Name: Ankita Ramvir Singh Yadav; CWID: 10457726

Team Project Proposal (Team 12) – Movie Recommendation System

Recommendation system has introduced a new way for websites to communicate with their users. Traditionally websites provided a static experience wherein users used to search for products and buy them however, in this modern big data era, websites have started offering personalized and a richer experience to users by recommending them products or services based on their past purchases, searches, ratings, and other user behaviors. This system provides increase interaction between the user and the service provider.

The movie recommendation system shall help in predicting movies based on historical data of users and their interests. It shall find a match between the user and the movie and imputes the similarities between the user and the movie for the recommendation. This shall improve user engagement, provide personalized content, and deliver the right choice of movie to the right user, and the user trust of the website shall increase leading to an increased retention rate of users.

For building a movie recommendation system I shall be using MovieLens Dataset, a movie recommendation service available on grouplens. This dataset describes 5-star rating and free-text tagging activity from MovieLens. It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996, and September 24, 2018. The dataset was generated on September 26, 2018. Users were selected at random for inclusion. All selected users had rated at least 20 movies. No demographic information is included. Each user is represented by an id, and no other information is provided. The data are contained in the files links.csv, movies.csv, ratings.csv, and tags.csv.

For building the recommendation system, I shall be using the alternating least square matrix factorization method. This method shall help in recommending the top 5 movies based on the user's past behavior.

Reference:

[1] https://towardsdatascience.com/build-recommendation-system-with-pyspark-using-alternating-least-squares-als-matrix-factorisation-ebe1ad2e7679

[2] https://medium.com/analytics-vidhya/movie-recommendation-with-collaborative-filtering-in-pyspark-8385dccecfca