

CREDIT SCORING

Home Credit Indonesia Data Scientist Virtual Internship Program
Rakamin Academy



<https://github.com/Aqshalmha>

Aqshal Maulana Hafiz Ardiantha



1. LATAR BELAKANG MASALAH

Banyak orang berusaha untuk **mendapatkan pinjaman** karena sejarah kredit yang tidak ada, namun hal ini sering kali **dimanfaatkan** oleh berbagai oknum atau perusahaan **pemberi pinjaman** yang **tidak dapat dipercaya**. Home Credit hadir dengan tujuan untuk **memperluas inklusi finansial** bagi **populasi unbanked** dengan memberikan pengalaman **pinjaman** yang **aman** dan **terpercaya**. Untuk memastikan bahwa **masyarakat** yang kurang terlayani ini memiliki pengalaman **pinjaman yang positif**, Home Credit menggunakan berbagai data alternatif termasuk data telekomunikasi serta informasi transaksional. Home Credit saat ini menggunakan berbagai **metode statistika** dan **machine learning** untuk memprediksi **kemampuan pembayaran klien mereka**.



2. TUJUAN

Meminimalkan jumlah klien yang pinjamannya **disetujui** namun **sebenarnya** mereka **tidak dapat melunasi pinjaman**

3. OBJEKTIF



Membuat **model prediksi** untuk menentukan **klien potensial** dan **klien default**.



4. METRICS BISNIS

Loss given default (LGD)

Loss given default (LGD) adalah perkiraan jumlah uang yang hilang dari lembaga peminjam ketika peminjam gagal membayar pinjaman.

5. MODEL EVALUATION

Accuracy dan nilai Area dibawah kurva ROC.



DATASET

HOME
CREDIT

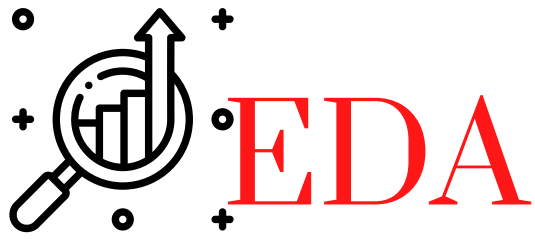
 **Rakamin**
Academy

Terdapat **3 Data** yang digunakan untuk **data training** pada **model** yang akan digunakan :

- **application_train.csv** : data untuk aplikasi kredit saat ini dan variabel data TARGET. Setiap pinjaman memiliki barisnya sendiri dan diidentifikasi oleh feature SK_ID_CURR.
- **bureau.csv** : Semua kredit klien sebelumnya disediakan oleh lembaga keuangan lain yang dilaporkan ke Biro Kredit untuk setiap pinjaman dalam sampel. Ada baris sebanyak jumlah kredit yang dimiliki klien di Biro Kredit sebelum tanggal pengajuan.
- **previous_application.csv** : Semuaajuan pinjaman sebelumnya untuk Home Credit loans dari klien yang memiliki pinjaman dalam sampel. Ada satu baris untuk setiapajuan sebelumnya yang terkait dengan pinjaman dalam sampel data.

Satu data target untuk **data target** pada **model** yang akan digunakan :

