Loan Prediction Based on Customer Behavior

Final Project Studi Independen Batch 5 Rakamin Academy

TEAM MEMBER

Team Leader

Wahyu Afriza

Hanif Zaki

M Fahrurrozi

M Rofi'i

Alyani Noor

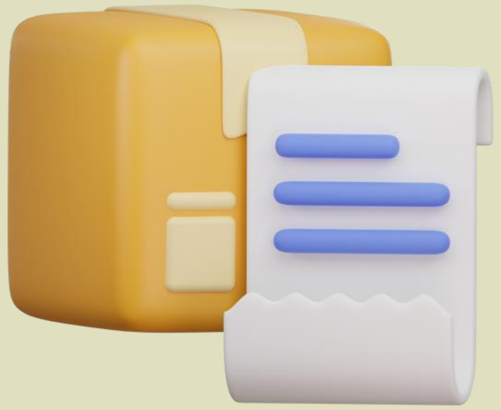
Bilqis Nafida

Windy Agelina

Project Overview

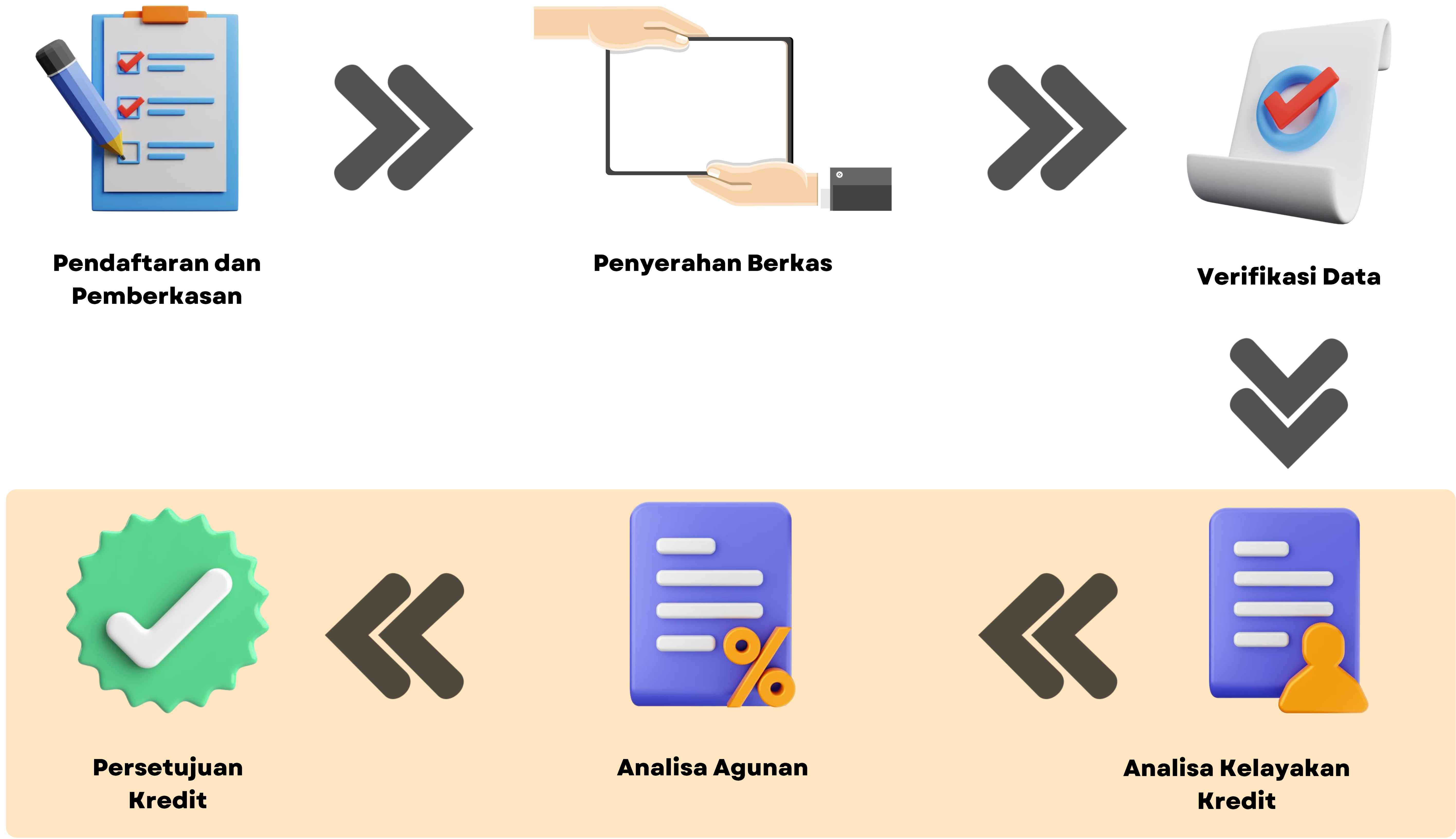
Product Team → **Loan Risk Analyst**

Risk Analyst

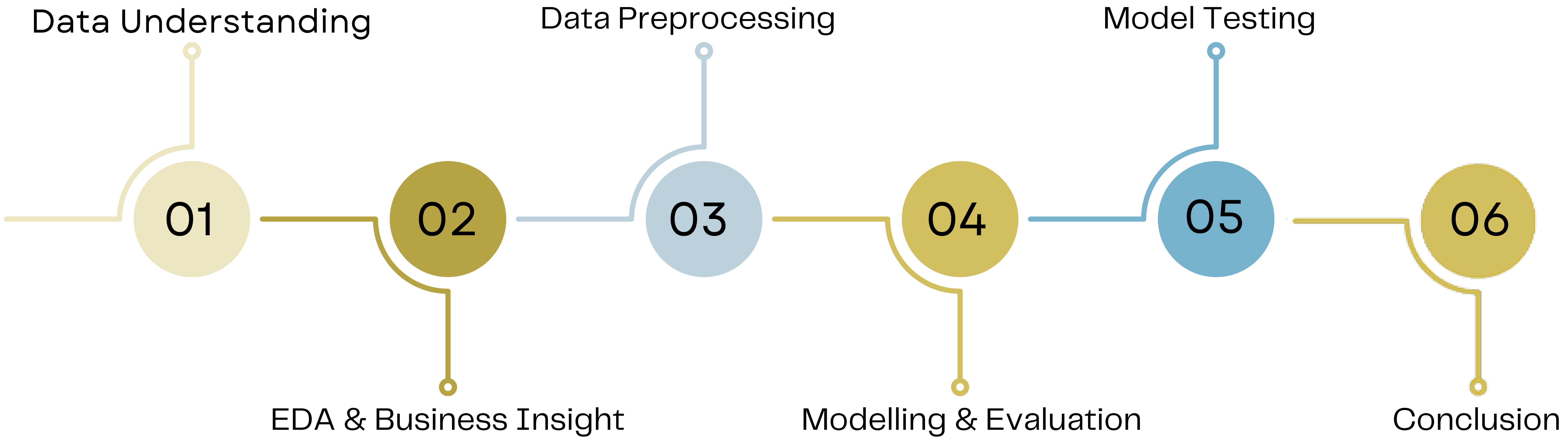


Customer Behavior » **Loan Prediction**

Business Process



Workflow


Data Understanding
01
02
EDA & Business Insight
Data Preprocessing
03
04
Modelling & Evaluation
Model Testing
05
06
Conclusion

Data Understanding

Dataset

Kolom 13

Baris 252.000

Tipe Data

Categorical

- House_Ownership
- Car_Ownership
- Profession
- CITY
- STATE
- Married/Single
- Risk Flag

**Targe
t**

Numerical

- ID
- Age
- Income
- Experience
- CURRENT_JOB_YRS
- CURRENT_HOUSE_YRS

Problem

Pengajuan pinjaman semakin banyak namun Loan Approval dilakukan secara manual sehingga membutuhkan waktu yang cukup lama.

Loan Approval dalam waktu yang lama secara manual membutuhkan banyak tenaga kerja sehingga biaya dan tenaga kerja semakin tinggi

Goals

Mengurangi biaya, usaha, dan waktu proses Loan Approval melalui analisis customer behavior.

Objective

Memberikan rekomendasi loan approval berdasarkan customer behavior

Business Metrics

- Default rate
- Workforce cost

Exploratory Data Analysis
&
Business Insight

#0

FEATURE EXTRACTION

Pengelompokkan peminjam atau nasabah menjadi beberapa kategori berdasarkan berbagai faktor seperti usia, pendapatan, pekerjaan, dan lainnya memiliki tujuan utama untuk mengidentifikasi pola perilaku dan risiko unik di setiap segmen.

Income Group

Low : 10310 - 200000
Medium: 200000 - 1000000
High : 1000000 - dst

Current Job Group

Junior : 0-3
Mid-level : 3-7
Senior : 7- dst

#0

FEATURE EXTRACTION

Pengelompokkan peminjam atau nasabah menjadi beberapa kategori berdasarkan berbagai faktor seperti usia, pendapatan, pekerjaan, dan lainnya memiliki tujuan utama untuk mengidentifikasi pola perilaku dan risiko unik di setiap segmen.

City Group

Metro: Mumbai, Delhi_city, Kolkata, Bangalore, Chennai, Hyderabad, Pune

Urban: Ahmedabad, Gurgaon, Noida, Navi_Mumbai, Thane, Jaipur, Vadodara, Vijayawada

Suburban: Lucknow, Nagpur, Indore, Kochi, Bhopal, Patna, Visakhapatnam, Coimbatore

Town: Srinagar, Jodhpur, Amritsar, Kota, Ajmer, Bikaner, Mysore.

Rural: Gorakhpur, Warangal, Siliguri, Dhanbad, Farrukhabad, Haldia, Gangtok

Age Group

Millennials:	21 - 40
Gen X:	40 - 55
Baby Boomers:	55 - 80

#0

FEATURE EXTRACTION

Pengelompokkan peminjam atau nasabah menjadi beberapa kategori berdasarkan berbagai faktor seperti usia, pendapatan, pekerjaan, dan lainnya memiliki tujuan utama untuk mengidentifikasi pola perilaku dan risiko unik di setiap segmen.

