





Home Credit Analysis and Prediction with Machine Learning

Kelompok 1

Google Colab

Dashboard







Our Team

Our Mentor: Ahmad Ridha Kelrey







Inayahtul As Shafikah 035



Mita Amaliyatul Hayat 037



M Alfan Alfian Matondang
039



Nurlita Kholishotunnisa 041

Table of Content

Home Credit Default Risk

1	Business Understanding
2	Data Understanding
3	Data Preparation

4	Machine Learning Model
5	Evaluation
6	Deployment
7	Business Solution



Business Understanding





Abstrak

Home Credit adalah
perusahaan keuangan
yang memiliki layanan
untuk memberikan
pinjaman yang aman bagi
masyarakat



Problem

Dari banyaknya pelanggan yang ingin mengajukan pinjaman, 8% pelanggan tidak mampu membayar kembali pinjamannya

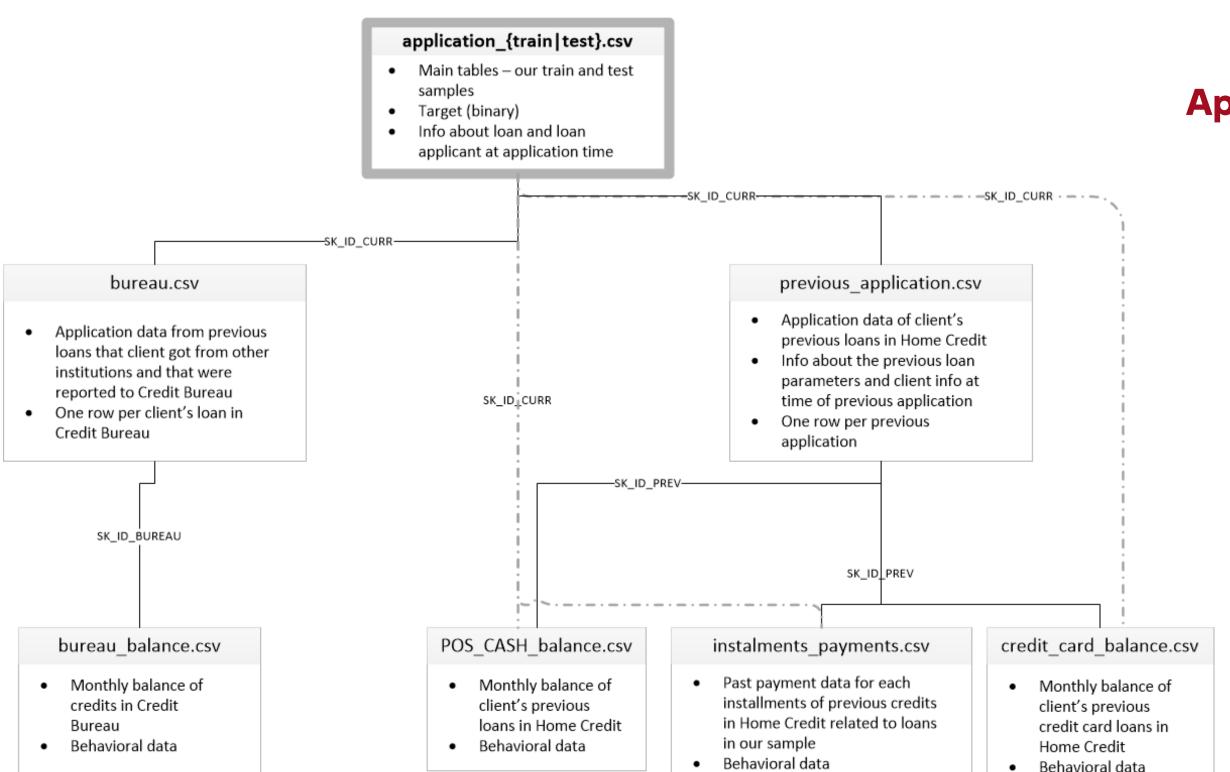


Goal

Membuat model untuk memprediksi seberapa tinggi kemungkinan pelanggan untuk membayar kembali pinjamannya



Data Understanding

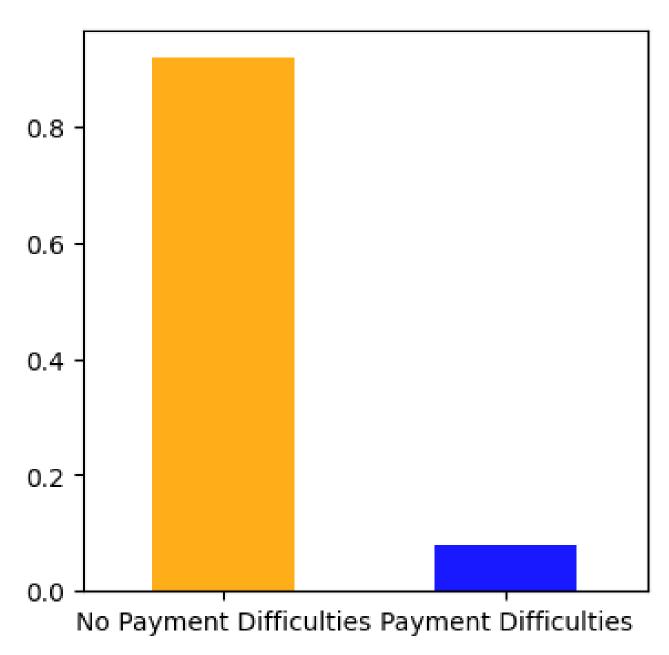


Application_(Train|Test).csv

- Dataset tersebut merupakan tabel utama yang terdiri dari 2 data. Train dengan variabel Target dan Test tanpa variabel Target
- Terdiri dari 122 rows, 307.511 Entries.

