

# HOME CREDIT SCORECARD MODEL

Bobby Kurniadi Widodo



# PROJECT BACKGROUND

Banyak orang berjuang untuk mendapatkan pinjaman karena kredit yang tidak mencukupi atau tidak ada Sejarah. Kredit Rumah berupaya memperluas inklusi keuangan bagi yang tidak memiliki rekening bank populasi dengan memberikan pengalaman meminjam yang positif dan aman. Untuk membuat yakin populasi yang kurang terlayani ini memiliki pengalaman pinjaman yang positif. Kredit Rumah membuat berbagai data alternatif untuk memprediksi kemampuan pembayaran klien mereka. Melakukan jadi akan memastikan bahwa klien yang mampu membayar kembali tidak ditolak dan bahwa pinjaman adalah diberikan dengan pokok, jatuh tempo, dan kalender pelunasan yang akan memberdayakan mereka klien untuk menjadi sukses.



## PROBLEM RESEARCH

## DATA SOURCE, OBJECTIVE, ACTIONS

### DATA SOURCE

Data yang digunakan adalah kereta aplikasi dan tes aplikasi. Ada meja utama kami, dipecah menjadi dua file untuk kereta api (dengan TARGET) dan uji (tanpa TARGET).

### OBJECTIVE

1. Identifikasi karakteristik klien potensial yang akan mengalami kesulitan membayar kembali pinjaman dan siapa yang tidak mau.
2. Memprediksi kemampuan pembayaran klien.

### ACTIONS

1. Lakukan pembersihan data, dan visualisasi untuk bisnis Wawasan.
2. Bangun model dengan pembelajaran mesin Algoritma.
3. Berikan rekomendasi agar perusahaan meningkat klien mereka berhasil dalam mengajukan pinjaman.





# DATA PREPROCESSING DATA TRAIN

EDA  
Temukan pola, dan  
struktur  
Dataset

Bivariat  
Visualisasi

Visualisasi dari hubungan  
antara 2 Fitur

Multivarian  
Visualisasi

Visualisasi dari  
hubungan lebih banyak  
dari 2 fitur

DATA CLEANING

Mendeteksi Duplikasi  
Tidak ada baris  
duplikat

Menangani Nilai yang Hilang

Ada beberapa kolom yang  
dijatuhkan dan sisanya  
diperhitungkan

Mendeteksi Outlier

Ada beberapa kolom yang  
memiliki outlier, tetapi diputuskan  
outlier tidak akan dihapus

