CREDIT SCORING

CREATED BY RAHMA YANTI

BACKGROUND

It is very important for a company to predict applicant who have a good score on loan credit, as it is **directly related to the loss and gain of income**.

On that occasion, we will make Machine Learning Modelling to predict **which application has payment difficulties or not with the credit score**.

Our objective is **Machine Learning modeling with AUC >0.7 and KS >0.3**, the value it is considered a good model on risk modeling.

As well, we will find **the insight to take effective action**.

Repository Link

DATASET OVERVIEW

Description

This data set contains **application record** for loan credit with payment difficulties, and includes others information such as how much amount credit that application have, family data status, application income, children and family, past behaviour credit, etc.
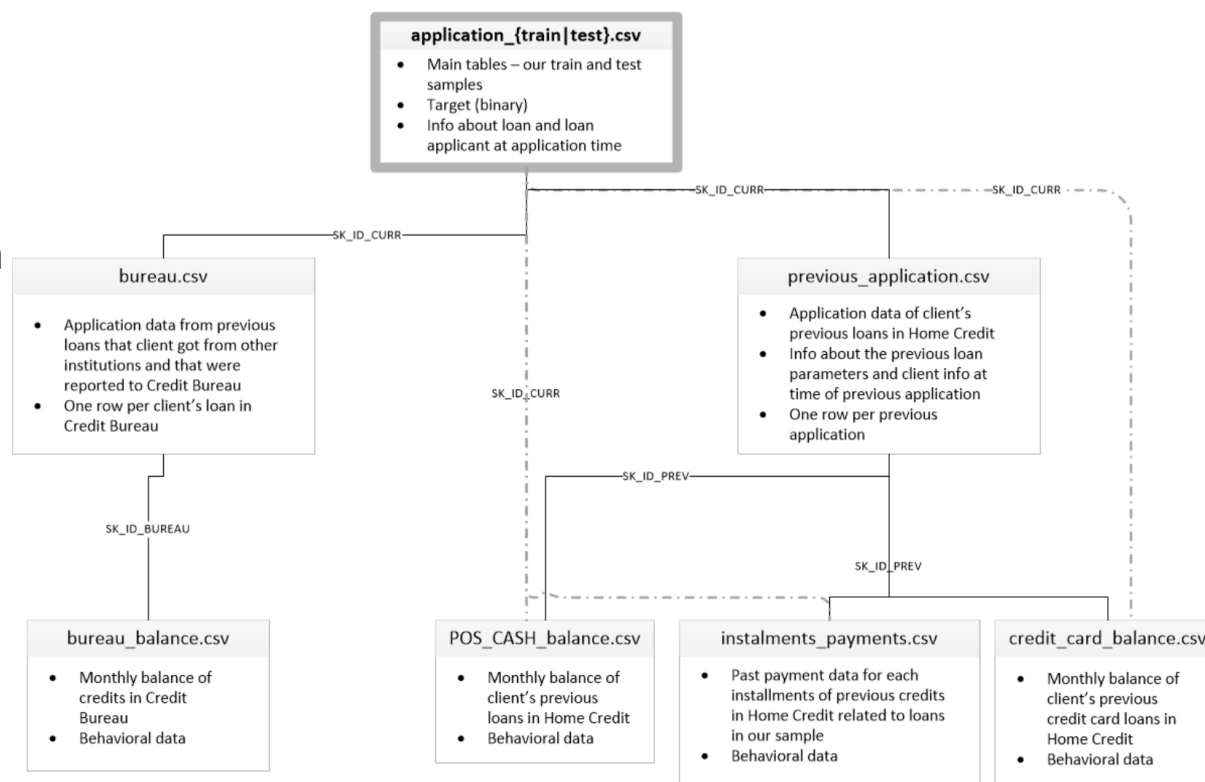
Shape

307511 data rows and **121** features

Missing Values

More than 50 features containing **>15%** missing data

Imbalanced Data

Target feature highly imbalance (92:8)

application_{train|test}.csv
- Main tables – our train and test samples
- Target (binary)
- Info about loan and loan applicant at application time

SK_ID_CURR

bureau.csv
- Application data from previous loans that client got from other institutions and that were reported to Credit Bureau
- One row per client's loan in Credit Bureau

previous_application.csv
- Application data of client's previous loans in Home Credit
- Info about the previous loan parameters and client info at time of previous application
- One row per previous application

