

Credit Scoring Card

Project Based Internship Program at
Home Credit Indonesia X Rakamin Academy

Presented by:
Sri Ranita



1

Company Profile

PT Home Credit Indonesia merupakan perusahaan pembiayaan berbasis teknologi seperti HP, Laptop, TV dan elektronik lainnya dengan menawarkan cicilan ringan tanpa beban, pembiayaan yang mudah, praktis dan cepat.

2

Goal

✓ Project ini bertujuan untuk **mengoptimalkan akurasi sistem credit scoring** menggunakan machine learning.

✓ Sasaran utamanya adalah **memastikan bahwa pemohon kredit yang layak tidak ditolak**, serta **memberikan pinjaman dengan ketentuan yang sesuai guna mendorong keberhasilan dalam pelunasan**.

3

Objective

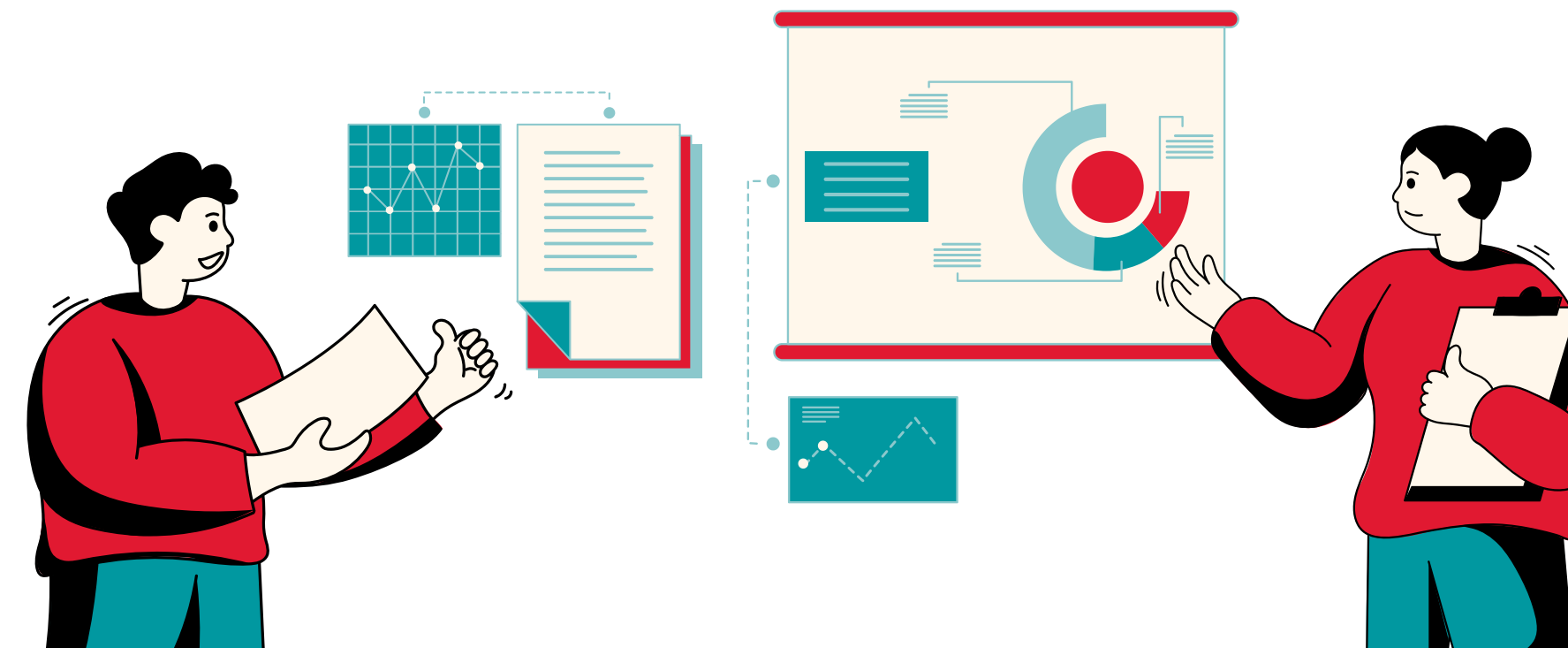
✓ Implementasi model machine learning, khususnya pada **Logistic Regression dan Random Forest** dalam memberikan prediksi credit scoring