MODEL BUILDING

Pengkodean Label

Pemilihan Fitur

Menangani Data yang  
Tidak Seimbang

Membangun Model

Evaluasi Model



# DATA PREPROCESSING DATA TEST

## DATA CLEANING

Mendeteksi Duplikasi  
Tidak ada baris  
duplikat

### Menangani Nilai yang Hilang

Ada beberapa kolom yang  
dijatuhkan dan sisanya  
diperhitungkan

### Pengkodean Label

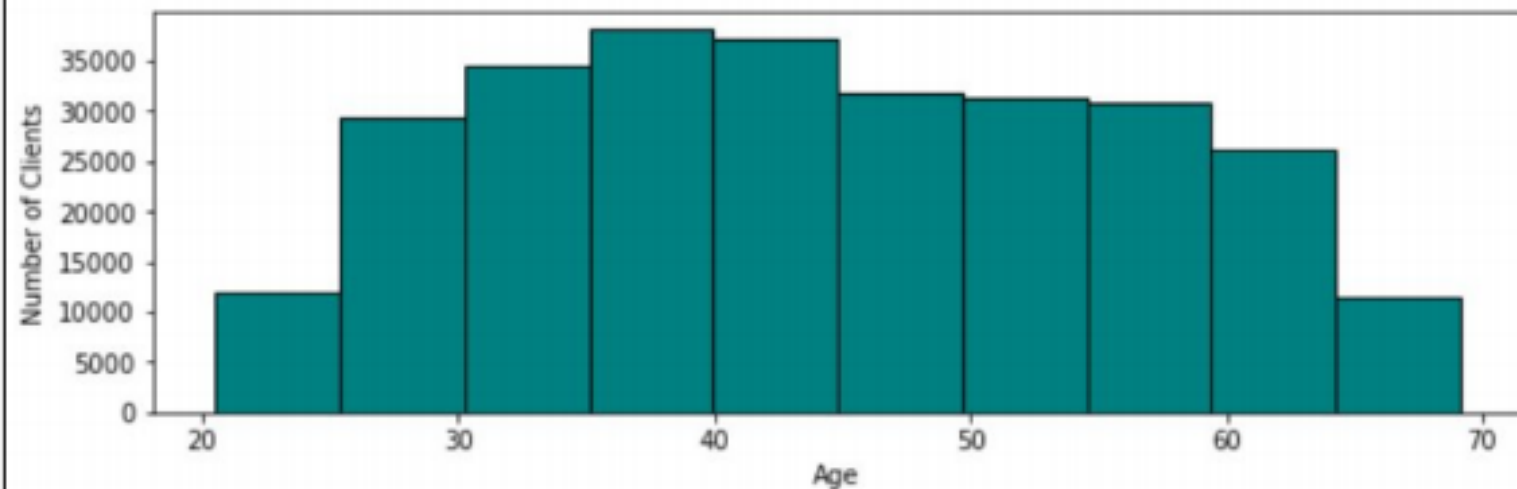
Ubah non-numerik menjadi  
numerik Label

## PREDICTION

Prediksi kemampuan  
pembayaran klien  
dengan yang terbaik  
model pembelajaran  
mesin yang diperoleh  
sebelumnya

