

# FINAL PROJECT MACHINE LEARNING MODELING

Yayang Nurul Aulia Azizah (Statistics Student)  
Virtual Internship Experience Data Scientist at Home Credit  
Indonesia by Rakamin Academy

Referensi: <https://github.com/yangswell12/fsb-project2>

Link Github: <https://github.com/yangswell12/vixrakamin>

# Problem Research

## **Business Question**

- Berapa persen client yang mengalami kesulitan pembayaran di Home Credit Indonesia?
- Kriteria client seperti apa yang tidak mengalami kesulitan pembayaran dengan syarat pinjaman sebelumnya diterima di Home Credit Indonesia?

## **Goals**

- Membangun model Machine Learning dengan tujuan agar client yang tidak mengalami kesulitan dalam pelunasan tidak tertolak saat mengajukan pinjaman.

# Data Preprocessing

Berikut tahapan yang dilakukan

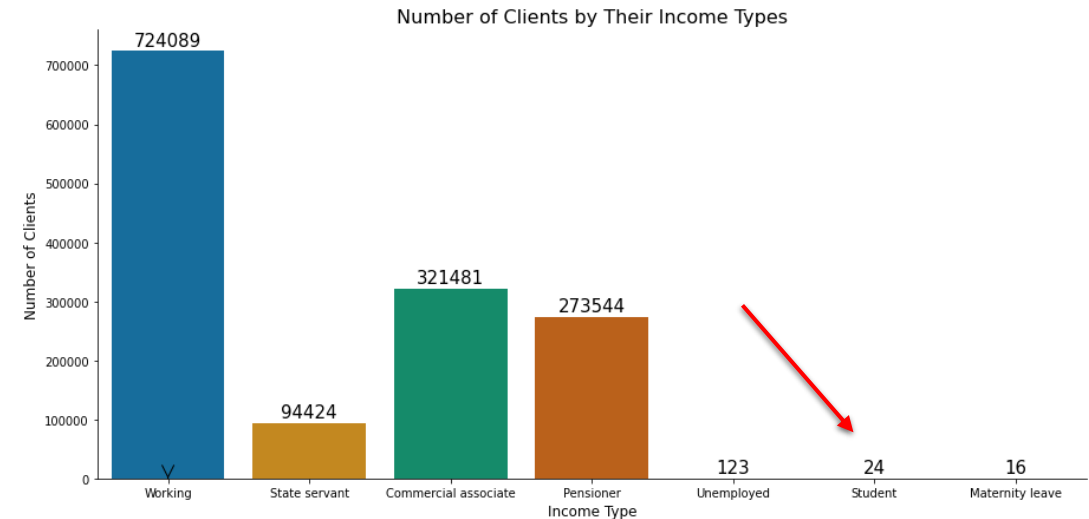
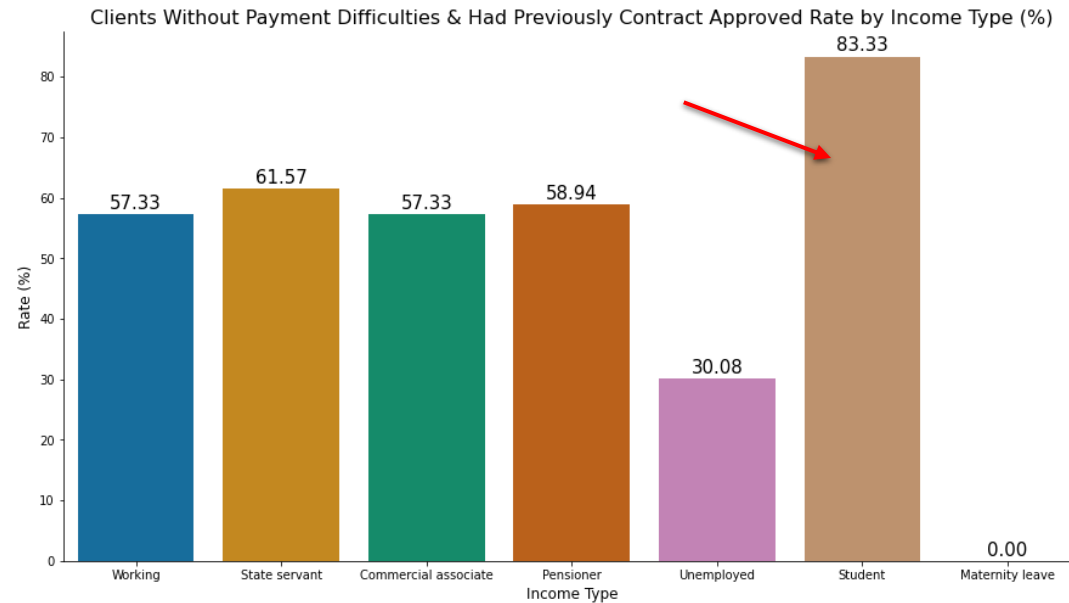
1. Check & Handling Missing Values (using median for numeric features and mode for categorical features).
2. Check & Handling Duplicated Values
3. Check & Handling Outliers (using z-score)
4. Check & Handling Inconsistent Data
5. Check & Handling Imbalanced Class (using SMOTE-NC)

