**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

# Segmentation-Aware Recommendation with Cluster-Specific Item Graphs Using Pointwise Mutual Information for Market Basket Analysis

**Khalifatur Rauf[1,*], Arief Hermawan[1], Donny Avianto[2]**

[1] Fakultas Pascasarjana ,Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Sleman, Indonesia
[2] Fakultas Sains dan Teknologi, Program Studi Informatika, Universitas Teknologi Yogyakarta, Sleman, Indonesia
Email: [1,*]khalifatur.6250211012@student.uty.ac.id, [2]ariefdb@uty.ac.id, [3]donny@uty.ac.id
Correspondence Author Email: khalifatur.6250211012@student.uty.ac.id
Submitted: **20/04/2026**; Accepted: **02/06/2026**; Published**: 05/06/2026**

**Abstract**−Traditional Association Rule-based recommendation methods often exhibit limited coverage and high redundancy when applied to sparse transactional data, thereby constraining their effectiveness for product discovery in e-commerce systems. This study proposes a hybrid recommendation framework that integrates customer behavioral segmentation with graph-based item representation learning to address these limitations. Customers are first grouped into behaviorally homogeneous clusters using historical transaction features. For each cluster, an item co-occurrence graph is constructed and weighted using pointwise mutual information to mitigate sparsity bias and emphasize informative associations. Graph-based representation learning is then applied using Node2Vec to generate low-dimensional product embeddings that capture both local structural proximity and higher-order relational patterns. The proposed framework explicitly restricts the candidate item space to the Top 100 most frequent products within each behavioral cluster, thereby focusing the recommendation task on improving localized discovery within high-frequency product segments rather than global catalog exploration. The objective of this research is to assess whether segmentation-aware graph embeddings can outperform traditional FP-Growth association rules under a strict temporal split between the Historical Training Set and the Hold-out Evaluation Set, ensuring realistic and leakage-free evaluation. Model performance is evaluated using precision, recall, normalized discounted cumulative gain, and intra-list diversity on the Hold-out Evaluation Set. Experimental results indicate that the proposed graph-based approach improves ranking quality and diversity within constrained high-frequency item spaces, demonstrating more effective localized discovery within Top 100 product segments compared to FP-Growth. These results demonstrate that graph-based embeddings are more robust to sparse behavioral patterns within high-frequency product segments and better suited for exploratory recommendation scenarios within dense product subsets. The proposed framework offers a scalable and temporally valid foundation for knowledge-driven recommender systems.

**Keywords**: Graph-Based Recommendation; Node2Vec; Customer Segmentation; Association Rule Mining; E-Commerce Transactions

## 1. INTRODUCTION

The rapid growth of e-commerce platforms has led to an exponential increase in transactional data, particularly in online grocery retail systems where users generate large-scale, high-dimensional, and highly sparse purchase histories. Within this context, market basket analysis plays a crucial role in uncovering latent purchasing patterns that support product recommendation, cross-selling strategies, and decision support systems in Information Systems (IS) [1]. However, extracting meaningful insights from such data remains challenging due to inherent sparsity, heterogeneity in customer purchasing behavior, and the difficulty of accurately modeling higher-order relationships among staple, high-frequency products within behaviorally segmented yet data-limited environments [2].

Traditional approaches to market basket analysis, particularly association rule mining methods such as Frequent Pattern Growth (FP-Growth), are widely adopted due to their interpretability and computational efficiency [3]. However, these methods rely heavily on minimum support thresholds, which often leads to the exclusion of infrequent yet potentially meaningful item associations. At the same time, even within high-frequency product subsets, rule-based methods remain limited in capturing higher-order co-occurrence structures and multi-hop relational dependencies, resulting in redundant and less context-sensitive recommendations across behavioral segments [4].

To address these limitations, graph-based representation learning has emerged as an alternative paradigm for modeling item relationships in recommendation systems. Techniques such as Node2Vec enable the transformation of item co-occurrence structures into continuous vector spaces, allowing the capture of latent and higher-order relationships beyond direct co-purchase patterns [5]. Despite their effectiveness, many existing graph-based recommender systems treat user populations homogeneously, without explicitly accounting for behavioral heterogeneity among customers [6].

Recent research has demonstrated a growing reliance on graph-based and embedding-driven recommendation models to address sparsity and complex item relationships in large-scale e-commerce data. Wang et al. [7] conducted a seminal study on Neural Graph Collaborative Filtering, where item-user interactions are modeled as a graph structure and optimized using neural message propagation, showing substantial improvements over traditional matrix factorization methods. Meanwhile, Wu et al. [8], in a comprehensive survey published in ACM Computing Surveys, systematically reviewed recent advances in graph-based recommender systems, highlighting the dominance of Graph Neural Networks, heterogeneous graphs, and self-supervised learning as the current research frontier. Their analysis emphasizes that most research models prioritize representational power and ranking accuracy, often at the expense of interpretability and deployment simplicity. In parallel, Liu et al. [9] proposed STAMP, a session-based recommendation model published in KDD, which leverages attention mechanisms to capture short-term user intent

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

from sequential purchase behavior. However, these approaches predominantly rely on global interaction modeling or session-level dynamics, with limited investigation into how localized, cluster-specific item graphs constructed over constrained high-frequency product spaces compare against traditional rule-based mining under strictly separated temporal evaluation settings. Consequently, despite strong predictive performance, existing state-of-the-art approaches remain limited in modeling localized purchasing structures and behavior-specific item associations under sparse transactional conditions.

To address this gap, this study proposes a segmentation-driven graph-based recommendation framework for market basket analysis. The proposed approach first segments customers based on purchasing behavior using clustering techniques, and subsequently constructs cluster-specific item co-occurrence graphs. This segmentation strategy enables the model to capture localized purchasing structures within behavioral groups rather than relying on a global interaction graph. The modeling process is explicitly constrained to the Top 100 most frequent products within each cluster, ensuring that representation learning focuses on dense, high-signal item interactions where higher-order relationships are more reliably inferred. Furthermore, edge weights are computed using Positive Pointwise Mutual Information (PPMI) to reduce sparsity bias and improve the robustness of item association strength estimation in sparse data settings [10].

In addition to graph-based modeling, this study employs FP-Growth as a rule-based baseline to provide a systematic comparison under identical segmentation conditions. Unlike prior comparisons conducted on global item spaces, this study evaluates both approaches within the same localized, cluster-specific Top 100 product subsets, enabling a controlled assessment of representational learning versus rule-based mining in dense interaction regimes. To ensure methodological rigor, all experiments are conducted using a strict temporal split between the Historical Training Set and the Hold-out Evaluation Set, where only past interactions are used for model construction and future interactions are reserved for evaluation. This design prevents data leakage and ensures that performance metrics reflect realistic predictive capability rather than memorization of observed transactions.

