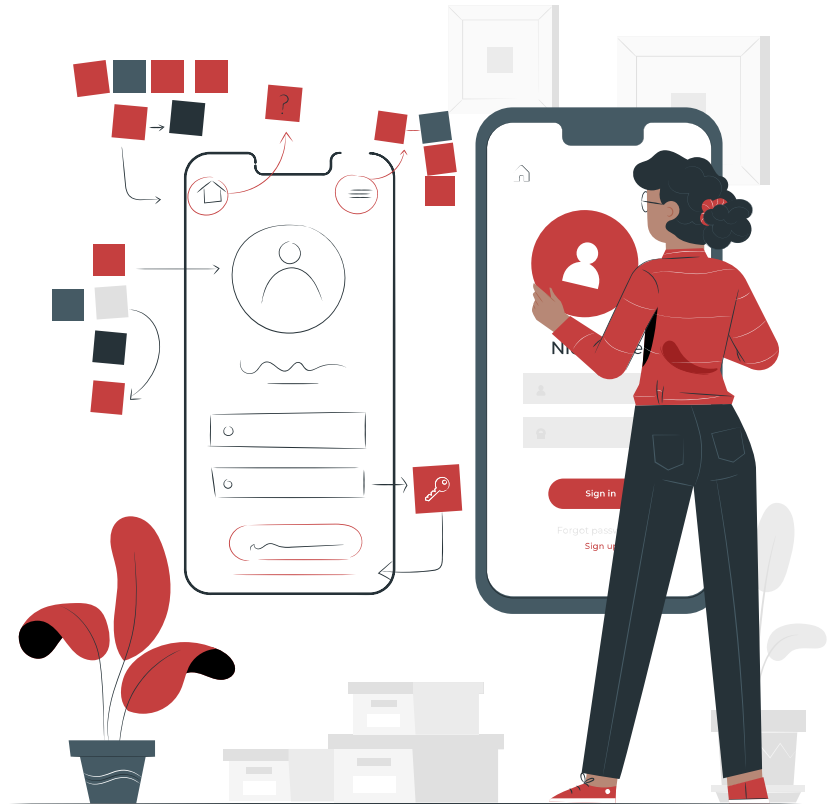


HOME CREDIT CUSTOMERS PREDICTION ANALYSIS

Tevani Oline Perdana

Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)



AGENDA

01 Project
Background

02 Data
Pre-processing

03 Business
Insight

Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)

04 Machine
Learning
Model

05 Business
Recommendation

Cek code selengkapnya disini :
[Home Credit Customers Prediction Analysis](#)



01

Problem Background

PROJECT BACKGROUND

- Banyak orang berusaha untuk mendapatkan pinjaman karena riwayat kredit yang tidak mencukupi atau tidak ada. Home Credit berupaya untuk memperluas inklusi keuangan bagi masyarakat yang belum tersentuh layanan perbankan dengan memberikan pengalaman peminjaman yang positif dan aman.
- Untuk memastikan suatu populasi yang kurang terlayani tersebut memiliki pengalaman pinjaman yang positif, Home Credit menggunakan berbagai data alternatif untuk memprediksi kemampuan pembayaran klien mereka. Melakukan hal itu akan memastikan bahwa klien yang mampu membayar kembali tidak ditolak dan bahwa pinjaman diberikan dengan pokok, jatuh tempo, dan kalender pembayaran yang akan memberdayakan klien mereka untuk berhasil.

Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)

DATA SOURCE

- Dataset yang digunakan yaitu set application train dan test.
- Di dalam data train terdapat kolom prediksi label (TARGET).
- Sedangkan, data tes tidak terdapat kolom prediksi

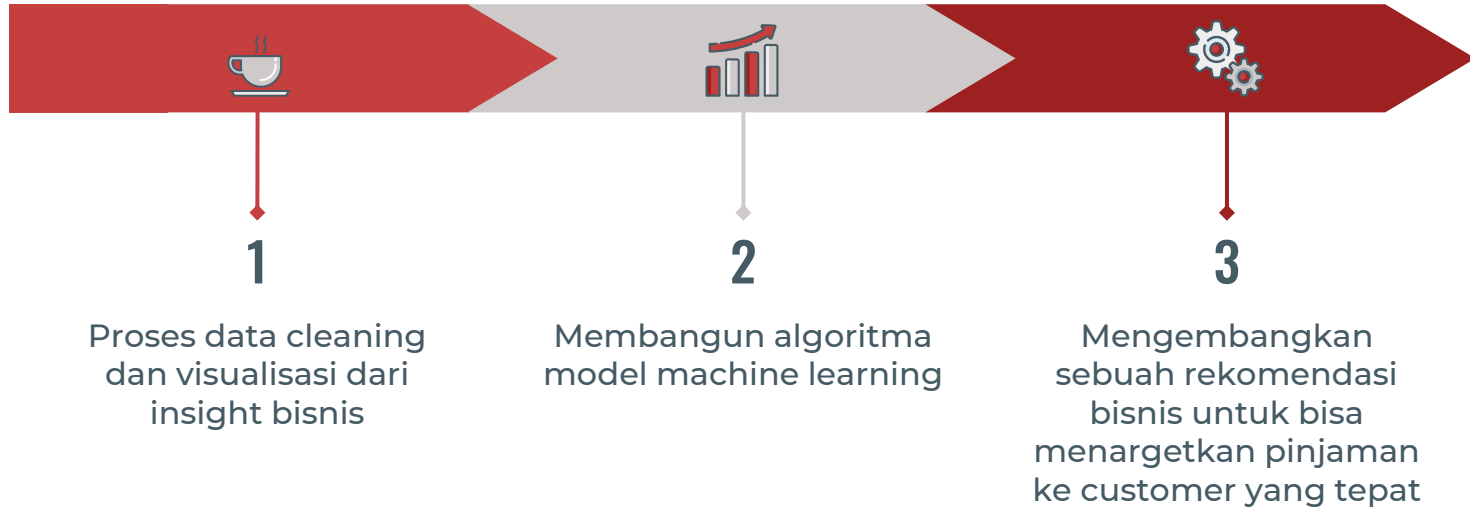
OBJECTIVE

- Mengidentifikasi ciri customer yang berpotensi kesulitan dalam pembayaran pinjaman atau tidak
- Memprediksi kemampuan customer dalam pembayaran kredit

Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)

ACTIONS



Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)



02 Data Preprocessing

Cek code selengkapnya disini :

[Home Credit Customers Prediction Analysis](#)

