The American Express Innovation Labs AI Hackathon (Singapore) 2024

Modeling Problem Statement

Recommendations to drive **incremental activation**
on relevant merchants for each customer

Merchant Recommender: Overview

Merchant Recommender is a capability to **connect** Amex **Merchants** (shops accepting Amex cards) & Amex Credit Card **Customers**

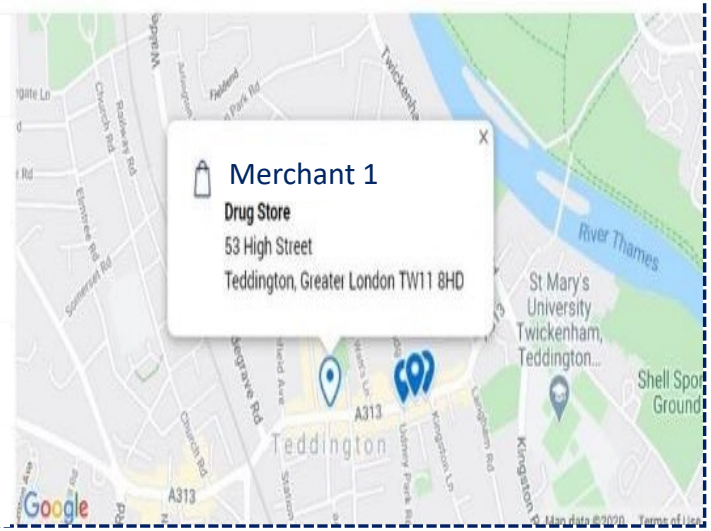
This capability recommends* **personalized** merchants to Credit Card Customers, to help Customers **discover** merchants near to them to shop, increase **spend on merchants** & benefit Amex enterprise by driving more **revenue**

Customer visits the channel

Recommender Algorithm

Personalized Merchants for that Customer

- Merchant 1
- Merchant 2
- Merchant 3
- Merchant 4
- Merchant 5

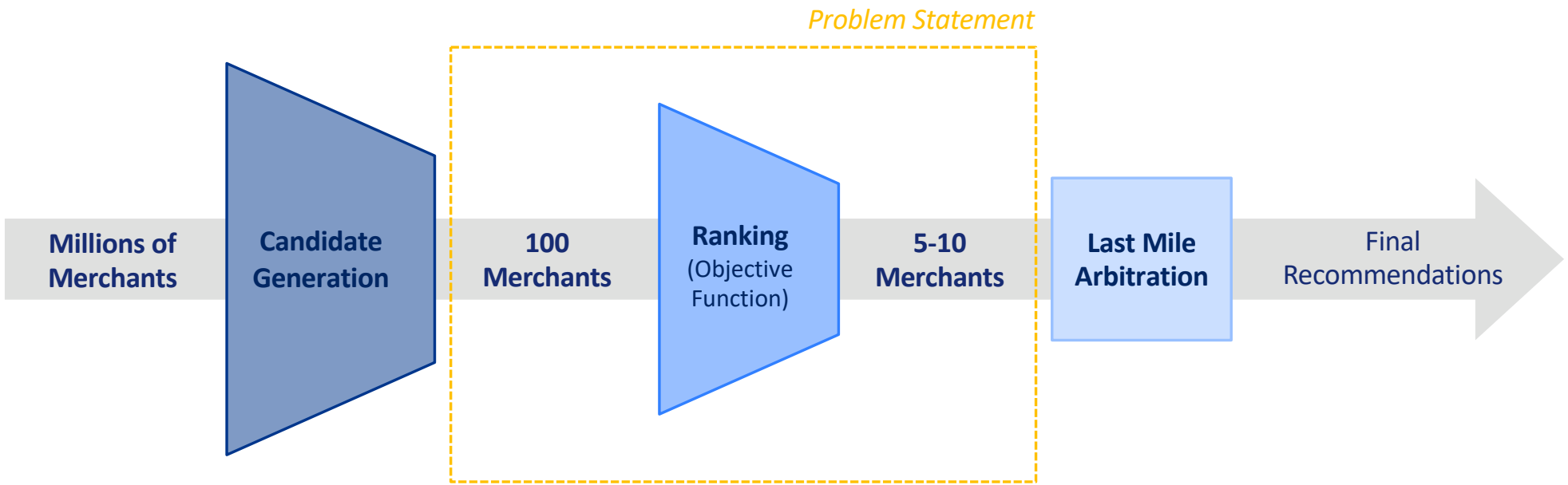


Sample logged in experience of Merchant recommender capability on Amex Website

*Recommendation means ***showing*** an Amex Merchant to an Amex Credit Card Customer on a marketing channel

