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:**CSVcontaining 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 |