

HOME CREDIT

DEFAULT RISK USING PREDICTION MACHINE LEARNING

Achmad Fathony

PROBLEM STATEMENT

Data ini terkait dengan customer yang meninggalkan layanan atau melunasi pinjaman yang tidak sesuai dari jangka waktu yang diharapkan dari sebuah perusahaan pembiayaan.

Problem:

- Menghadapi tantangan dalam meningkatkan layanan produk credit
- Credit Scoring (0 : lancar dan 1 : tersendat)

Goals:

- Meningkatkan efisiensi dan efektivitas dalam pemasaran produk credit
- Meningkatkan profitabilitas perusahaan dengan meningkatkan pendapatan dari produk credit



DATA OVERVIEW



DATA TRAIN

Data terdiri dari 307.511 baris dan 122 kolom. 116 data numerik dan 16 data categorical

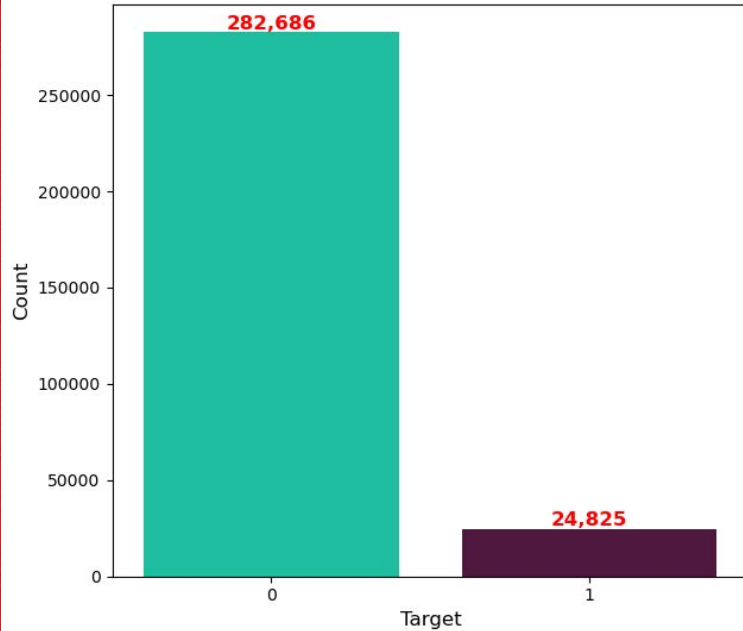


DATA TEST

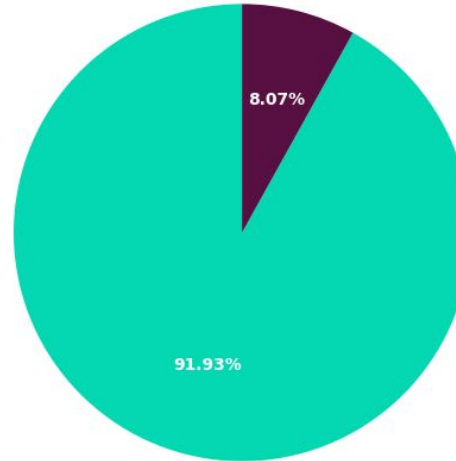
Data terdiri dari 48.744 baris dan 121 kolom. 115 data numerik dan 16 data categorical

