

Home Credit Scorecard

PROJECT-BASED INTERNSHIP: DATA SCIENTIST
HOME CREDIT INDONESIA X RAKAMIN ACADEMY
YULY ARTAMI SAGALA

Background

Industri keuangan semakin bergantung pada teknologi canggih untuk menilai risiko kredit secara lebih akurat. Risiko kredit adalah potensi kerugian yang mungkin terjadi karena peminjam tidak bisa membayar utangnya. Home Credit Indonesia menghadapi tantangan dalam menyaring calon peminjam yang layak untuk mendapatkan kredit sebagai salah satu penyedia layanan keuangan.

Problem:

Mencegah penolakan pelanggan yang layak dan menentukan pinjaman optimal.

Solusi:

Memprediksi scorecard nasabah home credit dengan menggunakan metode statistik dan Machine Learning

Dataset

- ❑ Dataset yang digunakan pada proyek ini adalah “application_train.csv”.
- ❑ Dataset terdiri dari 307511 baris dengan nilai unik berdasarkan ID pinjaman dalam “SK_ID_CURR”
- ❑ Dataset terdiri dari 122 kolom fitur yang terkait dengan demografi klien
- ❑ Kolom “TARGET” menunjukkan label dari masalah, dengan 0 = klien tanpa kesulitan pembayaran dan 1 = klien dengan kesulitan pembayaran

Project Workflow (1)

Data Preparation



Data Validation



Analisis

Data

Eksplorasi (EDA)

- Mengimpor library dan memasukkan dataset
- Memeriksa tipe data dan bentuk data

- Menghapus kolom yang memiliki data hilang $>50\%$
- Mengatasi data hilang menggunakan modus dan mean

- Memvisualisasikan data menggunakan diagram diagram batang
- Melihat wawasan bisnis

Project Workflow (2)

Data Preprocessing



Membuat Machine Learning Model



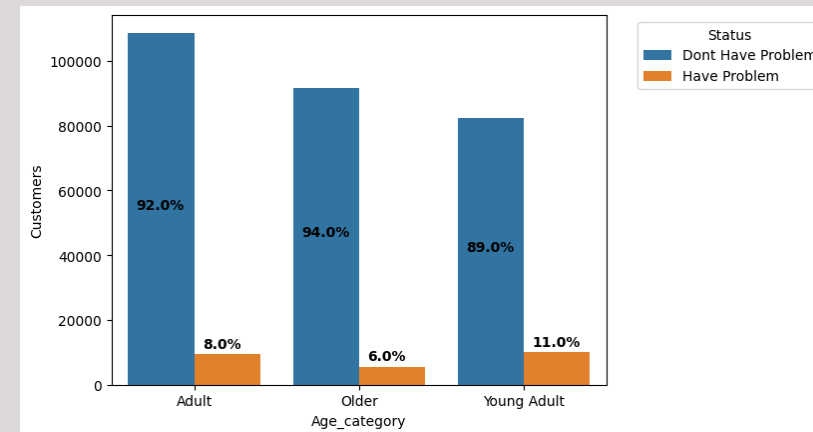
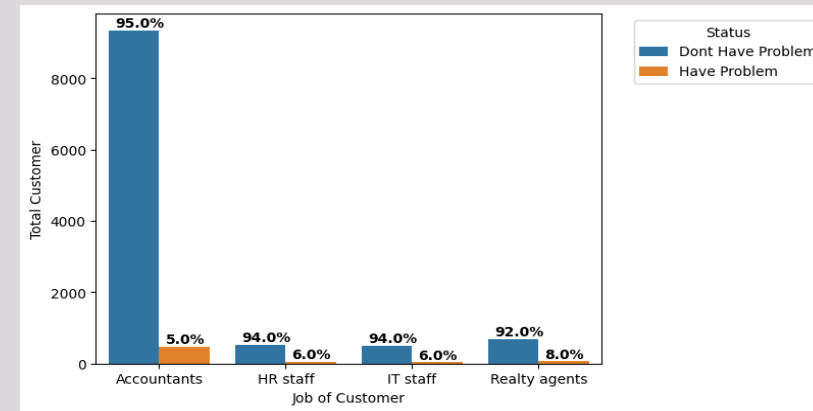
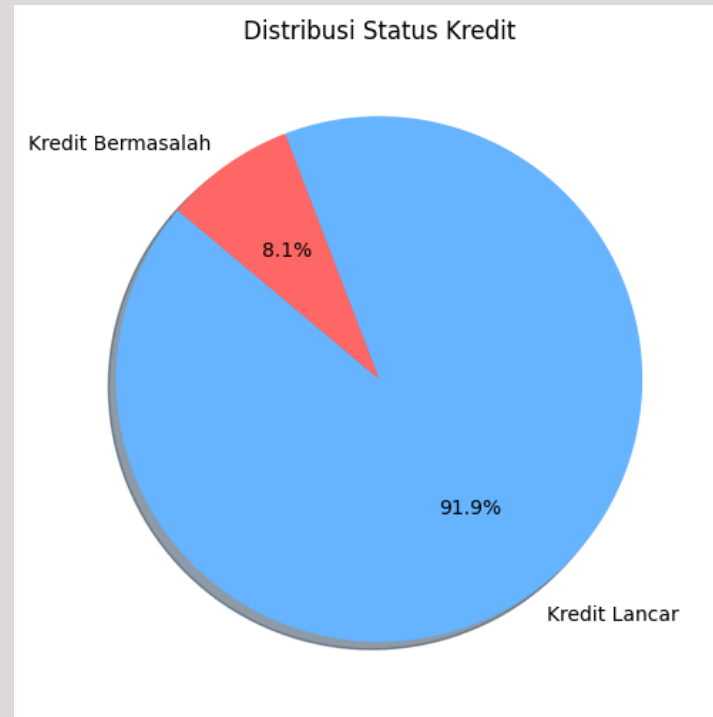
Model Evaluation

