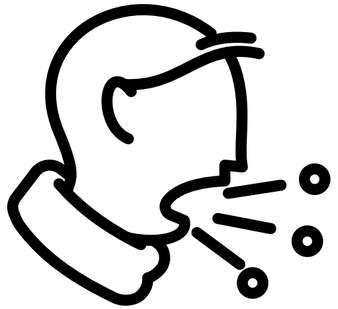
aihack
THAILAND

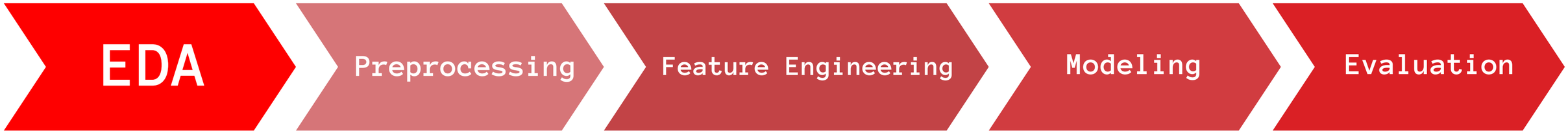


AI Crack Crack

Goal

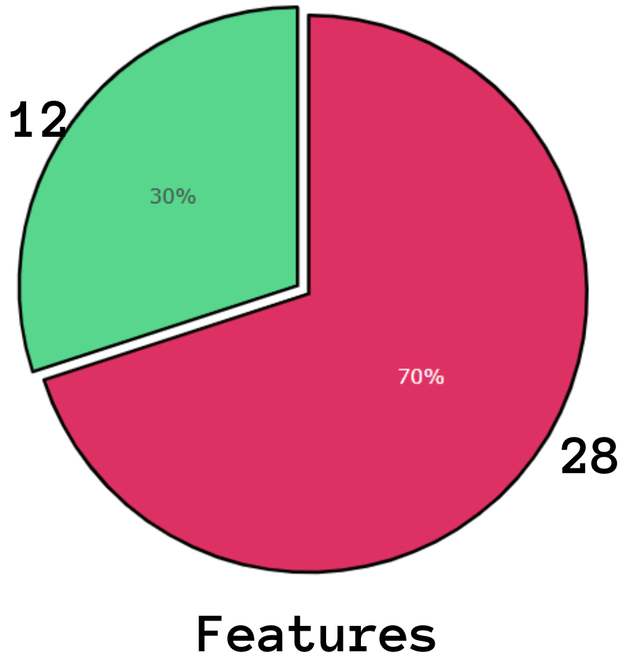
To create a forecast model of **Long Overdue Debtor** from AIFUL loan-applicants from their credit data with the **highest AUC**

