Lab Assignment 01:
Apriori vs FP growth performance analysis

Submitted by:

Sihan Tawsik

Roll: 05

Batch: 23

Submitted to:

Dr. Chowdhury Farhan Ahmed

Professor,

Dept of Computer Science and Engineering,

University of Dhaka

Introduction

For analyzing the performance of the algorithms, Kosarak, T10I4D100K, chess, retail, and the mushroom dataset is used with five levels of threshold.
Used module: memory profiler (for registering peak memory usage.

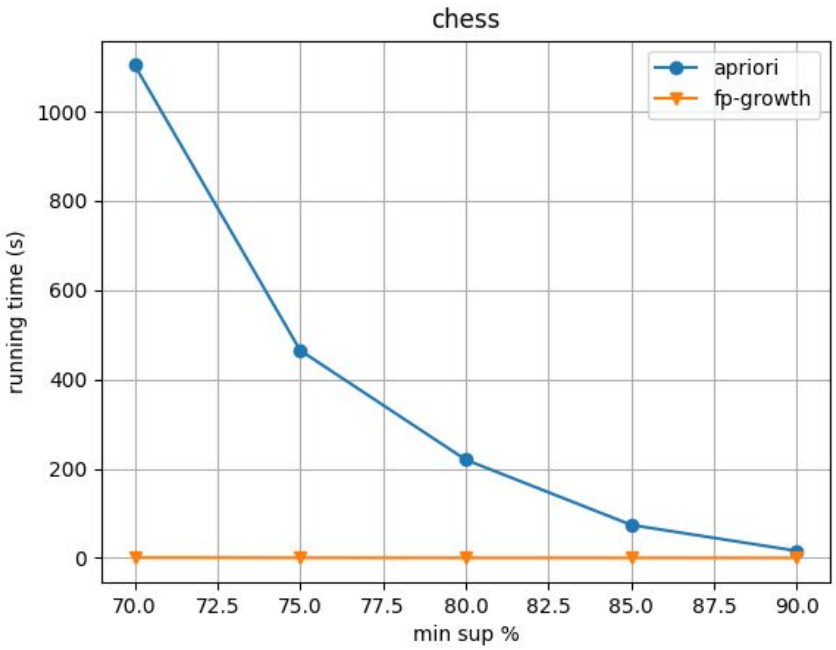
Notes

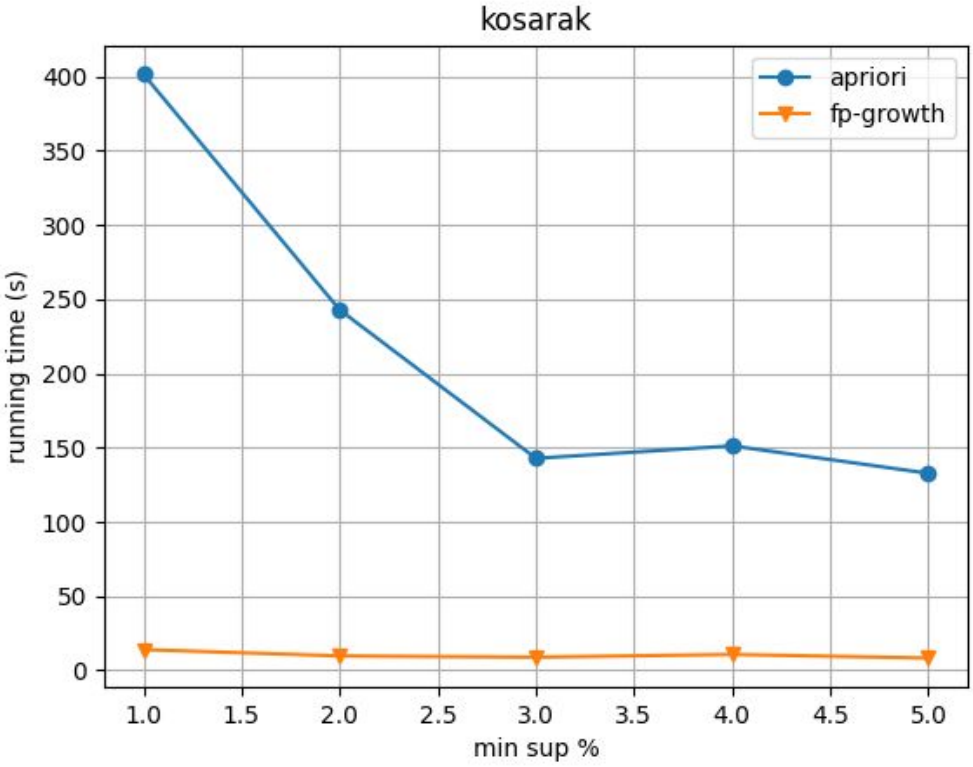
The database was preprocessed before using the algorithms and so, does not include the preprocessing time.

