CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (2) (2026), Pages 1969–1986
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Consensus Bayesian Network Learning for Market Basket Analysis in Indonesian Retail Transactions

Muhammad Sahid*, I Gede Nyoman Mindra Jaya, and Restu Arisanti

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia

Abstract

Understanding relationships among products in retail transactions was essential for supporting strategic decisions such as product placement and cross-selling. However, traditional market basket analysis was limited to pairwise associations and could not capture complex conditional dependencies among multiple items. This study presented an integrated, stability-aware Bayesian Network framework to model product relationships probabilistically using transaction data from an Indonesian retail outlet. Transaction data collected over one month were transformed into a binary dataset of 441 transactions across 31 product categories. The network structure was learned using a stochastic search algorithm with a Bayesian Dirichlet equivalent score, enhanced by a multi-seed strategy and edge support aggregation. Parameters were estimated using maximum likelihood, and additional sensitivity analyses were conducted to assess the stability of the core dependencies to prior specification, directional constraints, and parameter smoothing. The results indicated that relationships were both directional and context-dependent, with snack probabilities increasing with instant noodles but decreasing when cigarettes were also present. These core dependencies remained stable across all sensitivity checks. The LOIO accuracy was 0.6659, while the ROC-AUC was 0.5235, which was consistent with the model's focus on structural discovery rather than classification performance.

Keywords: Bayesian network; market basket analysis; probabilistic modelling; retail analytics; structure learning.

1. Introduction

The rapid growth of the retail industry leads to an increasing volume of transaction data, creating opportunities for data-driven decision-making [1, 2]. In modern retail environments, understanding consumer purchasing behaviour is essential for optimizing product placement, designing promotional strategies, and improving cross-selling performance. Transaction data, particularly in the form of shopping baskets, reflects complex purchasing patterns and interactions among products. As the scale and diversity of retail systems continue to expand, there is a growing need for analytical approaches that can extract structured and meaningful insights from such data [3–5].

Market Basket Analysis (MBA) is widely used to identify relationships among products based on their co-occurrence in transactions [6, 7]. Traditional MBA approaches rely on association-rule

*Corresponding author. E-mail: muhammad24568@mail.unpad.ac.id

Received: April 13, 2026 Revised: July 24, 2026 Accepted: August 28, 2026
DOI: https://doi.org/10.18860/cauchy.v11i2.41879

mining techniques, which utilize measures such as support, confidence, and lift to discover frequent itemsets and simple associations. While these methods are effective for capturing basic co-occurrence patterns, they are inherently limited in representing complex relationships. Specifically, the resulting rules are typically fragmented, lack a global structure, and do not explicitly model conditional dependencies or distinguish between direct and indirect relationships [4]. Additionally, the problem of rule explosion becomes more prominent as the number of product categories increases, reducing interpretability and practical usability.

To address these limitations, Bayesian Network (BN) provides a more comprehensive probabilistic framework for modelling relationships among products [3, 8]. In BN, product categories are represented as nodes in a Directed Acyclic Graph (DAG), where edges encode conditional dependencies. The representation enables the construction of a unified model that captures both direct and indirect relationships, while also supports multivariate probabilistic inference based on the joint distribution implied by the graph [3, 9]. Compared to association-rule approaches, BN offers a more compact and structured representation, allowing for deeper insights into consumer purchasing behaviour. The key differences between these approaches, particularly in terms of analytical requirements, are summarized in Table 1.

Table 1: Comparison of analytical requirements between association-rule MBA and Bayesian-network MBA (adapted from [3, 4])

Analytical requirement	Association-rule MBA	Bayesian-network MBA
Global representation of product relations	Produces separate rules and frequent itemsets	Produces one coherent dependency graph
Explicit conditional dependence	Limited to local conditional frequency summaries	Directly encoded through graph structure and conditional probabilities
Distinction between direct and indirect links	Difficult to establish	Naturally represented through network paths
Multivariate probabilistic inference	Not available as a joint model	Available from the joint distribution implied by the DAG
Sensitivity to rule explosion	High when categories increase	More compact because relations are summarized as edges
Parameter uncertainty and model regularization	Usually absent	Supported by Bayesian scoring and probabilistic estimation

Based on these considerations, this study presents an integrated, stability-aware Bayesian Network framework to model the relationships among product categories in retail transaction data. The framework combines a multi-seed stochastic search with edge support aggregation and Markov blanket reduction to ensure robust and interpretable structure learning [10]. Beyond producing a coherent representation of product relationships, the resulting model enables probabilistic interpretation through Conditional Probability Tables, offering both analytical insights and practical value for retail decision-making. The contribution of this work lies in the application and integration of these established techniques to sparse retail transaction data, providing a practical pipeline for extracting stable dependency structures from real-world basket data.

2. Methods

The research was based on primary retail transaction data collected from one of Indomaret outlets in Tasikmalaya, Indonesia, over a 32-day observation period. The dataset represented the complete population of all transactions recorded at the outlet during this period, thereby avoiding sampling bias and ensuring that the observed purchasing patterns reflect the actual behaviour of consumers at this location. The data were recorded at the transaction level and subsequently aggregated into product categories to align with the objective of modelling category-level dependencies rather than SKU-level variations. The original 1,540 SKUs were manually grouped into 65 product categories based on functional similarity and retail convention. For example, different brands and package sizes of instant noodles were consolidated under the single category MIE_INSTANT, while various snack products were aggregated under SNACK. This aggregation step was essential to reduce dimensionality and sparsity, enabling meaningful probabilistic modelling at the category level. The preprocessing stage played a critical role in ensuring that the data structure was suitable for probabilistic modelling, particularly in reducing sparsity and improving the reliability of conditional probability estimation [3].

