Rakamin Academy

Virtual Internship Experience

# HOME CREDIT

"Empowering Online Lending Business with Advanced Machine Learning: Revolutionizing Customer Prediction and Campaign Strategies for Enhanced Promotion Efficiency and Accurate Credit Assessment in the Digital Age"

Amarindra Ardinova, Data Science, Batch 33

VIX 3

# TABLE OF CONTENT


## Problem Identification

Goal, Object & Metrics

## Data PreProcessing

Handling Missing Value & Outliers, Feature Engineering

## Visualization and Business Insight

Univariate, Bivariate, Multivariate

## Machine Learning & Implementation

Logistic Regression, Credit ScoreCard & K-Means

## Business Insight : Potential Business Prospects

Prediction & Clustering

## Business Recommendation

Summary


https://github.com/amarindraa/HCI-VIX-2024

# Problem Identification

Established in 2013, Home Credit Indonesia, a multinational tech-based financial company, seeks modernization. Leveraging existing algorithms for credit assessment and marketing campaigns is pivotal for sustained growth and operational efficiency.

| GOAL | OBJECTIVES | METRICS |
| --- | --- | --- |
| The overarching goal is to leverage machine learning techniques to optimize and enhance various aspects of an online lending business, specifically reducing False Negative on customer prediction and campaign strategies. | **Algorithm for Credit Assessment Accuracy**<br><br>**Algorithm for Marketing Strategy :**<br>Campaign Strategy Optimization / Enhanced Promotion Efficiency | **False Negative**<br>The lower the best<br><br>**Evaluation Recall**<br>Decreasing Opportunity Loss<br><br>**K_Cluster**<br>Clusterized customers for their credit behaviour |

# Data PreProcessing

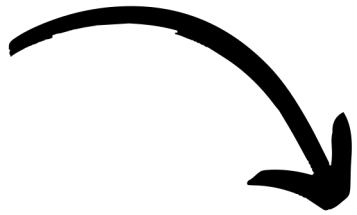
The dataset consists of a-targeted-dataset, a dataset without target, and a dataset that requires table joining.

Own Car Age fillna 0.

58 features with 40% missing values will be dropped.

The remaining will be fillna with median value.

## HANDLING MISSING VALUE



Negative values convert to positive values.

Create 20 new features and cleaning value < 0

## FEATURE ENGINEERING

apptrain_final.to_csv('data_unsupervised.csv', index=False)

**(48744, 81)**

apptest_final.to_csv('data_supervised.csv', index=False)

**(306972, 82)**

## SAVING AS NEWFILE

# Visualization and Business Insight

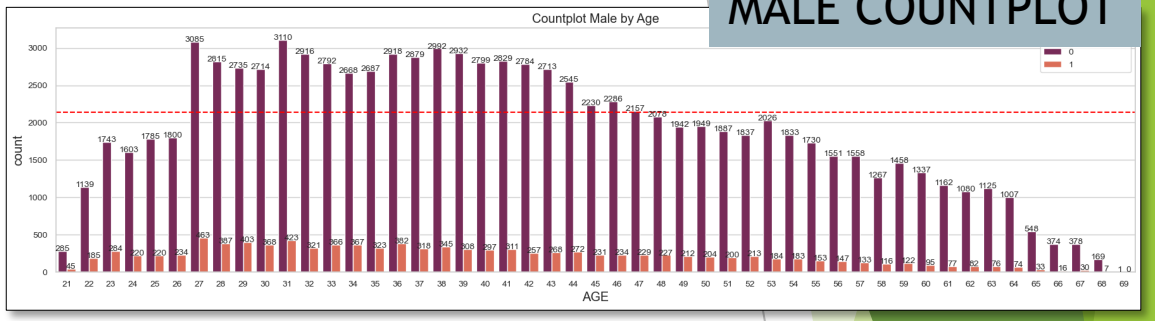
**Supervised_data**

Male customers exhibit an aggressive tendency in borrowing between the ages of 20 to 38, diminishing in the 39 onwards.

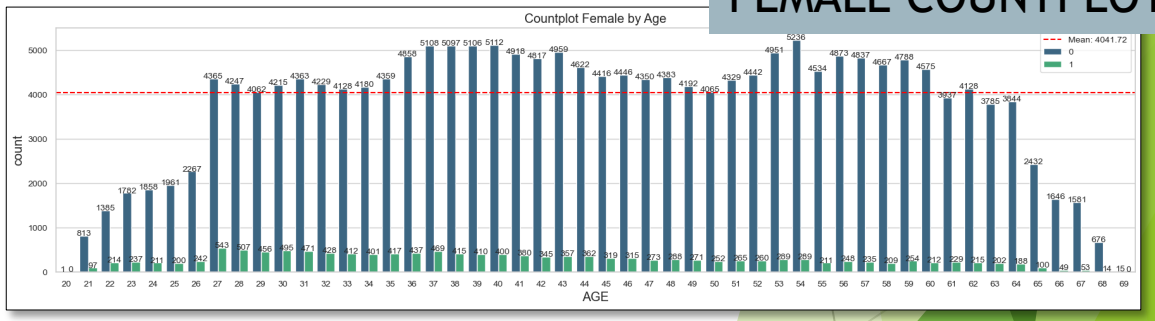
On the other hand, female customers display notable borrowing activity from 27 to 64.

