



# Home Credit Scorecard Machine Learning Model

By Akbar Adi Nugroho

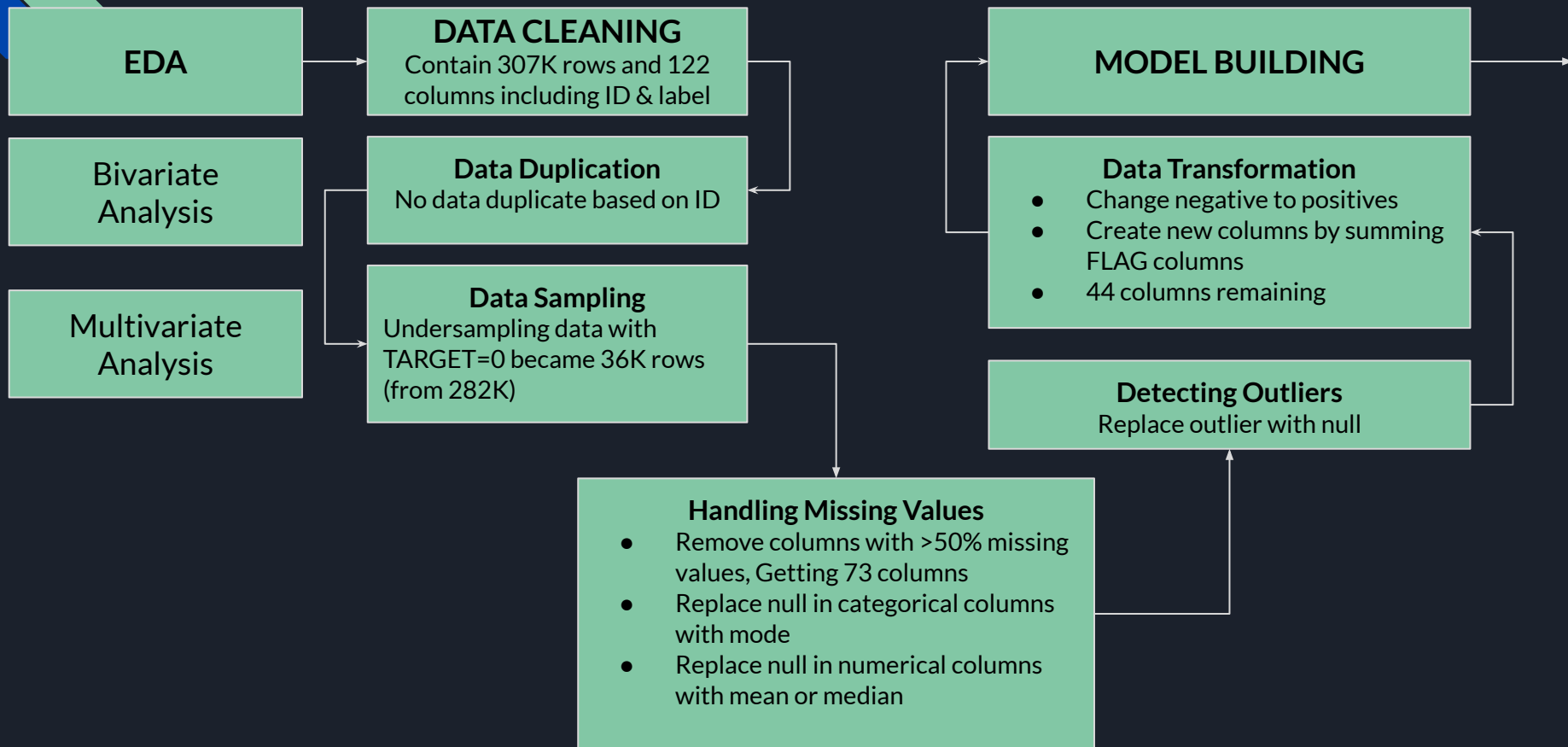
<https://github.com/akbaradie/home-credit-scoring>



# Problem Research

Many people face **challenges in obtaining loans** because they **lack sufficient or any credit history**. Home Credit is striving to expand financial inclusion for individuals who are not served by traditional banking services, by offering a favorable and secure borrowing process. To ensure a positive loan experience, Home Credit employs alternative data sources to **predict their clients' repayment capabilities**. This approach aims to prevent **deserving clients from being denied loans** and ensures that loans are provided with terms and repayment schedules that enable clients to succeed financially.