At the initial stage, the dataset contained 515 transactions, 2,745 item lines, and 65 aggregated product categories. Two filtering rules were applied to improve data quality. First, categories appearing in fewer than 20 transactions were removed to mitigate extreme sparsity, as rare categories could destabilize both structure learning and conditional probability estimation [8, 11]. The choice of this threshold was guided by the stability of conditional probability estimates. Under Maximum Likelihood Estimation (MLE), the conditional probability parameter for category i taking value k given parent configuration j was estimated as the empirical proportion:

$$\hat{\theta}_{ijk} = \frac{N_{ijk}}{N_{ij}}, \tag{1}$$

where N_{ijk} denotes the number of transactions in which $X_i = k$ and its parent configuration is j, and $N_{ij} = \sum_k N_{ijk}$ is the total number of transactions with parent configuration j. The asymptotic variance of this estimator was given by:

$$\mathrm{Var}(\hat{\theta}_{ijk}) = \frac{\theta_{ijk}(1-\theta_{ijk})}{N_{ij}}. \tag{2}$$

As shown in (2), the variance was inversely proportional to N_{ij}. Consequently, when a parent configuration is rarely observed, the variance inflated considerably, producing unstable estimates that can undermine the reliability of the learned structure. To illustrate this effect in the present dataset, consider a binary node with two parents, yielding $2^2 = 4$ parent configurations distributed across 441 transactions. Configurations supported by fewer than 20 transactions (e.g., $N_{ij} = 5$) yield standard errors exceeding 0.20 under moderate θ values, whereas configurations meeting the 20-transaction threshold reduce the standard error by more than half. This threshold also ensures that even for nodes with three parents—the maximum allowed in this study—the expected frequency per configuration remains above 10 transactions on average ($441/2^3 = 8$ configurations, yielding approximately 55 transactions per configuration, though unequally distributed in practice), thereby providing a reasonable minimum support for estimation. The threshold of 20 transactions was therefore selected to ensure that each retained category contributes reliable information to the dependency structure, while still preserving a diverse set of 31 product categories for meaningful analysis.

Second, transactions containing only a single category were excluded because they did not provide co-occurrence information. After these steps, the final dataset consisted of 441 transactions and 31 product categories. A summary of the data preparation process was presented in Table 2.

Following preprocessing, the dataset was transformed into a binary matrix representation. Let X_{ij} denote the presence of category j in transaction i, as defined in (3):

$$X_{ij} = \begin{cases} 1, & \text{if item } j \text{ is present in transaction } i, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

As defined in (3), each element of the matrix indicated whether a product appeared in a transaction [12]. To improve data quality, product categories with low frequency were removed, and transactions containing only a single item were excluded, as they did not contribute to the

Table 2: Data preparation summary

No.	Processing stage	Value
1	Unique transactions in the raw data	515
2	Item lines	2,745
3	SKUs	1,540
4	Aggregated product categories before filtering	65
5	Aggregated baskets before single-item filtering	507
6	Minimum category support threshold	20 transactions ($\approx 4\%$)
7	Transactions in the final analytic sample	441
8	Categories in the final analytic sample	31
9	Mean basket size after aggregation	4.80
10	Median basket size after aggregation	4
11	SNACK share in the final dataset	34.01%

analysis of relationships among products. Based on the filtered dataset, let D denoted the set of all transactions and I the set of product categories [13].

To provide a baseline comparison with traditional MBA, standard association-rule measures were also considered. Specifically, the support of an itemset $A \subseteq I$ was defined in (4) as [14, 15]:

$$\text{support}(A) = \frac{|\{t \in D \mid A \subseteq t\}|}{|D|}, \tag{4}$$

and the confidence of a rule $A \to B$, with $A, B \subseteq I$, was defined as:

$$\text{confidence}(A \to B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}. \tag{5}$$

As shown in (4) and (5), the measures captured co-occurrence relationships but did not account for conditional dependencies among multiple variables [6, 16]. Therefore, the research adopted a BN framework [3, 15].

Specifically, a BN was defined as a pair (G, Θ), where $G = (V, E)$ was a Directed Acyclic Graph (DAG) and Θ represented the set of parameters [3]. It should be noted that the learned structure represents one of the equivalent DAGs that encode the same conditional independence relationships. In the formulation, $V = \{X_1, X_2, \ldots, X_p\}$ denoted the set of random variables corresponding to product categories, and $E \subseteq V \times V$ represented the set of directed edges encoding dependencies. For each node X_i, the probabilistic relationship was defined conditionally on its parent set $\text{Pa}(X_i)$, as expressed in equation (6):

$$P(X_i \mid \text{Pa}(X_i)). \tag{6}$$

Equation (6) allowed the joint distribution to be factorized into local conditional probabilities, enabling the model to capture context-dependent purchasing behaviour [17, 18].

