



Home Credit Scorecard Model

By Muhammad Farras Rizki

LinkedIn [linkedin.com/in/mfarrasrizki](https://www.linkedin.com/in/mfarrasrizki)

GitHub https://github.com/FarrasRizki/HomeCredit_ScorecardModel

Latar Belakang



Manajemen Home Credit Indonesia ingin membuat membuat prediksi skor kredit. Dengan tujuan, dapat memastikan pelanggan yang mampu melakukan pelunasan tidak ditolak ketika melakukan pengajuan pinjaman, dan pinjaman dapat diberikan dengan principal, maturity, dan repayment calendar yang akan memotivasi pelanggan untuk sukses.

Tujuan Proyek

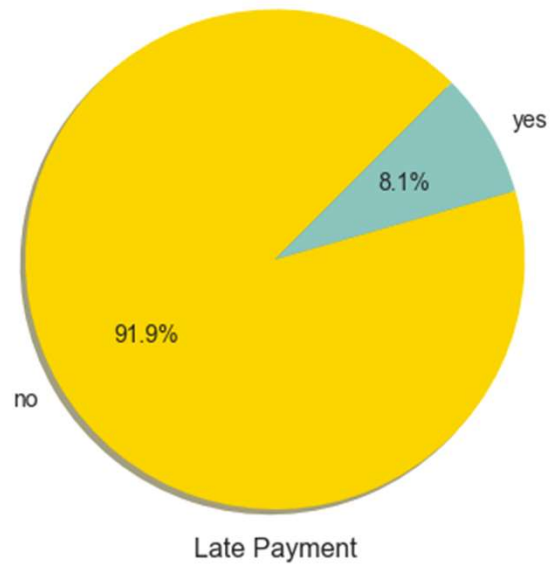
Membuat Model Machine Learning yang dapat memberikan sebuah score kelayakan pada client dalam memperoleh pinjaman.

Sumber Dataset

<https://rakamin-lms.s3.ap-southeast-1.amazonaws.com/vix-assets/home-credit-indonesia/home-credit-default-risk.zip>

Target Variabel

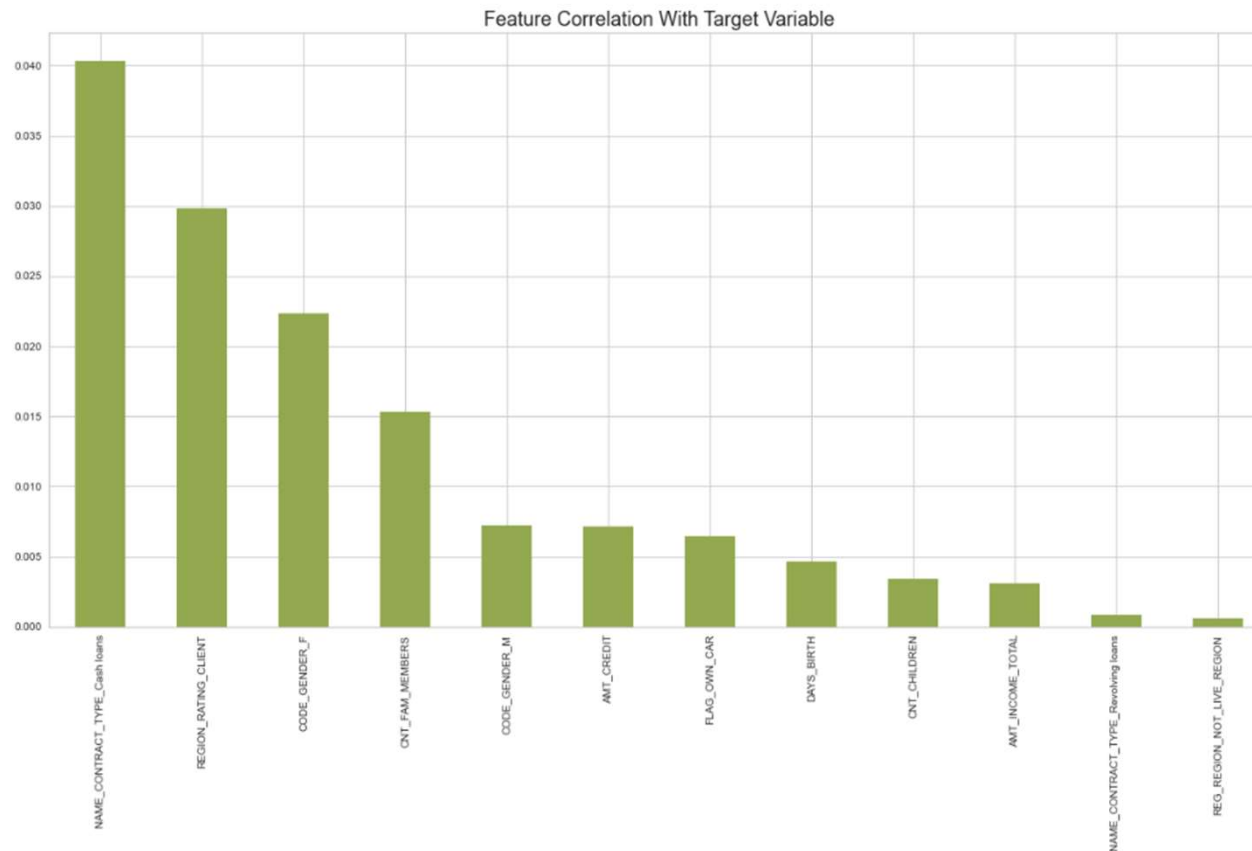
Persentase Pelanggan Berdasarkan Nilai TARGET



