**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 applications of machine learning technologies in business incorporated by companies like Amazon, Netflix, YouTube provides a recommendation of products to its 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

The dataset used for this project is taken from Goodreads.com (available on UCSD) which is a 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 of the meta-data of the 2.3M books, goodreads\_interactions.csv which consists of 223M user-book interactions (3) users\_id\_map.csv which contains 876K users metadata.

As a part of preprocessing the data we have removed all the users which have fewer than 10 user-book interaction an

1. Implementation