FINAL PROJECT – MOVIES RECOMMENDER SYSTEM

Saheli Kar Nathan Groom Shipra Ahuja

Department of Data Analytics, CUNY School of Professional Studies

saheli.kar@spsmail.cuny.edu nathan.groom@spsmail.cuny.edu shipra.ahuja@spsmail.cuny.edu

PROPOSED RECOMMENDER SYSTEM

We intend to build a collaborative/content based hybrid recommender system that will recommend movies to shoppers at online retailers such as Amazon.com. This recommender system will be delivered in the cloud using the Spark platform

BUSINESS GOALS

Our goal is to deliver online shoppers appropriate recommendations based on the movies they have previously searched for and rated. We believe that if the customers are satisfied with the recommended titles, sales will increase, thereby furthering the business goals of our online retailer. Similarly to such movie recommender systems as the one used by Netflix's streaming service, we intend for our business and audience's goals to align. That is, if enough users find our recommender system more than satisfactory, our business would expand due to an improved reputation, both online and through word of mouth.

TARGET AUDIENCE

Our specific target audience is the population that purchases DVDs and Blu Rays online but the recommender system we build could also be used for a streaming service such as Netflix or Hulu.

DATA SOURCE

Our project will use data from the movie ratings database CiaoDVD, available at the dvd.ciao.co.uk website. More specifically, we will use the datasets **movie_rating** and **review_rating**.

The movie_rating data includes over 72,000 movie ratings from over 17,000 users with the following columns: userID, movieID, movie-categoryID, reviewID, movieRating and reviewDate.

The review_ratings data includes ratings of reviews, which can be evaluated by users for their helpfulness, and has the following columns: userID, reviewID and review rating.

PROPOSED MODIFICATIONS TO THE ORIGINAL DATASETS

We intend to alter the dataset described above by adding both an age and location feature to each user in the movie_rating dataset. We will use these features to determine more precise recommendations based the tastes of other users of similar age and/or geographic location.