

Association Rule Mining for Market Basket Analysis: A Comparative Study of the Apriori and FP-Growth Algorithms

Saurabh Kumar¹, Ritu Kumari²

ABSTRACT

Association rule mining is a fundamental data-mining task that uncovers interesting relationships among items in large transactional databases, with market basket analysis being its most familiar application. The discovery of frequent itemsets, however, is computationally demanding, and the choice of algorithm strongly affects efficiency. This paper presents a comparative study of the two most widely used algorithms for association rule mining, namely Apriori and FP-Growth, applied to transactional data for market basket analysis. The algorithms were evaluated across a range of minimum-support thresholds in terms of execution time, memory usage, and the number of frequent itemsets and rules discovered. The results show that, while both algorithms generate the same rules for a given support and confidence, FP-Growth is substantially faster and more scalable, particularly at low support thresholds where the number of candidate itemsets for Apriori grows rapidly. Apriori remains simpler to implement and understand. The study confirms the efficiency advantage of the pattern-growth approach and offers practical guidance on algorithm selection for association rule mining.

KEYWORDS: *Association Rule Mining, Market Basket Analysis, Apriori, FP-Growth, Frequent Itemsets, Data Mining, Support and Confidence*

INTRODUCTION

The widespread digitisation of commerce has produced vast transactional databases that record the items purchased together by customers. Extracting useful patterns from such data is the province of association rule mining, which identifies relationships of the form that the purchase of certain items implies the purchase of others (1, 4).

The central computational challenge is the discovery of frequent itemsets, those sets of items that occur together at least as often as a specified minimum support. From these, rules satisfying a minimum confidence are derived. Because the number of possible itemsets grows exponentially with the number of items, efficient algorithms are essential (2, 5).

Two algorithms dominate the literature: Apriori, which generates and tests candidate itemsets level by level (2), and FP-Growth, which mines frequent patterns from a compact tree structure without candidate generation (3). This paper compares the two for market basket analysis, evaluating their efficiency and scalability across a range of support thresholds (8).

LITERATURE REVIEW

The foundations of association rule mining were laid by early work formalising the problem and introducing the support-confidence framework (1, 2). The Apriori algorithm exploited the downward-closure property, that all subsets of a frequent itemset must themselves be frequent, to prune the candidate space (2).

Subsequent research sought to overcome the cost of repeated database scans and candidate generation. The pattern-growth approach embodied in FP-Growth compresses the database into a frequent-pattern tree and mines it recursively (3), while other methods used vertical data formats and dynamic counting to improve efficiency (6, 7). Surveys have compared these algorithms and catalogued their trade-offs (8, 9).

CORE CONCEPTS IN ASSOCIATION RULE MINING

The literature defines a set of core concepts that underpin association rule mining (1, 4, 5). The principal ones can be summarised as follows:

- Support, the frequency with which an itemset occurs in the database.
- Confidence, the conditional probability of the consequent given the antecedent.
- Frequent itemsets, those whose support meets the minimum threshold.
- The downward-closure property exploited by candidate-generation methods.
- Lift and other measures used to assess the interestingness of rules.

RESEARCH GAP

Although both algorithms are well known, their relative performance depends strongly on the data and the support threshold, and concise comparisons that report execution time, memory, and rule counts together across a range of thresholds remain useful for practitioners.

This study addresses that need by comparing Apriori and FP-Growth on the same transactional data across several minimum-support thresholds, reporting efficiency and the patterns discovered.

OBJECTIVES

- To review the principles of association rule mining and frequent-itemset discovery.
- To implement the Apriori and FP-Growth algorithms for market basket analysis.
- To evaluate the algorithms across a range of minimum-support thresholds.
- To compare execution time, memory usage, and the number of itemsets and rules discovered.
- To provide guidance on algorithm selection for association rule mining.

METHODOLOGY

A transactional dataset representing retail purchases was used to compare the two algorithms. Both Apriori and FP-Growth were applied to the same data over a range of minimum-support thresholds at a fixed minimum confidence, so that differences in performance could be attributed to the algorithm rather than the data or parameters (2, 3).

Dataset And Parameters

The dataset consisted of retail transactions, each a set of purchased items. After removing incomplete records, the data were converted to the transactional format required by both algorithms. The minimum-support threshold was varied while the minimum confidence was held constant, as summarised below.

Table 1: Dataset and Mining Parameters

Parameter	Value
Data type	Retail transactions
Number of transactions	Approximately 9,800
Number of distinct items	120
Minimum support thresholds	0.5%, 1%, 2%, 3%, 5%
Minimum confidence	60%
Algorithms compared	Apriori, FP-Growth

Evaluation Procedure

For each support threshold, both algorithms were executed on the same hardware, and the execution time, peak memory usage, and the number of frequent itemsets and rules discovered were recorded. Because both algorithms are exact, they were expected to produce identical rules for a given support and confidence, allowing the comparison to focus on efficiency (3, 6).

RESULTS AND FINDINGS

The execution time of the two algorithms as a function of the minimum-support threshold is shown in Figure 1. The time required by both algorithms rose sharply as the support threshold was lowered, but FP-Growth was consistently faster than Apriori, and the gap widened markedly at low support, where the candidate space for Apriori expands rapidly (2, 3).