Analisis Kredit

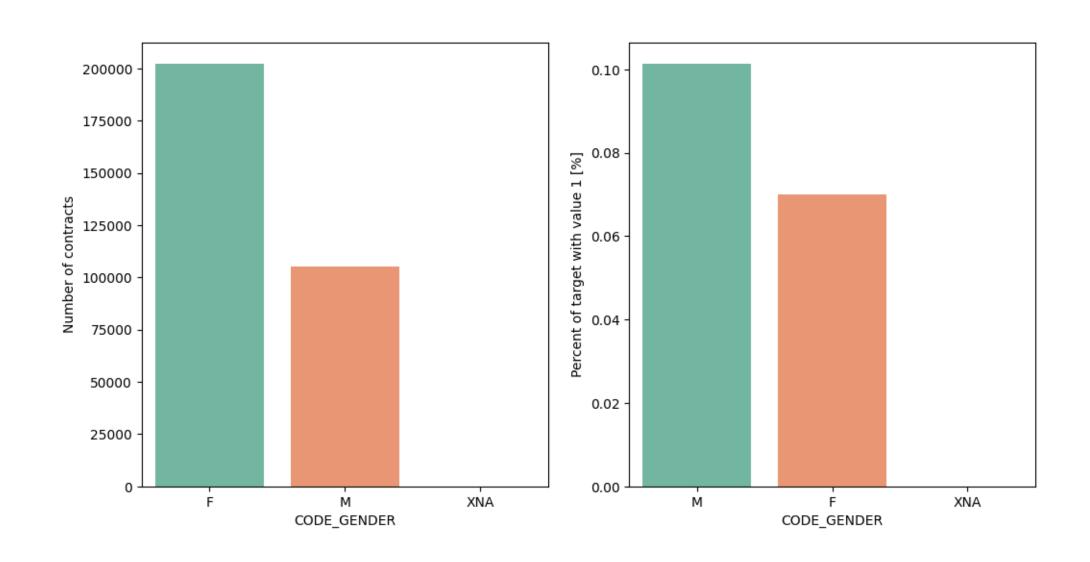




Insight

8% (24.825) pelanggan mengalami kesulitan dalam membayar pinjaman

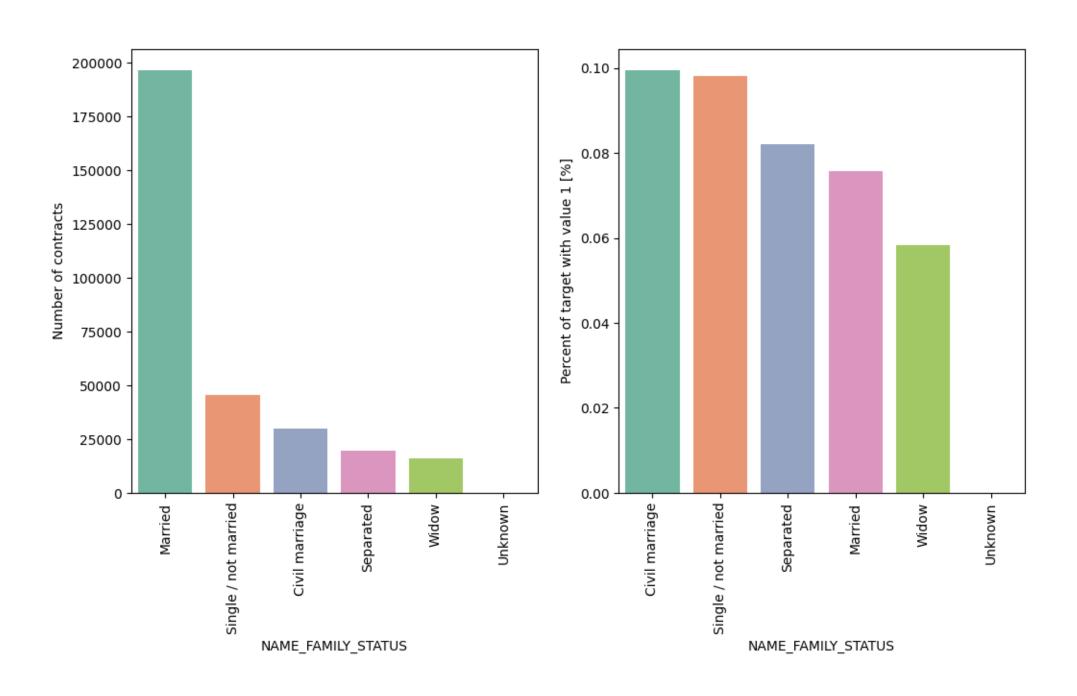
Kemampuan Bayar Berdasarkan Gender



Insight

Wanita lebih banyak mengajukan pinjaman. Pria memiliki risiko gagal bayar yang lebih tinggi dengan probabilitas sebesar 89%

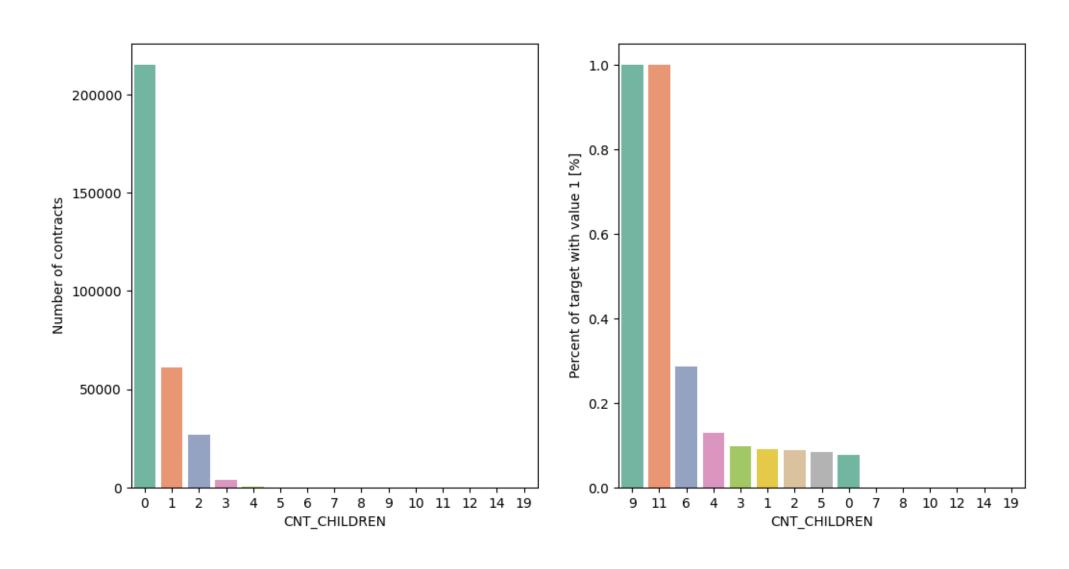
Kemampuan Bayar Berdasarkan Status Keluarga



