ELKOM (Jurnal Elektronika dan Komputer)
p-ISSN : 1907-0012 (print) e-ISSN : 2714-5417 (online)
Vol. 18, No. 2, Desember 2025
https://journal.stekom.ac.id/index.php/elkom

Uncovering the Fundamental Structure of Consumer Shopping Patterns

(*A Synergistic Approach Combining Statistical Analysis and Graph Topology*)

Riadi Marta Dinata[1*], Kurniawan Atmadja [2], Marhaeni[3], Lely Mustika[4]
[1*]Department of Informatics Engineering, Faculty of Applied Science and Technology, ISTN
[2]Department of Mathematics, Faculty of Applied Science and Technology, ISTN
[3]Department of Information Systems, Faculty of Applied Science and Technology, ISTN
[4]Department of Architectural Engineering, Faculty of Applied Science and Technology, ISTN
riadimrt@gmail.com, kurniawan_atmadja@istn.ac.id, marhaeni@istn.ac.id, mustika@istn.ac.id

Article Info

Article history:

Received September 3, 2025
Revised Desember 3, 2025
Accepted Januari 13, 2026

Keywords:

Market Basket Analysis
Association Rules
Minimum Spanning Tree
Hybrid Model
Consumer Purchase Patterns

ABSTRACT

Traditional association rule analysis is effective at uncovering co-purchase patterns but fails to provide a global structural view of the market, which often results in fragmented and isolated insights. This study proposes a hybrid framework that integrates the Apriori algorithm with a Minimum Spanning Tree (MST) in order to validate and contextualize association rules within a single structural backbone. Transaction data from a retail store are transformed into a weighted, undirected product graph using an inverse-support function, and an MST is then extracted to represent the market backbone, while frequent itemsets and strong rules are obtained using Apriori. Experimental results on 236 multi-item transactions show that the MST backbone comprises 10 products and 9 fundamental links, with 66.67% of these links being confirmed by strong association rules, indicating a substantial coherence between statistical and structural evidence. The proposed model identifies 41 Apriori patterns that can be embedded in the MST and ranks them using a new metric, Structural Distance, which enables the categorization of Core Patterns, Bridge Patterns, and Complex Patterns according to their structural tightness. This hybrid perspective distinguishes dense, strategically meaningful bundles from anomalous but frequent combinations that are structurally peripheral, thereby offering a more holistic and actionable alternative to conventional Market Basket Analysis. The validated framework can support various applications, including store layout optimization, cross-selling strategies, and the design of path-based recommender systems, and it opens avenues for future extensions based on dynamic graphs and Graph Neural Networks.

Corresponding Author:

Riadi Marta Dinata
Universitas Lampung
Jl. Lintas Sumatera No.1, Kota Bandar Lampung
Email: riadimrt@gmail.com

1. INTRODUCTION

Digital economic transformation has fundamentally reshaped the retail landscape and consumer behavior, resulting in transaction data volumes of an unprecedented scale [1]. These data are no longer merely sales archives, but have become a strategic asset that is crucial for achieving competitive advantage in an increasingly saturated market [2], [3]. The urgency to analyze such data arises from business needs for personalization, operational efficiency, and product innovation that are grounded in a deep understanding of purchasing patterns, even though the complexity and high volume of transactional data present substantial challenges [4]. To address these issues, Market Basket Analysis (MBA) has emerged as a key discipline for extracting actionable insights from consumers' digital footprints [5].

Historically, MBA has been dominated by frequency-based statistical approaches, with the Apriori algorithm as one of the most influential pioneering methods introduced by Agrawal and Srikant (1994)[1]. Operating on the support and confidence metrics, Apriori has proven highly effective in mining statistically significant association rules from large-scale databases, and it remains widely used in contemporary studies of sales patterns across various retail sectors [4], [5]. However, as data complexity increases, the limitations of purely frequency-based approaches become more evident, which has motivated the development of topology-based paradigms that leverage graph theory as a powerful alternative [6], [7]. In these paradigms, transactional data are represented as a consumer–product graph in which products are nodes and co-purchase relationships are edges, a concept popularized by Huang et al. (2007) [8] and further advanced in modern e-commerce settings [9].

To simplify such complex product networks and extract their underlying structure, the Minimum Spanning Tree (MST) technique has been adopted as a compact representation of the market backbone [6], [10], [11]. The long history and solid mathematical foundations of MST make it a reliable tool for visualizing the most efficient relational framework among products [7], [10]. Beyond this, the idea of combining statistical and structural models in a single hybrid approach has shown strong potential in several complex domains; studies in medical imaging, for example, report substantial gains in classification performance when statistical and structural pattern recognition are jointly exploited [12]. Likewise, network analysis in bioinformatics often relies on graph models to uncover protein–protein interactions, illustrating the broad power of topological modeling [13]. Early work in consumer analytics has also explored the notion of hybrid consumption paths, indicating considerable potential in integrating these two perspectives within retail analytics [14].

