

# Final Project Planning Document

*Samriti Malhotra & Vishal Arora*

*July 4, 2019*

## Final Project :- Recommender system

Planning Document Find an interesting dataset and describe the system you plan to build out. If you would like to use one of the datasets you have already worked with, you should add a unique element or incorporate additional data. (i.e. explicit features you scrape from another source, like image analysis on movie posters). The overall goal, however, will be to produce quality recommendations by extracting insights from a large dataset.

### Team Members

The members working on the final project group are:

*Vishal Arora.*

*Samriti Malhotra.*

### Objective and Approach

The main objective of the project is to build a recommender system using a dataset in a cloud-hosted Spark distributed computing environment. To explore various recommender techniques like UBCF/IBCF etc.. which we have fairly covered during our Project 1 to Project 4 journey but as part of our final project we want use a large dataset and explore other techniques like ALS and spark MLlib in our recommender system.

### Data Set

We will use the Good Read data set and concentrate specifically on the any one of the genres. This data set is also accessible through site by Julian McAuley at UC San Diego and also can be accessed through the link provided directly from google. The books data are categorized on Genres , for our project we will stick to only one genre dataset. We decided to work this dataset as this gives us exposure to clean & transform the data set before we can apply the various techniques learned as part of Recommender class such as UBCF|IBCF|Matrix Factorization(SVD|ALS)|Random|Popular|Hybrid System , Evaluating various models and finding the accuracy and best model.

### Challenges

One of the section where we think we will have to put more effort is the database and distributed computing environment administration. Spark LLib will be a good exposure while building the recommender system.