These will be target to marketing campaign

MALE COUNTPLOT




FEMALE COUNTPLOT



# Machine Learning & Implementation

## Supervised Learning

Handling outlier
- IQR

Train Test Split

scaler
- standard

ScoreCard

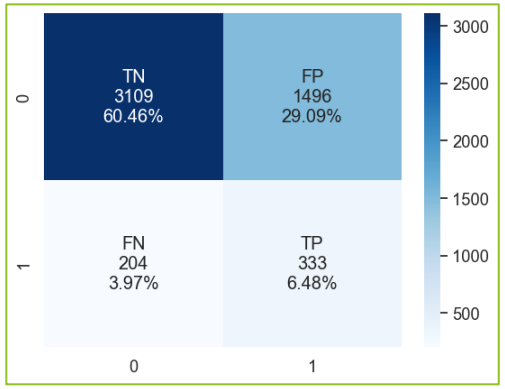
Features Selection
- Kendall

Handling Imbalance
- smote

Logistic Regression
- Accuracy, Recall, F1 Score

Save Model
- joblib.dump(logreg_model, 'logistic_regression_model.joblib')

Accuracy : 66.93
Recall : 62.01

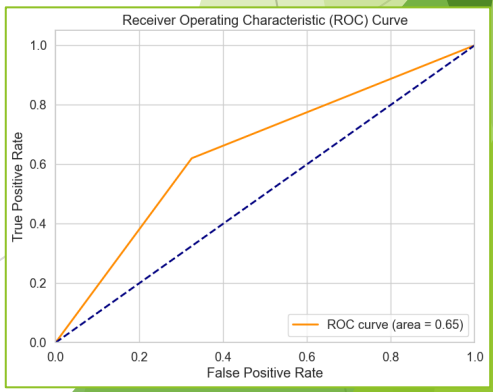



Scorecard

binning_process: BinningProcess estimator: LogisticRegression

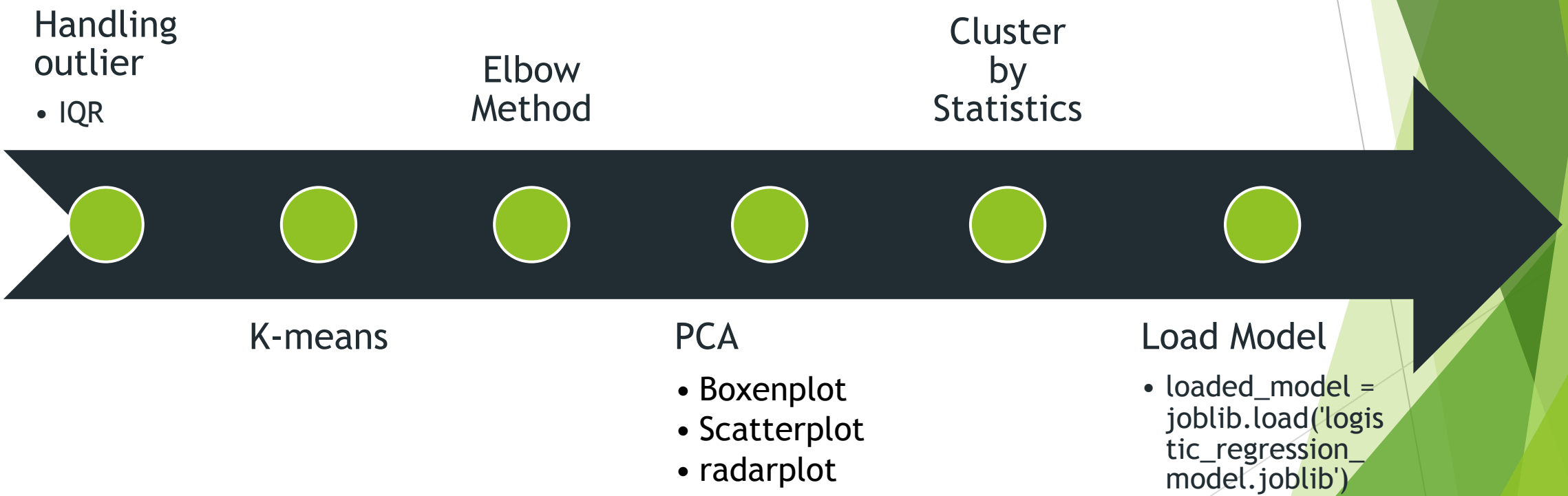
BinningProcess LogisticRegression

| | Variable | Bin | Points |
|---|---|---|---|
| 0 | AGE_CAT | (-inf, 0.50) | 30.42704773705491 |
| 1 | AGE_CAT | [0.50, 2.50) | 27.500841563657765 |
| 2 | AGE_CAT | [2.50, 3.50) | 22.91111337764535 |
| 3 | AGE_CAT | [3.50, inf) | 20.396038363530266 |
| 0 | ANNUITY_TO_CREDIT% | (-inf, 4.45) | 43.9271994669117 |
| 1 | ANNUITY_TO_CREDIT% | [4.45, 5.09) | 29.55932162170747 |
| 2 | ANNUITY_TO_CREDIT% | [5.09, 5.22) | 28.562825540265514 |
| 3 | ANNUITY_TO_CREDIT% | [5.22, 6.67) | 16.484557538919567 |
| 4 | ANNUITY_TO_CREDIT% | [6.67, inf) | 15.750584947752708 |