Framework


EDA
Preprocessing
Feature Engineering
Modeling
Evaluation

Data Overview

Train (40 features + Target)
Public Data (40 features,



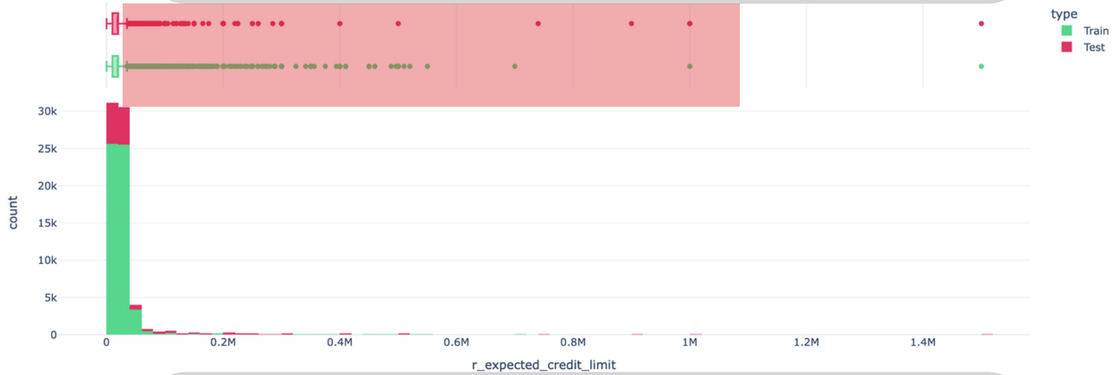
Target

EDA | Preprocessing | Feature Engineering | Modeling | Evaluation

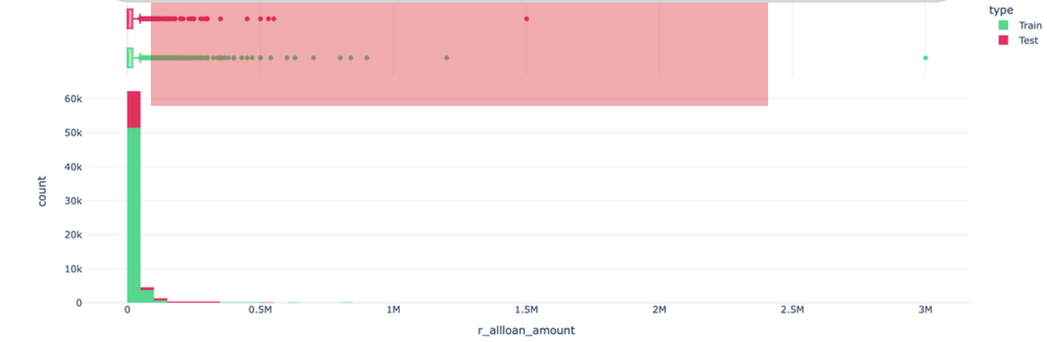
Outlier

monthly_salary

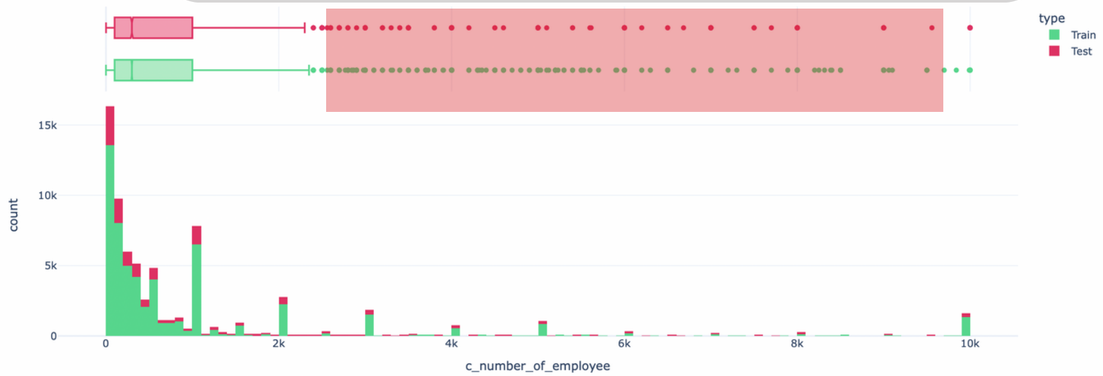
expected_credit_limit



allon_amount



number_of_employee

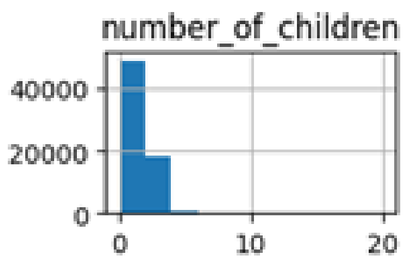


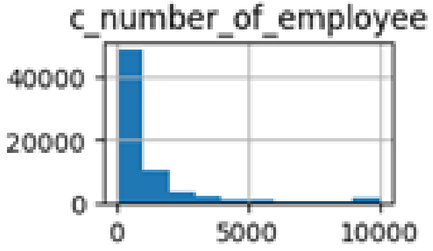
Solution

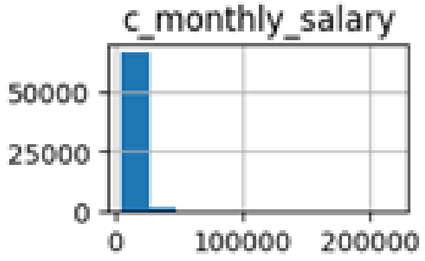
Set it to min-max range or using model that can handle

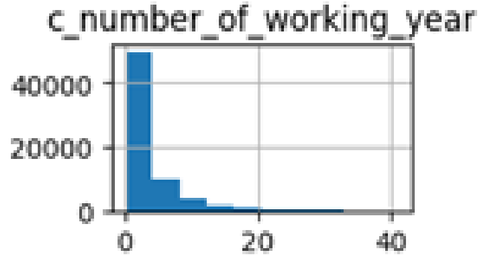
EDA → Preprocessing & Feature Engineering → Modeling → Evaluation

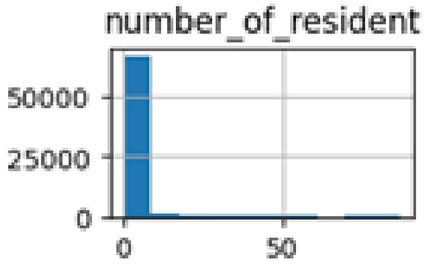
Skewness

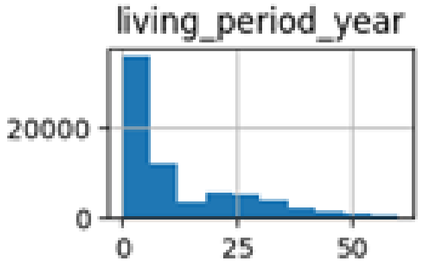


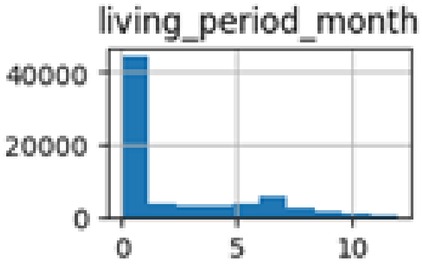












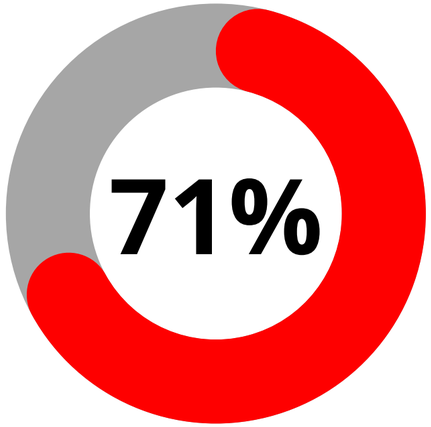
Solution

Determine range,change it to ordinal data and automatically run by acceptable model(bagging technique)

