



Credit Risk Analyst

Project Based Internship

https://github.com/WiefranVarenzo/HomeCredit_Rakamin



Project Overview

Home Credit menghadapi tantangan dalam memprediksi risiko kredit secara akurat untuk memastikan pelanggan yang mampu melunasi pinjaman tidak ditolak. Dengan menggunakan metode Machine Learning seperti Logistic Regression dan model lainnya, perusahaan bertujuan untuk meningkatkan persetujuan pinjaman yang tepat, mengurangi risiko gagal bayar, dan meningkatkan kepuasan serta kepercayaan pelanggan. Hal ini penting untuk pertumbuhan dan keberlangsungan bisnis.

Problem Research

- Analisis Faktor Pengajuan Pinjaman: Identifikasi faktor utama yang mempengaruhi keputusan pengajuan pinjaman.
- Segmentasi Pelanggan: Kelompokkan pelanggan berdasarkan karakteristik dan perilaku mereka.
- Prediksi Gagal Bayar: Memprediksi risiko gagal bayar pelanggan menggunakan model Machine Learning.

Tahap Data Understanding

- **Data Application Train**

122 Columns X 307,511 Rows

50 Kolom memiliki >30% Missing value

Beberapa Kolom memiliki tipe data yang salah (string-> integer)

- **Data Application Test**

122 Columns X 48,744 Rows

50 Kolom memiliki >30% Missing value

Beberapa Kolom memiliki tipe data yang salah (string-> integer)

Tahap Cleaning

Menghapus kolom dengan >30% missing value

Ada 50 Kolom dengan missing value >30% dihapus/drop karena merupakan data tidak berkualitas

Melakukan imputasi kolom dengan <30% missing value dengan nilai median

Menggunakan nilai median akan mencegah nilai dengan skewness yang condong ke samping

Melakukan Feature Engineering pada kolom Days_BIRTH menjadi AGE_Years

DAYs_BIRTH Bernilai jumlah hari yang negatif, membagi dengan 365 hari akan mempermudah mengetahui umur klien

Memastikan tidak ada value yang duplikat

Setelah di cek, Setiap kolom tidak memiliki nilai duplikat

Memperbaiki Tipe data ID yang awalnya integer menjadi string

Jika tipe data ID integer, maka takutnya akan ikut terproses saat Scaling kolom numerik

Drop Kolom yang tidak memberikan informasi yang penting, seperti satuan jam atau hari, serta kolom flag yang tidak penting

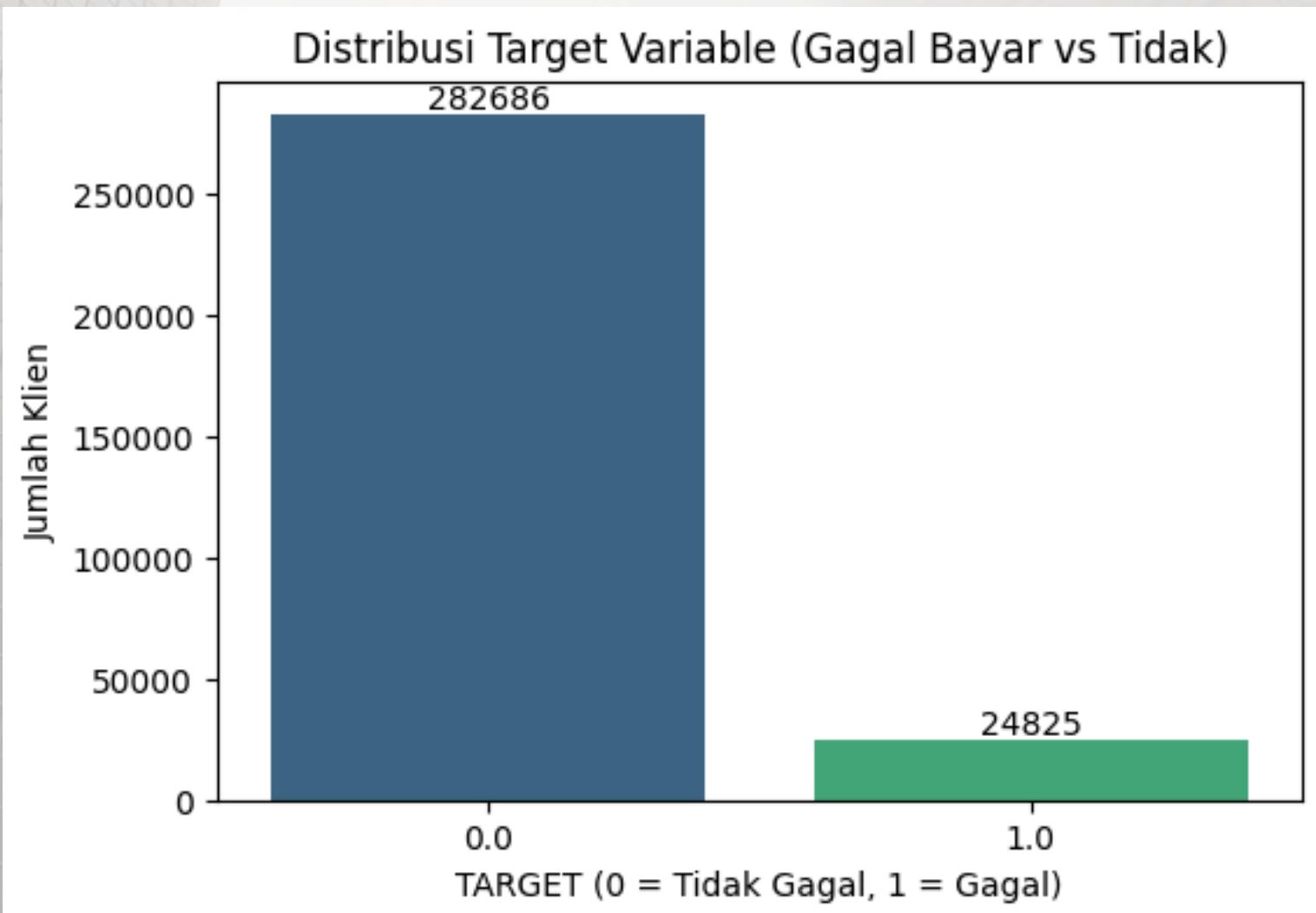
Satuan hari/jam/minggu bisa di gabungkan dengan rekapan perbulannya