Insight

Mayoritas pelanggan yang mengajukan pinjaman berstatus married. Sedangkan jika dilihat dari kemampuan pelanggan membayar pinjamannya. Pelanggan dengan status civil marriage memiliki lebih banyak kesulitan untuk membayar kredit

Kemampuan Bayar Berdasarkan Jumlah Anak



Insight

Pelanggan yang memiliki 9 atau 11 anak memiliiki risiko gagal bayar yang tinggi yakni sebesar 100%.

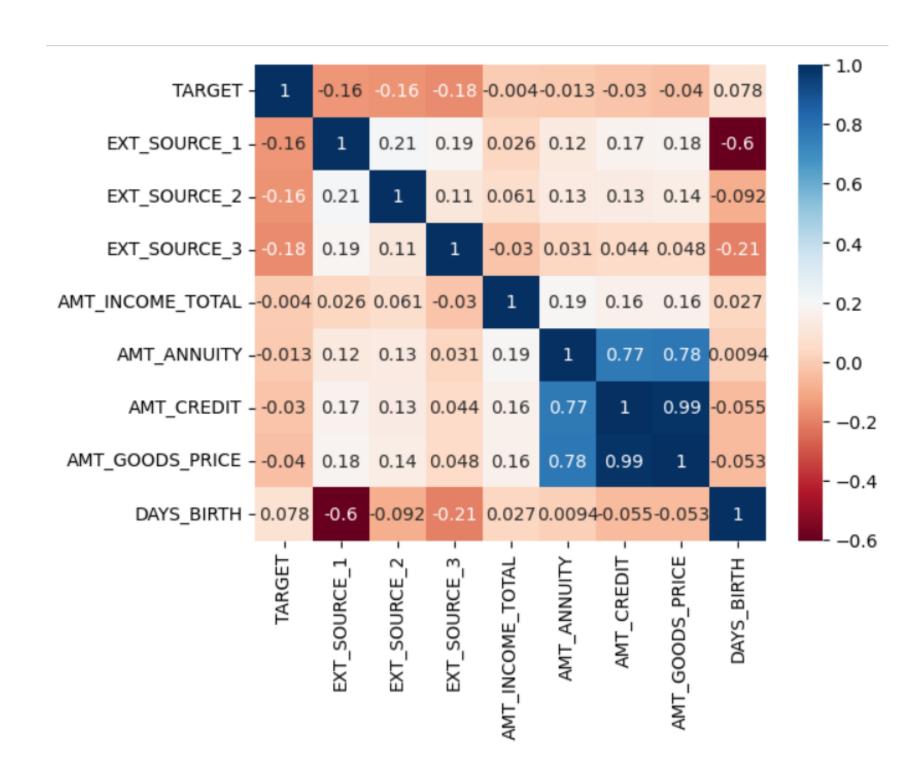
Kemampuan Bayar Berdasarkan Umur



Insight

Pelanggan yang berusia lebih muda cenderung memiliki kesulitan dalam membayar pinjamannya

