Data Project I

Lead Generation with Personalized Campaign Promotion

Tanapon Tanavejwasin
6610424023
MADT 6004

Agenda

Business Requirement

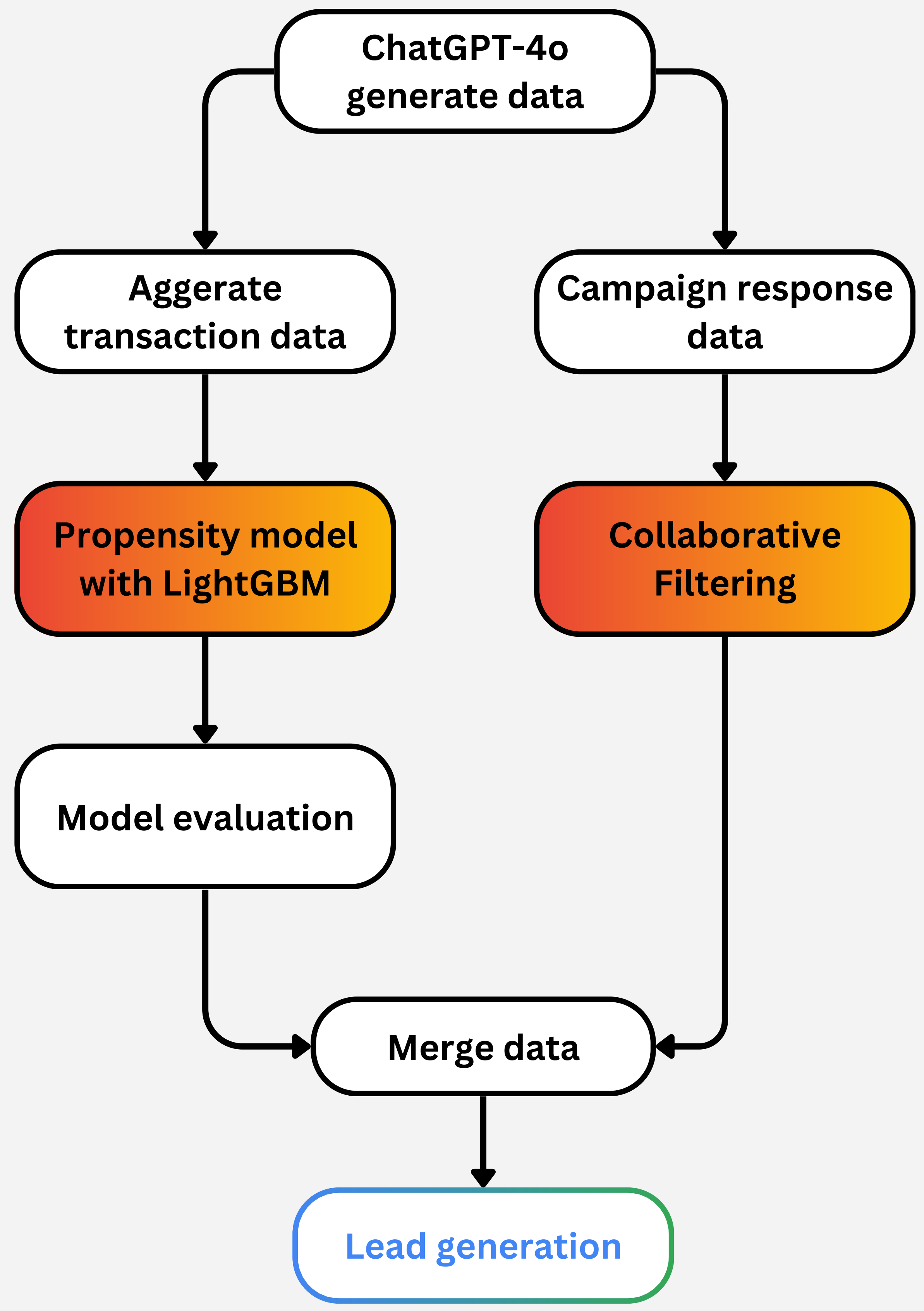
Business Requirement

- The Business Unit would like to send a campaign to increase credit card spending for customers who have a restaurant credit card.
- There is a limited budget for sending to all customers. Focus on the customers who are likely to respond to the campaign.

Project Architecture

Project Architecture

- **First step:** Generate two sources of data using ChatGPT-4o. One is campaign response data, used to perform collaborative filtering to suggest personalized campaigns. The other is an aggregate of credit card transaction data, used to perform a propensity model to predict customers who are likely to take the campaign next month.
- **Second step:** Perform modeling.
- **Third step:** Evaluate the model with the propensity model using the AUC score. model using the AUC score.
- **Fourth step:** Merge the results of both models.
- **Finally:** The leads for the marketing campaign are ready to be implemented in the business.



Collaborative Filtering

Data Sources

Campaign Response Data
for Content-Based Filtering

- **Simulated behavior** of credit card customers who use the Food Lover card type at restaurants they like.
- **1000 customers** x 50 campaign promotions.
- **Simulated population bias** (some promotions are more popular).
- **Simulated customer preferences** (some customers participate more).

ChatGPT-4o

	cust_id	pro_restrnt01	pro_restrnt02	pro_restrnt03	pro_restrnt04	pro_restrnt05
0	66000001	0.0	1.0	0.0	0.0	0.0
1	66000002	1.0	0.0	0.0	0.0	0.0
2	66000003	0.0	0.0	0.0	0.0	0.0
3	66000004	0.0	1.0	1.0	0.0	0.0
4	66000005	0.0	0.0	1.0	1.0	0.0