Tahap PreProcessing

Membagi dataset menjadi dua bagian yaitu data latih(train) dan uji(test).

Melakukan encoding terhadap feature kategorik dan scaling terhadap feature numerik.

Melakukan encoding terhadap feature kategorik dan scaling terhadap feature numerik.

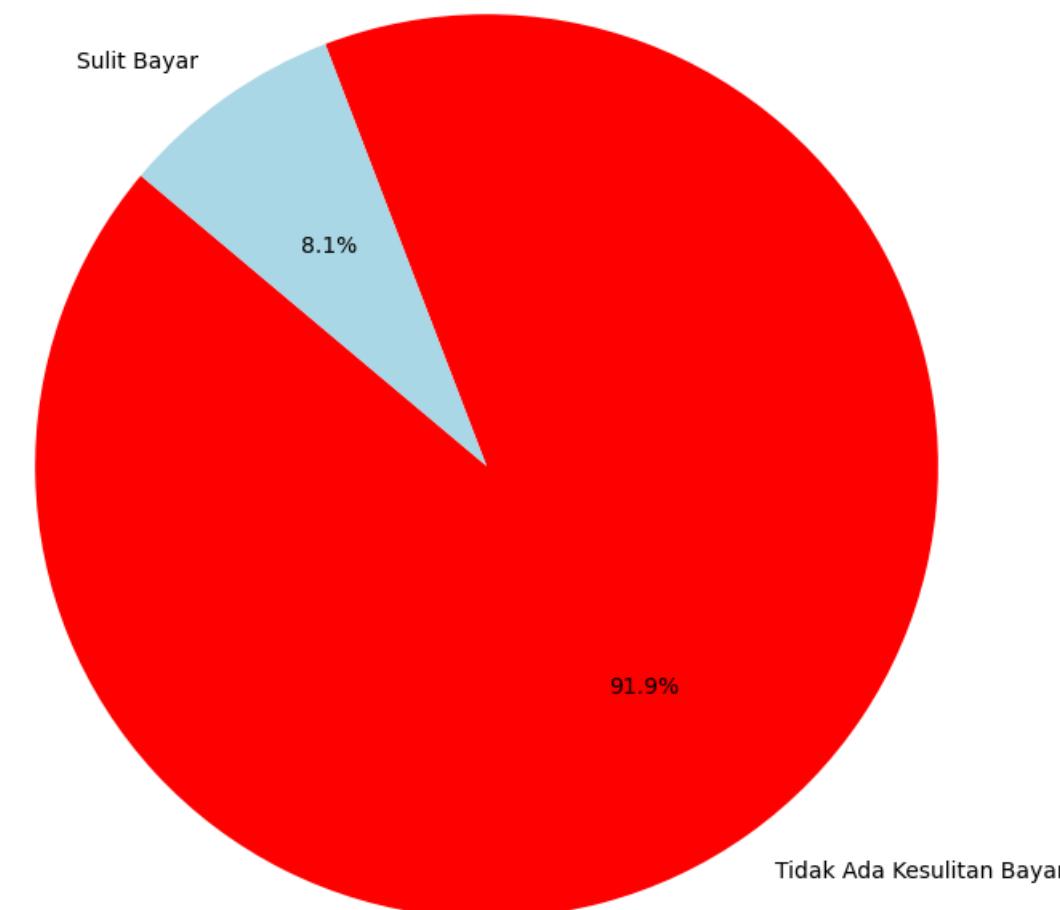


Tahap Data Visualization

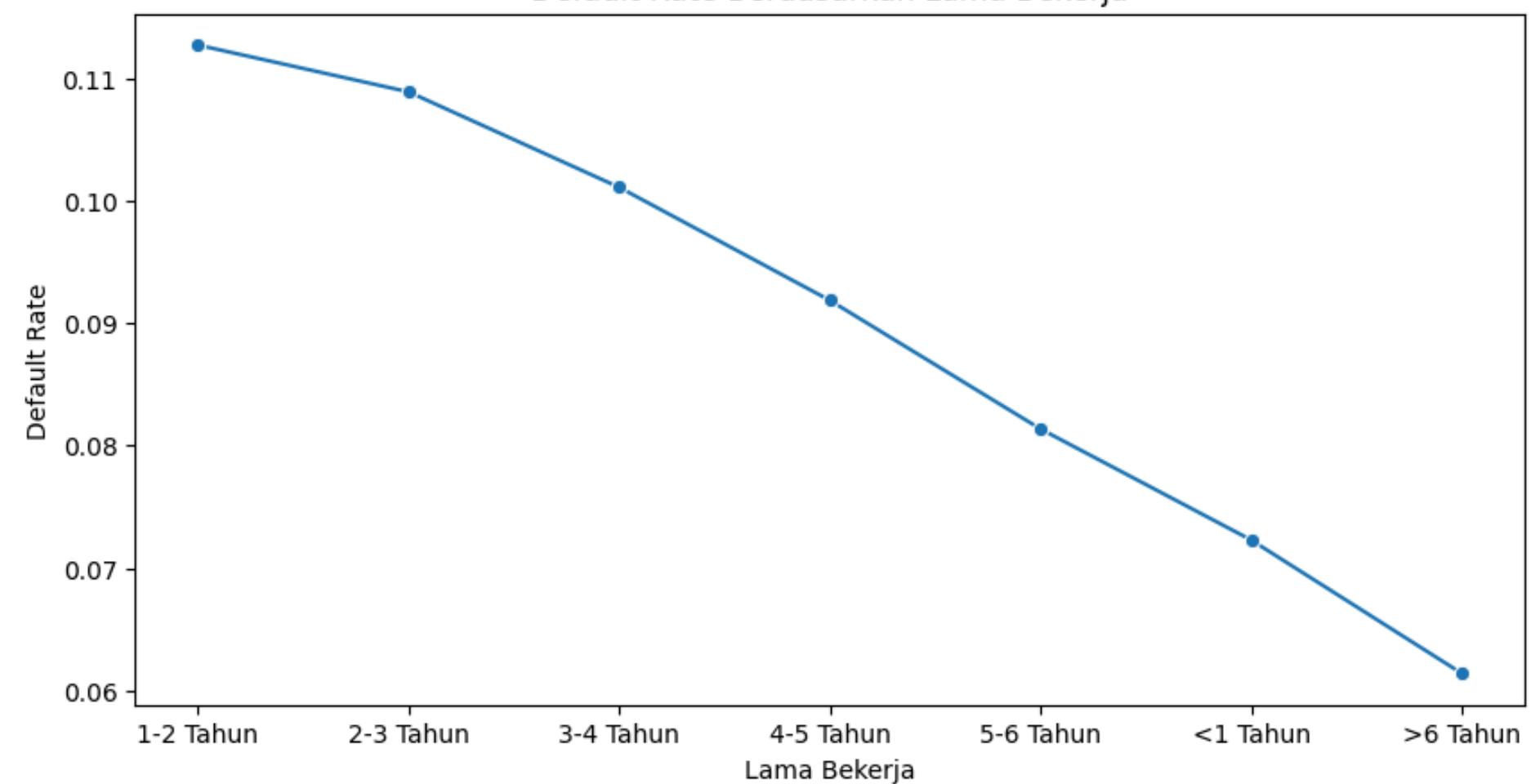
Distribusi Data dan Faktor Resiko Berdasarkan Lama Kerja

Default Rate : Nilai Rata-Rata klien yang sulit membayar

Distribusi Target Variable (Gagal Bayar vs Tidak)



