Prerequisites to run code:

1. pySpark
2. Python3
3. Movielens dataset

To run the code:

1. Files ratings.csv and movies.csv should be in the same directory.
2. Run -> spark-submit LSHJaccard.py ratings.csv

Note: The model is implemented for 25M datasets. Movielens 100K dataset works quicker.