

# HOME CREDIT

**Maulana Kavaldo**

 [Maulana Kavaldo](#)

 [Maulana Kavaldo](#)

 [maulana.kavaldo](#)

# Problem Research

1. Ingin mengetahui apa saja kriteria pelanggan yang tidak dan atau memiliki masalah dalam proses pelunasan kredit?
2. Membangun machine learning untuk membantu tim menentukan apakah pengajuan pinjaman dari nasabah akan mengalami kendala dalam proses pelunasan?

# Data Pre-Processing

**Dataset: application\_train.csv**

Terdapat 124 kolom dan 307 ribu baris lebih yang berisi data *customers* yang memiliki masalah dan yang tidak bermasalah dalam melakukan pembayaran/pelunasan pinjaman.

Data Cleaning

Missing Values  
Duplicate Data

Feature Reduction

Correlation

Data Transformation

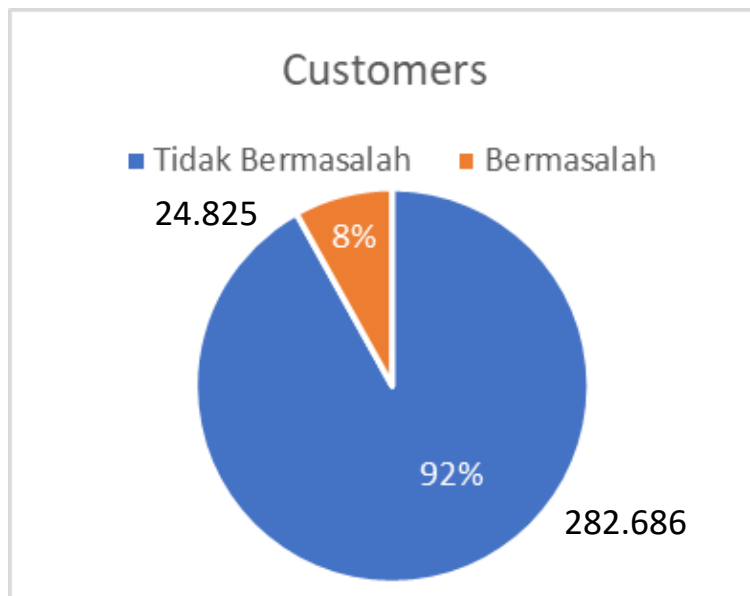
Features:  
Numerical: MinMax Scaller  
Categorical: OneHotEncoder