Merchant Recommender: Architecture

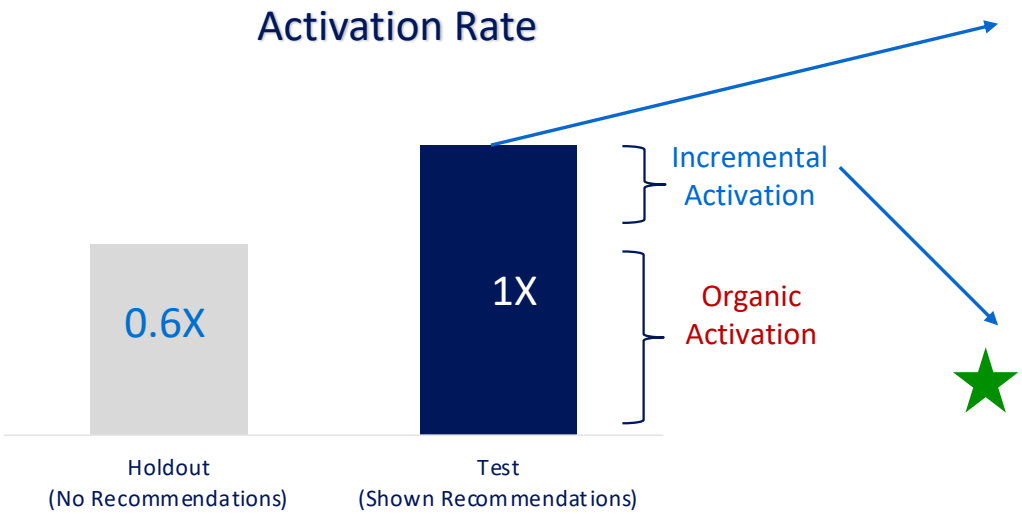
Merchant Recommender architecture is designed as per industry best practices, and comprises of 3 modules -



** Pre-select 100 nearest merchants for each customer that he/she hasn't visited for the past three months and recommend top 10 merchants out of 100*

Motivation

By activating ‘net-new’ merchants that the Customer would not have discovered otherwise, unless recommended



Naive Response Algorithm maximizes Test Activation Rate only

Objective Function

P (Y = 1 | Recommendation)

Uplift Algorithm maximizes the **difference** between Test & Holdout Activation Rate (or incremental activation)

Objective Function

P (Y = 1 | Recommendation) –
P (Y = 1 | no Recommendation)

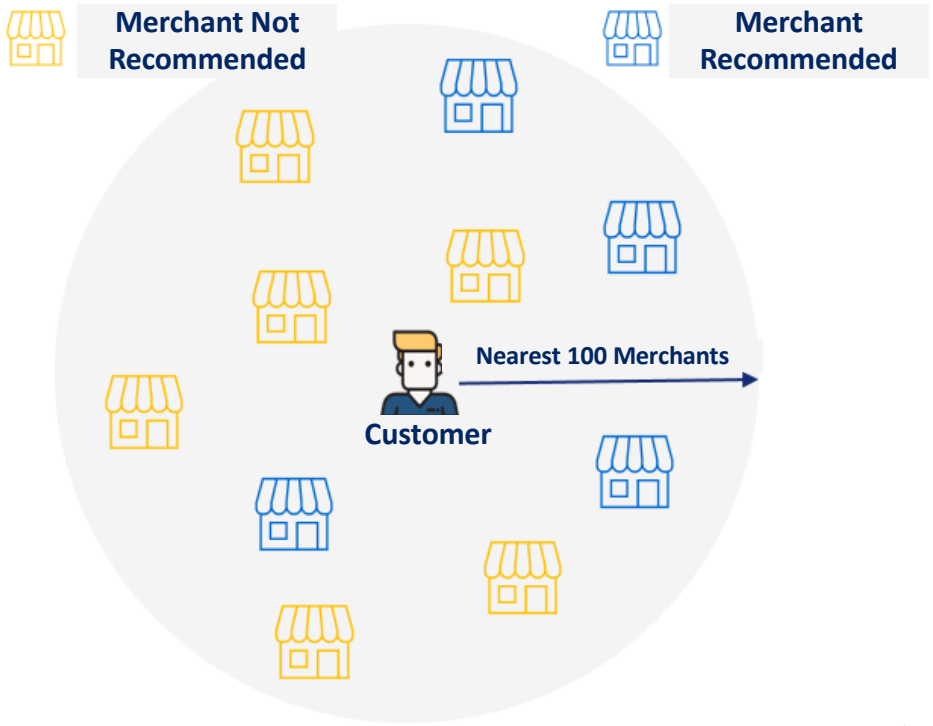
#Recommendations – Unique Customer x Merchant