The primary objective of this research is to examine whether segmentation-aware graph-based learning can improve product discovery and recommendation quality in sparse e-commerce environments. Specifically, the study focuses on improving localized discovery performance within high-frequency product segments, evaluating how effectively higher-order relationships among staple items can be leveraged to enhance recommendation relevance and diversity. The study evaluates performance in terms of recommendation accuracy, diversity, and catalog coverage, while also analyzing the interpretability of product communities derived from graph structures.

This study makes three primary contributions. First, it introduces a segmentation-first recommendation framework that integrates customer behavioral clustering with graph-based item representation learning under constrained, high-frequency product spaces. Second, it demonstrates the effectiveness of PPMI-weighted cluster-specific item graphs in capturing higher-order relationships among staple products, thereby improving localized discovery within dense interaction subsets. Third, it provides a controlled empirical comparison between graph-based representation learning and rule-based association mining under identical Top 100 product constraints and a strict temporal split between the Historical Training Set and the Hold-out Evaluation Set, offering new insights into the relative strengths of localized graph embeddings versus traditional rule-based methods in e-commerce recommendation systems.

# 2. RESEARCH METHODOLOGY

## 2.1. Research Stages

This subsection outlines the end-to-end methodological pipeline adopted in this study, which integrates data preprocessing, behavioral segmentation, graph-based representation learning, and rule-based benchmarking under a strict temporal evaluation protocol. The workflow is designed to ensure reproducibility, prevent information leakage, and enable a controlled comparison between competing recommendation paradigms. The overall research process is illustrated in Figure 1, which summarizes the sequential stages from data acquisition through evaluation.

This research follows a structured and sequential data mining pipeline designed to ensure methodological rigor, reproducibility, and realistic evaluation of recommendation performance in sparse e-commerce environments. The study begins with the acquisition of transactional data from the Instacart Online Grocery Basket Analysis dataset, which represents real-world online grocery purchase behavior characterized by high dimensionality and extreme sparsity. To maintain computational feasibility while preserving behavioral heterogeneity, a subset of 5,000 users is selected using uniform random sampling. This sampling strategy is adopted to preserve statistical sparsity characteristics and behavioral diversity across users while ensuring that graph construction and embedding learning remain computationally tractable at scale.

Following data acquisition, a strict temporal partitioning strategy is applied to prevent data leakage and ensure realistic evaluation. Transactions labeled as Historical Training Set are used exclusively for model construction, including feature engineering, customer segmentation, graph construction, and rule mining. Transactions labeled as Hold-out Evaluation Set are reserved solely for evaluation. This temporal separation ensures that all reported performance metrics reflect true predictive capability rather than retrospective pattern memorization.

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

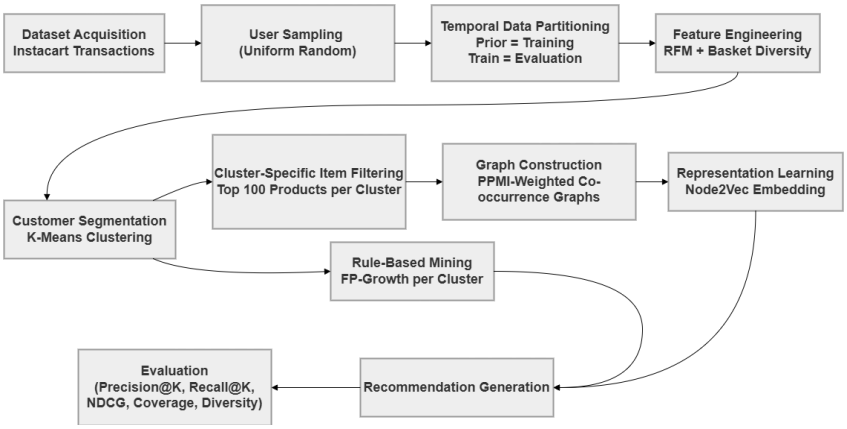


**Figure 1.** Flow diagram of the segmentation-based recommendation framework

Feature engineering is then performed on the Historical Training Set to characterize customer purchasing behavior. Recency, frequency, and monetary-related attributes are derived using a correct temporal definition of recency based on the time elapsed since a customer 's most recent order. In addition, basket diversity is computed as a normalized measure, defined as the ratio between the number of unique aisles and the total number of purchased items, in order to capture behavioral variety while mitigating multicollinearity with basket volume.

Based on these engineered features, customers are segmented using K-Means clustering. The optimal number of clusters is determined empirically using the elbow method and silhouette coefficient analysis, ensuring that the resulting segmentation reflects both compactness and separation quality of behavioral groups rather than being predefined arbitrarily. This segmentation stage constitutes a central design decision in the proposed framework, as all subsequent modeling steps are performed independently within each behaviorally homogeneous cluster.

For each cluster, the item space is constrained to the top 100 most frequently purchased products. This represents a deliberate head-focused modeling design, where the objective is to analyze dense, high-frequency product interactions rather than long-tail behaviors. By restricting the item universe, the study emphasizes localized discovery and higher-order relational learning among staple products that dominate transactional activity within each cluster. Cluster-specific item co-occurrence graphs are then constructed, where nodes represent products and edges represent co-occurrence relationships within baskets. Edge weights are computed using Positive Pointwise Mutual Information (PPMI), which reduces sparsity-induced bias and stabilizes association strength estimation.

On these weighted graphs, Node2Vec is applied to learn dense vector representations of products. Node2Vec is a graph representation learning technique that generates biased random walks to preserve both local neighborhood structure and higher-order connectivity, enabling the embedding space to encode latent item relationships beyond direct co-purchase signals. In parallel, FP-Growth is executed within each cluster using identical transactional inputs to serve as a rule-based baseline. FP-Growth is a classical frequent pattern mining method that constructs compact FP-trees to efficiently extract association rules, but its reliance on support thresholds limits its ability to capture complex relational structures in sparse or fragmented interaction spaces. This dual-model design enables a controlled comparison between graph-based representation learning and rule-based mining under identical segmentation and data constraints.

Finally, recommendations are generated using both approaches and evaluated on the Hold-out Evaluation Set. The evaluation framework employs top-N recommendation metrics, including Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG), alongside catalog coverage and intra-list diversity. These metrics collectively assess not only predictive accuracy but also the system 's ability to promote product discovery and long-tail exposure, which are essential objectives in modern e-commerce recommendation systems.

### 2.2. Dataset and Data Preprocessing

This study utilizes the Instacart Online Grocery Basket Analysis dataset from Kaggle [11], a large-scale transactional dataset that reflects real-world online grocery purchasing behavior. The dataset contains anonymized records of customer transactions, including order-level, product-level, temporal, and categorical attributes. Such data exhibits high dimensionality, extreme sparsity, and long-tail product distributions, which are characteristic challenges in e-commerce market basket analysis.