# Data Visualization and Business Insight

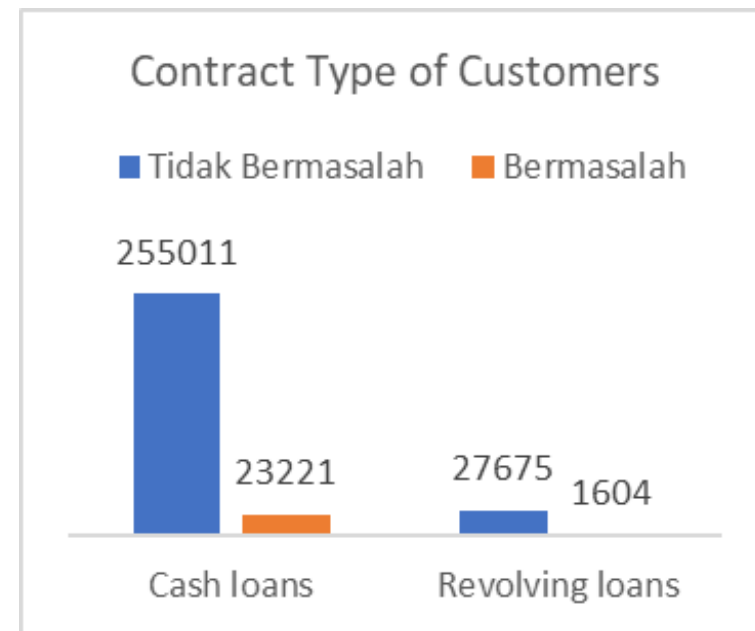
## General Information

### Total Customers Loan

**307511**

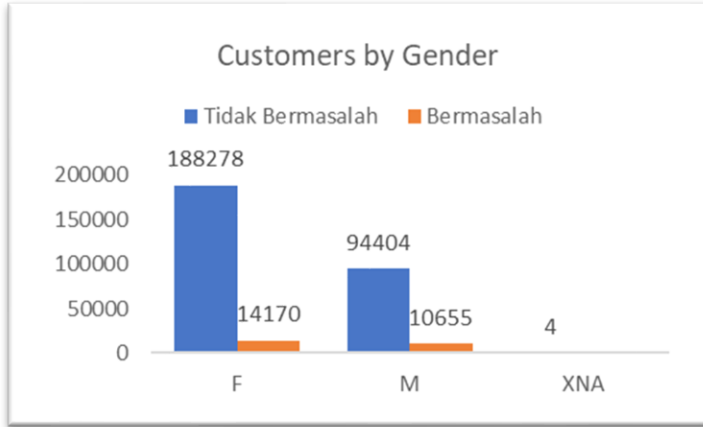


Customers dengan kesulitan pembayaran pinjaman sebanyak 24825 atau 8% dari data.

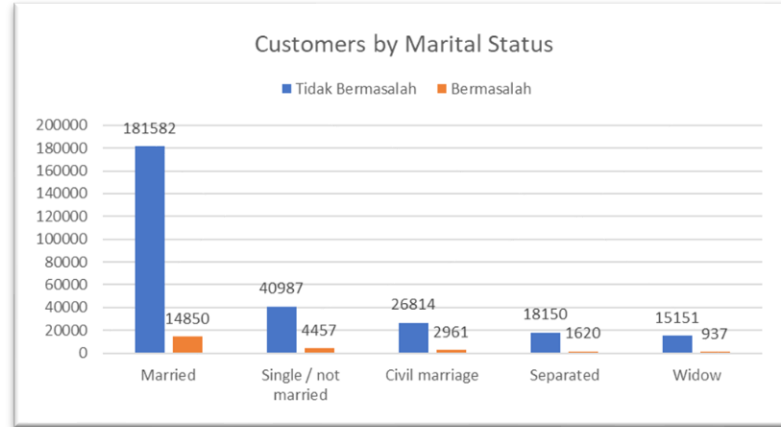


Customers terbanyak berdasarkan tipe kontrak. Tipe *Cash Loan* merupakan tipe terfavorit.

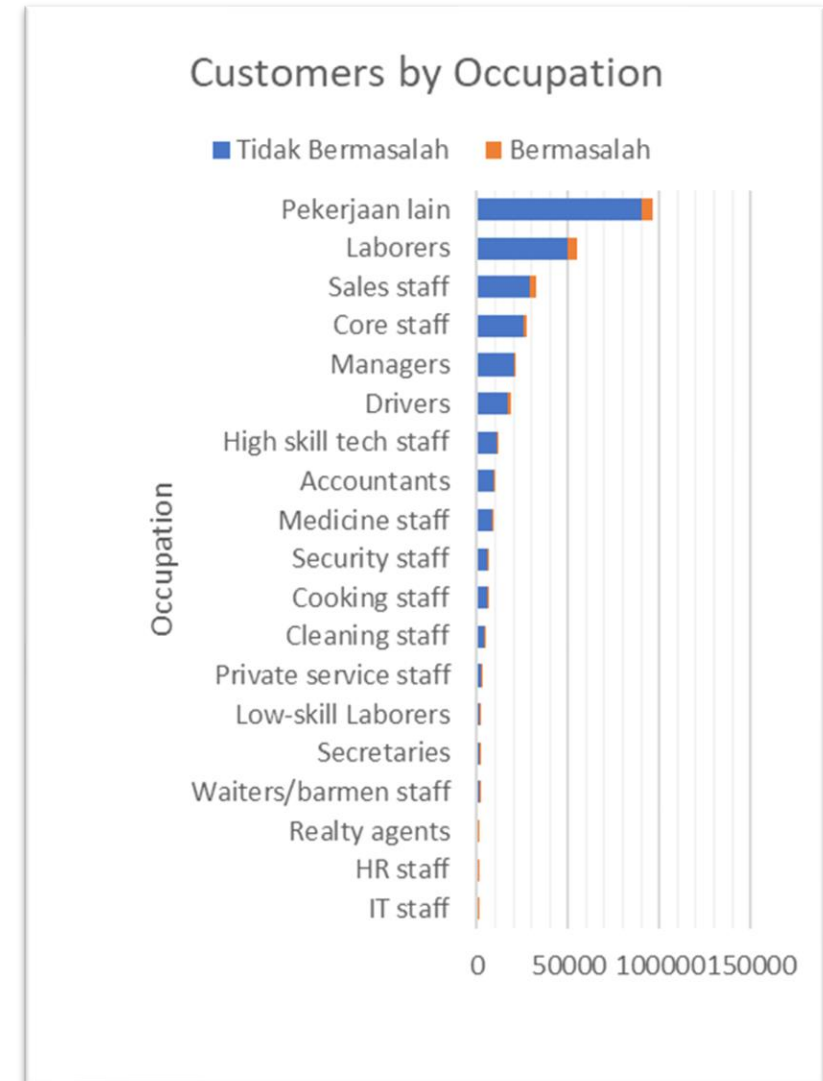
Continue...