The network structure was learned using a score-based stochastic search procedure implemented in the Bayesian Network Inference with Java Objects (BANJO) framework, version 2.2.0. The search was conducted using the Simulated Annealing algorithm with its default annealing schedule, employing an initial temperature of 1000, a cooling factor of 0.7, and a reannealing temperature of 800. Cooling was triggered after a maximum of 2,500 accepted networks or 10,000 proposed networks, whichever occurred first, and reannealing was initiated when the number of accepted networks fell below 500. The RandomLocalMove proposer was used to generate candidate structures through single-edge additions, deletions, and reversals, and the Metropolis acceptance rule determined whether to accept each proposed move.

Candidate structures were evaluated using the Bayesian Dirichlet equivalent (BDe) score with an Equivalent Sample Size (ESS) of $\alpha = 5.0$ [3]. The search was constrained to a maximum of three parents per node ($|\text{Pa}(X_i)| \leq 3$) and was terminated when either a maximum wall-clock time of 2 hours was reached or 2,000,000 candidate networks had been proposed. To avoid premature convergence, the algorithm was allowed up to 40 restarts, and the top 100 non-identical networks were retained from each run. A minimum of 1,000 networks was evaluated before any convergence checking was initiated. Computation was accelerated using a precomputed log-Gamma lookup table and a fast level-2 cache. The complete BANJO settings file used in this study is provided as Supplementary Material.

To further ensure robustness against local optima, the structure learning procedure was repeated across $R = 10$ independent runs, each initialized with a distinct random seed set to integer values from 1 to 10 using the `seedForStartingSearch` parameter. This multi-seed strategy was applied to every training fold in the 10-fold cross-validation as well as to the full-dataset analysis. The resulting set of candidate networks was aggregated using an edge support measure [10], computed as the proportion of runs in which a given edge appeared among the $R = 10$ executions. Only edges with support of at least 0.60 were retained to construct the final consensus network. The threshold of 0.60 was selected to balance structural stability and network sparsity, as lower thresholds resulted in dense and less interpretable structures, while higher thresholds removed meaningful dependencies. To ensure the consensus graph remained a valid Directed Acyclic Graph (DAG), the retained edges were assembled into a graph and any cycles that emerged from the independent thresholding procedure were resolved by iteratively removing the edge with the lowest support value within each detected cycle until acyclicity was achieved. In addition, outgoing edges from the SNACK category were restricted throughout the entire learning process to preserve its role as a dependent outcome rather than a driver variable.

The BDe score was a closed-form expression of the marginal likelihood $P(D \mid B_S)$ obtained by integrating out the conditional probability parameters under a Dirichlet prior [3], and it penalized model complexity implicitly through the marginalization over parameter space. For a network structure B_S with p nodes, the BDe score was given by:

$$P(D \mid B_S) = \prod_{i=1}^{p} \prod_{j=1}^{q_i} \left[\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \right] \prod_{k=1}^{r_i} \left[\frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})} \right], \tag{7}$$

where p denoted the number of nodes, q_i is the number of unique parent configurations for node X_i, $r_i = 2$ is the number of states for each binary variable, N_{ijk} is the count of transactions in which $X_i = k$ and its parent configuration is j, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$ is the total number of transactions with parent configuration j, α_{ijk} are the Dirichlet prior hyperparameters, $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$, and $\Gamma(\cdot)$ is the Gamma function.

Under the Bayesian Dirichlet equivalent uniform (BDeu) assumption, the hyperparameters were distributed uniformly across all states and parent configurations, yielding $\alpha_{ijk} = \alpha/(r_i \cdot q_i)$, where α is the Equivalent Sample Size (ESS) [3]. With binary variables ($r_i = 2$), this simplified to:

$$\alpha_{ijk} = \frac{\alpha}{2q_i}. \tag{8}$$

In this study, the ESS was set to $\alpha = 5$. This choice provided moderate regularization: the value is sufficiently small relative to the sample size of 441 transactions to avoid dominating the observed data, yet large enough to prevent extreme probability estimates of zero or one for rare parent configurations, which could otherwise destabilize the structure learning process. To evaluate the stability of the learned structure to the choice of prior, the full-data structure learning was repeated with equivalent sample sizes of 1, 5, 10, and 20, while retaining the same reduced variable set. The BDe score with this prior specification was used as the objective function in the Simulated Annealing search to identify the highest-scoring network structure. The stochastic optimization strategy was adapted from a robust strategy framework [10] to ensure stability in the search process.

To control model complexity and maintain interpretability, the maximum number of parents per node was restricted to three ($|\text{Pa}(X_i)| \leq 3$). This restriction serves as a structural regularization mechanism tailored to the sample size. For binary variables, the number of CPT parameters required for a node X_i grows exponentially with the number of its parents. Let $|\text{Pa}(X_i)|$ denote the number of parents for node X_i. The number of parameters for node X_i is:

$$\text{Parameters}(X_i) = 2^{|\text{Pa}(X_i)|}. \tag{9}$$

With the three-parent limit, each node requires at most $2^3 = 8$ parameters. Consequently, the total number of parameters in the network with $p = 31$ nodes is bounded above by:

$$\text{Total Parameters} \leq \sum_{i=1}^{31} 2^{|\text{Pa}(X_i)|} \leq 31 \times 2^3 = 248. \tag{10}$$

As indicated in (10), the maximum total number of parameters is 248. Given the sample size of 441 transactions, the observation-to-parameter ratio is approximately $441/248 \approx 1.78$, which provides adequate empirical support for parameter estimation. Without this restriction, the parameter count could grow substantially. As shown in (9), a single node with four parents would require $2^4 = 16$ parameters, and a node with five parents would require $2^5 = 32$ parameters, potentially pushing the total parameter count beyond 500 and causing the observation-to-parameter ratio to drop below 1. In such a scenario, the model would be at high risk of overfitting, producing spurious dependencies that do not generalize beyond the observed transactions. The three-parent limit is consistent with the principle of structural regularization in Bayesian Network learning, where model complexity is explicitly constrained to balance goodness-of-fit with generalizability [3, 8].

