SJSU CMPE-256 Large Scale Analytics Spring 2020

Group 10 Project Proposal

- **Project Title**: Recommendation system for products on an e-commerce website

- **Group Members**:

Name	Student ID
Atharva Munshi	014502397
Kapil Mulchandani	014501396
Karnavee Kamdar	014518517
Sathya Sri Pasham	014552460

- **Data Set:**

  For the dataset, we are using Amazon's products dataset which consists of customer reviews, metadata, product categories and many more attributes which build up a 233.1 million record dataset with product data reviews from May 1996 – October 2018. This provides us with sufficiently large amount of data to better train the recommendation system and get an overall view of users' preference.

- **Project Idea**:

  As E-Commerce retail is growing day-by-day, customers are tending to order more products online rather than going to a store and buying products. As a part of a survey, results showed that 63.15% of people surveyed preferred online ordering of items (ranging from electronics to even daily need items such as groceries) over going to the market and buying from a shop. Now, in order to increase product sales as well as to increase the customer satisfaction, E-commerce websites are extensively using recommendation systems to provide better user experience as well as to increase the sales of their products.

  At present, Amazon, one of the leaders in E-commerce, uses item-item collaborative filtering which caters to large datasets and efficiently recommends products that user might be interested to buy in real time. Item-item filtering uses user's order and reviews history and maps them to similar items and then combines those similar items in a new recommendation list for that specific user. This provides more personalization for each user as the recommendations vary based on the user activities. So, by taking these points into consideration we decided to build and test a product recommender system by combining and comparing different recommendation system algorithms and evaluating their performances.

- **Steps**:

  We will use the following approaches to build and compare the performance of recommendation systems:

  - Popularity based recommendation system which is based on frequency counts of purchased items. It is a non-personalized recommender system as the recommendation might not be suitable for the customer.
  - Content based recommendation system to recommend items to customers based on the similarity of their highly rated previously purchased items. We will use Local Sensitive Hashing to quickly find the similar items.
  - User-User and Item-Item Collaborative Filtering with different similarity measures like Jaccard similarity, Cosine similarity and Pearson Correlation to predict product ratings.
  - Latent Factor Model to capture the similarity between user and items and to handle the scalability and sparsity issue created by Collaborative Filtering and use Singular Value Decomposition to achieve minimum RMSE.
  - For evaluating predictions of recommendation systems, we will use root-mean-square error (RMSE), find the precision at top N recommendations and plot the ROC curve to visualize the trade-off between the false positives and false negatives.

- **References:**

  - https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf
  - http://wwwconference.org/proceedings/www10/papers/pdf/p519.pdf
  - http://cs229.stanford.edu/proj2017/final-reports/5230053.pdf
  - https://patents.google.com/patent/US6266649B1/en
  - https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75