Credit Risk Modeling

Helmy Naufal Aziz
Email | Linkedin | Github

Virtual Internship Experience - Home Credit Indone

OVERVIEW

Home Credit is currently using various statistical methods and Machine Learning to make credit score predictions. It can ensure that clients capable of repayment are not rejected and that loans are given with a principal, maturity, and repayment calendar that will empower their clients to be successful.

There are 7 datasets available in this project containing applicants data, credit card history and previous credit in Credit Bureau and Home Credit.

Goal and Objectives

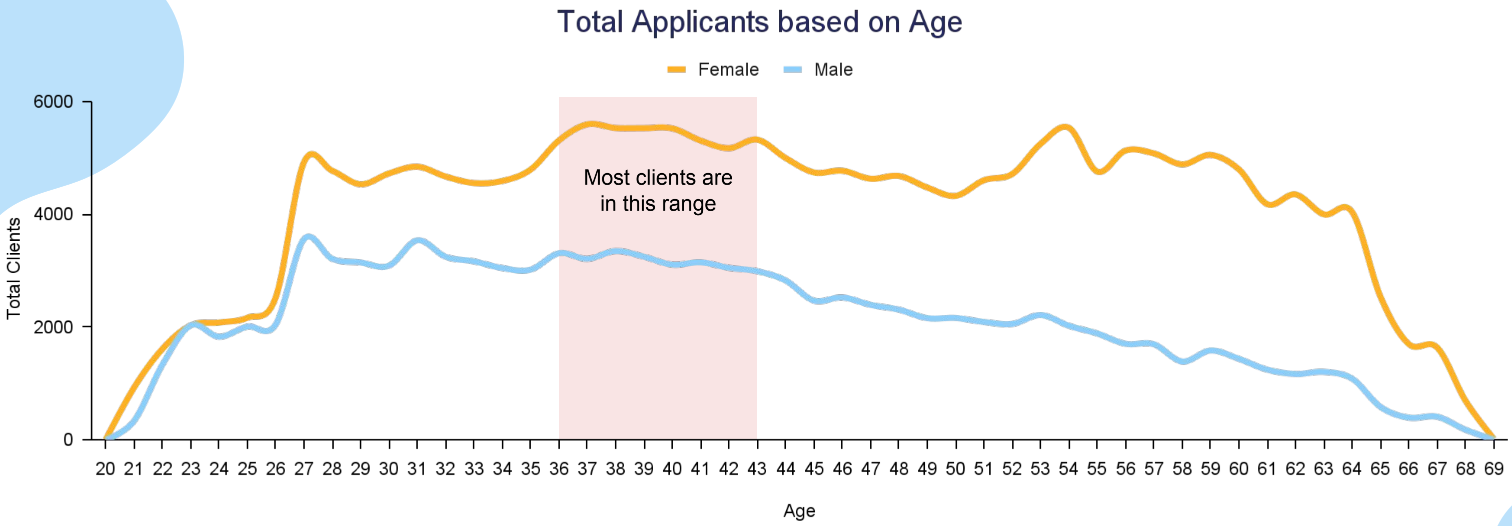
Goal

- Predict client's loan repayment capabilities

Objective