# Feature Engineering

Berikut tahapan yang dilakukan

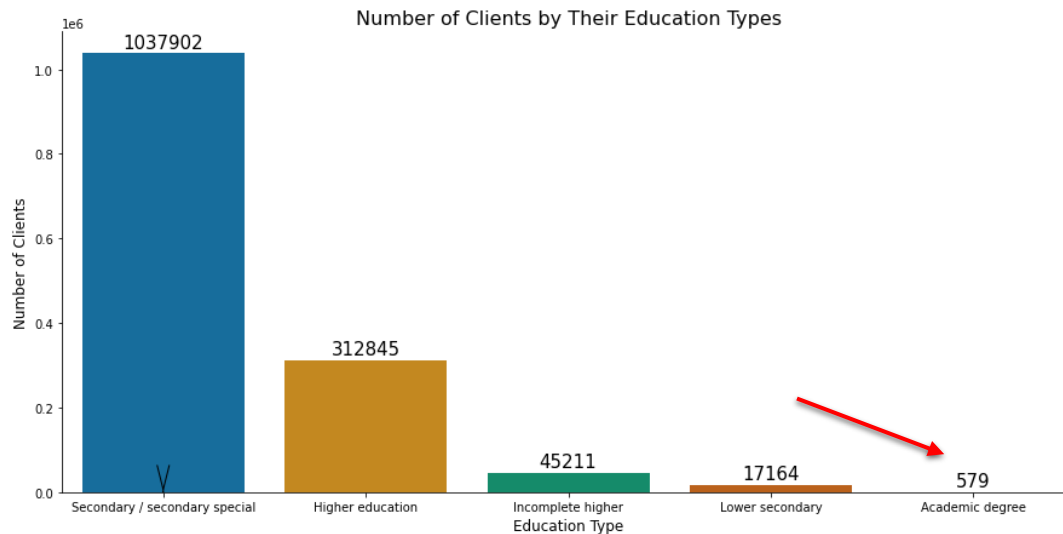
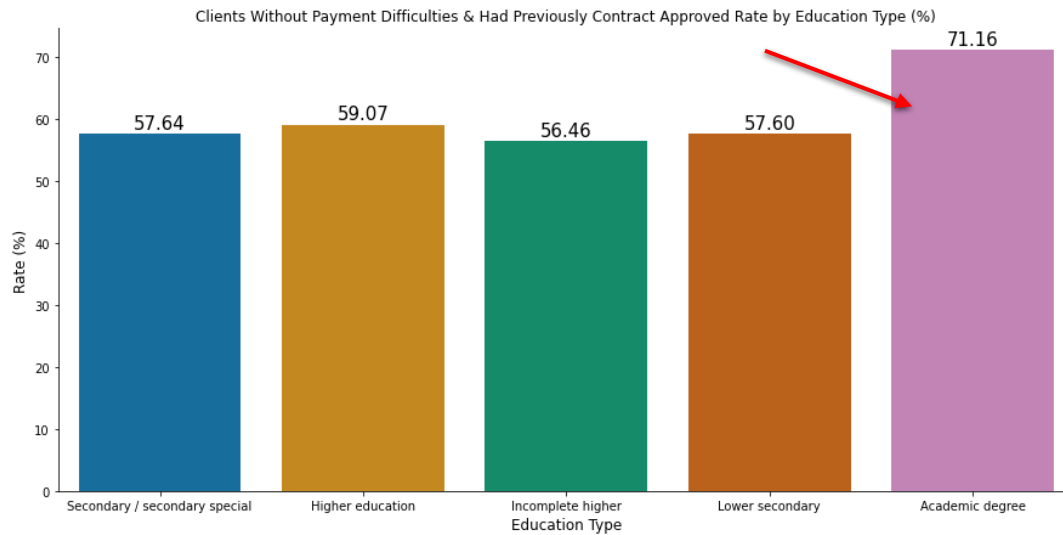
1. Feature Scaling (using z-score normalization)
2. Feature Encoding (using one-hot encoding)

