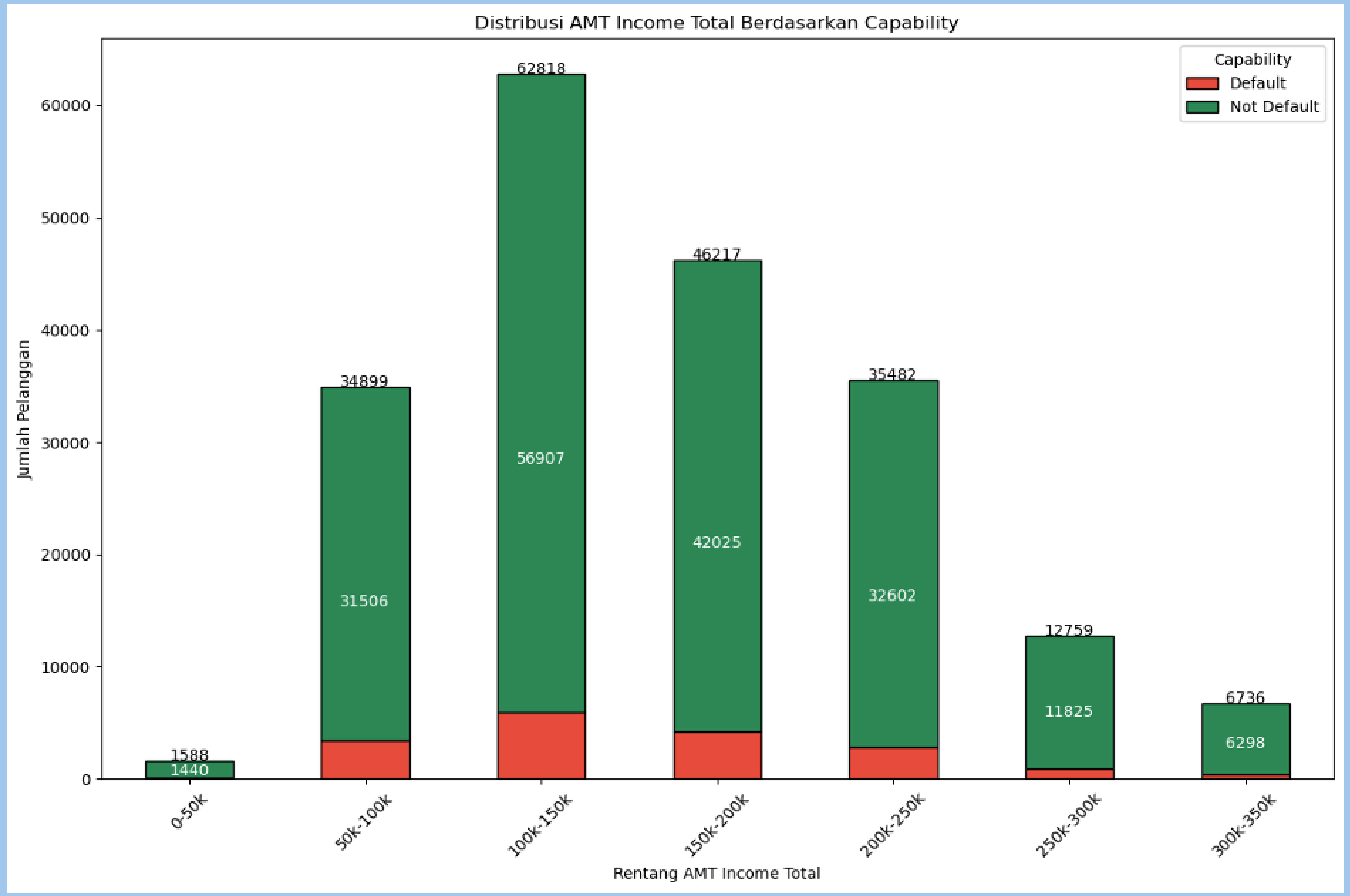
OCTAGRAM



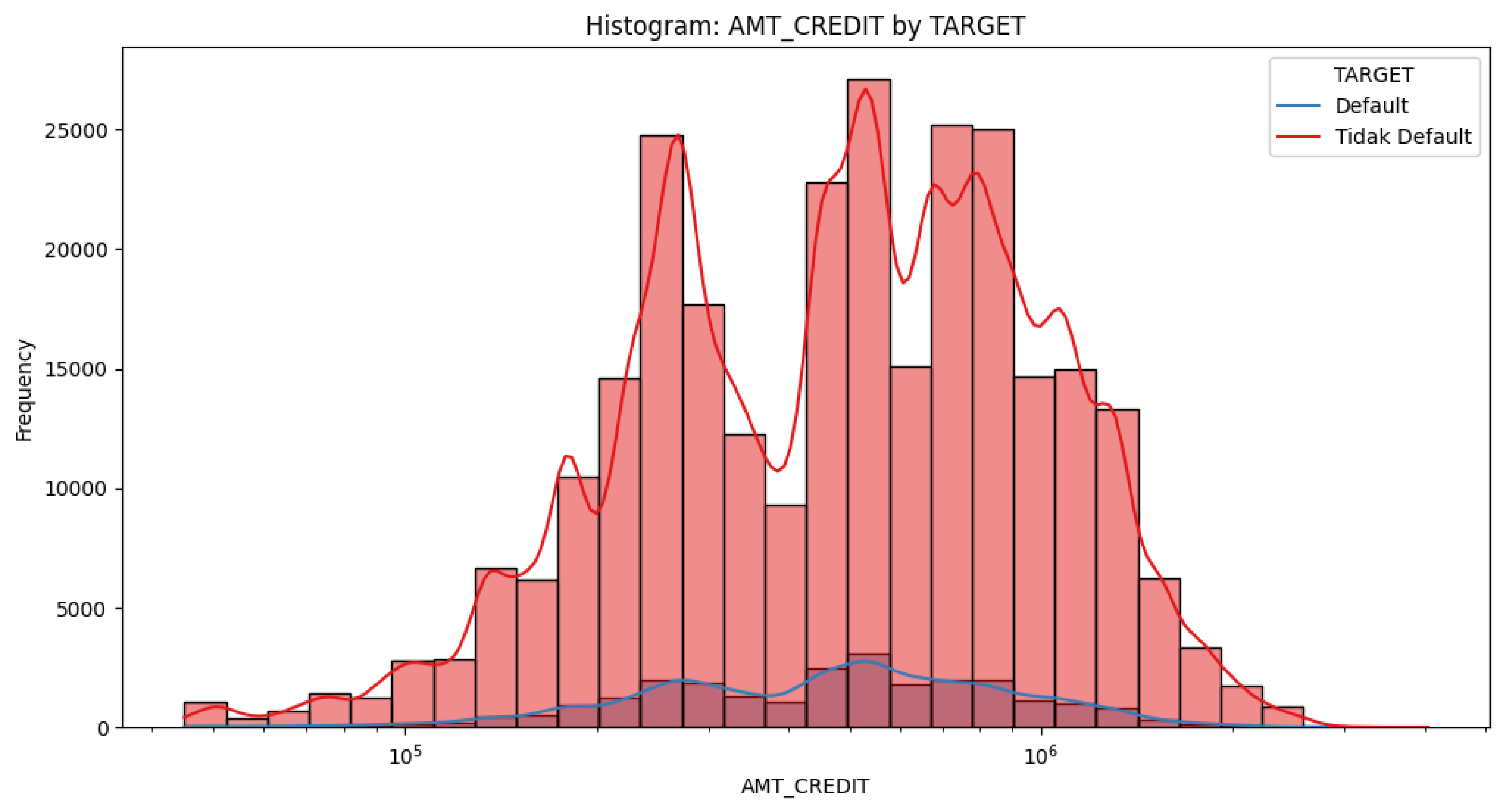




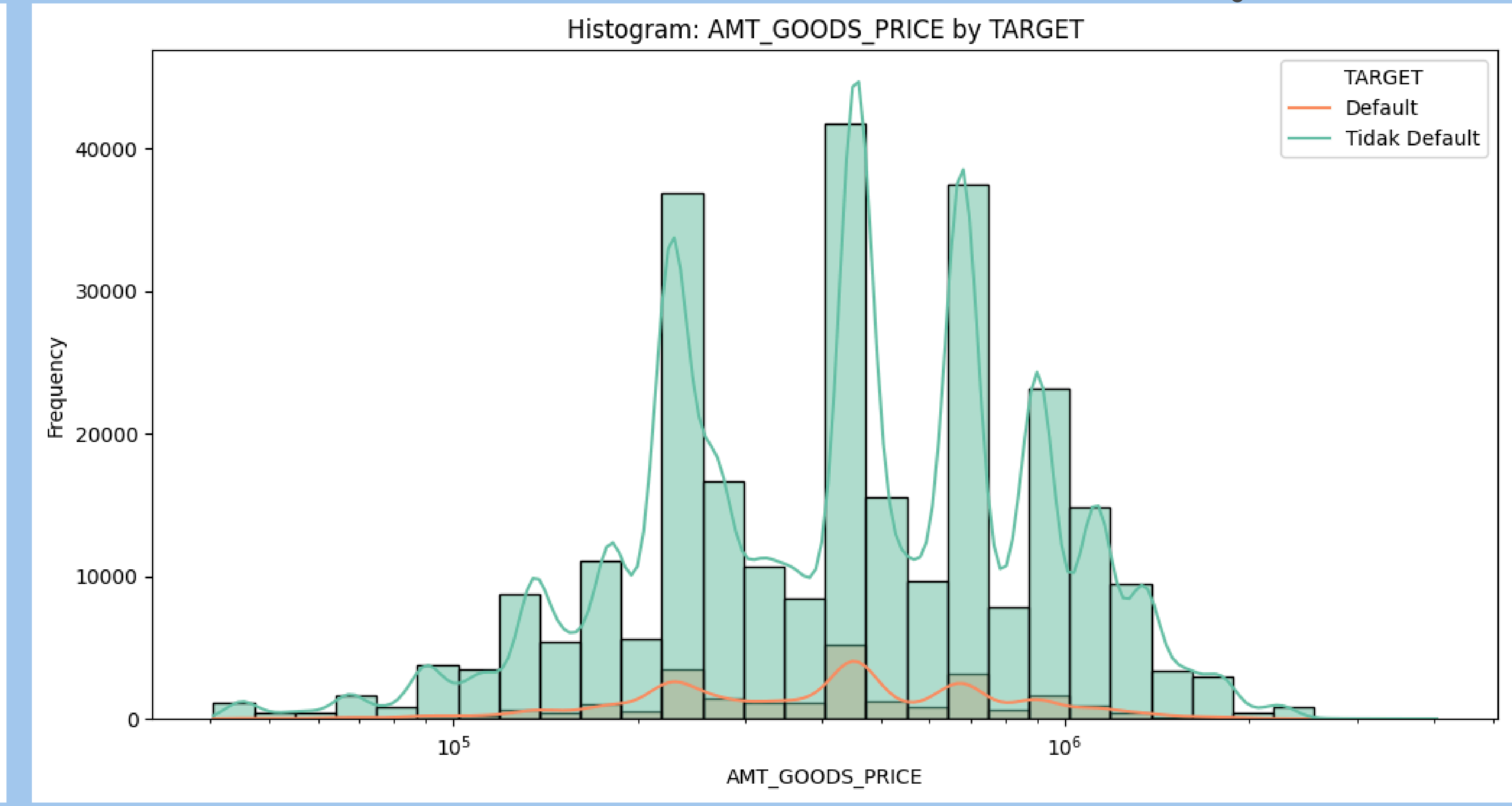

Jumlah Pelanggan Default vs Not Default Berdasarkan Jumlah Anak
Status
Default
Not Default
Jumlah Pelanggan
0
20000
40000
60000
80000
100000
120000
140000
138310
11984
50393
4597
22417
1966
No Children
1
2
Jumlah Anak


Distribusi AMT Income Total Berdasarkan Capability
Capability
Default
Not Default
Jumlah Pelanggan
0
10000
20000
30000
40000
50000
60000
1588
1440
34899
31506
62818
56907
46217
42025
35482
32602
12759
11825
6736
6298
0-50k
50k-100k
100k-150k
150k-200k
200k-250k
250k-300k
300k-350k
Rentang AMT Income Total