DISTRIBUTION CUSTOMER TARGET

Counts of Target



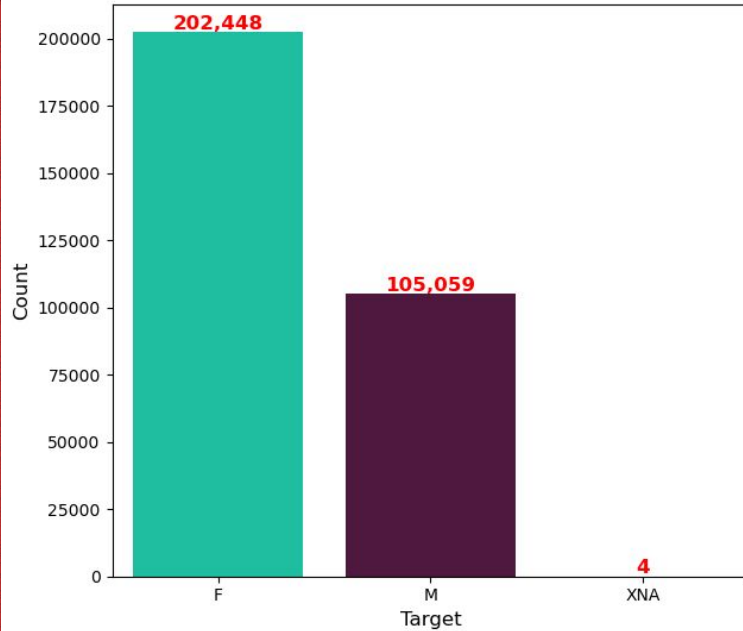
Percentages of Target



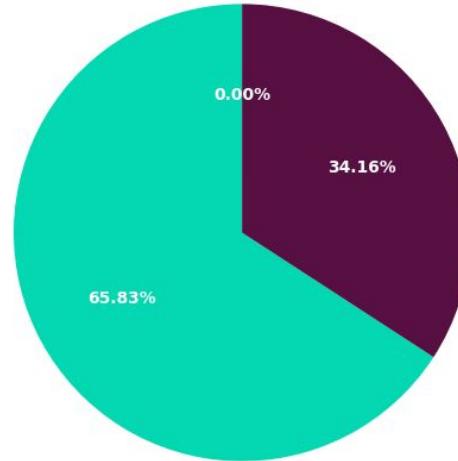
Dari grafik disamping sebaran data imbalance dimana customer yang tidak lancar credit sebesar 8,07% dan customer yang lancar credit sebesar 91,93%
https://github.com/fathony-ac hmad/prediction_credithome

DISTRIBUTION CUSTOMER GENDER

Counts of Gender Type



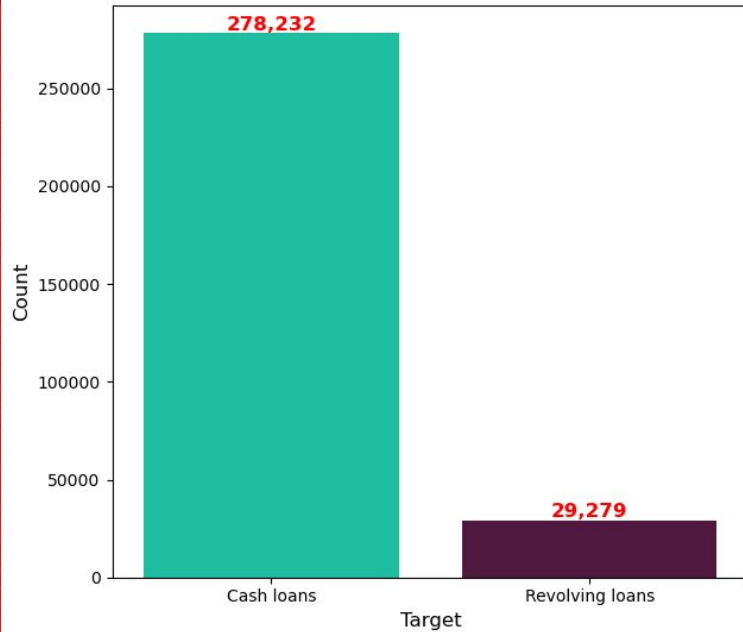
Percentages of Gender Type



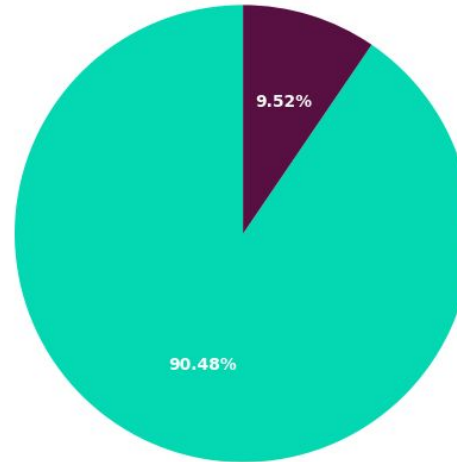
Dari grafik disamping sebaran wanita sebanyak 65.83% dan pria sebanyak 34,16%.