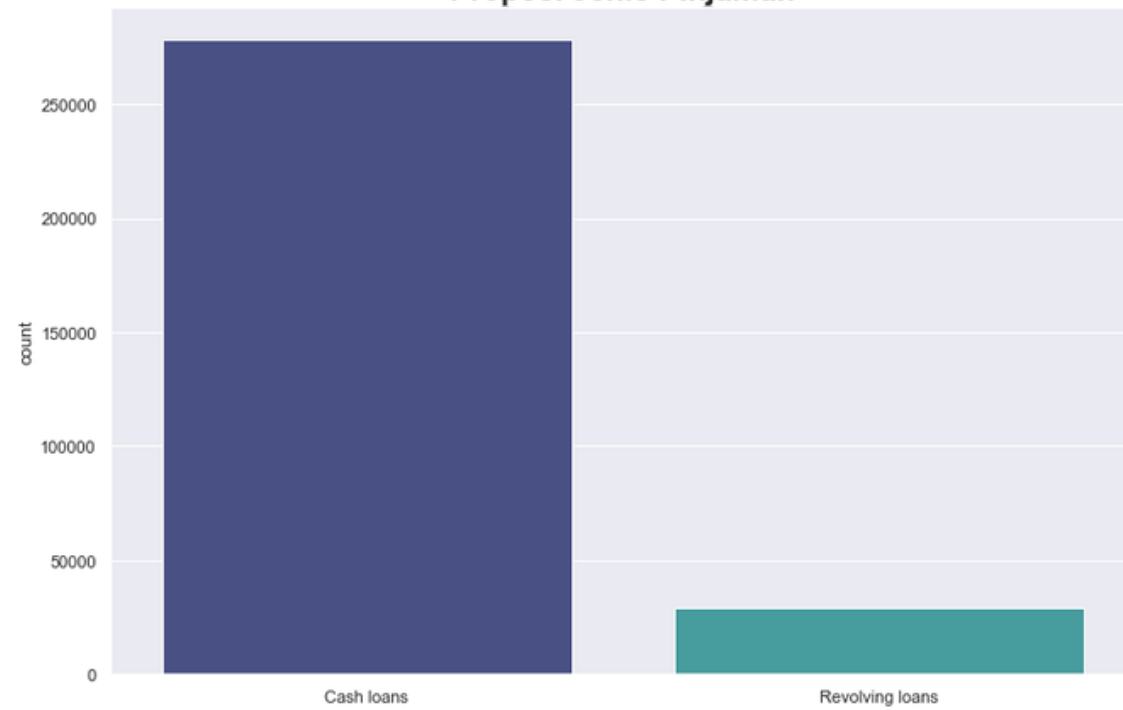
- **application_test.csv** : dataset yang akan digunakan untuk memprediksi probabilitas untuk variabel TARGET.



