

UNIVERSITY OF DHAKA
Department of Computer Science and Engineering

**CSE-4255 : Introduction to Data Mining and
Warehousing Lab**

Lab Assignment: Implementation and time-memory usage
comparison of Apriori and FP-Growth Algorithms on different
datasets

Submitted On:

April 30, 2025

Submitted To:

Dr. Chowdhury Farhan Ahmed

Professor
Department of Computer Science & Engineering
University of Dhaka

Md. Mahmudur Rahman

Assistant Professor
Department of Computer Science & Engineering
University of Dhaka

Submitted By:

Himel Chandra Roy

Roll No : 13

Registration No : 2020915623

Contents

1	Introduction	2
2	Results	2
2.1	Execution Time Comparison	2
2.2	Memory Consumption Comparison	2
3	Result Figures	3
3.1	Accidents	3
3.2	Connect	4
3.3	Mushroom	5
3.4	Retail	6
3.5	Kosarak	7
3.6	Chess	8
3.7	Pumsb	9
4	Result analysis	10

1 Introduction

Frequent pattern mining is a fundamental task in data mining, aiming to discover recurring relationships among items in large transactional datasets. Two of the most widely used algorithms for this purpose are the Apriori algorithm and the FP-Growth (Frequent Pattern Growth) algorithm. **Apriori** relies on a breadth-first search strategy and generates candidate itemsets, which can become computationally expensive with large or dense datasets. In contrast, **FP-Growth** employs a depth-first strategy and uses a compact data structure known as the FP-tree to mine frequent patterns without candidate generation, often resulting in improved performance.

This report presents a comparative analysis of the Apriori and FP-Growth algorithms across various datasets. The primary focus is on evaluating their execution time and memory consumption under different data characteristics, such as transaction size, itemset density, and dataset scale. Through empirical experiments, this study aims to highlight the strengths and limitations of each algorithm and provide guidance on their suitability for different data mining scenarios.

2 Results

To evaluate the performance of the Apriori and FP-Growth algorithms, experiments were conducted on multiple benchmark datasets with varying characteristics such as transaction volume, average transaction length, and itemset density. The key performance metrics analyzed were execution time and memory consumption, measured for each algorithm under identical support thresholds and computational environments.

2.1 Execution Time Comparison

Across all datasets, the FP-Growth algorithm consistently outperformed Apriori in terms of execution time. For smaller datasets with low itemset density, the difference was minimal; however, as dataset size and density increased, the performance gap widened significantly. In dense datasets, Apriori's execution time increased exponentially due to the costly candidate generation step, while FP-Growth maintained relatively stable performance by avoiding candidate generation through its FP-tree structure.

2.2 Memory Consumption Comparison

In terms of memory usage, the results were more nuanced. Apriori, despite its slower execution, showed lower memory usage in sparse datasets due to its simple data structures. However, as dataset complexity increased, Apriori's memory usage rose sharply due to the exponential growth in candidate itemsets. FP-Growth, on the other hand, required more initial memory to construct the FP-tree but used it efficiently during mining, resulting in better scalability.

