CSE-4225 Fall 2020 Data Mining Lab

Lab Assignment - 1

Assignment 1: Implementation and Analysis of Apriori and FP-Growth algorithms

Submitted by:
Mohd. Sayemul Haque
Roll: 53

Submitted To:

Dr. Chowdhury Farhan Ahmed

Professor

Department of CSE,
University of Dhaka

Implementation:

- Method: Trie Based Structure On Apriori, Fp Growth
- Language: Python
- Libraries Used: *tracemalloc, copy, time, os, errno*
- Processor: Intel Core i5-7300HQ, 8GB Ram, 64bit Operating System(Windows 10)

Datasets:

- Large : Accidents
- Sparse: Retail, Kosarak
- Dense: Chess, Mushroom

In case of dense graphs Apriori tends to take more memory and space as the minimum support(minsup) is decreased. This is because decreament of minsup increases the number of patterns. And Apriori scans the whole dataset for every length of patterns generated. Whereas Fp growth reads the whole dataset only a single time and uses a divide and conquer method to solve the problem.

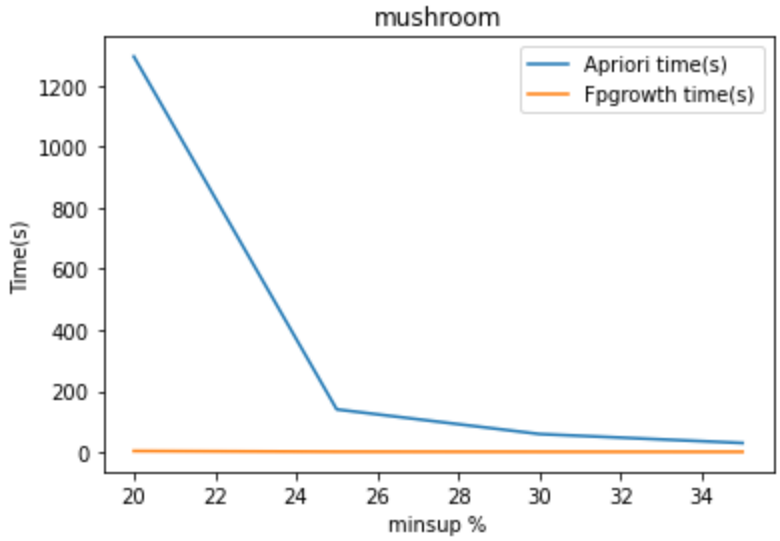
The runtime for sparse dataset is almost similar for both algorithm. But Fp growth needs more memory as it has to build a header table for every conditional pattern tree.

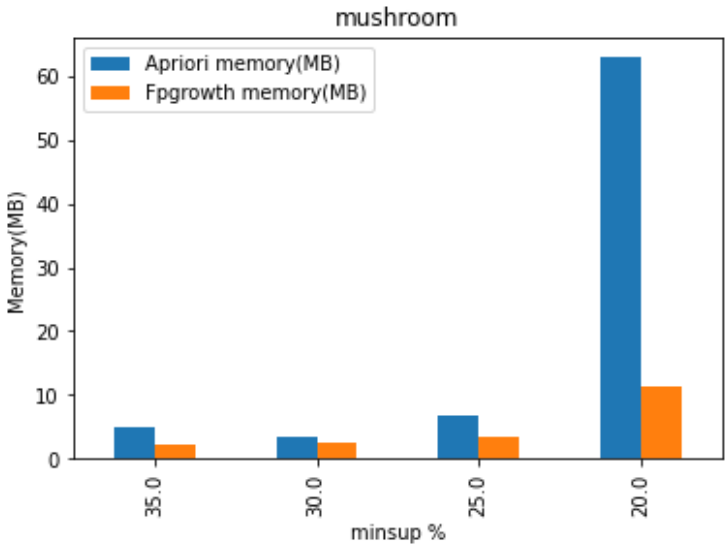
Fp growth's runtime is better than Apriori in case of large datasets like Accidents. But the memory consumption is larger than Apriori as usual.

The comparison of the performences of FastApriori and FPGrowth algorithms with respect to time and memory is given below.