**Data
Application Train**

307,511

Number of Columns

122

Number of Row

**Exploratory Data
Analysis (EDA)**

Visualisasi Bivariate

Visualisasi hubungan antara
2 fitur

Visualisasi Multivariate

Visualisasi hubungan lebih
dari 2 fitur

Data Cleaning

Hapus Duplicate Data

Tidak ada data duplikat : 0

Mengisi Missing Values

Hapus fitur yang mengandung
data kosong > 50% dan
melakukan imputasi

Handling Outliers

Outlier tidak dihapus

Model Flow Handling

Label Encoding

Mengubah fitur kategorikal jadi fitur numerical

Feature Selection

Mengambil 20 fitur terbaik

Handling Imbalance Data

Resample data supaya jumlah label seimbang

Model Building & Evaluation

Membangun model ML dengan beberapa
algoritma, dan memilih akurasi terbaik

**Data
Application Test**

48,744

Number of Columns

121

Number of Row

Data Cleaning

Hapus Duplicate Data

Tidak ada data duplikat : 0

Mengisi Missing Values

Hapus fitur yang mengandung data kosong > 50% dan melakukan imputasi

Handling Outliers

Outlier tidak dihapus

Model

Label Encoding

Mengubah fitur kategorikal jadi fitur numerical

Feature Selection

Mengambil 20 fitur terbaik

Model Building & Evaluation

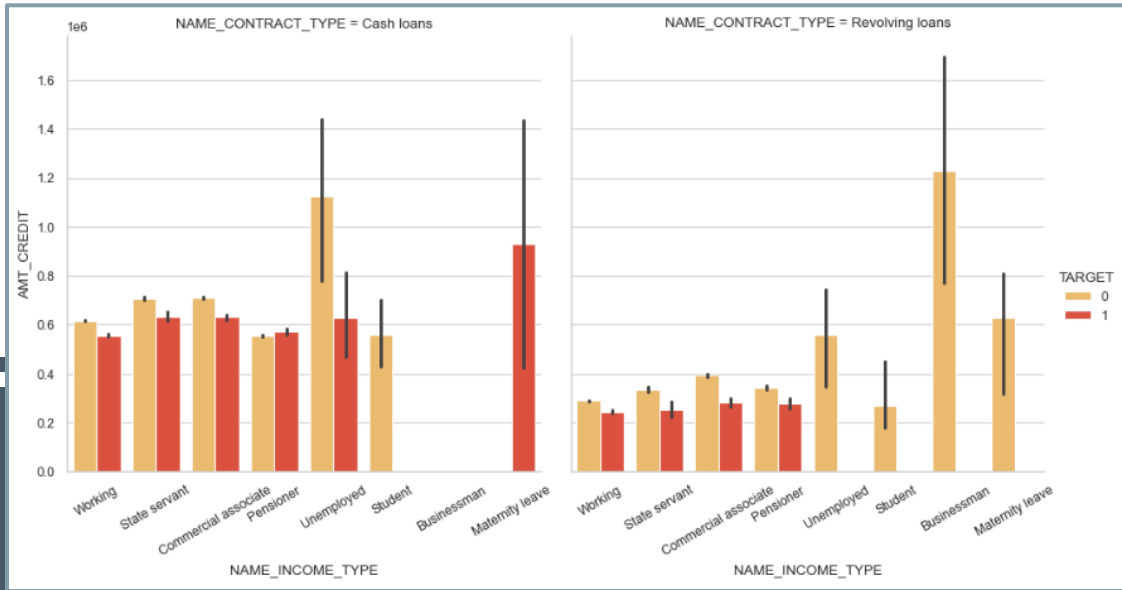
Membangun model ML dengan beberapa algoritma, dan memilih akurasi terbaik

Prediction

Memprediksi kemampuan customer dalam membayar pinjaman menggunakan model machine learning terbaik yang sudah diuji sebelumnya



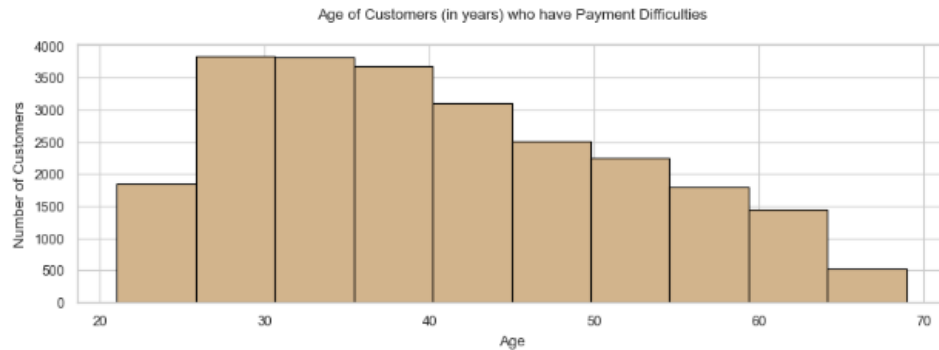
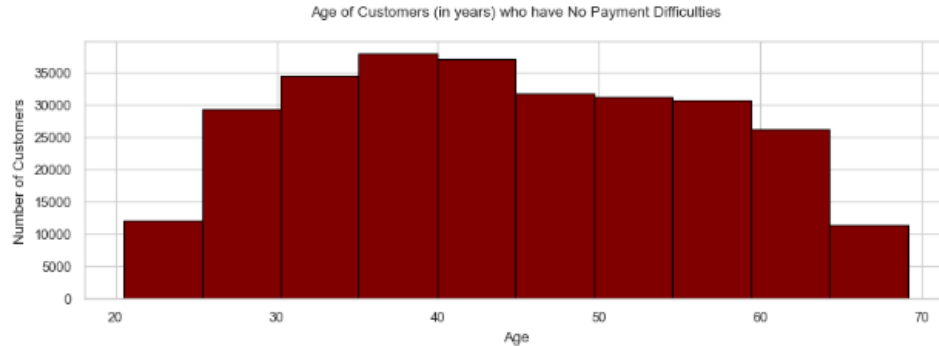
03 Business Insights