3 Result Figures

3.1 Accidents

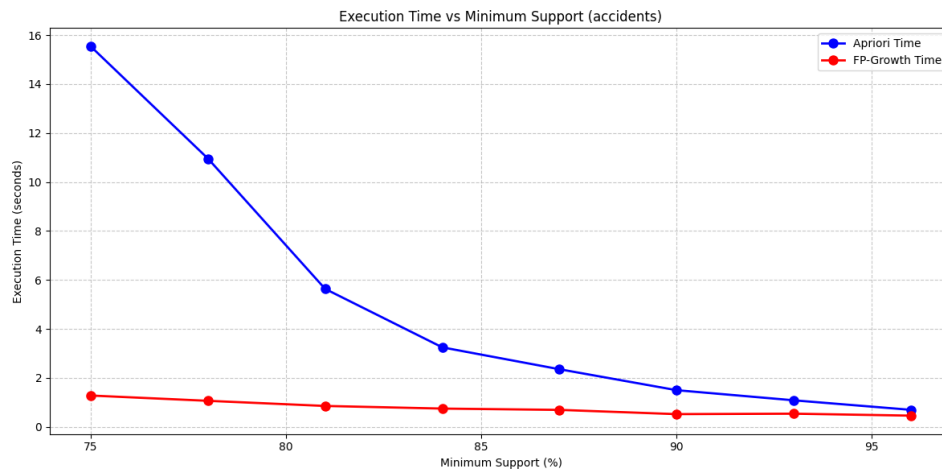


Figure 1: Execution time of Accidents dataset.

Figure 1 shows the execution time vs minimum support graph for accidents dataset for both Apriori algorithm and FP-Growth algorithm.

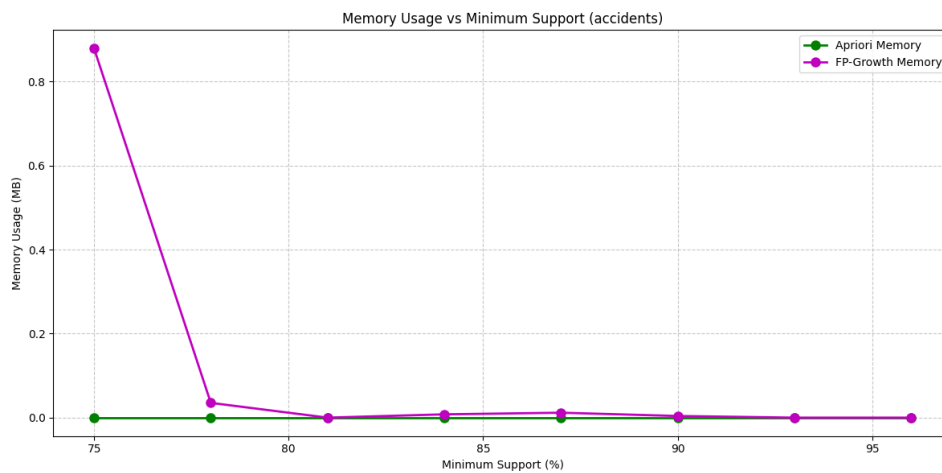


Figure 2: Memory usage of accidents dataset

Figure 2 shows the memory usage vs minimum support graph for accidents dataset for both Apriori algorithm and FP-Growth algorithm.

3.2 Connect

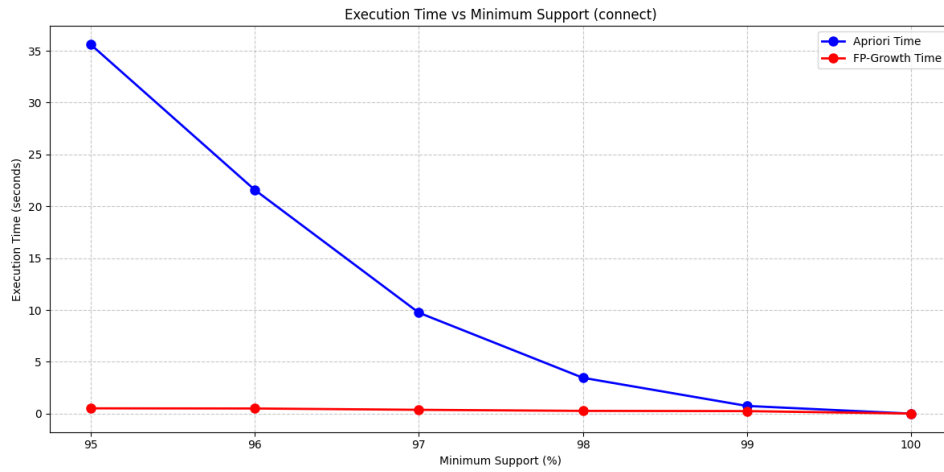


Figure 3: Execution time of Connect dataset.