✓ Mengidentifikasi segmen pelanggan berisiko rendah atau tinggi berdasarkan **faktor-faktor yang memengaruhi kelayakan kredit pelanggan dari data demografi**

4

Metrics

- ✓ Accuracy
- ✓ AUC-ROC
- ✓ Confusion Matrix



Data Understanding

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	...
0	100002	1	Cash loans	M	N	Y	0	202500.0	406597.5	24700.5	...
1	100003	0	Cash loans	F	N	N	0	270000.0	1293502.5	35698.5	...
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	135000.0	6750.0	...
3	100006	0	Cash loans	F	N	Y	0	135000.0	312682.5	29686.5	...
4	100007	0	Cash loans	M	N	Y	0	121500.0	513000.0	21865.5	...

Dataset yang digunakan pada project ini merupakan data application yang berisi informasi mengenai **pengajuan pinjaman**. Setiap baris merepresentasikan satu pengajuan pinjaman oleh peminjam. Selain itu, dataset ini mencakup berbagai informasi seperti demografi, pekerjaan, pendapatan, status kepemilikan properti/kendaraan, dan lainnya.

Dataset yang digunakan terdiri 2 jenis:

- Data **Application train** = Berisikan variabel Target (0 = tidak gagal bayar, 1 = gagal bayar) untuk melatih model
- Data **Application test** = Tanpa variabel Target sehingga digunakan untuk prediksi kemampuan pelanggan

Data application_train ini terdiri dari:

307.511

Observasi

122

Kolom

Data application_test ini terdiri dari:

48.744

Observasi

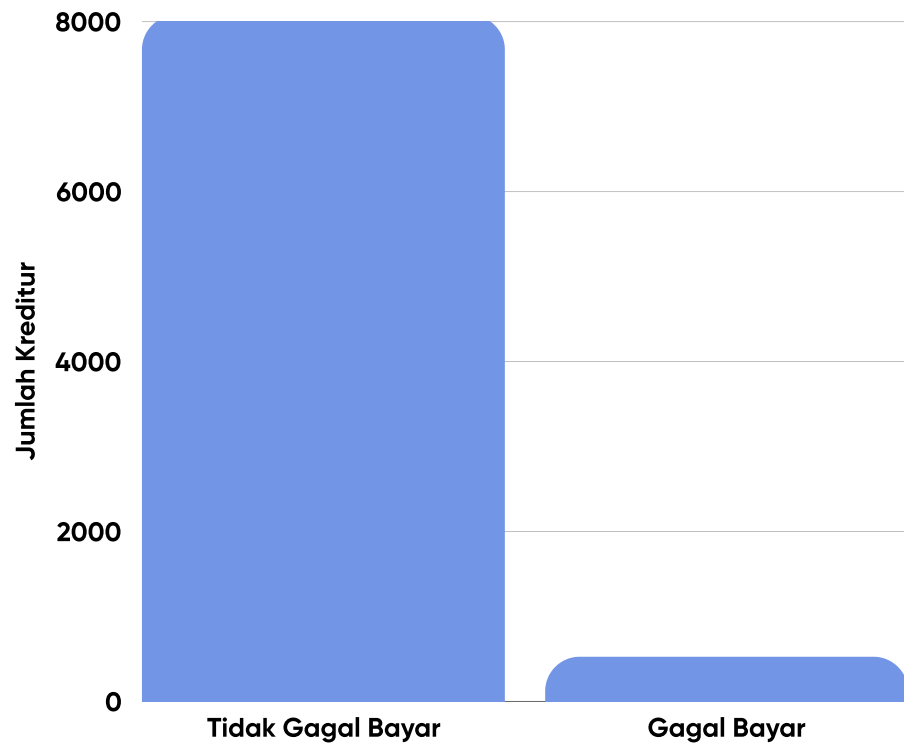
121

Kolom

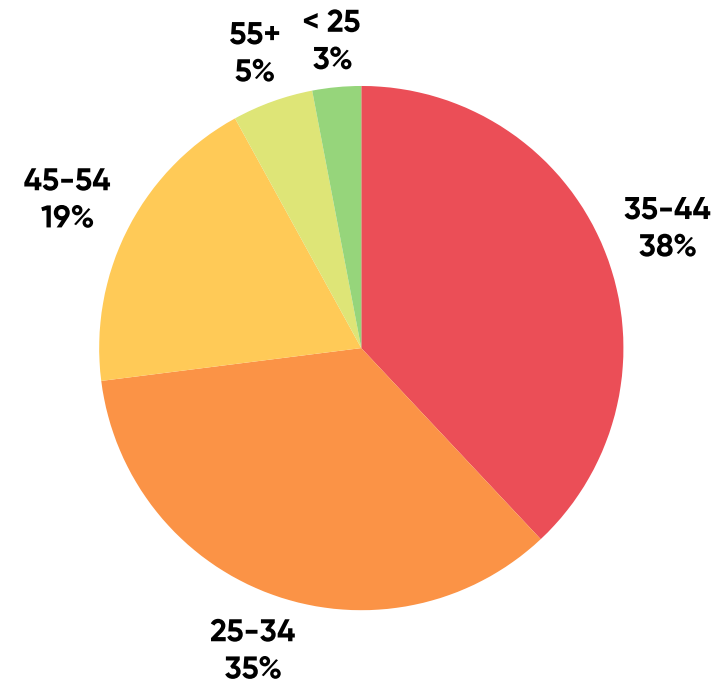
Exploratory Data Analyst



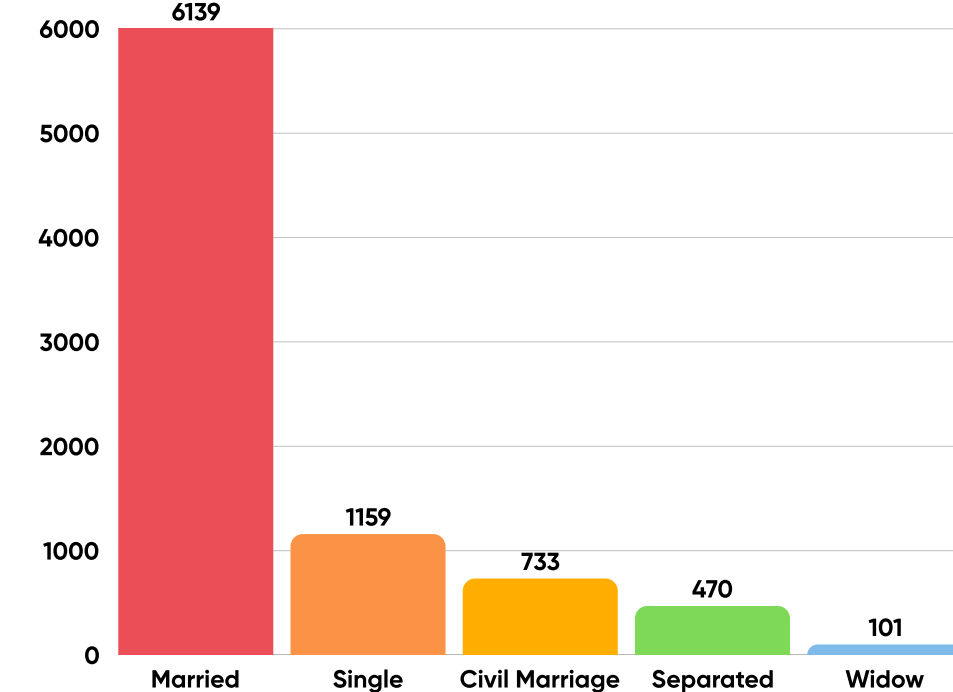
Jumlah Kreditur Berdasarkan Kemampuan Membayar