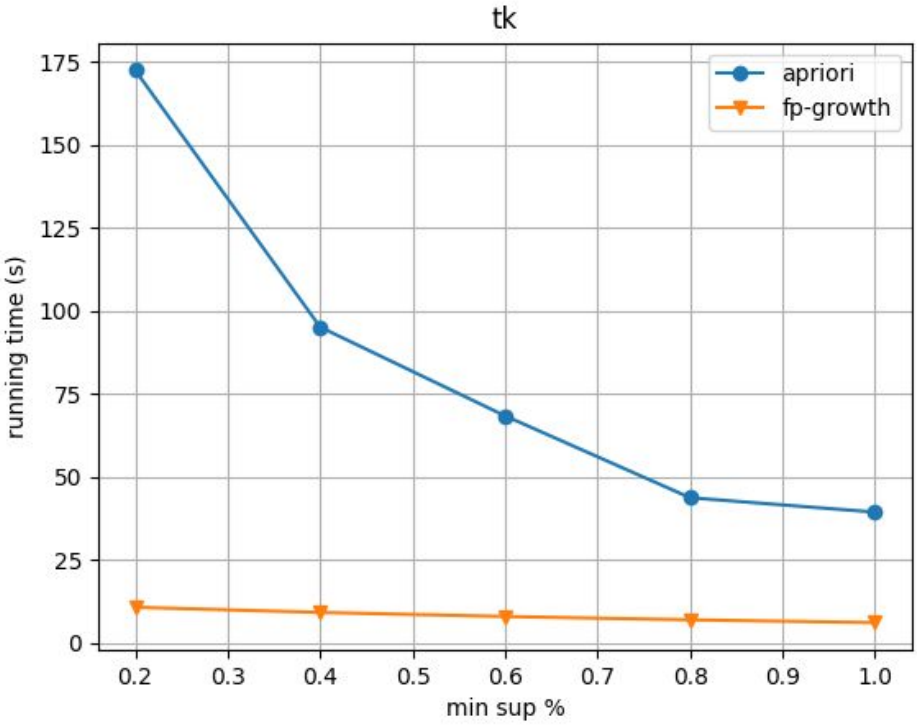
Result Analysis

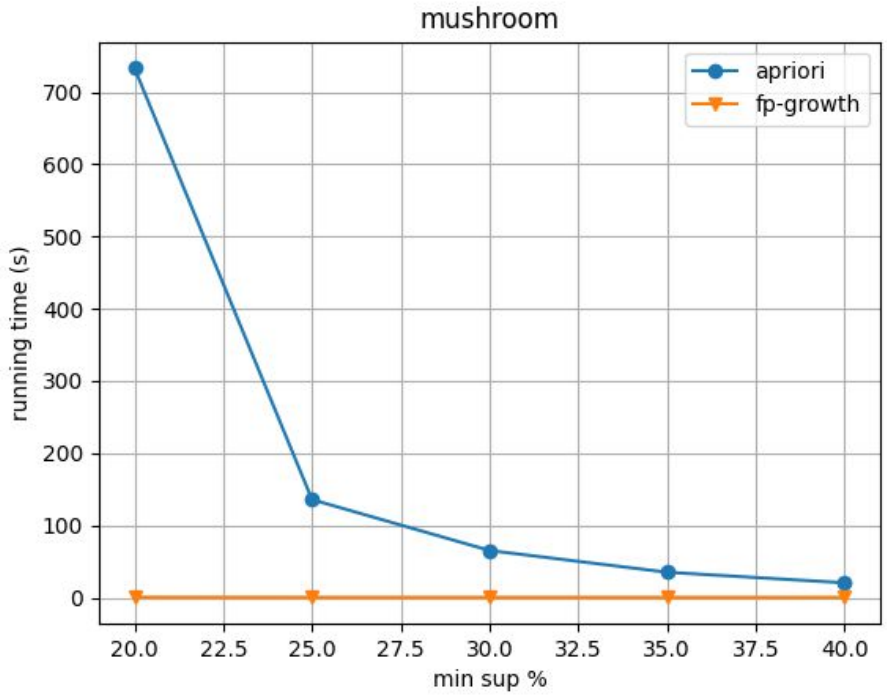
As we can see that the Apriori algorithm takes more time than the FP Growth algorithm for all thresholds in all the datasets. This is more evident in larger datasets. In comparison, the fp growth takes up more memory due to recursion stacks. Only the chess dataset is an exception. Here the fp growth takes up less space.

