Article Recommendation

In the previous module, we understood content-based filtering and its significance in the recommender system. As a part of the practice, Implement content-based filtering to provide recommendations for the same problem that we used in Assignment 1 & 2.

**Problem Statement**

Sérendipité is an article aggregation platform where articles from different domains such as technology, politics, news, and so on are shared by its users and then these articles are recommended on the basis of reading habits.

In Assignment 1, you used a simple popularity-based system with no personalization.  
In Assignment 2, you implemented collaborative filtering to provide the recommendations.  
  
As you already know they have a rating system for articles under which the users when they read the article rate it on a scale of 1 to 5

Now in this assignment, you are expected to take one step ahead and implement content-based filtering by accurately predicting ratings for each user article combination.

**Data Description**

**train.zip**

train.csv

|  |  |
| --- | --- |
| Variable | Definition |
| user\_id | Unique ID for the user |
| article\_id | Unique ID for the article |
| rating | Rating provided by the user (1-5) |

article\_info.csv

|  |  |
| --- | --- |
| Variable | Definition |
| article\_id | Unique ID for the article |
| website | Name of the website where the article is hosted |
| title | Title of the article |
| content | Complete text for the article |

**test.csv**

|  |  |
| --- | --- |
| Variable | Definition |
| user\_id | Unique ID for the user |
| article\_id | Unique ID for the article |

**sample\_submission.csv**

|  |  |
| --- | --- |
| Variable | Definition |
| user\_id | Unique ID for the user |
| article\_id | Unique ID for the article |
| rating | Predicted rating for the user article combinations in the test file |

**Evaluation Metric**

The evaluation metric for this problem is RMSE Score.

**Guidelines for Final Submission**

Please ensure that your final submission includes the following:

1. Solution file containing the predicted rating for each user article combination in the test set
2. Code file for reproducing the submission, note that it is mandatory to submit your code for a valid final submission

**Download**

You can download the dataset from the link given below (This is the same dataset that you used in Assignment 1&2). Once you have finished the assignment, submit the jupyter notebook and we will evaluate it.