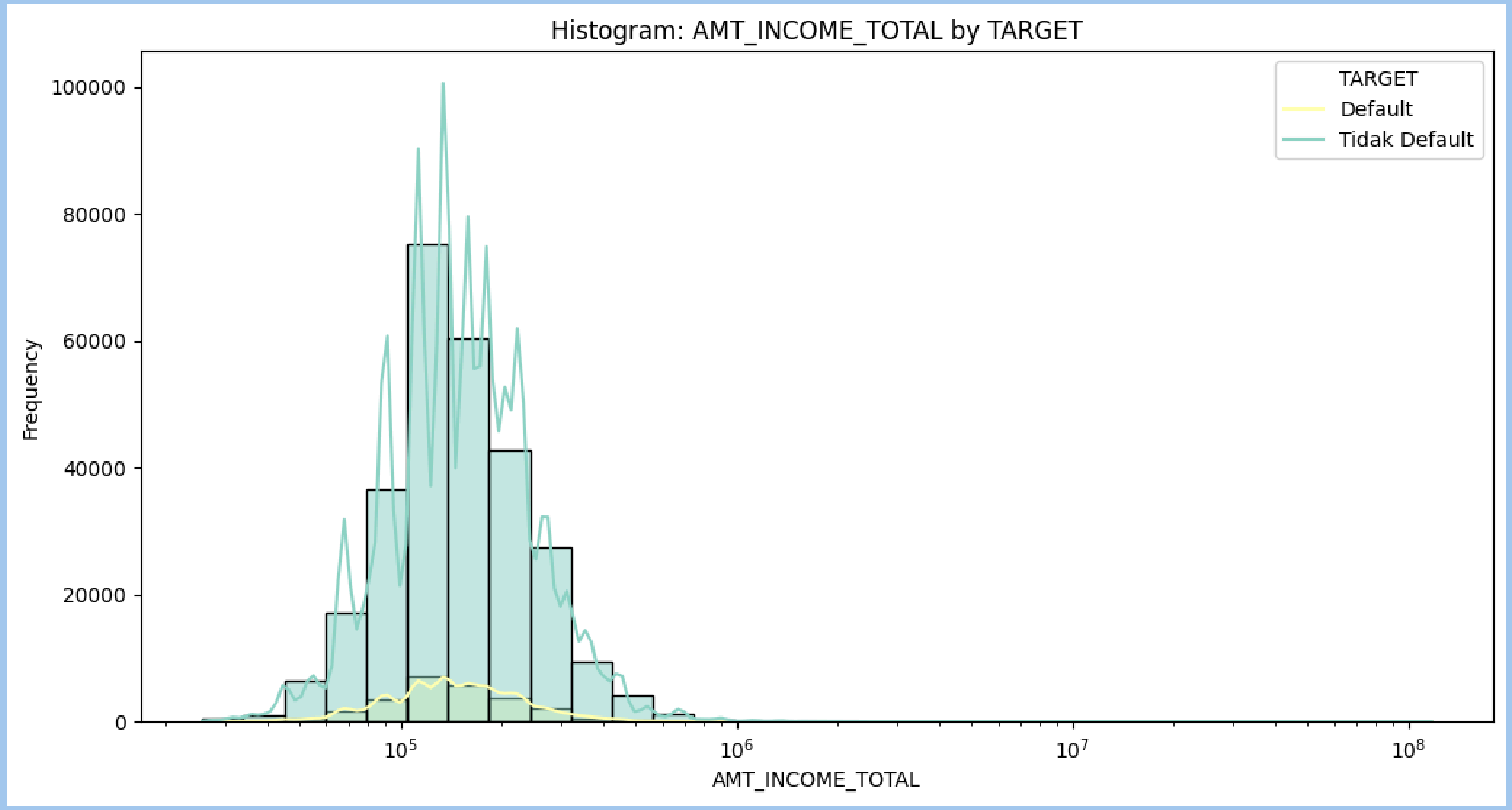
OCTAGRAM



Histogram: AMT_CREDIT by TARGET
TARGET
Default
Tidak Default
Frequency
AMT_CREDIT