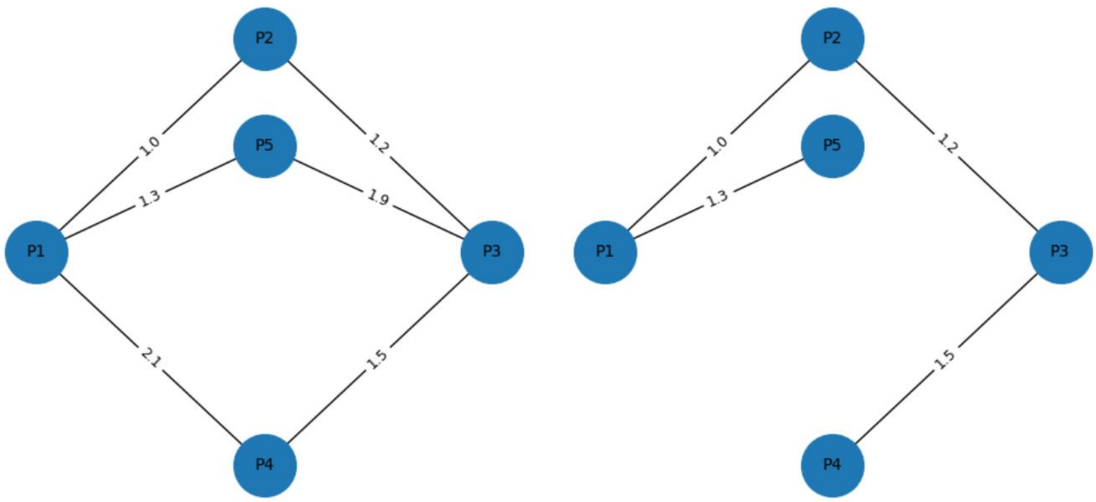

Figure 1. Hybrid Apriori–MST workflow

Figure 1. illustrates how a complex product co-purchase network can be simplified into a clearer structural representation. The left panel shows the full weighted co-purchase graph, where each node represents a product and each edge reflects how strongly two products are purchased together. Several cycles appear in this graph, indicating multiple possible pathways between products. The right panel presents the corresponding Minimum Spanning Tree (MST), which retains all product nodes but selects only the edges with the smallest total weight. This produces a cycle-free structure that highlights the

p-ISSN: 1907-0012
e-ISSN: : 2714-5417

core relationships among products. The MST reveals the most essential co-purchase links and identifies intermediate bridge products, providing a simplified backbone that supports further analysis such as confirming association pathways or detecting hidden product connections.

Despite the extensive body of work on each of these pillars, there remains a substantial methodological gap between them [8], [9]. Statistical approaches such as Apriori, although quantitatively accurate, tend to produce fragmented and isolated insights, essentially providing a list of facts without a unifying map [1]. In contrast, purely graph-based approaches can offer a global map of relationships but often lack strong statistical validation for each individual connection, which makes them vulnerable to spurious links driven mainly by product popularity [6]. Existing hybrid attempts, although promising, frequently remain at the level of enrichment rather than achieving a truly synergistic integration in which the two perspectives formally validate one another [12]. Alternative methods such as clustering, as demonstrated by Apriyanto and Sitio (2025)[2], can group products effectively but still fail to reveal connective pathways and the bridging role of specific products across clusters, and there is still no framework that systematically uses the structural efficiency of MST to contextualize association rules while at the same time exploiting the statistical significance of Apriori to validate the MST structure itself [8], [11].

In response to this gap, the present study proposes a hybrid framework for the structural validation of association rules [1], [8]. The primary objective is to design and implement a model that integrates Apriori and the Minimum Spanning Tree in a mutually reinforcing way, thereby producing a market insight map that is both holistic and operationally useful [6], [11]. The main contribution of this study lies in a cross-validation mechanism in which the MST structure constructed from support-based metrics is re-evaluated using additional association measures, enabling a clear distinction between Confirmed Pathways and Hidden Bridges [12]. The remainder of this article is organized as follows: Section 2 details the research methodology, Section 3 presents the experimental results and discussion, and Section 4 provides the concluding remarks of this study [5].

2. METHOD

This study follows an experimental quantitative design that integrates association rule mining and graph-based analysis into a single hybrid workflow. The research procedure consists of four main stages: data acquisition and preprocessing, extraction of statistically significant patterns using the Apriori algorithm, construction of a weighted product graph and Minimum Spanning Tree (MST), and hybrid validation and categorization of purchasing pathways.