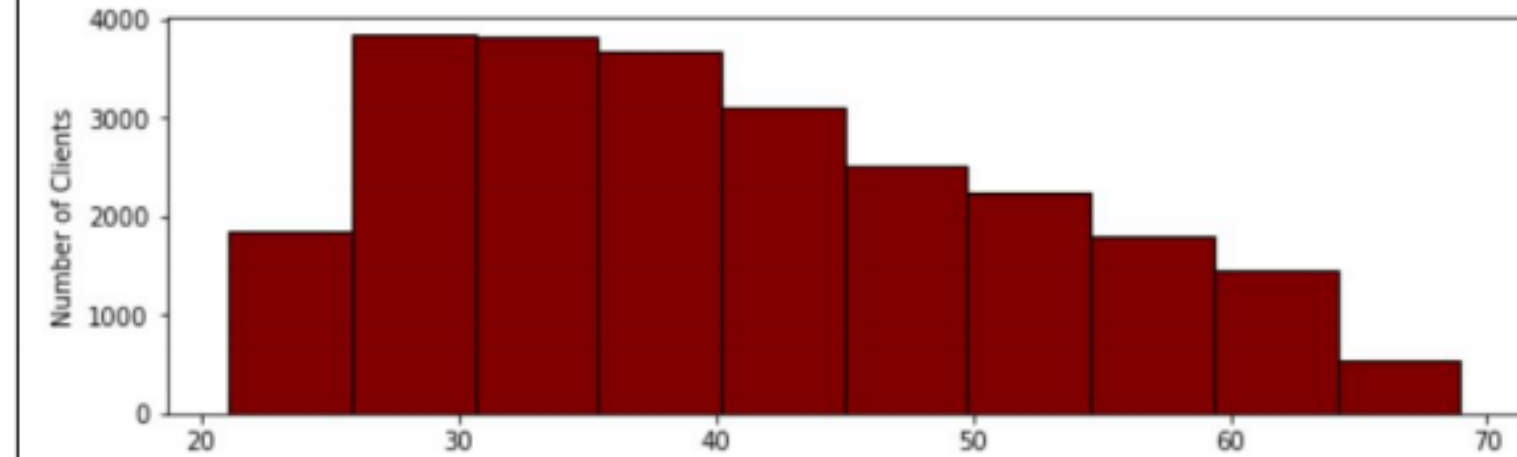
# BUSINESS INSIGHTS

Age of Client (in years) who have No Payment Difficulties



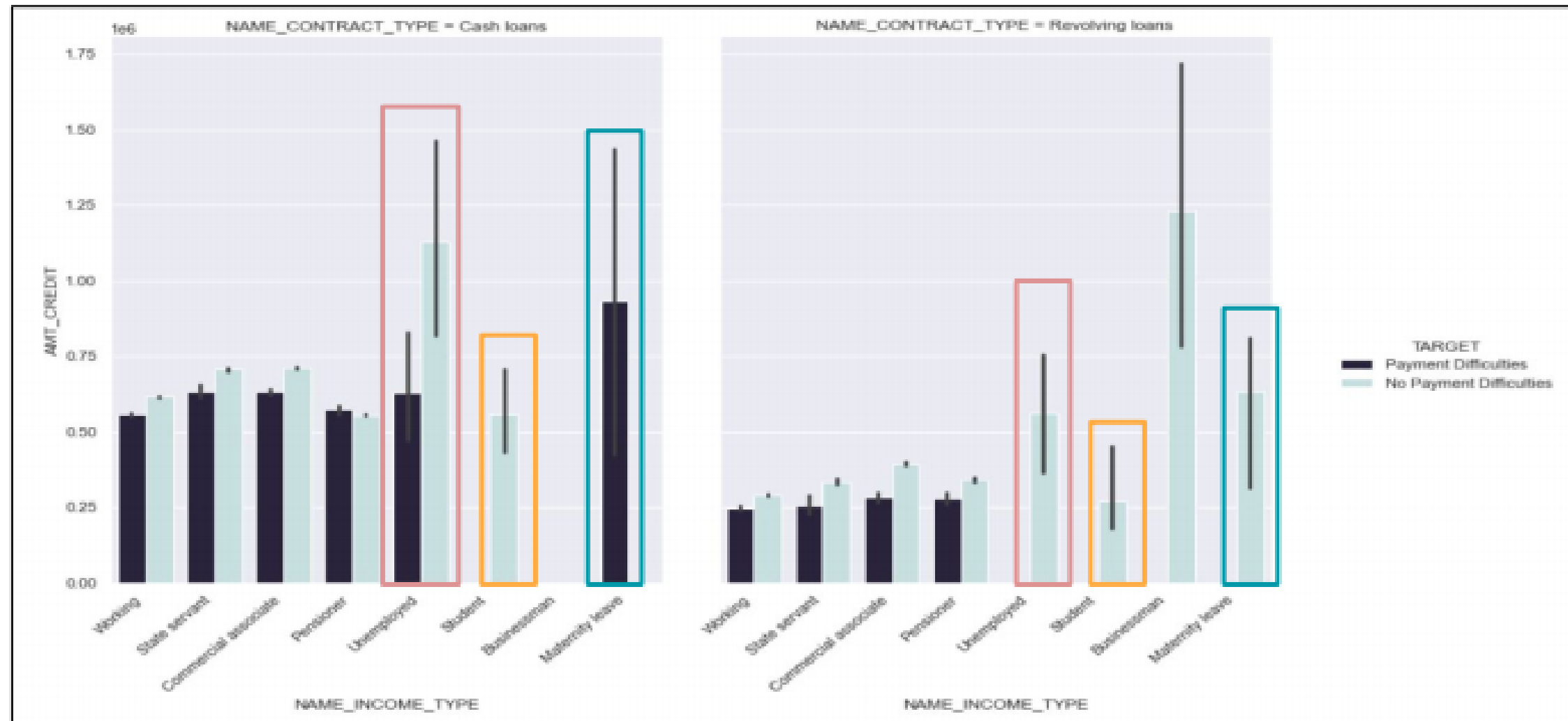
- Sebagian besar jumlah klien yang ajukan pinjaman berada di kisaran 35-40 tahun.
- Sementara itu, jumlah pelamar untuk klien berusia <25 tahun atau usia >65 tahun sangat rendah.

