# Data Description

Table 1: Description of attached files

|  |  |
| --- | --- |
| File Name | Description |
| training.csv | Training Data: Contains the samples needed to train the model  Record count: 96,406  Positive samples: 1,465  Negative samples: 94,941 |
| validation.csv | Validation Data: Contains the samples for validating the model.  Record count: 80,008  Positive samples: 620  Negative samples: 793,88 |
| interest\_topics.csv | Contains the topic label and topic description. |

### Interest Topics Dataset

Each row in this dataset represents one of the interest topics.

# of record in the file: 1,411

Table 2: Schema for interest\_topics.csv

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| topic\_id | Integer | Numerical identifier of the topic |
| topic\_name | String | Interest topic |

### Training and Validation Dataset

Each row in this dataset represents a shopper’s long-term and short-term interests in one of the 1,411 categories that we have.

Table 3: Schema for files training.csv and validation.csv

|  |  |  |
| --- | --- | --- |
| Column Name | Data Type | Description |
| userID | Integer | User identifier. Unique for each row in this csv file. |
| inAudience | Boolean | This column represents whether or not the shopper has converted in the past.  TRUE: Shopper has converted  FALSE: Shopper has not converted |
| ltiFeatures | Dictionary  Key: String  Value: Double | Long Term Interests.  The dictionary’s keys are one of interest topics. The value represents the proportional interest the user has in that topic.  For example, if the *ltiFeatures* of one of the users is {‘34’: 0.7, ‘41’:0.3}, it can be inferred as follows. The keys represent the *topic\_id* in the interest\_topics.csv file.  Topic 34 is */Arts & Entertainment/Movies*  Topic 41 is */Games/Computer & Video Games*  This user has 70% interest in */Arts & Entertainment/Movies* and 30% interest in */Games/Computer & Video Games*. |
| stiFeatures | Dictionary  Key: String  Value: Double | Short Term Interests  Description is same as above, however, for shopper’s short term interest. |