2.1 Data acquisition and preprocessing

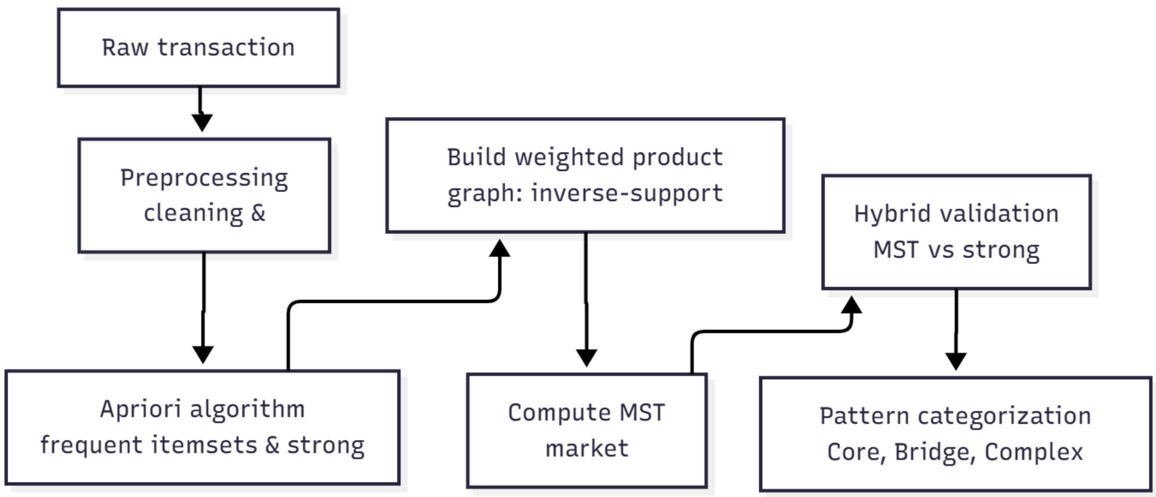

Figure 2. Overall hybrid Apriori–MST workflow

Figure 2 presents the overall workflow of the proposed hybrid Apriori–MST model, from raw transaction data to hybrid pattern insights. The process begins with data acquisition and preprocessing, where noisy or single-item transactions are removed and products are encoded into consistent identifiers. The cleaned transactions are then used as input to the Apriori algorithm to generate frequent itemsets and statistically strong association rules.

The empirical evaluation is conducted on a retail transaction dataset consisting of 300 purchase receipts collected from a single store. All receipts that contain only a single item are removed, resulting in 236 valid multi-item transactions that are subsequently used as the input for pattern mining and network construction.Each transaction is represented as a set of product identifiers, forming a standard transaction–item incidence structure suitable for Market Basket Analysis. The preprocessing stage includes: (1) cleaning inconsistent or missing product codes, (2) mapping all products to integer labels, and (3) storing the cleaned transactions in a format compatible with the Apriori implementation (e.g., a binary incidence matrix or list of itemsets).

In the next stage, a weighted, undirected product graph is constructed by connecting co-purchased products and assigning edge weights based on an inverse-support function. On top of this graph, a Minimum Spanning Tree is extracted using a greedy algorithm in order to obtain the market backbone. Finally, the MST is cross-validated with strong association rules and used to compute Structural Distance, which allows the identification of Confirmed Pathways, Hidden Bridges, and ranked pattern categories in the hybrid insight map.

2.2 Association rule mining with Apriori

In the second stage, frequent itemsets are extracted using the Apriori algorithm, which iteratively generates larger candidate itemsets based on the Apriori principle and evaluates their support against a predefined minimum threshold. The minimum support is set to 0.05, which yields 75 frequent itemsets in total, including 45 two-itemsets, 20 three-itemsets, and several higher-order itemsets.

From these frequent itemsets, association rules are derived and evaluated using support, confidence, and lift. Rules with confidence greater than 0.3 and lift greater than 1.0 are retained as strong rules and serve as the statistical backbone for subsequent hybrid validation. These strong rules represent pairs or bundles of products that exhibit both high co-occurrence and meaningful deviation from independence.

2.3 Construction of weighted product graph and MST

The third stage transforms the transaction data into a weighted, undirected product graph, where each node represents a unique product and each edge represents a co-purchase relationship between two products. For every pair of products that appears together in at least one transaction, the empirical support of the corresponding 2-itemset is computed and then mapped to an edge weight using an inverse-support function.