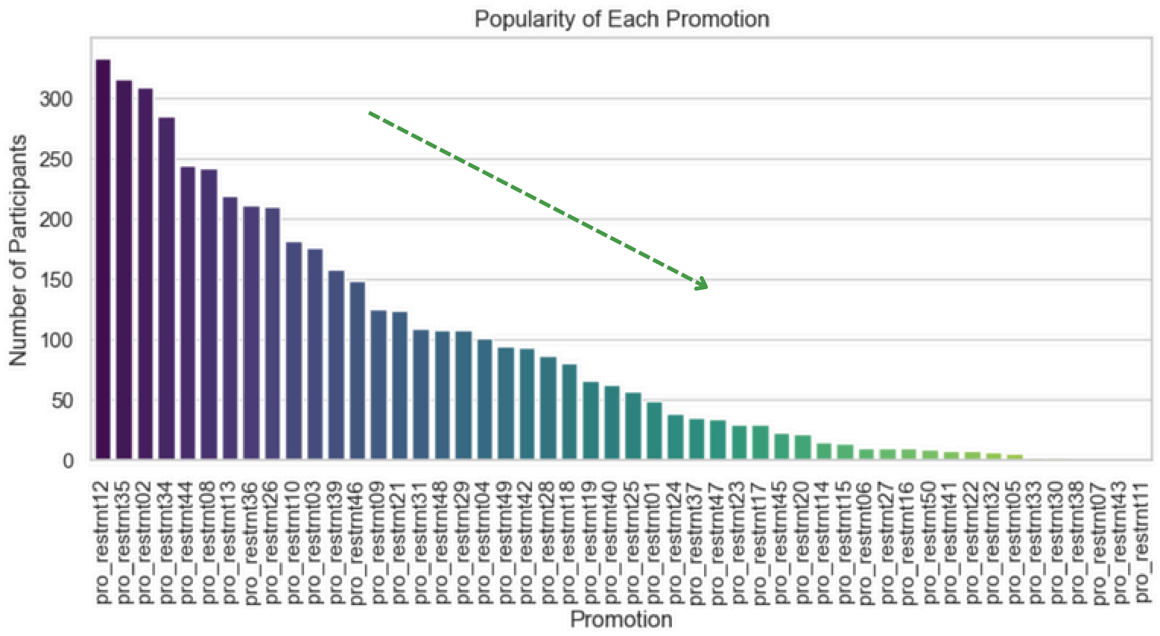
5 rows × 51 columns

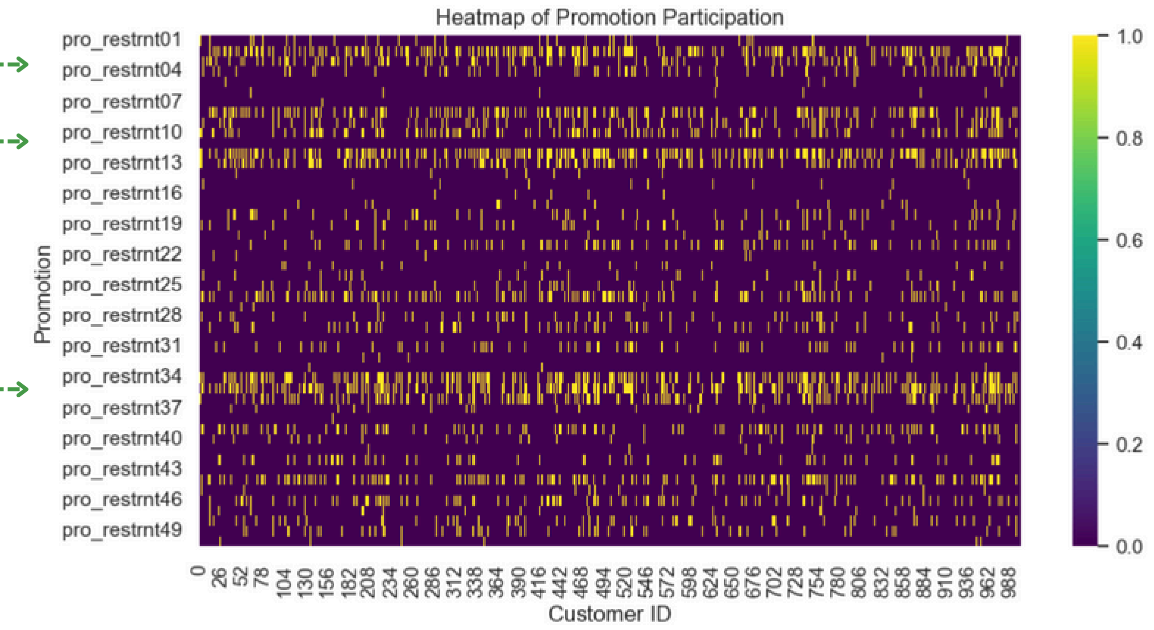
Example Data

Collaborative Filtering

Exploratory Data Analysis

Most customers never response campaign





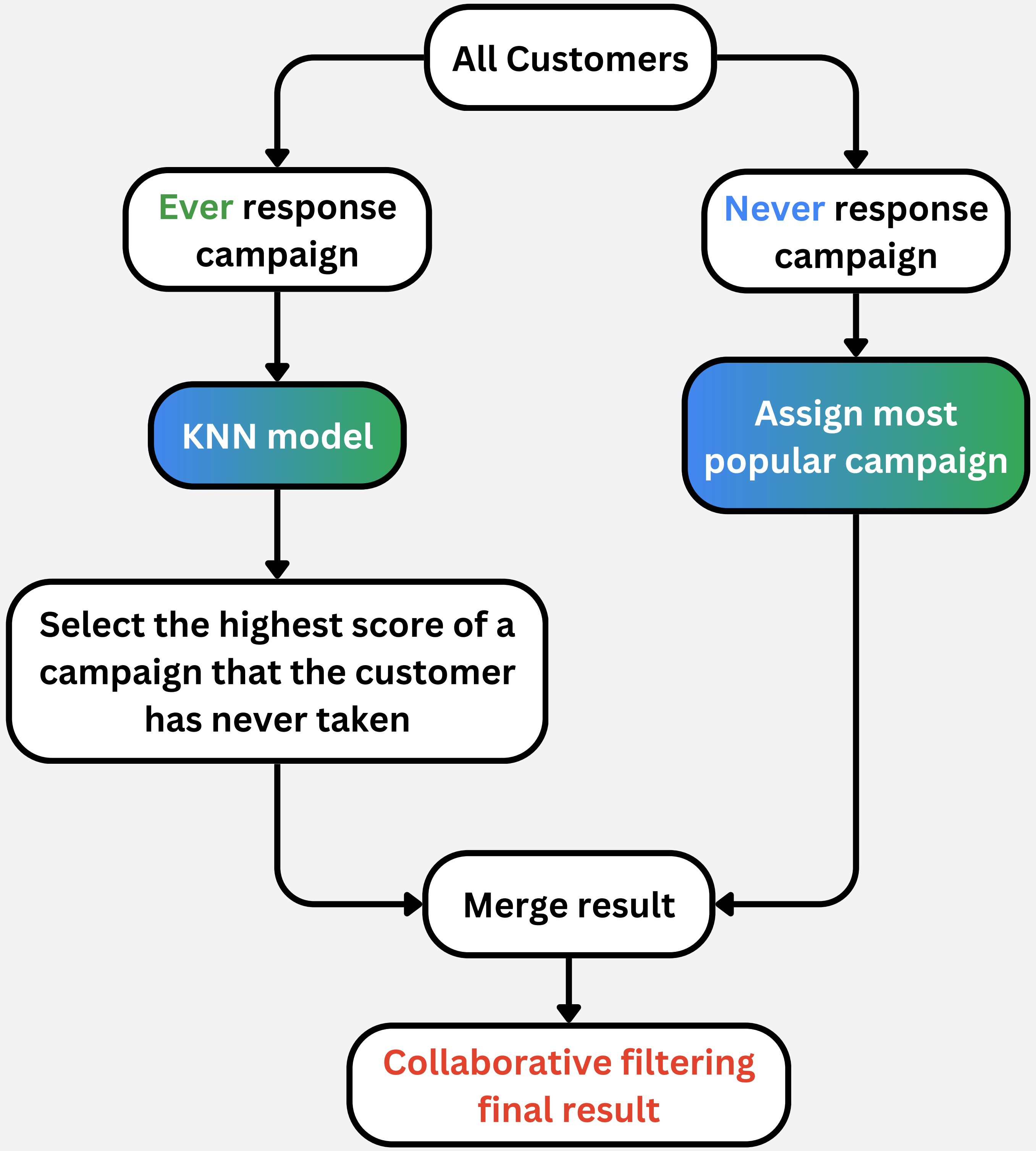
Some promotions are liked by customers, while others are not.