1. **Mushroom Dataset:**

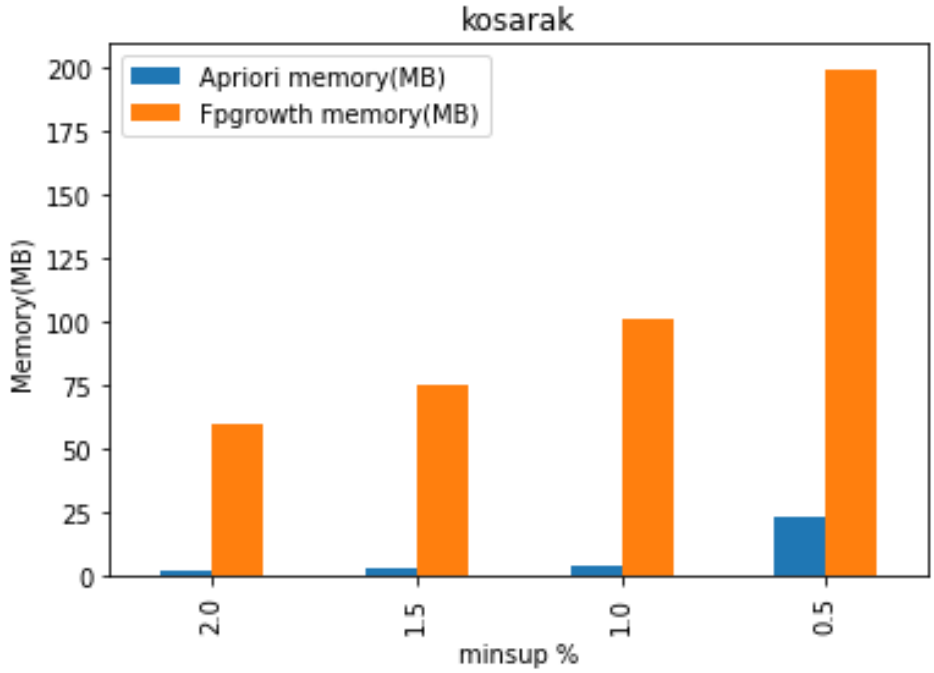
Support	FastApriori		FPGrowth	
	Time(S)	**Memory(MB)**	**Time(S)**	**Memory(MB)**
35	29.7999	4.911668	0.8697	2.28204
30	59.0705	3.373821	0.9375	2.698249
25	139.7329	6.782377	1.2676	3.477877
20	1295.072	62.81261	3.6143	11.28331