To ensure experimental control and computational feasibility, a subset of 5,000 users is selected using uniform random sampling at the user level. This sampling design preserves heterogeneous purchasing trajectories while maintaining statistical sparsity patterns typical of real-world e-commerce systems, and it also ensures computational tractability for graph construction and embedding learning. As a result, the sampled dataset maintains representative behavioral diversity under constrained computational resources.

The raw transactional data are merged across multiple tables to form a unified transaction-level dataset. Each record corresponds to a product added to a specific order and includes user identifiers, temporal ordering information,

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

reorder indicators, and product category metadata. A sample of the merged transaction dataset is presented in Table 1 to illustrate the structure and attributes used in this study.

**Table 1**. Sample of Merged Transactions Dataset

| Order id | Product id | Add to cart order | Reordered | … | Order number | Order dow | Orderhourof day | Days since prior order | Aisle id | Department id |
|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 35108 | 1 | 0 | … | 29 | 3 | 13 | 6 | 36 | 16 |
| 28 | 40593 | 2 | 1 | … | 29 | 3 | 13 | 6 | 108 | 16 |
| 28 | 17461 | 3 | 0 | … | 29 | 3 | 13 | 6 | 35 | 12 |
| 28 | 22825 | 4 | 1 | … | 29 | 3 | 13 | 6 | 24 | 4 |
| 28 | 25256 | 5 | 1 | … | 29 | 3 | 13 | 6 | 84 | 16 |

A strict temporal partitioning strategy is applied to separate model construction and evaluation data. Transactions labeled as Historical Training Set are used exclusively for all preprocessing and modeling stages, including feature engineering, customer segmentation, graph construction, and association rule mining. Transactions labeled as Hold-out Evaluation Set are reserved solely for evaluation. This temporal separation is enforced throughout the experimental pipeline to prevent data leakage and ensure that all predictive assessments reflect forward-looking performance.

Before modeling, customer-level behavioral features are engineered from the Historical Training Set. These features summarize purchasing behavior across multiple dimensions, including temporal recency, purchase frequency, basket volume, and category-level diversity. Recency is computed as the elapsed time since a customer 's most recent order, ensuring a correct temporal interpretation. Diversity measures are normalized by total item count to avoid multicollinearity with basket size and to provide scale-independent indicators of exploratory purchasing behavior. A sample of the engineered and normalized customer features used for segmentation is shown in Table 2.

**Table 2.** Sample of Corrected and Normalized Engineered Customer Features

| Recency | Frequency | Total items | Aisle diversity index | Dept diversity index |
|---|---|---|---|---|
| 17 | 23 | 437 | 0.096110 | 0.029748 |
| 30 | 6 | 59 | 0.203390 | 0.101695 |
| 8 | 3 | 41 | 0.512195 | 0.195122 |
| 21 | 8 | 79 | 0.240506 | 0.113924 |
| 12 | 3 | 37 | 0.378378 | 0.189189 |

All preprocessing steps are performed exclusively on the Historical Training Set to preserve the integrity of the evaluation framework. No information from the Hold-out Evaluation Set is incorporated during feature computation or data transformation. Through these procedures, the dataset is transformed into a clean, temporally consistent, and analytically robust structure that supports subsequent customer segmentation, graph construction, and recommendation modeling.

### 2.3. Customer Segmentation

Customer segmentation is applied to explicitly model heterogeneity in purchasing behavior prior to recommendation modeling. Rather than relying on a single global interaction model, customers are grouped into behaviorally homogeneous segments to ensure that subsequent item relationship learning reflects localized purchasing patterns [12].

Segmentation is performed using the K-Means clustering algorithm on standardized customer-level behavioral features derived exclusively from the Historical Training Set. These features represent temporal purchasing activity, purchase volume, and normalized category-level diversity, enabling the clustering process to differentiate between routine-oriented and exploratory shopping behaviors. Feature standardization is applied to ensure balanced contribution across attributes and to prevent scale dominance. The process of determining the optimal number of clusters is illustrated in Figure 2, which reports both elbow and silhouette analysis results used to guide model selection.

The optimal number of clusters is determined through a combined evaluation of the elbow method and silhouette analysis. While the silhouette score reaches its maximum at two clusters, this configuration provides overly coarse behavioral separation that is insufficient to capture meaningful heterogeneity in purchasing patterns. Increasing the number of clusters beyond three results in diminishing returns, as indicated by marginal reductions in within-cluster variance and a consistent decline in silhouette values. Based on this trade-off between cohesion, separation quality, and behavioral interpretability, the number of clusters is selected using this empirical diagnostic procedure rather than being fixed a priori.

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

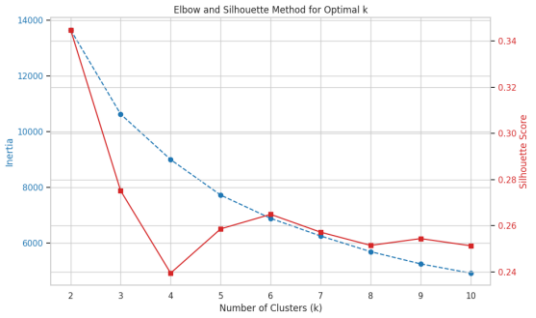


**Figure 2.** Elbow and silhouette analysis for determining the optimal number of customer clusters

Clustering is conducted solely on historical transactions, with no exposure to Hold-out Evaluation Set, thereby preserving temporal validity and preventing data leakage. All subsequent modeling stages, including item graph construction, representation learning, and association rule mining, are executed independently within each customer cluster [13]. This cluster-specific modeling design is essential for capturing localized relational structures among frequently purchased products, which are often obscured when modeled under a global interaction assumption [14].

### 2.4. Cluster-Specific Item Graph Construction

Following customer segmentation, item relationship modeling is performed independently within each customer cluster to capture behavior-specific purchasing structures. Instead of constructing a single global item graph, this study adopts a cluster-specific graph construction strategy to ensure that item associations reflect localized co-purchase patterns within homogeneous behavioral groups.

For each cluster, transactional data from the Historical Training Set are aggregated into basket-level representations, where each basket corresponds to a single order containing a set of purchased products. Item co-occurrence is defined based on joint appearance within the same basket. To control graph density and computational cost, the item space in each cluster is restricted to the Top 100 most frequently purchased products. This restriction constitutes a deliberate head-focused modeling choice, emphasizing dominant and high-frequency product interactions while enabling the analysis of strong co-purchase structures within staple items rather than long-tail exploration. This design preserves the most informative purchasing signals while maintaining tractable graph construction [15].

An undirected weighted graph is then constructed for each cluster, where nodes represent products and edges represent co-occurrence relationships. Edge weights are computed using Positive Pointwise Mutual Information (PPMI), which quantifies the statistical association between item pairs relative to their independent occurrence probabilities. By retaining only positive PMI values, the graph suppresses spurious associations caused by item rarity and removes unreliable negative correlations in sparse transactional data [15]. Figure 3 provides a visual illustration of a PPMI-weighted item graph constructed for a representative customer cluster, highlighting the resulting localized interaction structure.