Korelasi Beberapa fitur dengan TARGET

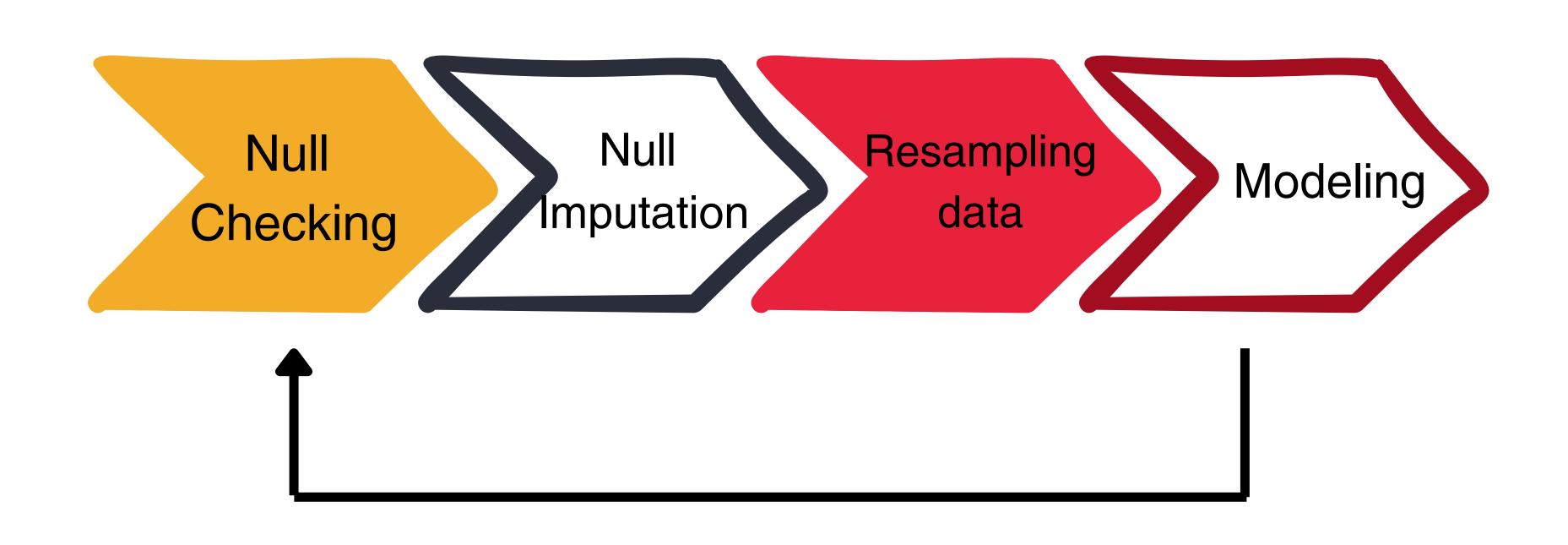


Insight

Fitur DAYS_BIRTH (umur) adalah fitur yang memiliki korelasi tertinggi dengan TARGET



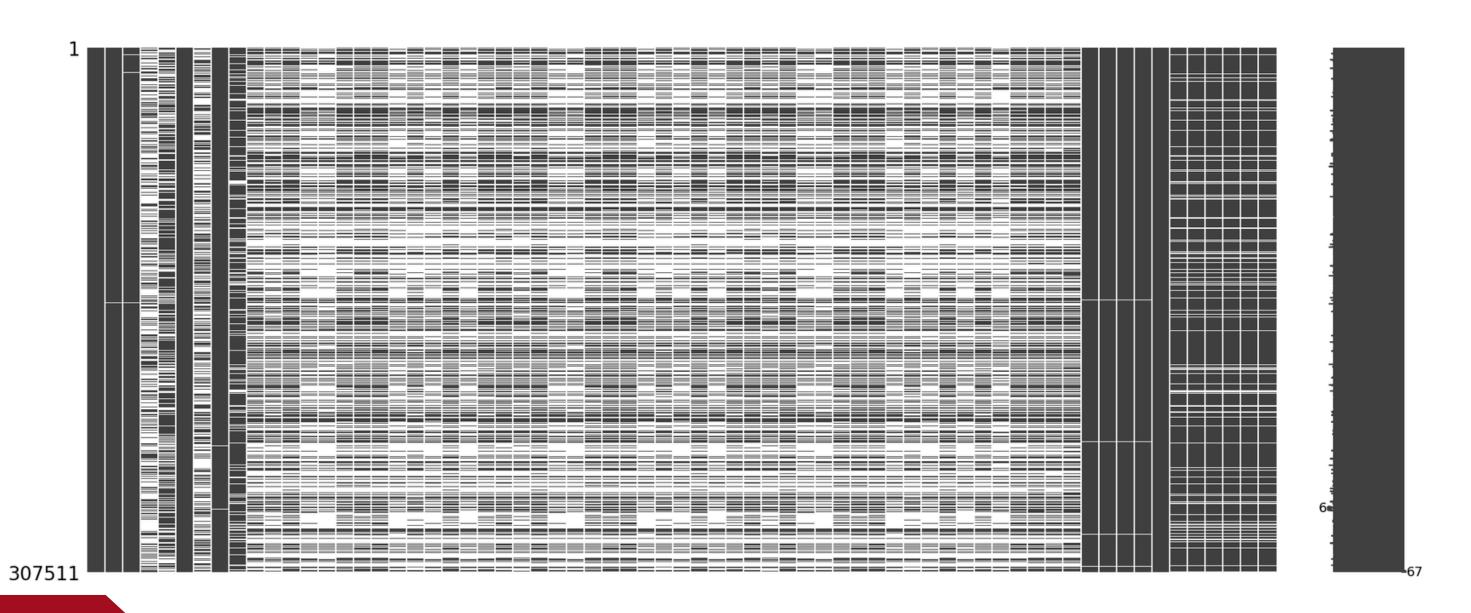
Data Preparation Flow



Data Preparation Flow

Insight

ada 67 fitur memiliki nilai null dan kebanyakan memiliki relasi dengan fitur tetangga



Data Preparation Flow

Null values Solution

Menggunakan imputasi null dengan nilai -1 yang mengindikasikan bahwa nilai tersebut memang null