Proporsi Kreditur Berdasarkan Kelompok Umur



Jumlah Kreditur Berdasarkan Status Perkawinan



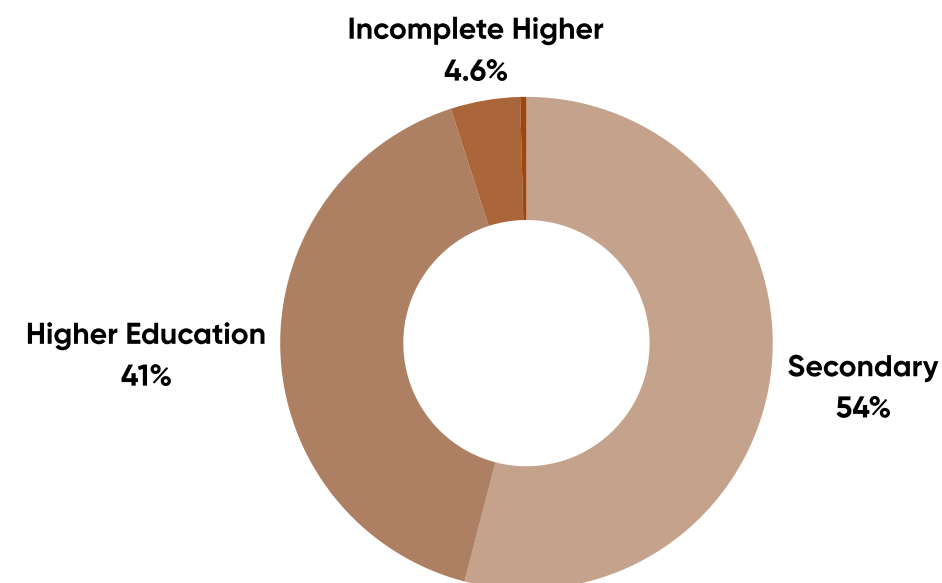
Proporsi Kreditur Berdasarkan Jenis Kelamin