Unemployed pada **Cash Loans**, jumlah customers yang membayar pinjaman tinggi dan tidak memiliki masalah pembayaran itu jumlahnya lebih tinggi dibandingkan dengan yang kesulitan membayar, namun **Unemployed** pada **Revolving Loans** tidak ada customer yang memiliki kesulitan dalam pembayaran

Maternity Leave pada **Cash Loans**, customer yang memiliki income tersebut memiliki jumlah pinjaman yang lebih banyak dan **mengalami kesulitan pembayaran kredit** dibandingkan dengan maternity leave pada **Revolving Loans** yang sama sekali tidak ada customer yang mengalami kesulitan dalam pembayaran.

Student, baik itu pinjaman jenis **Cash** maupun **Revolving** keduanya tidak ada yang memiliki kesulitan dalam pembayaran.



- Customers yang paling banyak mengajukan pinjaman kredit berumur sekitar **35-40** tahun, kemudian diikuti oleh customer dengan umur **41-45** tahun

- Customers yang tidak memiliki masalah atau kesulitan dalam pembayaran yaitu yang berumur sekitar **35-40** tahun
- Customer yang memiliki masalah dan kesulitan pembayaran kredit yaitu yang berumur sekitar **25-35** tahun. Hal ini bisa terjadi karena di umur tersebut merupakan masa awal dalam merintis karir sehingga belum mengalami financial stable, sehingga masih kesulitan dalam membayar pinjaman kredit.