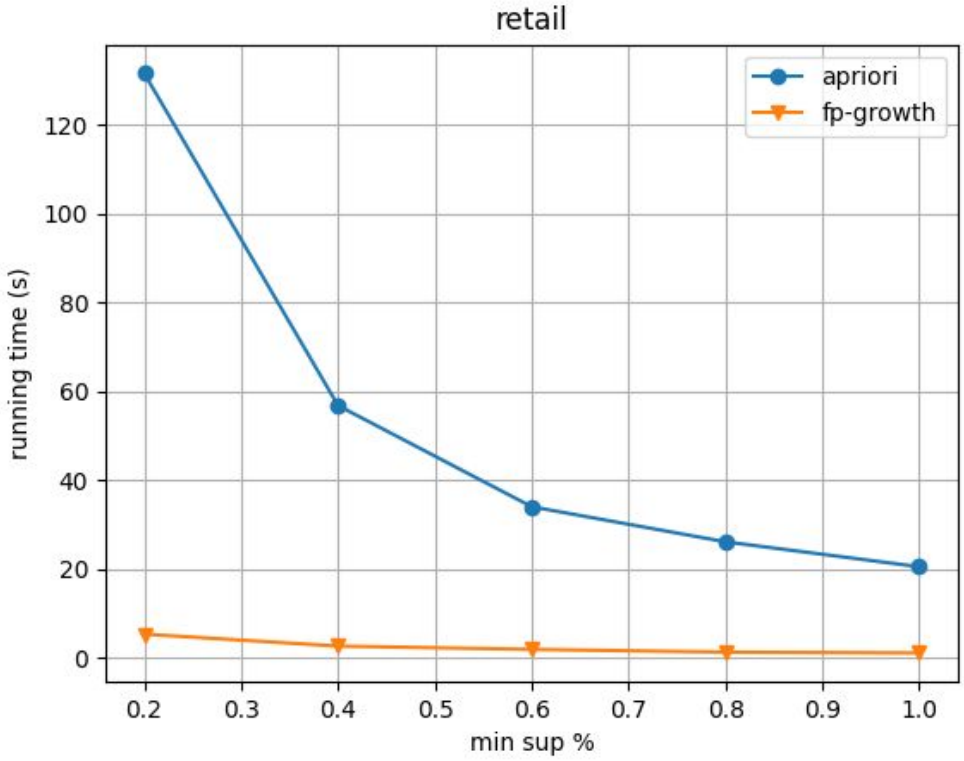
Runtime comparison



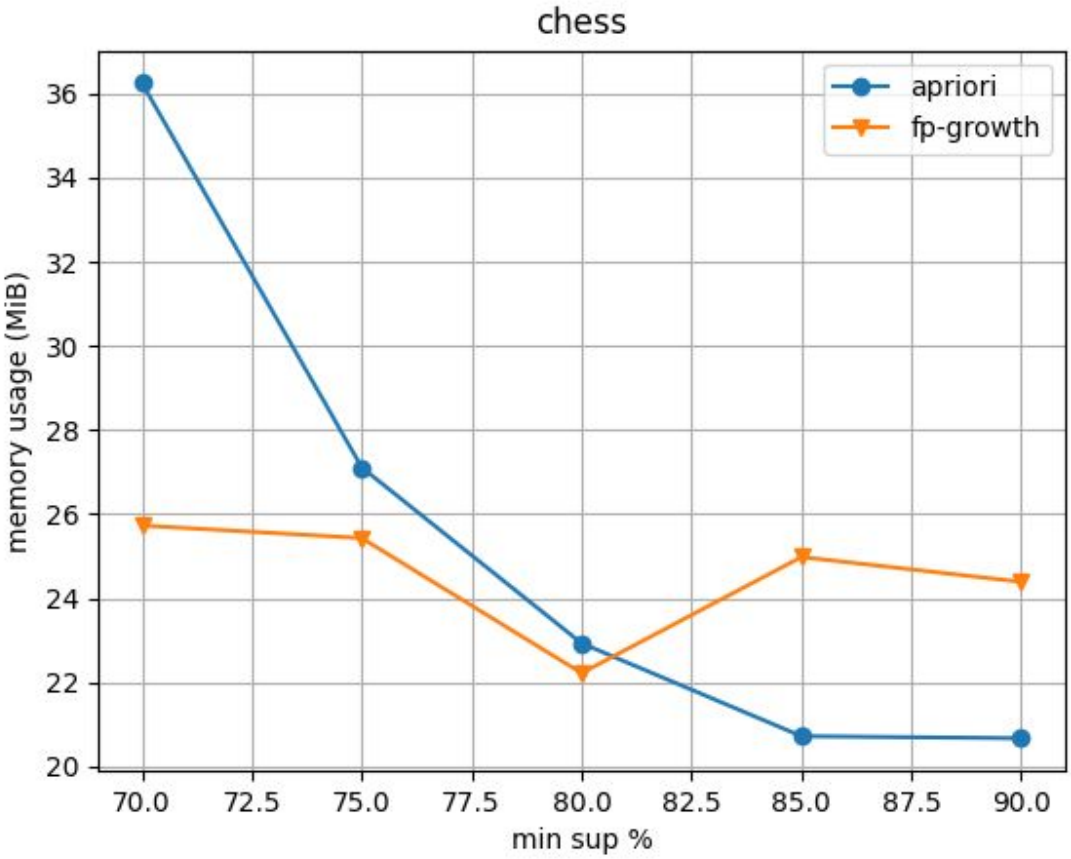


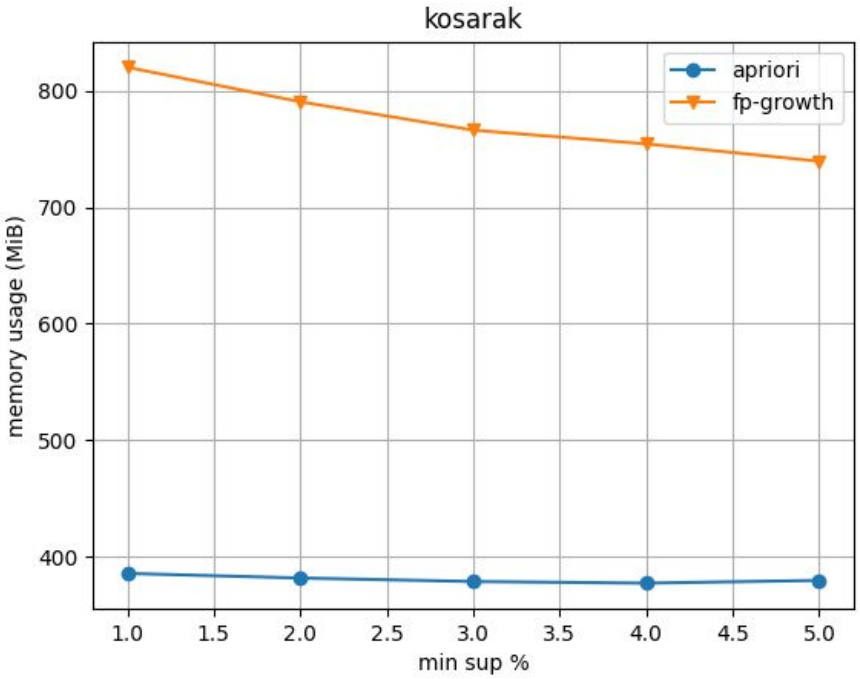

tk
apriori
fp-growth
running time (s)
min sup %
0
25
50
75
100
125
150
175
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