Figure 3 presents a visualization of the initial PPMI-weighted item graph for a representative cluster. To improve interpretability, only the top 50 nodes by degree centrality are displayed. Node size reflects connectivity strength, while edge thickness encodes association intensity. This visualization highlights hub products and densely connected item groups, providing an intuitive representation of localized purchasing structures learned from historical data.

All co-occurrence statistics and PPMI computations are derived exclusively from Historical Training Set within each cluster, ensuring temporal consistency and preventing data leakage. The resulting cluster-specific item graphs serve as the structural foundation for subsequent representation learning and recommendation modeling.

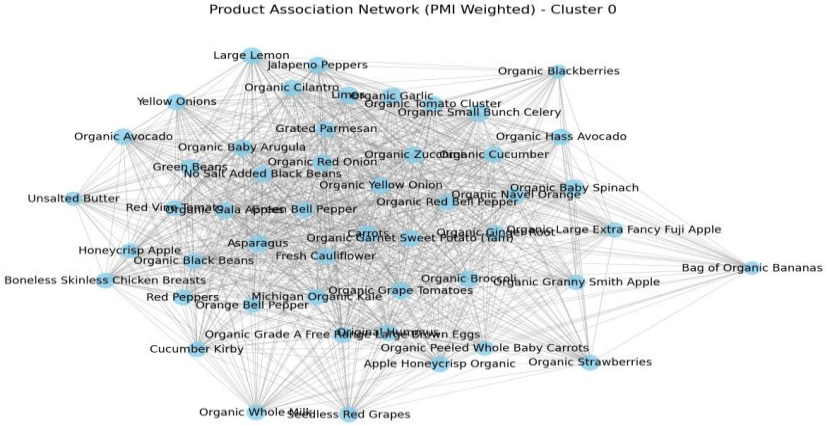


**Figure 3.** Sample of PPMI-Weighted Item Graph for a Customer Cluster

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

### 2.5. Graph-Based Representation Learning using Node2Vec

After constructing cluster-specific PPMI-weighted item graphs, graph-based representation learning is applied to encode products into dense vector spaces. This study employs the Node2Vec algorithm to learn low-dimensional embeddings that preserve both local neighborhood structures and higher-order connectivity patterns within each cluster-specific graph.

Node2Vec operates by simulating biased random walks over the item graph to generate node sequences that capture structural context. These sequences are subsequently processed using a skip-gram optimization objective to learn vector representations in which products with similar co-occurrence and connectivity patterns are embedded in close proximity. By balancing breadth-first and depth-first exploration strategies, Node2Vec captures both homophily-based relationships and structural equivalence among products [5]. This property is particularly relevant in sparse e-commerce interaction graphs, where direct co-purchase signals are limited and higher-order connectivity provides additional relational signal for embedding quality.

Node2Vec is applied independently to each cluster-specific graph using only data from the Historical Training Set, ensuring strict temporal validity and preventing information leakage from the Hold-out Evaluation Set. The resulting embeddings provide continuous representations of item relationships, which are particularly advantageous in sparse transactional environments where direct co-occurrence signals are limited [16]. To qualitatively assess the learned embedding space, Figure 4 presents a two-dimensional t-SNE projection of product embeddings for a representative customer cluster, illustrating their structural organization.

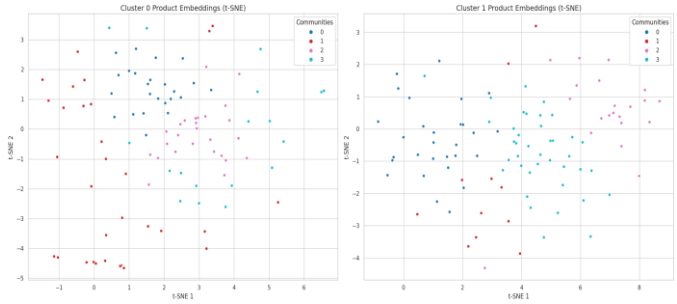


**Figure 4.** t-SNE projection of Node2Vec product embeddings for a representative customer cluster

To qualitatively examine the structural properties of the learned embeddings, Figure 4 presents a two-dimensional projection of product embeddings for a representative customer cluster using t-distributed Stochastic Neighbor Embedding (t-SNE). As shown in Figure 4, products that frequently co-occur or share similar graph neighborhoods are positioned in close proximity, forming coherent local groupings. These groupings correspond to intuitive product themes, such as fresh produce, dairy items, and beverage-related products, despite the absence of explicit category labels during embedding learning. Moreover, items belonging to the same graph-based community tend to cluster spatially, indicating that the Node2Vec embeddings successfully preserve higher-order and transitive relationships encoded in the PPMI-weighted item graph. This visualization serves as an exploratory validation of embedding quality, while all quantitative evaluations are conducted under the strict temporal evaluation framework described in Section 2.7.

### 2.6. Rule-Based Baseline using FP-Growth

To provide a systematic and controlled baseline for comparison, this study employs Frequent Pattern Growth (FP-Growth) as a rule-based association mining approach. FP-Growth is selected due to its widespread adoption in market basket analysis and its ability to efficiently mine frequent itemset without explicit candidate generation [17]. It constructs a compact FP-tree representation of transactional data, enabling efficient discovery of frequent co-occurrence patterns under minimum support constraints.

FP-Growth is applied independently within each customer cluster using the same training transactions that are utilized for graph construction and representation learning. This cluster-level execution ensures that both graph-based and rule-based models operate under identical behavioral segmentation conditions, enabling a fair and meaningful comparison between modeling paradigms [18].

Association rules are derived from frequent itemset based on predefined minimum support thresholds, with rules representing direct co-occurrence relationships between products. Unlike graph-based models, FP-Growth relies exclusively on observed item co-purchases that satisfy frequency constraints, without capturing higher-order or transitive relationships [3]. Consequently, its representational capacity is primarily limited to first-order co-occurrence structures, which may be restrictive when modeling dense relational dependencies even within high-frequency product segments. As a result, rule generation is inherently constrained by data sparsity and long-tail item distributions.

All rule mining procedures are performed solely on the Historical Training Set. No information from the Hold-out Evaluation Set is used during itemset extraction or rule generation, thereby maintaining temporal validity and preventing data leakage. Recommendations are generated by matching observed trigger items to antecedents of mined rules within each cluster.

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

By incorporating FP-Growth as a baseline under identical segmentation and evaluation settings, this study enables a controlled assessment of the relative strengths and limitations of rule-based association mining and graph-based representation learning in sparse e-commerce recommendation scenarios.

### 2.7. Recommendation Generation and Evaluation Protocol

Recommendation generation and evaluation are conducted under a strict temporal validation framework to ensure methodological rigor and to prevent data leakage. Unlike random or cross-sectional splits, this study adopts a forward-looking evaluation design in which only historical transactions are used for model construction, while future transactions are reserved exclusively for performance assessment.