04 Business Insights

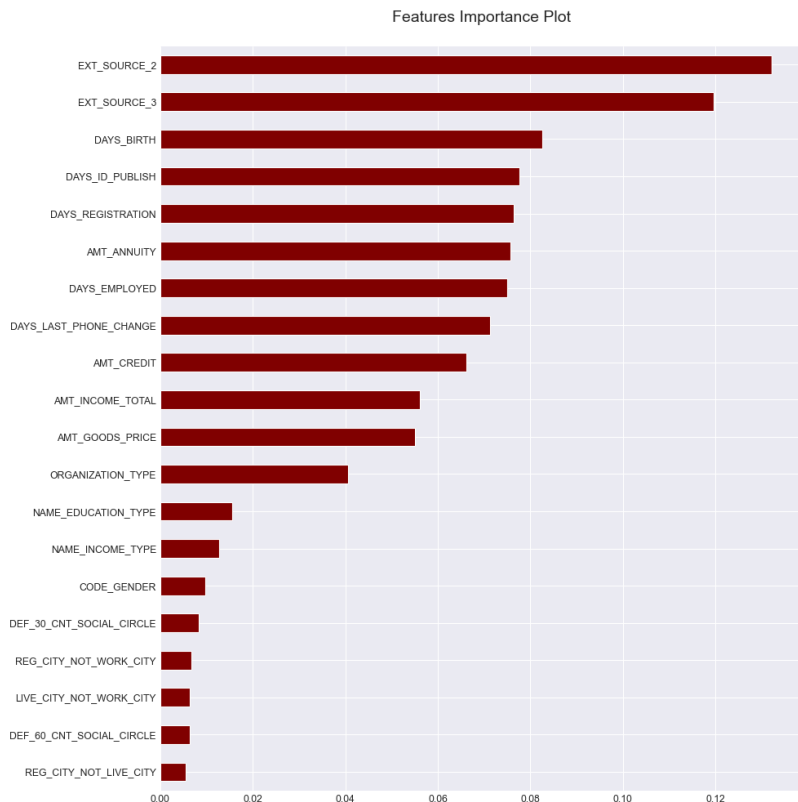
MODEL COMPARISON

Algorithm	Training Accuracy Score	Testing Accuracy Score	ROC Score
Random Forest	1.0000	0.9966	0.9966
Decision Tree	1.0000	0.8721	0.8722
K-Nearest Neighbor	0.9157	0.8803	0.8801
Logistic Regression	0.6721	0.6718	0.6718

Akurasi prediksi dari data train dan test menunjukkan bahwa model **Random Forest** memiliki nilai yang tidak berbeda signifikan, sehingga bisa dikatakan bahwa model tersebut sangat baik untuk pengujian selanjutnya

karna tidak menunjukkan adanya **overfitting** atau **underfitting**. Jadi, Random Forest akan dipilih untuk menguji data tes kemampuan pembayaran kredit customer.

FEATURES IMPOPRTANCE



Algorithm

Random Forest Classifier

Performance

100%

Model Random forest pada data training

99.66%

Model Random forest pada data test

5 Top Features IMportance

EXT_SOURCE_2, EXT_SOURCE_1, DAYS_BIRTH
DAYS_ID_PUBLISH, DAYS_REGISTRATION



05

**Business
Recommendation**

1. Melakukan Survey Langsung dan Interview

- Melakukan analisis mendalam seperti melakukan survey langsung dan interview bagi calon customer yang pendapatannya dari **Maternity Leave, Working, Unemployed, Businessman**, atau **Pensioners**. Untuk **Businessman** bisa melakukan survey ke tempat usahanya dan interview mengenai kestabilan profitnya.
- Untuk **Student** meskipun data menunjukkan tidak adanya kesulitan dalam pembayaran, kita harus melakukan interview mengenai pendapatannya atau bisa menerapkan batasan umur untuk peminjaman credit

2. Mengadakan Perjanjian Jaminan

- Mengadakan sebuah jaminan apabila customer telat melakukan pembayaran atau mengalami kemacetan pembayaran seperti perjanjian bunga/denda, hal ini dilakukan agar customer tetap ingat target dalam pengembalian pinjamannya

3. Melakukan Campaign Pada Usia Dengan Pendapatan Stabil

- Pada data menunjukkan bahwa pengajuan pinjaman dilakukan oleh customer yang berumur sekitar 35-40 tahun. Ini bisa menjadi pendukung untuk melakukan campaign kepada yang berumur 35-45 tahun untuk menawarkan pengambilan kredit, karena data menunjukkan bahwa range umur sekian tidak menunjukkan adanya kesulitan dalam pembayaran kredit. Meskipun nantinya akan dilakukan filter kembali mengenai ketentuan-ketentuan lainnya.

2. Membuat Golongan Nasabah Prioritas Utama Untuk Campaign/Pemberian Kredit

- Memiliki pendapatan setiap bulannya karena dia adalah businessman, student, pensioner
- Memiliki pendidikan terakhir di **Academic Degree** atau **Higher Education**
- Memiliki **Office Apartment** atau **House/apartment** atau **Co-op apartment**
- Memiliki pekerjaan sebagai **Accountants, High Skill Tech Staff, Managers, Secretaries, HR Staff, Core Staff, IT Staff, Private Service Staff, Medicine Staff**
- Dibawah organisasi **Trade type:4, Industry: type 12, Transport Type: 1, Trade type:6, Military, Police,Bank, University, Insurance, Hotel**
- Berumur antara **35-45 tahun** karena biasanya sudah memiliki karir dan umur masih termasuk produktif
- Memiliki salah satu aset kendaraan pribadi atau realty flat/home
- Berstatus Married, Civil Marriage

THANKS

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