SK_ID_BUREAU

SK_ID_PREV

SK_ID_PREV

bureau_balance.csv
- Monthly balance of credits in Credit Bureau
- Behavioral data

POS_CASH_balance.csv
- Monthly balance of client's previous loans in Home Credit
- Behavioral data

instalments_payments.csv
- Past payment data for each installments of previous credits in Home Credit related to loans in our sample
- Behavioral data

credit_card_balance.csv
- Monthly balance of client's previous credit card loans in Home Credit
- Behavioral data

Repository Link

HOME CREDIT

VIRTUAL INTERSHIP EXPERIENCE PROGRAM

DATASET
PRE PROCESSING

DROP UNUSED FILTER
>15% null data

FEATURE ENCODING

MISSING VALUES IMPUTATION

FEATURE SELECTION
WoE and IV
K Select

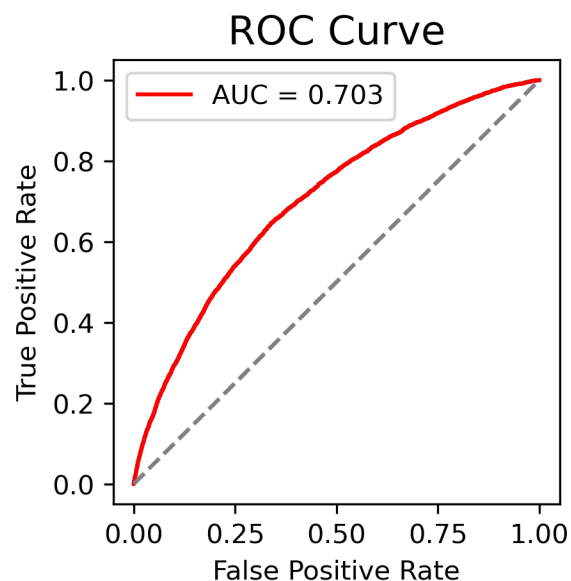**FEATURE ENGINEERING**
Days to years

HANDLING IMBALANCE
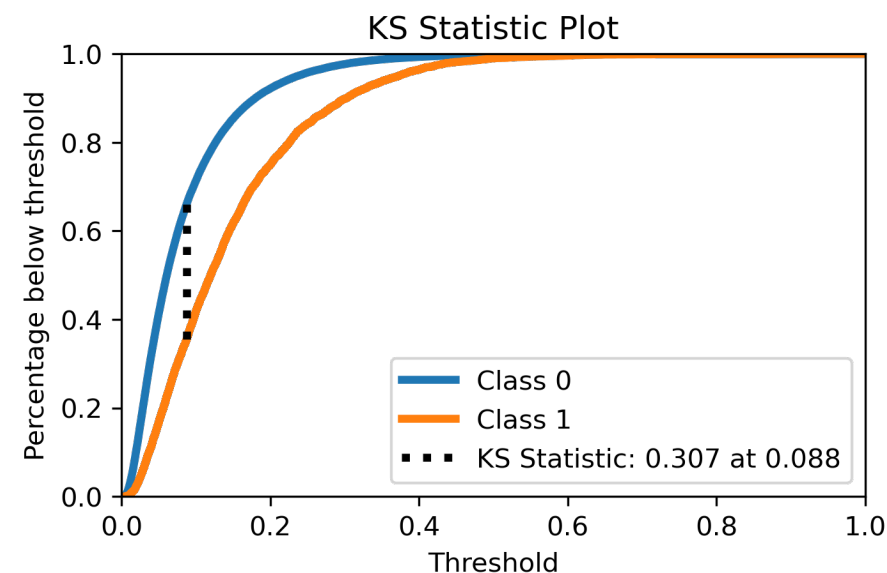SMOTE (Oversampling)