# Data Preprocessing (1/2)



# Data Preprocessing (2/2)

## MODEL BUILDING

### PCA

Reduce dimension

- Expluatation
- Floorsmax
- Social Circle
- Family Member & Child
- Region Rating
- Amount of Credit & Goods
- 35 columns remaining

### Label & One-Hot Encoding

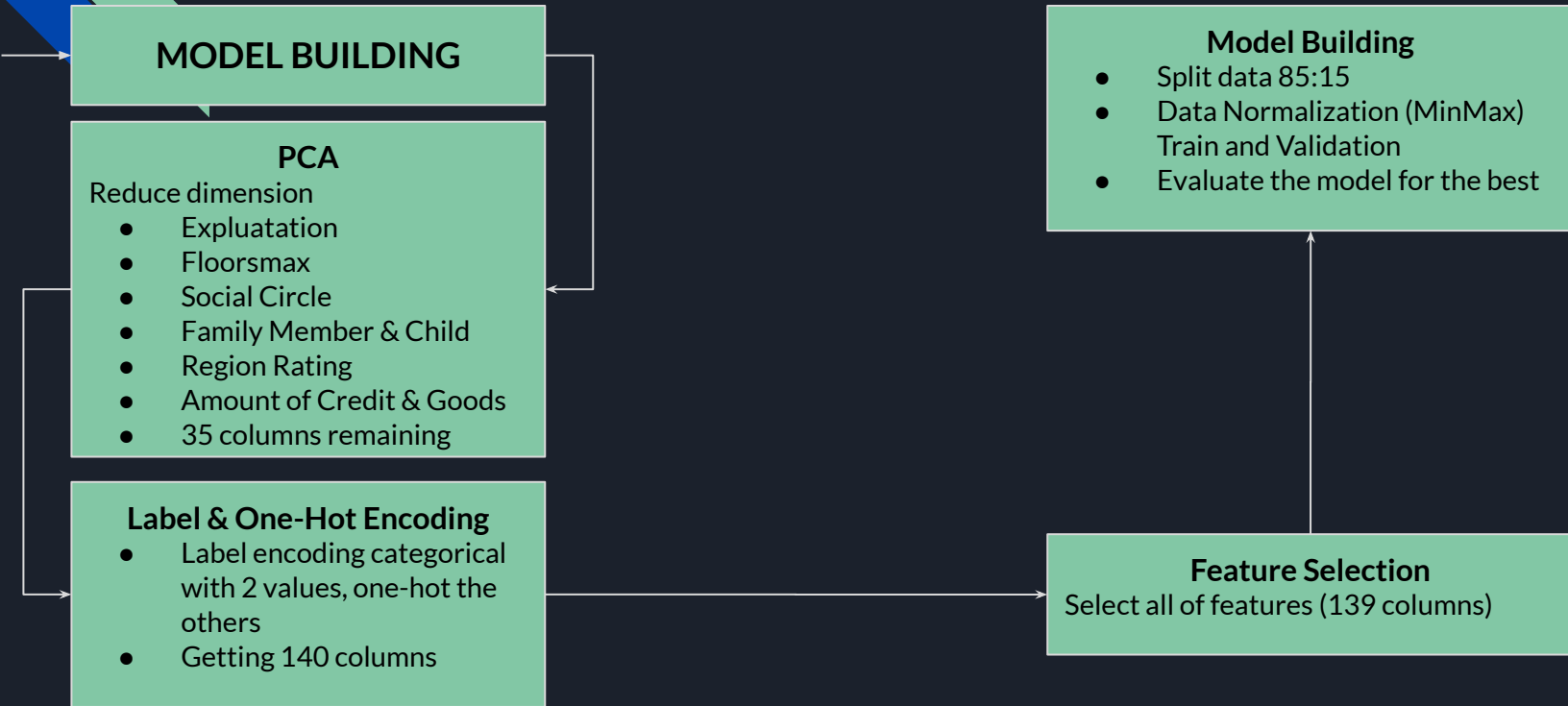
- Label encoding categorical with 2 values, one-hot the others
- Getting 140 columns