DISTRIBUTION CONTRACT TYPE

Counts of Contract Type



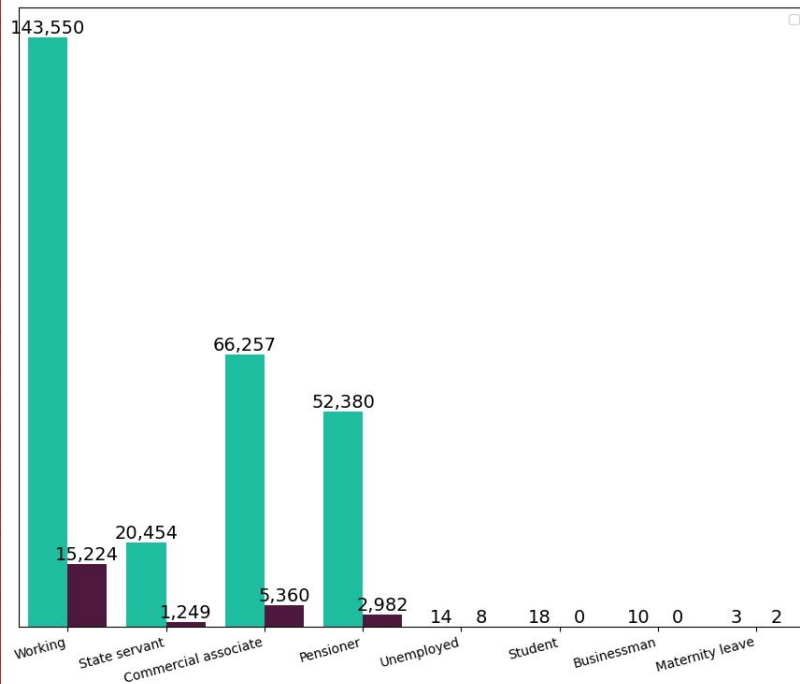
Percentages of Contract Type



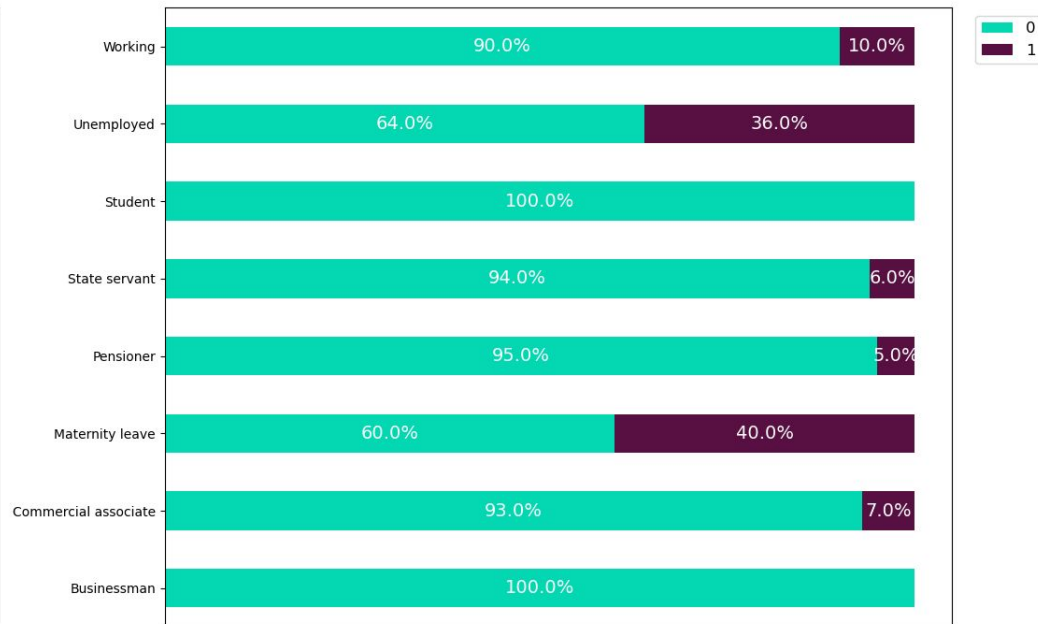
Dari grafik disamping tipe kontrak didominasi oleh pinjaman tunai sebesar 90,49% dari keseluruhan data