Default Rate Berdasarkan Lama Bekerja

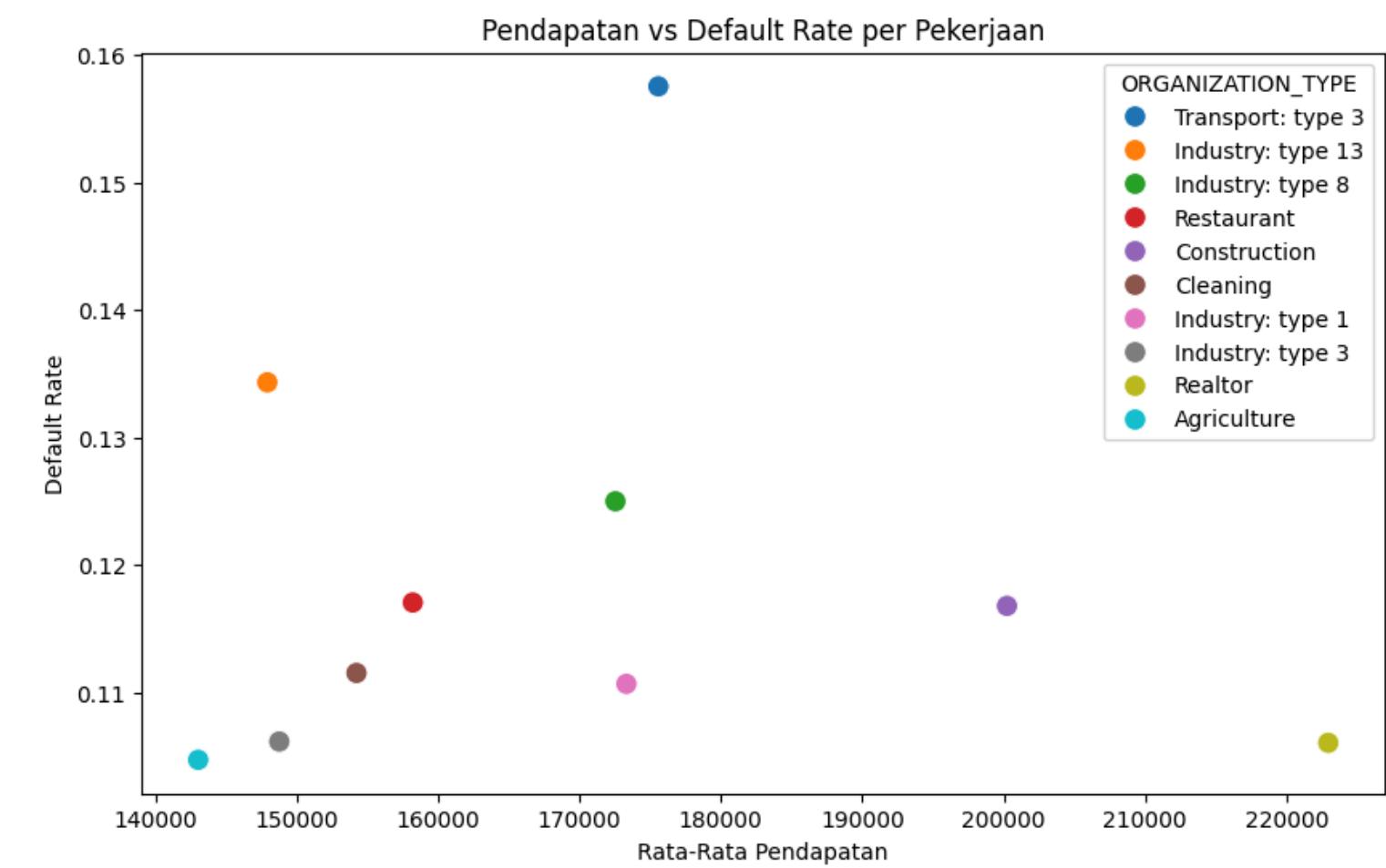