Repository
Link

CREDIT SCORE MODELLING

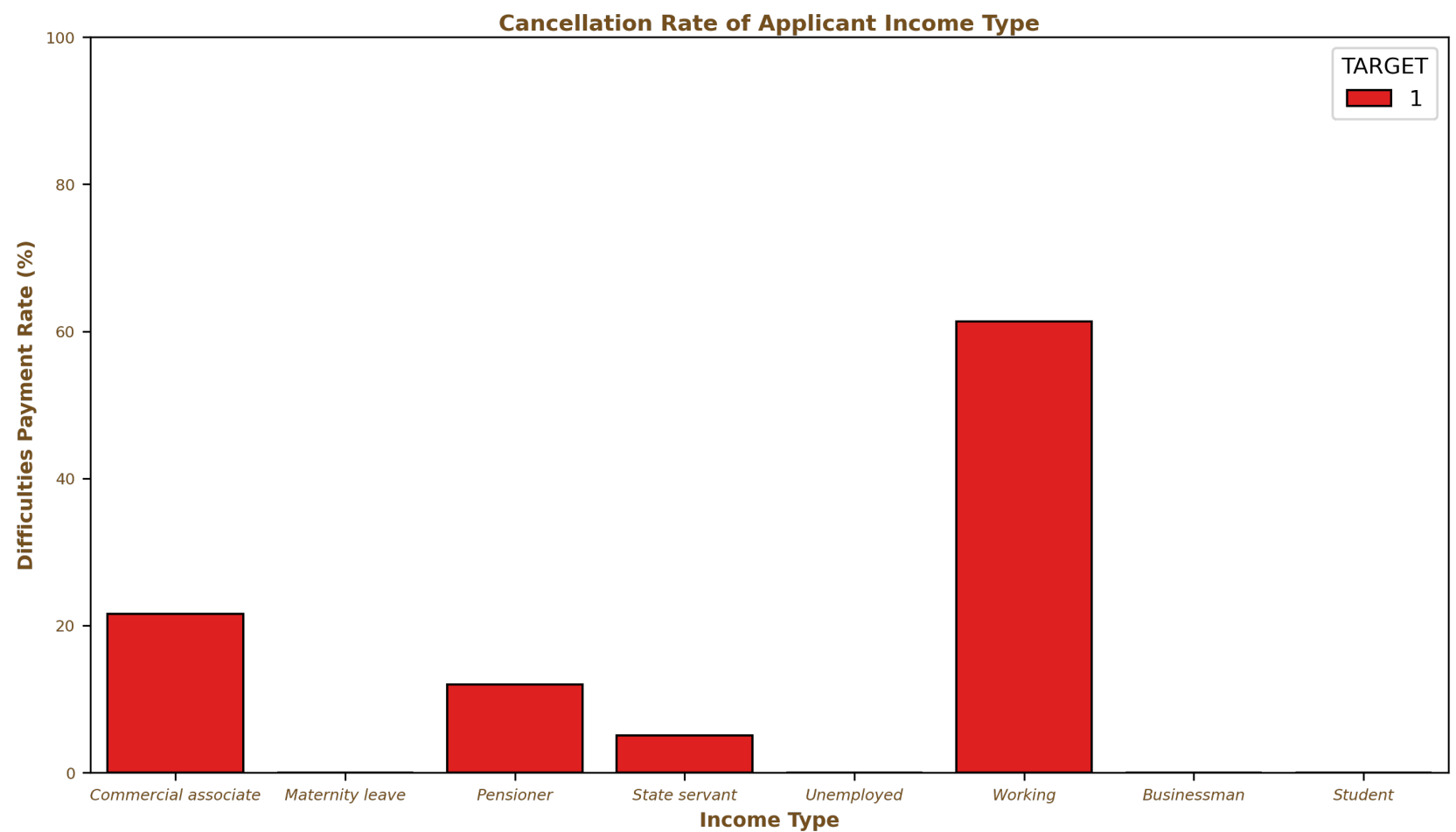We choose **Logistics Regression** beside of Decision Tree


AUC = **0.703**

KS = **0.307**

Generally with >0.7 AUC and >0.3 KS in risk modelling, it is considered as good model

Repository Link

BUSINESS INSIGHT

Cancellation Rate of Applicant Income Type

Working income type is the highest difficulties payment for loan credit. With the lowest or near zero is Businessman and Student. It can be happen because **Working Income type has potential to lay off from job.** The recommendation for marketing team is focus to businessman and student for loan credit, and make several screening for Working income type like how long to be employed.