INCOME TYPE BY TARGET

Target by Income Type



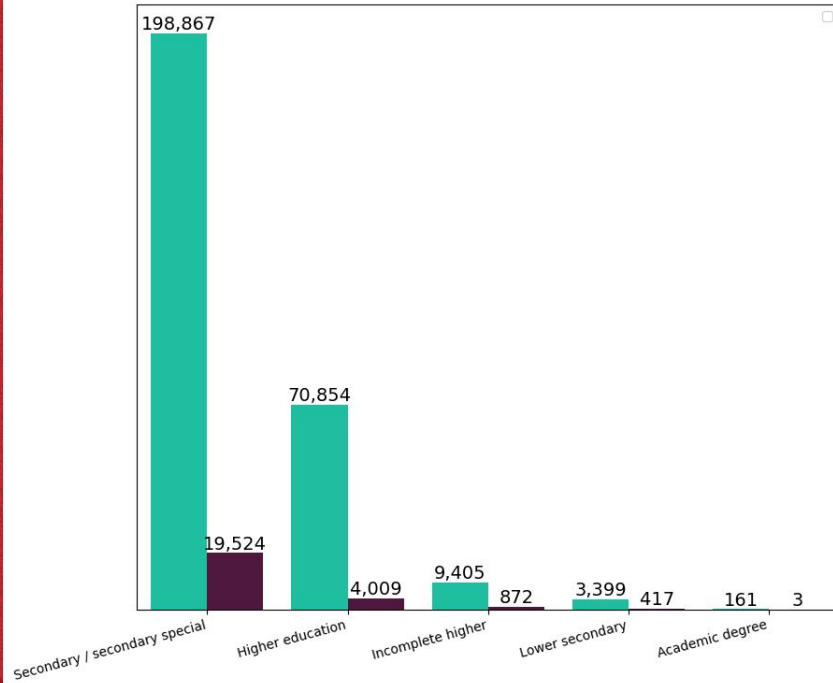
Ratio of Target by Income Type



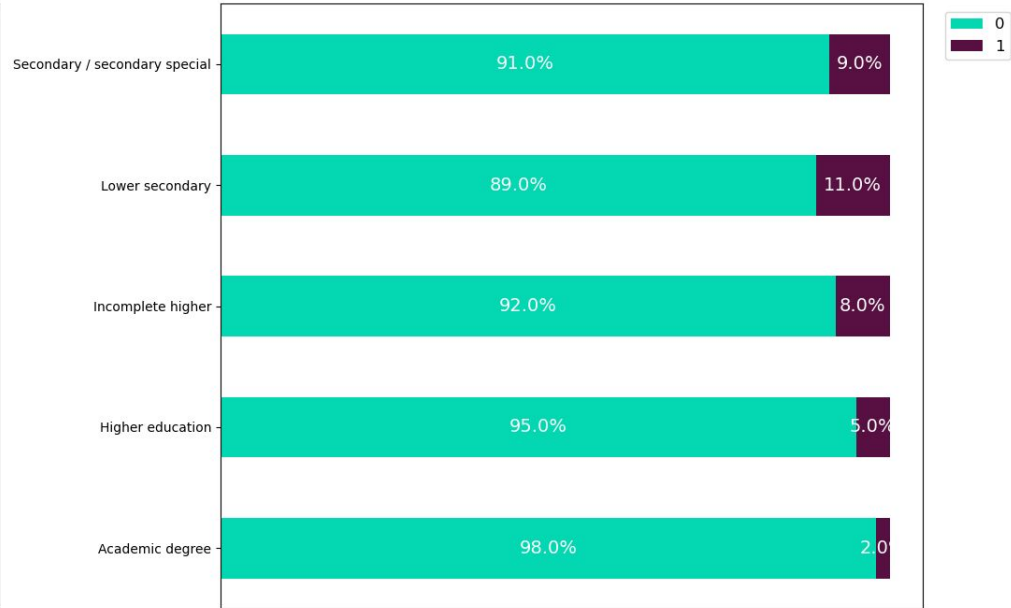
Dapat dilihat bahwa klien dengan tipe income working memiliki total kredit tidak lancar terbanyak sebanyak 15.224 orang dan tipe income commercial associate memiliki total kredit tidak lancar sebanyak 5.360 orang. Dari rasio target berdasarkan tipe income maternity leave dan unemployed memiliki perbandingan lebih banyak melakukan kredit tidak lancar yaitu 40% dan 36%.