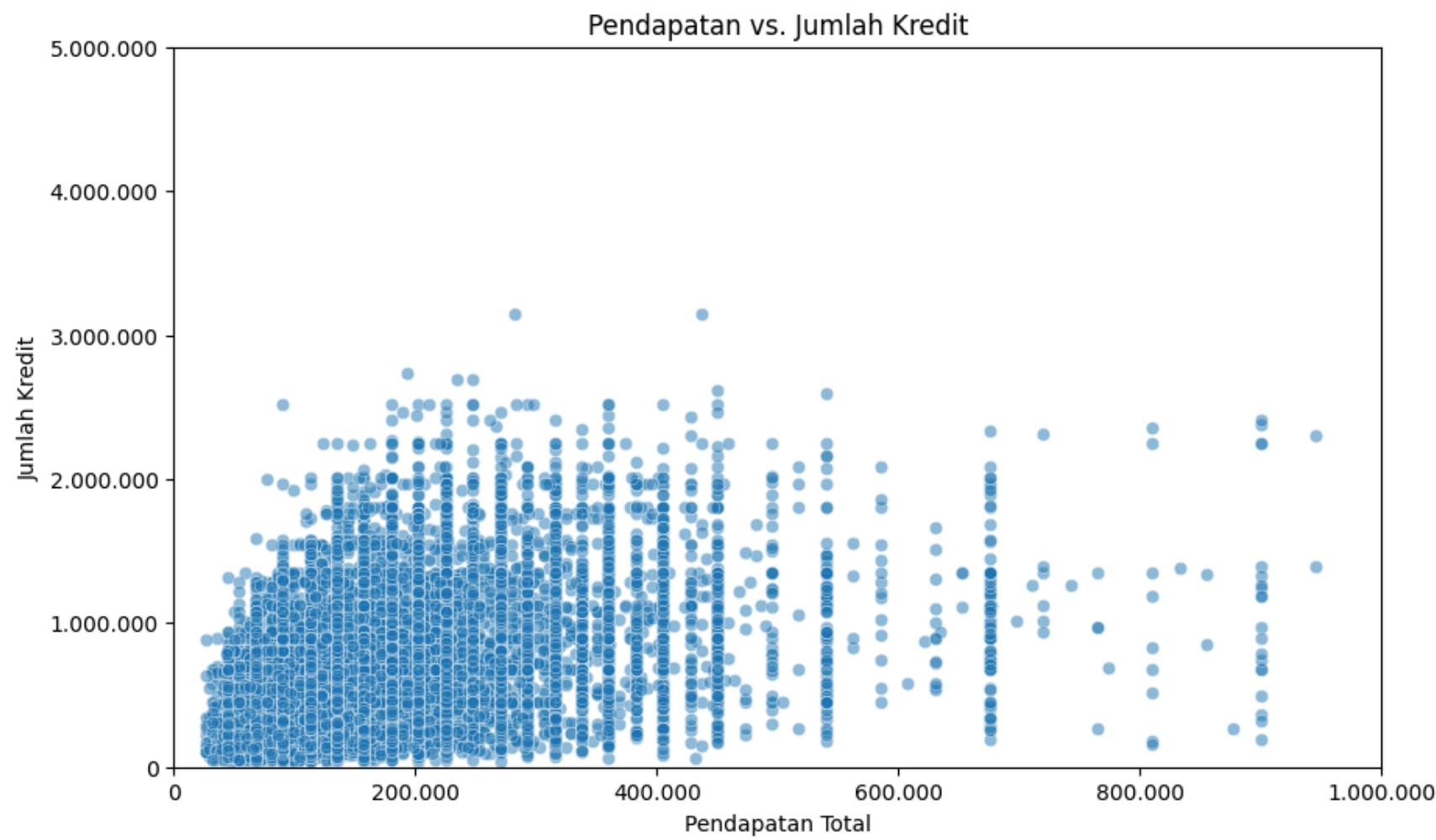