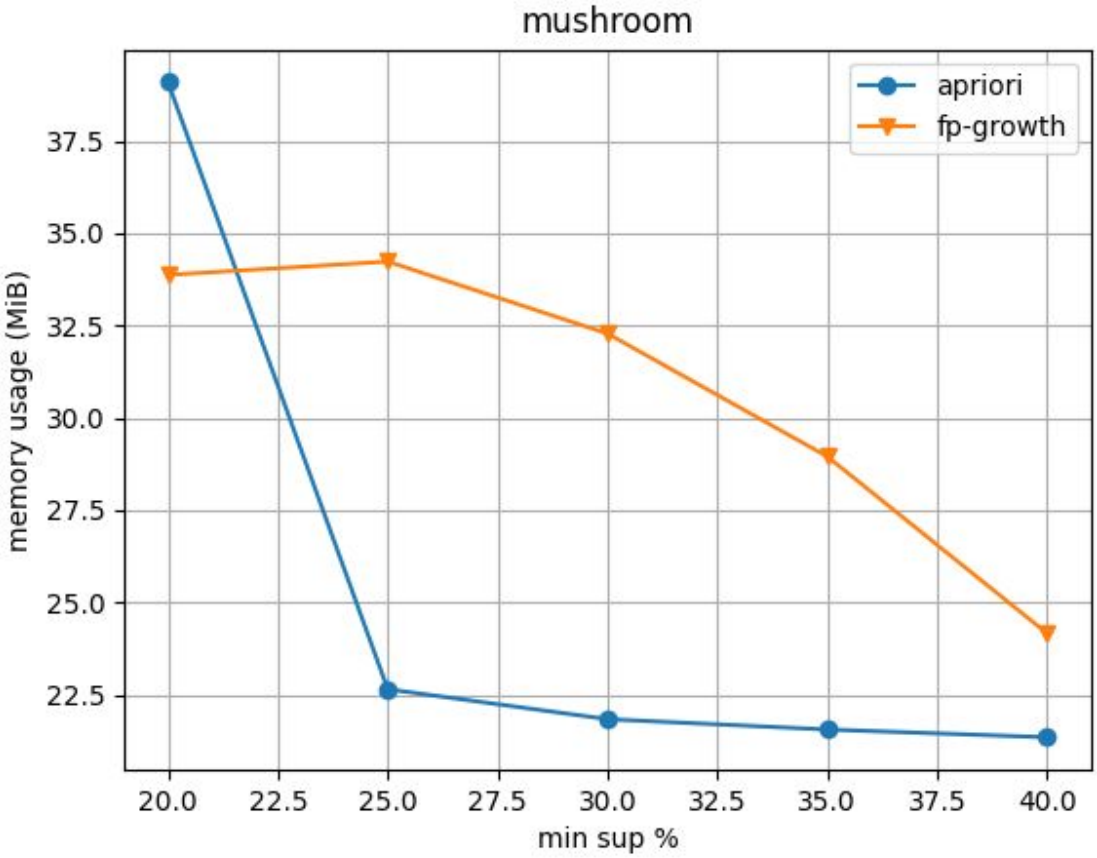

mushroom
apriori
fp-growth
running time (s)
min sup %
0
100
200
300
400
500
600
700
20.0
22.5
25.0
27.5
30.0
32.5
35.0
37.5
40.0