Profession Group

Healthcare:	Physician, Psychologist, Dentist, Surgeon, dan Research
Research, Engineering & IT:	Statistician, Web_designer, Engineer, Computer_hardware_engineer, Drafter, Scientist, Industrial_Engineer, Mechanical_engineer, Chemical_engineer, Biomedical_Engineer, Petroleum_Engineer, Technology_specialist, Design_Engineer, Civil_engineer, Software_Developer, Computer_operator, Technical_writer, Graphic_Designer, Web_designer, Architect, Technician, Microbiologist, Geologist, Statistican, dan Surveyor.
Finance:	Financial_Analyst, Economist, Analyst, dan Chartered_Accountant
Excecutive:	Magistrate, Consultant, Official, Politican, dan Lawyer
Public Service:	Civil_servant, Police_officer, Army_officer, Hotel_Manager, Flight_attendant, Air_traffic_controller, Aviator, Firefighter, Chef, dan Librarian
Art & Entertainment:	Comedian, Artist, dan Fashion_Designer

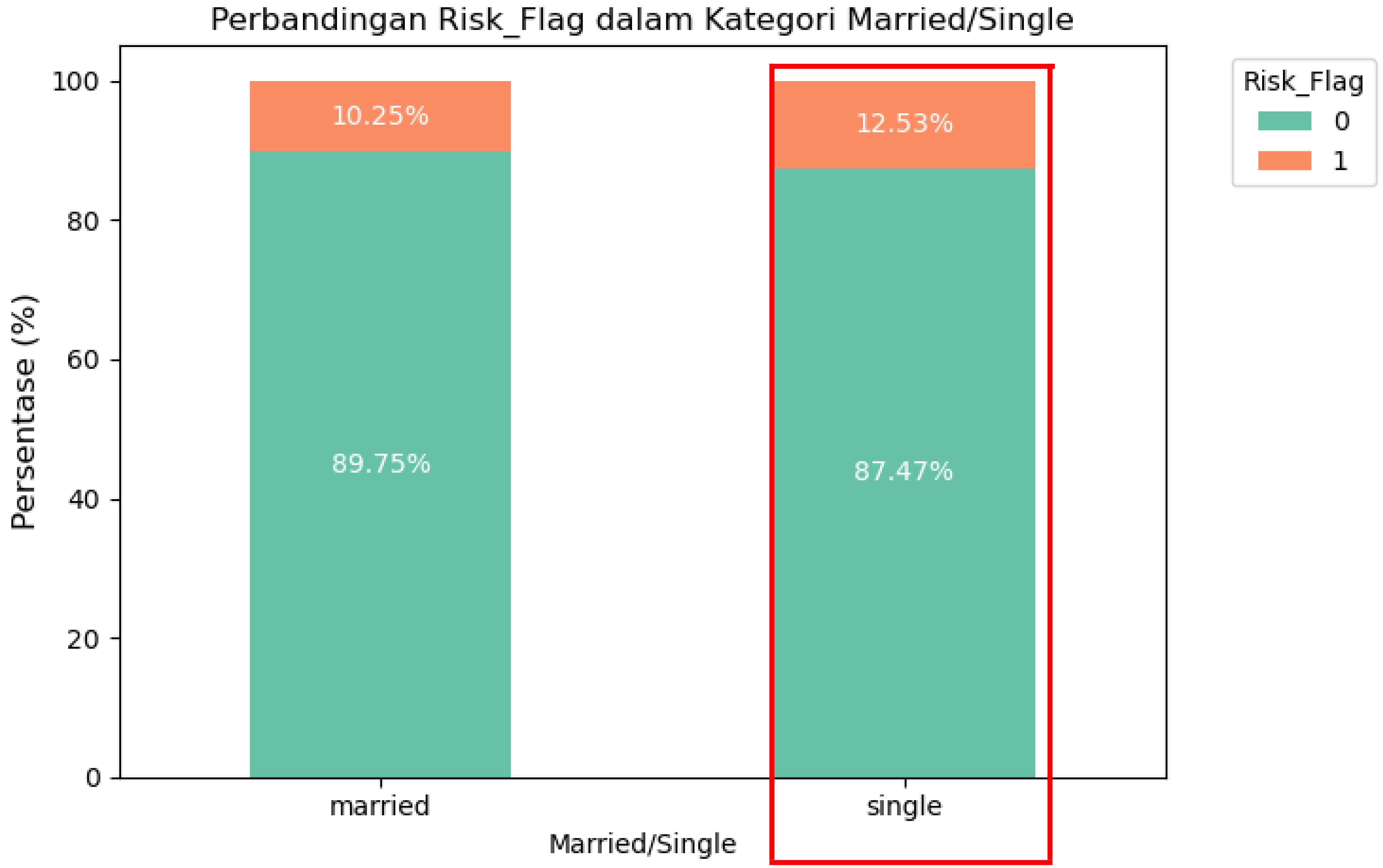
#1 STATUS PERNIKAHAN BERDASARKAN RISK FLAG

Insight

- Kelompok nasabah yang memiliki risiko gagal membayar pinjaman lebih tinggi adalah kelompok nasabah yang memiliki status single.

Potensial Masalah

Kurangnya rasa tanggung jawab



#2 PENDAPATAN BERDASRKAN RISK FLAG

Insight

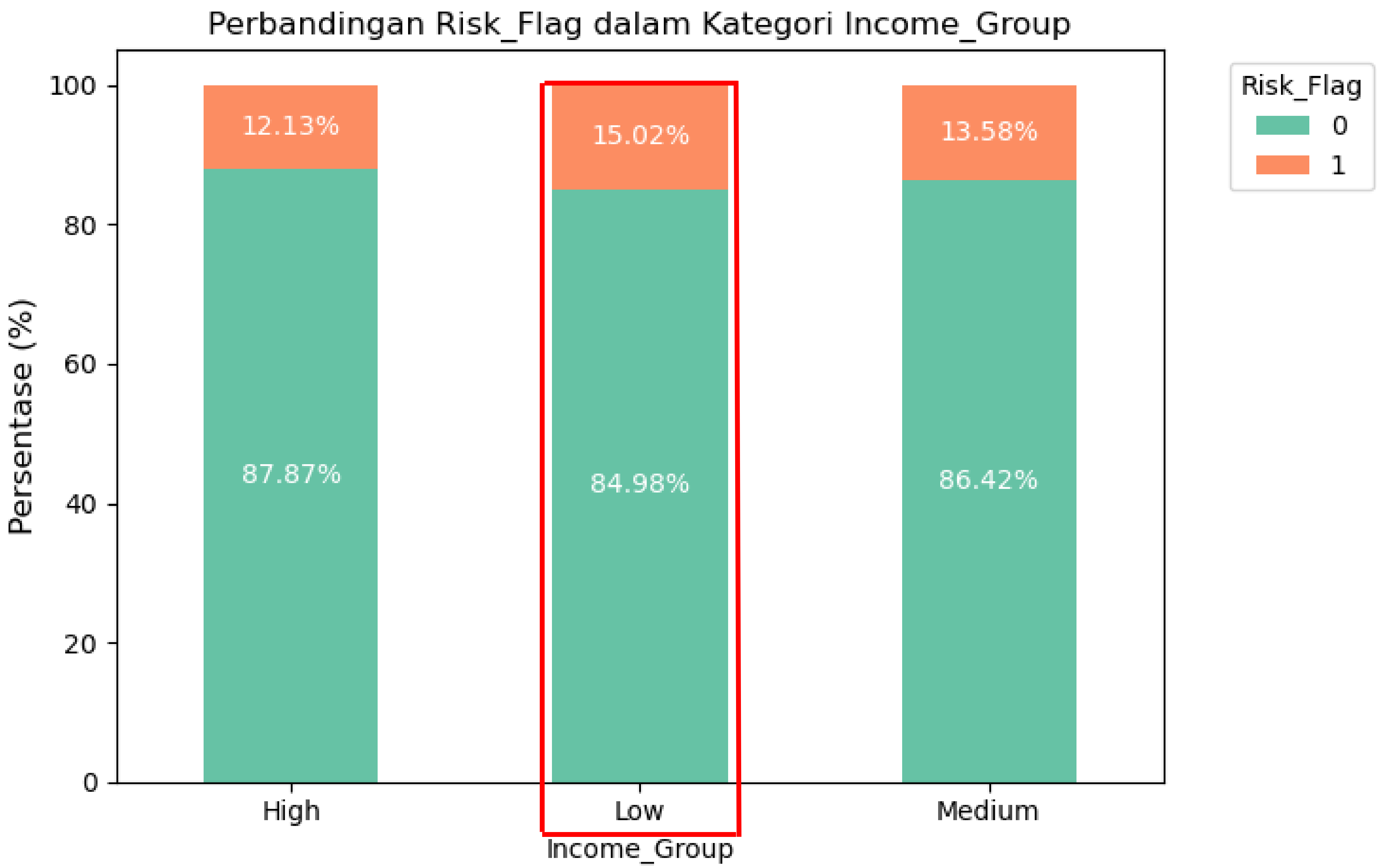
- Semakin rendah pendapatan nasabah maka kecenderungan berpotensi untuk gagal membayar pinjaman semakin tinggi.

Potensial Masalah

Kebutuhan biaya pribadi tinggi

Tekanan Finansial

Low:
Income 10310 - 200000



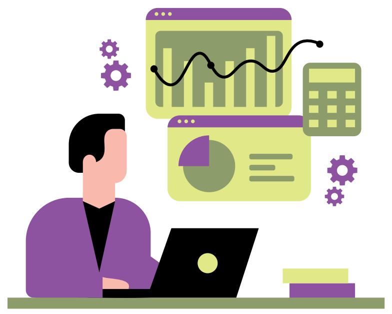
#3 CURRENT JOB YEARS BERDASARKAN RISK FLAG

Insight

- Nasabah yang memiliki pengalaman kerja yang lebih rendah cenderung berpotensi untuk gagal membayar pinjaman.