Figure 3 shows the execution time vs minimum support graph for the Connect dataset for both Apriori algorithm and FP-Growth algorithm.

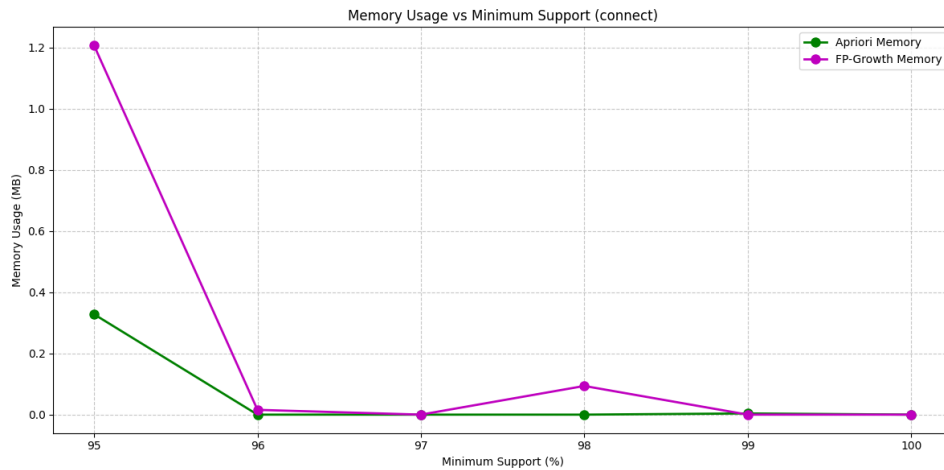


Figure 4: Memory usage of Connect dataset.

Figure 4 shows the memory usage vs minimum support graph for the Connect dataset for both Apriori algorithm and FP-Growth algorithm.

3.3 Mushroom

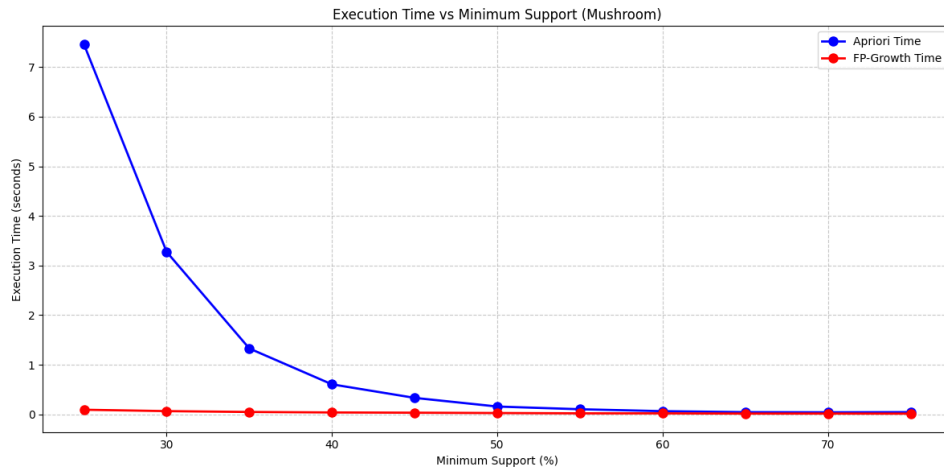


Figure 5: Execution time of Mushroom dataset.

Figure 5 shows the execution time vs minimum support graph for the Mushroom dataset for both Apriori algorithm and FP-Growth algorithm.