To further ensure robustness, the learning process was repeated across multiple random seeds. The resulting set of candidate networks was aggregated using an edge support measure [10]. Edge support was computed as the relative frequency of an edge across multiple runs of the structure learning procedure, as defined in (11) [16]:

$$\text{ES}(E) = \frac{\sum_{r=1}^{R} \mathbb{I}_r(E)}{R}. \tag{11}$$

As indicated in (11), edges with higher support values were more stable across runs. In the research, only edges with support of at least 0.60 were retained to construct the final consensus network. The threshold of 0.60 was selected to balance structural stability and network sparsity, as lower thresholds resulted in dense and less interpretable structures, while higher thresholds removed meaningful dependencies.

In addition, outgoing edges from the SNACK category were prohibited throughout the structure learning process. This constraint reflects the substantive nature of SNACK as a snack food category: in a convenience-store setting, snack purchases are typically impulsive or complementary to other purchases rather than being the primary driver of a shopping trip. It is substantively implausible that the purchase of snack items would *cause* the purchase of categories such as baby diapers (POPOK), dish soap (SABUN_PIRING), or cigarettes (ROKOK). From a methodological standpoint, this constraint also addresses a known limitation of score-based Bayesian Network learning from observational data. The BDe score is Markov equivalent: it assigns identical scores to structures that represent the same conditional independence relationships, even when they differ in edge direction. Without the constraint, the search algorithm could arbitrarily direct edges away from SNACK, producing a network that is statistically equivalent but substantively less interpretable. By fixing SNACK as a terminal node, the constraint resolves this ambiguity in favour of a structure that aligns with domain knowledge. It is important to note that this constraint does not predetermine which variables appear as parents of SNACK; those relationships remain entirely data-driven, as evidenced by the variation in SNACK's parent sets observed across cross-validation folds (see Table 3). The influence of this constraint on the resulting dependency structure was further investigated by comparing the constrained network to an unconstrained version in which outgoing edges from SNACK were permitted.

After the network structure was established, the model parameters were estimated using Maximum Likelihood Estimation (MLE) based on empirical relative frequencies [3], as defined in (6). MLE was chosen for CPT estimation for two reasons. First, the primary objective of the CPTs in this study was descriptive. They served to illustrate the conditional dependence relationships encoded in the learned network structure, rather than to support out-of-sample prediction or inference in a production system. For this descriptive purpose, the empirical frequencies provided a direct and interpretable summary of the observed co-occurrence patterns. Second, the structure learning phase already employed a Bayesian approach through the BDe score with a Dirichlet prior (ESS = 5), which regularizes the search process and penalizes overly complex structures. The CPTs estimated via MLE thus reflect the dependencies present in the data after the structure has been regularized.

It was acknowledged that MLE can produce extreme probability estimates of zero or one for parent configurations with limited observations. In the present dataset, this occurred for one configuration in the CPT of SABUN_PIRING (Fig. 1), where $P(\text{SABUN\_PIRING} = 1 \mid \text{MIE\_INSTANT} = 0, \text{ROKOK} = 1) = 0.000$, reflecting the absence of this combination in the 441 transactions. To assess the robustness of the parameter estimates to zero-frequency cells, the CPTs were additionally re-estimated using Laplace smoothing with a pseudo-count of one, and the resulting probabilities and predictive performance were compared to the MLE estimates. This sensitivity analysis is reported in the Results section.

For each variable X_i and parent configuration u, the conditional probability was defined as shown in (12):

$$P(X_i = x \mid \text{Pa}(X_i) = u) = \frac{N(x, u)}{N(u)}. \tag{12}$$

As expressed in (12), the conditional probability tables (CPT) quantified how the presence of certain categories influenced others, particularly for the SNACK category as the focal outcome.

Model evaluation was conducted using a two-stage validation strategy. First, stratified 10-fold cross-validation was applied to assess structural stability by examining variations in network characteristics across folds [8]. In each transaction, the value of SNACK was treated as unknown, and the values of its parent variables in the learned network (MIE_INSTANT and ROKOK) were used as evidence. The predicted probability $P(\text{SNACK} = 1 \mid \text{MIE\_INSTANT}, \text{ROKOK})$ was obtained from the conditional probability table of SNACK. If this probability exceeded 0.6, SNACK was predicted as present; otherwise, it was predicted as absent. The LOIO accuracy was then computed as the proportion of transactions for which the predicted presence or absence of SNACK matched the actual value, as defined in (13):

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}}. \tag{13}$$

This per-transaction accuracy reflects the model's ability to infer a hidden item given the observed context of its parent categories. In addition to LOIO accuracy, the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) was evaluated as a complementary diagnostic measure. Using the predicted probabilities $P(\text{SNACK} = 1)$ from the CPT without applying a decision threshold, the ROC-AUC quantified the model's ability to discriminate between transactions that contain SNACK and those that do not. However, as the primary objective of this study is to model dependency structures rather than to optimize classification performance, the ROC-AUC served a secondary, illustrative role.