A few promotions were responded to by almost all customers.

Collaborative Filtering

Modeling

- **First step:** Split customers into two groups: those who have ever taken a campaign and those who have never taken a campaign. Feed the former group to the KNN model, and assign the most popular campaign to the latter group.
- **Second step:** Perform the KNN model on the group of customers who have ever taken a campaign, using the cosine metric and k = 5.
- **Finally:** Create a table for all customers containing customer ID, predicted campaign promotion for each customer, and recommendation score.



	Customer_ID	Promotion	Recommendation_Score
0	66000001	pro_restrnt12	0.8
380	66000002	pro_restrnt35	0.4
434	66000003	pro_restrnt02	0.8
512	66000004	pro_restrnt21	0.4
1	66000005	pro_restrnt12	0.6

Example Result

Collaborative Filtering

Campaign Dict

Campaign Number	Campaign Description
pro_restrnt01	Indulge in a delightful dinner at a top Chinese restaurant in Bangkok with 8 renowned hotel brands. Enjoy a 10% cash back and earn 100 points on every meal.
pro_restrnt02	Savor a luxurious lunch at an Italian restaurant in Central World. Get a free dessert with every main course and earn double points.
pro_restrnt03	Start your day with a Japanese breakfast at Siam Paragon. Get a 15% discount and earn 50 points on your bill.
pro_restrnt04	Experience a traditional Indian dinner at Terminal 21. Enjoy a 5% cash back and earn 150 points.
pro_restrnt05	Enjoy a Thai lunch at MBK Center with a free drink and a 10% discount on your meal.
...	...
pro_restrnt50	Savor Italian restaurant lunch at Central Ladprao with a free appetizer and earn 90 points.

Propensity Model

Data Sources

Aggerate Transaction Data
for Propensity Model

- **Simulated credit card customers** who use the Food Lover card type.
- **1000 customers x 24 features** (demographic and transaction data lagged by 2 months) x 1 target x 12-month period with time series data
- **Simulated trend for the target**: customers with high frequency, high utilization, good payment history, and spending in the dinner category are more likely to respond.
- **Simulated customer preferences** (some customer participate more)

ChatGPT-4o

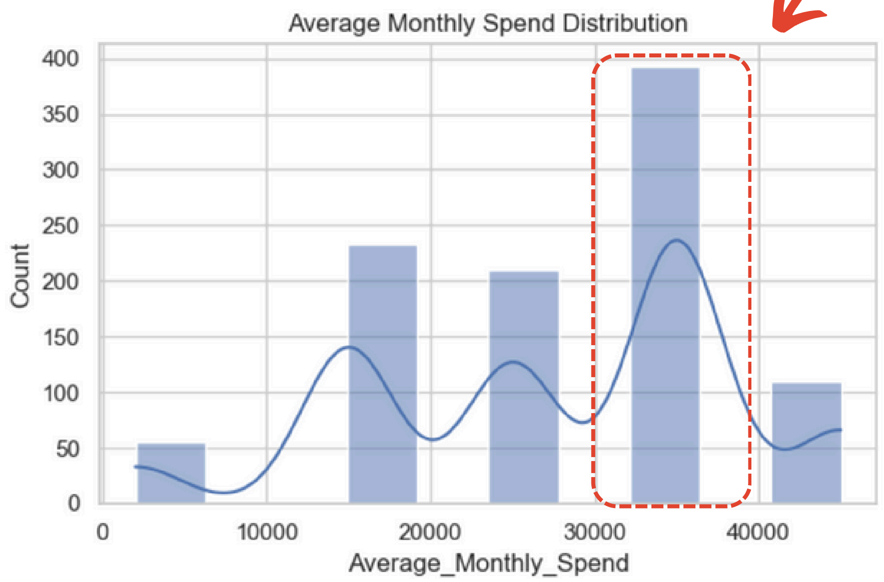
	Customer_ID	Feature_Date	Age	Gender	Income	Marital_Status	Number_of_Dependents	Days_as_Customer
0	66000001	2023-01-31	56	Female	15000	Married	1	2199
1	66000001	2023-02-28	56	Female	15000	Married	1	2199
2	66000001	2023-03-31	56	Female	15000	Married	1	2199
3	66000001	2023-04-30	56	Female	15000	Married	1	2199
4	66000001	2023-05-31	56	Female	15000	Married	1	2199

Avg_Days_Between_Transactions	Times_Credit_Limit_Exceeded	Previous_Month_Payment	Times_Payment_Missed	Target	Target_Date
1.250000	0	1201.088429	0	0.0	2023-03-31
2.142857	0	950.426671	0	0.0	2023-04-28
1.200000	0	4173.537581	0	0.0	2023-05-31
1.304348	0	24625.451156	0	0.0	2023-06-30
2.500000	0	11521.271854	0	0.0	2023-07-31