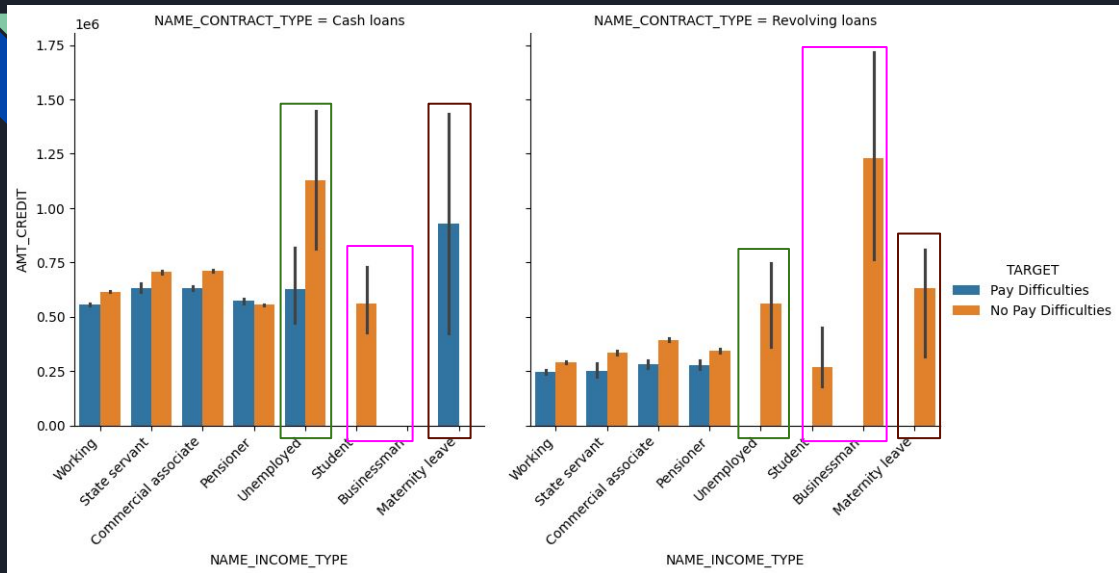
## Model Building

- Split data 85:15
- Data Normalization (MinMax) Train and Validation
- Evaluate the model for the best

## Feature Selection

Select all of features (139 columns)

