

Recommendation System Document

➤ Why Recommendation System ?

Recommendation systems help the users to get personalized recommendations, helps users to take correct decisions in their online transactions, increase sales and redefine the users web browsing experience. Even if the buyer buys one item from the recommended than it is like profit to the company and also discounts are provided .

Ex – If someone buys Mobile Phone then it will recommend Charger, USB pin, Headphones etc.

1. Understanding the real world dataset .

The dataset consists of the reviews of fine food from amazon. The data span a period of more than 10 years including all 500,000 reviews up to October 2012. It includes product and user information, ratings and a plain text review. It also includes reviews from all other Amazon categories.

2. Download the Dataset.

Ex – Online web site, Kaggle etc.

3. Data Preprocessing .

Drop N.A values, removal of stop words case of NLP.

4. Featurization .

Create a data frame and select the specific columns like Product ID and Score for further recommendation, you can sort the values as per use.

5. Plot the records for better understanding of the data.

6. Calculate the mean rating for a Product and count of the product.

7. After calculation plot the mean and count of the ratings separately.

8. Now, analyse the correlation between the similar products or movies depending on the dataset.

➤ Why Coefficient of Correlation ?

Correlation will give you the exact dependency of the columns between each other. Shows the linear relationship between the two variables.

When Correlation is 1 Both columns are dependent on each other else there is no relation between them.

Here we can also use **Collaborative Filtering** approach for the recommendation aside from **Truncated SVD**.

➤ Truncated SVD

It is a factorization technique that factors a matrix in to three matrices U , Σ and V . SVD diagonalize a matrix in to special matrices that are easy to manipulate and to analyze as compared to PCA

9. Feature Reduction.

We use Truncated SVD to reduce the large no of features to convert them in to smaller subset of features, that useful for our prediction problem. It can be also called as Dimensionality reduction. You can also use **PCA** for the same.

10. Divide the data in to training and testing format and pass it to correlation matrix for finding the exact correlation and build the model for the same.

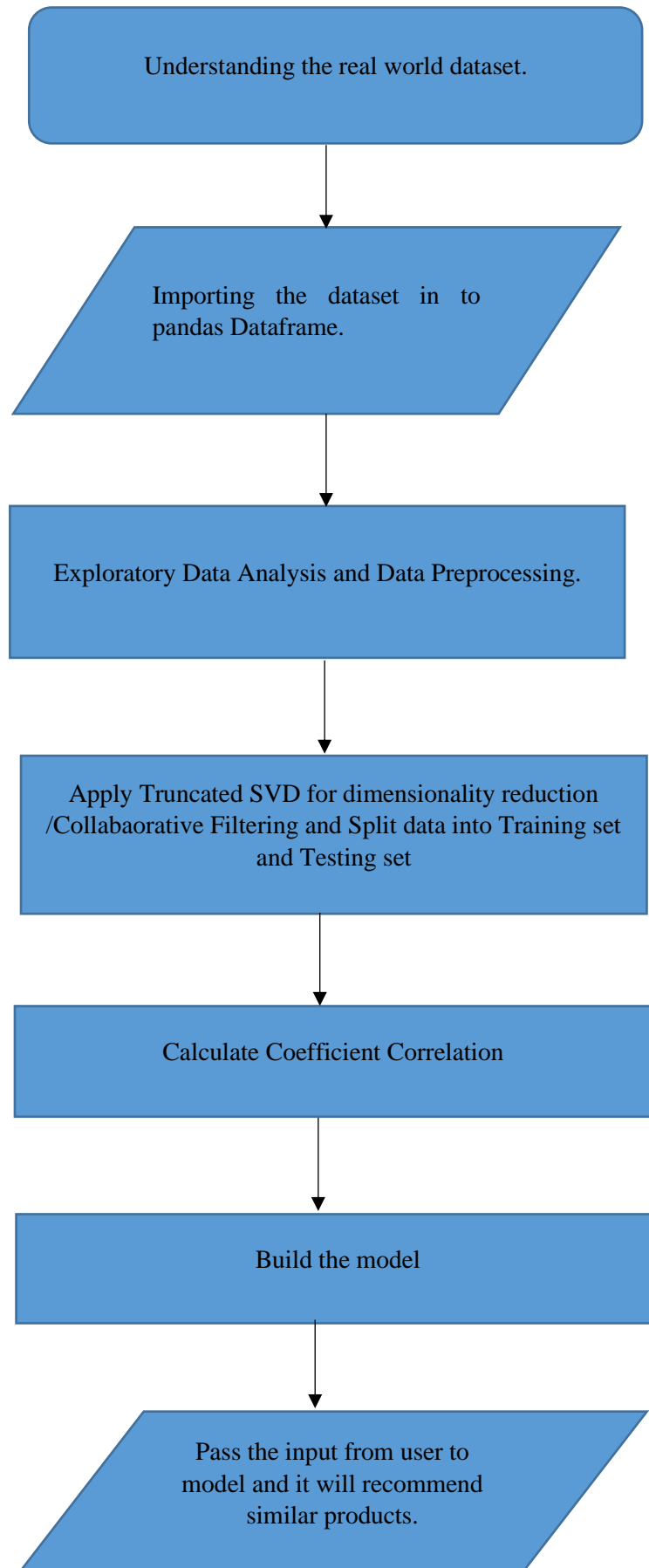
11. Accept any product from the user. Take the Product ID and pass the Product Id to the correlation matrix for the same and keep the probability of the product more than 0.50

12. Performing the calculation it will recommend the similar products to the user.

➤ **Confirmation of the product.**

If the person buys the process an online random OTP is developed with import random.random . By confirming the OTP we can process the transaction and confirm the delivery of the product.

Flow Chart



➤ **Conclusion**

For optimal classifier performance we need to be sure that our model does not have any large dimension and not too small. It may lead to underfitting and overfitting.