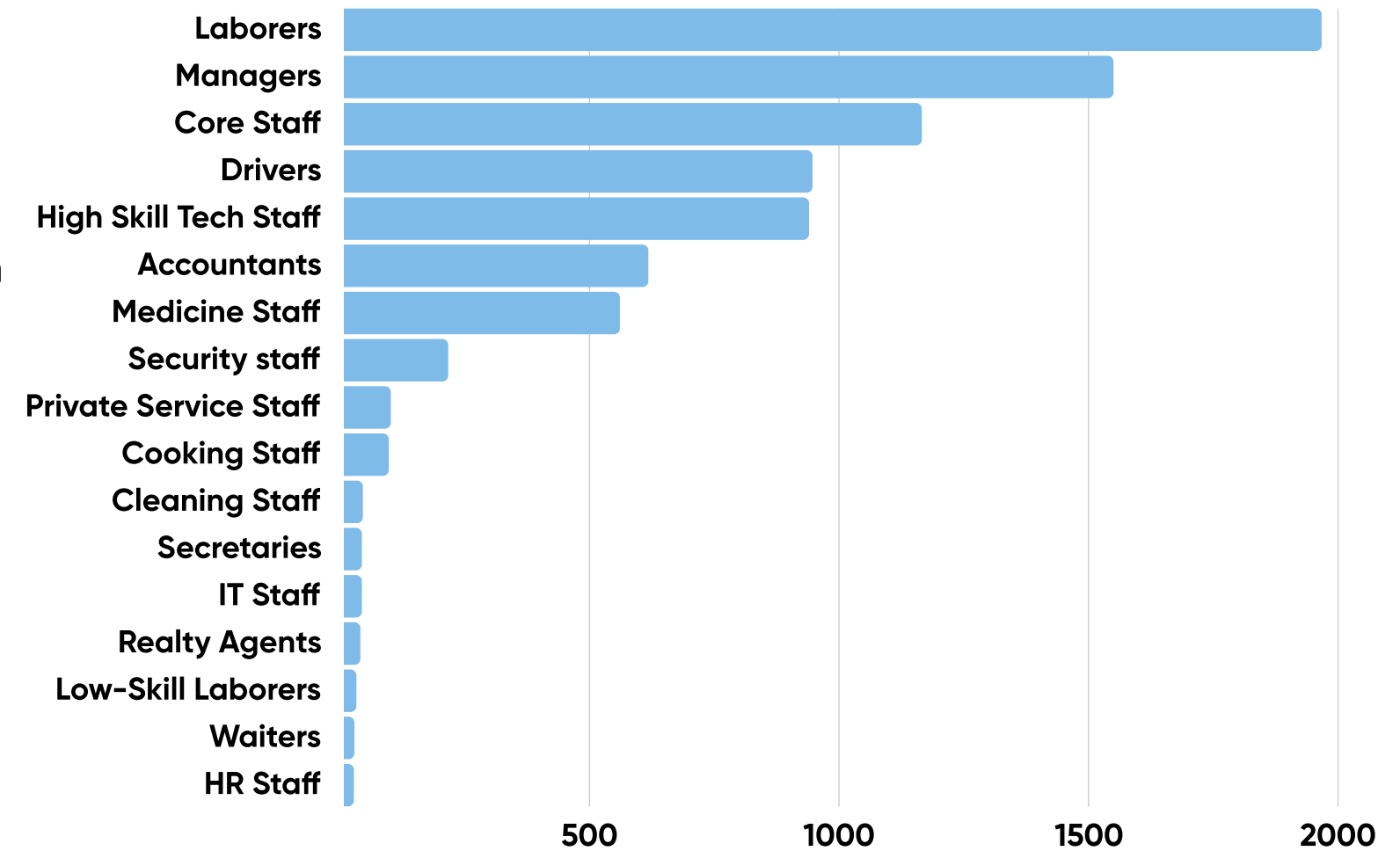
Rata-Rata Income Kreditur

Rp2.227.282

Jumlah Kreditur Berdasarkan Tingkat Pendidikan



Jumlah Kreditur Berdasarkan Pekerjaan



Data Preprocessing

1. Handling Missing Value

- ✓ Menghapus nilai yang mengandung NA atau kosong, sebanyak 41.519 baris

2. Feature Selection

- ✓ Menghapus variabel yang kurang relevan untuk modeling

3. Feature Engineering

- ✓ Merubah variabel yang lebih mudah diinterpretasi, seperti Usia dan Durasi Kerja

4. Splitting Data

- ✓ Membagi data training dan data test untuk modeling

5. Normalisasi Data

- ✓ Menstandarkan variabel numerik kedalam skala 0-1.

