**Book Recommendation System using Spark**

**DSGA 1004 BIG DATA**

Hitesh Patel

New York University

hlp276@nyu.edu

Sumedha Rai

New York University

@nyu.edu

1. Introduction:

Recommender systems are one of the most successful and widespread application of machine learning technologies in business incorporated by companies like Amazon, Netflix, YouTube to provide recommendation of products to their user. A Recommender System predicts the likelihood that a user would prefer an item. Based on previous user interaction with the data source that the system takes the information from (besides the data from other users, or historical trends), the system can recommend an item to a user. For this project we built a book recommendation system using Sparks’ alternating least squares (ALS) method to learn latent factor representations for users and items and use those to make recommendations for those users.

1. Data

Dataset used for this project is taken from Goodreads.com (available on UCSD) which is social cataloging website that allows individuals to search freely its database of books, annotations, and reviews. For building a recommender system we have got access to 3 files i.e. (1) books\_id\_map.csv consisting the meta-data of the books, goodreads\_interactions.csv which consists of user-book interactions (3) users\_id\_map.csv which contains user metadata.

1. Implementation