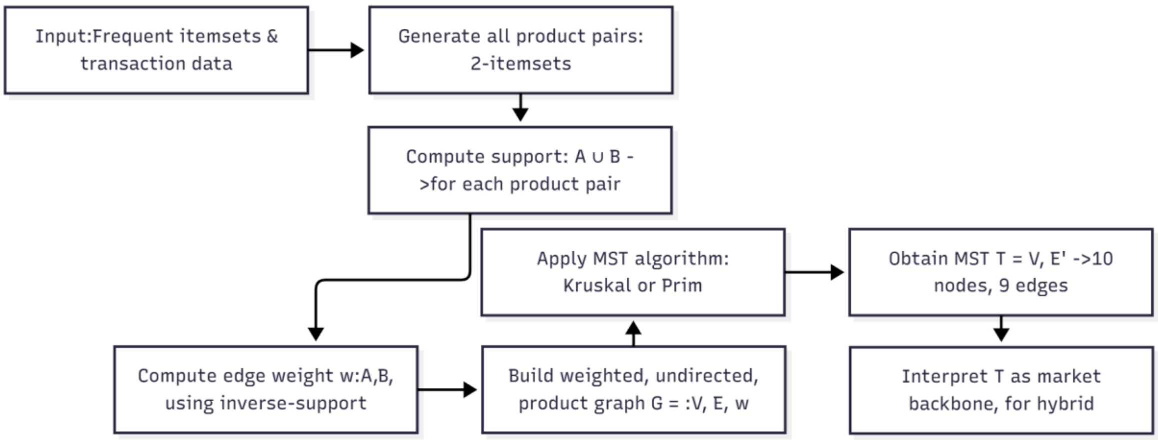

Figure 2. Construction of weighted product graph and MST

Figure 2. illustrates the construction of the weighted product graph and the subsequent extraction of the Minimum Spanning Tree. The procedure starts from frequent itemsets and transaction data, generates all co-purchased product pairs, and computes the empirical support for each pair. The

support values are then transformed into edge weights through an inverse-support function, which is used to build a weighted, undirected product graph. Finally, a greedy MST algorithm such as Kruskal or Prim is applied to obtain a tree with minimum total weight, which is interpreted as the core market backbone for the hybrid analysis.

Formally, if A and B are two products with joint support support$(A \cup B)$, the weight $w(A, B)$ is defined as a monotonically decreasing function of this support, with a small constant ϵ added to avoid division by zero. This mapping ensures that product pairs with higher co-purchase frequency correspond to shorter "distances" in the graph and are therefore more likely to be selected when extracting the MST.

After all edge weights are computed, a Minimum Spanning Tree is generated from the weighted graph using a greedy algorithm such as Kruskal or Prim, which is guaranteed to find the tree with minimum total edge weight. The resulting MST contains 10 nodes and 9 edges and is interpreted as the core market backbone that captures the most efficient set of structural relationships among products.

2.4 Hybrid validation and pathway categorization

In the final stage, the MST structure is cross-validated with the strong association rules obtained from Apriori to distinguish structurally important and statistically confirmed relationships from purely structural or purely statistical ones. Each MST edge is checked against the set of strong rules; edges that correspond directly to strong rules are labeled as Confirmed Pathways, while edges that do not have a matching strong rule are labeled as Hidden Bridges.

Beyond edge-level validation, all frequent itemsets (particularly 2-itemsets and 3-itemsets) are projected onto the MST to compute their Structural Distance, defined as the sum of edge weights along the shortest path connecting all products in the itemset within the tree. Using this distance and the number of intermediate nodes (bridges) in the path, each pattern is assigned to one of three hybrid categories: Core Patterns (direct links with zero bridges), Bridge Patterns (paths with exactly one bridge), and Complex Patterns (paths involving multiple bridges and longer structural distances).

To evaluate the coherence of the hybrid model, the Rule Overlap metric is calculated as the proportion of MST edges that are confirmed by strong association rules; in the reported experiment, 6 out of 9 MST edges are confirmed, yielding a Rule Overlap of 66.67%. This quantitative indicator, together with the Structural Distance–based ranking of itemsets, forms the basis for interpreting the resulting hybrid insight map and for comparing the contributions of purely statistical, purely structural, and integrated perspectives.

2.5 Mathematical formulation

This subsection summarizes the main mathematical definitions used in the proposed hybrid model. Let T be the set of transactions with $N = |T|$, and let $n(X)$ denote the number of transactions that contain an itemset X. The empirical support of X is defined as

$$\text{support}(X) = \frac{n(X)}{N} \tag{1}$$

For an association rule $A \rightarrow B$, confidence is defined as

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)} = \frac{n(A \cup B)}{n(A)}, \tag{2}$$

and lift is defined as

$$\text{lift}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)\,\text{support}(B)}. \tag{3}$$

These three measures are used to identify and filter statistically significant rules.

To model the structural relations between products, a weighted, undirected graph $G = (V, E, w)$ is constructed, where V is the set of product nodes, E is the set of undirected edges representing co-purchase relationships, and $w: E \rightarrow \mathbb{R}^+$ assigns a positive weight to each edge. Given two products A and B with joint support support$(A \cup B)$, the corresponding edge weight is defined through an inverse-support function

$$w(A,B) = \frac{1}{\text{support}(A \cup B) + \epsilon}, \quad (4)$$

where $\epsilon > 0$ is a small constant to avoid division by zero and to stabilize the weights. This formulation makes frequently co-purchased product pairs correspond to shorter distances in the graph.

