HOME CREDIT SCORECARD MODEL

FINAL PROJECT - VIX RAKAMIN ACADEMY X HOME CREDIT

ISAAC DWADATTUSYAH HAIKAL AZZIZ

1

PROBLEM

THE PROBLEM WHAT WE WANT TO SOLVE
SOLUTION FOR THE PROBLEM

PROBLEM

Problem

Home credit is a company that runs in the financing sector that provides financing services for customers who shop online or offline. Home Credit also offers money lending services where incoming customers can borrow and pay instalments regularly. The challenge of the home credit company itself is to determine whether the client can be given a loan or not. because providing loans to troubled clients can cause losses for the company.

Solution

We can model and draw characteristic insights from past client data that can predict the client's condition in future loans.

Objective

Draw insight and Identify client characteristic from each class (have problem with paying loans and not have any problem with loans). ALso create good model from provided train data

2

DATASET & STEP

THE DATASET THAT WE USE
STEP FOR SOLVING THE PROBLEM

DATASET & STEP

1 IMPORT DATASET

The dataset that i used was application_train as train and validation dataset and Test application_test as test dataset.

2 BASIC DATA EXPLORATION & VISUALIZATION

Draw insight & visualization from train_data, Make basic recommendation based on train_data.

3 DATA CLEANING

Drop Duplicate, Missing data treatment, Cleaning impossible value & outlier, Drop unneeded feature

4 MODEL BUILDING

Label Encoding, Build Logistic Regression, Random Forest, Decision Tree, and KNN Classification Model. Also choose the best model and feature importance

5 PREDICT TEST DATA

Predicting Test data target from best model from training.

3

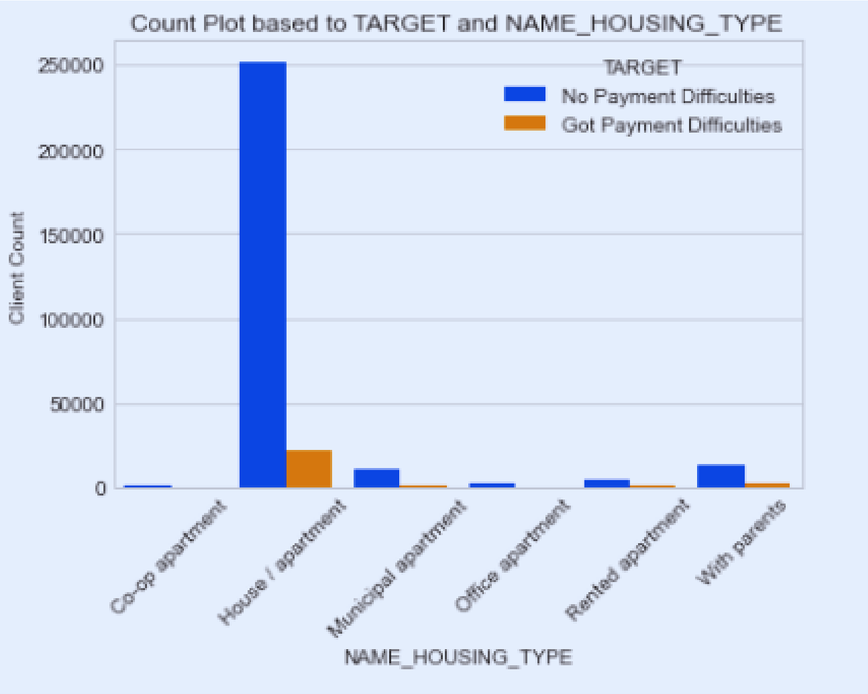
INSIGHT & BEST MODEL

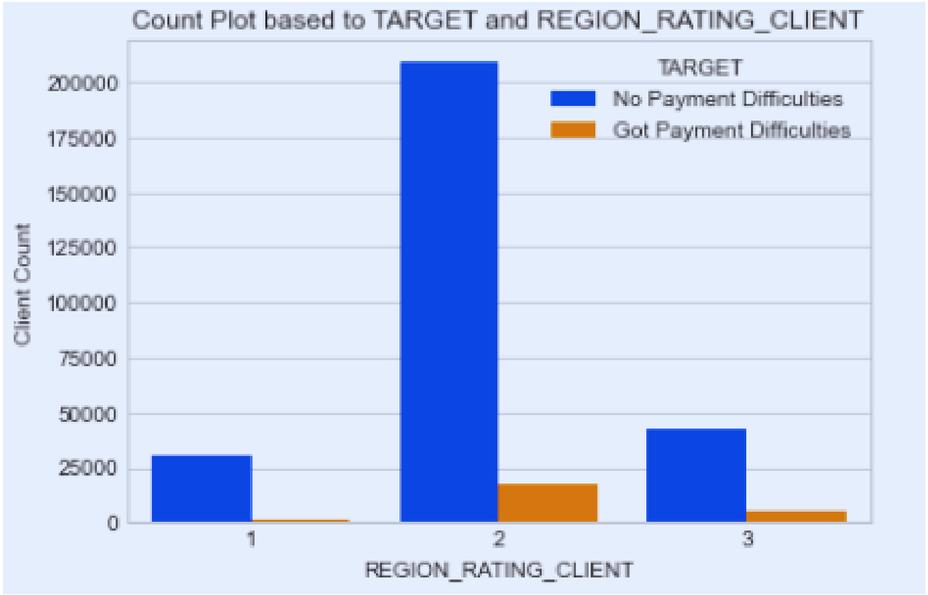
INSIGHT FROM DATA

BEST MODEL CHOOSED

INSIGHT

From the plot and relative proportion, Clients who still live in "Rented Apartment" & "With Parent" have a higher chance of difficulties paying back loans. Therefore, avoid lending funds to these groups.