All transactions labeled as Historical Training Set are used as the training set. This subset serves as the sole source of information for customer segmentation, cluster-specific graph construction, Node2Vec representation learning, and FP-Growth rule mining. No information from future orders is incorporated at any stage of model development. By restricting all learning processes to historical interactions, the framework mirrors realistic deployment scenarios in which recommendations must be generated without access to future purchases.

The evaluation set consists of transactions labeled as Hold-out Evaluation Set, corresponding to the next observed orders of the same sample users. These transactions represent unseen future behavior and are used exclusively for testing recommendation performance. To maintain consistency with the segmentation-driven design, cluster assignments derived from the Historical Training Set are transferred to users in the evaluation set, ensuring that recommendations are generated within the appropriate behavioral cluster without re-estimating the segmentation model.

For each evaluation transaction, a single item is treated as the observed trigger, while the remaining items in the same basket constitute the ground truth set. Recommendations are generated independently for each model based on the trigger item and compared against the held-out items. This evaluation strategy simulates a realistic recommendation scenario in which a system suggests complementary products after observing an initial purchase.

Performance is assessed using top N recommendation metrics, including Precision@K, Recall@K, and Normalized Discounted Cumulative Gain (NDCG), which measure ranking quality and relevance [19]. In addition to accuracy-oriented metrics, catalog Coverage and intra-list Diversity are computed to evaluate the system 's ability to promote product discovery and to expose users to a broader range of items [20]. In this study, these discovery-oriented metrics are interpreted specifically as indicators of localized product exploration within high-frequency (Top 100) item segments, rather than global catalog expansion.

By enforcing strict temporal split and isolating training and evaluation processes, the proposed evaluation protocol ensures that all reported results reflect genuine predictive capability rather than retrospective pattern fitting. This design constitutes a critical methodological safeguard and strengthens the validity of comparative conclusions drawn between graph-based and rule-based recommendation approaches.

## 3. RESULT AND DISCUSSION

This section presents and analyzes the experimental results of the proposed segmentation-driven graph-based recommendation framework. The evaluation focuses not only on predictive accuracy, but also on recommendation diversity and catalog coverage, which are essential indicators of knowledge discovery capability in sparse transaction environments. All experiments are conducted under a strict temporal evaluation protocol to ensure realistic generalization and to prevent data leakage.

### 3.1. Comparative Recommendation Performance Across Clusters

To facilitate a direct comparison between graph-based and rule-based recommendation performance, Figure 5 summarizes the evaluation results across clusters under the Hold-out Evaluation Set, while Table 3 reports the corresponding quantitative metrics within the Top 100 item space per cluster.

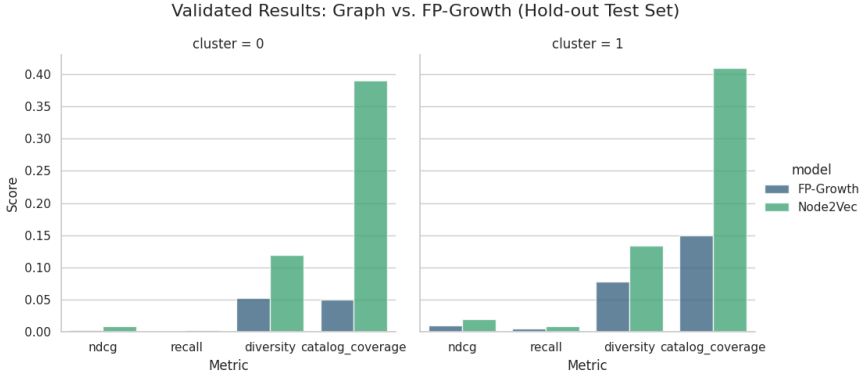


**Figure 5.** Validated Performance Comparison per Cluster (Graph vs FP-Growth)

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

The numerical results corresponding to Figure 5 are reported in Table 3.

**Table 3.** Cluster-wise Validated Recommendation Performance (Hold-out Evaluation Set)

| Cluster | Model | Precision | Recall | NDCG | Diversity | Lift | Catalog Coverage |
|---|---|---|---|---|---|---|---|
| 0 | FP-Growth | 0.004000 | 0.001000 | 0.002921 | 0.052569 | 1.039457 | 0.050000 |
| 0 | Node2Vec | 0.008000 | 0.002333 | 0.008559 | 0.119305 | 1.000000 | 0.390000 |
| 1 | FP-Growth | 0.008000 | 0.005000 | 0.009407 | 0.078072 | 1.083880 | 0.150000 |
| 1 | Node2Vec | 0.024000 | 0.008881 | 0.019703 | 0.133883 | 1.000000 | 0.410000 |

Across clusters, FP-Growth exhibits strong sensitivity to sparse co-occurrence structures, particularly in Cluster 0, where precision and recall remain extremely low even within the constrained Top 100 item space. This indicates that rule-based mining is unable to reliably extract stable association rules when purchase behavior is highly fragmented. In contrast, Node2Vec consistently improves ranking quality across all clusters, with substantial gains in NDCG and more stable performance under sparse conditions.

The degradation of FP-Growth in Cluster 0 is especially notable, as it reflects the inability of minimum-support-based mining to generate meaningful rules when co-purchase repetition is insufficient. This limitation persists even under head-item restriction, demonstrating that sparsity operates at the behavioral level rather than solely at the catalog scale. In contrast, Node2Vec maintains robustness by leveraging higher-order graph connectivity, allowing indirect relational signals to contribute to embedding formation.

Importantly, improvements in catalog coverage and diversity must be interpreted strictly within the localized Top 100 item boundary per cluster, not the global product catalog. Within this constrained head space, Node2Vec achieves substantially higher coverage, indicating more balanced utilization of frequently purchased items rather than over-concentration on a small subset of dominant products. Similarly, higher diversity reflects broader exploration within structurally related high-frequency items.

### 3.2. Diversity and Catalog Coverage Analysis

To further investigate the exploratory behavior of the recommendation models, Figure 6 presents a comparative visualization of diversity and catalog coverage differences between FP-Growth and Node2Vec across customer clusters, evaluated strictly within the Top 100 item space per cluster under the Hold-out Evaluation Set. The figure highlights how graph-based representation learning modifies the distributional characteristics of recommended items compared to rule-based mining, particularly in how frequently purchased (head) items are distributed within localized recommendation lists.

