

STEPS TO FOLLOW:

Spark 2.3.1

Scala 2.11

Python 2.7

TASK1

- To run the code which is in the Jar file for task1 write the below command:
 - `./spark-submit --class Task1 /Path/of/your/Jar/File /Path/of/your/rating_file /Path/of/your/testing_file`
 - My Jar File name is Krish_Mehta_hw2.jar so the path will be as `/Path/where/your/jar/file/ Krish_Mehta_hw2.jar`
 - E.g. `spark-submit --class Task1 "/Users/krishmehta/Desktop/Krish_Mehta_hw2.jar" "/Users/krishmehta/Desktop/DataMining/Krish_Mehta_HW2/hw2/Data/train_review.csv" "/Users/krishmehta/Desktop/DataMining/Krish_Mehta_HW2/hw2/Data/test_review.csv"`
 - Please note at the end it will take some time to get the output file, so the code might stop for few seconds during that process.

TASK2

- Task 2 is done in python.
 - `spark-submit "/Path/to/Folder/Krish_Mehta_UserBasedCF" "/Train path" "/Test/path"`
 - E.g. `spark-submit "/Users/krishmehta/Desktop/DataMining/Krish_Mehta_HW2/Krish_Mehta_UserBasedCF.py" "/Users/krishmehta/Desktop/DataMining/Krish_Mehta_HW2/hw2/Data/train_review.csv"`

"/Users/krishmehta/Desktop/DataMining/Krish_Mehta_HW2/hw2/D
ata/test_review.csv"

Table for task1:

≥ 0 and ≤ 1 : 29011

≥ 1 and ≤ 2 : 12854

≥ 3 and < 3 : 2246

≥ 4 : 38

RMSE: 1.070291359090508

Time: 18

Table for task2:

≥ 0 and ≤ 1 : 29071

≥ 1 and ≤ 2 : 12706

≥ 3 and < 3 : 2850

≥ 4 : 63

RMSE: 1.09704910746

TIME: 100.032000065