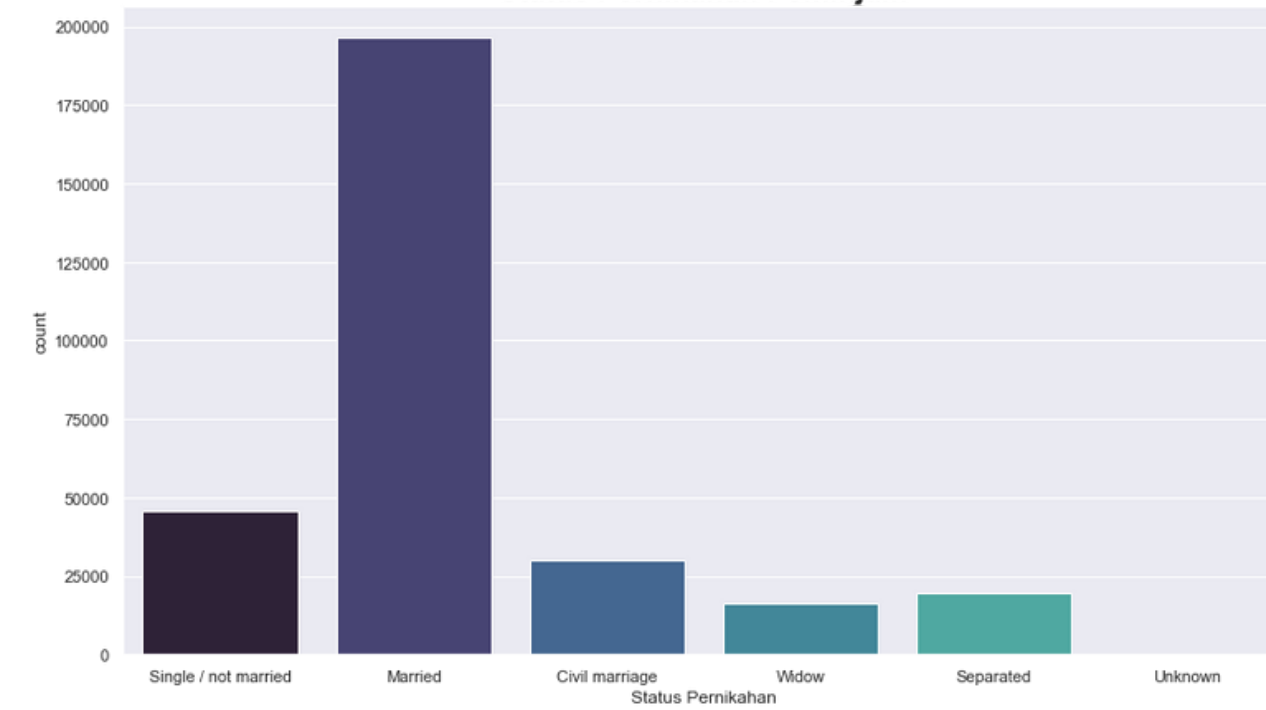
HOME
CREDIT

Rakamin
Academy

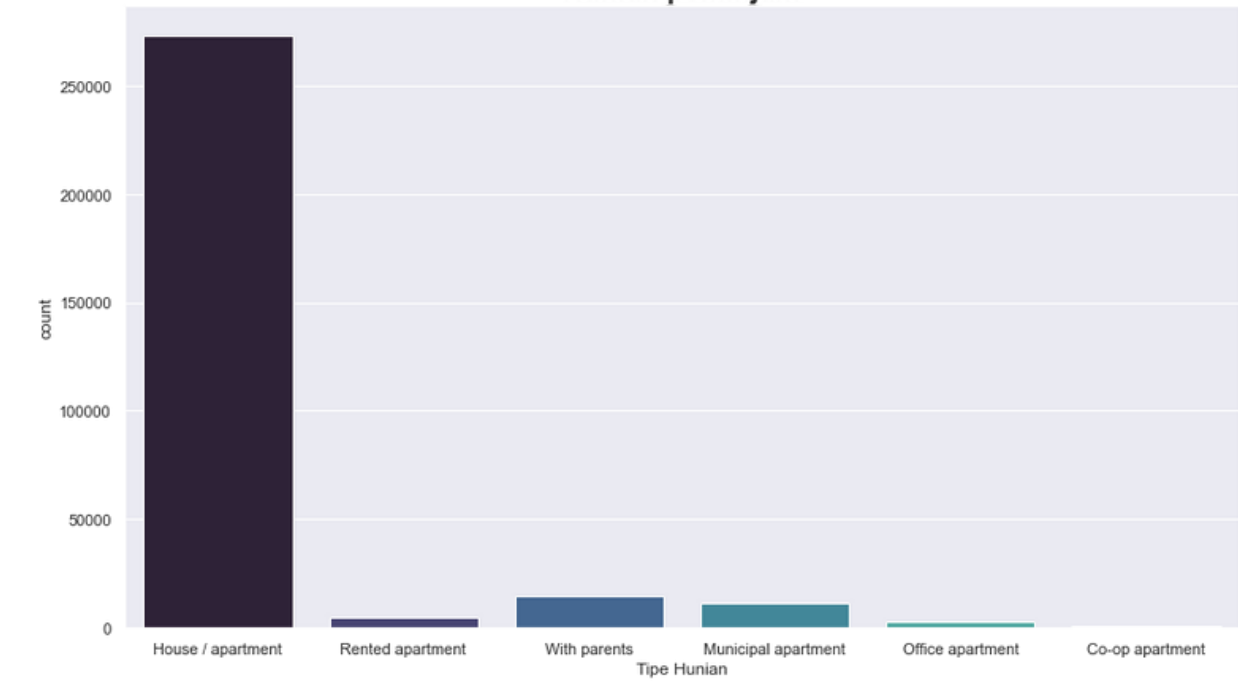
Proposi Jenis Pinjaman