Age of Client (in years) who have Payment Difficulties



- Klien yang tidak memiliki pembayaran kesulitan adalah klien di kisaran 35-45 tahun. Kamu bisa targetkan klien ini sebagai milik Anda Prioritas.
- Sementara klien yang memiliki kesulitan pembayaran adalah klien kisaran 25-35 tahun.

# BUSINESS INSIGHTS



Semua klien siswa tidak mengalami kesulitan membayar kembali pinjaman apakah dengan uang tunai pinjaman atau bergulir pinjaman untuk yang rendah ke kredit menengah jumlah pinjaman.

- Untuk jenis pendapatan cuti hamil dengan pinjaman tunai, semua klien memiliki masalah dalam membayar kembali pinjaman untuk jumlah kredit menengah dari pinjaman. Sementara semua klien dengan cuti hamil dan pinjaman bergulir tidak mengalami kesulitan membayar kembali pinjaman.
- Untuk klien yang menganggur dengan pinjaman tunai, lebih dari 50% klien memiliki masalah dalam membayar kembali pinjaman dengan media jumlah kredit pinjaman. Sementara semua klien yang menganggur dengan pinjaman bergulir tidak mengalami kesulitan untuk membayar kembali pinjaman tersebut.



Algoritma	Akurasi Skor Training	Akurasi Skor Test	Error Margin	Skor ROC
Logistic Regression	67.16%	67.29%	0.13%	0.6728
Gaussian Naive Bayes	60.24%	60.39%	0.15%	0.604
Decision Tree	100%	83.9%	11.74%	0.8826
Random Forest	100%	99.65%	0.35%	0.9965
K-Nearest Neighbor	91.56%	88.07%	3.79%	0.8806
Neural Network	70.01%	69.48%	0.58%	0.6948

Akurasi prediksi kereta api dan data pengujian dalam model Random Forest memiliki nilai yang tidak banyak berbeda, dapat dikatakan bahwa modelnya sangat bagus, yaitu tidak ada underfitting atau overfitting. Jadi model Random Forest dipilih sebagai model terbaik untuk memprediksi kemampuan pembayaran klien.