#Activations - First transaction (irrespective of $ amount) on Merchant within 30 days post seeing the recommendation

Activation Rate - #Activations / #Recommendations

Merchants that are activated without recommendation is organic activation

Data Collection and Objective Function



Data Collection

- A Customer logs into Amex website
- Out of the nearest 100 merchants to the Customer, we **randomly** recommend some merchants and don't recommend remaining merchants
- Out of the recommendations, we have four types of outcomes:
  - Event A: merchant activated (Y=1) given it is recommended,
  - Event B: merchant activated (Y=1) given it is not recommended → refer to as ***organic activation***
  - Event C: merchant not activated (Y=0) given it is recommended,
  - Event D: merchant not activated (Y=0) given it is not recommended.

Conditional Activation Rate for Each Customer:

- Activation rate for recommended merchants: AR1 = $\frac{no.\,of\ A}{no.of\ A+no.of\ C}$
- Activation rate for non-recommended merchants/ organic activation rateAR2 = $\frac{no.\,of\ B}{no.of\ B+no.of\ D}$

Goal: to **maximize Incremental Activation Rate** on Merchants for **each** Customer**.**

Incremental activation rate
= P [(Y = 1 | Recommendation) – (Y = 1 | no Recommendation)
= AR1 – AR2

Evaluation Criteria

A good solution should be able to rank order well with Incremental Activation Rate

Illustration

Rank within Customer	Activation Rate (%)		
	Recommended	Not-Recommended	Incremental
1 - 10	0.55%	0.21%	0.33%
11 - 20	0.20%	0.11%	0.09%
21 - 30	0.14%	0.09%	0.05%
31 - 40	0.10%	0.08%	0.02%
41 - 50	0.08%	0.07%	0.01%
51 - 60	0.08%	0.07%	0.01%
61 - 70	0.08%	0.09%	-0.01%
71 - 80	0.09%	0.11%	-0.02%
81 - 90	0.12%	0.19%	-0.06%
91 - 100	0.19%	0.35%	-0.16%

This table is created at dataset level post collating ranks of all Customers

Steps:

- Data is scored at Customer x Merchant level
- Ranks are created **within each Customer** basis **predicted** score
- **Incremental Activation rate** is calculated within the top 10 merchants (ranked by the predicted scores) for each customer

Evaluation Metric:

- ❑ Incremental Activation Rate on Top 10 merchants by prediction for each Customer

*Evaluation code to be provided to participants

Data

Illustration

Sample of the Modeling Dataset

Primary Key Columns

Customer *(masked)*	**Merchant** *(masked)*	***ind_recommended***	***activation***	**X Features (67)**
C1	L1	1	0	….
C1	L2	1	1	….
…	…	…	…	….
C1	L6	0	1	….
C1	L7	0	0	….
…	…	…	…	….

Variable Name	Description
Customer (Masked)	Masked Customer Identifier
Merchant (Masked)	Masked Merchant Identifier
ind_recommended	1 if Merchant was recommended to Customer, 0 otherwise.
activation	1 if Customer transacts at Merchant within 30 days post recommendation, 0 otherwise. This cannot be used as independent feature directly or indirectly.
X Features (67)	67 Ready-to-use Independent Features are provided (details on next slide)

*ind_recommended & activation are only going to be provided for Training data

Details about Independent Predictors – X Features

Feature Type	# Features
Numerical	66
Categorical (Merchant_Profile_01)	1
	67

- For some of the features, raw feature descriptions have also been provided. For others, it's same as the feature names.
- None of the features have been normalized or imputed by any value

Feature Category	# Features	Feature Names
Customer Digital activity	22	Customer_Digital_activity_1 … Customer_Digital_activity_22
Customer Spend	14	Customer_Spend_1 … Customer_Spend_14
Merchant Spend	11	Merchant_Spend_1 … Merchant_Spend_11
Distance	5	Distance_1 … Distance_5
Customer Industry Spend	5	Customer_Industry_Spend_1 … Customer_Industry_Spend_5
Customer Profile	4	Customer_Profile_1 … Customer_Profile_4
Customer Merchant	3	Customer_Merchant_1 … Customer_Merchant_3
Merchant Profile	3	Merchant_Profile_1 … Merchant_Profile_3
	67	

Datasets

TBD

Data		Column Details	# Rows	# Columns	Location/Name
Train Data		Customer, Merchant, ind_recommended, activation + 67 X Features	12,229,978	71	*To be downloaded from the unstop website (available to all registered candidates)*
Round 1 Submission Data	Leaderboard Evaluation & Closed Evaluation	Customer, Merchant + 67 X Features	12,604,600	69	*To be downloaded from the unstop website as a single file (available to all registered candidates)*