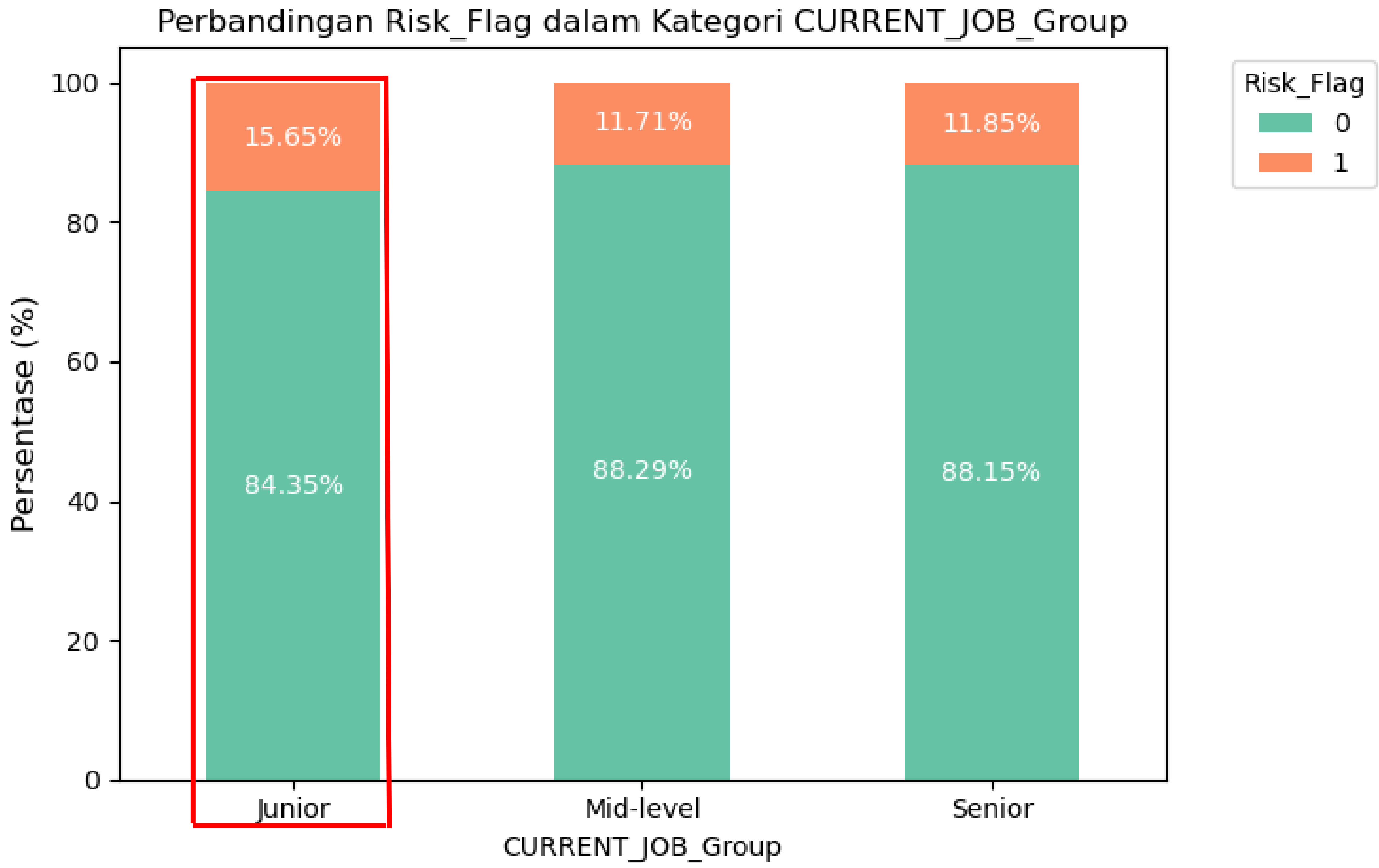
Potensial Masalah

Rendahnya Stabilitas keuangan



Tinggi nya kebutuhan pribadi

Junior:
Current job years 0 - 3



#4 GRUP PROFESI DENGAN TINGKAT RISIKO KREDIT TERATAS

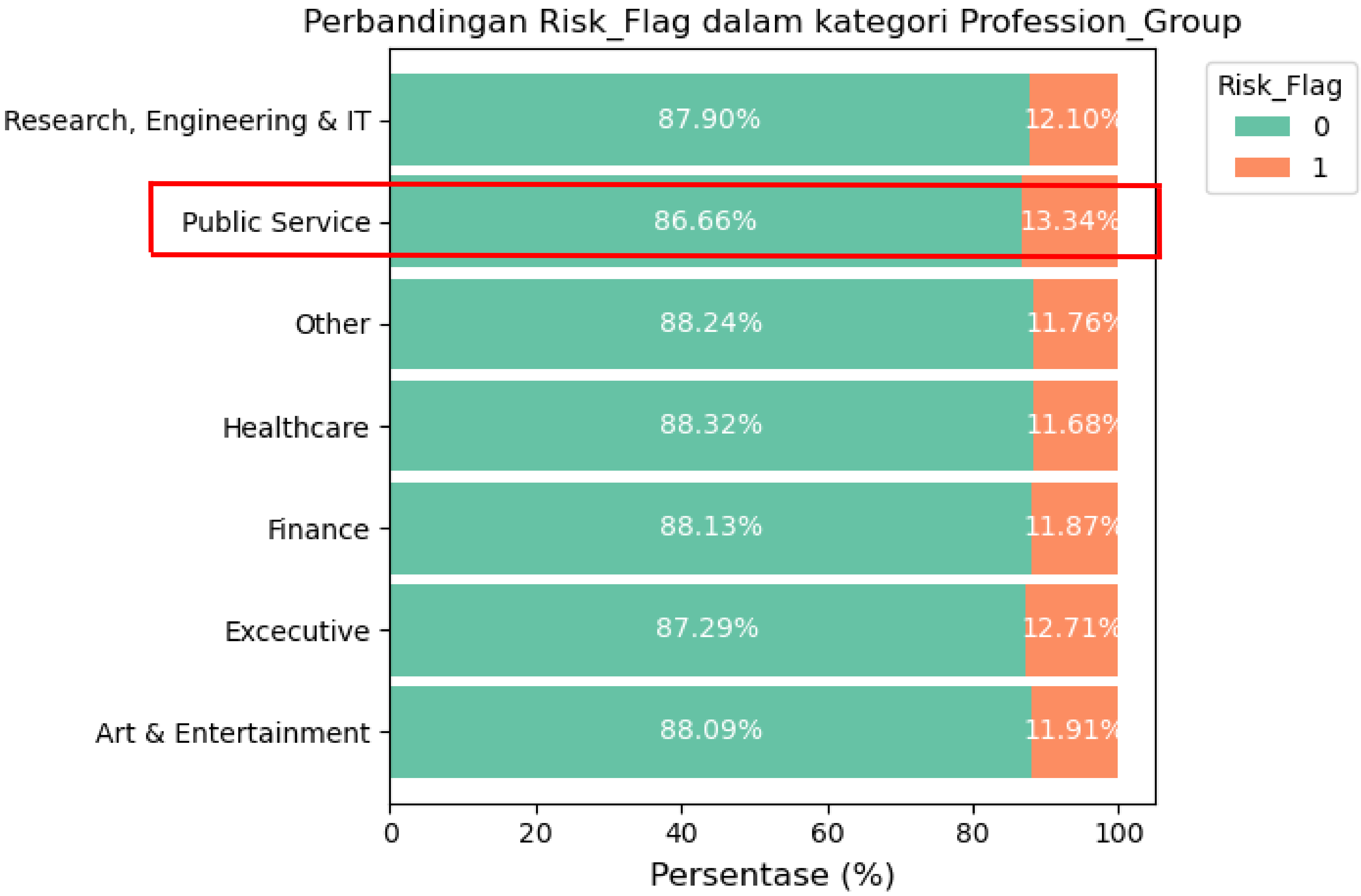
Insight

- Pelanggan dalam sektor **Pelayanan Publik** menunjukkan tingkat risiko teratas dalam pengajuan pinjaman.

Potensial Masalah

Gaji Rendah

Tanggungan Finansial Tinggi



Public Service:
Civil servant, Police officer, Army officer, Hotel Manager, Flight_attendant, Air_traffic_controller, Aviator, Firefighter, Chef, and Librarian

#5 TINGKAT RESIKO KREDIT BERDASARKAN KARAKTERISTIK KOTA

Insight

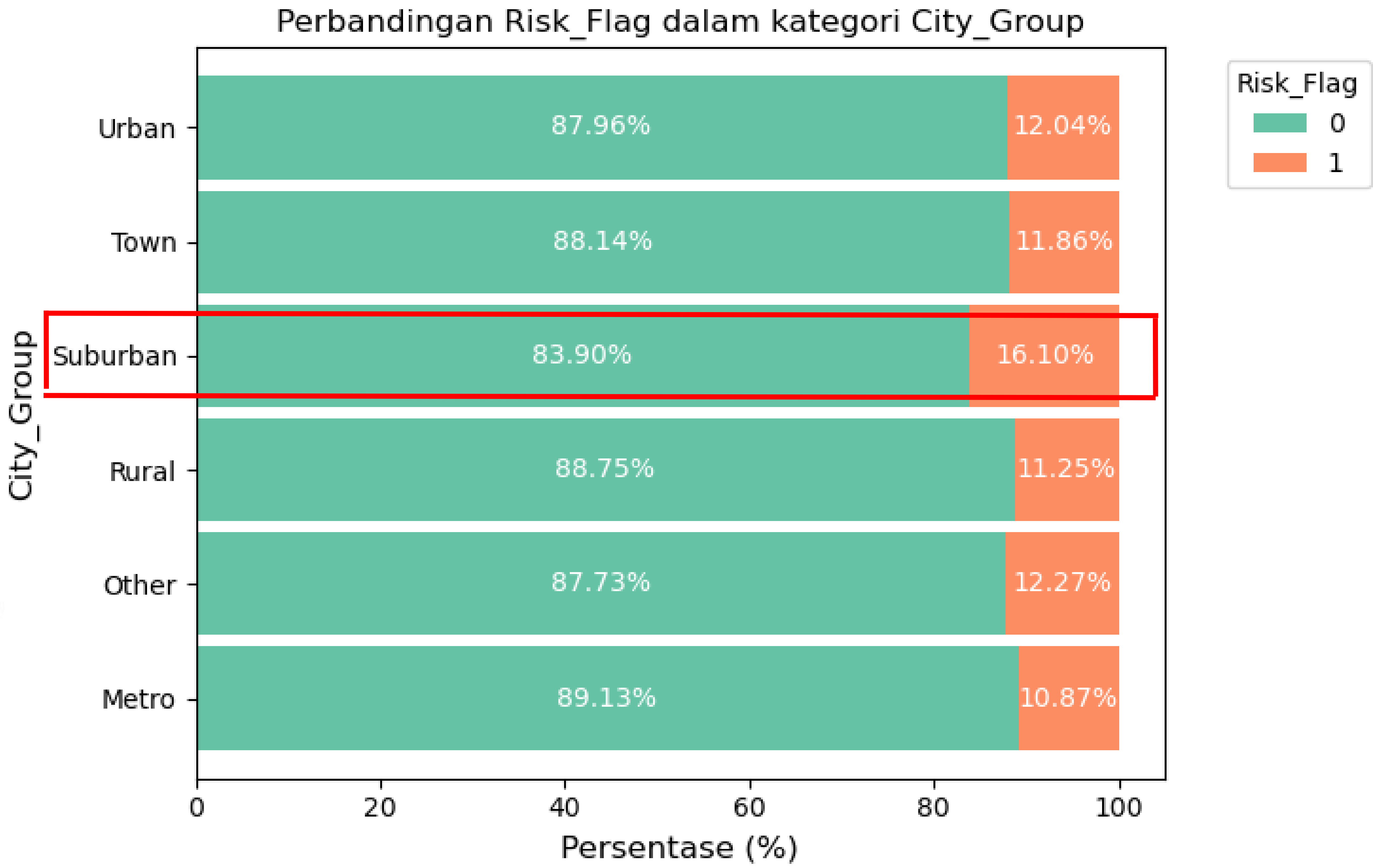
- Meskipun berada di antara kota dan pemukiman, peminjam **Suburban** menunjukkan risiko yang lebih tinggi daripada yang diharapkan.