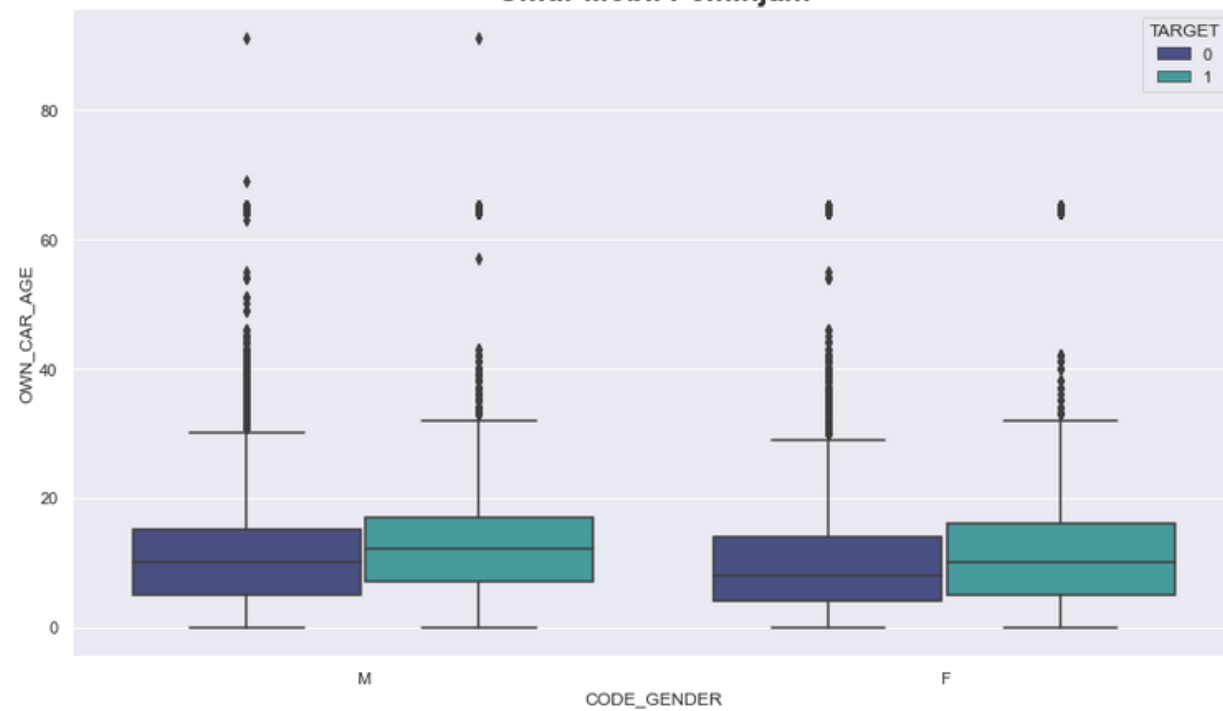
Status Pernikahan Peminjam



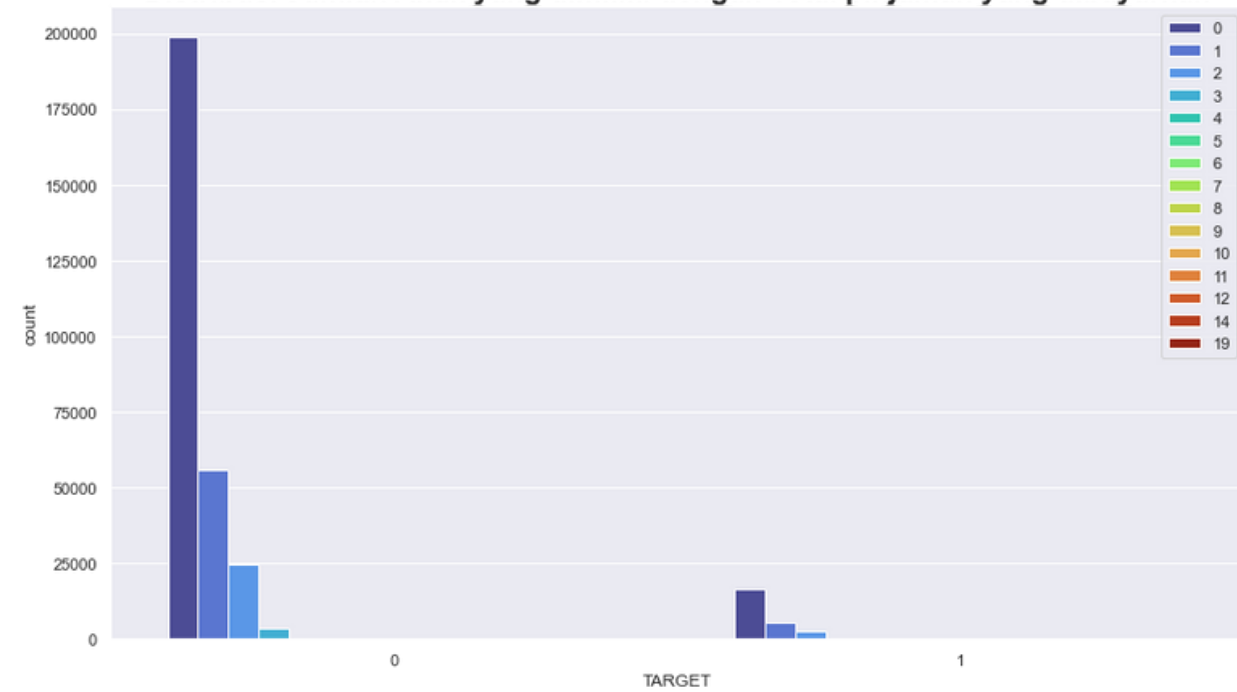
Hunian peminjam



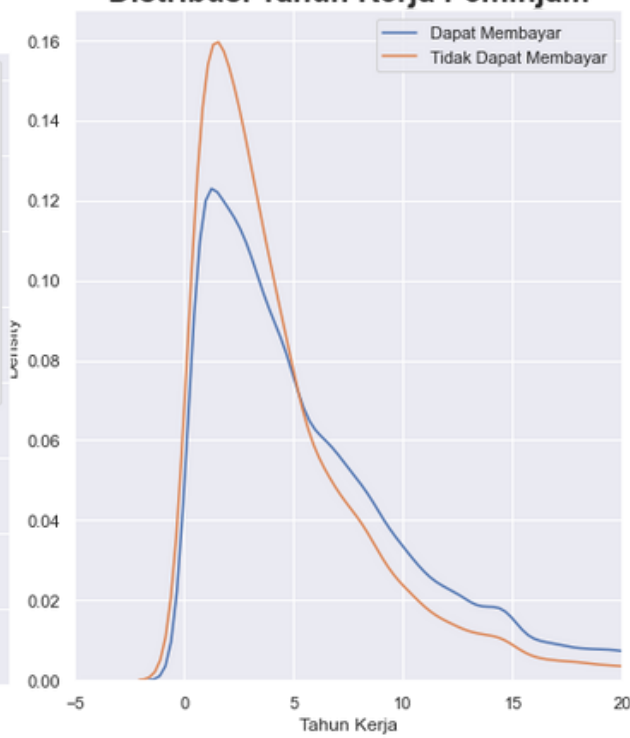