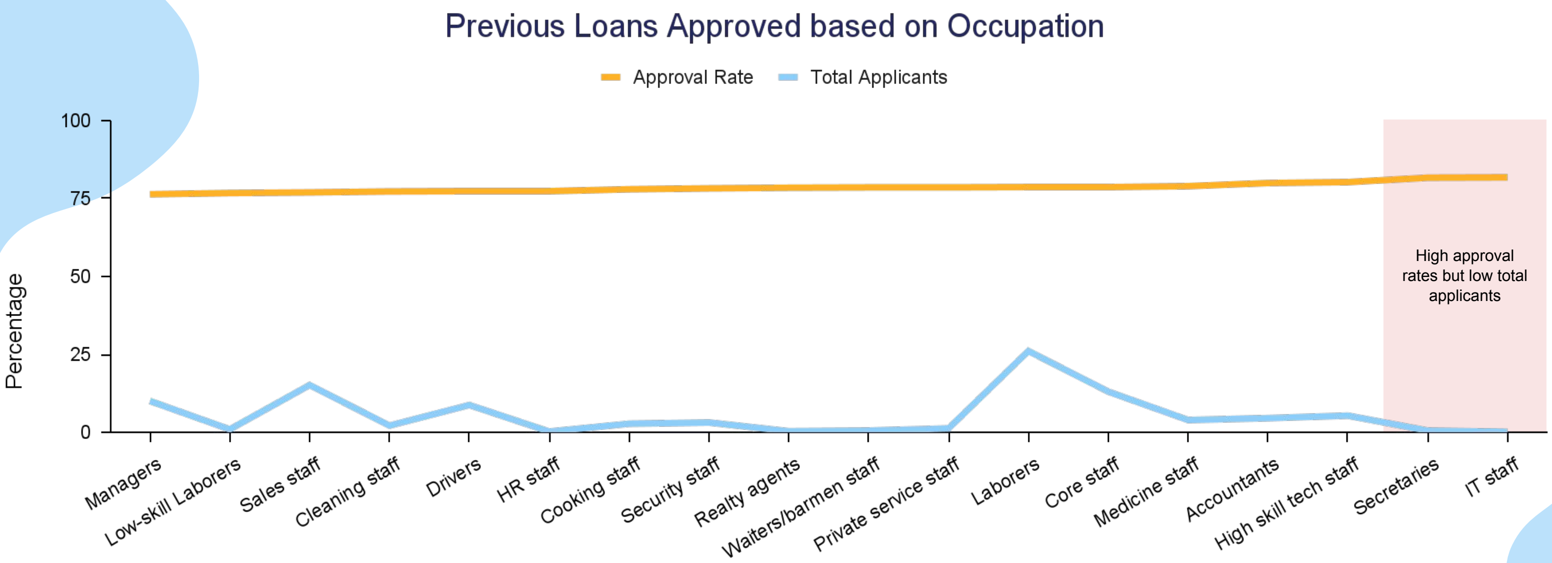
- Develop a machine learning model to predict potential default clients
- Create credit scorecard for loan approval

DATA EXPLORATION



Most applicants are 36 to 43 years old, with 22.3% of the total clients, and females dominate the distribution. We can create marketing strategies to engage more applicants in this age range.

DATA EXPLORATION



Applicants who work as Secretaries and IT Staff have the highest approval rates. However, this category only has a few applicants. We can create marketing strategies to engage more applicants with these occupations.

Data Preprocessing

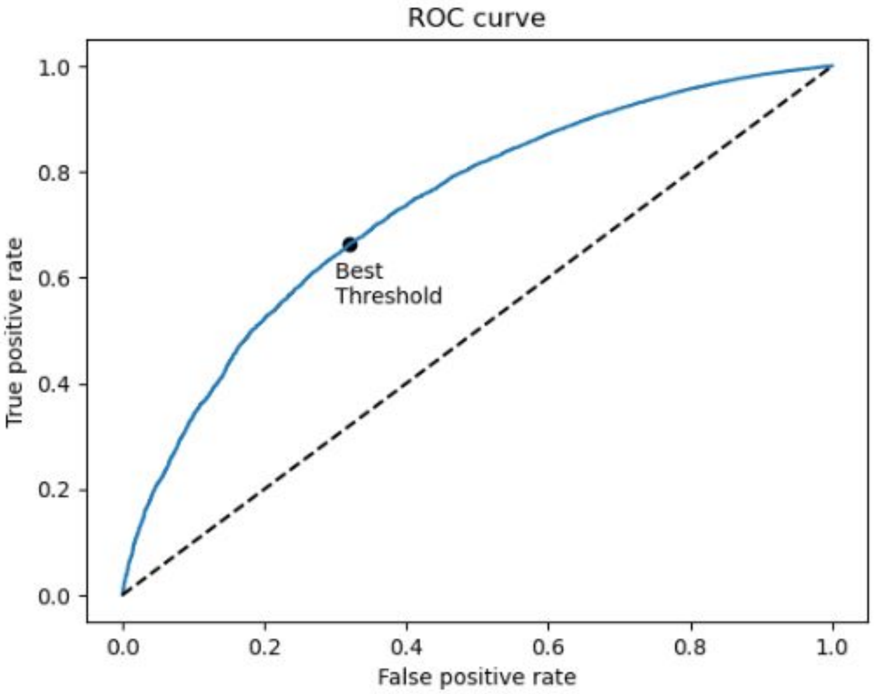
1
Data Cleaning
Clean data from any missing values