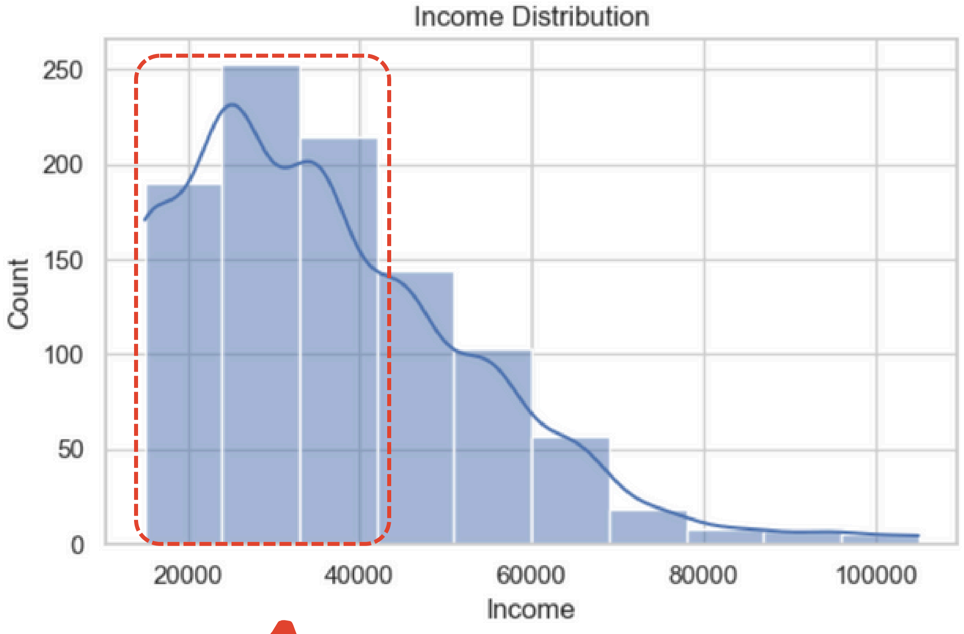
Example Data

Propensity Model

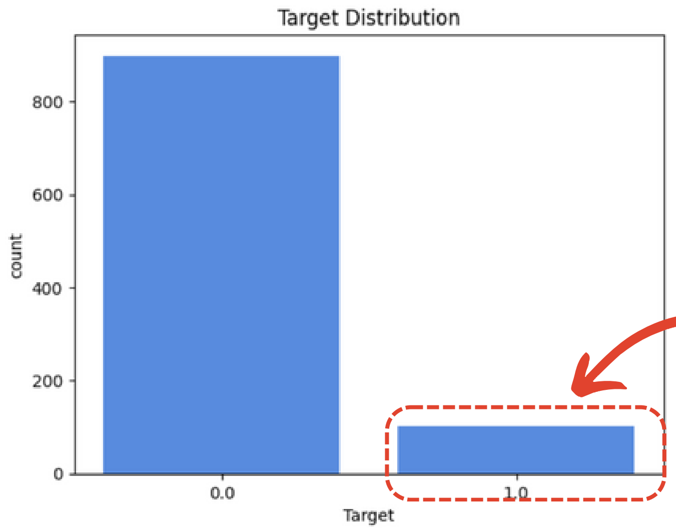
Exploratory Data Analysis



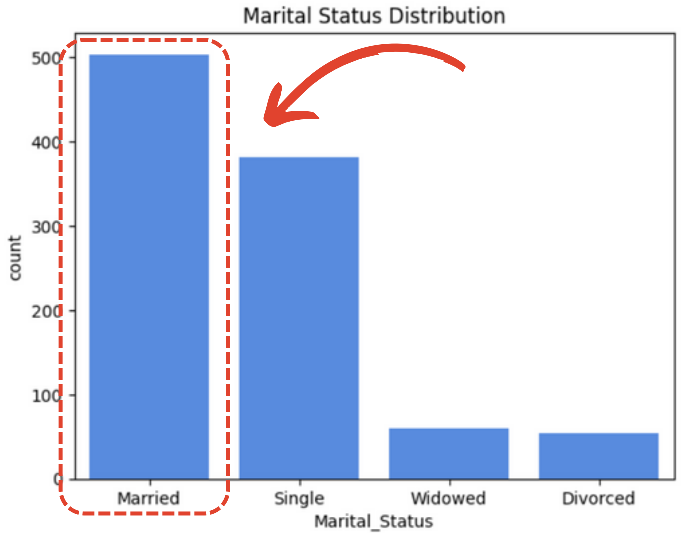
Almost 40% of the total customers spent around 33k



More than haft of the customers have an income around 20k-40k

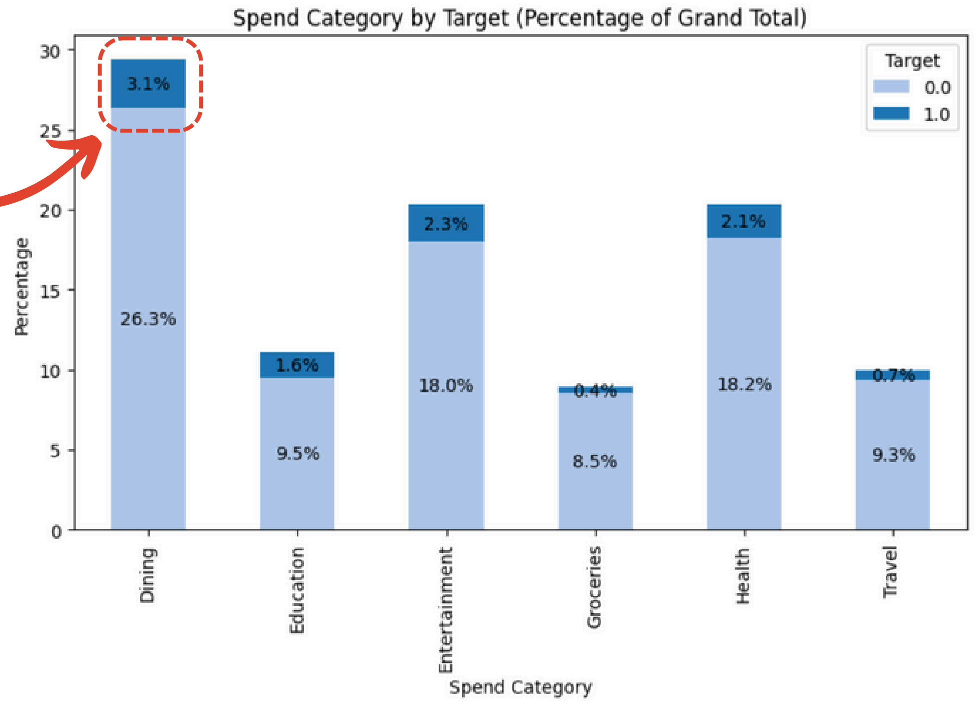


Target is imbalance data



Almost half of the portfolio customers are married.