Target Variabel dari penelitian ini adalah 'TARGET' yaitu status akhir hasil pengajuan pinjaman nasabah sebelumnya, apakah mengalami gagal bayar (keterlambatan pembayaran) atau tidak.

Di antara 307511 pelanggan, terdapat 24825 pelanggan (8,1%) yang mengalami keterlambatan dalam pembayaran kredit (gagal bayar)

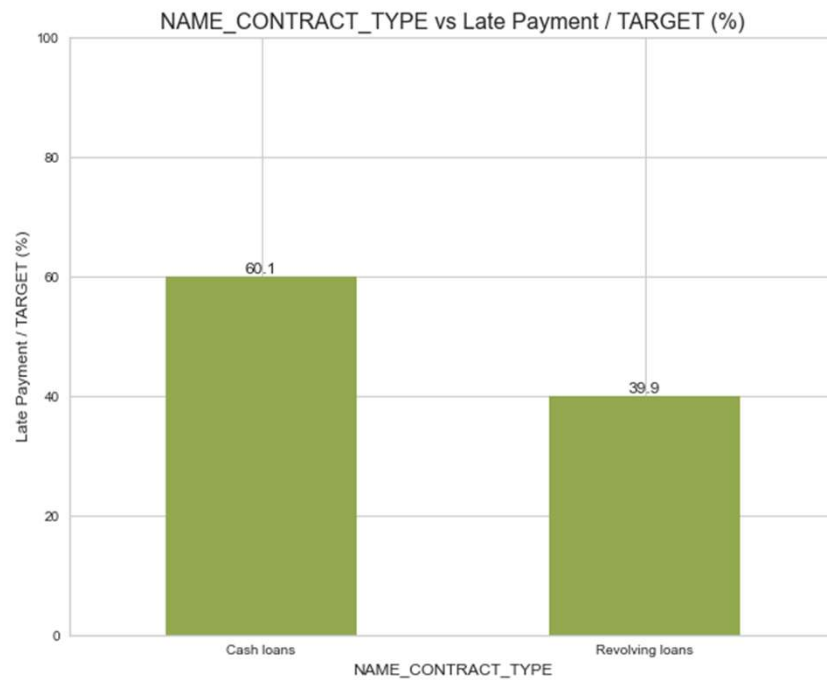
Feature Selection



3 Feature yang paling berkorelasi dengan target variable (TARGET) adalah

1. NAME_CONTRACT_TYPE
2. REGION_RATING_CLIENT
3. CODE_GENDER

Type Contract



Feature Type Contract adalah Identifikasi apakah jenis pinjaman nasabah tunai atau berjangka.

Dari Analisa yang telah dilakukan, dapat dinyatakan bahwa pelanggan yang memilih tipe kontrak Pinjaman Tunai (Cash) lebih sering mengalami keterlambatan dalam pembayaran kredit dibandingkan dengan pelanggan yang menggunakan jenis pinjaman berjangka.