- Memasukan input (X) dan target (Y) secara terpisah
- Melakukan feature selection

- Mengatasi ketidakseimbangan data menggunakan oversampling
- Menggunakan Machine Learning model dengan Logistic Regression, XGBoost, Naives Bayes

- Mengevaluasi model menggunakan akurasi
- Memilih model terbaik

EDA (Analisis Data Eksplorasi)



Wawasan Bisnis (1)

- ❑ Persentase nasabah yang mengalami kesulitan pembayaran dalam melunasi pinjaman adalah 8,1% atau 24.825.
- ❑ Sebanyak 8% dari nasabah yang tidak menikah memiliki masalah kesulitan bayar.
- ❑ Jenis kontrak yang ditawarkan kepada nasabah terbagi menjadi 2 jenis yaitu pinjaman tunai dan pinjaman bergulir. Sebanyak 8% pinjaman tunai dan 5% pinjaman bergulir mengalami kesulitan pembayaran.

Wawasan Bisnis (2)

- ❑ Nasabah dengan pekerjaan Staf HR, Staf IT, dan Agen Properti memiliki persentase yang besar (92%-94%) tidak kesulitan membayar, tetapi jumlah nasabah dengan pekerjaan tersebut yang mengajukan pinjaman kurang dari 1000 nasabah.
- ❑ Usia dewasa (35-50 tahun) merupakan kategori usia dengan pengajuan pinjaman terbesar Dimana 92% dari pinjaman tidak mengalami masalah pembayaran.

Hasil Akhir

Model	Accuracy
Logistic Regression (oversampling)	94%
Logistic Regression (undersampling)	68%
XGBoost (oversampling)	95%
Naives Bayes (undersampling)	96%
Random Forest (oversampling)	95%
Random Forest (undersampling)	100%

Didapatkan model terbaik adalah Random Forest

Rekomendasi Bisnis

- ☐ Home Credit Indonesia dapat membuat campaign dengan menargetkan nasabah dengan pekerjaan Staf HR, Staf IT, dan Agen Properti untuk mengajukan pinjaman
- ☐ Menargetkan orang dewasa dengan usia 35-50 tahun untuk mengajukan pinjaman
- ☐ Meningkatkan system pengawasan di Perusahaan misalnya memonitor pembayaran pinjaman, batas maksimum pinjaman.

Thank You

<https://github.com/yulyartamisagala>