6. Label Encoder

- ✓ Mengubah variabel kategorik menjadi numerikal dengan encoder

7. SMOTE

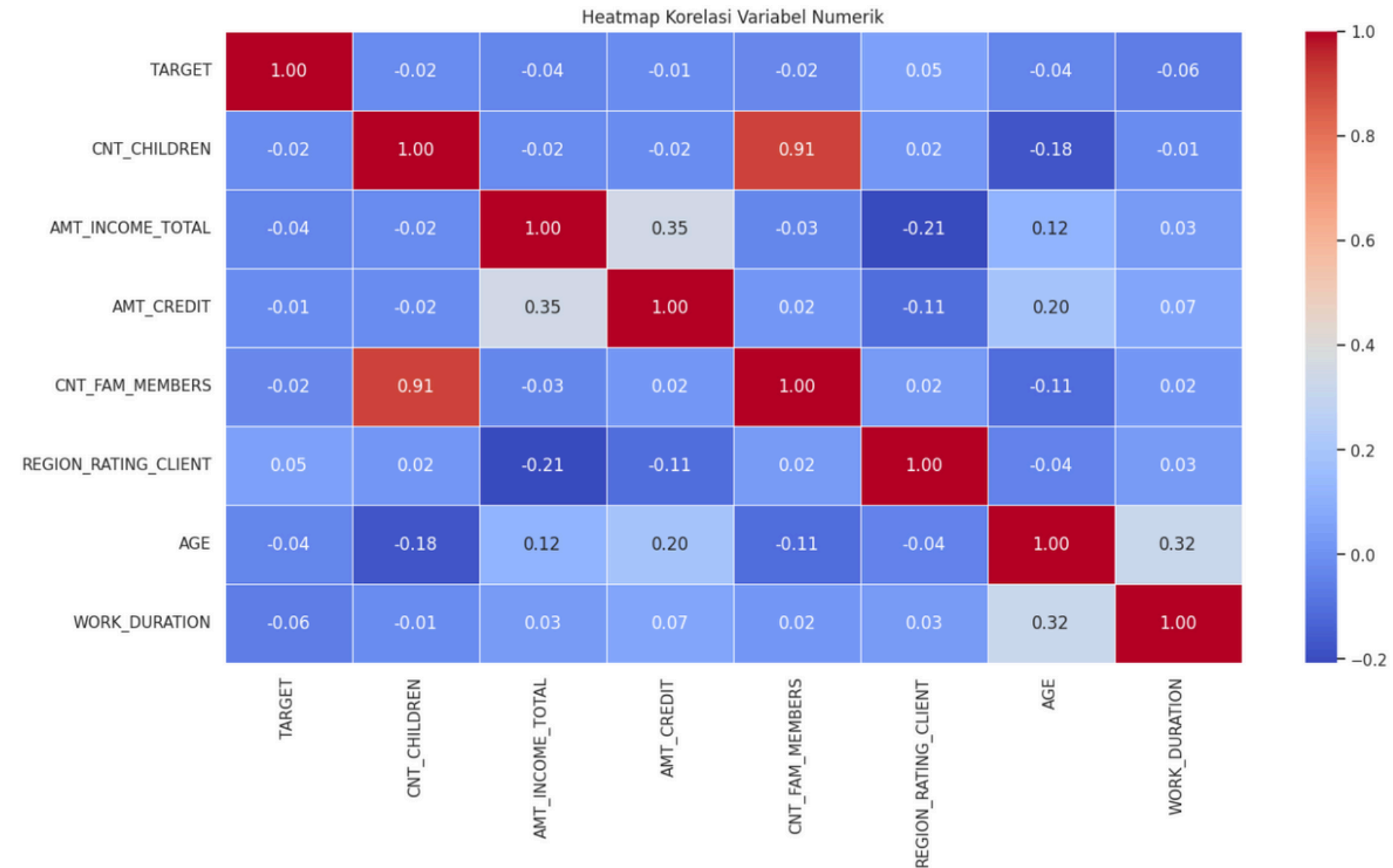


- ✓ Berdasarkan visualisasi disamping, data mengalami imbalance, sehingga perlu ditangani dengan menyeimbangkan variabel Target dengan teknik oversampling SMOTE