Sebanyak 8,1% klien mengalami kesulitan pembayaran, dan lama bekerja adalah salah satu faktor risikonya. Data visual menunjukkan korelasi negatif: semakin lama seseorang bekerja, semakin kecil risikonya. Contohnya, tingkat kesulitan pembayaran klien yang bekerja 1-2 tahun adalah 11-12%, menurun menjadi 9-10% pada 3-4 tahun, dan 6-7% setelah lebih dari 6 tahun bekerja. Ini adalah peluang untuk meningkatkan pangsa pasar dengan memprioritaskan klien dengan profil masa kerja yang baik guna meminimalkan risiko gagal bayar dan meningkatkan profitabilitas.

Data Visualization & Business Insight

Faktor Resiko Berdasarkan Pendapatan Debitur

Default Rate : Nilai Rata-Rata klien yang sulit membayar

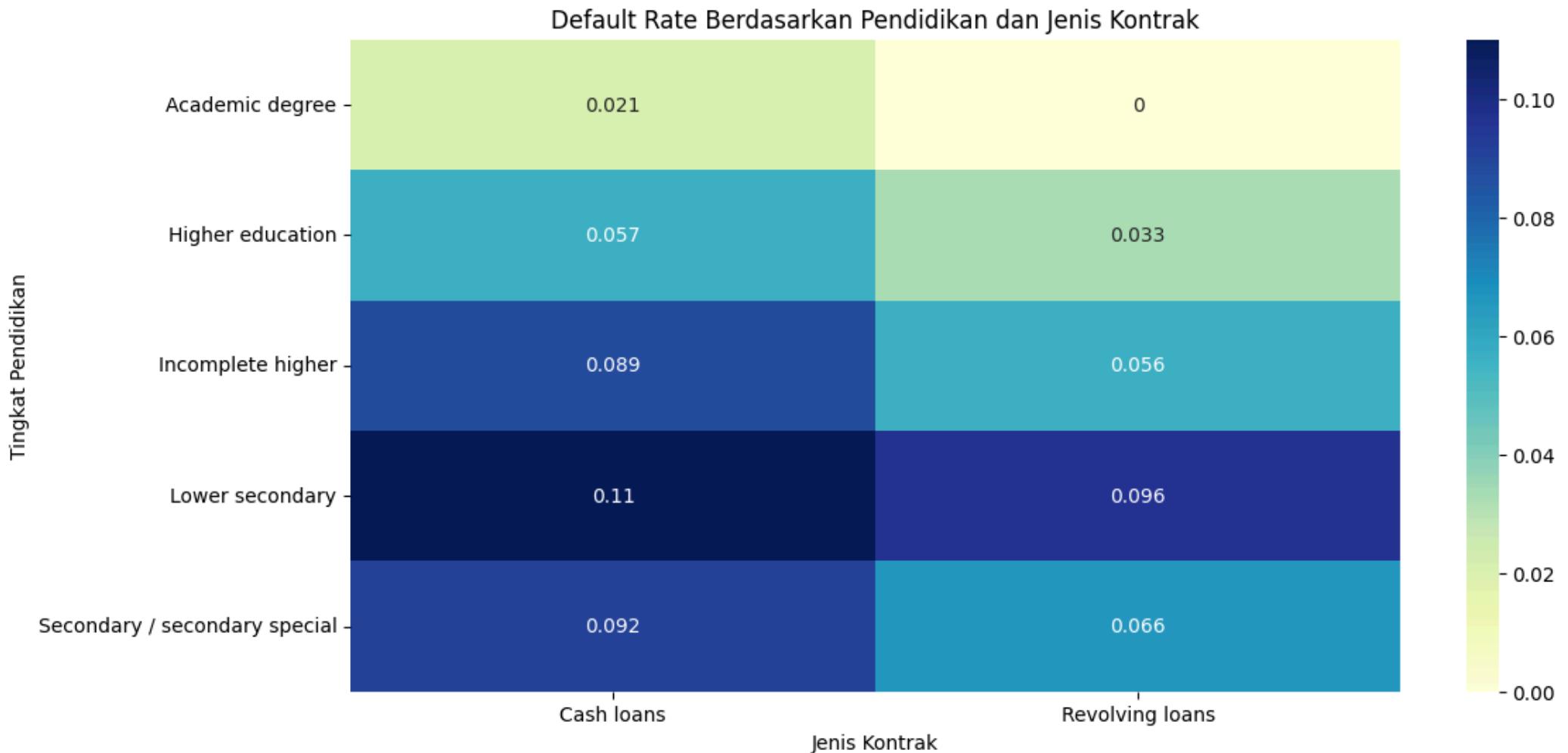
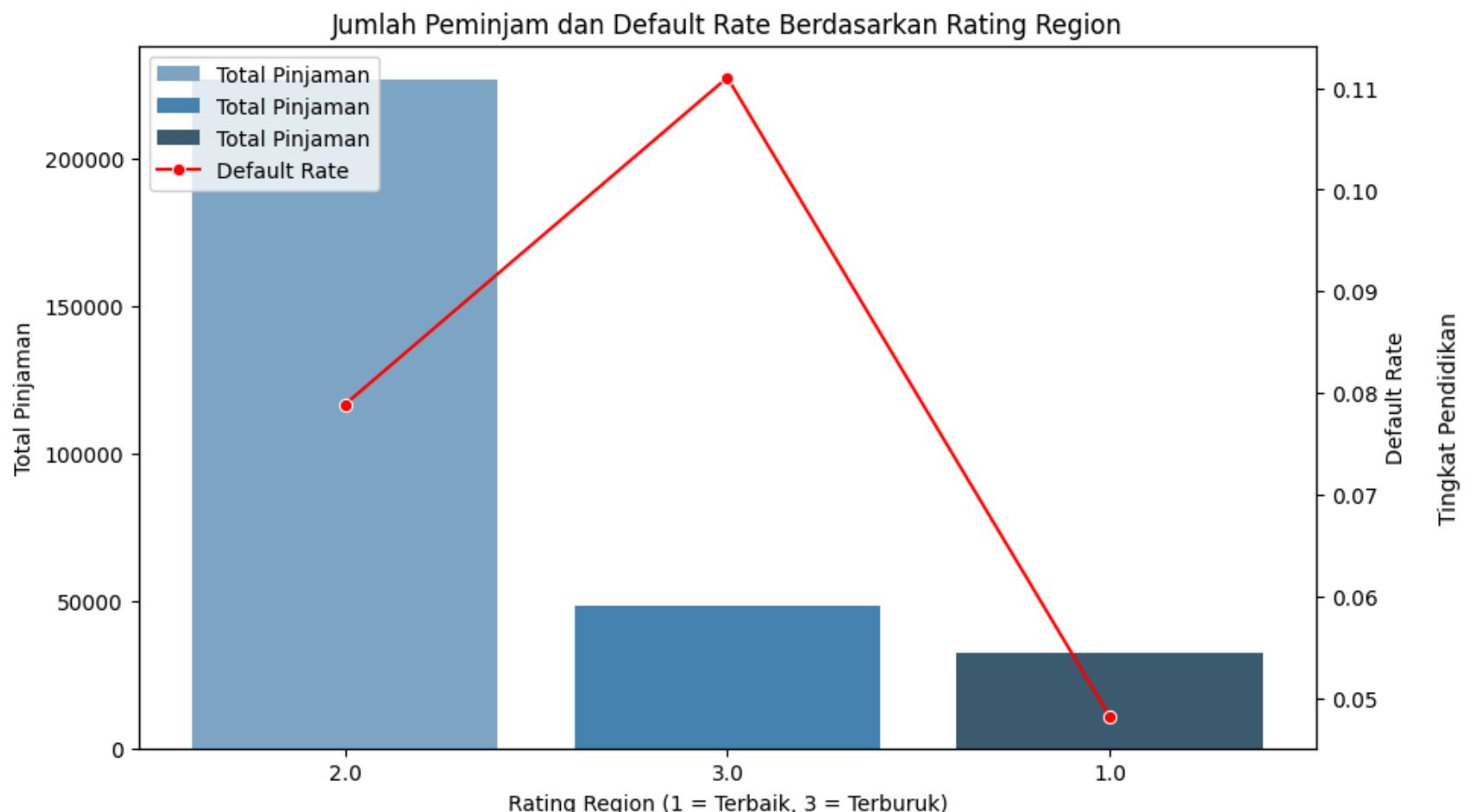


Klien dengan pendapatan rendah cenderung mengambil kredit lebih kecil. Setelah pendapatan melebihi Rp100.000, beberapa klien mulai meningkatkan jumlah kreditnya. Namun, mereka dengan pendapatan rata-rata Rp140.000-Rp180.000 menunjukkan tingkat gagal bayar tertinggi sebesar 0,16. Data ini menunjukkan korelasi antara pendapatan, jumlah kredit, dan risiko gagal bayar, serta dapat digunakan untuk menargetkan calon klien dengan profil keuangan yang lebih baik dan merancang strategi pengelolaan risiko guna mengurangi gagal bayar dan meningkatkan profitabilitas.

