Problem Statement

Recent years have witnessed a surge in the number of internet savvy users. Companies in the financial services domain leverage this huge internet traffic arriving at their interface by strategically placing ads/promotions for cross selling of various financial products on a plethora of web pages. The digital analytics unit of Best Cards Company uses cutting edge data science and machine learning for successful promotion of its valuable card products. They believe that a predictive model that forecasts whether a session involves a click on the ad/promotion would help them extract the maximum out of the huge clickstream data that they have collected. You are hired as a consultant to build an efficient model to predict whether a user will click on an ad or not, given the following features:

* Clickstream data/train data for duration: (2nd July 2017 – 7th July 2017)
* Test data for duration: (8th July 2017 – 9th July 2017)
* User features (demographics, user behaviour/activity, buying power etc.)
* Historical transactional data of the previous month with timestamp info (28th May 2017– 1st July 2017) (User views/interest registered)
* Ad features (product category, webpage, campaign for ad etc.)
* Date time features (exact timestamp of the user session)

Data

**Train Data**

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| session\_id | Unique ID for a session |
| DateTime | Timestamp |
| user\_id | Unique ID for user |
| product | Product ID |
| campaign\_id | Unique ID for ad campaign |
| webpage\_id | Webpage ID at which the ad is displayed |
| product\_category\_1 | Product category 1 (Ordered) |
| product\_category\_2 | Product category 2 |
| user\_group\_id | Customer segmentation ID |
| gender | Gender of the user |
| age\_level | Age level of the user |
| user\_depth | Interaction level of user with the web platform (1 - low, 2 - medium, 3 - High) |
| city\_development\_index | Scaled development index of the residence city |
| var\_1 | Anonymised session feature |
| is\_click | 0 - no click, 1 - click |

**Historical User logs**

|  |  |
| --- | --- |
| **Variable** | **Definition** |
| DateTime | Timestamp |
| user\_id | Unique ID for the user |
| product | Product ID |
| action | view/interest (view - viewed the product page, interest - registered interest for the product) |

Evaluation Metric

The evaluation metric for this competition is AUC-ROC score.

Public and Private Split

Test data is further randomly divided into Public (30%) and Private (70%) data.

* Your initial responses will be checked and scored on the Public data.
* The final rankings would be based on your private score which will be published once the competition is over.

Solution Checker

1. Adding comments is mandatory for the use of solution checker.
2. Comments will help you to refer to a particular solution at a later point in time.

Hackathon Rules

1. Setting final submission is mandatory. Without a final submission, your entry will not be considered.
2. You are only allowed to make 20 submissions a day
3. Use of external datasets is not allowed
4. Code file is mandatory while sending final submission. For GUI based tools, please upload a zip file of snapshots of steps taken by you, else upload code file.
5. The code file uploaded should be pertaining to your final submission.
6. Use of session\_id variable in any way is not allowed as part of the model/solution