Korelasi Antar Variabel

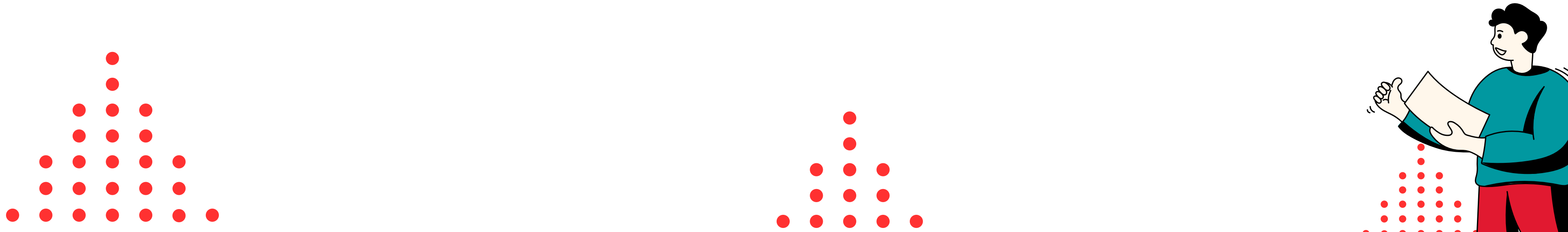
Heatmap menunjukkan bahwa

- Keluarga dengan banyak anak (CNT_CHILDREN) cenderung memiliki anggota keluarga lebih banyak (CNT_FAM_MEMBERS) dengan **korelasi sangat tinggi** (0.89), sehingga perlu pertimbangan khusus dalam penilaian kredit.
- Pendapatan (AMT_INCOME_TOTAL) **berkorelasi moderat** dengan jumlah pinjaman (AMT_CREDIT), tetapi tidak cukup kuat, menunjukkan bahwa faktor lain seperti stabilitas pekerjaan mungkin lebih berpengaruh.
- Usia (AGE) dan lama bekerja (WORK_DURATION) **berkorelasi positif** (0.25), mengindikasikan bahwa debitur yang lebih tua cenderung lebih stabil.
- Wilayah dengan rating rendah (REGION_RATING_CLIENT) **berkorelasi negatif** dengan pendapatan, menandakan potensi risiko kredit yang lebih tinggi di daerah tersebut.



Algoritma	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.59	0.90	0.59	0.69	0.50
Random Forest	0.85	0.89	0.85	0.87	0.51

Random Forest jelas **lebih unggul** dalam **mendeteksi risiko kredit**, dengan **akurasi 85%**, recall 85%, dan F1-Score 0.87, menunjukkan kemampuannya memprediksi default secara lebih tepat dan seimbang dibanding Logistic Regression (akurasi 59%, recall 59%, F1-Score 0.69). Meski AUC-ROC kedua model setara, precision tinggi (89-90%) menjamin minimnya nasabah baik yang salah diklasifikasikan.



Business Recommendations



1. Penggunaan Model Random Forest sebagai Model Utama

Berdasarkan hasil evaluasi, **model Random Forest** menunjukkan **performa yang paling stabil dan akurat** dibandingkan Logistic Regression. Dengan nilai akurasi mencapai 85%, serta F1-score dan recall yang juga tinggi, Random Forest mampu menangkap pola risiko gagal bayar secara lebih efektif. Oleh karena itu, disarankan agar model ini dijadikan model utama dalam sistem credit scoring Home Credit Indonesia.

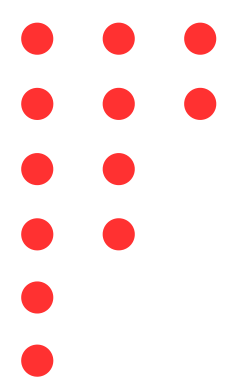
2. Penerapan Kebijakan Kredit Lebih Ketat untuk Segmen Risiko Tinggi

Hasil analisis menunjukkan bahwa nasabah berusia muda (<25 tahun), berpendidikan menengah (secondary), serta yang bekerja di sektor berisiko tinggi seperti laborer, driver, dan low-skilled staff memiliki **kecenderungan lebih besar untuk gagal bayar**. Perusahaan disarankan menerapkan kebijakan kredit yang lebih ketat, seperti peningkatan verifikasi dan peninjauan tenor, khususnya untuk segmen ini.

3. Pemanfaatan Informasi Lokasi dan Stabilitas Pekerjaan

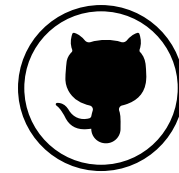
Variabel seperti **`REGION_RATING_CLIENT`** dan **lama bekerja berkontribusi** terhadap **stabilitas pembayaran**. Nasabah dari wilayah dengan rating lebih rendah atau yang memiliki riwayat kerja yang singkat cenderung memiliki risiko lebih tinggi. Oleh karena itu, pengambilan keputusan kredit sebaiknya mempertimbangkan kombinasi faktor lokasi dan riwayat pekerjaan sebagai indikator tambahan dalam penilaian risiko.





//////

Thank You Very Much



Final Project HCI X Rakamin Academy.



www.linkedin.com/in/sriranita