On top of this weighted graph, a Minimum Spanning Tree (MST) is extracted to obtain the core market backbone. Formally, for a connected weighted graph $G = (V, E, w)$, an MST is a subgraph $T = (V, E')$ with $E' \subseteq E$ that connects all nodes without cycles and minimizes the total edge weight

$$\min_{T \subseteq E'} \sum_{e \in T} w_e. \quad (5)$$

Greedy algorithms such as Kruskal's and Prim's algorithms are used to compute this optimal tree, which then serves as the structural reference for defining Confirmed Pathways, Hidden Bridges, and Structural Distance in the hybrid analysis

Table 1. Summary of symbols and mathematical definitions used in the hybrid model

Notation	**Description**	**Role in the model**
T, (N =	T	)
X, A, B	Individual item or itemset	Building blocks of frequent patterns and association rules
$n(X)$	Number of transactions containing itemset X	Used to compute support
support(X)	Empirical frequency of itemset Xin the dataset	Thresholding frequent itemsets and rules
confidence($A \rightarrow B$)	Conditional probability of Bgiven A	Measuring predictive strength of rules
lift($A \rightarrow B$)	Deviation from independence between $A$and B	Filtering rules with meaningful positive association
$G = (V, E, w)$	Weighted, undirected product graph	Structural representation of co-purchase relationships
V	Set of product nodes	Universe of items considered in the network
E	Set of undirected edges between co-purchased products	Encodes observed co-occurrence in transactions
$w(A, B)$	Edge weight derived from inverse support of $A \cup B$	Translates statistical strength into graph distance
$T = (V, E')$	Minimum Spanning Tree on G	Market backbone capturing the most efficient product connections
$\sum_{e \in T} w_e$	Total weight of edges in the MST	Objective function minimized by MST algorithms
Structural Distance	Sum of weights along the MST path connecting items in a pattern	Basis for hybrid ranking and pattern categorization

RESULT

The model was implemented on a dataset consisting of 300 transaction receipts from a retail store. After preprocessing steps that removed transactions containing only a single item, 236 valid multi-item transactions remained and served as the basis for subsequent analysis. Running the Apriori algorithm with a minimum support threshold of 0.05 yielded 75 frequent itemsets, including 45 two-itemsets, 20 three-itemsets, and several higher-order combinations.

Table 2. Summary of the Hybrid Analysis of Consumer Shopping Patterns

Itemset Apriori	Nilai Support	Pola Hibrida	Jml Jembatan	Jarak Struktural	Kategori Pola
{8, 3}	0.186	Direct links	0	5.36	Core Pattern
{2, 10}	0.153	Direct links	0	6.56	Core Pattern
{8, 2}	0.144	Direct links	0	6.94	Core Pattern
{9, 2}	0.144	Direct links	0	6.94	Core Pattern
{3, 7}	0.144	Direct links	0	6.94	Core Pattern
{8, 6}	0.144	Direct links	0	6.94	Core Pattern
{10, 4}	0.140	Direct links	0	7.15	Core Pattern
{1, 4}	0.131	Direct links	0	7.61	Core Pattern
{2, 5}	0.127	Direct links	0	7.87	Core Pattern
{2, 3}	0.140	2 → 8 → 3	1	12.30	Bridge Pattern
{3, 6}	0.140	3 → 8 → 6	1	12.30	Bridge Pattern
{8, 7}	0.136	8 → 3 → 7	1	12.30	Bridge Pattern
{8, 10}	0.131	8 → 2 → 10	1	13.50	Bridge Pattern
{9, 10}	0.106	9 → 2 → 10	1	13.50	Bridge Pattern
{2, 4}	0.127	2 → 10 → 4	1	13.71	Bridge Pattern
{2, 6}	0.123	2 → 8 → 6	1	13.88	Bridge Pattern
{8, 9}	0.131	8 → 2 → 9	1	13.88	Bridge Pattern
{10, 5}	0.102	10 → 2 → 5	1	14.42	Bridge Pattern
{1, 10}	0.127	1 → 4 → 10	1	14.76	Bridge Pattern
{8, 5}	0.114	8 → 2 → 5	1	14.81	Bridge Pattern
{9, 5}	0.123	9 → 2 → 5	1	14.81	Bridge Pattern
(8, 3, 7)	0.072	Complex Pathway	2	12.30	Complex Pattern
(2, 10, 4)	0.051	Complex Pathway	2	13.71	Complex Pattern
(8, 2, 5)	0.051	Complex Pathway	2	14.81	Complex Pattern
(8, 3, 6)	0.068	Complex Pathway	3	17.67	Complex Pattern
(8, 2, 3)	0.055	Complex Pathway	3	19.25	Complex Pattern
(8, 9, 2)	0.051	Complex Pathway	3	20.82	Complex Pattern
(1, 10, 4)	0.051	Complex Pathway	3	21.92	Complex Pattern
(8, 3, 5)	0.059	Complex Pathway	4	25.54	Complex Pattern
(8, 6, 7)	0.051	Complex Pathway	4	26.19	Complex Pattern
(1, 4, 9)	0.055	Complex Pathway	4	28.26	Complex Pattern
(3, 6, 7)	0.059	Complex Pathway	5	31.55	Complex Pattern
(8, 10, 3)	0.055	Complex Pathway	5	32.36	Complex Pattern
(8, 9, 3)	0.055	Complex Pathway	5	33.13	Complex Pattern
(1, 2, 4)	0.051	Complex Pathway	5	35.03	Complex Pattern
(1, 10, 6)	0.051	Complex Pathway	5	35.20	Complex Pattern
(8, 5, 6)	0.059	Complex Pathway	5	36.56	Complex Pattern
(1, 3, 7)	0.055	Complex Pathway	6	40.57	Complex Pattern
(9, 5, 7)	0.055	Complex Pathway	6	41.92	Complex Pattern
(8, 1, 9)	0.051	Complex Pathway	8	56.52	Complex Pattern
(8, 1, 3)	0.055	Complex Pathway	9	61.89	Complex Pattern