To provide a quantitative benchmark for the Bayesian Network, an independence model was constructed as a baseline. The independence model assumed that all product categories are mutually independent, representing the limiting case of association-rule methods that capture pairwise co-occurrence without modelling multivariate dependencies. Under this model, the probability of a transaction $X^{(t)}$ is computed as the product of the marginal probabilities of each category:

$$P_{\text{Ind}}(X^{(t)}) = \prod_{j=1}^{p} \hat{p}_j^{x_{tj}} (1 - \hat{p}_j)^{1-x_{tj}}, \tag{14}$$

where $\hat{p}_j$ is the empirical frequency of category j in the training data, $x_{tj} \in \{0,1\}$ indicates the presence or absence of category j in transaction t, and $p = 31$ is the number of product categories. Both the BN and the independence model were evaluated using the average predictive log-likelihood on the held-out test sets from the 10-fold cross-validation:

$$\text{Predictive Log-Likelihood} = \frac{1}{N_{\text{test}}} \sum_{t \in \text{Test}} \log P(X^{(t)} \mid \text{Model}), \tag{15}$$

where N_{test} denotes the number of transactions in the test set. The predictive log-likelihood quantifies how well a model predicts unseen data, with higher values (closer to zero) indicating better predictive performance. The results of this comparison are presented in Table 6. All log-likelihood computations for both the Bayesian Network and the independence model were performed on the identical set of 31 product categories across all folds. To avoid undefined values when computing logarithms for extremely small probabilities, a small constant $\varepsilon = 10^{-10}$ was added to every probability term, ensuring that $\log(0)$ never occurred.

Finally, to enhance interpretability and focus the analysis on the most relevant dependencies, the model was refined using a Markov blanket reduction strategy centered on the SNACK category [3]. In a BN, the Markov blanket of a node X_i, denoted as $\text{MB}(X_i)$, consisted of its parents, its children, and the parents of its children [19]. The set represented the minimal subset of variables that rendered X_i conditionally independent of all other variables in the network, as expressed:

$$X_i \perp V \setminus (\text{MB}(X_i) \cup \{X_i\}) \mid \text{MB}(X_i). \tag{16}$$

The property implied that, given its Markov blanket, no additional variables contributed further information about X_i. In the research, cross-validated structure learning indicated that SNACK was embedded within a localized dependency structure. Therefore, the analysis focused on its Markov blanket and second-degree neighbors to construct a reduced network that preserved the most relevant relationships while eliminated less informative variables. The reduction improved interpretability while maintained the essential probabilistic structure of the model [20, 21].

3. Results and Discussion

The filtered dataset exhibited a typical retail distribution: a small number of categories appeared frequently while the majority remained rare, justifying the 20-transaction support threshold. Within the dataset, SNACK emerged as the most prominent category, appearing in 150 out of 441 transactions (34.01%). The remaining 291 transactions did not contain SNACK, resulting in a class distribution that was sufficiently balanced for validation procedures such as stratified cross-validation. Moreover, the average basket size of 4.80 items indicated that transactions typically involved multiple categories, supporting the use of a multivariate modelling framework rather than pairwise association analysis [6, 16].

The structure learning process revealed that the global network configuration was relatively stable across validation folds, with 28–30 nodes and 42–51 edges per fold [8]. Central dependencies remained stable while peripheral relations varied with sample composition. The local structure around SNACK provided deeper insight: ROKOK consistently appeared as a parent of SNACK in eight out of ten folds, while MIE_INSTANT appeared in four, suggesting that these categories formed a stable component of the dependency structure surrounding SNACK even before the final model was estimated. The Markov blanket of SNACK was compact, ranging from one to three variables [3], indicating that its probabilistic behaviour was governed by a small, interpretable subset of categories and supporting the feasibility of model reduction without significant loss of information.

From a predictive perspective, the model demonstrated moderate performance in reconstructing missing items within transactions [3, 22]. The LOIO accuracy reached 0.6659, indicating that approximately two-thirds of hidden items could be correctly inferred from the remaining basket context. In contrast, the ROC-AUC value was only slightly above 0.50, suggesting limited discriminative ability as a binary classifier. The discrepancy highlighted that the model was better suited for capturing dependency structures and contextual relationships rather than optimizing classification performance.

Table 3: Validation and structural diagnostics

Diagnostic	Empirical result	Interpretation
Nodes per cross-validation fold	28 to 30	Core variable set is stable across resamples
Edges per cross-validation fold	42 to 51	Some peripheral relations vary with the training sample
Most frequent SNACK parent across folds	ROKOK in 8 of 10 folds	Strong evidence of local structural relevance
Second most frequent SNACK parent across folds	MIE_INSTANT in 4 of 10 folds	Recurrent but less dominant than ROKOK
SNACK Markov blanket size	1 to 3 variables	Dependence around SNACK is local and interpretable
Most stable cross-validation edge	ROKOK → SNACK, support 0.80	Robust relation across repeated learning runs
LOIO reconstruction accuracy	0.6659	See Table 8 for baseline comparison
ROC-AUC	0.5235	Confirms model is not a classifier; secondary diagnostic

Table 3 summarizes the validation and structural diagnostics. Edge stability analysis confirmed that ROKOK → SNACK was the most robust relation (support = 0.80), followed by ROKOK → KOPI and BUMBU → MIE_INSTANT with lower support values [10].

