Versions-

Python: 2.7

Spark: 2.2.1

Task1 - LSH

bin/spark-submit Ankitha_Radhakrishna_task1_Jaccard.py <rating file path>

recall : 0.901960529158
precision : 1.0
running time : 230 seconds

Task2 –

Model Based CF

bin/spark-submit Ankitha_Radhakrishna_task2_ModelBasedCF.py <rating file path> <testing file path>

** The values of rank, iterations and lambda have to be updated in variables – rank, numIterations, lambdaVal

ml-latest-small/ratings.csv

rank : 5

iterations : 10

lambda : 0.1

ml-20m/ratings.csv

rank : 10

iterations : 10

lambda : 0.1

User Based CF

bin/spark-submit Ankitha_Radhakrishna_task2_UserBasedCF.py <rating file path> <testing file path>

	TASK1		TASK2
	SMALL	LARGE	SMALL
>=0 AND <1	13053	3238125	15101
>=1 AND <2	4224	701444	4204
>=2 AND <3	1065	92837	802
>=3 AND <4	287	12349	136
>=4	104	1576	13
RMSE	1.13628642252	0. 830283334596	0.954509141811
TIME	12.3378210068	1506	33.25