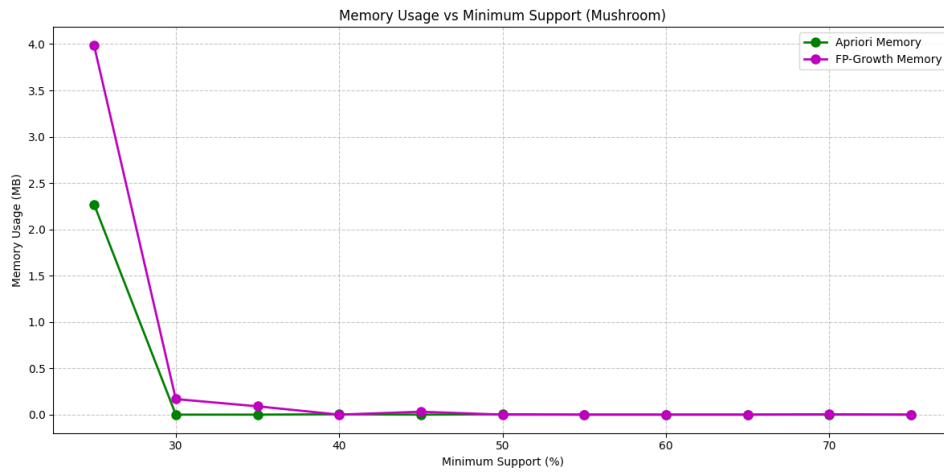


Figure 6: Memory usage of Mushroom dataset.

Figure 6 shows the memory usage vs minimum support graph for the Mushroom dataset for both Apriori algorithm and FP-Growth algorithm.

3.4 Retail

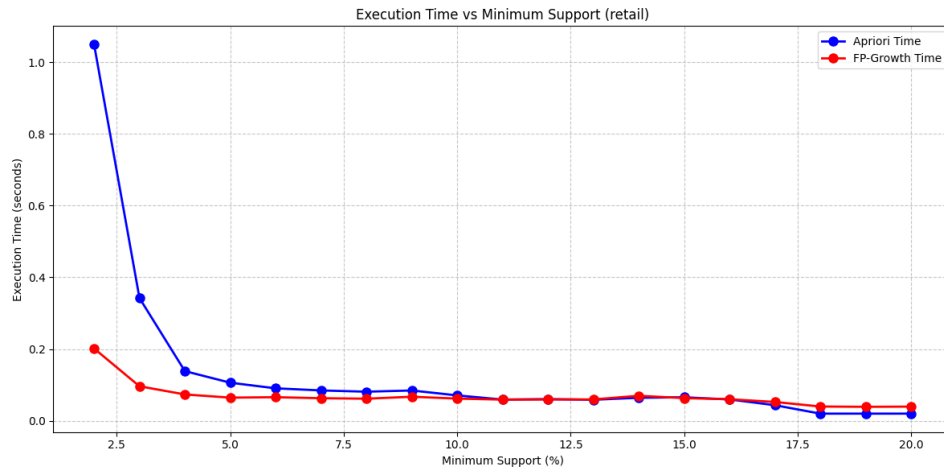


Figure 7: Execution time of Retail dataset.

Figure 7 shows the execution time vs minimum support graph for the Retail dataset for both Apriori algorithm and FP-Growth algorithm.