Data Visualization & Business Insight

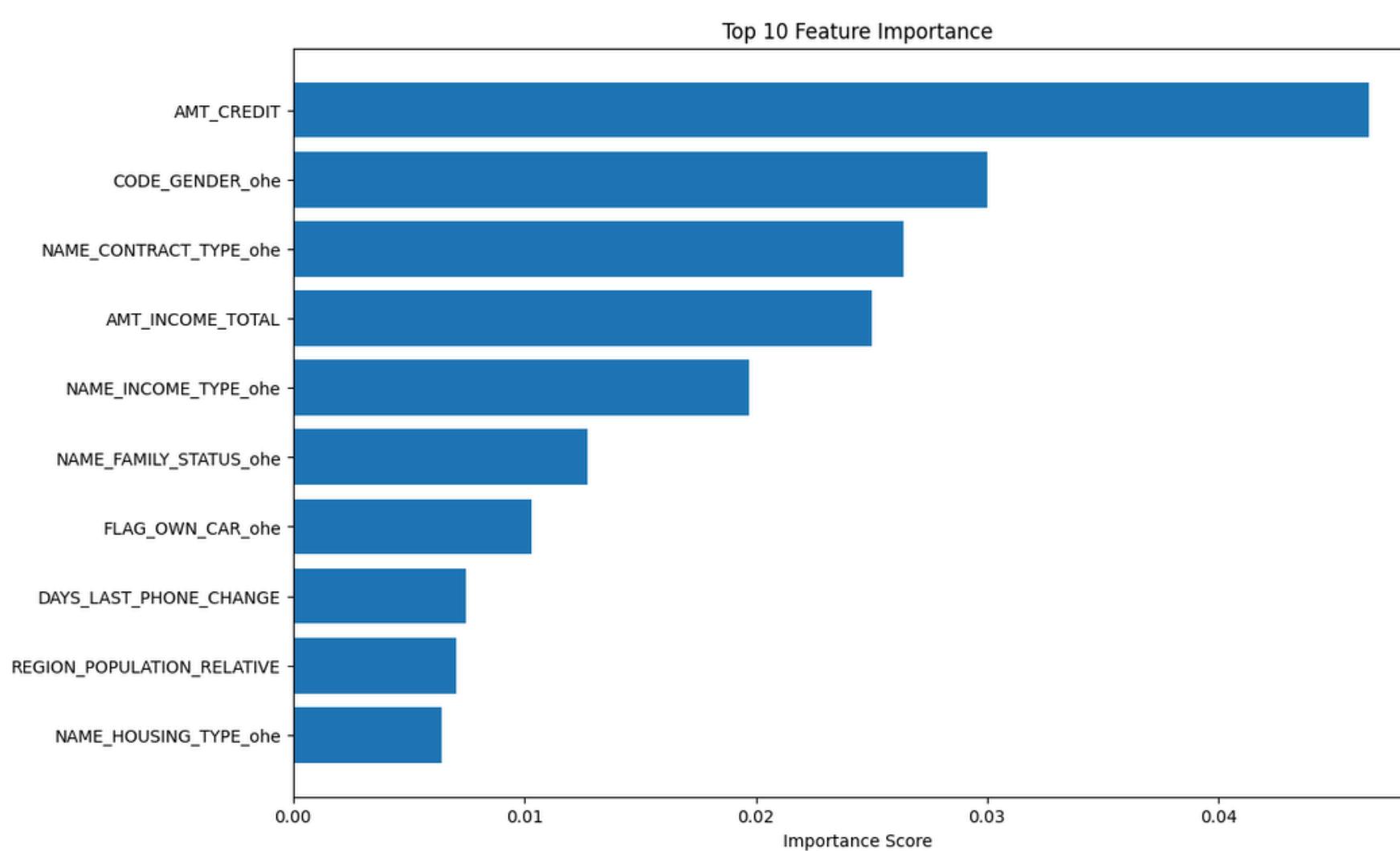
Faktor Resiko Berdasarkan Pendapatan Debitur