Table 2 summarizes a representative subset of these patterns along with their hybrid characteristics. For each frequent itemset, the table reports its empirical support, the corresponding structural pathway in the MST (for example, direct connections or multi-step routes such as 2 → 8 → 3), the number of intermediate bridge nodes along that path, the resulting Structural Distance, and the assigned hybrid category. The hybrid model organizes the patterns into three main groups: Core

Patterns (direct links with zero bridges), Bridge Patterns (paths that include exactly one bridge), and Complex Patterns (multi-bridge, longer-distance bundles).

Table 3. Summary of Structural Metrics of the Generated MST

Structural Metric	Value	Description
Number of Nodes	10	Total number of unique products analyzed.
Number of Edges	9	Fundamental co-purchase relationships forming the market backbone.
Tree Diameter	5	The longest shortest path between any two products within the MST structure.
Average Path Length	2.82	The average number of steps required to connect each pair of products.

Table 3 provides a brief overview of the structural properties of the Minimum Spanning Tree (MST) derived from the transaction data. The number of nodes and edges reflects the size of the product network and the essential co-purchase relationships preserved in the tree. The tree diameter indicates the maximum distance between any two products, highlighting how far apart the most separated items are within the market structure. Meanwhile, the average path length summarizes the overall connectivity efficiency, showing how many steps are typically required to move from one product to another within the MST. Together, these metrics describe the backbone structure of consumer purchasing patterns in a compact and interpretable form.

At the core of the hybrid analysis lies the Minimum Spanning Tree constructed over the weighted product graph, which forms a market backbone with 10 nodes (products) and 9 edges (fundamental relationships). Cross-validation between this MST structure and the set of strong association rules (confidence greater than 0.3 and lift greater than 1.0) shows that 6 out of 9 tree edges are confirmed statistically, while the remaining 3 edges are not supported by strong rules. This yields a Rule Overlap of 66.67%, indicating a substantial alignment between the structurally efficient connections selected by the MST and the statistically significant associations discovered by Apriori.