Customers dengan jenis kelamin wanita memiliki rasio yang lebih besar dengan masalah dalam pembayaran pinjaman.



Customers dengan status menikah merupakan kebanyakan dari customers yang tidak memiliki masalah dalam pembayaran.

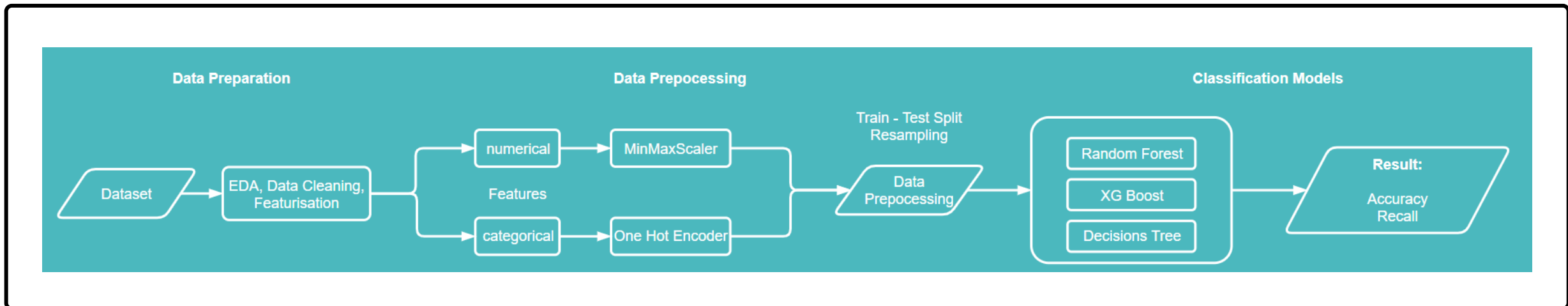


Customers didominasi dengan jenis pekerjaan lain (tanpa profesi yang spesifik) .

# Machine Learning Implementation and Evaluation

Pada pemodelan menggunakan model machine learning:

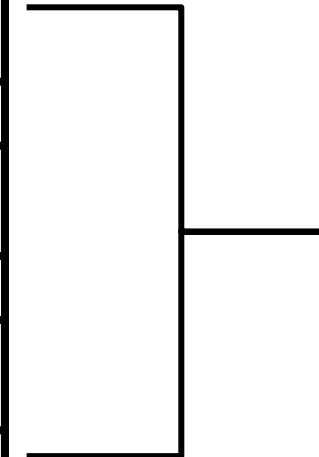
- a) XGBoost
- b) Decision Tree
- c) Random Forest



*Machine Learning Work Flow*

# Comparison of Model

Model	Data	Train		Test	
		Accuracy (%)	Recall (%)	Accuracy (%)	Recall (%)
Random Forest	Original Data	100	100	92	100
	Random Under Sampling	100	100	100	100
XGB Classifier	Original Data	92	100	92	100
	Random Under Sampling	50	100	50	100
Decision Tree	Original Data	100	100	84	90
	Random Under Sampling	100	100	100	100
Logistic Regression	Original Data	92	100	92	100
	Random Under Sampling	50	100	50	100



**Random Forest  
Decision Tree**

Model Random Forest dan Decision Tree memiliki akurasi dan recall yang baik. Model ini dapat digunakan

# Business Recommendation

- Buat iklan atau promosi untuk staf SDM, staf TI, dan agen properti untuk mengajukan kredit.
- Membuat kampanye untuk pelanggan yang berprofesi sebagai laborer karena merupakan jumlah pelanggan terbesar dan persentase pembayaran yang berhasil.
- Random Forest dan Decision Tree merupakan model *machine learning* yang dipilih untuk membantu tim menentukan apakah pelanggan memiliki masalah dalam melunasi pinjaman atau tidak.





***Thank you!***

*Repo Project:*



[VIX-HomeCreditIndonesia](https://github.com/VIX-HomeCreditIndonesia)