When the network is re-estimated using the full dataset and subsequently is reduced based on the Markov blanket, the resulting structure became substantially more interpretable [3]. The initial full-sample network contained 29 nodes and 46 edges, which provides a comprehensive representation but is difficult to interpret directly. After reduction, the final network consisted of only 8 nodes and 9 edges, focusing on the most relevant dependencies associated with the SNACK category.

The reduction process effectively filtered out less stable and peripheral relationships while preserving the core dependency structure identified during cross-validation, thereby mitigating the risk of overfitting in sparse transactional data. The contrast between the full network (29 nodes, 46 edges) and the reduced network (8 nodes, 9 edges) highlighted the trade-off between completeness and interpretability: many connections in the full network contributed marginally to explaining SNACK, whereas the reduced network emphasized a smaller subset of variables with stronger, more stable associations. The dependency system surrounding SNACK was organized around a limited number of central nodes—particularly ROKOK and MIE_INSTANT—reinforcing their roles as key variables and justifying the Markov blanket–based reduction.

Overall, the final structure retained only the most informative relationships directly relevant to SNACK. The compactness of this network—from 46 edges to 9—demonstrated that the purchasing behaviour surrounding snack products could be captured by a parsimonious set of conditional dependencies. This finding reinforced the value of the Markov blanket approach as a principled method for variable selection in high-dimensional basket data. Rather than relying on all 31 product categories, the analysis showed that a small subset of interconnected items was sufficient to characterize the probabilistic environment of SNACK. These results also confirmed that the multi-seed consensus strategy successfully isolated stable dependencies from incidental co-occurrences, producing a network structure that was both interpretable and analytically tractable. Fig. 1 showed the reduced BN structure.

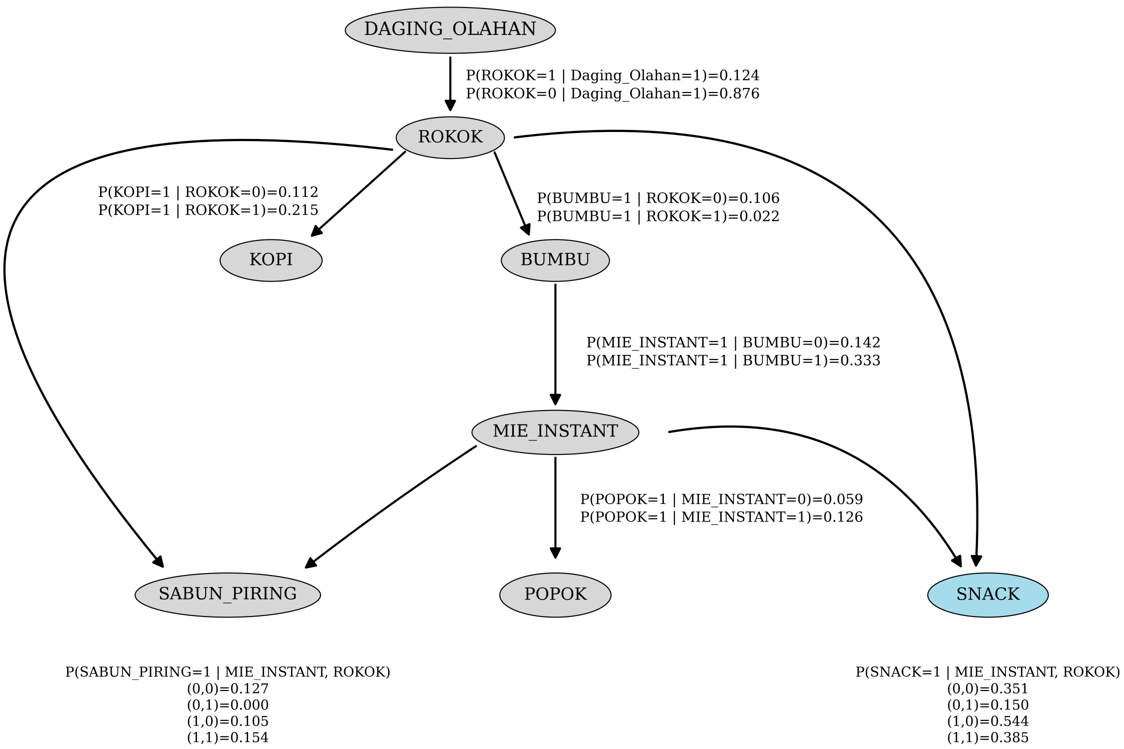

Fig. 1: Reduced Bayesian Network structure highlighting the Markov blanket of SNACK and its second-order dependencies.

In Fig. 1, edge labels represented conditional probabilities for category presence. ROKOK occupied a central position, receiving input from DAGING_OLAHAN and being conditionally associated with BUMBU, KOPI, SABUN_PIRING, and SNACK. MIE_INSTANT served as a bridging node to SNACK and household items.

The probabilistic structure further indicated that the purchasing behaviour associated with SNACK is embedded within a broader consumption context that spans food, beverages, and non-food categories. In particular, the conditional probability of SNACK was substantially higher given the presence of MIE_INSTANT than given the presence of ROKOK. This suggested that MIE_INSTANT was more directly associated with SNACK in terms of conditional dependence, whereas ROKOK, despite its structural centrality in the graph, exhibited a more context-dependent pattern of association.