Distribution Label Solution

Menggunakan Under-Sampling untuk menyetarakan distribusi

	index	TARGET
0	No Payment Difficulties	0.919271
1	Payment Difficulties	0.080729



Perbandingan Model

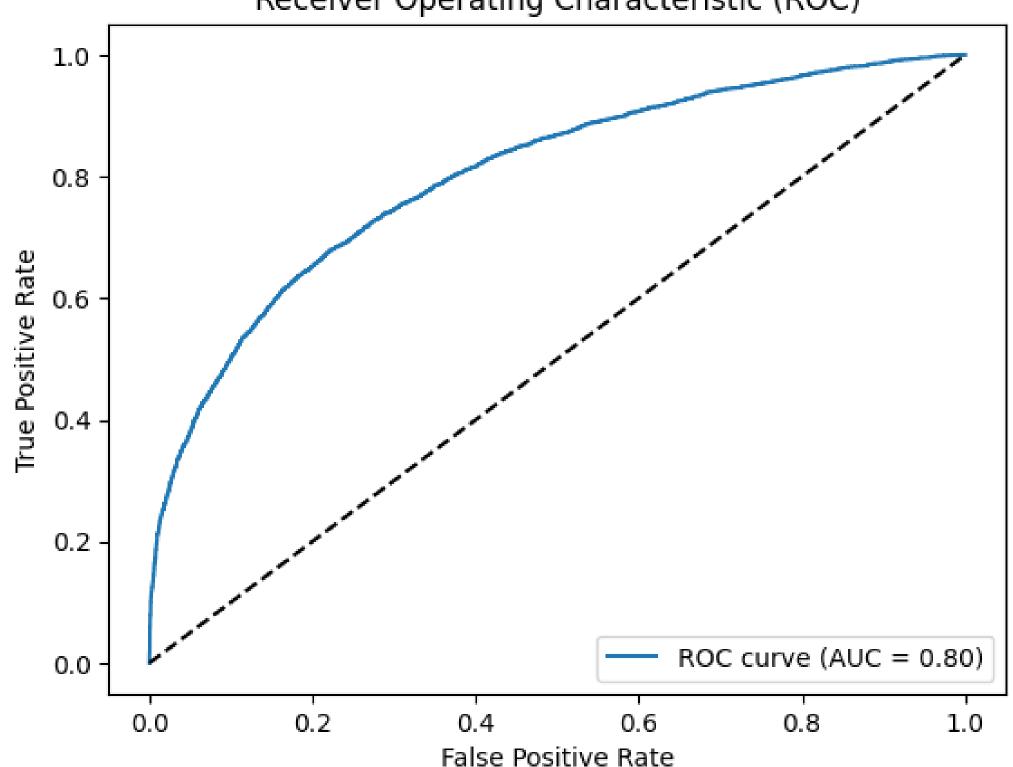
		Accuracy	Presicion	Recall	F1-Score	AUC-ROC	Confusion Matrix	Execution Time
	Naïve Bayes	0.56566	0.655832 0	.276334	0.388834	0.56566	[[4245, 720], [3593, 1372]]	0.175427
	Decision Tree	0.621652	0.618246 0	.636052	0.627023	0.621652	[[3015, 1950], [1807, 3158]]	5.537.551
	Random Forest	0.697583	0.703022 0	.684189	0.693478	0.697583	[[3530, 1435], [1568, 3397]]	22.063.154
Γ	Gradient Boosting	0.725579	0.727642 0	.721047	0.72433	0.725579	[3526, 1439], [1950, 3580]]	96.455.277
	k-Nearest Neighbors (k-NN)	0.658711	0.67692 0	.607251	0.640195	0.658711	[[3526, 1439], [1950, 3015]]	14.192.813

Data train dan test pada model **Gradient Boosting** memiliki nilai performa yang paling baik sekitar 72% dalam memprediksi serta memiliki nilai yang tidak jauh berbeda sehingga menjadi **Model Terbaik**.