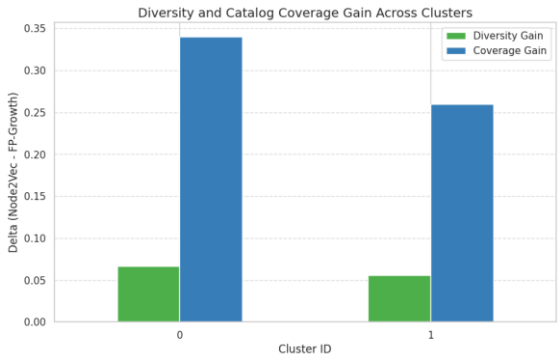


**Figure 6.** Diversity and Catalog Coverage Gain Across Clusters

The corresponding quantitative improvements are summarized in Table 4, which reports the absolute deltas in diversity and catalog coverage achieved by Node2Vec relative to FP-Growth.

**Table 4.** Diversity and Coverage Gain of Graph Model over FP-Growth

| Cluster | Diversity Delta | Coverage Delta |
|---|---|---|
| 0 | +0.066735 | +0.34 |
| 1 | +0.055810 | +0.26 |

Across both clusters, Node2Vec consistently improves diversity and catalog coverage within the constrained Top 100 item boundary. These gains do not indicate expansion into a global product catalog, but rather reflect more balanced utilization of frequently purchased items within each cluster-specific recommendation space. In other words, improvements correspond to reduced over-concentration on a small subset of head items and more even distribution across high-frequency products.

Cluster 0 shows the strongest effect in coverage (+0.34), which is consistent with its higher behavioral sparsity. In this setting, FP-Growth tends to generate narrow and repetitive rule sets due to insufficient co-occurrence structure, whereas Node2Vec leverages higher-order graph connectivity to propagate similarity through anchor-mediated

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

relationships, enabling broader item activation within the Top 100 space. This results in more diversified recommendations despite limited transactional signals.

Cluster 1 exhibits smaller but stable gains, indicating that even in relatively denser behavioral conditions, FP-Growth remains biased toward high-support item combinations. Node2Vec mitigates this concentration by distributing relevance across connected neighborhoods in the embedding space, leading to improved diversity without sacrificing ranking structure.

Overall, these results highlight a fundamental difference between the two approaches under localized head-item constraints: FP-Growth is restricted by explicit co-occurrence frequency, leading to redundancy in recommendation outputs, while Node2Vec captures latent structural relationships that enable smoother distribution across related items. Importantly, all improvements should be interpreted strictly as enhanced localized discovery within the Top 100 item space, not as global catalog expansion.

### 3.3. Structural Interpretation via Anchor Products

To provide an explanatory basis for the observed performance differences between FP-Growth and Node2Vec, a structural analysis of the cluster-specific item graphs is conducted. This analysis focuses on identifying anchor products, defined as high-centrality nodes that function as connectivity hubs within the PPMI-weighted item graphs. These anchor items are critical because they mediate information flow between otherwise weakly connected product groups, thereby shaping the structure of learned embeddings and downstream recommendation behavior.

Figure 7 presents a visualization of a representative PPMI-weighted item graph for a selected customer cluster, illustrating the emergence of highly connected anchor products within the localized head-item space.

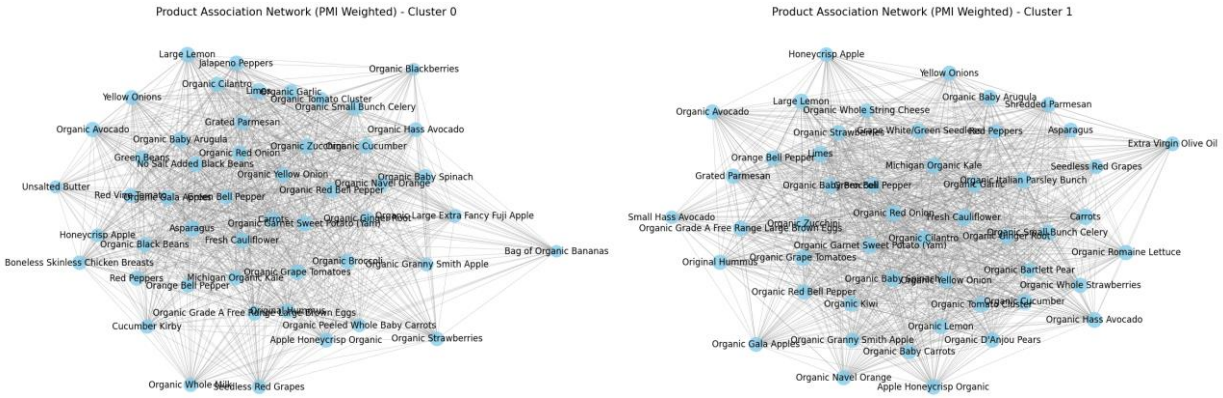


**Figure 7.** PMI-Weighted Item Graph Visualization for a Representative Cluster

The visualization in Figure 7 highlights that the cluster-specific graphs are not uniformly connected but instead exhibit a hub-and-spoke structure, where a small subset of products concentrates a disproportionately large number of connections. These hub nodes correspond to frequently co-purchased staple goods, particularly fresh produce and organic grocery items, which dominate interaction density within the Top 100 item constraint. The top anchor products identified through degree centrality analysis are summarized in Table 5.

**Table 5.** Top Anchor Products Identified via Graph Centrality Analysis

| Cluster | Representative Anchor Products | Degree |
|---|---|---|
| 0 | Organic Grape Tomatoes | 0.707071 |
| 0 | Organic Baby Spinach | 0.696970 |
| 0 | Limes | 0.666667 |
| 0 | Asparagus | 0.666667 |
| 0 | Organic Avocado | 0.656566 |
| 0 | Large Lemon | 0.646465 |
| 0 | Original Hummus | 0.646465 |
| 0 | Fresh Cauliflower | 0.646465 |
| 0 | Jalapeno Peppers | 0.646465 |
| 0 | Organic Gala Apples | 0.646465 |
| 1 | Organic Baby Spinach | 0.858586 |
| 1 | Organic Garlic | 0.808081 |
| 1 | Fresh Cauliflower | 0.808081 |
| 1 | Organic Zucchini | 0.79798 |
| 1 | Carrots | 0.79798 |
| 1 | Organic Red Onion | 0.79798 |
| 1 | Organic Italian Parsley Bunch | 0.79798 |

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

| Cluster | Representative Anchor Products | Degree |
|---|---|---|
| 1 | Organic Grape Tomatoes | 0.787879 |
| 1 | Original Hummus | 0.787879 |
| 1 | Michigan Organic Kale | 0.787879 |

Table 5 shows that anchor products are consistently high-frequency staple goods, especially fresh produce and basic grocery ingredients. These items act as structural bridges because they co-occur with many other products, resulting in high degree centrality and strong connectivity within the PPMI-weighted network.

In Node2Vec, these anchors play a central role during biased random walks, where they are frequently visited and therefore strongly influence embedding learning. This produces two effects: (1) more stable representations due to repeated contextual exposure, and (2) transitive similarity propagation, where items not directly co-purchased become related through shared anchor neighborhoods.