- Both datasets have been created on different time periods and are unique at Customer x Merchant level
- All datasets have no overlap with respect to Customer, Merchant or Customer x Merchant
- Datasets are comma-separated (,)
- Please note that *ind_recommended* & *activation* columns will not be provided with Round 1 Submission data

Sample of a scored Round 1 submission data

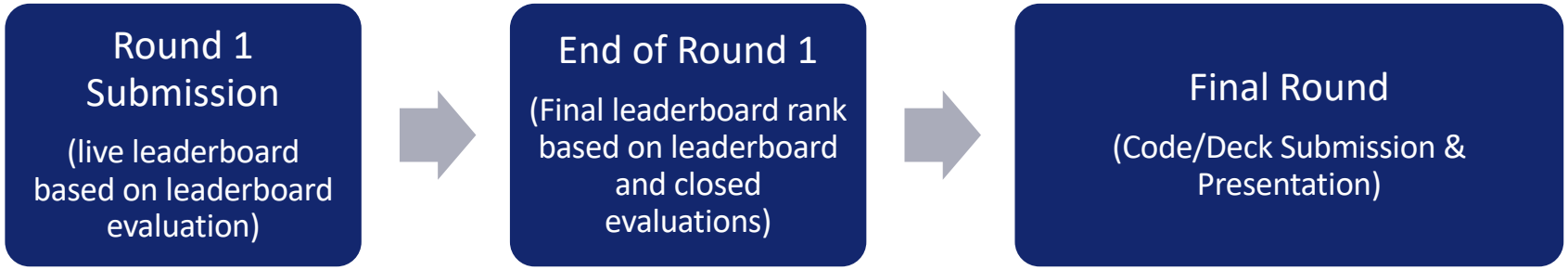
Primary Key Columns

Customer *(masked)*	**Merchant** *(masked)*	**Predicted Score**
C1	L1	0.5
C1	L2	0.01
…	…	….
C1	L6	0.2
C1	L7	0.09
…	…	….

Predicted Score:

- Score to be calculated at each row i.e., unique combination of **Customer x Merchant**
- This predicted score will be used for evaluation
- Scored submission data unique Customer x Merchant counts should match with Submission Data Round 1 counts mentioned in previous slide

Stages of Competition



Round 1 Guidelines

- Participants will train their model using labeled Training Data
- They are free to split the training data into Train & Out-of-sample/In-time data into whatever ratio they deem fit. They are also free to use any sampling technique on Train data*
- An evaluation custom code (in python) to calculate "Incremental Activation Rate (at dataset level) on Top 10 merchants (by prediction) for each Customer" would be provided to participants to run on their training/out-of-sample data scores to mimic the exact evaluation process that will run on their submitted Round 1 data. Directions to run the code have been mentioned in the beginning of the code. This code cannot be used on scored Round 1 submission data as it doesn't have the 'activation' column.
- Once they have the best model according to them, they can score the Round 1 submission data and upload it
- Scored Round 1 submission data should be a csv & delimited by ",". The first row should be the header with column names: *customer, merchant & predicted_score*. None of these 3 columns can be missing/null/na etc.
- Scored Round 1 submission data should have only 3 columns and unique Customer x Merchant count should be 12,604,600
- Top 15 Teams with Highest "Incremental Activation Rate on Top 10 merchants by prediction for each Customer (Leaderboard + Closed evaluation datasets)" will be shortlisted for Round 2

*Train data has already been sampled once before sharing, hence will contain < 100 merchants per Customer

Round 2 Guidelines

- Top 15 teams from Round 1 will be asked to upload
    1. Code used for the task
    2. Deck to communicate details of the approach to arrive at the final solution

- The deck should detail their approach including (but not limited to) sampling technique, Feature Engineering/Selection, Algorithm/Modelling Framework used and Model Intuition.

- Teams participating in this round may be given an interview opportunity to work/intern at AMEX Singapore

- Finally, the Top 5 teams with the best code and deck will be shortlisted and invited for the final round

Final Round Guidelines

- Top 5 teams from Round 2 will be invited to the American Express Singapore office (One Marina Boulevard) to present their work to a panel of judges

- Performance in this round will determine the placement of the top 3

DISCLAIMER:
Please note that some forms of recordings may take place during the event, including photography, video and/or audio recordings. By attending the event, you will hereby give your consent for your image or voice ('data') to be recorded at the event and for such data to be used or quoted in internal printed materials at American Express and/or external media channels (including social media sites).

All the Best!!!