

Project 1: Mining Frequent Itemset

By: Dhvani Patel

Id: 110028937

In this project, I have implemented two efficient algorithms A-priori and PCY to find frequent pairs in given dataset.

Implementation:

- Each algorithm is implemented separately as python classes.
- There are in total four algorithm implemented to find frequent pairs efficiently. They are A-priori, PCY, Multistage version of PCY, and Multihash version of PCY.

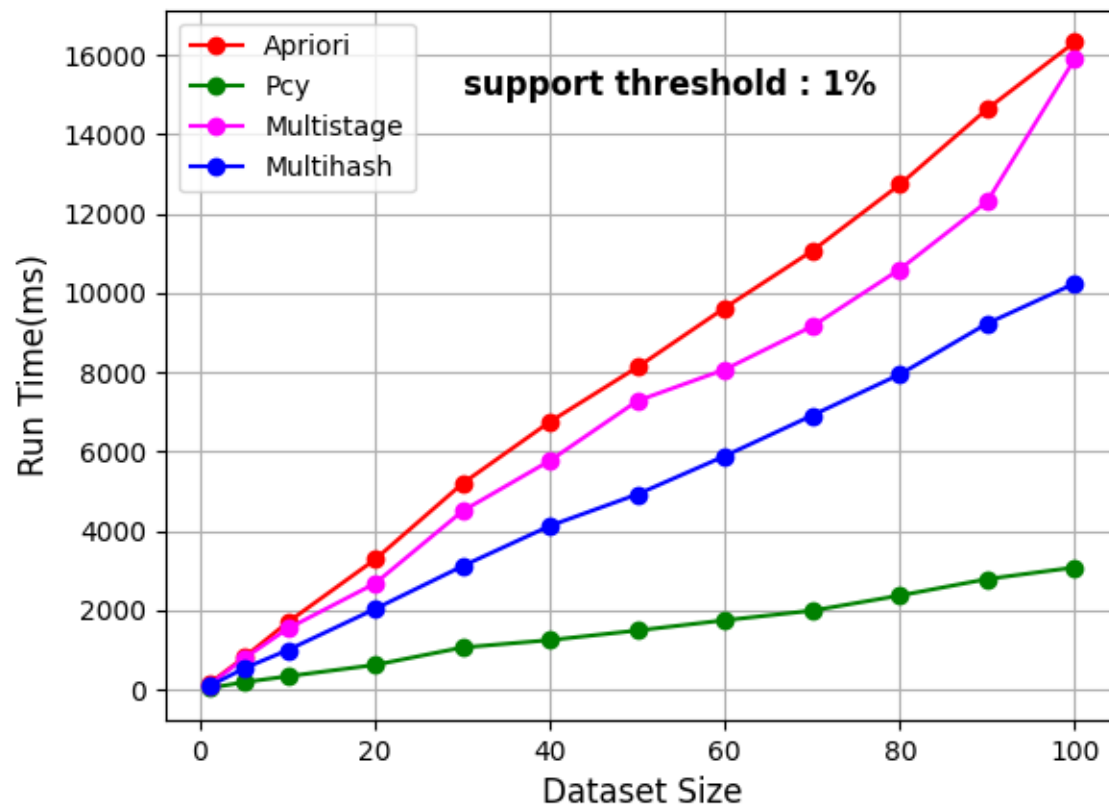
Testing:

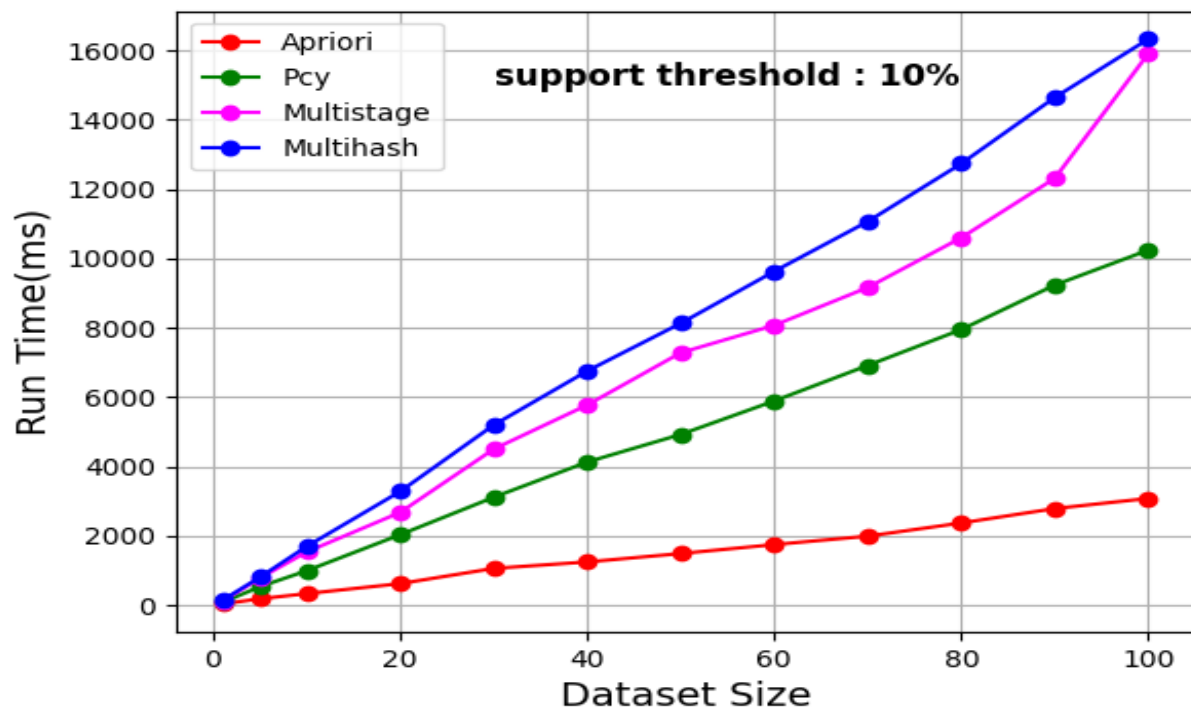
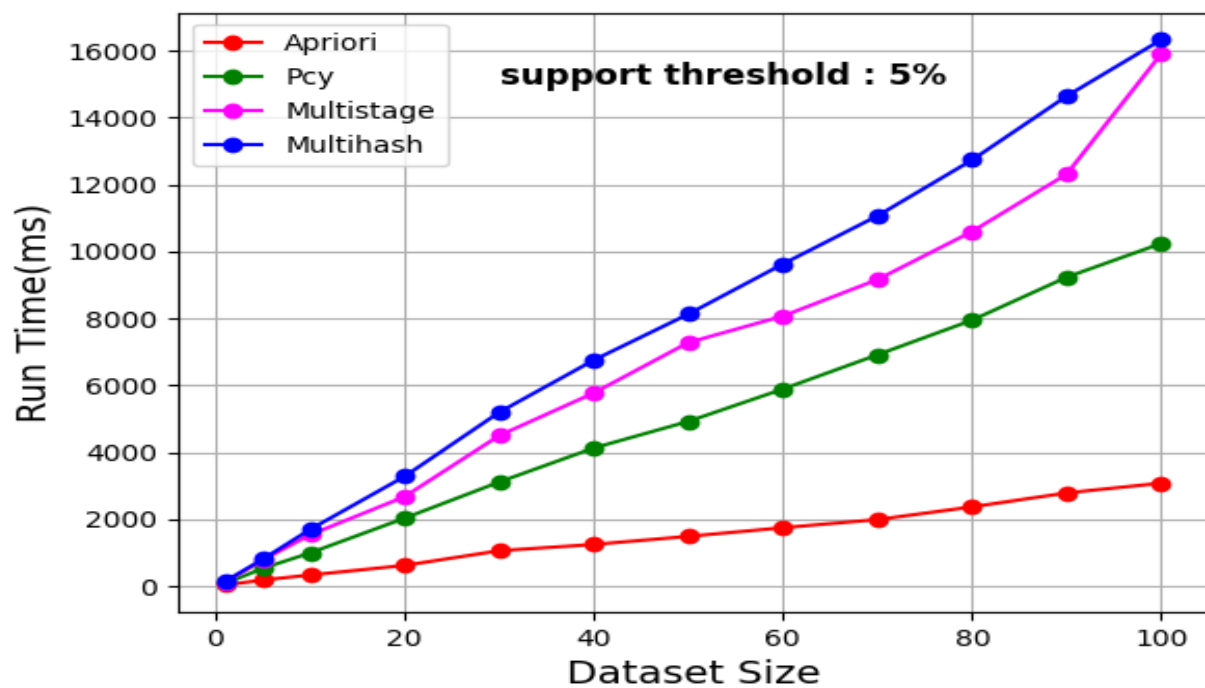
Command to run each algorithm:

- `python <algorithm_filename> -f <dataset_file> -d <data_chunks_percent> -t<support_percent>`

Experimental results:

- Line Chart for 1%, 5% and 10% threshold:





Computer Specification:

- **Operating Systems** – Windows 10
- **Ram** – 16 GB
- **Processor** – Intel-i7-7700HQ
- **CPU** – 2.8GH

Observations:

- From the results, it can be observed that candidate pairs and frequent pairs in PCY < candidate pairs and frequent pairs in apriori for all data chunks and threshold.

TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE

PS C:\Users\Dhawani\Documents\Comp 4250\Frequent_Itemset_Prediction> python pcy.py

Enter the data file name: retail.txt

Total candidate pairs: 2291

Total frequent pairs: 87

Total false positives: 2204

runtime: 1242.98

Total candidate pairs: 1957

Total frequent pairs: 67

Total false positives: 1890

runtime: 6317.27

Total candidate pairs: 1941

Total frequent pairs: 58

Total false positives: 1883

runtime: 16588.84

Total candidate pairs: 10

Total frequent pairs: 9

Total false positives: 1

runtime: 1392.07

Total candidate pairs: 10

Total frequent pairs: 9

Total false positives: 1

runtime: 6678.02

Total candidate pairs: 14

Total frequent pairs: 7

Total false positives: 7

runtime: 14410.63

Total candidate pairs: 8

Total frequent pairs: 4

Total false positives: 4

runtime: 2486.9

Total candidate pairs: 8

Total frequent pairs: 4

Total false positives: 4

runtime: 6886.31

Total candidate pairs: 7

Total frequent pairs: 4

Total false positives: 3

```
PS C:\Users\Dhawani\Documents\Comp 4250\Frequent_Itemset_Prediction> python apriori.py
Enter the data file name: retail.txt
Total items: 8346
Total candidate pairs: 2647
Total frequent pairs: 87
Total false positives: 2560
runtime: 478.2
-----
Total items: 13958
Total candidate pairs: 2270
Total frequent pairs: 67
Total false positives: 2203
runtime: 2076.61
-----
Total items: 16470
Total candidate pairs: 2404
Total frequent pairs: 58
Total false positives: 2346
runtime: 4684.9
-----
Total items: 8346
Total candidate pairs: 10
Total frequent pairs: 9
Total false positives: 1
runtime: 178.28
-----
Total items: 13958
Total candidate pairs: 10
Total frequent pairs: 9
Total false positives: 1
runtime: 921.55
-----
Total items: 16470
Total candidate pairs: 15
Total frequent pairs: 7
Total false positives: 8
runtime: 1913.76
-----
Total items: 8346
Total candidate pairs: 10
Total frequent pairs: 4
Total false positives: 6
runtime: 173.64
-----
Total items: 13958
```