EDA → Preprocessing & Feature Engineering → Modeling → Evaluation

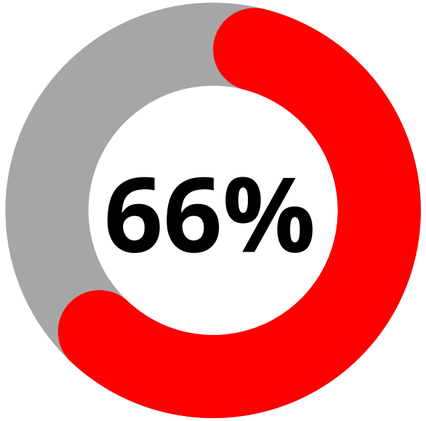
Train Data Insights

Default Customer Characteristics (default_12month = 1)



has POOR credit (FICO Score<580)

Top 3 Default Risk by Occupation are factory worker, labor and desk work.



is clerk (The lower position, the more default)

Preprocessing

aihack THAILAND

categorical data

c_business_type

1	agriculture
2	resource
3	architecture
4	manufacturing
5	merchant
6	import-export
7	business service
8	individual service
9	human transportation
10	communication
11	finance
12	insurance
13	non profit organization
14	education
15	hospital
16	government
17	unemployment

one-hot encoding →

Binary values(0&1)

cast to bool type to reduce memory space and make the process run faster

EDA | Preprocessing & Feature Engineering | Modeling | Evaluation

Drop feature

aihack THAILAND

trial and error method

-use general business sense and research

-then try to drop feature in multiple ways to see efficiency result

Unrelated

-no -app_shop_name

-postal_code -c_postal_code

-c_date_of_salary_payment

Repeat

-date_of_birth_week (with date_of_birth_year)

Too much missing data

-r_generalcode1 -r_generalcode2

Add new feature

- income_per_person = c_monthly_salary/number_of_resident
- total_income = c_monthly_salary+r_additional_income+r_spouse_income
- credit_utilization = r_expected_credit_limit/total_income
- income_to_creditlimit = c_monthly_salary/r_expected_credit_limit

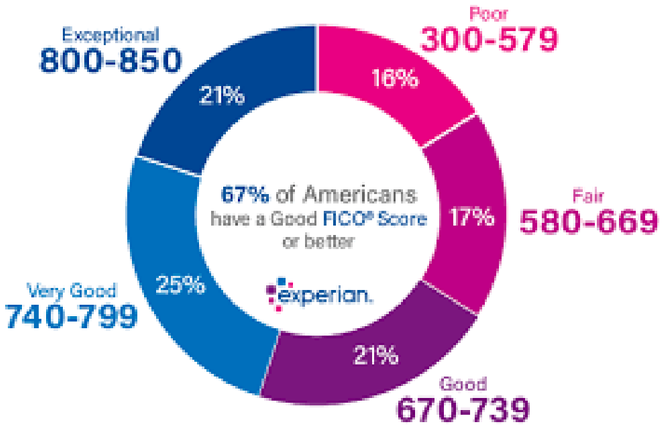
These new features refer to financial ratio indicator

Grouping

applicant_age_group by 0,25,45,65,100

fico_score_category → from the research fico score

No Credit Info(0)
Poor(<580)
Fair(580-669)
Good(670-739)
Very Good(740-799)
Exceptional(800+)



EDA → **Preprocessing & Feature Engineering** → Modeling → Evaluation

Modeling

Tree Based Model (Catboost, XGboost, LightGBM)

1. **Compare every model**
2. **Choose high perfomance model**

Evaluation

aihack THAILAND

Final Model

Catboost
train auc = 0.747029,
validtion auc = 0.660990

Evaluation

Tree Based Model (Catboost, XGboost, LightGBM)

Overfiting ->
cross validation, bagging,
ensemble

Business Application

1. Detect Bad Debtors (who default within 12 months) out faster - should be reject or reconsider by human.

2. Deep dive to important feature and set the new standard of screening criteria

3. In furthermore , we can improve customer application forms that include only potential factors. (To serve customer friendly that reduce time when customer apply)

Members and Roles

Tee Stat #4

- data preprocessing
- feature engineering & modeling

Dom Stat #4

- data analysis
- presentation making
- pitching

Get Eng-Com #2

- feature engineering & modeling
- model tuning

Bank Stat #2

- explore the definition of data and categorize data
- pitching

Gymmy Stat #2

- explore the definition of data and categorize data
- pitching

Thank You