16CA437: Big Data Analytics and Visualization

Apache Mahout – Assignment 1

Create a User-based recommender using the provided Sample_Data.csv data file. This csv file consists of three columns and the first column is “user_id”, second column is “item_id” and the third column is “rating”. The item_id can represent anything like movies, music, hotels, books, online shopping items, etc. The third column indicates the ratings provided by the users to each items and it ranges from 1 to 5. Here 1 is considered as the worst rating and 5 as best rating. In Mahout, there are different similarity measures available like **PearsonCorrelationSimilarity**, **EuclideanDistanceSimilarity, TanimotoCoefficientSimilarity, LogLikelihoodSimilarity,** etc. You can call these similarity measures in your java program as follows;

```
UserSimilarity similarity = new PearsonCorrelationSimilarity(model);

UserSimilarity similarity = new LogLikelihoodSimilarity(model);

UserSimilarity similarity = new TanimotoCoefficientSimilarity(model);

UserSimilarity similarity = new EuclideanDistanceSimilarity(model);

UserSimilarity similarity = new GenericUserSimilarity(model);

UserSimilarity similarity = new SpearmanCorrelationSimilarity(model);
```

Generate three recommendations to the user with user_id 2 using all the above similarity measures and compare your outputs.