Potensial Masalah

Pendapatan

Ketersediaan Pekerjaan

Biaya Hidup



Suburban:
Lucknow, Nagpur, Indore, Kochi, Bhopal, Patna, Visakhapatnam, dan Coimbatore

#6 TINGKAT RESIKO KREDIT BERDASARKAN GENERASI

Insight

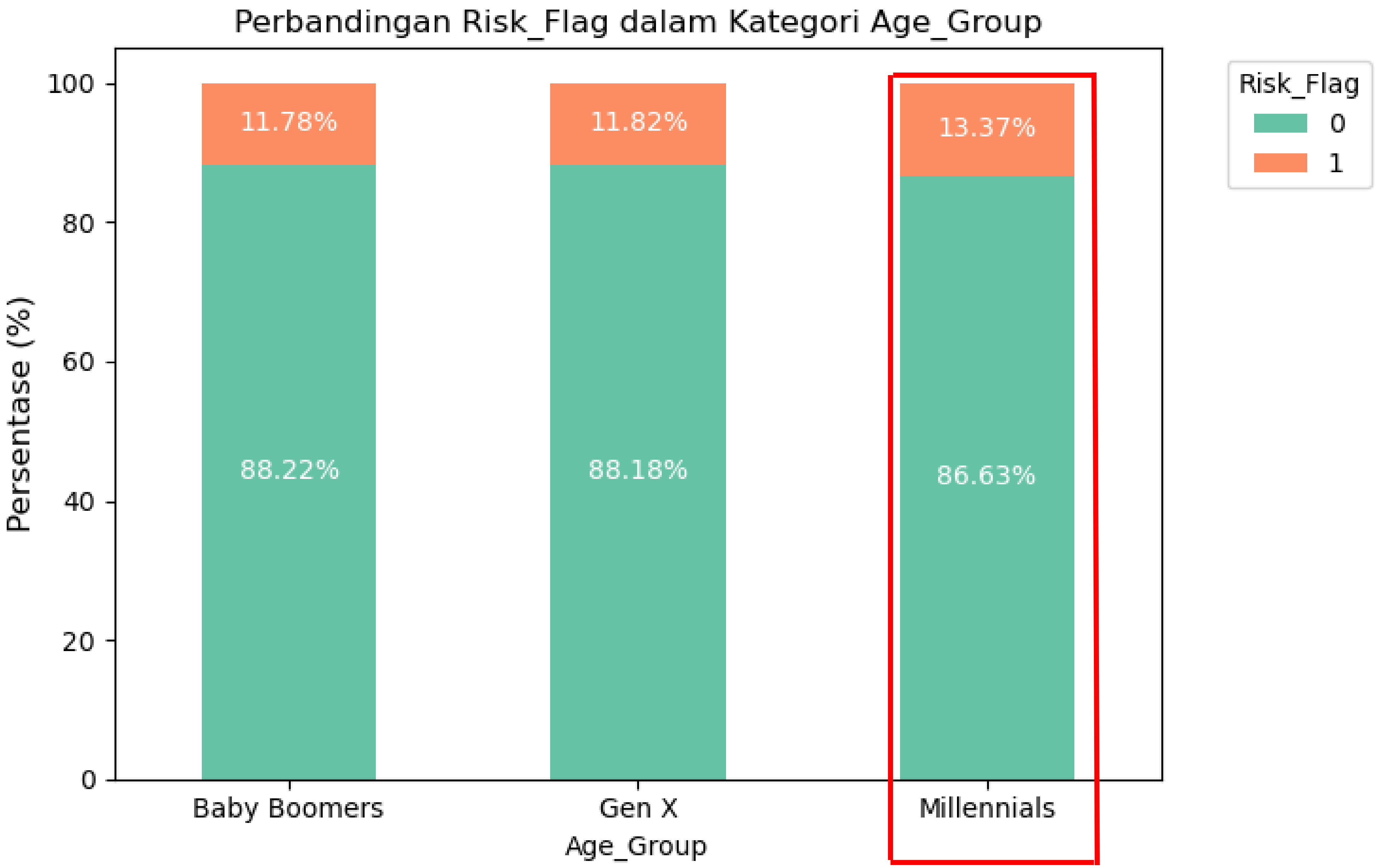
- Secara umum, terlihat bahwa risiko peminjam cenderung menurun seiring dengan peningkatan usia generasi.

Potensial Masalah

Konsumtif

Persaingan
Dunia Kerja

Millennials:
Rentang usia 21 - 40 tahun



Data Preprocessing

Data PreProcessing

Data Cleaning

Handle Duplicate Data

- Menghapus kolom id
- Drop data duplicate

Handle Outliers

- Dilakukan pada Kolom Numerical yaitu:
- Tidak Terdapat Outliers

'Income' ,
'Age'
,'Experience',
'Current_House_Yrs'
'Current_Job_Yrs'

Data PreProcessing

Data Transformation

Feature Transformation

Normalisasi pada kolom 'Current_House_Yrs'

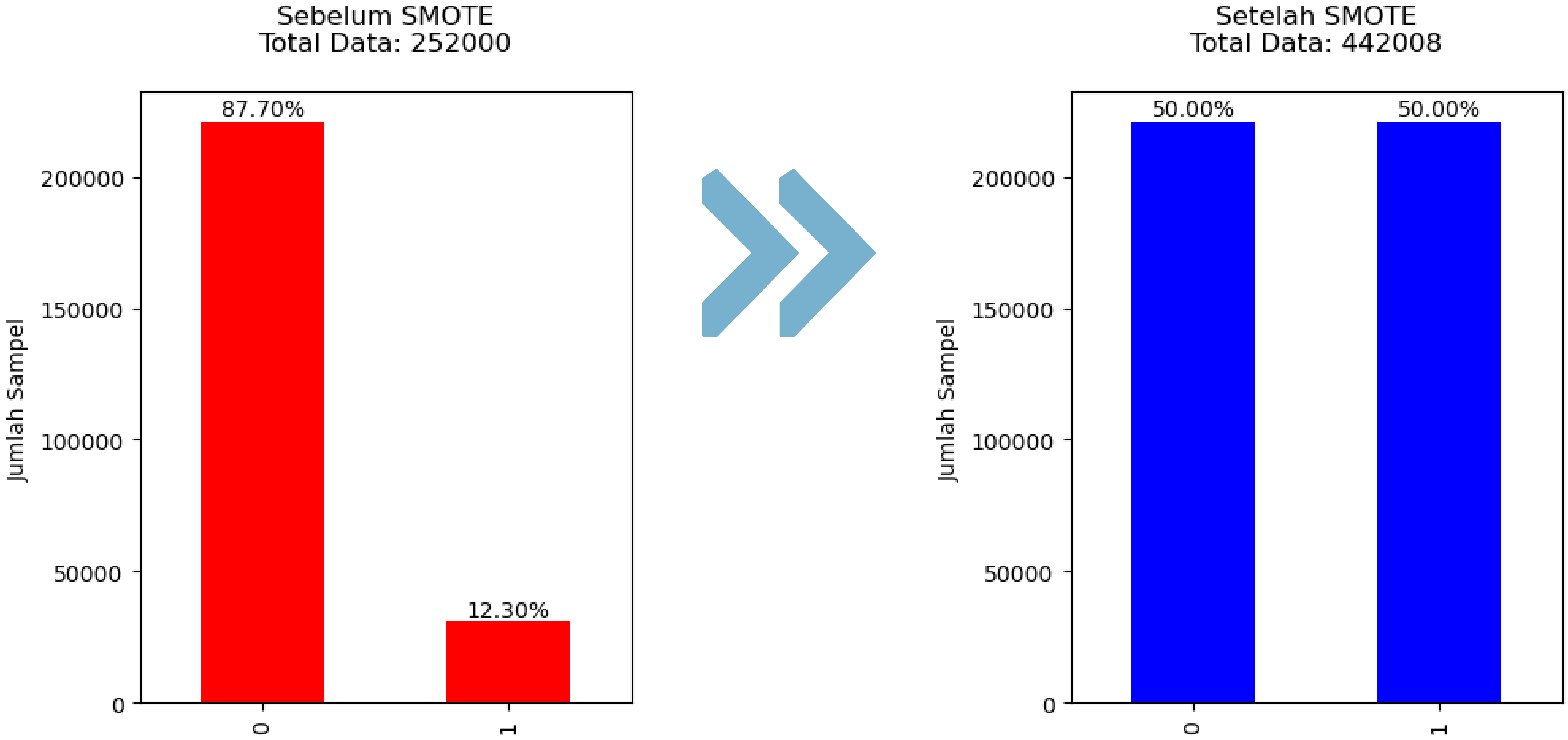
Feature Encoding

Transformasi data pada kolom kategorikal menggunakan OrdinalEncoder.
'Married/Single', 'Profession_Group', 'Age_Group', 'Income_Group', 'Experience_Group', 'CURRENT_JOB_Group', 'State_Zone', 'City_State'