EDUCATION TYPE BY TARGET

Target by Education Type



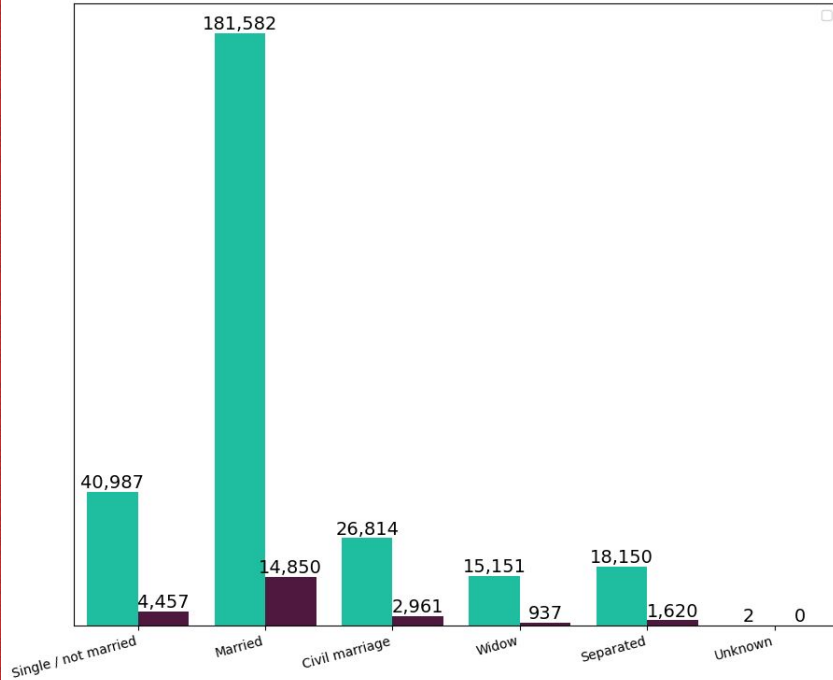
Ratio of Target by Education Type



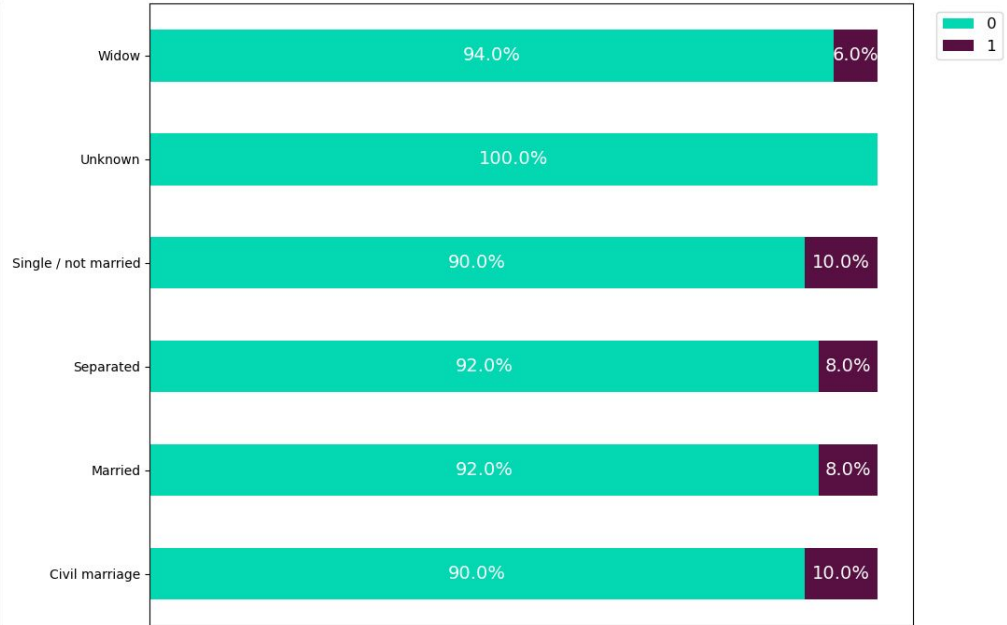
Dapat dilihat bahwa klien dengan tipe pendidikan secondary special memiliki total kredit tidak lancar terbanyak sebanyak 19.524 orang dan tipe pendidikan higher education memiliki total kredit tidak lancar sebanyak 4.009 orang. Dari rasio target berdasarkan tipe pendidikan lower secondary dan secondary special memiliki perbandingan lebih banyak melakukan kredit tidak lancar yaitu 11% dan 9%.