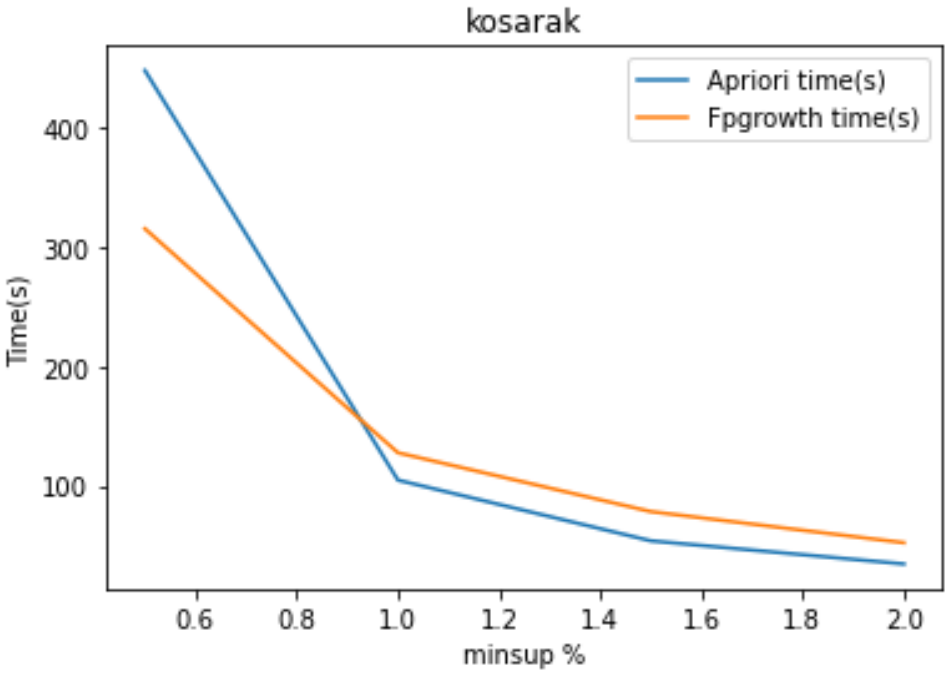




2. KOSARAK Dataset:

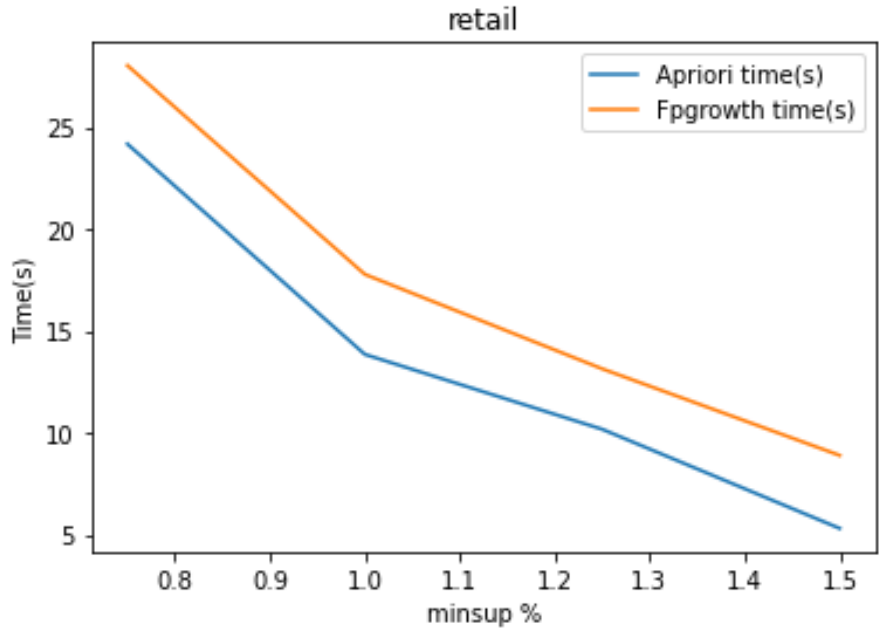
Support	FastApriori		FPGrowth	
	Time(S)	**Memory(MB)**	**Time(S)**	**Memory(MB)**
2	34.5435	2.057109	52.3013	59.8332
1.5	53.883	2.539661	78.3437	75.05755
1	104.6837	4.059005	127.5817	101.2485
0.5	447.3773	23.03505	314.9841	199.3641

