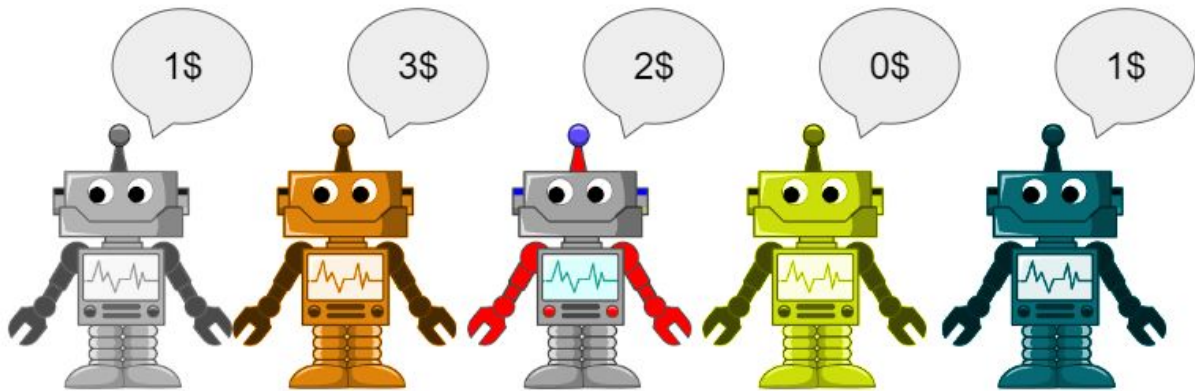


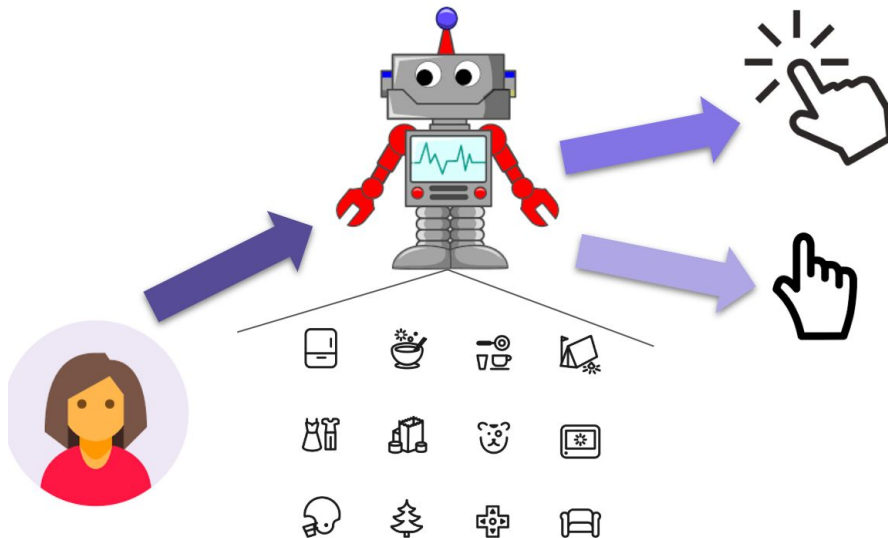
Click Prediction Challenge

Understanding the CPC (cost per click) market place

Every time a user opens a mobile app, an auction starts between several agents. The winner of the auction gets to display an ad to the user.



However, the agent gets paid only if the user clicks on the ad



In order to make an informed decision, each agent needs to take the users features into account and make a prediction.

The Dataset

The data is available for download at this link:

https://drive.google.com/file/d/1Gtda7RggD_FX3f8OAN73ZOc-pK-cGhJr/view

The data is given in CSV format, with the last column being our target variable (click / skip).

The data is taken from an ad marketplace, between August 2018 and November 2018, and the data is filtered on bid-requests of Android users from the United states.

Each row is a bid-request has the following user features:

App id	User state	User isp	Device maker	Device model	Device osv	Device height	Device width
-----------	------------	----------	--------------	--------------	------------	---------------	-----------------

As well as 2 special fields:

- **Utc_time** - the number of mili seconds that passed since January 1st 1970
 - A common time format, also known as “epoch”
- **Click** - Our target variable
 - 1 - if the user clicked
 - 0 - if the user ignored our ad

Task

You are required to design a bidding strategy (and eventually make predictions) given the dataset at hand.

We would like you to at least follow these steps:

1. Load the dataset, clean the data
2. Explore the data (EDA)
3. Rank existing features importance
4. Engineer new features (e.g. day-of-week)
5. *Optional*: Enrich the dataset with external sources (e.g. timezones, app categories)
6. Predict whether the user clicked the ads or not
7. Evaluate the model with metrics

Tips:

1. If you split the data to train and test sets, do it wisely and not at random
2. A combination of features could be more helpful than a single feature prediction.
3. If you filtered data out, make sure you have enough records to have a statistical significant prediction.
4. Make sure to use evaluation metrics which are useful for evaluating the model.
5. Explore on the Internet for further resources and deepen your analysis.

Submission

- The language used should be Python 3.
- **You should always document your code, detail your thoughts process and elaborate on your logic.**
- The submitted files should at least include a jupyter notebook (format .ipynb).
- Your graphs should always include titles, legends, labels and be well formatted.
- You may use any library of your choice.