# Data Visualization & Insight



Berdasarkan diagram batang di atas, Student Client memiliki rate tertinggi sebesar 83.3% untuk client yang pinjaman sebelumnya diterima & tidak mengalami kesulitan pembayaran. Namun, total Student Client yang mengajukan pinjaman hanya 24 orang atau sekitar 0.0017% dari total client yang mengajukan pinjaman. Sehingga diperlukan promosi atau campaign kepada Student Client lainnya agar semakin banyak yang tertarik untuk mengajukan pinjaman di Home Credit Indonesia.

# Data Visualization & Insight



Berdasarkan diagram batang di samping, Client dengan gelar akademik memiliki rate tertinggi sebesar 71,16% untuk client yang pinjaman sebelumnya diterima & tidak mengalami kesulitan pembayaran. Namun, total Client dengan gelar akademik yang mengajukan pinjaman hanya 579 orang atau sekitar 0,04% dari total client yang mengajukan pinjaman. Sehingga diperlukan promosi atau campaign kepada Client dengan gelar akademik lainnya agar semakin banyak yang tertarik untuk mengajukan pinjaman di Home Credit Indonesia.

## Data Modelling & Evaluation I

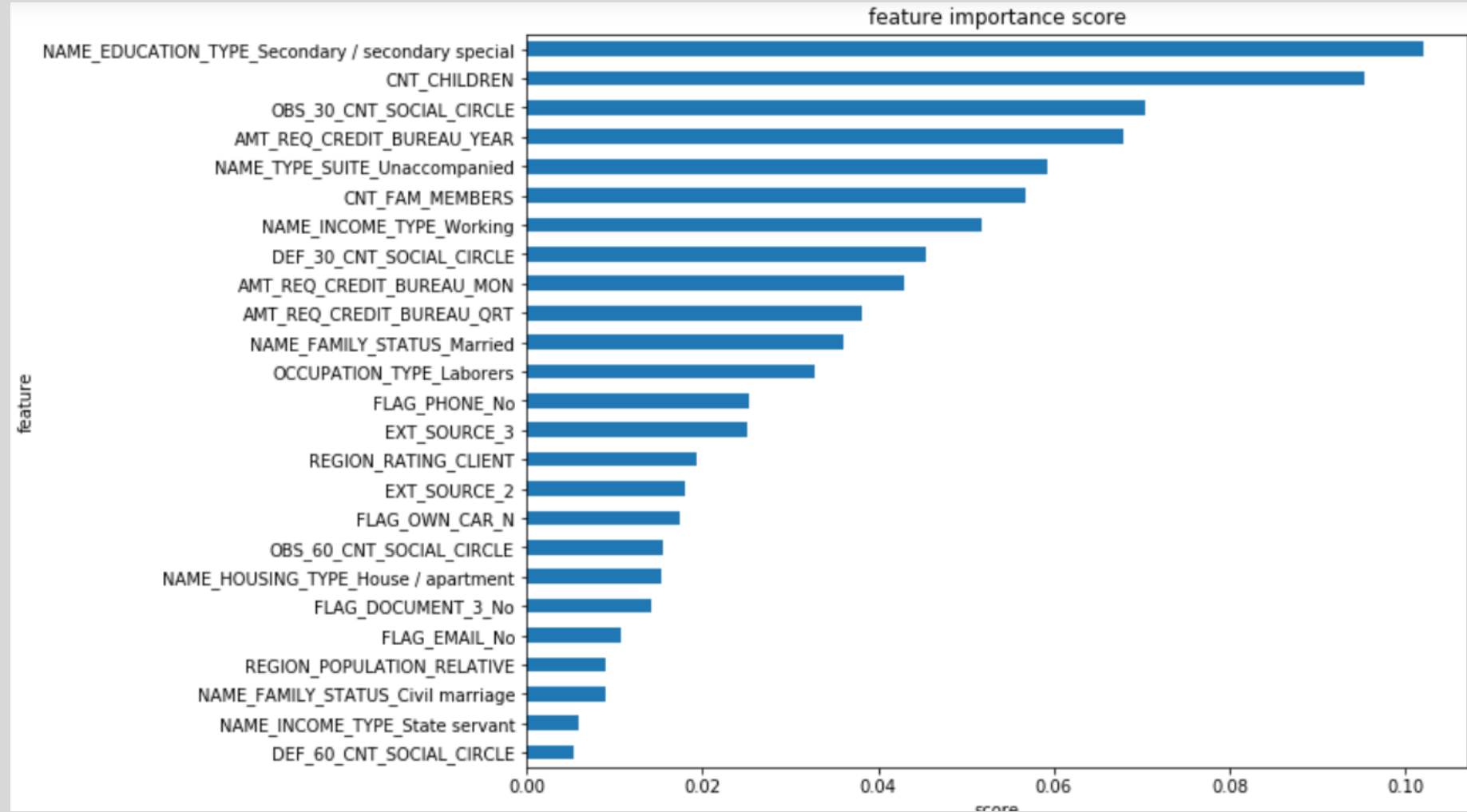
- Semua feature dimasukkan ke dalam model kecuali SK\_ID\_CURR & ORGANIZATION\_TYPE
- Evaluasi model menggunakan Precision dengan tujuan untuk mengurangi kesalahan false positive, yakni client yang bisa melunasi pembayaran terprediksi sebagai client yang tidak bisa melunasi pembayaran.