Default Rate : Nilai Rata-Rata klien yang sulit membayar



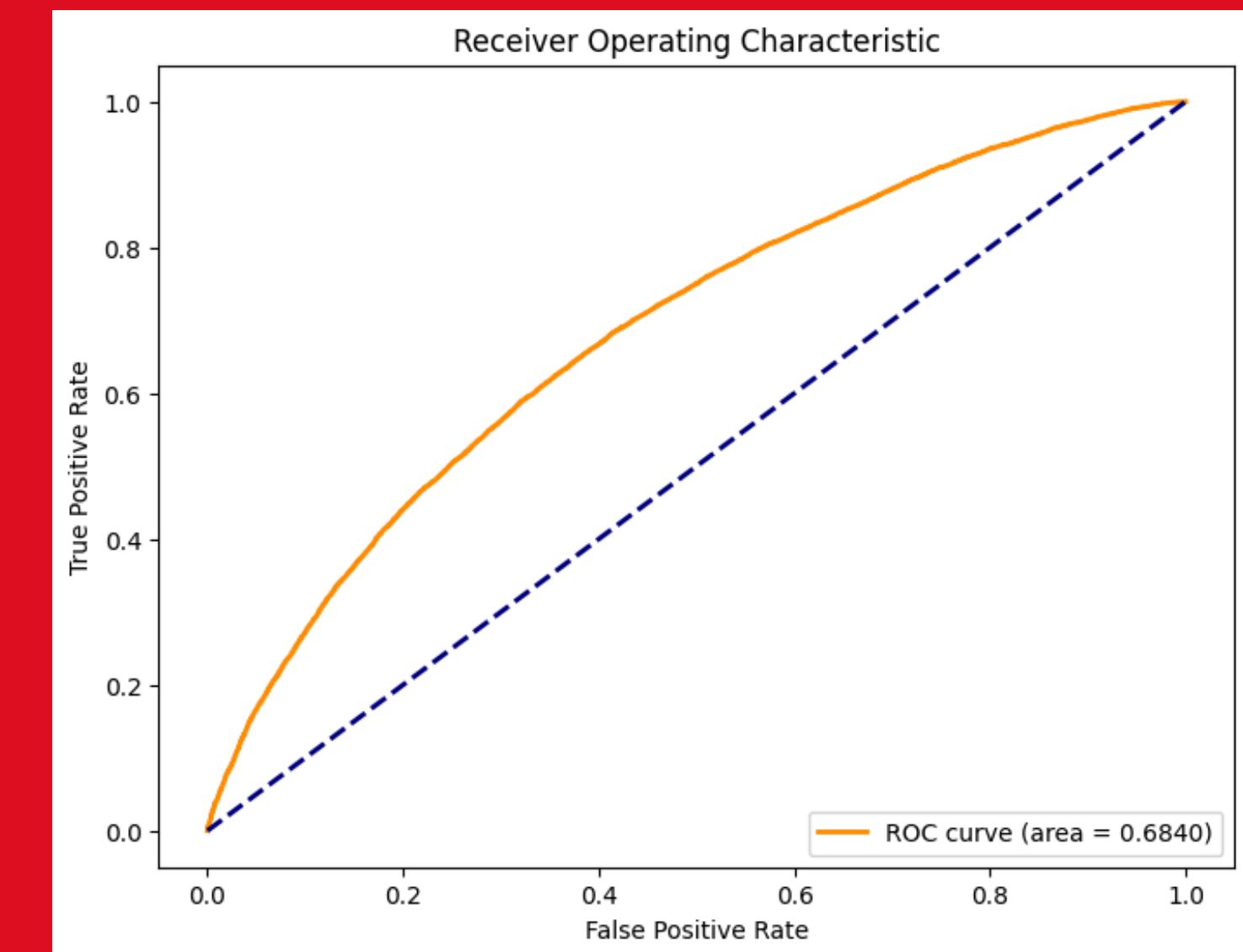
Dari grafik terlihat bahwa debitur di region dengan rating 2 mendominasi dengan jumlah di atas 200.000 debitur, sedangkan region dengan rating 3 hanya memiliki sekitar 5.000 debitur, dan region dengan rating 1 memiliki di bawah 5.000 debitur. Menariknya, region dengan rating 3 memiliki tingkat gagal bayar tertinggi di atas 0,11. Hal ini menunjukkan pentingnya mempertimbangkan penerimaan calon debitur berdasarkan region. Selain itu, tingkat pendidikan juga mempengaruhi peluang gagal bayar, di mana debitur dengan pendidikan lower secondary memiliki tingkat gagal bayar tinggi baik untuk pinjaman tunai maupun kredit bergulir. Informasi ini penting untuk menyusun strategi pengelolaan risiko yang lebih baik.

Model 1 - Faktor paling berpengaruh



Berikut adalah [Video Penjelasan](#)

Model 2 - Prediksi Resiko



Best Random Forest Model:
NumTrees: 100
MaxDepth: 7
Akurasi: 0.6840

Business Insight

1. Klien dengan masa kerja >6 tahun memiliki tingkat gagal bayar 6-7%, jauh lebih rendah dibandingkan mereka yang bekerja 1-2 tahun (11-12%), sehingga perusahaan dapat memprioritaskan debitur dengan masa kerja lebih lama melalui suku bunga lebih rendah atau skema kredit khusus bagi pekerja baru dengan mitigasi risiko tambahan.
2. Klien dengan pendapatan Rp140.000-Rp180.000 memiliki tingkat gagal bayar tertinggi (0,16), sehingga perusahaan perlu menerapkan evaluasi kredit lebih ketat di segmen ini, menyesuaikan batas kredit sesuai daya bayar mereka, serta memberikan edukasi keuangan untuk meningkatkan kemampuan mengelola utang.
3. Region rating 3 memiliki tingkat gagal bayar >0,11, tertinggi dibandingkan region lain, sehingga perusahaan perlu menerapkan kebijakan kredit lebih ketat seperti persyaratan agunan atau suku bunga berbasis risiko serta melakukan analisis lebih lanjut untuk memahami penyebabnya dan menyusun strategi mitigasi yang lebih efektif.
4. Fitur "NAME_CONTRACT_TYPE_ohe" dan "CODE_GENDER_ohe" termasuk dalam 10 faktor terpenting yang memengaruhi gagal bayar, sehingga perusahaan dapat menyesuaikan strategi pemasaran berdasarkan jenis kontrak dengan risiko lebih rendah serta mengoptimalkan skema kredit berbasis gender untuk meningkatkan inklusivitas tanpa meningkatkan risiko gagal bayar.