Umur Mobil Peminjam



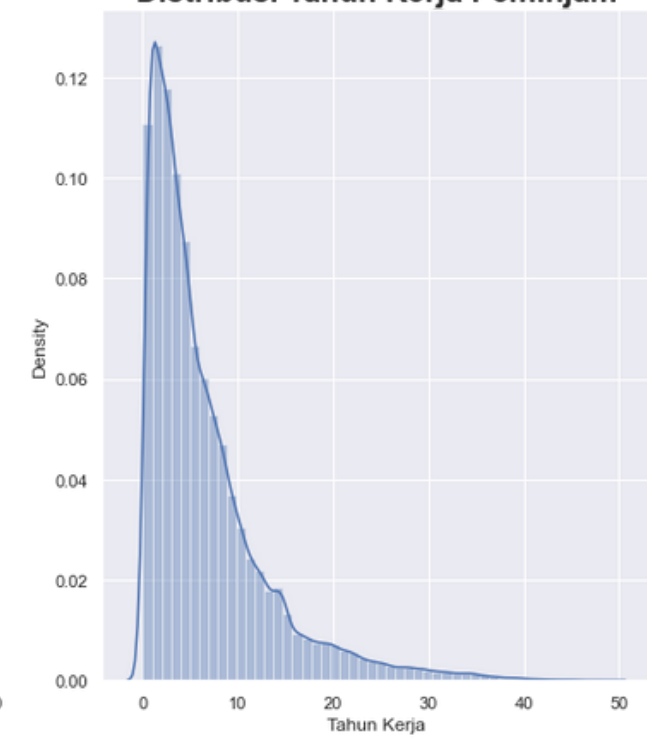
Distribusi Jumlah Anak yang dimiliki dengan Total pinjaman yang dibayarkan