	Model	Precision
0	Logistic Regression	0.8059
1	Random Forest	0.9802
2	XGBoost	0.9951

	Model	Training Score	Test Score
0	Logistic Regression	0.8066	0.8059
1	Random Forest	1.0000	0.9802
2	XGBoost	0.9978	0.9951

Pada hasil evaluasi (tabel kiri), model XGBoost memiliki nilai Precision yang lebih tinggi daripada dua model lainnya. Lalu pada hasil evaluasi (tabel kanan) model Random Forest mengalami overfitting, yakni score data trainingnya data testingnya mengalami perbedaan yang paling besar dibandingkan dua model lainnya. Sehingga model terbaik yang dipilih adalah model XGBoost yang memiliki nilai Precision sangat tinggi dan tidak terjadi overfitting.

## Feature Importances berdasarkan model terbaik (XGBoost)





# Data Modelling & Evaluation II

Feature yang dimasukkan ke dalam model terbaik (XGBoost) adalah 25 feature yang memiliki skor feature importance tertinggi.

	Model	Training Score	Test Score
0	XGBoost	0.9992	0.9982

Berdasarkan hasil pada tabel di atas, terlihat skor Precision model XGBoost yang menggunakan 25 feature mengalami peningkatan dibandingkan sebelumnya yang menggunakan keseluruhan feature. Model ini merupakan model akhir yang akan digunakan untuk memprediksi client yang mampu melunasi pembayarannya agar tidak tertolak saat mengajukan pinjaman.

# Summary & Recommendation

- Sebesar 8.19% client mengalami kesulitan dalam pembayaran di Home Credit Indonesia.
- Student Client memiliki rate tertinggi sebesar 83.3% untuk client yang pinjaman sebelumnya diterima & tidak mengalami kesulitan pembayaran. Namun, total Student Client yang mengajukan pinjaman hanya 24 orang atau sekitar 0.0017% dari total client yang mengajukan pinjaman. Sehingga diperlukan promosi atau campaign kepada Student Client lainnya agar semakin banyak yang tertarik untuk mengajukan pinjaman di Home Credit Indonesia.
- Client dengan gelar akademik memiliki rate tertinggi sebesar 71,16% untuk client yang pinjaman sebelumnya diterima & tidak mengalami kesulitan pembayaran. Namun, total Client dengan gelar akademik yang mengajukan pinjaman hanya 579 orang atau sekitar 0,04% dari total client yang mengajukan pinjaman. Sehingga diperlukan promosi atau campaign kepada Client dengan gelar akademik lainnya agar semakin banyak yang tertarik untuk mengajukan pinjaman di Home Credit Indonesia.
- Model terbaik yang dipilih adalah model XGBoost yang memiliki nilai Precision tertinggi dan tidak terjadi overfitting.