Histogram: AMT_GOODS_PRICE by TARGET
TARGET
Default
Tidak Default
Frequency
AMT_GOODS_PRICE



Histogram: AMT_INCOME_TOTAL by TARGET
TARGET
Default
Tidak Default
Frequency
AMT_INCOME_TOTAL

Data Preprocessing

How we cleaning the data

1. Handling Null value

- Null Percentage over 40%: Drop the features.
- Null Percentage around 20% - 40%: Fill null value with median or mode.
- Null Percentage under 20%: Drop null value (rows), or filling null value with median or mode.

2. Handling Duplicated Data

- Keep the first data and delete the duplicated data.

3. Handling Outliers

- No handling outliers.

4. Merging Data

- Drop categorical features.
- Aggregate every dataset on SK_ID_CURR and merge all dataset to Application_{train | test}.

Data Preprocessing

How we cleaning the data

5. **Feature Transformation**

- Applying the Yeo-Johnson method.
- Applying Standardization.

6. **Feature Encoding & Engineering**

- Applying label encoding for distinct and ordinal value.
- Application_{train | test}: DEBT_TO_INCOME, ANNUITY_TO_INCOME, AGE, NUM_DOCUMENTS.
- Bureau & Bureau Balance: DAYS_DURATION_CREDIT.
- Installment: DAYS_LATE, FLAG_LATE_PAYMENT, FLAG_UNDERPAYMENT.
- Credit Card Balance: LIMIT_RATIO, TOTAL_DRAWINGS, TOTAL_DRAWINGS_COUNT.
- Pos Cash Balance: PAYMENT_PROGRESS_RATIO, REMAINING_PAYMENT.

Data Preprocessing

How we cleaning the data

8. Train Test Split

- Train test split on df_train.
- Test_size: 0.2 and random state: 42.
- y is label ''TARGET''.
- Train test split into:
  - X_train, X_val, y_train, y_val.
- X_test is for testing the choosen model.

```
X_train dimension:  (246008, 339)
y_train dimension:  (246008,)
X_val dimension:  (61503, 339)
y_val dimension:  (61503,)
X_test dimension:  (48744, 339)
```

Image 1.4 - Variable train, val, and test

9. Handling Imbalance Class

- Handling imbalance class using "Synthetic Minority Over-sampling Technique"

```
TARGET
0    226132
1     19876
Name: count, dtype: int64
```

Image 1.5 - Imbalance Class of "TARGET"

Data Preprocessing

How we cleaning the data

8. Feature Selection

- There are 339 Features used in model.
- Using SelectKBest for feature reduction.
- Features used in this model are 339 features because all features might be important.

