Gender Prediction for E-Commerce

With the evolution of the information and communication technologies and the rapid growth of the Internet for the exchange and distribution of information, Electronic Commerce (e-commerce) has gained massive momentum globally, and attracted more and more worldwide users overcoming the time constraints and distance barriers.

It is important to gain in-depth insights into e-commerce via data-driven analytics and identify the factors affecting product sales, the impact of characteristics of customers on their purchase habits.

It is quite useful to understand the demand, habits, concern, perception, and interest of customers from the clue of genders for e-commerce companies.

However, the genders of users are in general unavailable in e-commerce platforms. To address this gap the aim here is to predict the gender of e-commerce's participants from their product viewing records.

About Data Source:

PAKDD 2015 Conference

Data Dictionary

Train file: CSV containing the product viewing data with gender as label

Variable	Definition
session_id	Session ID
startTime	Start time of the session
endTime	End Time of the session
ProductList	List of products viewed
gender	(Target) male/female

Product list contains list of products viewed by the user in the given session and it also contains the category, sub category, sub-sub category and the product all encoded and separated with a slash symbol. Each consecutive product is separated with a semicolon.

Test file: CSV containing sessions for which gender prediction is to be submitted

Variable	Definition
session_id	Session ID

startTime	Start time of the session
endTime	End Time of the session
ProductList	List of products viewed

Submission file format

Variable	Definition
session_id	Session ID
gender	(Target) Male/Female

How to Make a Submission?

Evaluation Metric

Submissions are evaluated on accuracy between the predicted and observed gender for the sessions in the test set.

Public and Private Split

Test sessions are further divided into Public (40%) and Private (60%)

- 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.

Guidelines for Final Submission

Please ensure that your final submission includes the following:

1. Solution file containing the predicted gender (male/female) in the test dataset (format is given in sample submission csv)
2. Code file for reproducing the submission, note that it is mandatory to submit your code for a valid final submission

How to Set Final Submission?

Hackathon Rules

1. *The final standings would be based on private leaderboard score*
2. Setting the final submission is recommended. Without a final submission, the submission corresponding to best public score will be taken as the final submission
3. Use of external dataset or labels from open source will lead to disqualification from the leaderboard
4. Entries submitted after the contest is closed, will not be considered
5. The code file pertaining to your final submission is mandatory while setting final submission
6. Throughout the hackathon, you are expected to respect fellow hackers and act with high integrity.