Data PreProcessing

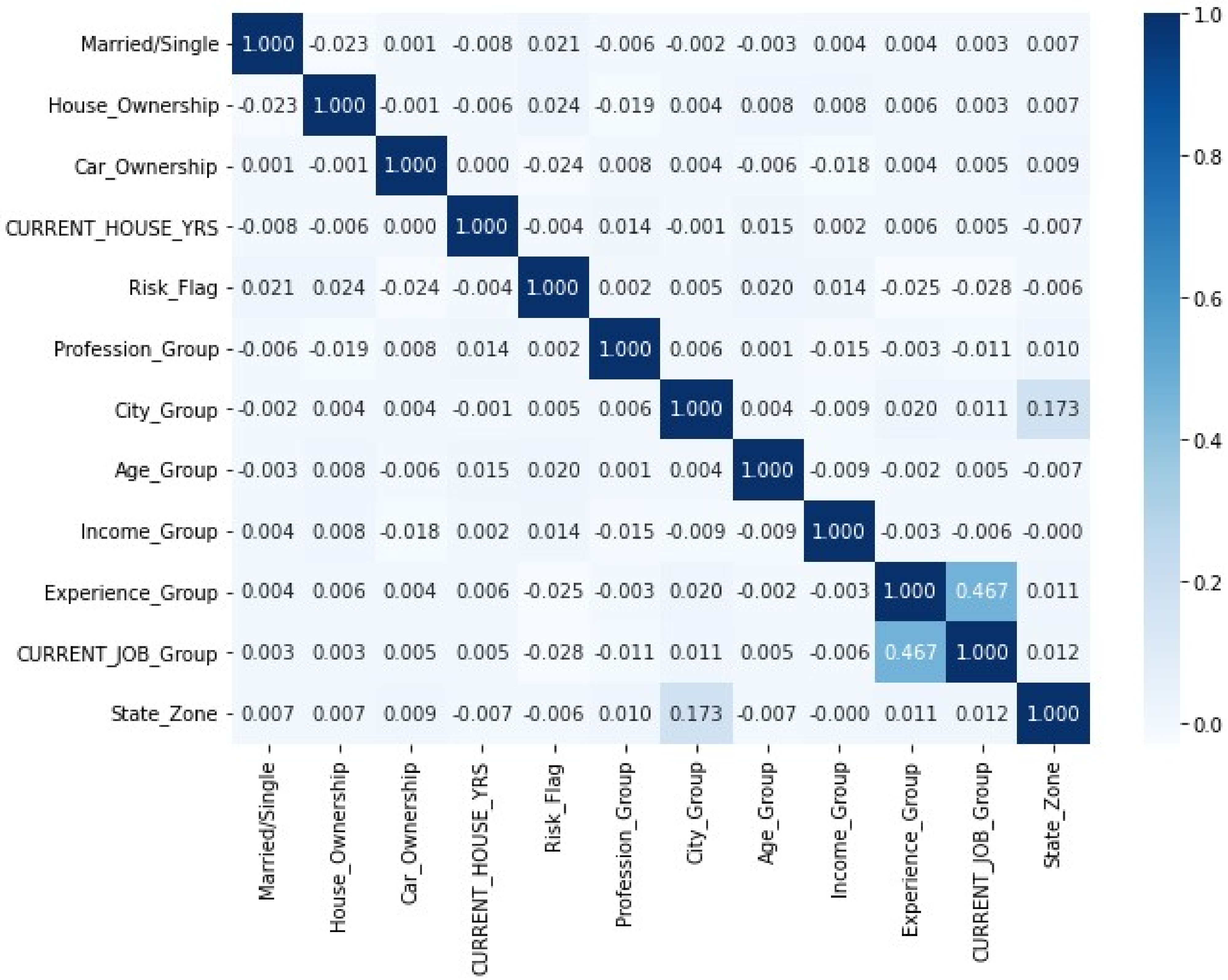
DATA REDUCTION

Handle Class Imbalance Oversampling SMOTE



Feature ENGINEERING

Data Nilai Korelasi Antar Feature




```python
# melihat nilai korelasi feature dengan feature target
data_corr.corr()['Risk_Flag'].sort_values()
```

```
CURRENT_JOB_Group    -0.028410
Experience_Group     -0.025108
Car_Ownership        -0.024036
State_Zone           -0.005805
CURRENT_HOUSE_YRS    -0.004375
Profession_Group      0.002047
City_Group            0.004921
Income_Group          0.013840
Age_Group             0.020129
Married/Single        0.021092
House_Ownership       0.023622
Risk_Flag             1.000000
Name: Risk_Flag, dtype: float64
```

FEATURE Yang DAPAT DIPERTAHANKAN

- CURRENT_JOB_Grou
  p
- Experience_Group
- Car_Ownership
- House_Ownership
- Married/Single

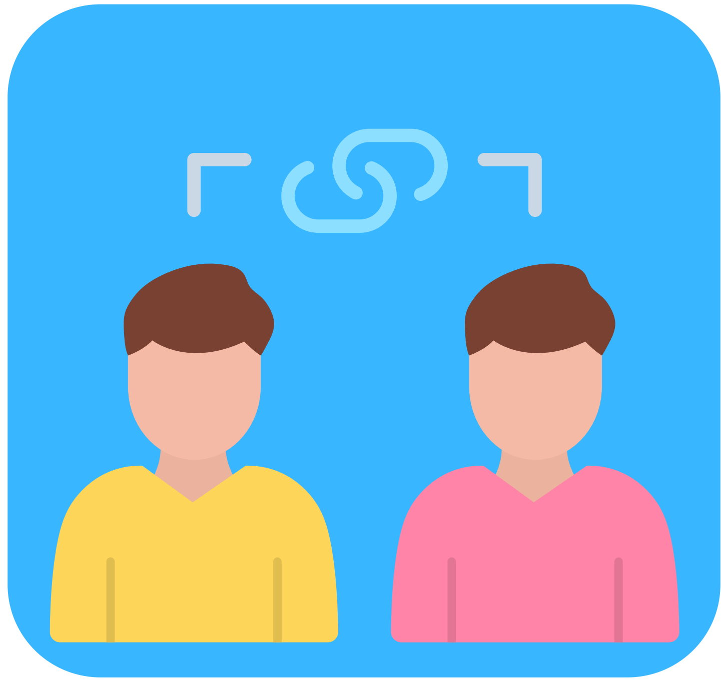
FEATURE YANG DAPAT TIDAK DIGUNAKAN

- City_Group
- State_Zone

FEATURE TAMBAHAN

Pendidikan

Kepemilikan Asuransi



Jumlah Tanggungan

Jumlah Pinjaman

Machine Learning Model

MODELLING

Process of Choosing the Algorithm

SPLITTING

ALGORITHM SCORE

No	Algoritma	Precision	Recall	F1-Score
1	Logistic Regression	0.52	0.53	0.53
2	Decision Tree	0.51	0.58	0.54
3	Random Forest	0.60	0.53	0.57

MODELLING

Hyperparameter Tuning

Grid Search CV

Parameter Terbaik

{'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 100}

F1 score Terbaik

0.5685075251318994

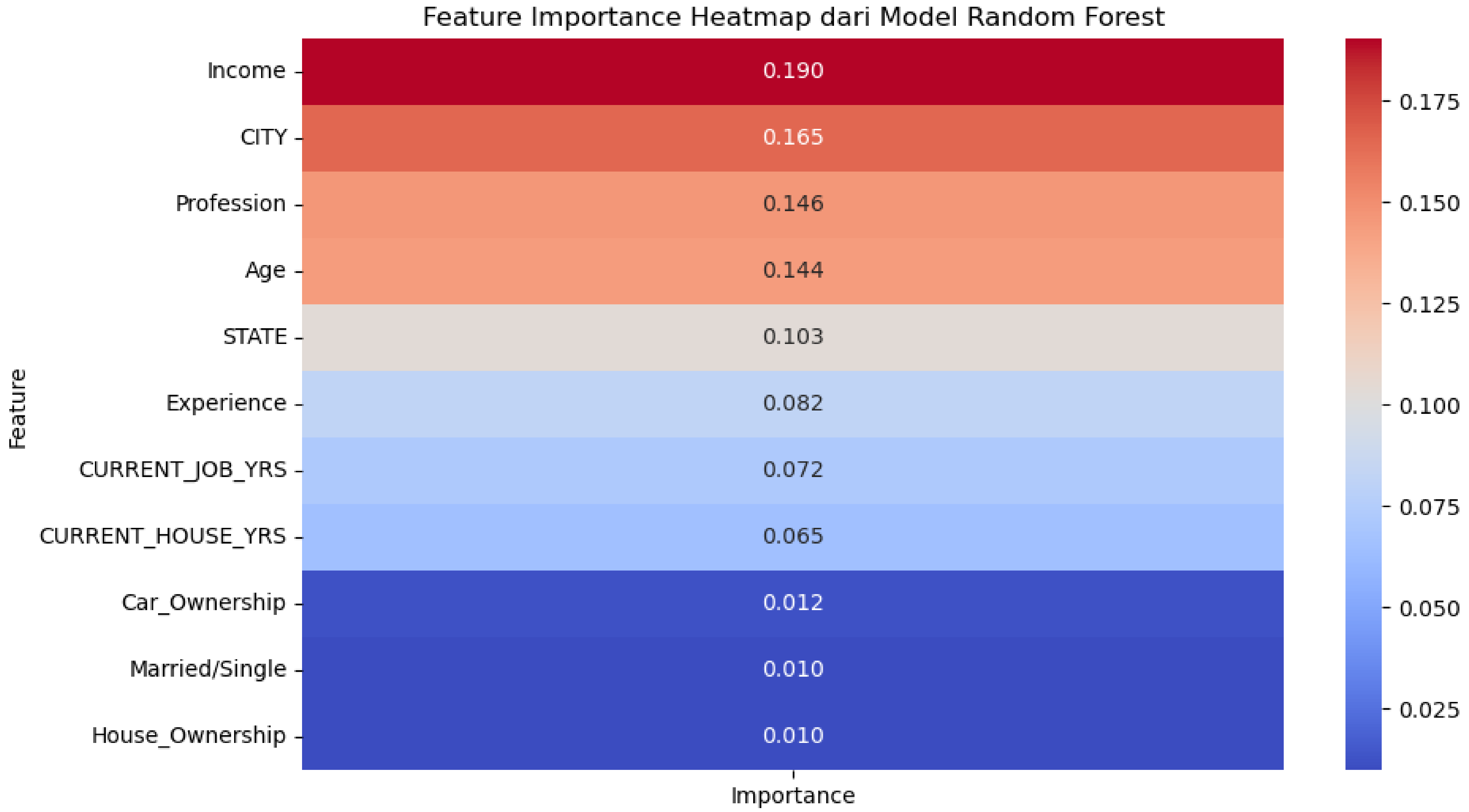
MODELLING

Evaluasi Model Performance

No	Data		Performance Score
1	Data Group	Train	67.99%
		Test	67.87%
2	Data Mentah	Train	93.64%
		Test	89.89%