Dining is slightly higher than other categories



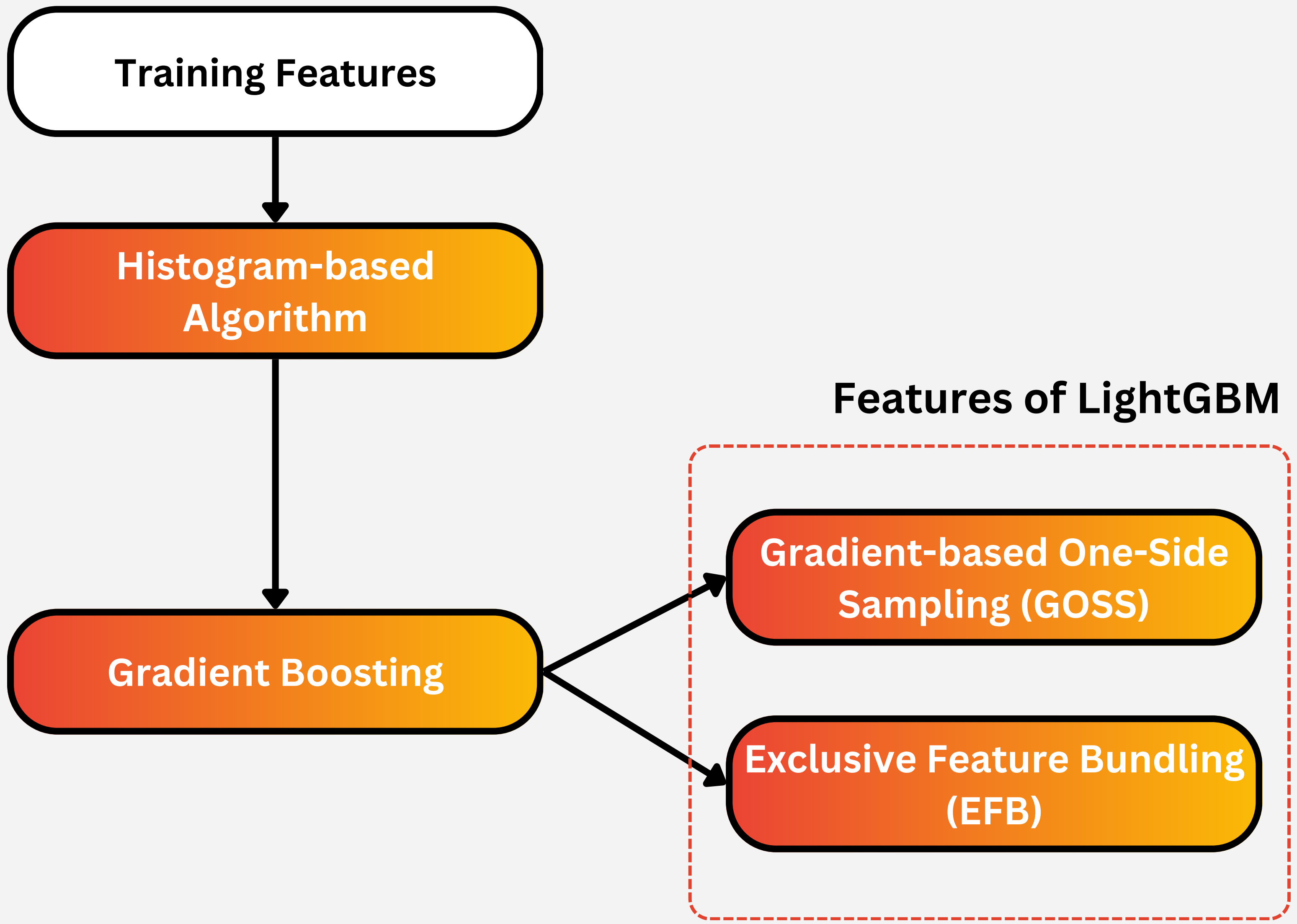
PD Model

Modeling

LightGBM Algorithm Explanation

LightGBM use tree-based learning algorithm

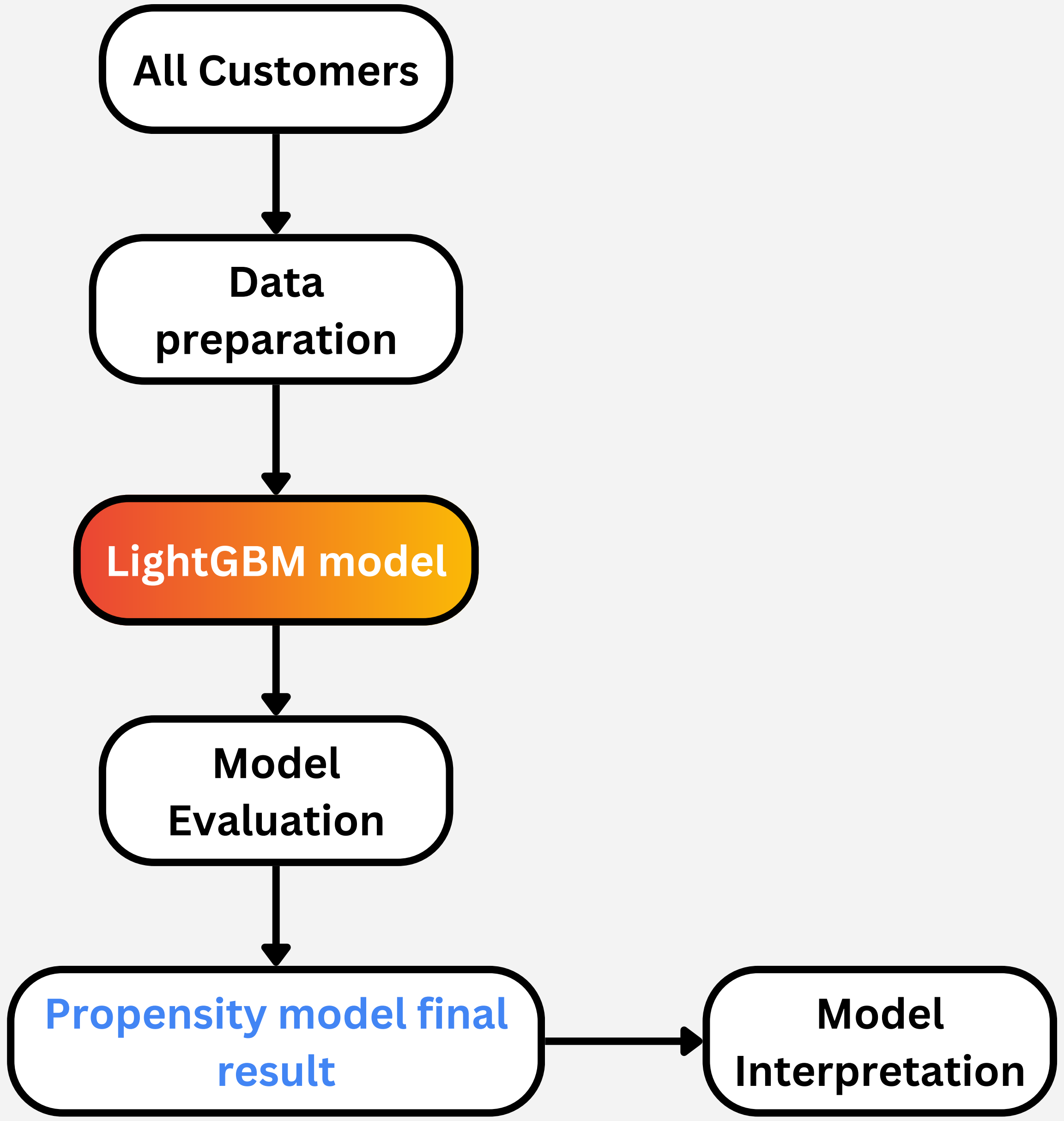
- **Gradient Boosting:** Involves training models sequentially, where each new model aims to reduce the residual errors of the combined ensemble of previous models.
- **Enhance two techniques:** Gradient-based One-Side Sampling (GOSS) to retains instances with large gradients while randomly sampling from instances with small gradients and Exclusive Feature Bundling (EFB) to bundle mutually exclusive features (features that rarely take non-zero values).
- **Histogram-based Algorithm:** use a histogram-based algorithm for finding the best split, which reduces the computational cost and speeds up training.



Propensity Model

