Predict Click-Through Rates Challenge

Jampp – UTDT

1. Overview

In display and mobile advertising, the most relevant technical development in recent years is the growth of Real-Time Bidding (RTB). Billions of display ad impressions are purchased on a daily basis through a public auction hosted by RTB exchanges. The auction is per impression and the process usually occurs less than 100 milliseconds before the ad is placed.

In this context, demand side platforms (DSPs) were created to help advertisers manage their campaigns and optimise their real-time bidding activities. One of their main tasks is to predict the click-through rate (CTR) for ads and calculate the bid price according to the estimated CTR.

For this challenge, Jampp, one of the key players in the mobile adtech industry, has provided a week's worth of data to build and test models predicting CTR. Can you find good strategy for making these predictions?

2. Data

File descriptions

- **train**: The training set consists of a portion of Jampp's click-through data over a period of 7 days, not necessarily chronologically ordered (however a time column is provided).
- **test**: The test set consists of Jampp's traffic on the day following the training period (computed/sampled in the same way as the training set).
- sample_submission.csv: A sample submission file in the correct format.

Data Fields

- **Label**: Target variable that indicates if the ad was clicked (1) or not (0).
- **action_categorical_0**: business unit id. 1st level. These ids have a one-to-many relationship (a 1st level id can have many 2nd level, but a 2nd level can have only a single 1st level parent id)
- **action_categorical_1**: business unit id. 2nd level
- **action_categorical_2**: business unit id. 3rd level
- **action_categorical_3**: business unit id. 4th level
- **action_categorical_4**: business unit id. 5th level
- **action_categorical_5**: a categorical variable
- **action_categorical_6**: a categorical variable
- **action_categorical_7**: a categorical variable
- **action_list_0**: a list of categories related to the action
- **action_list_1**: a list of entity ids related to the action
- **action_list_2**: a list of entity ids related to the action
- **auction_time**: the time in which the auction took place. Format is unix time (a.k.a. epoch time).
- **auction_age**: the user's age
- **auction_bidfloor**: bidfloor (the minimum value you can bid for)

- **auction_boolean_0**: a boolean variable regarding an attribute of the auction
- **auction_boolean_1**: a boolean variable regarding an attribute of the auction
- **auction_boolean_2**: a boolean variable regarding an attribute of the auction_not_track
- **auction_categorical_0**: an entity id related to a component of the auction
- **auction_categorical_1**: an entity id related to a component of the auction
- **auction_categorical_2**: a categorical variable
- **auction_categorical_3**: a categorical variable
- **auction_categorical_4**: a categorical variable
- **auction_categorical_5**: a categorical variable
- **auction_categorical_6**: a categorical variable
- **auction_categorical_7**: an entity id related to a component of the auction
- **auction_categorical_8**: an entity id related to a component of the auction
- **auction_categorical_9**: an entity id related to a component of the auction
- **auction_categorical_10**: a categorical variable
- **auction_categorical_11**: an entity id related to a component of the auction
- **auction_categorical_12**: a categorical variable
- **auction_list_0**: a list of categories related to the auction
- **creative_categorical_0**: business unit id
- **creative_categorical_1**: a categorical variable
- **creative_categorical_10**: a categorical variable
- **creative_categorical_11**: a categorical variable
- **creative_categorical_12**: a categorical variable
- **creative_categorical_2**: a categorical variable
- **creative_categorical_3**: a categorical variable
- **creative_categorical_4**: a categorical variable
- **creative_categorical_5**: business unit id
- **creative_categorical_6**: a categorical variable
- **creative_categorical_7**: a categorical variable
- **creative_categorical_8**: a categorical variable
- **creative_categorical_9**: a categorical variable
- **creative_height**: the creative's height (in pixels)
- **creative_width**: the creative's width (in pixels)
- **device_id**: the device id. This is a unique identifier for a device (or close to unique)
- **device_id_type**: the type of device id (there are many types, also a device can have many different types, with a different id for each of them)
- **gender**: gender
- **has_video**: a boolean value indicating if the banner contains a video.
- **timezone_offset**: timezone offset in hours based on the auction's country and region.

3. Evaluation

Submission Format

The submissions should contain the predicted probability of click for each ad impression in the test set, one per line:

```
id,pred
1,0.6591843
2,0.5749322
3,0.3478105
4,0.2685858
```

```
5,0.2997332
...
```

4. About Jampp

Jampp is a programmatic marketing platform for acquiring and retargeting mobile app customers that is growing at a fast pace. The company is currently at the forefront of programmatic technology, developing groundbreaking products to serve a global client base from offices in San Francisco, London, Berlin, São Paulo, Singapore, Cape Town and Buenos Aires.

Jampp's world-class engineers have built a scalable platform from the ground up: their proprietary in-house tools can process over 100 terabytes of data per day to make intelligent bids in real-time. The machine learning algorithms feed off billions of these data points to determine, in less than 80 milliseconds, which banner is the best fit for a given user (and this is done more than 500,000 times per second!). And it doesn't end there: all this big data and machine learning wizardry is achieved with a tech team of only about 20 people!