MODELLING

Feature Importance



Fitur paling berpengaruh dalam pengambilan keputusan dari Algoritma RandomForest adalah **Income**, **CITY**, dan **Profession**

Perbandingan Data Test sebelum dan sesudah prediksi

	ID	Income	Age	Experience	Married/Single	House_Ownership	Car_Ownership	Profession	CITY	STATE	CURRENT_JOB_YRS	CURRENT_HOUSE_YRS
0	1	7393090	59	19	single	rented	no	Geologist	Malda	West Bengal	4	13
1	2	1215004	25	5	single	rented	no	Firefighter	Jalna	Maharashtra	5	10
2	3	8901342	50	12	single	rented	no	Lawyer	Thane	Maharashtra	9	14
3	4	1944421	49	9	married	rented	yes	Analyst	Latur	Maharashtra	3	12
4	5	13429	25	18	single	rented	yes	Comedian	Berhampore	West Bengal	13	11
...	...	...	...	...	...	...	...	...	...	...	...	...
27995	27996	9955481	57	13	single	rented	no	Statistician	Eluru[25]	Andhra Pradesh	5	10
27996	27997	2917765	47	9	single	rented	no	Technical writer	Ratlam	Madhya Pradesh	9	14
27997	27998	8082415	24	5	single	rented	no	Lawyer	Mira-Bhayandar	Maharashtra	4	13
27998	27999	9474180	51	13	single	rented	yes	Chartered Accountant	Bhilai	Chhattisgarh	13	14
27999	28000	9250350	42	9	single	rented	no	Chef	Navi Mumbai	Maharashtra	4	10

28000 rows × 12 columns

Test Data.csv (Kaggle) original

	Income	Age	Experience	CURRENT_JOB_YRS	CURRENT_HOUSE_YRS	Married/Single	House_Ownership	Car_Ownership	Profession	CITY	STATE	Predicted_Risk_Flag
0	0.739054	0.655172	0.95	0.285714	0.75	1.0	2.0	0.0	26.0	181.0	28.0	0
1	0.120596	0.068966	0.25	0.357143	0.00	1.0	2.0	0.0	24.0	131.0	14.0	0
2	0.890037	0.500000	0.60	0.642857	1.00	1.0	2.0	0.0	30.0	290.0	14.0	0
3	0.193614	0.482759	0.45	0.214286	0.50	0.0	2.0	1.0	1.0	171.0	14.0	0
4	0.000312	0.068966	0.90	0.928571	0.25	1.0	2.0	1.0	12.0	39.0	28.0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
27995	0.995562	0.620690	0.65	0.357143	0.00	1.0	2.0	0.0	44.0	90.0	0.0	0
27996	0.291051	0.448276	0.45	0.642857	1.00	1.0	2.0	0.0	47.0	249.0	13.0	0
27997	0.808059	0.051724	0.25	0.285714	0.75	1.0	2.0	0.0	30.0	190.0	14.0	0
27998	0.947381	0.517241	0.65	0.928571	1.00	1.0	2.0	1.0	7.0	47.0	4.0	0
27999	0.924975	0.362069	0.45	0.285714	0.00	1.0	2.0	0.0	8.0	212.0	14.0	0

28000 rows × 12 columns

Test Data.csv (Kaggle) setelah ditransformasi dan diprediksi dengan model Random Forest

Perbandingan Risk_Flag data_mentah VS Data Test

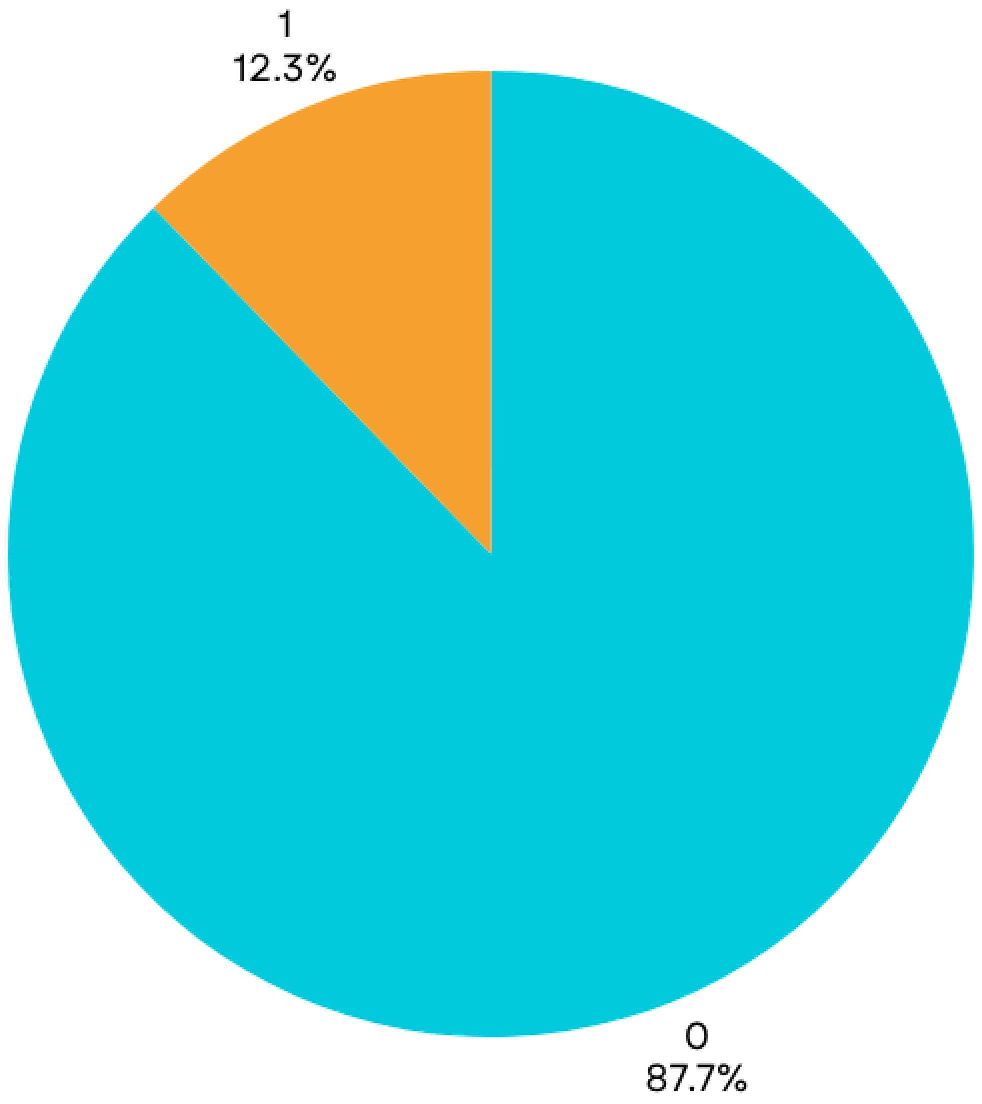
```
Unique values kolom Risk_Flag di data_mentah
0    221004
1     30996
Name: Risk_Flag, dtype: int64
```

Komposisi Risk_Flag pada data_mentah

```
Unique values kolom Predicted_Risk_Flag di Test Data
0    24945
1     3055
Name: Predicted_Risk_Flag, dtype: int64
```