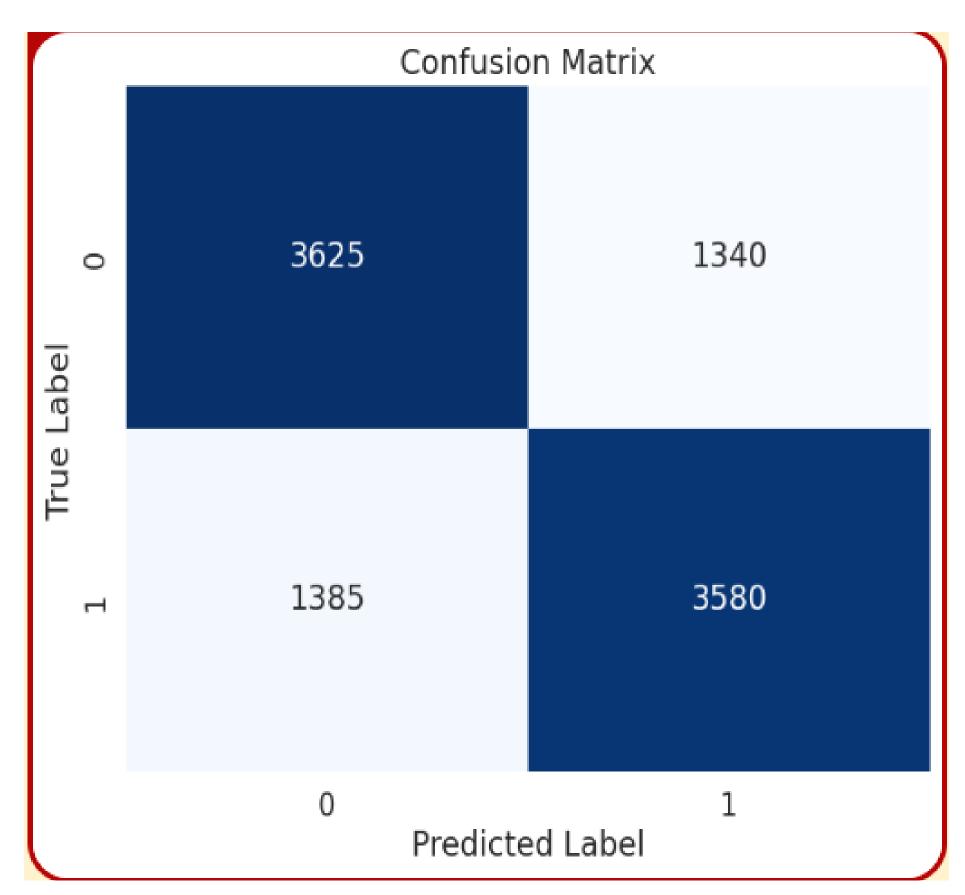
AUC - ROC Curve





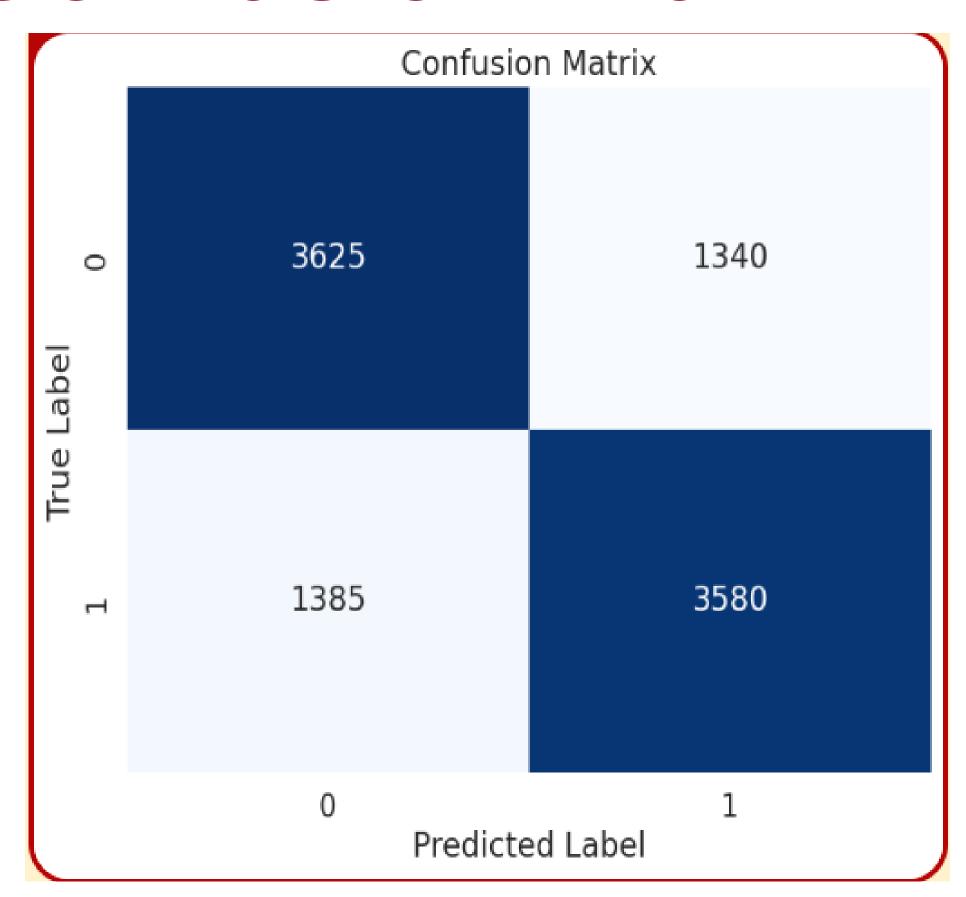
- Hubungan antara tingkat True
- Positive Rate (TPR) dan False
- Positive Rate (FPR) pada
 - berbagai ambang batas
 - (threshold)
- Persentase Kurva AUC-ROC : "72,57%"
 - Semakin dekat dengan sudut kiri
- atas, semakin baik model dalam membedakan antara kelas (+)
 - dan (-).

Confusion Matrix



- True Positive = 'Sesuai Prediksi '
- False Positive = 'Tidak sesuai Prediksi '
- False Negative = 'Tidak sesuai Prediksi '
- True Negative = ' Sesuai Prediksi '

Confusion Matrix



- True Positive = 'Sesuai Prediksi '
- False Positive = 'Tidak sesuai Prediksi '
- False Negative = 'Tidak sesuai Prediksi '
- True Negative = 'Sesuai Prediksi '

Main Point:

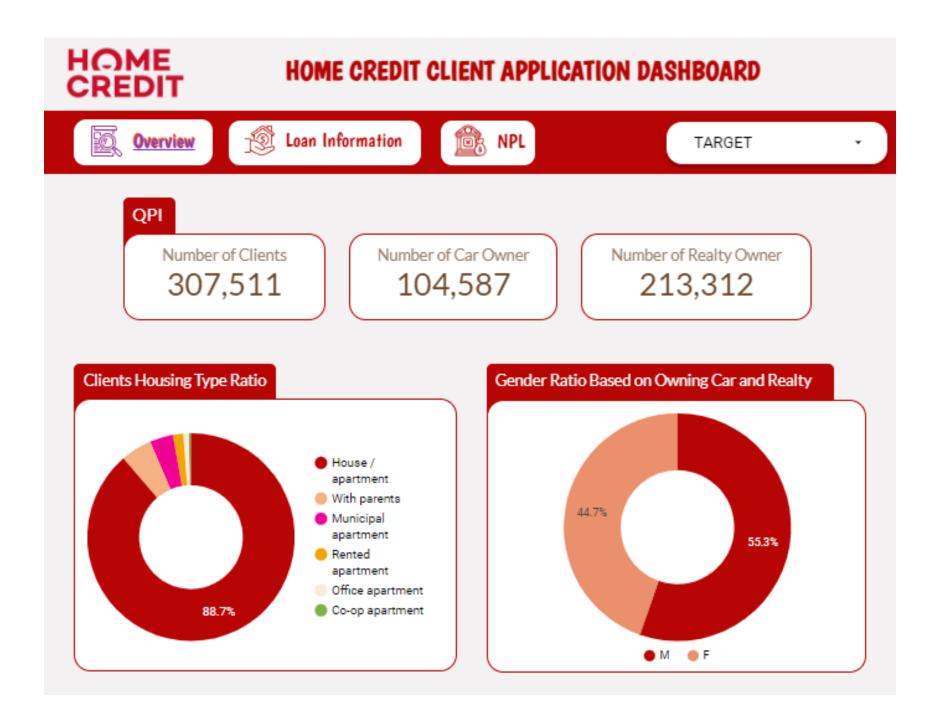
'Prediksi "1385" tidak sesuai/kesalahan fatal dan menyebabkan **kerugian**'