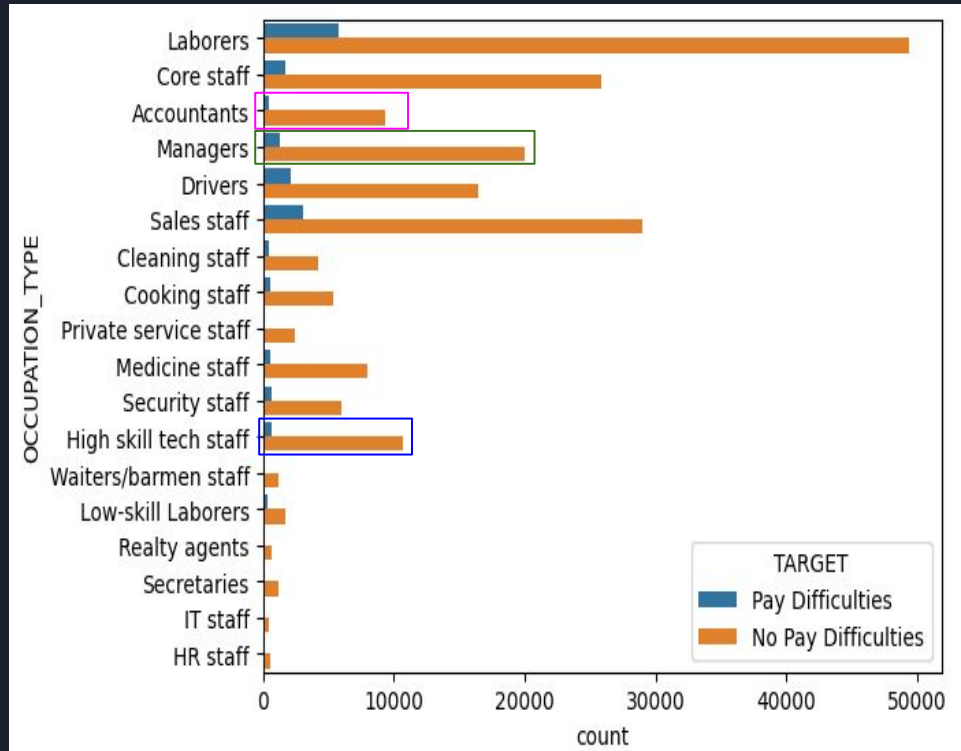
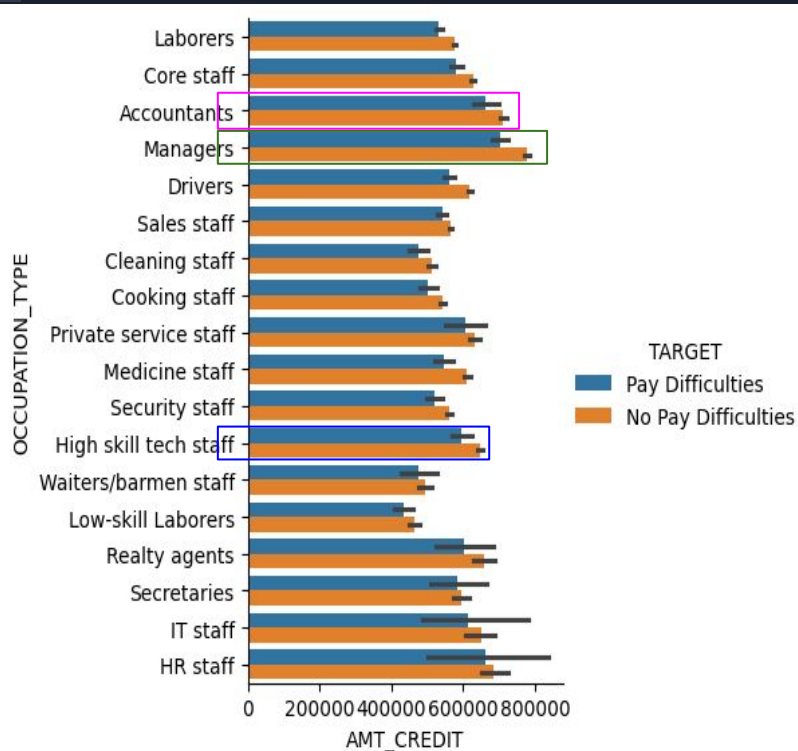




All **students** and **businessman** clients have **no problems** repaying the loans for a **low to medium** and **medium to high** credit amount of the loan respectively. All **businessman** clients prefer **revolving loans**.

All clients with **maternity leave** as income type and **cash loans** have **problems** repaying the loans for a **medium to high** amount of the loan. But, **NOT** for the **revolving loans**.

Only **unemployed** clients with **cash loans**, more than **50%**, who have **problems** in repaying the loans with **medium** credit amount.



Client who work as **accountants** can be said as client who **capable for repaying** the loans (95%) with **medium** credit amount of loan. But, **accountants** only **4,64%** of total applicants. So do **High skill tech staff** and **managers** (94%). They are **capable** but **only a few** applications from them, **10,1%** and **5,4%** of total respectively.