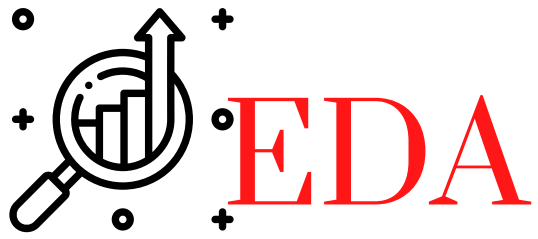


Distribusi Tahun Kerja Peminjam



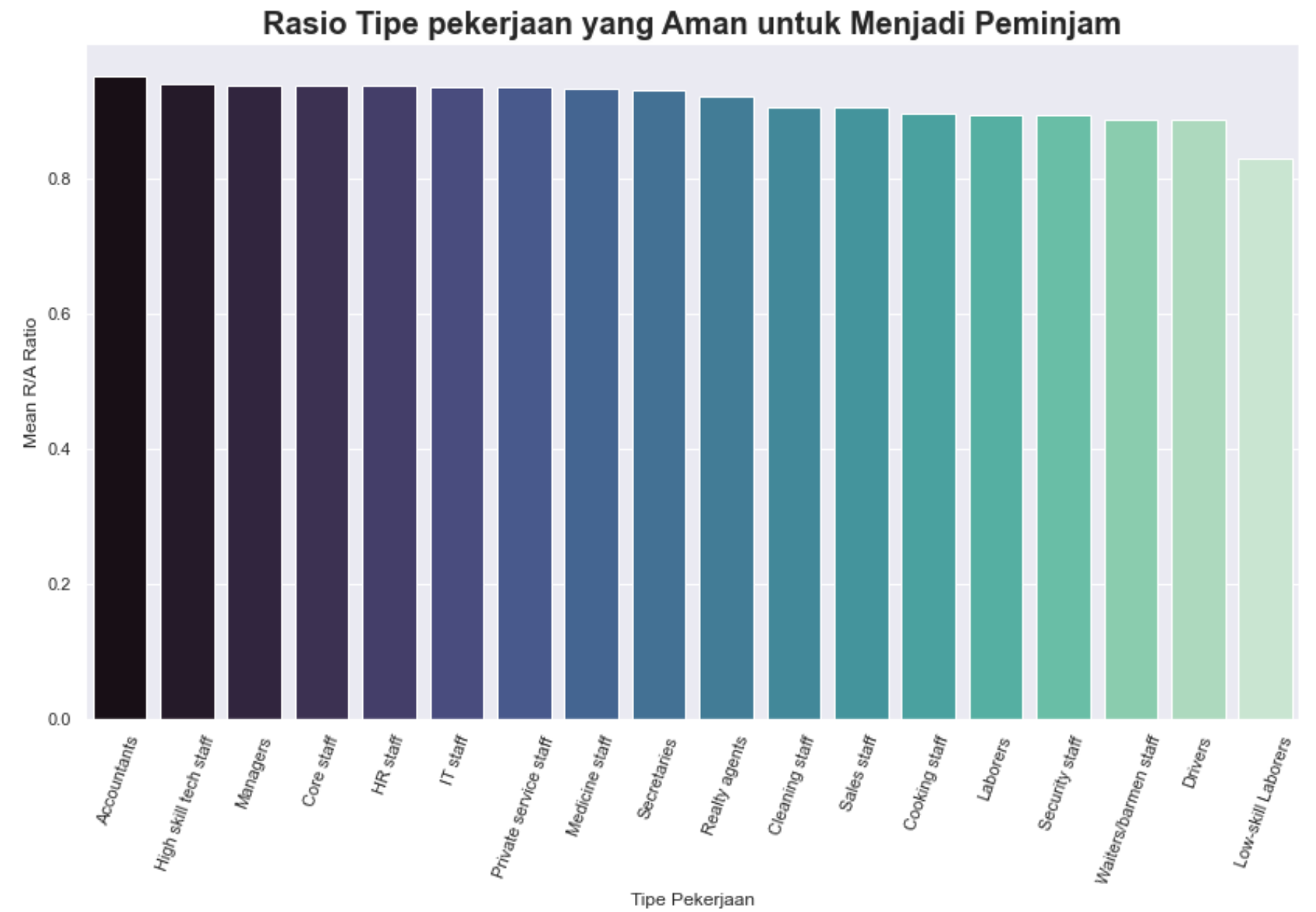
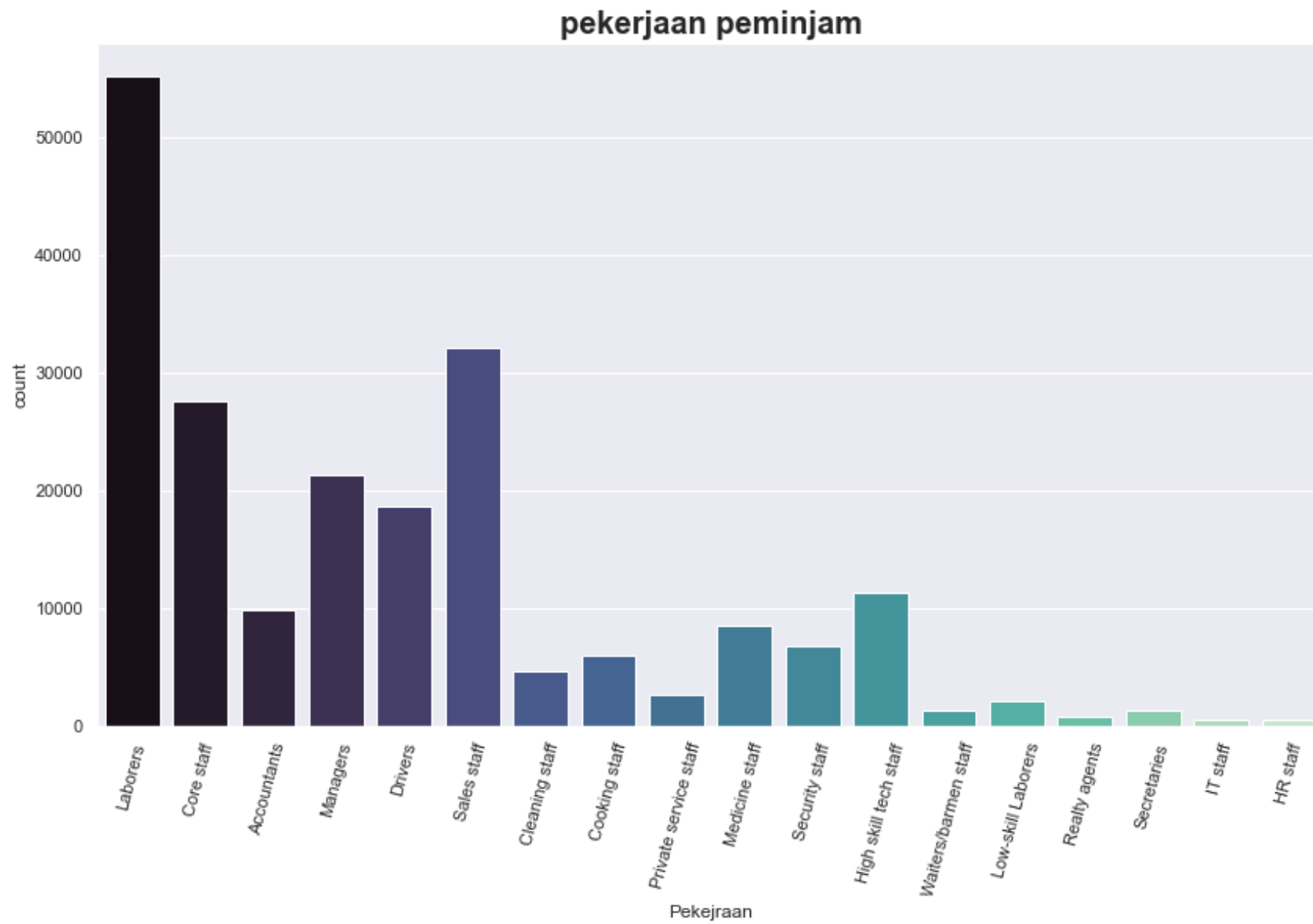
Distribusi Tahun Kerja Peminjam