# MACHINE LEARNING MODEL



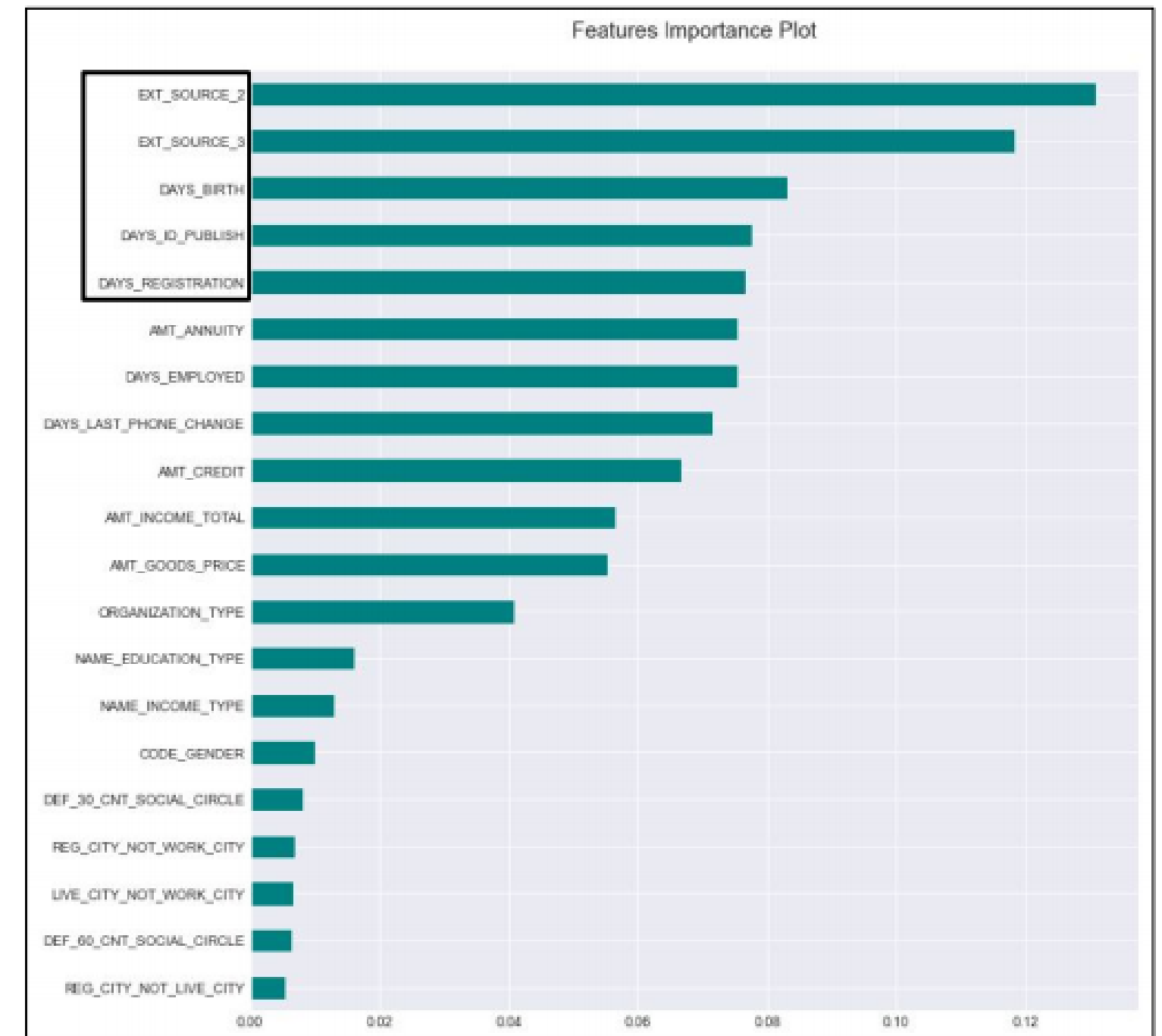


# BEST MODEL

- Algorithm : Random Forest Classifier
- Performance : Random forest model gives 100% correct results Performance, There is 0.35% error margin

The 5 most important features :

- Score from external data source 2
- Score from external data source 3
- Client's age in days
- Days ID publish
- Days registration



# BUSINESS RECOMMENDATION

Inilah cara kita mencapai sasaran dengan platform baru yang lebih terintegrasi.

- **Buat kampanye sehingga lebih banyak siswa, akuntan, staf teknologi keterampilan tinggi, manajer yang tertarik untuk mengajukan pinjaman.**

- **Perlu analisis lebih lanjut, Anda dapat mensurvei untuk mengetahui apakah ada masalah jika klien dengan cuti hamil atau pengangguran mengambil kontrak pinjaman tunai. Jadi, di masa depan, jika ada klien dengan jenis pendapatan itu, Anda dapat merekomendasikan jenis kontrak yang tepat sehingga aplikasi mereka akan disetujui**

- **Klien yang tidak memiliki kesulitan pembayaran adalah klien dalam kisaran 35-45 tahun. Anda dapat menargetkan klien ini sebagai prioritas Anda.**

**Link github : <https://github.com/bobbykurniadi/Home-Credit-Score-Card-Model>**