Modeling

- **First step:** Prepare the data by checking for missing values and assigning data types before feeding it to the LightGBM model.
- **Second step:** Split the data into training, testing, and validation sets, with the validation set consisting of out-of-time 1 and out-of-time 2 periods.
- **Third step:** Train the model using LightGBM.
- **Fourth step:** Evaluate the model using the AUC score.
- **Fifth step:** Create a bin table, rearranging scores from bin 10 (highest score) to bin 1 (lowest score).
- **Finally:** Create a table for all customers containing customer ID, predicted score, actual result, and bin number.

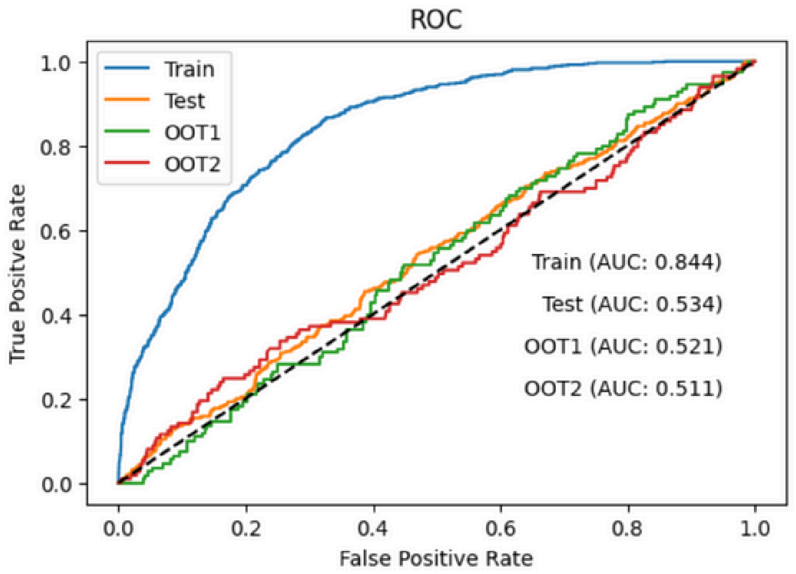


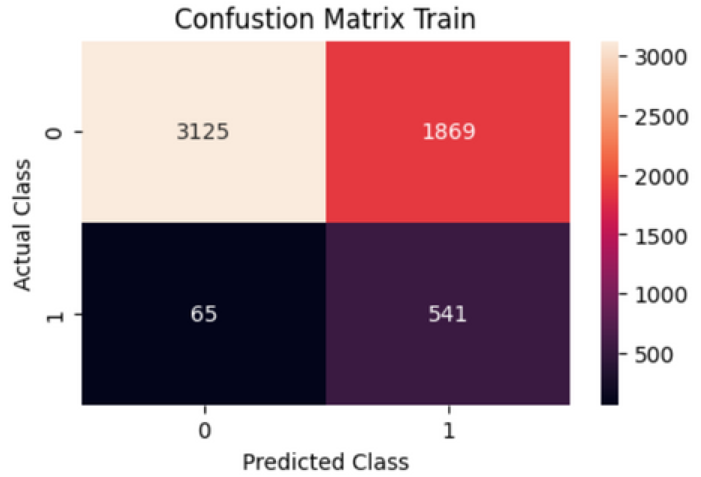
	Customer_ID	Predicted_Score	Actual	Bin
0	66000477	0.62	0	10
1	66000253	0.62	0	10
2	66000440	0.62	0	10
3	66000607	0.62	0	10
4	66000241	0.61	0	10