The most informative result is obtained from the conditional probability table (CPT) of SNACK, which quantifies the joint influence of its parent nodes, MIE_INSTANT and ROKOK. The estimated probabilities revealed that when neither parent is present, the probability of SNACK is 0.3505. The value increases substantially to 0.5439 when MIE_INSTANT is present without ROKOK, indicating a strong positive association between instant noodle purchases and snack consumption.

Table 4: Final reduced BN for the SNACK category

Parent node	Child node
DAGING_OLAHAN	ROKOK
ROKOK	BUMBU
ROKOK	KOPI
BUMBU	MIE_INSTANT
MIE_INSTANT	POPOK
MIE_INSTANT	SABUN_PIRING
ROKOK	SABUN_PIRING
MIE_INSTANT	SNACK
ROKOK	SNACK

Table 5: Conditional probability table for SNACK

MIE_INSTANT	ROKOK	$P(\text{SNACK} = 0)$	$P(\text{SNACK} = 1)$
0	0	0.6495	0.3505
0	1	0.8500	0.1500
1	0	0.4561	0.5439
1	1	0.6154	0.3846

The CPT of SNACK (Table 5) revealed that the probability of purchasing SNACK increased from 0.3505 (neither parent present) to 0.5439 when MIE_INSTANT was present without ROKOK, but dropped to 0.3846 when both parents were present.

To provide a quantitative comparison with traditional association-rule approaches, two complementary evaluations were conducted. First, the predictive performance of the learned BN was benchmarked against an independence model.

Table 6: Predictive log-likelihood comparison between the independence model and the Bayesian Network across 10-fold cross-validation

Fold	Independence Model	Bayesian Network	Difference
1	−11.8881	−6.1447	+5.7434
2	−11.4336	−11.1555	+0.2781
3	−11.9491	−4.8670	+7.0822
4	−12.2302	−6.6126	+5.6175
5	−11.6702	−9.3735	+2.2967
6	−10.7398	−1.0787	+9.6611
7	−11.0386	−5.3153	+5.7233
8	−10.9313	−4.7812	+6.1501
9	−11.1650	−8.8445	+2.3206
10	−10.1245	−5.5279	+4.5966
Mean (±SD)	−11.3170 (±0.6085)	−6.3701 (±2.6974)	+4.9470

The BN achieved a higher predictive log-likelihood than the independence model across all ten folds, with an average improvement of +4.9470 (Table 6), confirming that the multivariate dependency structure provided a substantially better fit to out-of-sample data.

Second, to directly compare the BN with an association-rule method, the Apriori algorithm was applied to the same dataset with a minimum support threshold of 0.05 and a minimum confidence threshold of 0.50. The rule MIE_INSTANT → SNACK was identified with a confidence of 0.5143 and a lift of 1.5120, indicating that the probability of purchasing SNACK increases when MIE_INSTANT is present. The BN provided a more detailed decomposition of this relationship: $P(\text{SNACK} = 1 \mid \text{MIE\_INSTANT} = 1, \text{ROKOK} = 0) = 0.5439$ and $P(\text{SNACK} = 1 \mid \text{MIE\_INSTANT} = 1, \text{ROKOK} = 1) = 0.3846$.

Table 7: Quantitative comparison between Apriori confidence and Bayesian Network conditional probabilities for the relationship MIE_INSTANT → SNACK

Method	Condition	$P(\text{SNACK} = 1 \mid \text{MIE\_INSTANT} = 1)$
Apriori (confidence)	(unconditional)	0.5143
Bayesian Network (CPT)	ROKOK = 0	0.5439
Bayesian Network (CPT)	ROKOK = 1	0.3846

As shown in Table 7, the Apriori confidence of 0.5143 masked substantial heterogeneity that the BN revealed through conditioning on ROKOK. The probability of SNACK given MIE_INSTANT was 0.5439 without ROKOK but only 0.3846 with ROKOK. Together, the log-likelihood results and the conditional probability comparison provided both global and local quantitative evidence complementing the qualitative comparison in Table 1.

The LOIO accuracy for reconstructing SNACK, evaluated using the protocol described in the Methods section, was 0.6659. To contextualize this result, two baseline methods were evaluated on the same data. The first baseline was a marginal-frequency predictor that always predicts the majority class (SNACK = 0), which yields an accuracy of 0.6599, matching the proportion of transactions without SNACK in the dataset. The second baseline was an association-rule predictor based on the Apriori rule MIE_INSTANT → SNACK (confidence = 0.5143), which predicts SNACK = 1 whenever MIE_INSTANT = 1 and SNACK = 0 otherwise, yielding an accuracy of 0.6644.

Table 8: LOIO accuracy of the Bayesian Network compared with baseline methods

Method	LOIO Accuracy
Marginal-frequency baseline (always predict SNACK = 0)	0.6599
Apriori rule (MIE_INSTANT → SNACK)	0.6644
Bayesian Network (CPT-based prediction)	0.6659

Based on Table 8, the BN achieved a modest improvement over both baselines, with the Apriori-based predictor already performing well (0.6644) and the BN improving slightly (0.6659) by additionally conditioning on ROKOK. The limited margin reflected the imbalanced class distribution (34% SNACK presence) and the inherent noise in individual item prediction from basket data. The ROC-AUC was 0.5235, which was close to 0.5 and confirmed that the BN was not optimized for classification. This value served only as a secondary diagnostic check, reinforcing that the primary contribution lay in structural stability and conditional-dependence analysis rather than predictive discrimination.

