← Go Back to Elective Project

≣ Course Content

Project FAQ - Amazon Product Recommendation System

1. When applying the ranking products function and sorting it based on the corrected ratings, I am getting the below error.

```
KeyError: 'rating_count'
```

How to resolve this error?

Below are the possible reasons and the solutions for getting this error:

- **Wrong naming**: When you access the column name "rating_count" from the data frame, it should be accessed with the correct name. By the correct name I mean to say that upper case, lower case, and correct spelling of the column have to be taken care of. Along with this, the name of dataframe has also to be correct.

The correct function code: The correct code for the function "ranking_products" has to be used. If a slight change is made in the function then it is not expected to work properly. Below is the correct code for the "ranking_products" function.

```
def ranking_products(recommendations, final_rating):
  # Sort the products based on ratings count
  ranked_products = final_rating.loc[[items[0] for items in recommendations]].sort_values('rating_count', ascend:
  # Merge with the recommended products to get predicted ratings
  ranked_products = ranked_products.merge(pd.DataFrame(recommendations, columns = ['prod_id', 'predicted_ratings
  # Rank the products based on corrected ratings
  ranked_products['corrected_ratings'] = ranked_products['predicted_ratings'] - 1 / np.sqrt(ranked_products['rat:
  # Sort the products based on corrected ratings
  ranked_products = ranked_products.sort_values('corrected_ratings', ascending = False)
  return ranked_products
```

2. While recommending the top 5 items with 50 and 100 interactions for the Rank Based Recommendation System, I am getting the duplicate results. How to resolve this?

To resolve this issue, the "top_n_products" function has to be used with the correct code.
Below is the correct code for the function.

```
# Defining a function to get the top n products based on the highest average rating and minimum interactions
def top_n_products(final_rating, n, min_interaction):
  # Finding products with minimum number of interactions
  recommendations = final_rating[final_rating['rating_count'] > min_interaction]
  # Sorting values w.r.t average rating
  recommendations = recommendations.sort_values(by = 'avg_rating', ascending = False)
  return recommendations.index[:n]
```

3. How to install surprise in google colab?

To install surprise in Google Colab, you can use the following line of code -

```
!pip install surprise
```

4. Is google colab mandatory for this project?

Yes, Google Colab is mandatory for this project. In this project a considerable amount of data has to be used and to operate on that, high GPU and RAM will be needed. This is properly incorporated in Google Colab and hence it is mandatory.

5. What is the purpose of checking missing values in the data in this project?

While acquiring data it is common to miss some records. Such records are called as missing values. The basic purpose of checking missing values is to understand the count of Nulls and None in the data. There are few methods of checking missing values in python that are given below -

If df is the name of the dataframe then to check missing values below code can be used.

```
df.info()
```

Another way of doing that is

```
df.isnull().sum()
```

6. While working with grid search cv, I am getting a different optimum model than the notebook. How to resolve this?

The one and only reason for this is you are passing a different dataset to the GridSearchCV code. Trying checking the dataset in terms of number of records, features, and whether correct set of records are passed or not.

7. In the below line of code -

```
rows, columns = df_final___
```

How to find the number of rows and columns in a single line of code?

For this, the most suitable method in python is the "shape" method. df_final.shape will give the requisite output in form of a tuple where the first element is the number of rows and the second element is the number of columns.

8. What is meant by a baseline model? Is it same for different algorithms?

A baseline model is one with default parameters. For different set of algorithms the model with default parameters will be different and so will be the baseline models.

9. In the final dataset, I am having one less number of records. How to resolve this?

For this project, the dataset is huge and the number of records used by code depends on how much data was uploaded to GPU/ RAM. Once the dataset is loaded to colab completely, on running the code the whole dataset should appear in the notebook.

10. Can the info method used to check the datatypes of the features and also the count of null values in different features?

Yes, it can be used because it gives both datatype and the count of null values from a dataframe as its output.

11. What does the term "interactions" mean in this project?

If a user has rated a product it is considered as an interaction between the user and the product.

12. What is the meaning of summary statistics of a dataframe in this project (it is asked to find the summary statistics of rating column in the notebook)?

The summary of a column in a dataframe includes the count of records, average value, standard deviation, minimum value, the first, second, and the third quartiles, and the maximum value of the column. In python to do so we can use the "describe" method on the dataframe/series.

Previous

Next

Help