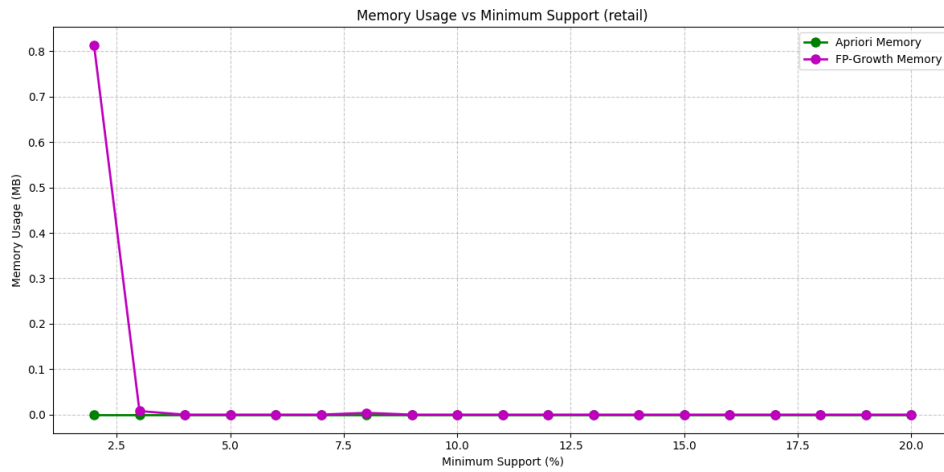


Figure 8: Memory usage of Retail dataset.

Figure 8 shows the memory usage vs minimum support graph for the Retail dataset for both Apriori algorithm and FP-Growth algorithm.

3.5 Kosarak

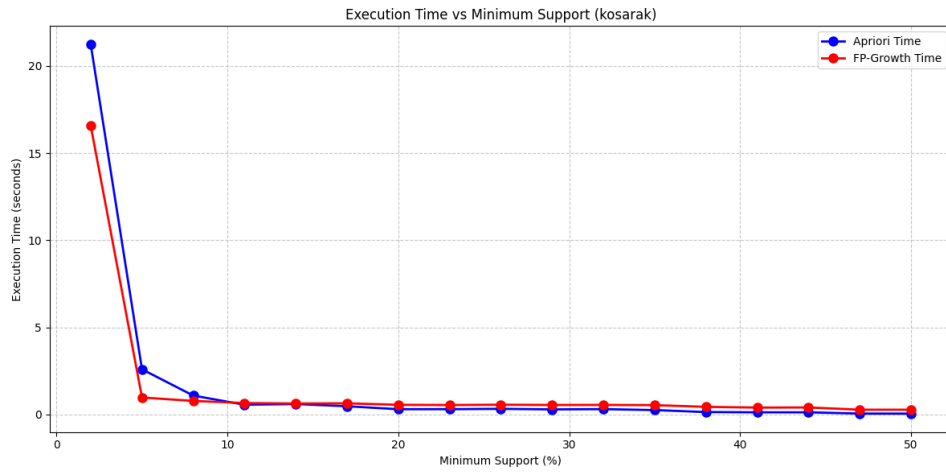


Figure 9: Execution time of Kosarak dataset.

Figure 9 shows the execution time vs minimum support graph for the Kosarak dataset for both Apriori algorithm and FP-Growth algorithm.

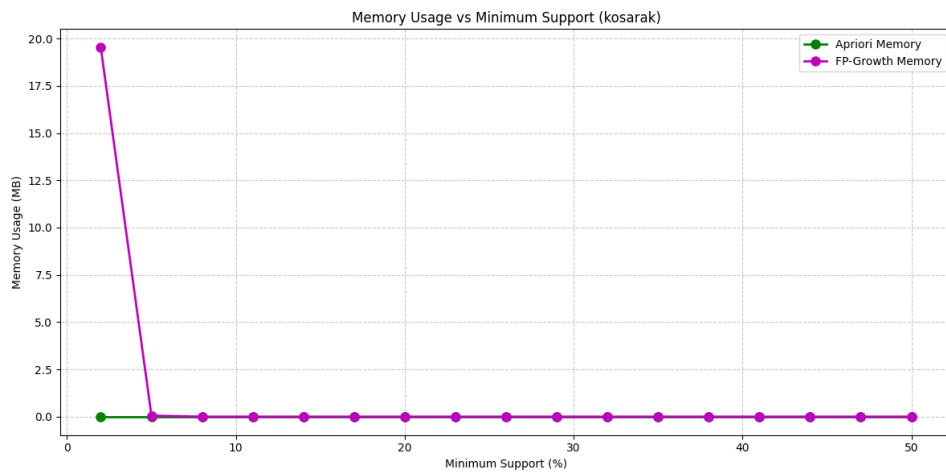


Figure 10: Memory usage of Kosarak dataset.