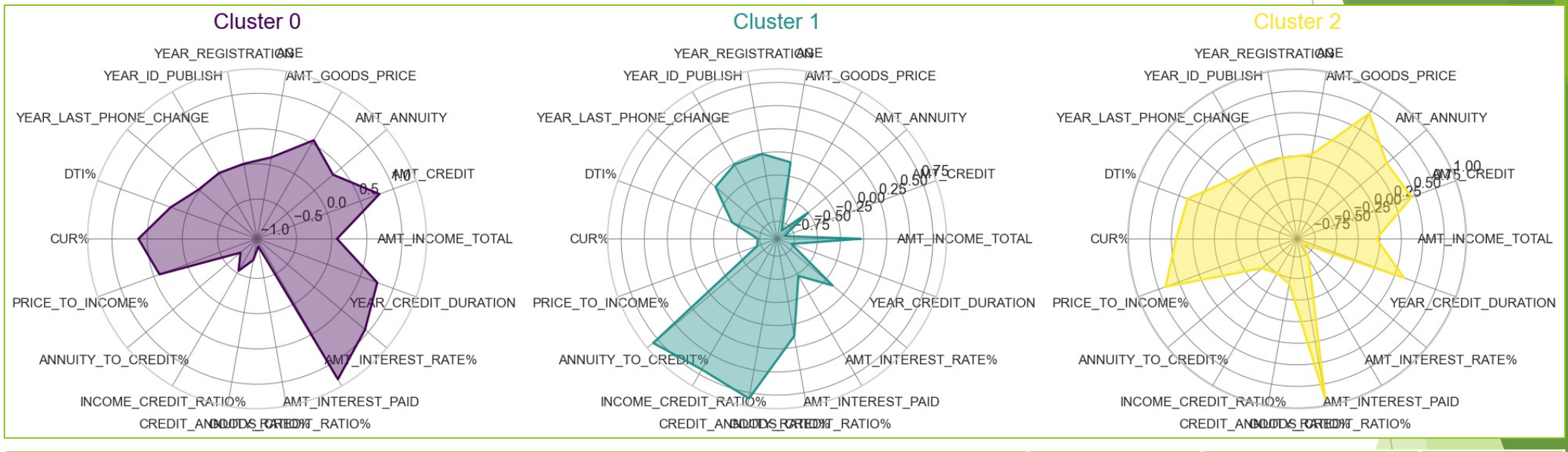


# Machine Learning & Implementation

## Unupervised Learning



# Machine Learning & Implementation




| Description | K_Cluster | Prediction | Count |
|---|---|---|---|
| Cluster 0 represents customers taking large and long-term loans with medium credit value. | 0 | approved(0) | 604 |
| | 0 | disapproved(1) | 10725 |
| Cluster 1 consists of customers typically making selective purchases, benefiting from low-interest rates, and preferring to settle debts promptly. | 1 | approved(0) | 684 |
| | 1 | disapproved(1) | 14899 |
| Cluster 2 comprises customers engaging in substantial-value shopping, benefiting from low-interest rates, and opting for a moderate repayment period. Consequently, they end up paying a significant amount of interest due to the extended loan duration. | 2 | approved(0) | 529 |
| | 2 | disapproved(1) | 9011 |

# Business Insight : Potential Business Prospects

The **THIRTIES** age range proves to be the most consumptive.

However, data analysis reveals that the highest approval rates are for those with **Higher Education**, while **Secondary/Secondary Special** experiences more rejections.