2
Split Data
Split data into train set and test set

3
Feature Selection
Identify the most suitable feature for modeling

4
Feature Engineering
Create new categorical feature based on Weight of Evidence

MODELING RESULT



Default Threshold: 0.5

Accuracy	Precision	Recall	AUROC
0.674048	0.958418	0.674697	0.732514

Best Threshold: 0.307662

Accuracy	Precision	Recall	AUROC
0.862836	0.938551	0.910397	0.732514

CREDIT SCORECARD

MIN: 300

MAX: 850

Accept Threshold: 475

Feature Name	Category	Score
INTERCEPT (BASE)		536
CODE_GENDER	M	-11
	F	14
NAME_EDUCATION_TYPE	Lower secondary	-41
	Secondary / secondary special	-30
	Incomplete higher	-13
	Academic degree	72
NAME_FAMILY_STATUS	Civil marriage	-5
	Single / not married	-2
	Separated	-5
	Married	10
	Widow	5
NAME_INCOME_TYPE	Work_Unemp_Stud_MaterLeave	-7
	Comm_assoc / Businessman	-1
	State servant	12
	Pensioner	-1

Feature Name	Category	Score
EXT_SOURCE_2	missing	-20
	<0.256	-56
	0.256-0.427	-23
	0.427-0.513	-8
	0.513-0.598	2
	0.598-0.684	14
	0.684-0.769	33
	>0.769	63
EXT_SOURCE_3	missing	-9
	<0.269	-64
	0.269-0.359	-28
	0.359-0.448	-13
	0.448-0.538	7
	0.538-0.627	24
	0.627-0.717	36
	>0.717	50

CREDIT SCORECARD

Feature Name	Category	Score
YEAR_BIRTH (AGE)	<30	-5
	30-40	-11
	40-50	1
	50-60	9
	>60	10
REGION_RATING_CLIENT	1	9
	2	3
	3	-8
YEAR_LAST_PHONE_CHANGE	<2	-9
	2-4	-2
	4-6	5
	>6	3
YEAR_ID_PUBLISH	<5	-6
	5-9	-3
	9-13	5
	>13	7

Feature Name	Category	Score
FLAG_DOCUMENT_3	0	12
	1	-8
REG_CITY_NOT_LIVE_CITY	0	9
	1	-6
YEAR_REGISTRATION	<8	-1
	8-17	0
	17-30	3
	30-40	-1
	>40	2
REGION_POPULATION_RELATIVE	<0.0292	0
	0.0292-0.0436	-3
	>0.0436	7
AMT_CREDIT	<40995	0
	40995-445500	30
	445500-846000	17
	>846000	31

CREDIT SCORECARD

IMPLEMENTATION

Feature Name	Category	Score
INTERCEPT (BASE)		536
CODE_GENDER	M	-11
NAME_EDUCATION_TYPE	Academic Degree	72
NAME_FAMILY_STATUS	Single	-2
NAME_INCOME_TYPE	Student	-7
EXT_SOURCE_2	0.6	14
EXT_SOURCE_3	0.5	7
YEAR_BIRTH	24	-5
REGION_RATING_CLIENT	1	9
YEAR_LAST_PHONE_CHANGE	1	-9
YEAR_ID_PUBLISH	2	-6
FLAG_DOCUMENT_3	1	-8
REG_CITY_NOT_LIVE_CITY	0	9
YEAR_REGISTRATION	2	-1
REGION_POPULATION_RELATIVE	0.03	-3
AMT_CREDIT	500000	17

Accept Threshold: 475

Total Score: 612

Total Score > Accept Threshold

ACCEPT