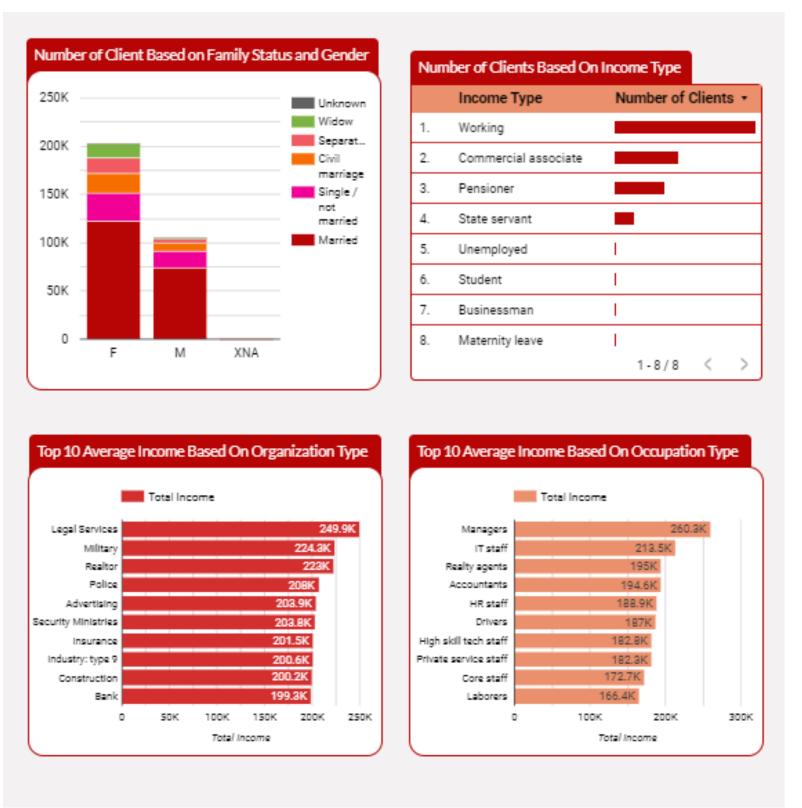
Feature Important

	variable	dropout_loss
219	_baseline_	0.50
218	AMT_GOODS_PRICE	0.27
217	EXT_SOURCE_3	0.24
216	EXT_SOURCE_2	0.23
215	AMT_INCOME_TOTAL	0.21
214	AMT_CREDIT	0.20
213	EXT_SOURCE_1	0.20
212	DAYS_BIRTH	0.20
211	FLAG_WORK_PHONE	0.20
210 NAME_CONTRACT	_TYPE_Revolving loans	0.20