MARRIED TYPE BY TARGET

Target by Married Type



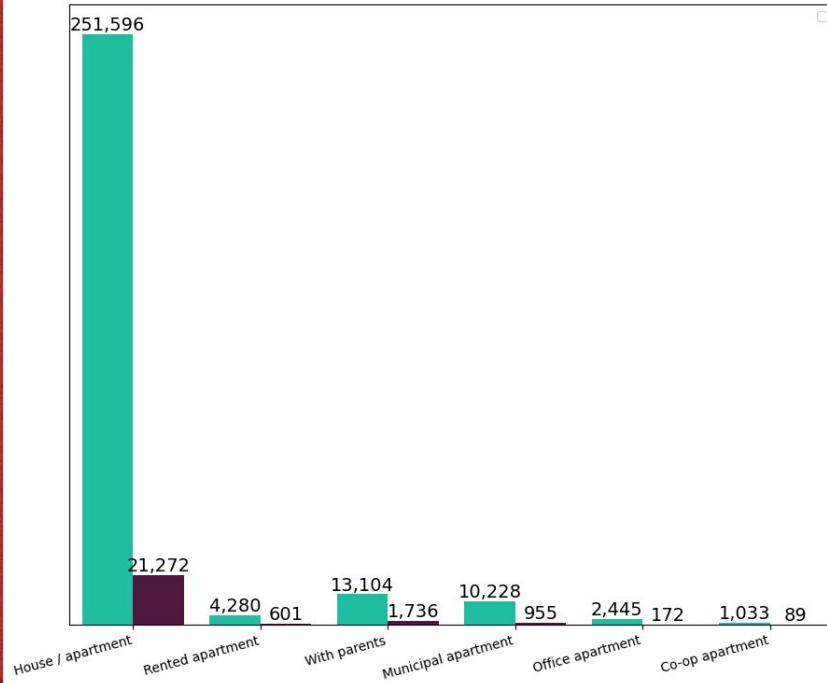
Ratio of Target by Married Type



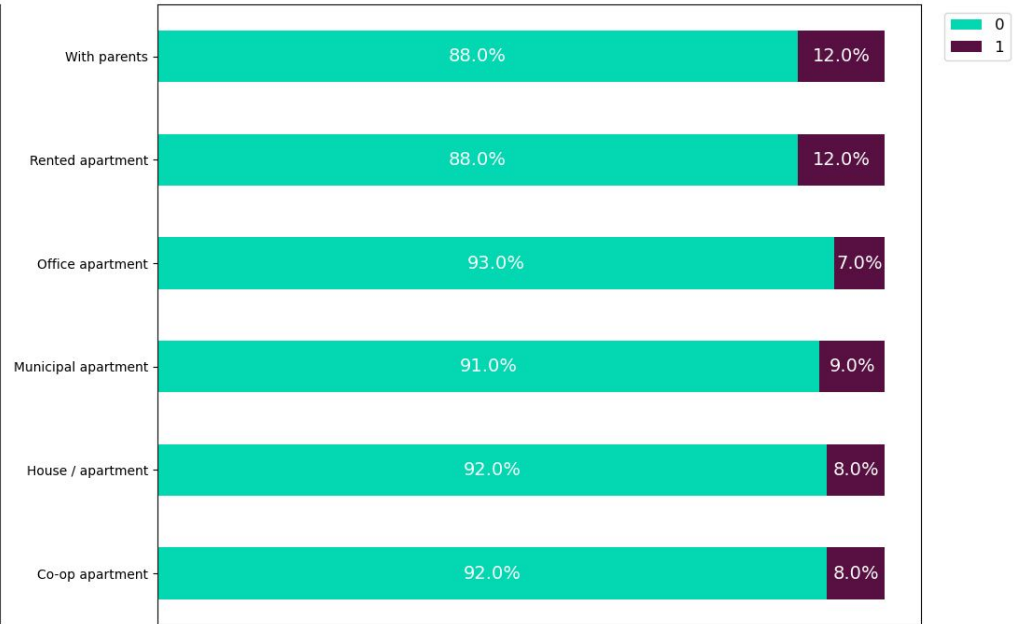
Dapat dilihat bahwa klien dengan tipe married, married memiliki total kredit tidak lancar terbanyak sebanyak 14.850 orang dan single memiliki total kredit tidak lancar sebanyak 4.457 orang. Dari rasio target berdasarkan tipe married, civil married dan married memiliki perbandingan lebih banyak melakukan kredit tidak lancar yaitu 10% dan 10%.

HOUSING TYPE BY TARGET

Target by Housing Type



Ratio of Target by Housing Type



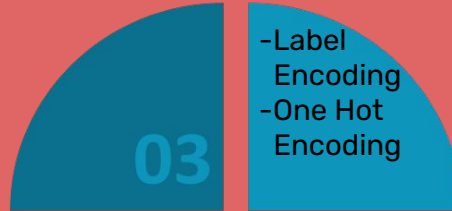
Dapat dilihat bahwa klien dengan tipe tempat tinggal, house/apartment memiliki total kredit tidak lancar terbanyak sebanyak 21.272 orang dan with parents memiliki total kredit tidak lancar sebanyak 1.736 orang. Dari rasio target berdasarkan tipe tempat tinggal, rented apartment dan with parents memiliki perbandingan lebih banyak melakukan kredit tidak lancar yaitu 12% dan 12%.