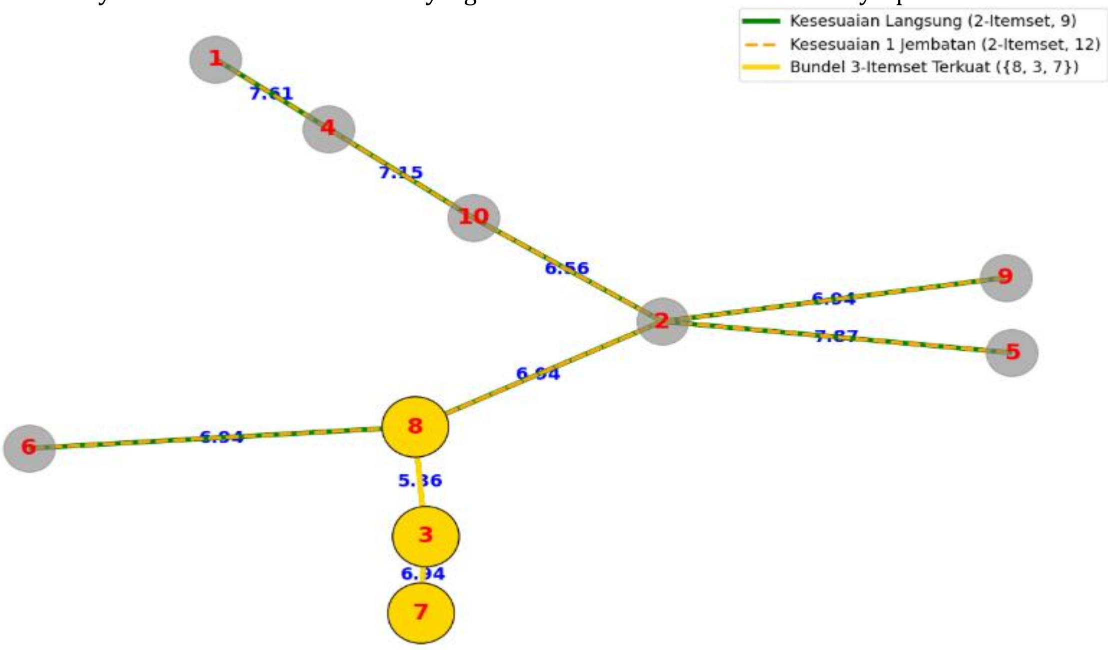

Figure 3. Hybrid Insight Map (Validated MST Structure Derived from Transactions)

Figure 3 visualizes the validated MST-based insight map, highlighting Confirmed Pathways and Hidden Bridges. In this visualization, solid green edges denote tree connections that are supported by strong association rules, whereas dashed red edges indicate structurally important but statistically unconfirmed links.

From all frequent itemsets identified by Apriori, 41 patterns (including both two-itemsets and three-itemsets) are mapped successfully onto the MST and thus receive a well-defined Structural Distance. These 41 patterns are then ranked and categorized based on their distance and number of bridges, which provides a graded view of how tightly each bundle is embedded within the market backbone. Table 2 summarizes key topological metrics of the extracted MST: the structure comprises 10 nodes and 9 edges, has a tree diameter of 5 (the longest shortest-path between any two products), and an average path length of 2.82 steps between product pairs.

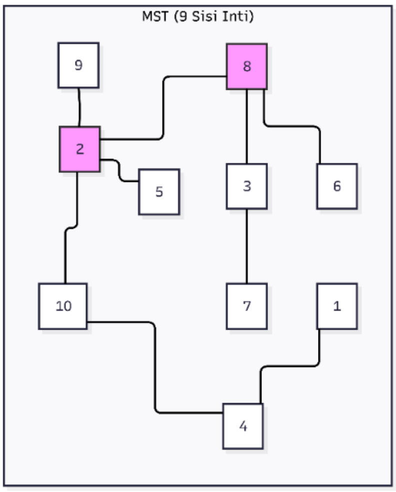

Figure 4. Minimum Spanning Tree (MST) Framework / Pattern for MBA

Figure 4 focuses on the MST skeleton itself, showing the market backbone extracted from all frequent pairs and emphasizing that the hybrid model is able to project and analyze 41 Apriori patterns within this structure, with the MST backbone consisting of exactly 10 nodes and 9 edges.

3. DISCUSSION/CONCLUSION

4.1 Discussion

Quantitatively, the hybrid model demonstrates a strong capacity to provide structural context for Apriori's statistical findings. Instead of returning a flat list of frequent itemsets, the framework ranks all detected patterns according to their position in the MST backbone, which makes it possible to distinguish structurally central bundles from more peripheral ones.

The hybrid categorization summarized in Table 3 effectively organizes 41 Apriori patterns into distinct strategic tiers. There are 9 Core Patterns, each corresponding to a direct MST edge with zero intermediate bridges, which represent the most fundamental and stable product pairs; among them, the pair {8,3} emerges as the strongest link, with the lowest Structural Distance of 5.36. In addition, 12 Bridge Patterns capture relationships that are not directly adjacent in the backbone but are connected through exactly one intermediate product. A notable example is the pair {2,3}, which Apriori identifies simply as a frequent co-purchase, while the hybrid model reveals that this association is structurally mediated by product 8, a critical insight for cross-promotion and shelf placement strategies.

The strength of the hybrid approach becomes even more apparent in the analysis of three-item bundles. Although Apriori marks all 20 three-itemsets as frequent, the MST-based Structural Distance differentiates between compact and anomalous bundles. The pattern {8,3,7} appears as the most cohesive bundle, with a distance of only 12.30, indicating that all three products lie close to each other on the market backbone. In contrast, the pattern {8,1,3}, while frequent in the Apriori sense, is revealed as structurally anomalous: its Structural Distance reaches 61.89 and spans 9 bridges, essentially traversing the entire backbone to connect the products. This ability to objectively separate dense, strategically meaningful bundles from loose, opportunistic co-occurrences is a key contribution of the hybrid model and aligns with prior arguments on the importance of graph analysis for understanding product relationships.

Despite its strengths, the implemented MST–Apriori model has several limitations that need to be acknowledged. First, the analysis is static and operates on aggregated historical transactions, delivering only a snapshot of consumer behavior without capturing temporal dynamics such as seasonality or trend shifts. Second, the MST is, by design, a simplified representation: it preserves a single, cycle-free backbone and intentionally omits other potentially strong parallel connections that may exist in the full product graph. These limitations indicate that while MST is highly effective for extracting a market skeleton and ranking patterns structurally, complementary analyses on the full graph are required to fully characterize dense local clusters and overlapping communities of products.