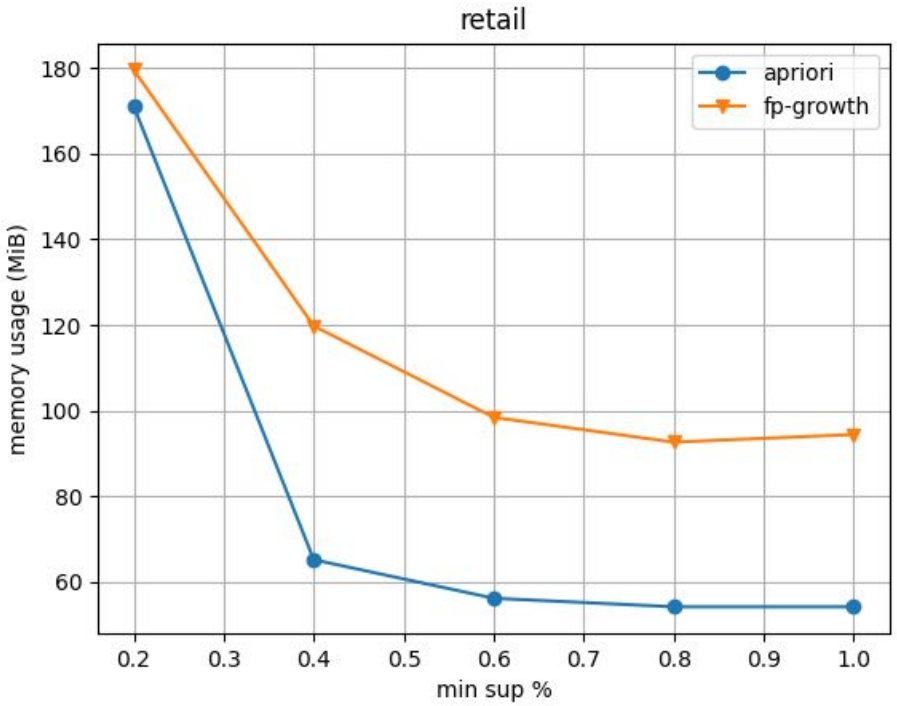

retail
apriori
fp-growth
running time (s)
0
20
40
60
80
100
120
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
min sup %

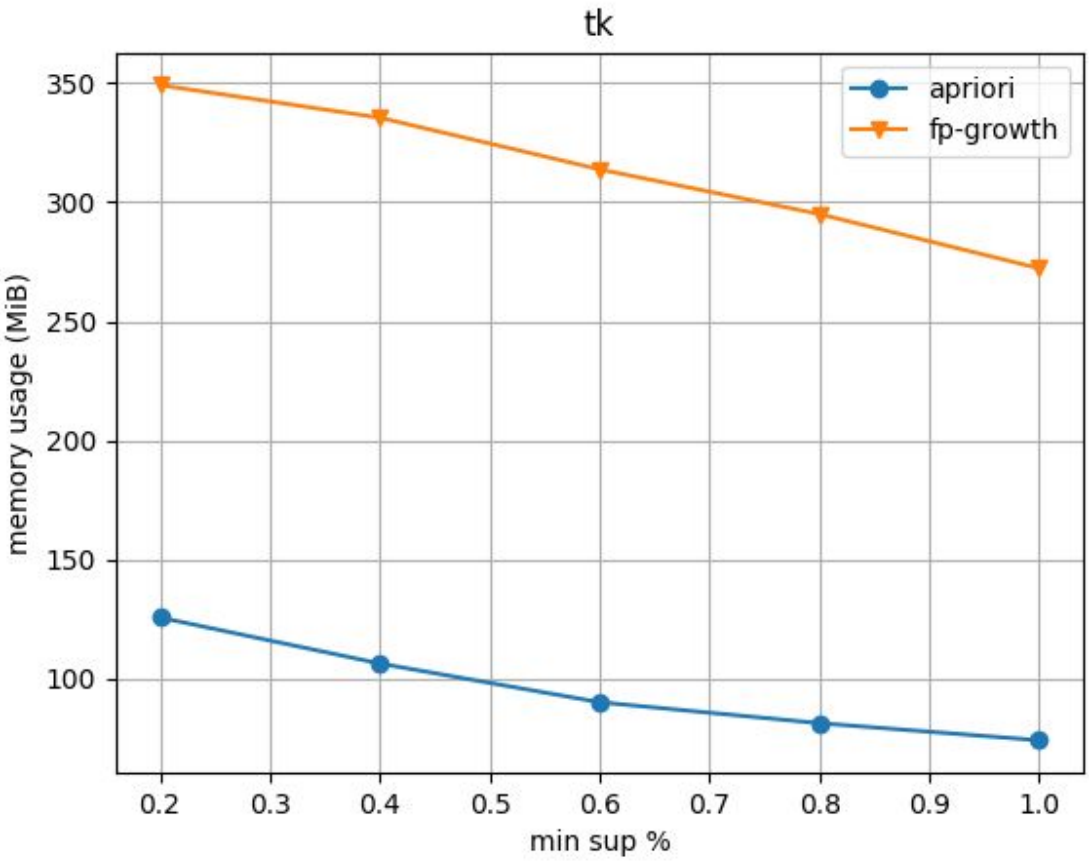
Memory Usage




kosarak
apriori
fp-growth
memory usage (MiB)
800
700
600
500
400
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
min sup %


mushroom
apriori
fp-growth
memory usage (MiB)
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
22.5
25.0
27.5
30.0
32.5
35.0
37.5
40.0
min sup %


retail
apriori
fp-growth
memory usage (MiB)
180
160
140
120
100
80
60
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
min sup %


tk
apriori
fp-growth
memory usage (MiB)
350
300
250
200
150
100
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
min sup %