Region Rating



Feature Region Rating Client adalah peringkat wilayah tempat tinggal klien. Peringkat ini dibuat oleh perusahaan Home Credit sendiri.

Dari Analisa yang telah dilakukan, dapat dinyatakan bahwa pelanggan Pelanggan yang tinggal dalam wilayah peringkat 3, lebih sering mengalami keterlambatan dalam pembayaran kredit. Dan pelanggan yang tinggal dalam wilayah peringkat 1 paling jarang mengalami keterlambatan pembayaran.

Gender



Feature Code Gender adalah Identifikasi dari jenis kelamin pelanggan.

Dari Analisa yang telah dilakukan, dapat dinyatakan bahwa pelanggan mengalami keterlambatan dalam pembayaran kredit dibandingkan dengan pelanggan perempuan

Modeling Data



Pendahuluan

Variabel yang paling diinginkan untuk optimal adalah **F1-score** karena ingin meminimalisir false positive maupun false negative.

Prediksi / Hipotesis yang dibuat berdasarkan target variabel :

0: Pelanggan mampu melakukan pelunasan (Positive)

1: Pelanggan tidak mampu melakukan pelunasan (Negative)

False Positive: Diprediksi pelanggan mampu melakukan pelunasan, Namun aktualnya pelanggan tidak mampu melakukan pelunasan

False Negative: Diprediksi pelanggan tidak mampu melakukan pelunasan, Namun aktualnya pelanggan mampu melakukan pelunasan

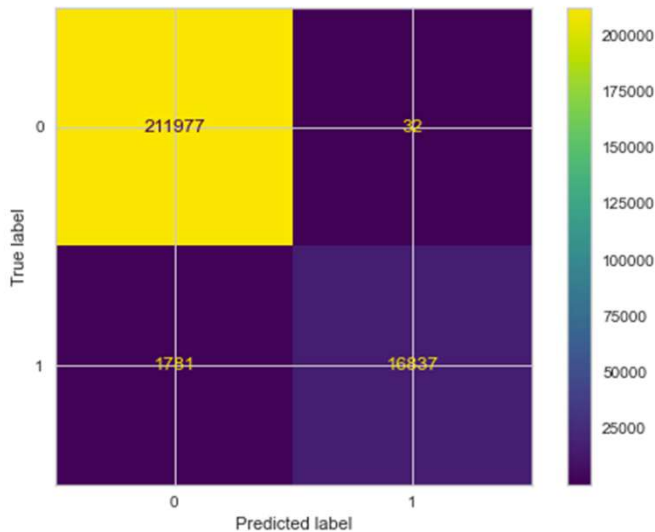
Modeling Data

Model Machine Learning yang akan diuji antara lain:

1. Logical Regression
2. K-NN
3. Naive Bayes
4. Decision Tree Classification

Modeling Data

Recall Score : 0.9043398861317005
Precision Score : 0.998103029225206
F1 Score : 0.9489108687688449
ROC-AUC Score : 0.9995353910019146



Dari observasi yang dilakukan, maka dapat disimpulkan bahwa model terbaik untuk memprediksi skor kredit pelanggan adalah **Model**

Decision Tree karena memiliki nilai F1-Score terbaik. Dengan hasil konfigurasi sebagai berikut:

Recall: $\pm 90\%$

Precision: $\pm 99\%$

F1-Score: $\pm 94\%$

ROC-AUC Score: $\pm 99\%$

Recommendation



Modeling Data

Dari beberapa model terbaik untuk memprediksi skor kredit pelanggan adalah **Model Decision Tree**, karena memiliki f-1 score yang paling tinggi yaitu 94%. Selain itu, model juga memiliki ROC-AUC Score yang tinggi yaitu 99%, sehingga model memiliki kemampuan untuk membedakan antara 0 dan 1 dengan baik.

Business Insight

Selain model, dari analisa yang telah dilakukan terdapat beberapa insight yang dapat digunakan perusahaan dalam menentukan pinjaman pada pelanggan, antara lain :

1. Pelanggan yang memilih tipe kontrak Pinjaman Tunai (Cash) lebih sering terlambat dalam pembayaran kredit
2. Pelanggan yang tinggal dalam wilayah peringkat 3, lebih sering terlambat dalam pembayaran kredit
3. Pelanggan laki-laki lebih sering terlambat dalam pembayaran kredit