4.2 Conclusion and future work

This study set out to design and evaluate a hybrid framework that integrates statistical association analysis and graph-based topology in order to map consumer purchasing patterns more holistically. The main findings show a strong coherence between the core market structure extracted by the MST and the association rules that are statistically significant according to support, confidence, and lift. More importantly, the model can effectively distinguish between Confirmed Pathways, which are robust both structurally and statistically, and Hidden Bridges, which represent strategic bridging relationships that remain invisible to purely statistical methods.

The fundamental contribution of this work lies in the introduction of a cross-validation framework and a new ranking metric, Structural Distance, which allows Apriori patterns to be evaluated by their structural importance rather than frequency alone. From a theoretical standpoint, this validates the feasibility of a genuine synergy between statistical and graph-based methods for deeper Market Basket Analysis. From a practical perspective, the proposed approach provides a methodology for transforming a long list of association rules into a visual and interpretable insight map that can directly inform strategic decisions, ranging from store layout optimization and product bundling to the design of path-based recommender systems.

At the same time, the static nature of the current model opens promising avenues for future research. One extension is to use the extracted graph structures and MST-derived features as inputs to Graph Neural Networks, thereby equipping the framework with predictive capabilities for next-basket or next-item recommendation. Another direction is to develop dynamic graph variants of the model that track the evolution of the market backbone over time, enabling the detection of emerging trends and shifts in consumer behavior. Finally, future work can explore the construction of personalized Hybrid Insight Maps for different customer segments, which would pave the way for more fine-grained and adaptive recommendation systems tailored to heterogeneous buying journeys.

REFERENCES

[1] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB ’94)*, San Francisco, CA: Morgan Kaufmann, 1994, pp. 487–499.

[2] B. Apriyanto and S. Sitio, “Penerapan k-means dalam menganalisis pola pembelian pelanggan pada data transaksi e-commerce,” *Bit-Tech*, vol. 7, no. 3, pp. 790–797, 2025, doi: 10.32877/bt.v7i3.2195.

[3] S. Fuaddah, “Dampak perubahan sosial dan ekonomi terhadap pola konsumsi masyarakat urban di era digital,” *OSF Preprints*, 2023, doi: 10.31219/osf.io/8npaq.

[4] R. Fadillah, L. Qadriah, and M. Rizal, “Market basket analisis data mining untuk mengetahui pola penjualan pada Cerry Mart Beureunueun menggunakan algoritma apriori,” *JRR*, vol. 5, no. 1, pp. 234–239, 2023, doi: 10.47647/jrr.v5i1.1152.

[5] A. Pramudita, Y. Hardianto, M. Marcellino, F. Thiodoris, and C. Kristi, “Analisis dataset estimasi penjualan produk pada Café For Good Luck menggunakan algoritma apriori,” *JCSIT*, vol. 2, no. 1, pp. 116–125, 2024, doi: 10.70248/jcsit.v2i1.1739.

[6] R. Campos and M. Ricardo, “A fast algorithm for computing minimum routing cost spanning trees,” *Computer Networks*, vol. 52, no. 17, pp. 3229–3247, 2008.

[7] R. L. Graham and P. Hell, “On the history of the minimum spanning tree problem,” *IEEE Annals of the History of Computing*, vol. 7, no. 1, pp. 43–57, 1985.

[8] Z. Huang, D. D. Zeng, and H. Chen, “Analyzing consumer–product graphs: Empirical findings and applications in recommender systems,” *Manage Sci*, vol. 53, no. 7, pp. 1146–1164, 2007.

[9] S. Shah, “Analysis of e-commerce product graphs,” *TechRxiv*, 2020, doi: 10.36227/techrxiv.12814244.

p-ISSN: 1907-0012
e-ISSN: : 2714-5417

[10] Wamiliana, "Solving the degree constrained minimum spanning tree problem using Tabu and modified penalty search methods," *Jurnal Teknik Industri*, vol. 6, no. 1, pp. 1–9, 2004.
[11] Wamiliana, *Minimum spanning tree dan desain jaringan*. Bandarlampung, Indonesia: Pusaka Media, 2022.
[12] E. Özdemir and Ç. Gündüz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Trans Med Imaging*, vol. 32, no. 2, pp. 474–483, 2013.
[13] V. Deshmukh, C. Cannings, and A. Thomas, "Estimating the parameters of a model for protein–protein interaction graphs," *Mathematical Medicine and Biology*, vol. 23, no. 4, pp. 279–295, 2006.
[14] S. Rachmat and A. Antoni, "Analisis pola pembelian secara online: Pendekatan kuantitatif untuk memahami perilaku konsumen dan pengaruhnya pada pasar e-commerce," *PESHUM*, vol. 4, no. 3, pp. 4300–4310, 2025, doi: 10.56799/peshum.v4i3.8929.