This mechanism explains why Node2Vec outperforms FP-Growth in the Top 100 localized item space. FP-Growth relies strictly on explicit co-occurrence and minimum support thresholds, meaning that anchor products only contribute when direct rules exist. In contrast, Node2Vec leverages their structural role in the graph, enabling indirect relationships to be captured even without explicit co-purchase frequency.

Anchor products are also relatively consistent across clusters, although their influence is stronger in Cluster 0 due to higher sparsity. In Cluster 1, anchors coexist with more moderately connected items, resulting in a less centralized structure. This difference helps explain why Node2Vec shows stronger relative gains in sparser behavioral settings.

Overall, anchor nodes expand the effective reach of the embedding space by connecting multiple item neighborhoods, allowing Node2Vec to explore a broader portion of the high-frequency Top 100 item graph compared to FP-Growth, which remains restricted to direct co-occurrence patterns.

### 3.4. Overall Model Comparison and Robustness Evaluation

To provide a consolidated view of recommendation effectiveness under the Hold-out Evaluation Set, Figure 8 illustrates the cumulative hit rate across different recommendation list lengths (K), enabling an assessment of how quickly each model retrieves relevant items within the cluster-specific Top 100 item space. This metric is particularly relevant for e-commerce systems, where early-ranked relevance strongly affects user interaction quality.

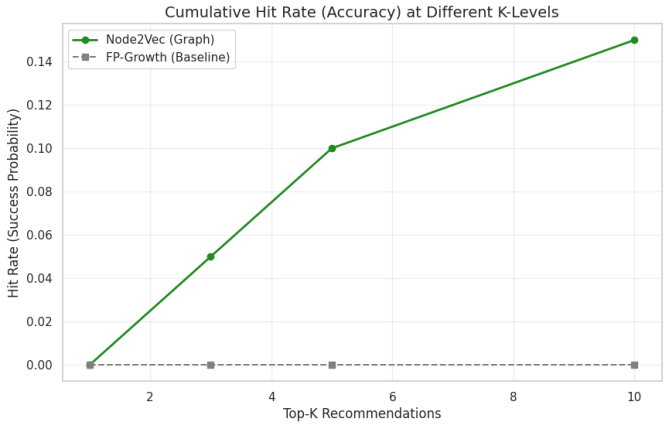


**Figure 8.** Cumulative Hit Rate at Different Recommendation List Lengths

Figure 8 shows a clear separation between the two approaches. The Node2Vec-based model begins retrieving relevant items at K=3 and steadily improves as K increases, while FP-Growth maintains a zero hit rate across all evaluated cutoffs. This indicates that FP-Growth is unable to generate effective ranked recommendations under the current sparse and constrained evaluation setting, even within the Top 100 item space.

This limitation stems from FP-Growth's dependence on exact co-occurrence rules derived from the Historical Training Set, which becomes restrictive under sparse transactional conditions. In contrast, Node2Vec leverages graph-based embeddings that allow similarity inference beyond explicit co-purchase patterns, enabling partial matching even when direct rules are absent. The overall performance across clusters is summarized in Table 6.

**Table 6.** Overall Performance Comparison Across All Clusters

| Model | Precision | Recall | NDCG | Diversity | Catalog Coverage |
|---|---|---|---|---|---|
| FP-Growth | 0.006 | 0.003 | 0.006164 | 0.065321 | 0.10 |
| Node2Vec | 0.016 | 0.005607 | 0.014131 | 0.126594 | 0.40 |

Across all ranking metrics, Node2Vec consistently outperforms FP-Growth. Precision improves by approximately 2.7 times, recall by 1.9 times, and NDCG more than doubles, indicating stronger ranking quality under

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

the Hold-out Evaluation Set. These improvements confirm that graph-based embeddings provide more informative item representations than rule-based co-occurrence mining in sparse behavioral settings.

Improvements in diversity and catalog coverage are also significant within the localized Top 100 item boundary per cluster. Node2Vec nearly doubles diversity and increases coverage from 0.10 to 0.40, indicating more balanced utilization of frequently purchased items rather than concentration on a narrow subset of head products. These gains should be interpreted strictly as improved utilization of the constrained item space, not as expansion of the global catalog.

From a structural perspective, Node2Vec benefits from higher-order connectivity in the PPMI-weighted graph, allowing similarity signals to propagate through multi-hop relationships. FP-Growth, by contrast, is limited to explicit co-occurrence patterns and cannot generalize beyond observed frequent itemsets in the Historical Training Set, which explains its zero-hit behavior in Figure 8.

Regarding robustness, Node2Vec is more stable under sparse interaction conditions because it leverages graph connectivity rather than strict frequency thresholds. This makes it more resilient in sparse clusters (e.g., Cluster 0), where FP-Growth often fails to generate sufficient rules. In such cases, anchor-mediated connectivity in the graph preserves meaningful structure, allowing Node2Vec to maintain consistent recommendation performance across heterogeneous clusters without manual threshold tuning.

### 3.5. Discussion

The experimental results demonstrate that customer segmentation combined with cluster-specific graph modeling provides a consistent advantage over global rule-based recommendation strategies. Across all clusters, the improvements obtained by Node2Vec are primarily driven by more effective modeling of localized purchasing structures within the constrained Top 100 high-frequency item space, rather than any expansion beyond this head-item boundary. Accordingly, the observed gains should be interpreted as improved organization and retrieval quality among frequently purchased products under behaviorally segmented conditions.

Improvements in diversity and catalog coverage are particularly important in this context because they are computed strictly within the cluster-specific Top 100 item set, not the global product catalog. Within this localized space, Node2Vec distributes recommendations more evenly across high-frequency items, reducing over-concentration on a limited subset of dominant products. This effect is consistent with the structural role of PPMI-weighted graphs, where anchor products act as hubs that propagate relational signals across multiple item neighborhoods.

In contrast, FP-Growth is constrained by its reliance on explicit co-occurrence counts and minimum support thresholds. In sparse clusters, this leads to weak or incomplete rule generation even among frequent items, resulting in limited coverage and less diverse recommendations. This limitation persists even under head-item restriction, indicating that sparsity remains a structural issue at the behavioral level, not only at the catalog level.

From a modeling perspective, Node2Vec outperforms FP-Growth because it captures both direct and higher-order relationships through biased random walks on the item graph. This allows the model to infer similarity between items that are not directly co-purchased but are connected through shared neighbors. FP-Growth, by contrast, cannot leverage such transitive structure and remains restricted to first-order co-occurrence patterns.

The role of anchor products further explains these differences. High-centrality items such as staple grocery products connect multiple item communities, enabling Node2Vec to propagate similarity signals across the graph. FP-Growth does not exploit this structural property, which limits its ability to generalize beyond frequent item pairs.

From a practical standpoint, graph-based models are more suitable for exploratory segments where user behavior is less repetitive and recommendation value depends on discovering relevant alternatives within frequently purchased items. FP-Growth remains useful in highly repetitive scenarios but is less effective under sparse or heterogeneous conditions.

