Data Description:

Amazon Reviews data (data source) The repository has several datasets. For this case study, we are using the Electronics dataset.

Domain:

E-commerce

Context:

Online E-commerce websites like Amazon, Flipkart uses different recommendation models to provide different suggestions to different users. Amazon currently uses item-to-item collaborative filtering, which scales to massive data sets and produces high-quality recommendations in real-time.

Attribute Information:

- userId : Every user identified with a unique id
- productId : Every product identified with a unique id
- Rating : Rating of the corresponding product by the corresponding user
- timestamp : Time of the rating (ignore this column for this exercise)

Learning Outcomes:

- Exploratory Data Analysis
- Creating a Recommendation system using real data
- Collaborative filtering

Objective:

Build a recommendation system to recommend products to customers based on the their previous ratings for other products.

Steps and tasks:

1. Read and explore the given dataset. (Rename column/add headers, plot histograms, find data characteristics) - (2.5 Marks)
2. Take a subset of the dataset to make it less sparse/ denser. (For example, keep the users only who has given 50 or more number of ratings) - (2.5 Marks)
3. Split the data randomly into train and test dataset. (For example, split it in 70/30 ratio) - (2.5 Marks)
4. Build Popularity Recommender model. - (20 Marks)
5. Build Collaborative Filtering model. - (20 Marks)
6. Evaluate both the models. (Once the model is trained on the training data, it can be used to compute the error (RMSE) on predictions made on the test data.) - (7.5 Marks)
7. Get top - K (K = 5) recommendations. Since our goal is to recommend new products for each user based on his/her habits, we will recommend 5 new products. - (7.5 Marks)
8. Summarise your insights. - (7.5 marks)

References:

- Recommeneder systems and its applications
- Use cases of Recommendation systems