

**Spark Version: 2.2.1 Python Version: 2.7.15**

### **Jaccard Similarity based LSH**

1. Go to command prompt on Windows OS.
2. In command line change directory to where spark is installed and then run command-  
**"bin\spark-submit Snehal\_Shirgure\_task1\_Jaccard.py <ratings file>"**

And hit enter

3. Output saved to –

**Snehal\_Shirgure\_SimilarMovies\_Jaccard.txt**

in ascending order with pair similarity  $\geq 0.5$ .

**Precision: 1.0**

**Recall: 0.982997295838**

**Time taken: 196.178999901 sec**