Overall, integrating behavioral segmentation with graph-based representation learning produces a more robust recommendation framework that improves localized discovery, enhances utilization of high-frequency items, and better handles sparsity compared to rule-based association mining.

## 4. CONCLUSION

This study proposed a segmentation-driven graph-based recommendation framework to address the limitations of traditional market basket analysis in sparse e-commerce environments by first partitioning users into behaviorally coherent groups and then learning item relationships within each cluster using graph-based representation learning under a strict temporal evaluation protocol. The experimental results under the Hold-out Evaluation Set demonstrate that the Node2Vec-based approach consistently outperforms FP-Growth in ranking quality and in discovery-oriented metrics such as diversity and catalog coverage; however, these improvements are explicitly measured within a constrained Top 100 high-frequency item space per cluster, meaning that all gains reflect improved localized discovery among head products rather than broader catalog expansion. Cluster-level analysis further shows that FP-Growth becomes unstable in sparse behavioral segments due to insufficient co-occurrence structure for reliable rule extraction, whereas PMI-weighted graph embeddings remain more robust by leveraging higher-order and transitive relationships mediated through anchor products in the item graph. A key methodological limitation of this study is the intentional

**Building of Informatics, Technology and Science (BITS)**
Volume 8, No 1, June 2026 Page: 83−94
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9707

restriction of the candidate item universe to the Top 100 items per cluster, which improves computational tractability and focuses analysis on dominant purchasing behavior but inherently excludes modeling of lower-frequency items and limits generalization to full-catalog recommendation scenarios. Additionally, user subsampling and the use of static embedding representations constrain the scope of behavioral coverage and temporal adaptivity. Overall, despite these limitations, the proposed framework provides a rigorous and interpretable approach for localized product discovery within high-frequency product segments in e-commerce systems, offering a structured alternative to rule-based association mining under sparse transactional conditions.

# REFERENCES

[1] M. Kholod and N. Mokrenko, "Market Basket Analysis Using Rule-Based Algorithms and Data Mining Techniques," *arXiv preprint*, 2024, doi: 10.48550/arxiv.2412.18699.
[2] M. Rahman, S. Mushfik, M. A. Rupak, M. N. Hasan, M. B. Farukee, and S. K. Suter, "Exploring Challenges and Innovations in E-Commerce Recommendation Systems: A Comprehensive Review," in *Lecture Notes in Computer Science*, 2024, pp. 123–130. doi: 10.1007/978-981-99-9040-5_8.
[3] D. Brancato, "Apriori Versus FP-Growth for Recommendation System," in *Lecture Notes in Computer Science*, 2023, pp. 155–162. doi: 10.1007/978-981-19-9493-7_16.
[4] A. B. Prasetio, "Interpretable Product Recommendation through Association Rule Mining: An Apriori-Based Analysis on Retail Transaction Data," *International Journal of Informatics and Information Systems*, vol. 8, no. 2, pp. 67–74, 2025, doi: 10.47738/ijiis.v8i2.252.
[5] M. Keskin, E. Teper, and A. Kurt, "Comparative Evaluation of Word2Vec and Node2Vec for Frequently Bought Together Recommendations in E-Commerce," in *Proceedings of the 6th International Conference on Computer Science and Engineering (UBMK)*, 2024, pp. 1–5. doi: 10.1109/ubmk63289.2024.10773398.
[6] K. Ammar, W. Inoubli, S. Zghal, and E. M. Nguifo, "Graph Representation Learning for Recommendation Systems: A Short Review," in *Lecture Notes in Computer Science*, 2024, pp. 33–48. doi: 10.1007/978-3-031-51664-1_3.
[7] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural Graph Collaborative Filtering," in *Proceedings of the 42nd International ACM SIGIR Conference*, 2019, pp. 165–174. doi: 10.1145/3331184.3331267.
[8] Y. Wu, X. He, Y. Wang, Y. Liu, and M. Wang, "A Survey on Graph-Based Recommender Systems," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–37, 2023, doi: 10.1145/3535101.
[9] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference*, 2018, pp. 1831–1839. doi: 10.1145/3219819.3219950.
[10] C. Cai, H. Chen, Y. Liu, D. Chen, X. Zhou, and Y. Lin, "Graph-Based Feature Crossing to Enhance Recommender Systems," *Mathematics*, vol. 13, no. 2, p. 302, 2025, doi: 10.3390/math13020302.
[11] T. Meisen and Y. H., "Instacart Online Grocery Basket Analysis Dataset." [Online]. Available: https://www.kaggle.com/datasets/yasserh/instacart-online-grocery-basket-analysis-dataset
[12] T. Meisen, "A Review on Customer Segmentation Methods for Personalized Customer Targeting in E-Commerce Use Cases," *Information Systems and e-Business Management*, 2023, doi: 10.1007/s10257-023-00640-4.
[13] K. Yuliandari, R. Anggraini, and B. Amaliah, "Cluster-Aware Next-Basket Recommendation Using a Hybrid Autoencoder-RNN and Graph Transformer Approach," in *Proceedings of ICERA*, 2025, pp. 1–6. doi: 10.1109/icera66156.2025.11087374.
[14] C. Arizmendi and K. M. Gates, "Clustering Individuals Based on Similarity in Idiographic Factor Loading Patterns," *Multivariate Behav. Res.*, pp. 1–25, 2024, doi: 10.1080/00273171.2024.2374826.
[15] A. M. Jørgensen, "Exploratory Analysis of Grocery Product Networks," *Journal of Management Analytics*, vol. 9, no. 2, pp. 169–184, 2022, doi: 10.1080/23270012.2022.2072779.
[16] S. Balcisoy and B. Bozkaya, "A Link Prediction-Based Recommendation System Using Transactional Data," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-34055-5.
[17] M. M. Lawal and O. T. Matthew, "FP-Growth Algorithm: Mining Association Rules without Candidate Sets Generation," *KASU Journal of Computer Science*, vol. 1, no. 2, pp. 392–411, 2024, doi: 10.47514/kjcs/2024.1.2.0016.
[18] F. Nuraeni, D. Tresnawati, Y. H. Agustin, and G. Fauzi, "Optimization of Market Basket Analysis Using Centroid-Based Clustering Algorithm and FP-Growth Algorithm," *Jurnal Teknik Informatika*, vol. 3, no. 6, pp. 1581–1590, 2022, doi: 10.20884/1.jutif.2022.3.6.399.
[19] A. Jadon and A. Patil, "A Comprehensive Survey of Evaluation Techniques for Recommendation Systems," *arXiv preprint*, 2023, doi: 10.48550/arxiv.2312.16015.
[20] B. Alhijawi, A. Awajan, and S. Fraihat, "Survey on the Objectives of Recommender Systems: Measures, Solutions, Evaluation Methodology, and New Perspectives," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–38, 2022, doi: 10.1145/3527449.