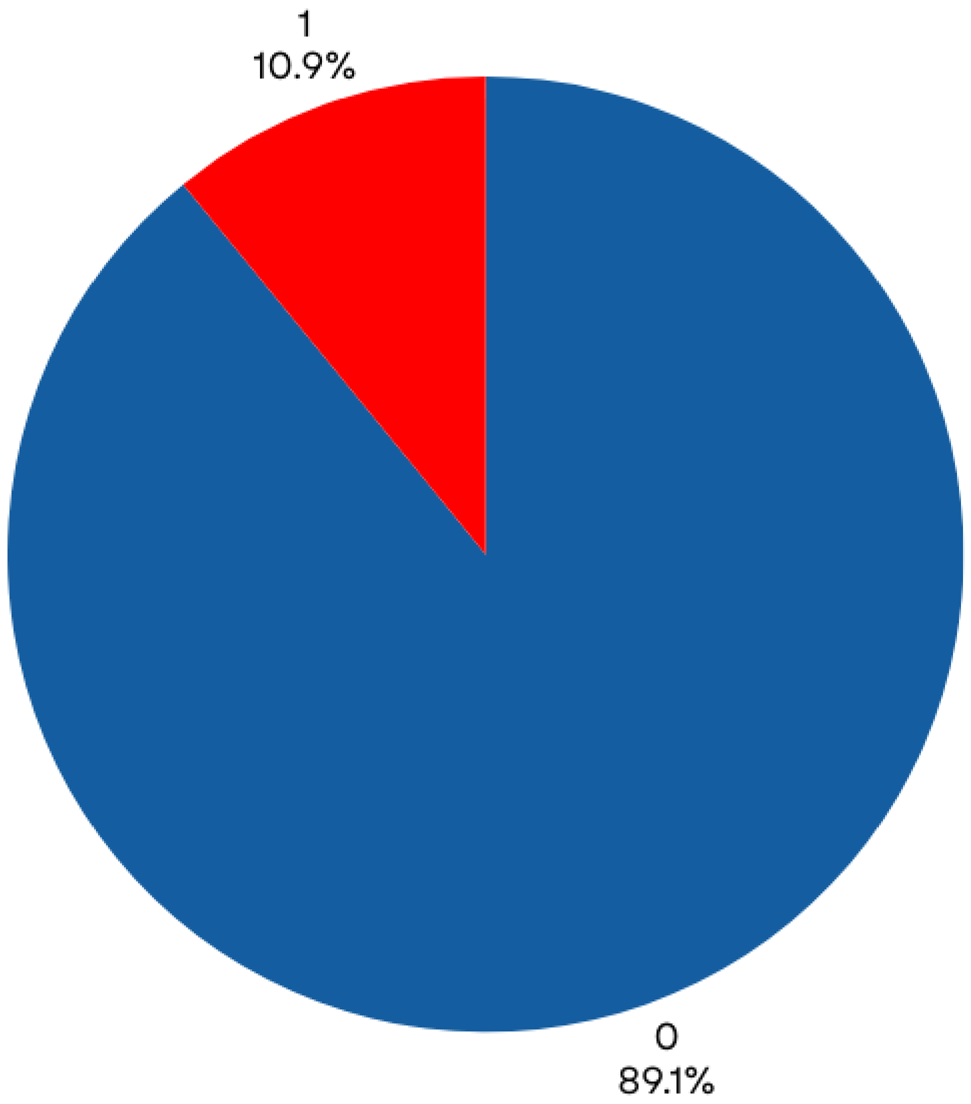
Komposisi Predicted_Risk_Flag di Test Data setelah melalui predict dengan model Random Forest

data_mentah



Total data_mentah row: 252000

Test Data



Total Test Data row: 28000

Conclusion

RECCOMENDATION – SUGGESTION

Dengan Machine Learning dari data behavior, dapat direkomendasikan untuk perusahaan untuk mempertimbangkan data sebagai berikut :

Income

200,000 s.d 1,000,000

City

Town: Srinagar, Jodhpur, Amritsar, Kota, Ajmer, Bikaner, Mysore.
Rural: Gorakhpur, Warangal, Siliguri, Dhanbad, Farrukhabad, Haldia, Gan

Profession

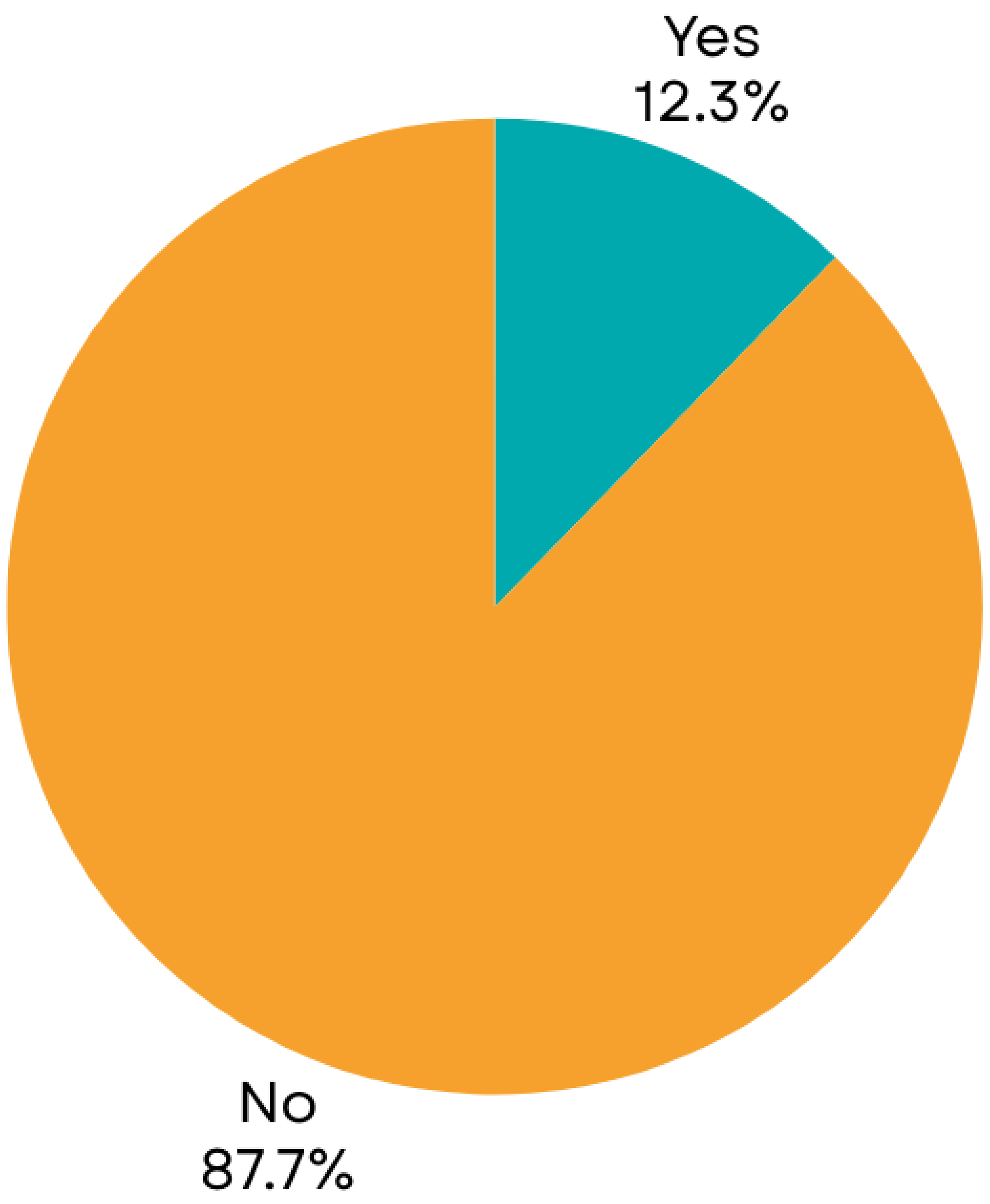
Research, Engineering, & IT: Statistician, Web_designer, Engineer, Computer_hardware_engineer, Drafter, Scientist, Industrial_Engineer, Mechanical_engineer, Chemical_engineer, Biomedical_Engineer, Petroleum_Engineer, Technology_specialist, Design_Engineer, Civil_engineer, Software_Developer, Computer_operator, Technical_writer, Graphic_Designer, Web_designer, Architect, Technician, Microbiologist, Geologist, Statistican, dan Surveyor.

Bussines Metrics

Default rate

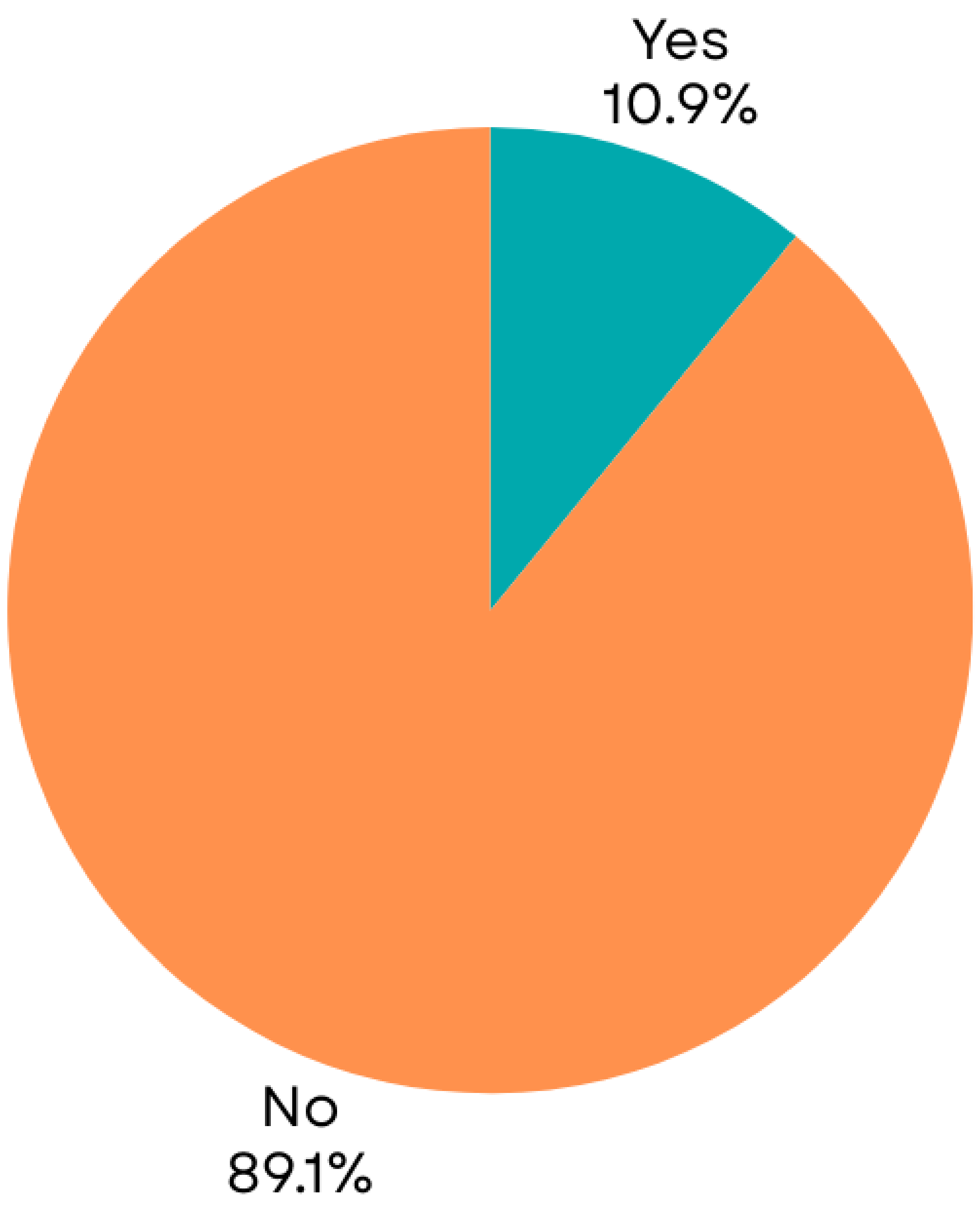
Default rate :
Terdapat **87.7%** dari total **252.000 nasabah**