Figure 10 shows the memory usage vs minimum support graph for the Kosarak dataset for both Apriori algorithm and FP-Growth algorithm.

3.6 Chess

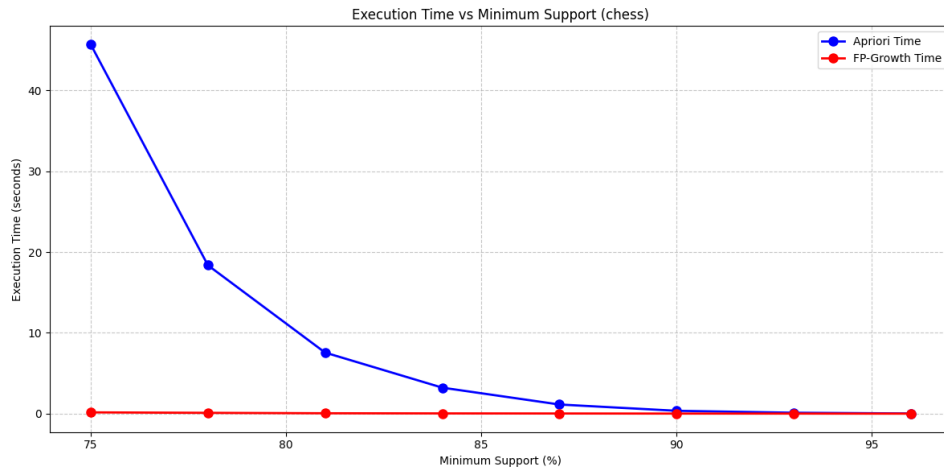


Figure 11: Execution time of Chess dataset.

Figure 11 shows the execution time vs minimum support graph for the Chess dataset for both Apriori algorithm and FP-Growth algorithm.

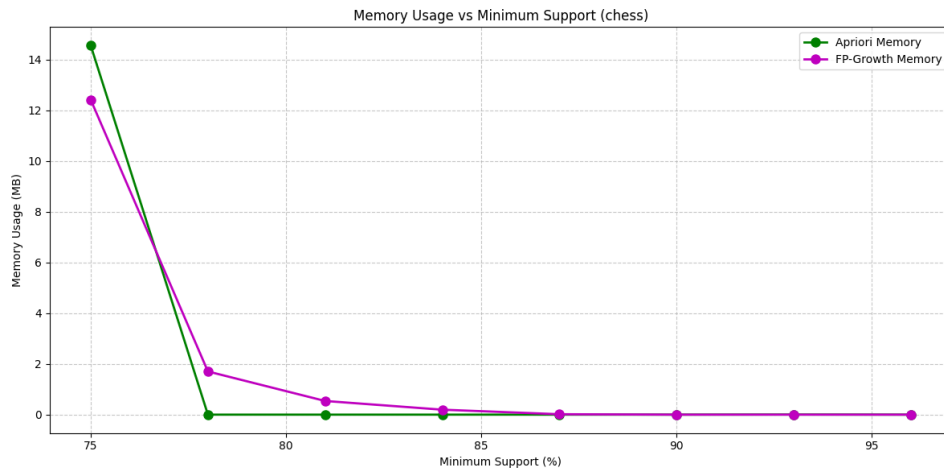


Figure 12: Memory usage of Chess dataset.

Figure 12 shows the memory usage vs minimum support graph for the Chess dataset for both Apriori algorithm and FP-Growth algorithm.