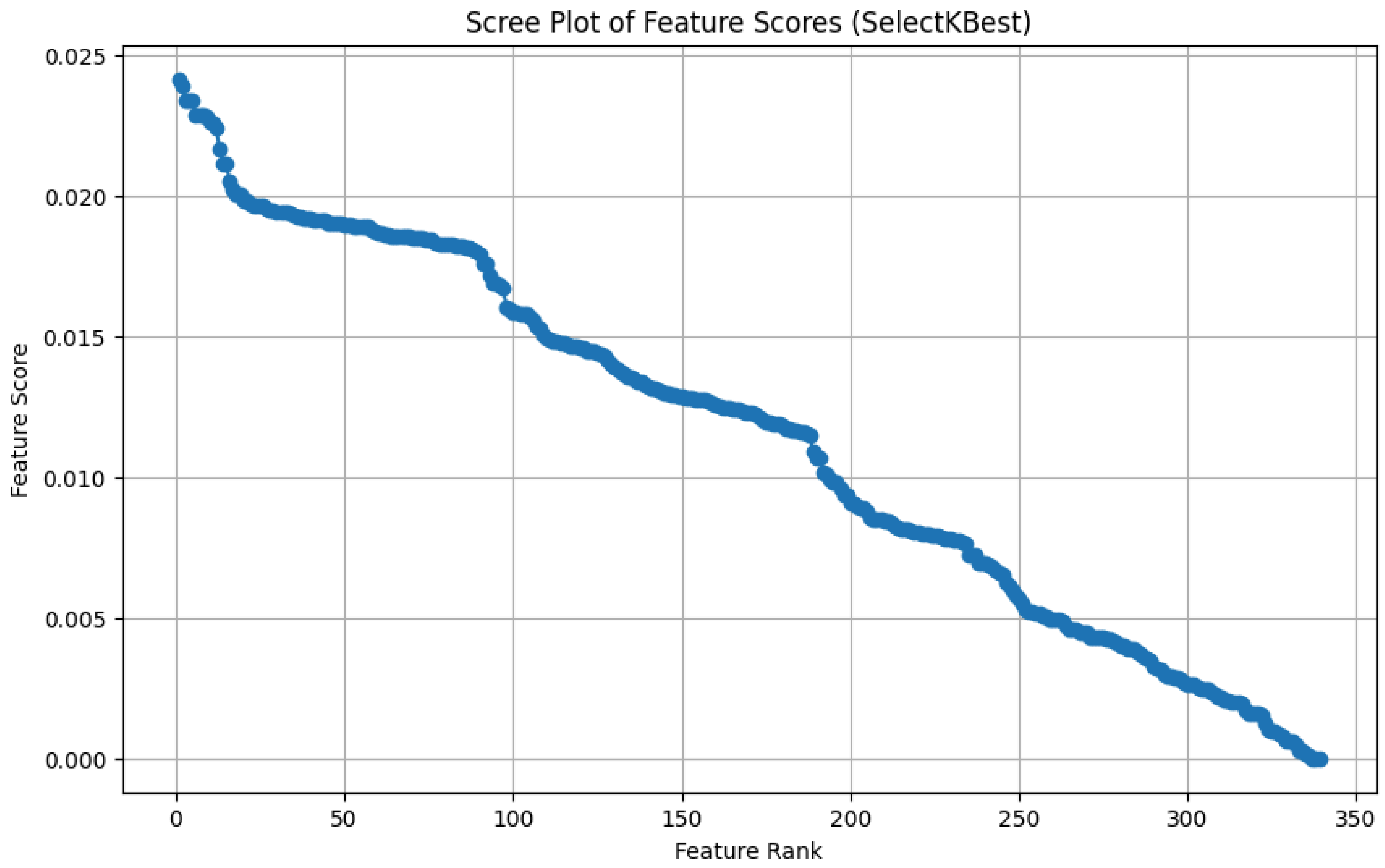


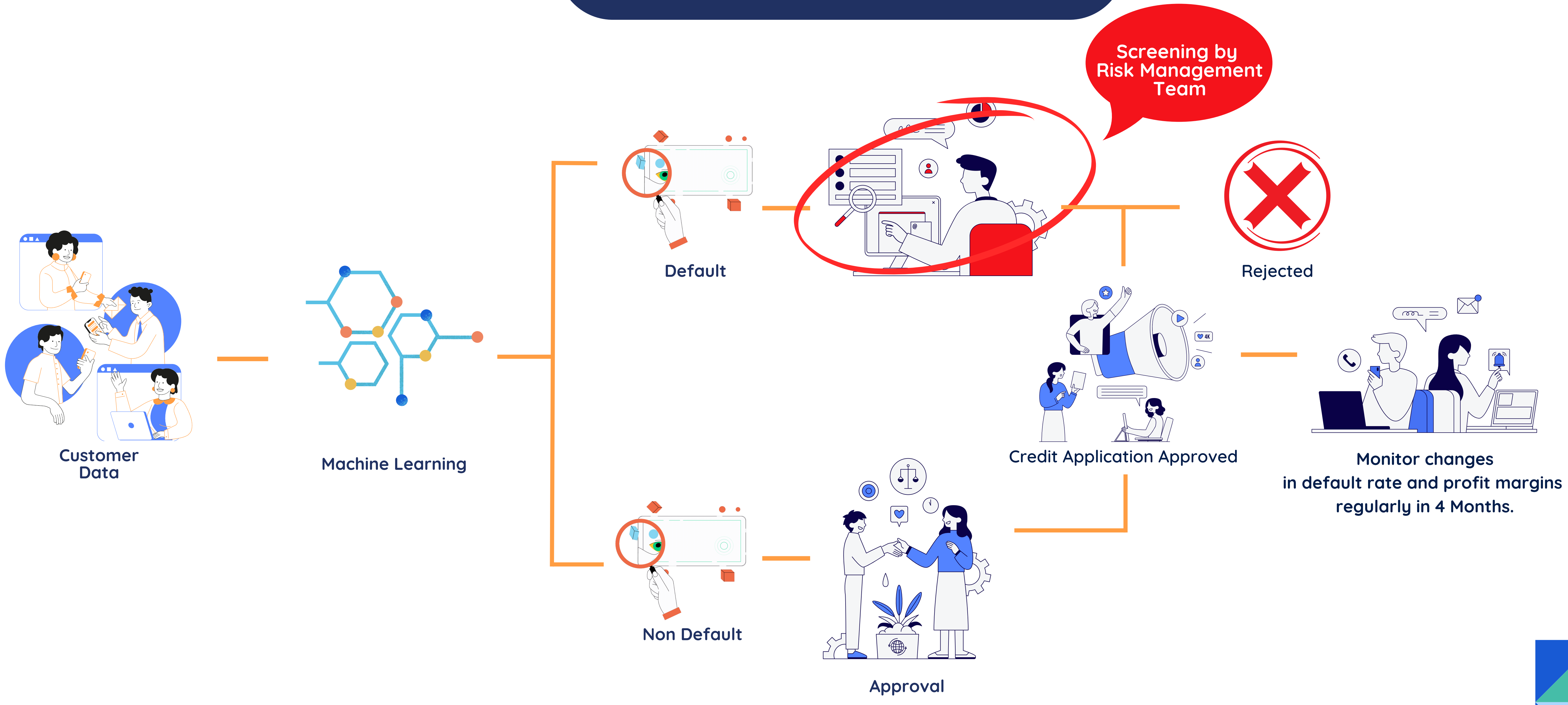
Image 1.6 - Scree Plot of Feature Scores (SelectKBest)

OCTAGRAM

OCTAGRAM

Business Recomendation

WORKFLOW



Simulasi Perhitungan Kerugian dalam Penerapan Machine Learning

OCTAGRAM

- **Rata-Rata Kredit: $596.564**
- **ROC AUC: 0.74**
- **Total Kreditur: 61.503**
- **Kreditur gagal bayar : 4.949**

kerugian per orang : $\dfrac{\textit{Rata - rata kredit}}{\textit{Kreditur gagal bayar}}$

Kerugian per orang

Before ML: $\dfrac{596.564}{4.949} = \$120.54$

After ML: $\dfrac{596.564}{(61.503 \times 1 - 0.74)} = \$37.31$

Kerugian per total data

Before ML: $\dfrac{120.54}{61.503} = \$7.413.715$

After ML: $\dfrac{37.31}{61.503} = \$2.294.476$

Persentase penurunan kerugian

69%

Risk Based Pricing

OCTAGRAM

- Misal customer A ingin meminjam uang Rp 100.000.000
- Riwayat Kredit: Baik (skor kredit 700)
- Pendapatan Bulanan: Rp10.000.000
- Total Utang yang Dimiliki: Rp30.000.000
- Rasio Utang terhadap Pendapatan (Debt-to-Income Ratio): 30% (Rp30.000.000 / Rp10.000.000)
- Klasifikasi Risiko: Sedang

- Biaya Modal: Misalkan biaya modal bank untuk mendapatkan dana adalah 4%.
- Margin Keuntungan: Bank ingin mendapatkan margin keuntungan sebesar 2%.
- Biaya Risiko: Berdasarkan analisis risiko, bank memperkirakan biaya risiko untuk peminjam dengan klasifikasi risiko sedang adalah 1,5%.

Suku Bunga=Biaya Modal+Margin Keuntungan+Biaya Risiko

Suku Bunga = 4%+2%+1,5%

Suku Bunga = 7,5%