BUSINESS INSIGHT

Cancellation Rate of Age
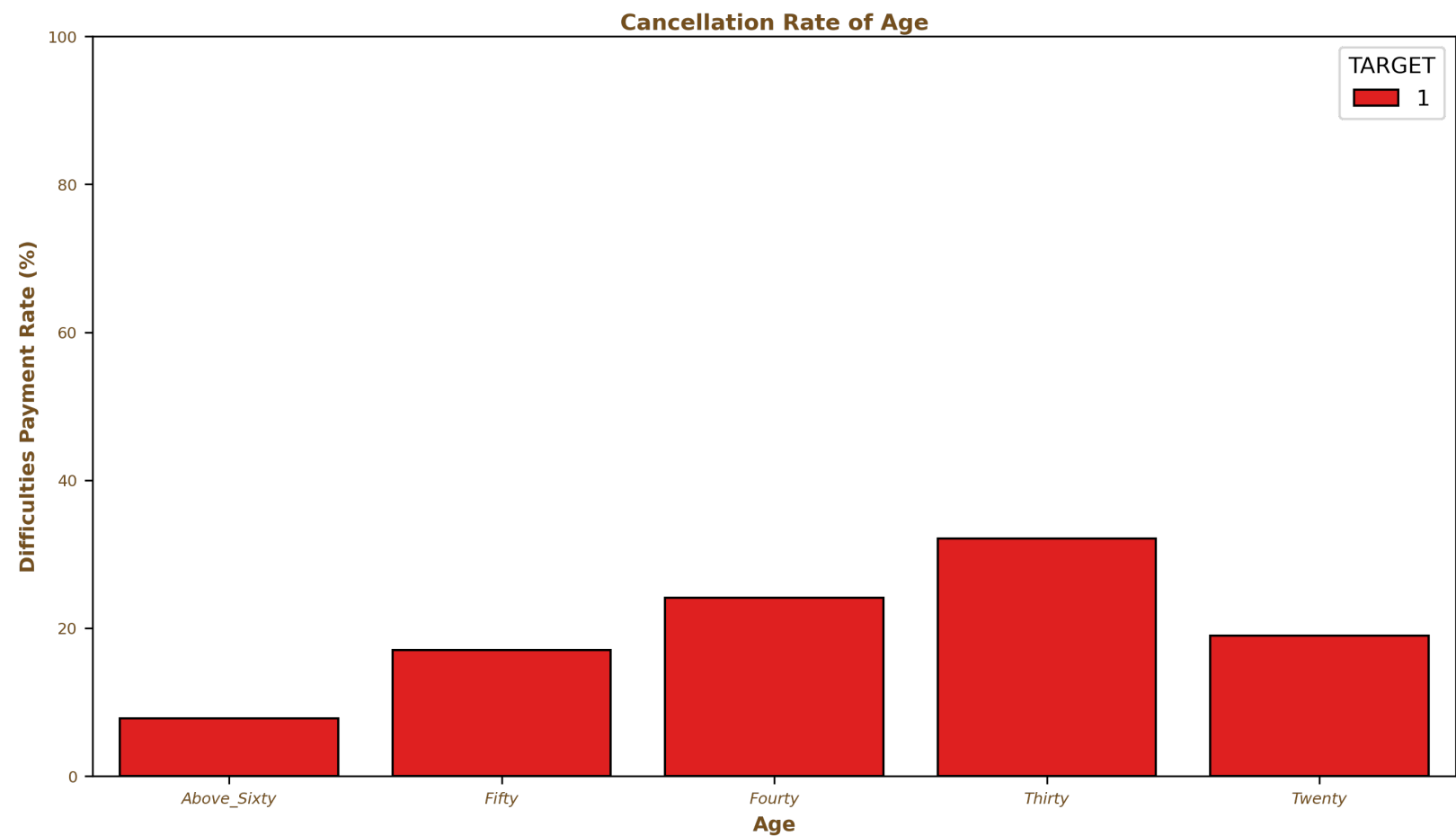
Payment difficulties does not differ significantly between the age group, however as we can see **Thirty age groups (21-30 years) is the highest difficulties payment**. The recommendation is to **concentrate on marketing fifty and above sixty** because they have stable finances.

THANK YOU!

CONTACT

https://github.com/rahmars18

linkedin.com/in/rahmars/