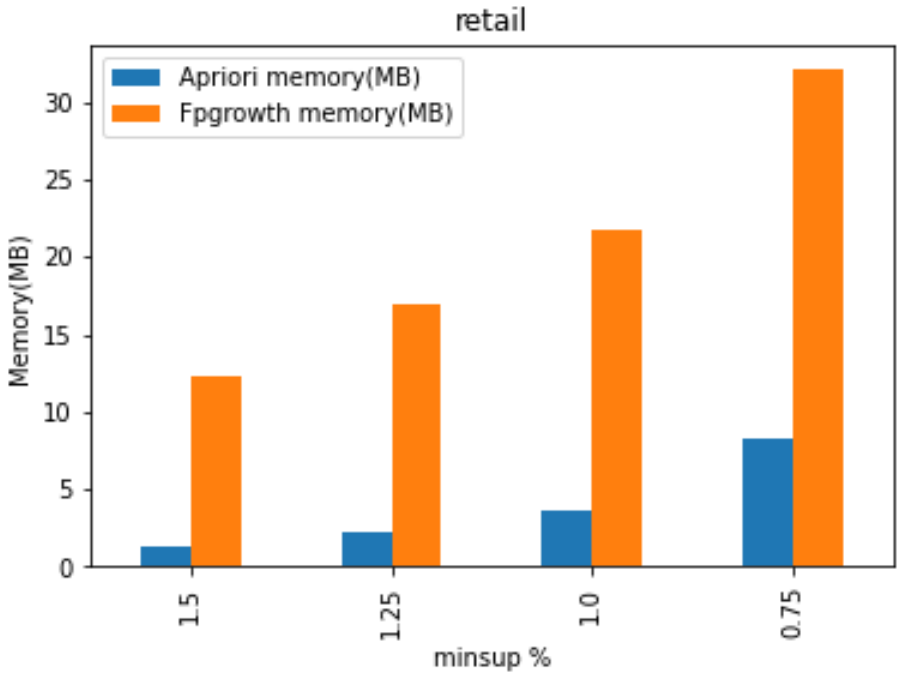




3. Retail Dataset:

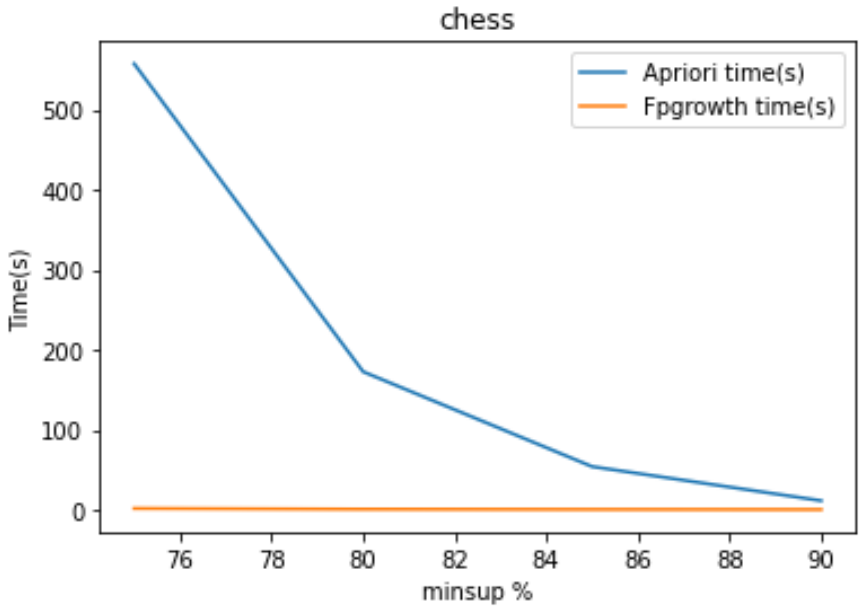
Support	FastApriori		FPGrowth	
	Time(S)	**Memory(MB)**	**Time(S)**	**Memory(MB)**
1.5	5.3109	1.364689	8.8938	12.35905
1.25	10.1745	2.279317	13.1404	16.89675
1	13.8496	3.582913	17.7767	21.80526
0.75	24.1729	8.202833	28.0259	32.07018

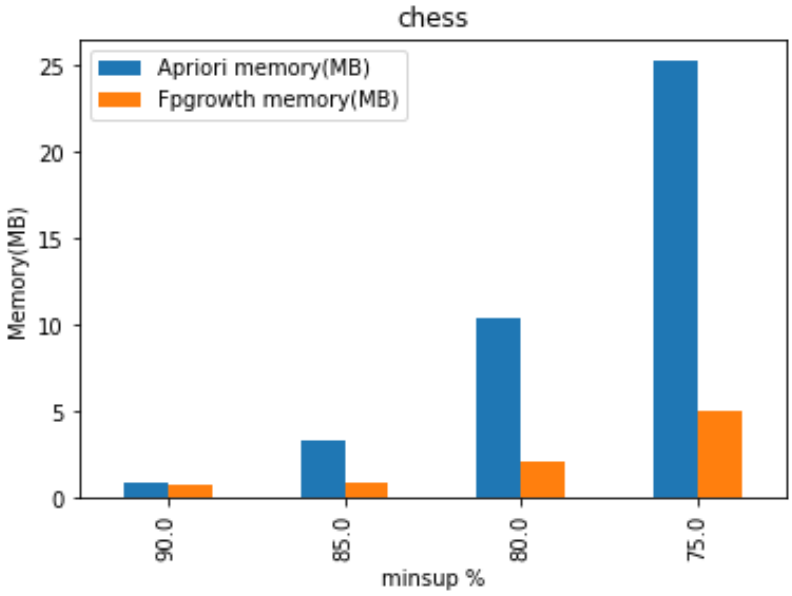




4. Chess Dataset:

Support	FastApriori		FPGrowth	
	Time(S)	**Memory(MB)**	**Time(S)**	**Memory(MB)**
90	11.3108	0.921385	0.2314	0.814376
85	53.9281	3.346877	0.3669	0.853177
80	172.5347	10.37275	0.6812	2.108409
75	557.5314	25.24292	1.6108	5.079177





5. **Accidents Dataset:**

Support	FastApriori		FPGrowth	
	Time(S)	**Memory(MB)**	**Time(S)**	**Memory(MB)**
90	42.8517	0.239565	13.8755	46.10311
85	111.6989	0.323061	16.7157	47.05286
80	260.498	0.476365	20.2983	67.06005
75	622.8345	0.768413	23.9777	69.57206