

# Product Recommendation Systems

---

**Domain** - E-commerce

**Context** - Everyday a million products are being recommended to users based on popularity and other metrics on e-commerce websites. The most popular e-commerce website boosts average order value by 50%, increases revenues by 300%, and improves conversion. In addition to being a powerful tool for increasing revenues, product recommendations are so essential that customers now expect to see similar features on all other eCommerce sites.

**Data Description** -

Data columns- First three columns are **userId**, **productId**, and **ratings** and the fourth column is timestamp. You can discard the timestamp column as in this case you may not need to use it.

**Source** - Amazon Reviews data (<http://jmcauley.ucsd.edu/data/amazon/>) The repository has several datasets. For this case study, we are using the Electronics dataset.

**Learning Outcomes** -

- Exploratory Data Analysis
- Data Wrangling
- Build a Popularity recommender model
- Build Collaborative Filtering model

**Objective** - To make a recommendation system that recommends at least five(5) new products based on the user's habits.

## **Steps and tasks -**

1. Read and explore the given dataset. ( Rename column/add headers, plot histograms, find data characteristics) ( 3 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 ) -(5 Marks)
3. Build Popularity Recommender model. ( 15 marks)
4. Split the data randomly into a train and test dataset. ( For example, split it in 70/30 ratio) ( 2 marks)
5. Build Collaborative Filtering model. ( 20 marks)
6. Evaluate the above model. ( Once the model is trained on the training data, it can be used to compute the error (like RMSE) on predictions made on the test data.) You can also use a different method to evaluate the models. ( 5 marks)
7. Get top - K ( K = 5) recommendations. Since our goal is to recommend new products to each user based on his/her habits, we will recommend 5 new products. ( 10 marks)
8. Summarise your insights. ( 10 marks)

### **Please Note -**

- If you are facing any memory issue while working on this project, create a small subset (Let's say 10% of data) and work on it.
- If you are stuck at the model evaluation part of this project.

Please refer to below links -

1. <https://surprise.readthedocs.io/en/stable/accuracy.html>
2. <http://surpriselib.com/> - Getting started, example

## Metrics

predicted	actual
2.3	2
4.2	3
4.8	5
2.1	4
3.5	1
3.8	4

- MAE (mean absolute error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - p_i|$$

- RMSE (root mean square error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - p_i)^2}$$