HOME
CREDIT

Rakamin
Academy



- Pekerjaan **Managers** dan **core staff** merupakan top 5 untuk total peminjam yang mengajukan pinjaman, dan 5 **teratas** dalam hal kemampuan untuk **membayar kembali ajuan pinjaman**,
- Pekerjaan **Buruh (Laborers)** sejauh ini jumlah **pengaju pinjaman** yang **paling banyak**, tetapi mereka tampaknya memiliki lebih banyak **kesulitan** untuk **membayar kembali pinjaman**.
- Pekerjaan **Accountants** merupakan pekerjaan yang bisa dibilang sebagai **peminjam yang aman (dapat membayarkan kembali pinjaman mereka)**, tetapi mereka **tidak masuk** dalam **5 pekerjaan teratas** sebagai **pekerjaan yang mengajukan pinjaman**.
- Kebanyakan peminjam yang tidak memiliki anak tidak mengalami kesulitan untuk membayar kembali pinjaman mereka, namun **memiliki anak** sangat terlihat **memperhambat peminjam** untuk **membayarkan pinjamannya**.
- Peminjam dengan **masa kerja < 5 tahun**, terutama pekerja dengan **masa kerja < 2 tahun** cenderung untuk **tidak dapat membayarkan pinjaman** yang ada, sebaliknya peminjam dengan **masa kerja > 5 tahun** dapat **membayarkan pinjamannya**.
- Peminjam yang **sulit membayar pinjaman** mereka kebanyakan memiliki **mobil yang sudah tua**, hal ini mungkin bisa menjadi asumsi jika mereka **mengeluarkan uang** mereka untuk **menyewa mobil** karena **mobil** yang mereka miliki **sudah tua**, atau bahkan mereka kebanyakan **menghabiskan uang** mereka untuk **memperbaiki mobil tua** mereka.



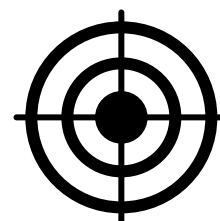
DATA CLEANING AND DATA PREPROCESSING

1. **Cleansing data** yang meliputi :

- **Handle Duplicated Data** dengan menghapus data yang berduplikat.
- **Handle Missing Value** dengan mean, modus, dan median sesuai dengan distribusi data setiap variabel
- **Handle Missing Value** menghapus feature yang memiliki missing value kosong $> 30\%$, dan feature yang terlalu banyak memiliki unique value
- **Handle Outlier** menggunakan Z-Score atau IQR

2. **Scaling Numerikal Feature** menggunakan MinMaxScaler

3. **OneHot Encoding dan Label Encoding** pada data kategorikal



MODELING AND MODEL EVALUATION

Ada **6 Model** yang digunakan pada problem kali ini :



MODEL	MODEL PERFORMANCE	
	Accuracy	AUC
XGBoost	0.894	0.74
ADABOOST	0.876	0.72
Random Forest	0.873	0.72
Decision Tree	0.747	0.56
Logistic Regression	0.761	0.66

Dapat dilihat dari model evaluasi dengan nilai **Accuracy** dan **AUC**, dapat dikatakan jika model yang terbaik adalah model **XGBoost**.



METICS BISNIS

Loss given default (LGD)

Loss given default (LGD) adalah perkiraan jumlah uang yang hilang dari lembaga peminjam ketika peminjam gagal membayar pinjaman.

Asumsi : kerugian karena default yang hanya dihitung dari total kredit klien.

XGBoost Model	Before Modeling	After Modeling
Total Peminjam yang Tidak Bisa Membayarkan Pinjamannya	4984	3875
	(Dengan data target = 1)	Total False Negatif
Total Kredit dari Peminjam yang Tidak Bisa Membayarkan Pinjamannya	1,132,249,716.16	909,029,970
Penurunan LGD	- 2,234,667.15	
	-19,73%	



KESIMPULAN DAN SARAN

- Untuk **mempertahankan klien** yang bekerja sebagai **buruh (Laborers)**, Home Credit dapat merancang **produk pinjaman** yang **lebih ramah** bagi mereka, misalnya dengan **menurunkan tingkat bunga**, dan **meningkatkan periode pembayaran**, serta lebih menekankan pada **pendapatan tahunan** mereka sebelum **menyetujui pinjaman** yang diajukan.
- **Mempertahankan klien** yang bekerja sebagai **Managers** dan **Core Staff** dengan memberi produk pinjaman yang spesial untuk mereka, seperti **potongan bunga pinjaman**, serta **bundle-bundle** dari **produk** yang **menguntungkan mereka**.
- Home Credit dapat **mengarahkan** dan **menyusun strategi campaign** untuk **mengakomodasi** lebih banyak **akuntan (Accountants)**, **HR Staffs**, dan **IT Staffs**, untuk **mengajukan** lebih **banyak pinjaman**.
- **Model terbaik** yang didapat sebenarnya masih **belum** memiliki **performa yang sangat baik**. Meskipun telah mampu **menurunkan LGD** sebesar **19,73%**, namun masih memungkinkan untuk dapat **meningkatkan performa model** saat ini dengan **menambahkan fitur penting** lainnya dari **dataset lain** selain yang dipilih pada pemodelan **saat ini**.

THANK YOU