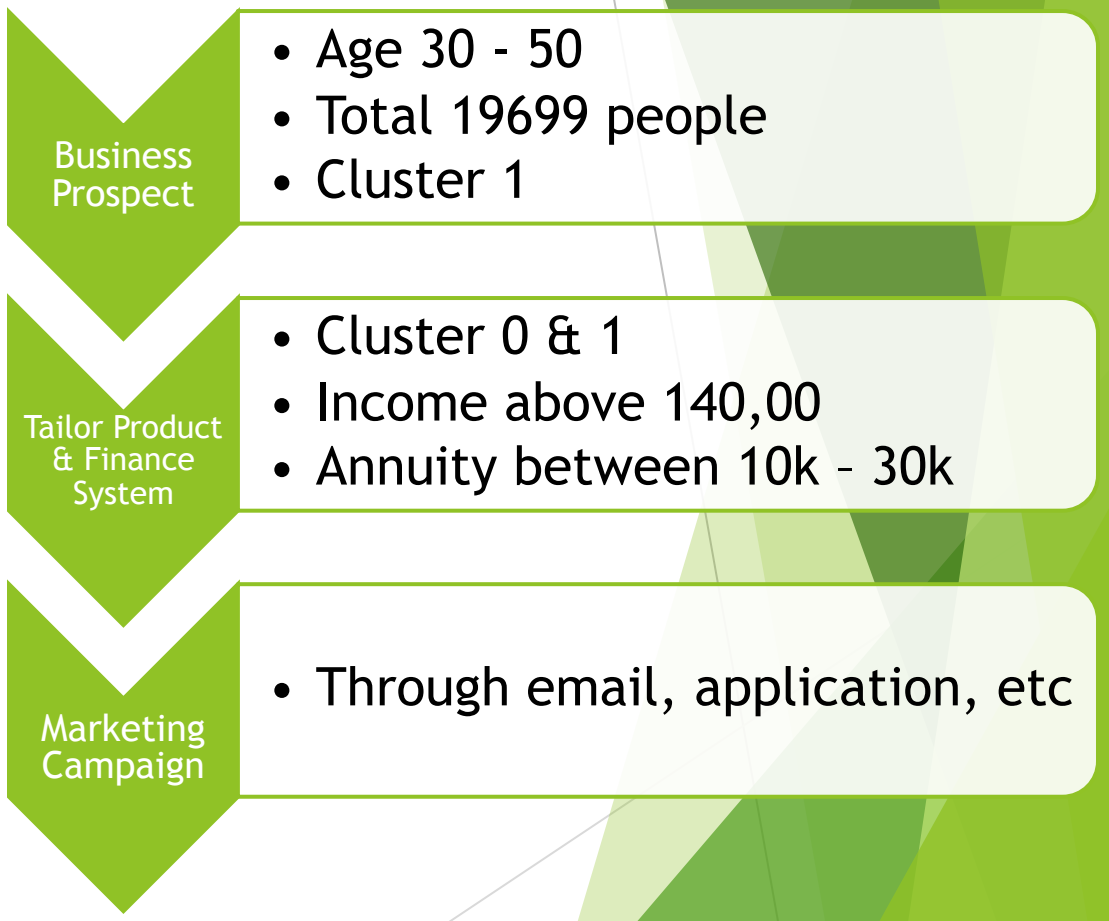
Therefore, marketing campaigns can be targeted at the THIRTIES group with Higher Education to stimulate transactions.

| | AGE_CAT | COUNT |
|---|---|---|
| 1 | TWENTIES | 5634 |
| 2 | THIRTIES | 9506 |
| 3 | FOURTIES | 7995 |
| 4 | FIFTIES | 6856 |
| 5 | SIXTIES & ABOVE | 4644 |

| NAME_EDUCATION_TYPE | Total | CI 0 | CI 1 | CI 2 |
|---|---|---|---|---|
| 1. Secondary / secondary special | 6549 | 1928 | 2957 | 1664 |
| 2. Higher education | 2475 | 641 | 1138 | 696 |
| 3. Incomplete higher | 366 | 81 | 179 | 106 |
| 4. Lower secondary | 109 | 38 | 43 | 28 |
| 5. Academic degree | 7 | 2 | 1 | 4 |

# Business Recommendation

In the age range of 30 to 50, which sees the highest loan activity, there is significant untapped potential in the **Secondary/Secondary Special** education category.

Approximately **19,699** customers, predominantly rejected as Cluster 1 *(customers making selective purchases with low-interest rates and a preference for prompt debt settlement)*, present an opportunity.

The business team can create a tailored product and financing system aligned with Cluster 0 & 1 characteristics and execute a targeted marketing campaign.