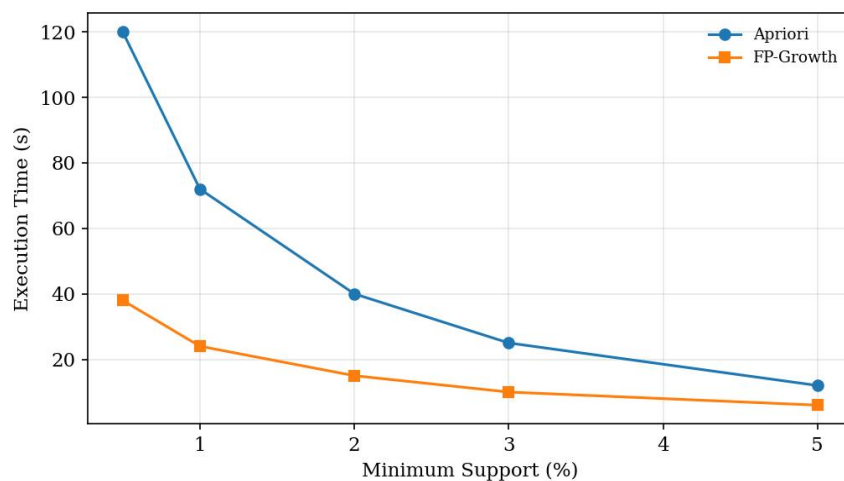


Figure 1: Execution Time of Apriori and FP-Growth versus Minimum Support

Table 2 compares the two algorithms at a representative low support threshold. Both discovered the same number of frequent itemsets and rules, confirming their equivalence in output, but FP-Growth required far less time and memory, reflecting its avoidance of explicit candidate generation.

Table 2: Comparison of Apriori and FP-Growth at 1% Minimum Support

Metric	Apriori	FP-Growth
Frequent itemsets discovered	1,240	1,240
Association rules generated	318	318

Execution time (s)	72	24
Peak memory usage (MB)	210	95
Database scans	Multiple	Two

DISCUSSION

The results confirm the well-known efficiency advantage of the pattern-growth approach (3). Apriori must generate and test candidate itemsets at each level and scan the database repeatedly, so its cost grows rapidly as the support threshold falls and the number of frequent itemsets increases (2, 8). This behaviour is clearly visible in the steep rise of its execution time at low support.

FP-Growth avoids candidate generation by compressing the database into a frequent-pattern tree and mining it recursively, requiring only two scans of the data (3). This explains its substantially lower execution time and memory usage, particularly at low support where Apriori is most disadvantaged (6, 9).

Both algorithms are exact and produced identical rules for a given support and confidence, so the choice between them is governed by efficiency and implementation considerations rather than by the patterns discovered (5, 8). Apriori remains valuable for its simplicity and intuitive operation, but FP-Growth is preferable for large or dense datasets and low support thresholds.

LIMITATIONS

- A single retail transaction dataset was used.
- Only two algorithms were compared.
- Execution time and memory depend on the implementation and hardware used.
- Rule interestingness beyond support and confidence was not deeply explored.

FUTURE SCOPE

- Comparison on multiple and larger transactional datasets.
- Inclusion of additional algorithms such as vertical-format methods.
- Evaluation of parallel and distributed implementations for big data.
- Use of additional interestingness measures such as lift and conviction.

- Extension to incremental and stream-based association rule mining.

CONCLUSION

This study compared the Apriori and FP-Growth algorithms for association rule mining in market basket analysis across a range of minimum-support thresholds. Both algorithms produced identical frequent itemsets and rules for a given support and confidence, but FP-Growth was substantially faster and more memory-efficient, with its advantage greatest at low support.

The findings confirm the efficiency benefit of the pattern-growth approach and provide practical guidance: Apriori is suitable for small datasets and instructional use, while FP-Growth is preferable for large, dense data and low support thresholds where scalability is essential.

REFERENCES

1. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 207-216.
2. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, 487-499.
3. Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1-12.
4. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
5. Tan, P. N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Pearson.
6. Zaki, M. J. (2000). Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3), 372-390.
7. Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 255-264.

8. Hipp, J., Guntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining: a general survey and comparison. *ACM SIGKDD Explorations Newsletter*, 2(1), 58-64.
9. Borgelt, C. (2005). An implementation of the FP-growth algorithm. *Proceedings of the Workshop on Open Source Data Mining*, 1-5.

***Author for Correspondence**

Saurabh Kumar

E-mail: saurabh.kumar@gmail.com

¹Assistant Professor, Department of Computer Science and Engineering, Magadh Institute of Engineering and Technology, Gaya, Bihar

²Research Scholar, Department of Computer Science and Engineering, Magadh Institute of Engineering and Technology, Gaya, Bihar

Received Date: April 6, 2026

Accepted Date: April 20, 2026

Published Date: April 28, 2026

Citation: Saurabh Kumar, Ritu Kumari. Association Rule Mining for Market Basket Analysis: A Comparative Study of the Apriori and FP-Growth Algorithms. *Journal of Data Mining and Knowledge Engineering*. 2026; 11(1): 25-31p.