3.7 Pumsb

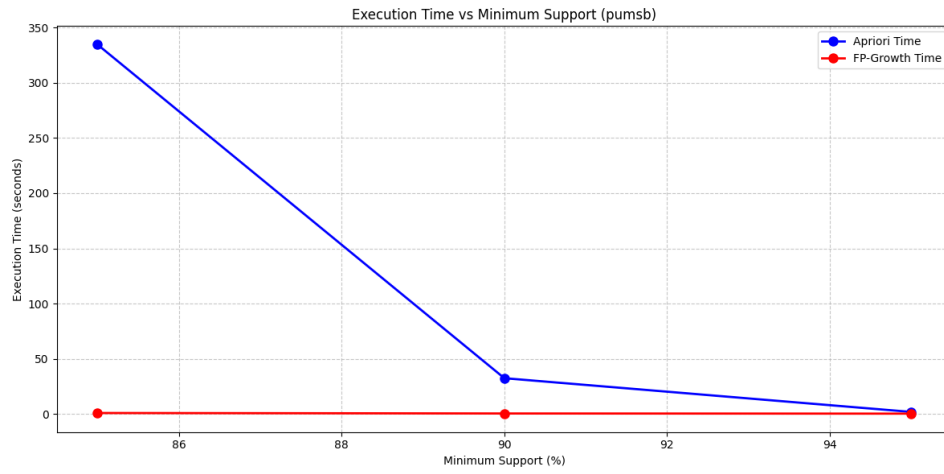


Figure 13: Execution time of PUMSB dataset.

Figure 13 shows the execution time vs minimum support graph for the PUMSB dataset for both Apriori algorithm and FP-Growth algorithm.

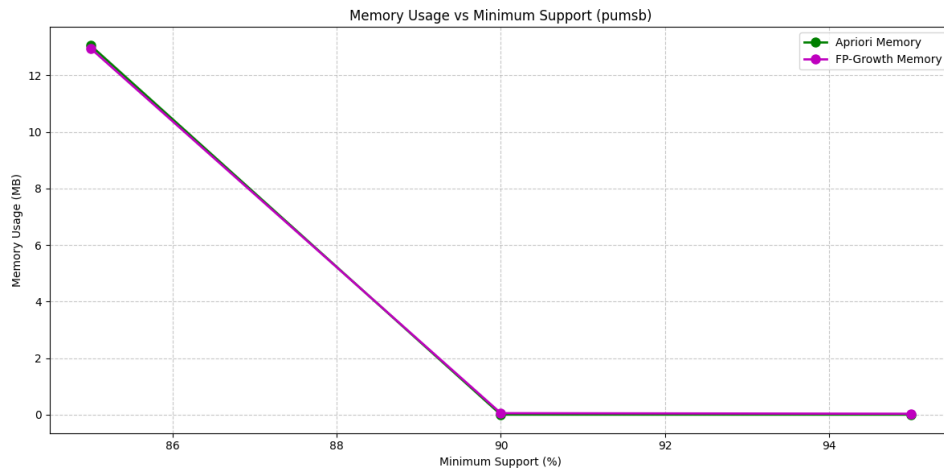


Figure 14: Memory usage of PUMSB dataset.

Figure 14 shows the memory usage vs minimum support graph for the PUMSB dataset for both Apriori algorithm and FP-Growth algorithm.

4 Result analysis

The experimental results clearly demonstrate the trade-offs between the Apriori and FP-Growth algorithms under varying data conditions. Apriori's performance deteriorates significantly with the increase in dataset size and itemset density, primarily due to its reliance on candidate generation and repeated scans of the database. This behavior leads to both longer execution times and higher memory consumption in complex datasets. Conversely, FP-Growth maintains a consistent advantage in performance by eliminating candidate generation and using a compact FP-tree structure, which allows for faster pattern mining and better scalability. While FP-Growth incurs a higher initial memory overhead due to tree construction, this is offset by its reduced processing time and lower memory growth rate as dataset complexity increases. These observations highlight that FP-Growth is better suited for large-scale or dense transactional datasets, whereas Apriori may remain a viable option for simpler, sparse datasets with fewer frequent patterns.