Spark Version: 2.3.1

Commands:

Spark-submit --class ModelBasedCF Dhiraj_Ramnani_hw2.jar <rating file path><testing file path>

Spark-submit --class ItemBasedCF Dhiraj_Ramnani_hw2.jar <rating file path><testing file path>

Item-based Collaborative Filtering (Task 2)

>= 0 and < 1: 29518

>= 1 and < 2: 12317

>= 2 and < 3: 2735

>= 3 and < 4: 581

>= 4: 85

RMSE: 1.0923990750778807

Time: 382 sec

Model-based Collaborative Filtering (Task 3)

>=0 and <1: 29245

>=1 and <2: 12543

>=2 and <3: 2286

>=3 and <4: 338

>=4: 48

RMSE: 1.075678442144526

Time: 19 sec

Improvements

- To start with, I performed Item-Item Collaborative Filtering.
- To boost the performance of the system, I employed the nearest neighbor technique and obtained top 20 most similar items for every item.