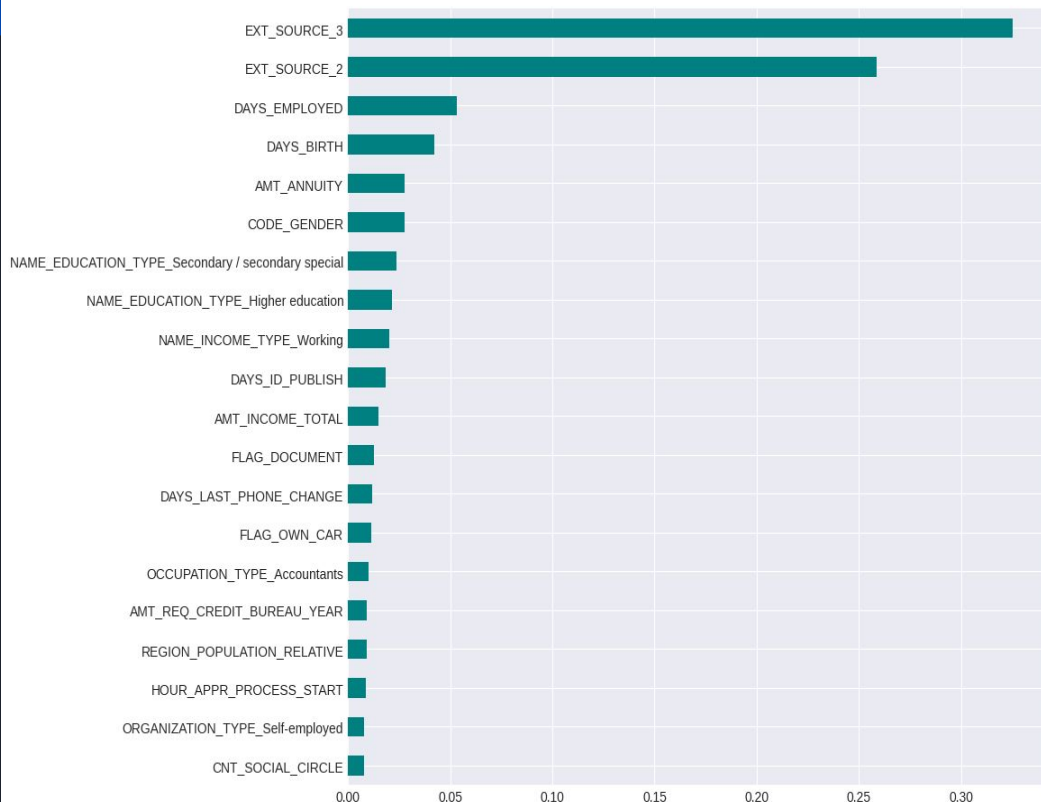


No	Model Name	Test Accuracy	Error Margin	Recall	Precision	AUC
1	Logistic Regression	69.4%	0.2%	49.4%	65.7%	73.8%
2	Decision Tree	56.3%	44%	49.4%	45.8%	55.2%
3	Ada Boost	68.6%	0.5%	42.3%	67.3%	72.4%
4	Gaussian NB	61.1%	0.6%	15.2%	55.5%	57.4%
5	XgBoost	66.7%	11.8%	52.8%	59.6%	70.3%
6	Random Forest	69%	31.8%	43.5%	67.6%	72.1%
7	Gradient Boosting	69.5%	0.6%	47.3%	66.9%	73.9%
8	Light GBM	69%	3.6%	52.3%	63.9%	73%

Accuracy value in train and test set in **Gradient Boosting** is not much different (0.06%) which is the model is very good, **no overfitting** and **underfitting**. So, we choose **Gradient Boosting** model to predict client's repayment abilities.

# Best Model

Features Importance Plot



Algorithm

Light Gradient  
Boosting Machine

## Performance

- 69.5% accuracy
- 73.9% ROC

Top 5 most  
important  
features

Source from external  
data source 3

Source from external  
data source 2

Days employed

Days birth

Amount of annuity





# Recommendation (1/2)

01

A Client with **student** as income type is **the most capable** for repaying the loans (100%) both cash or revolving loans. But, **students** are only 18 out of 307.511 applicants (0,006%).

02

A client who work as an **accountant** can be said as client who **capable** for repaying the loans (95%). But, an accountant only **4,64%** of total applicants. So do **high skill tech staff** and **managers**. They are **capable** but only **a few** applications from them.



Create a **campaign** to get more **students, accountants, high skill tech staff, and manager** interested in applying for a loan.



# Recommendation (2/2)

- 01 All clients with **maternity leave** as income type and **cash loans** have **problems** repaying the loans for a **medium to high** amount of the loan. But, **NOT** for the **revolving loans**.
- 02 Only **unemployed** clients with **cash loans**, more than 50%, who have **problems** in repaying the loans with **medium** credit amount.



Do a **research/survey** for **unemployed** and **maternity leave** clients about their problems in repaying the loans with **cash loans** type. So **in the future**, we can give recommendation about the **right contract type** for clients who have that type of income.