From the plot and relative proportion, There's a clear-cut relation between "Area Rating" and "Proportion of Clients" who have difficulties paying loans.

Clients who live in Higher rating (3) areas have a higher chance of difficulty paying loans. Avoid or reduce the lending funds to these groups.

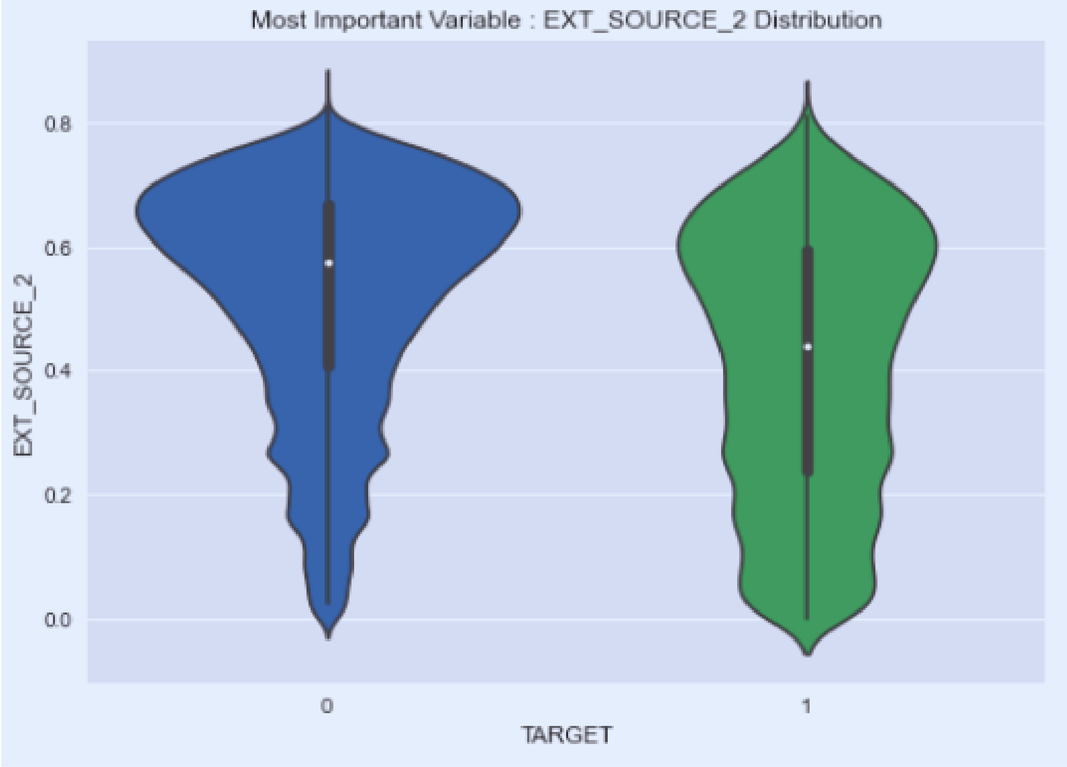
BEST MODEL

I tried 4 machine learning classification model combined with 3 resample method. Here the best score from each model

Model	Resampling Method	F1-Score Train	F1-Score Test
Logistic Regression	Undersampling	0.5709	0.5613
Decision Tree	Oversampling	1	0.9531
Random Forest	Oversampling	1	0.9977
K Nearest Neighbour	Oversampling	0.9138	0.88

For the best model, I chose random forest with the oversampling method based on F1-Score in both datasets. 3 of the most important features in this dataset were "EXT_SOURCE_2", "EXT_SOURCE_3", and "DAYS_BIRTH". This model also will be used to predict the test dataset

Most Important Features Plot



From the above plot, Client with Target = 0 or doesn't have any problem paying credit mostly has a high value of "EXT_SOURCE_2". This also applies to "EXT_SOURCE_3", which has a similar graph.

RECOMMENDATION

These are some business recommendation based on insight and model from both training dan test data:

1. From the visualization, avoid giving loans to clients who still live in "Rented Apartment" & "With Parent" and Clients who live in rating 3 areas and areas_city. Clients with this characteristic tend to have loans problem.
2. From the model's feature importance, prioritize clients with a high value on external sources 2 and 3. Clients with this characteristic tend to don't have a problem in pay loans.
3. From the ratio of clients with problems. Make the campaign appealing to students and businessmen. These two income categories have a high chance of not having problems paying loans (currently 100% paid) but have low counts of clients.
4. From Clients in test data, 100% of test data clients predicted to be don't have any problem in paying back loans (hopefully); based on the model, it's recommended to apply these client's loans proposal.

Github Repo:

THANK YOU

https://github.com/Isaacdha

https://www.linkedin.com/in/isaacdha/