Sementara pada data test (Test Data), persentase nasabah yang dapat disetujui adalah **89.1%** dari **28.000 calon nasabah**

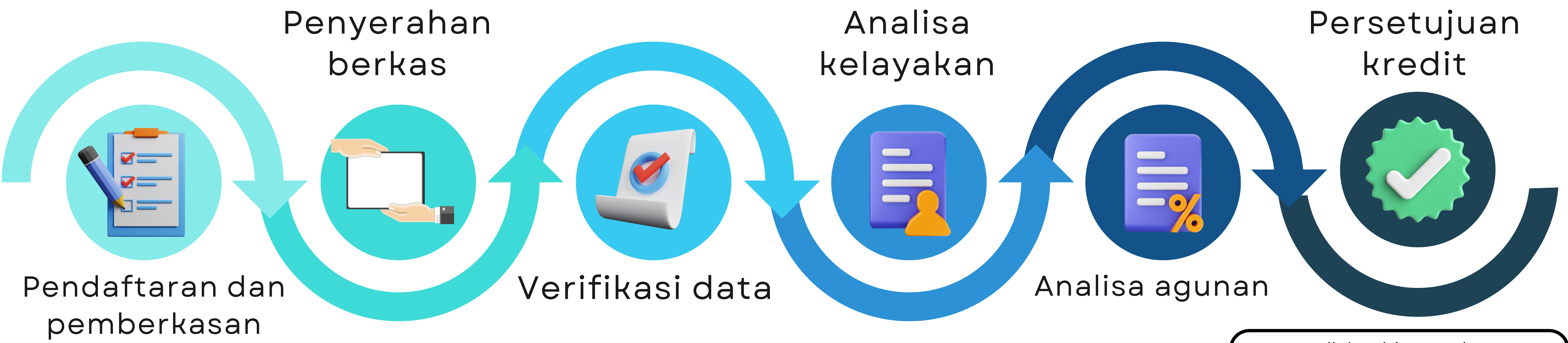


Data mentah

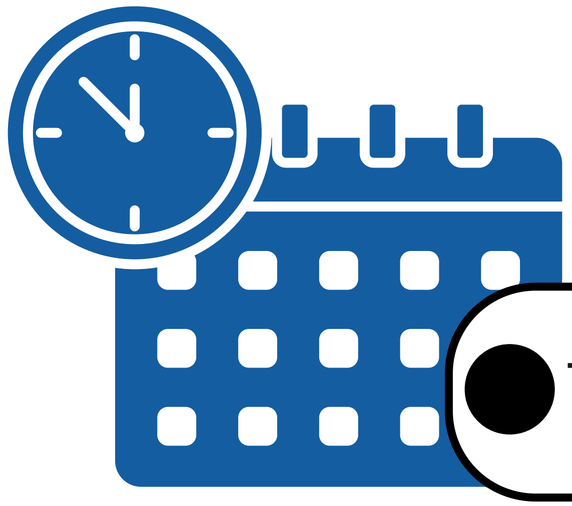
Test Data



WORKFORCE TIME
MANUAL

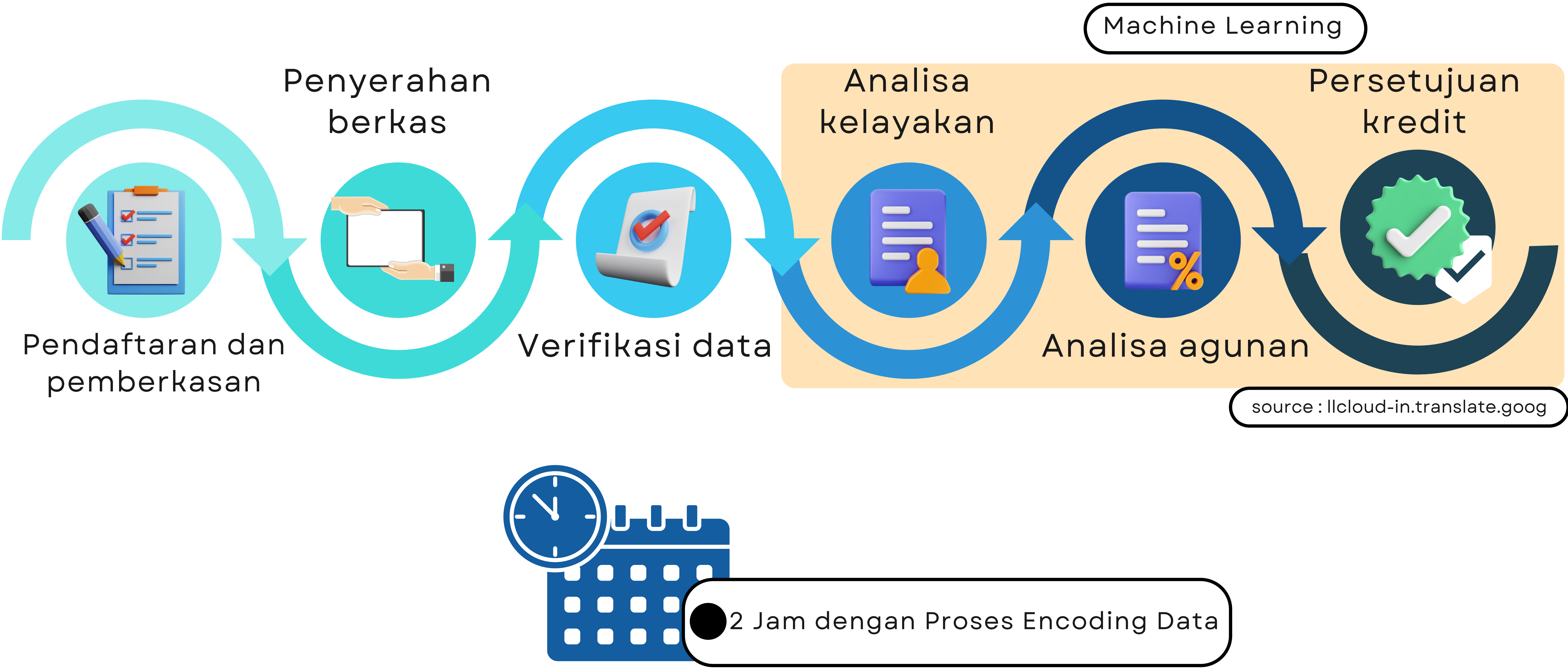


source : llcloud-in.translate.goog



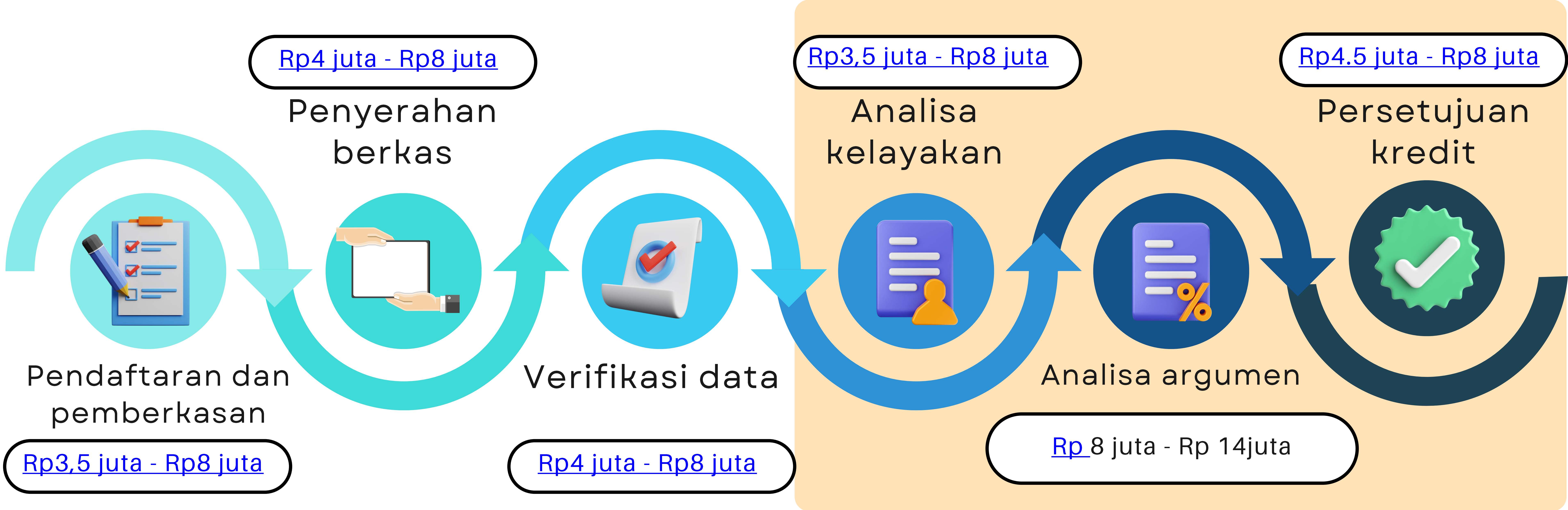
1 Minggu atau beberapa bulan

WORKFORCE TIME
MANUAL & MACHINE LEARNING



WORKFORCE COST

cost



RP 27,5JUTA → **ML Process = 16 JUTA**
Data Scientist= 9 - 10 JUTA → **RP 20,5 -21,5 JUTA**

https://www.detik.com/

The Efficiency

- **TIME** : 6 hari kerja menjadi 3 hari dan 3 jam kerja
- **COST** : 27,5 Juta menjadi 20,5-21,5 Juta
- **MAN** : 1 orang per proses (6 orang) menjadi 3 orang dan 1 Data Scientist (4 orang)

#1

44%

Time

#2

25,4%

Cost

#3

33%

Man

Thank You