PRE-PROCESSING

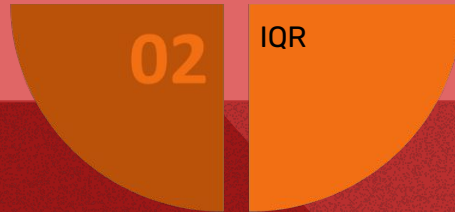
HANDLING MISSING VALUE



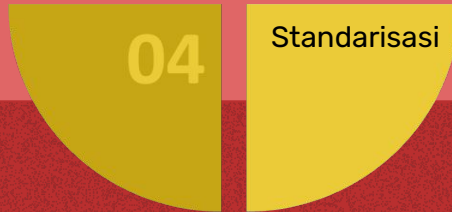
FEATURE ENCODING



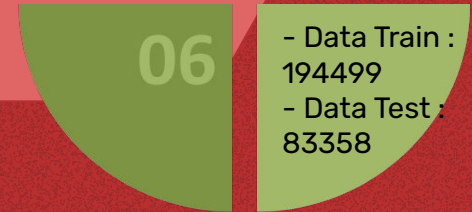
FEATURE SELECTION



HANDLING OUTLIER



FEATURE TRANSFORMATION



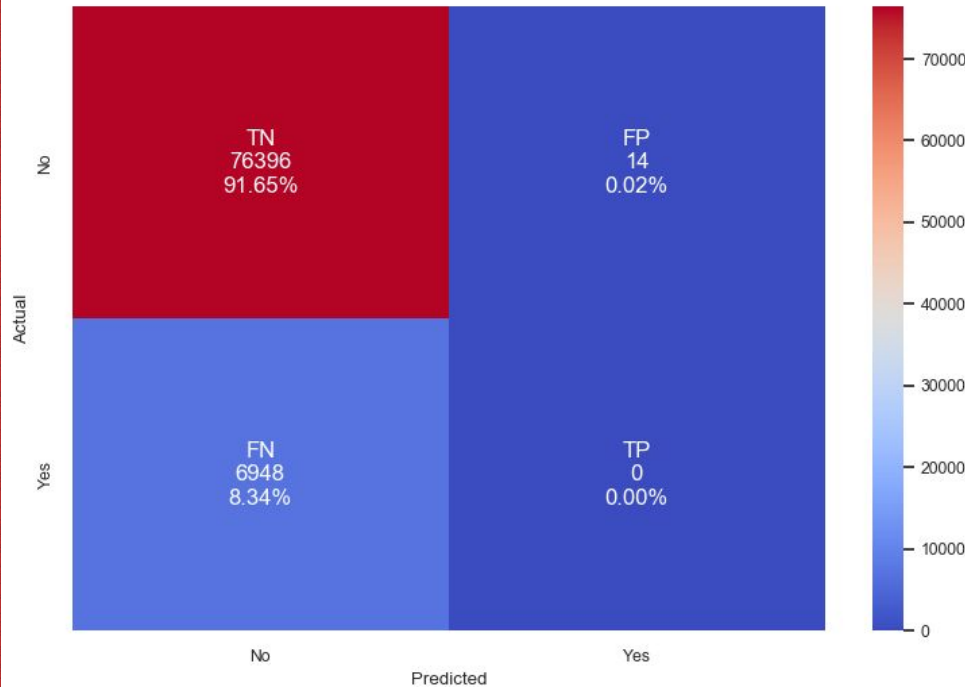
SPLIT DATA TRAIN & TEST (70:30)

MACHINE LEARNING MODEL

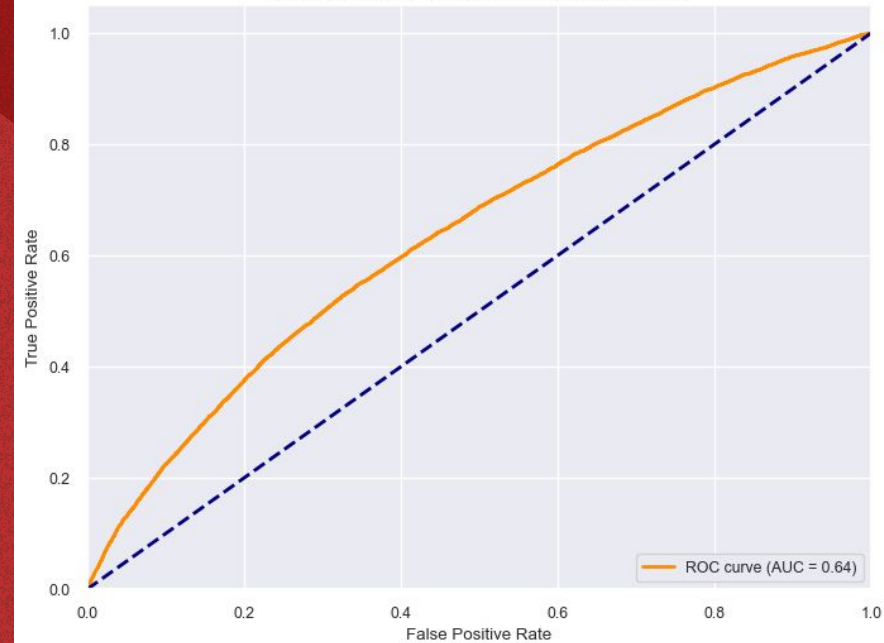
Model				
	Logistic Regression	Random Forest	Decision Tree	XgBoost
Accuracy (Test Set)	0,92	0,92	0,84	0,92
Precision (Test Set)	0	0,24	0,11	0
Recall (Test Set):	0	0	0,14	0
F1-Score (Test Set):	0	0	0,12	0
roc_auc (test-proba)	0,48	0,58	0,52	0,64
roc_auc (train-proba)	0,48	1	1	0,77
roc_auc (crossval test):	0,48	0,57	0,51	0,59
roc_auc (crossval train)	0,48	1	1	0,75