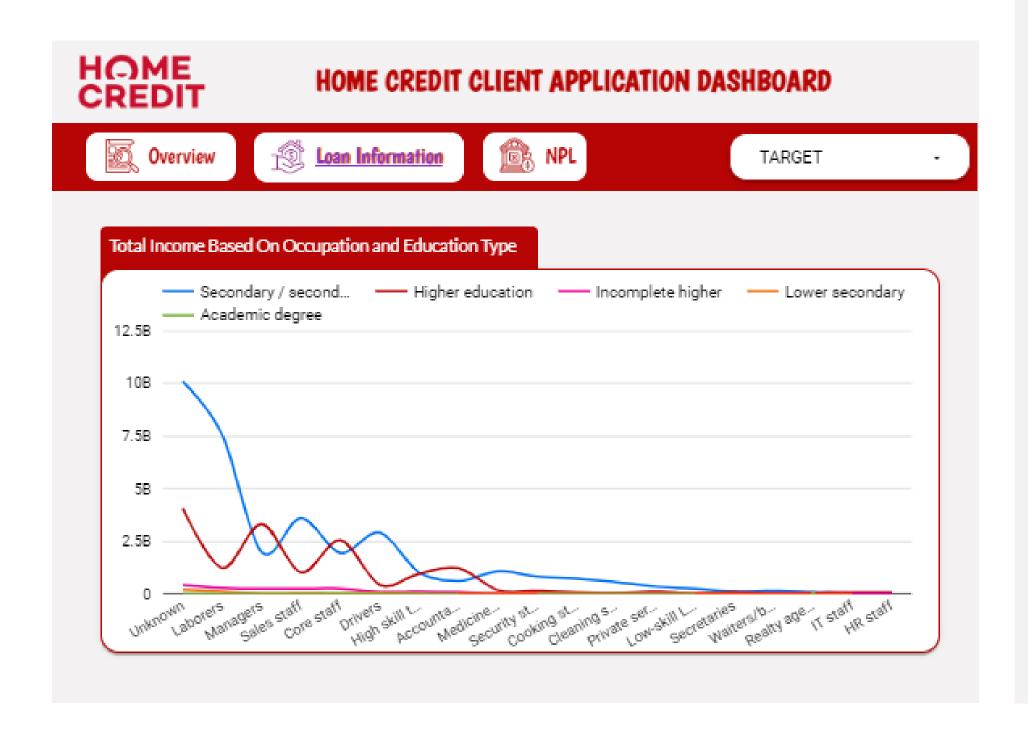
Page 1: Overview

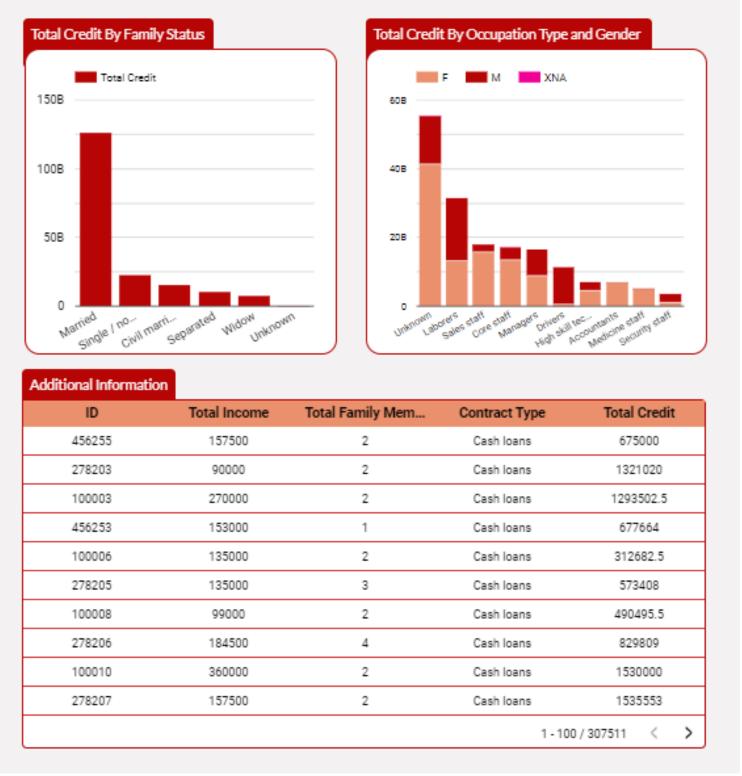




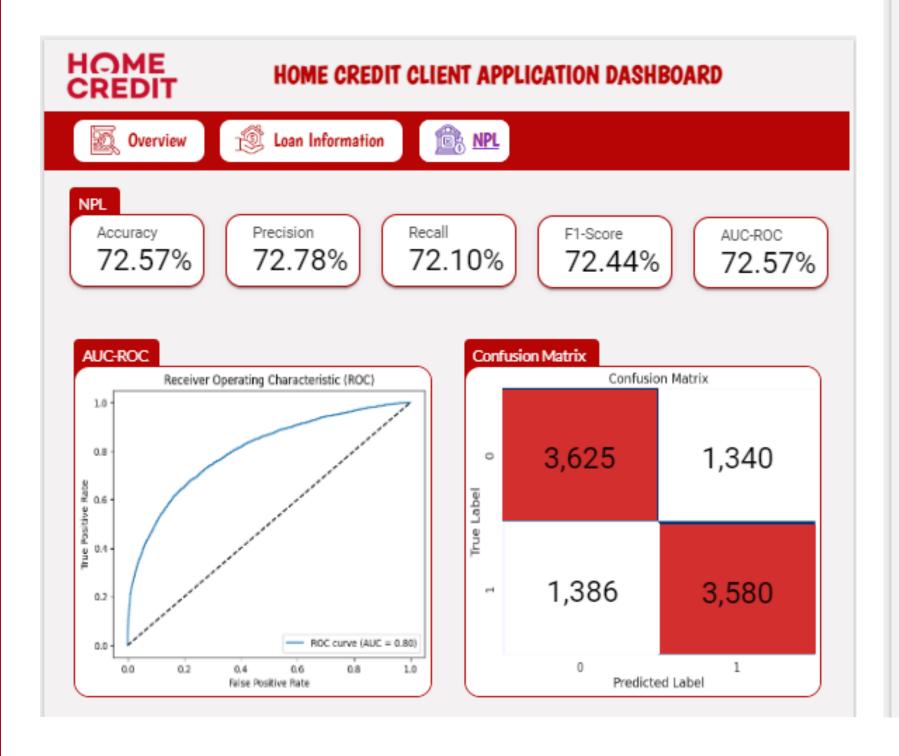
https://lookerstudio.google.com/reporting/93240a76-4270-46ad-a711-4752a7c5695c

Page 2: Loan Information





Page 3: NPL



Top 10 Feature		Bottom 10 Feature
Top Variable	Dropout Loss •	Bottom Variable
baseline	0.5	FLAG_OWN_REALT\
AMT_GOODS_PRICE	0.26	FLAG_OWN_REALTY
EXT_SOURCE_3	0.24	FLAG_PHONE
EXT_SOURCE_2	0.23	ORGANIZATION_TY
AMT_INCOME_TOTAL	0.21	ORGANIZATION_TY
EXT_SOURCE_1	0.2	ORGANIZATION_TY
AMT_CREDIT	0.2	ORGANIZATION_TY
DAYS_BIRTH	0.19	BASEMENTAREA_N
DAYS_EMPLOYED	0.19	AMT_REQ_CREDIT_
NAME_CONTRACT_TYPE_Revo	0.19	AMT_REQ_CREDIT_

DOLLOIT TO FEATURE	
Bottom Variable	Dropout Loss -
FLAG_OWN_REALTY_Y	0.19
FLAG_OWN_REALTY_N	0.19
FLAG_PHONE	0.19
ORGANIZATION_TYPE_Security	0.19
ORGANIZATION_TYPE_Kinder	0.19
ORGANIZATION_TYPE_Busine	0.19
ORGANIZATION_TYPE_Military	0.19
BASEMENTAREA_MEDI	0.19
AMT_REQ_CREDIT_BUREAU	0.19
AMT_REQ_CREDIT_BUREAU_Y	0.19

Model Evaluation					
Algorithm •	Accuracy	Precision	Recall	F1-Score	AUC-ROC
k-Nearest Neighbors (k-NN)	65.87%	67.69%	60.73%	64.02%	65.87%
Random Forest	69.17%	69.76%	67.69%	68.71%	69.17%
Naive Bayes	56.57%	65.58%	27.63%	38.88%	56.57%
Gradient Boosting	72.57%	72.78%	72.10%	72.44%	72.57%
Decision Tree	62.47%	62.09%	64.05%	63.05%	62.47%

https://lookerstudio.google.com/reporting/93240a76-4270-46ad-a711-4752a7c5695c



Business Solution



1 2 4

Perusahaan dapat memberikan bunga yang lebih rendah kepada orang yang berumur 30-40 tahun

Perusahaan dapat lebih berfokus dalam menargetkan wanita pada penawaran home creditnya

Perusahaan
disarankan
memberikan
penawaran kepada
pelanggan yang
berusia 50-60
tahun

Perusahaan
disarankan untuk
menolak
pelanggan yang
tidak ingin mengisi
data secara
lengkap







TERIMA KASIH

