INF 553 – Minh Tran

Please include the following content in your description file:

1. Mention the Spark version and Python version

   Spark-2.3.3 and Python 3.6

2. Same baseline table as mentioned in task 2 to record your accuracy and run time of programs in task 2:

   Model-based CF

RMSE	1.101
Time(sec)	48

   User-based CF

RMSE	1.047
Time(sec)	122

   Item-based CF

RMSE	1.032
Time(sec)	128

3. If you make any improvement in your recommendation system, please also describe it in your description file.

Using LSH-based results from task 1 reduces the computational time compared to the item-based approach in task 2 part 3. This is because the LSH helps reducing the number of pair of items to calculate the distance between. LSH hashes similar input items into the same "buckets" with high probability. Then we can calculate the distance between items of same buckets to reduce the computation time significantly.