CONFUSION MATRIX & ROC AUC

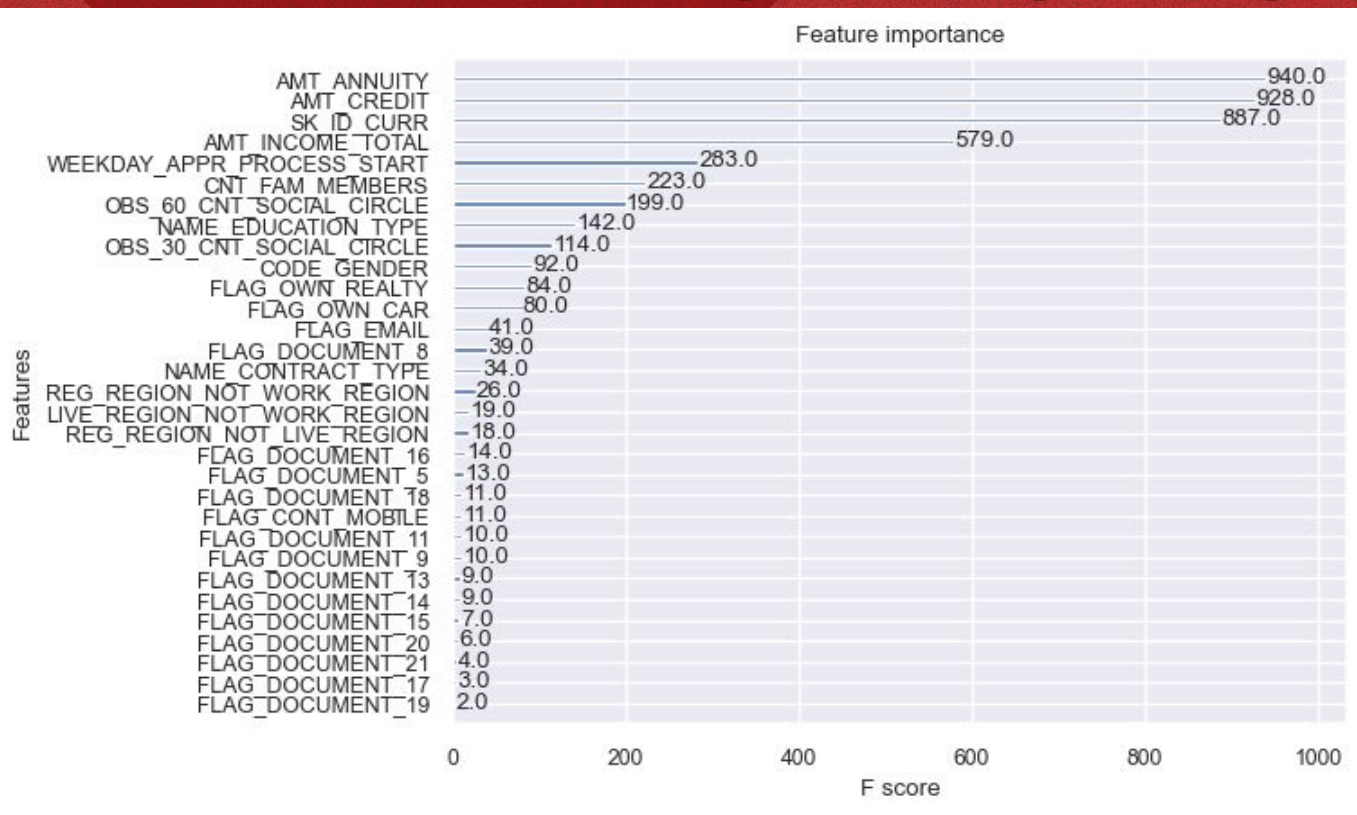
Confusion Matrix XGBoost



Receiver Operating Characteristic (ROC) Curve XgBoost



FEATURE IMPORTANCE



Feature **AMT_ANNUITY**,
AMT_CREDIT, dan
AMT_INCOME_TOTAL
menjadi fitur paling penting
berdasarkan hasil modeling

BUSINESS RECOMMENDATION



- Target segmentasi berdasarkan pendapatan total klien
- Mencari target klien yang memiliki tempat tinggal dan memiliki mobil
- Buat promosi untuk klien yang sudah menikah untuk melakukan kredit rumah

The background of the image is a vibrant red with a sunburst or starburst pattern. Numerous thin, dark red lines radiate from the center towards the edges, creating a dynamic, energetic feel. The lines are evenly spaced and extend to the corners of the frame.

THANK YOU