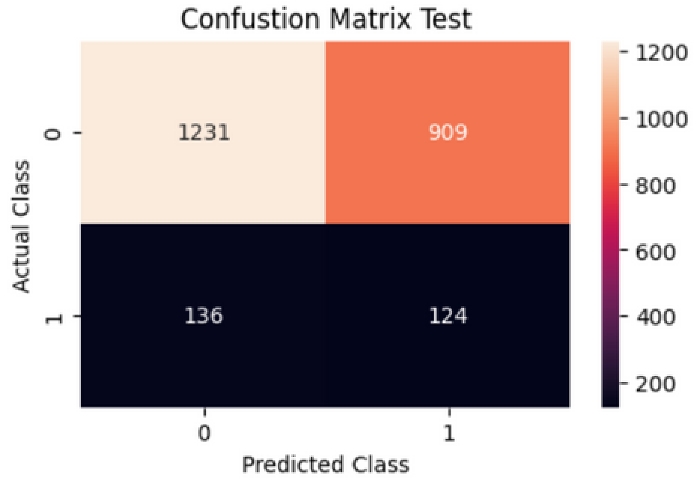
Example Result

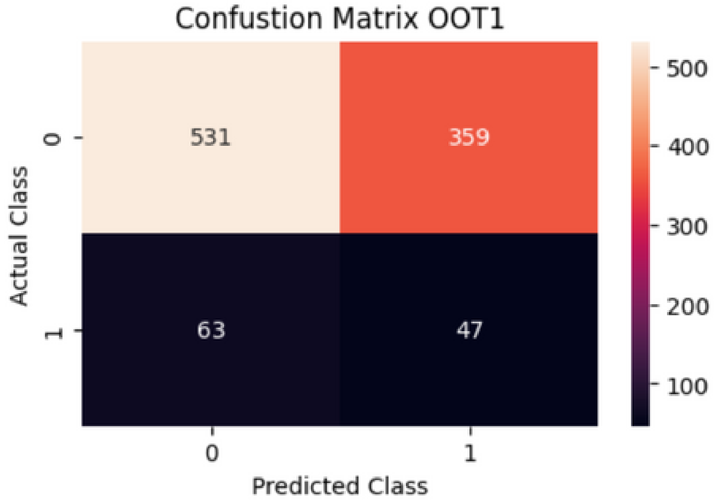
Propensity Model

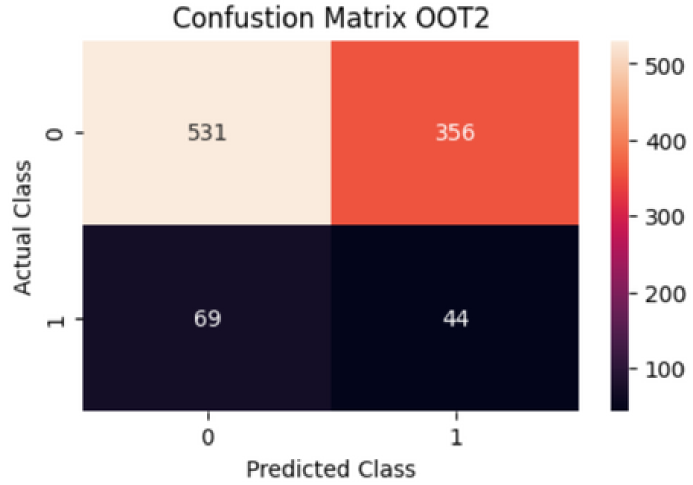
Interpretation









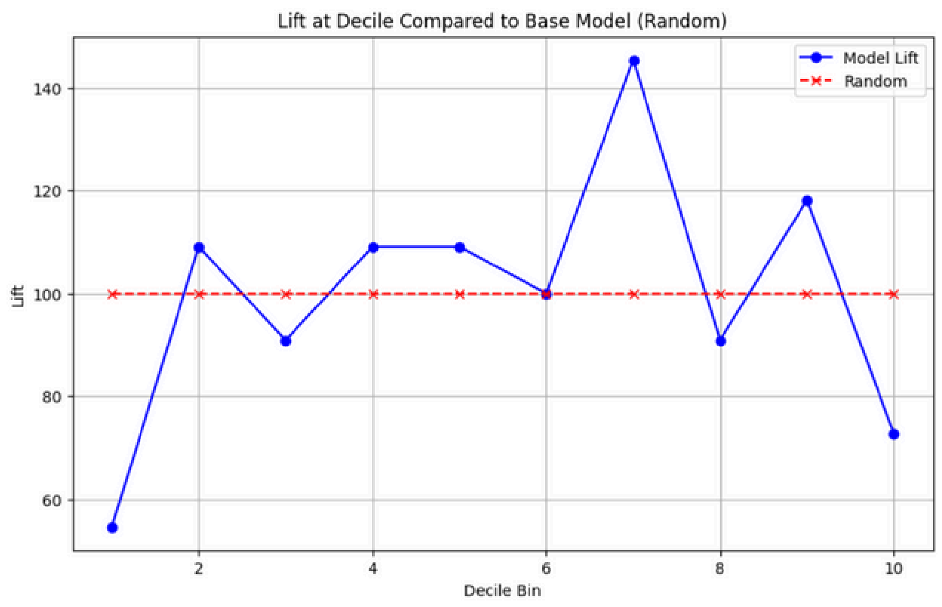


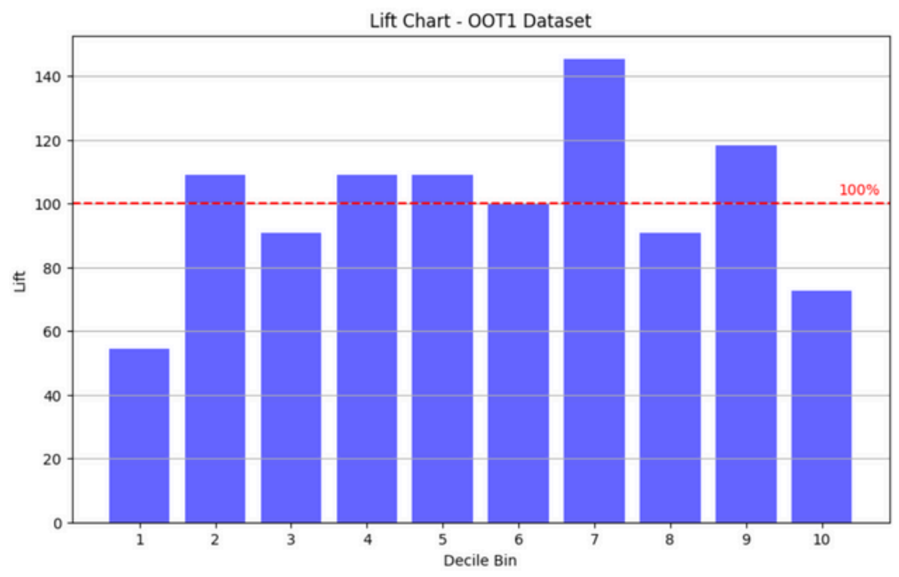
	model_name	parameters	rows	columns	%_target	accuracy	auc	precision	recall	f1-score
0	Train: LightGBM	<bound method LGBMModel.get_params of LGBMClas...	5600	18	10	0.66	0.84	0.22	0.89	0.36
1	Test: LightGBM	<bound method LGBMModel.get_params of LGBMClas...	2400	18	10	0.56	0.53	0.12	0.48	0.19
2	OOT1: LightGBM	<bound method LGBMModel.get_params of LGBMClas...	1000	18	11	0.58	0.52	0.12	0.43	0.18
3	OOT2: LightGBM	<bound method LGBMModel.get_params of LGBMClas...	1000	18	11	0.57	0.51	0.11	0.39	0.17

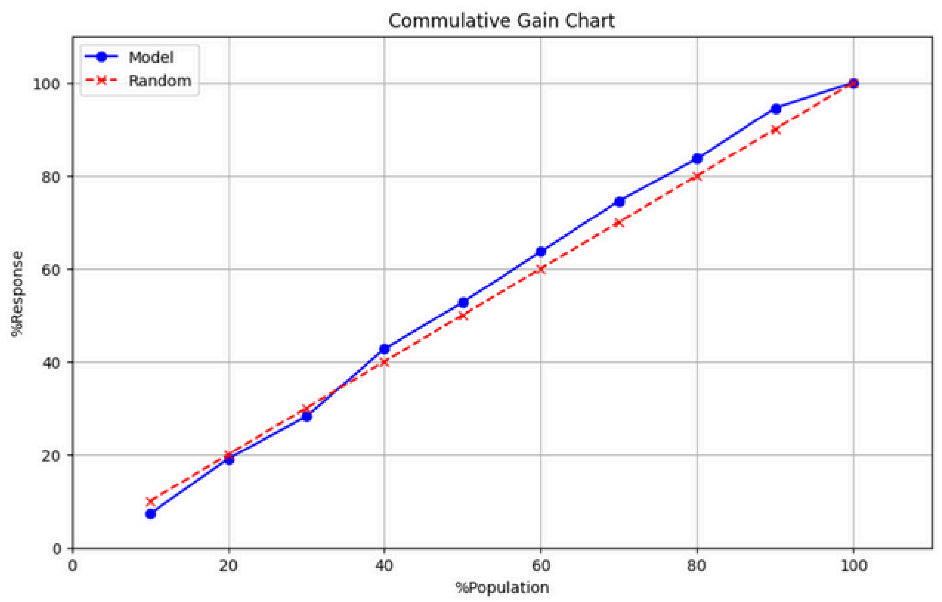
- This model cannot be used for production properly because the AUC is quite low.
- The model is overfitting to the training data and is not good at distinguishing the target. The numerical data was generated randomly, and despite attempts to bias some features towards the target, there is still no discernible pattern.
- Recall is better than other metrics, meaning the model remembers the target for almost half of the total.

Propensity Model

Interpretation







- In general, if the model distinguishes the target efficiently, the percent lift will be higher than the baseline and highest at bin 10. However, in this case, the model is not good at distinguishing the target, as the target is distributed quite randomly.
- On the Cumulative Gain Chart, if we send the campaign to 80% of the customers, the response would be around 82% of the target, which is only slightly better than randomly sending the campaign.

Model Result Usage

Model Result Usage

Propensity Model

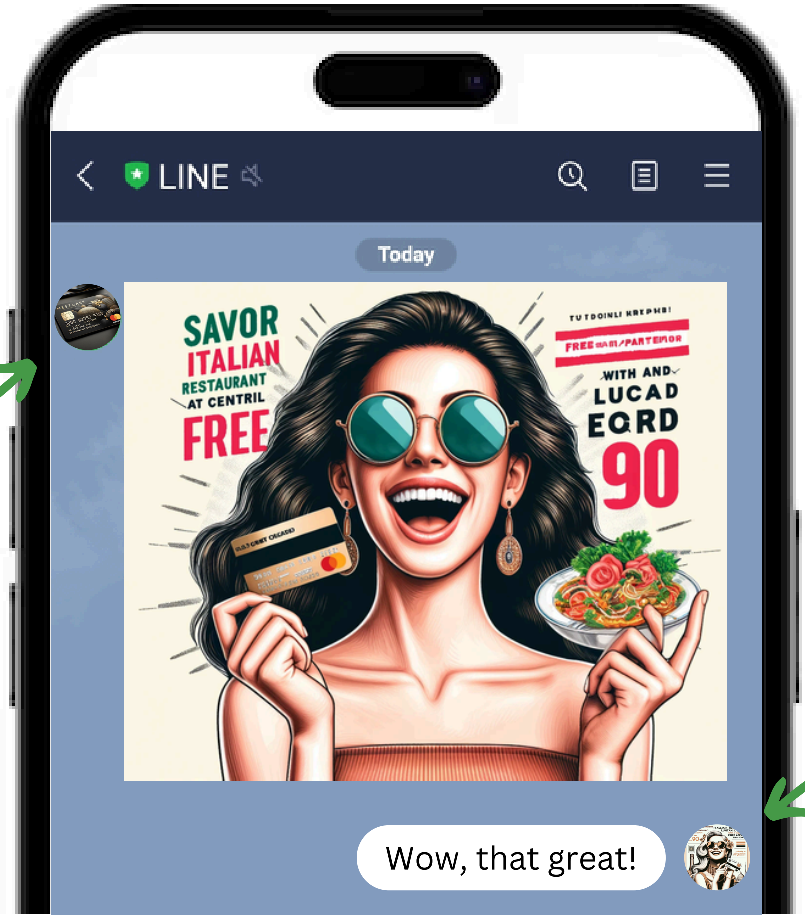
Collaborative Filtering

Combination

Marketing Lead

<u>Example:</u> sending campaign to customer from recommendation system and propensity score though line channel

Recommend: Pro_restrnt50



Customer ID 66000287

pro_restrnt50	Savor Italian restaurant lunch at Central Ladprao with a free appetizer and earn 90 points.

Summary

Summary

The Propensity Model is used for selecting credit card customers who are most likely to take the campaign if there is a limited budget.

Collaborative Filtering is used for recommending campaigns that customers have never taken before to increase credit card spending.

In this dataset, the results of the propensity model are not good enough for production use. It needs fine-tuning, adding more features, or using more realistic features and targets.

Currently, this system can recommend campaigns to customers who are likely to take them. In the future, the optimal time to offer the campaign can be determined to send it when customers are most likely to spend their money.