The stability of the final consensus network was examined through three complementary sensitivity analyses, all conducted on the full reduced dataset with 441 transactions.

First, the prior sensitivity was assessed by varying the Equivalent Sample Size (ESS) of the BDeu score. The structure learning procedure was repeated with ESS values of 1, 5, 10, and 20, each using 10 independent random seeds (with 20 seeds for ESS=1). The edge ROKOK → SNACK remained present with a support of 1.00 across all ESS configurations (Table 9). The edge MIE_INSTANT → SNACK also persisted with a support of 1.00 for ESS = 5, 10, and 20, while its support decreased to 0.50 at ESS = 1. This slight decrease under a very weak prior (ESS=1) was expected, as minimal regularization allowed weaker dependencies to fluctuate; nevertheless, the edge still exceeded the adopted threshold of 0.50. These results confirmed that the principal dependencies were not artifacts of the chosen prior strength.

Second, the influence of the directional constraint on SNACK was evaluated by removing the restriction on outgoing edges from SNACK. In the unconstrained network, ROKOK → SNACK remained highly stable with a support of 1.00, confirming that this relationship was genuinely driven by the data. However, the edge MIE_INSTANT → SNACK was replaced by the reverse direction SNACK → MIE_INSTANT, with a support of 1.00. The latter direction was substantively implausible, as snack purchases were unlikely to drive instant noodle purchases in a convenience-store setting. Thus, the constraint only resolved directional ambiguity among Markov-equivalent structures and did not artificially induce the parent set of SNACK.

Third, the robustness of the parameter estimates to zero-frequency cells was verified using Laplace smoothing (pseudo-count = 1). The conditional probability estimates for SNACK given its parents (MIE_INSTANT and ROKOK) differed by less than 0.01 from the MLE estimates (Table 10). Consequently, the LOIO accuracy changed only marginally from 0.6659 (MLE) to 0.6553 (Laplace), and the ROC-AUC remained virtually unchanged from 0.5235 to 0.5219. These results confirmed that the parameter estimation method did not materially affect the predictive conclusions or the core conditional relationships. The zero-probability issue in the CPT of SABUN_PIRING, which lay outside SNACK's Markov blanket, did not influence any inference involving the focal category.

Table 9: Stability of principal edges under varying Equivalent Sample Size (ESS)

ESS	ROKOK → SNACK Support	MIE_INSTANT → SNACK Support
1	1.00	0.50
5	1.00	1.00
10	1.00	1.00
20	1.00	1.00

Note: Support was computed as the proportion of independent runs (10 for ESS=5,10,20; 20 for ESS=1).

Table 10: Comparison of MLE and Laplace-smoothed conditional probabilities for SNACK

MIE_INSTANT	ROKOK	MLE $P(\text{SNACK} = 1)$	Laplace $P(\text{SNACK} = 1)$
0	0	0.3505	0.3496
0	1	0.1500	0.1552
1	0	0.5439	0.5417
1	1	0.3846	0.3889

Together, these sensitivity analyses confirmed that both the network structure and the estimated parameters were robust to reasonable variations in prior specification, directional constraints, and parameter estimation methods.

Having established the predictive capability of the BN through quantitative comparisons with both the independence model and the Apriori baseline, the discussion now turns to the substantive interpretation of the learned dependency structure. Beyond its predictive performance, the value of the BN lies in its ability to reveal the conditional dependence patterns among product categories, offering insights into consumer purchasing behaviour that are not accessible through traditional association-rule methods. The following subsections examine these patterns, their methodological implications, and their practical relevance for retail decision-making.

The final network structure suggests that SNACK purchases are embedded in at least two overlapping retail consumption patterns. The first pattern corresponds to a convenience-food dependency structure, in which BUMBU, MIE_INSTANT, and SNACK form a chain of conditional associations. This structure is substantively plausible, as instant noodles and seasonings are commonly purchased together for quick-meal preparation, and snack items tend to co-occur with these categories in the same transactions. The relatively high conditional probability of SNACK given the presence of MIE_INSTANT reinforces the interpretation, indicating that snack purchases are more likely to occur within the specific consumption context.

In contrast, the second pattern is centred around ROKOK, which occupies a structurally central position in the network, being conditionally associated with multiple categories, including KOPI, DAGING_OLAHAN, SABUN_PIRING, and SNACK. Unlike the food-based pathway, the configuration is more heterogeneous and does not correspond to a single, clearly defined consumption routine. One plausible interpretation is that ROKOK represents a broader convenience-store shopping mission, where cigarettes are purchased alongside various discretionary or replenishment items. However, the conditional probabilities reveal a more nuanced relationship: the presence of ROKOK is associated with a lower probability of SNACK when considered independently of MIE_INSTANT. This suggests that structural centrality does not necessarily imply a positive conditional influence on the focal outcome. The inferred relationships should be interpreted as probabilistic dependencies rather than causal effects, as the model is learned from observational transaction data.

The distinction is analytically important because it illustrates the advantage of probabilistic graphical modelling over traditional association-rule approaches [3, 4]. While rule-based methods may identify frequent co-occurrence between categories, they do not capture how such relationships change under different contextual conditions. The BN framework, in contrast, reveals an asymmetric pattern of conditional dependence: the probability of SNACK is substantially higher when MIE_INSTANT is present, but is lower when ROKOK is present, relative to the baseline probability. This demonstrates that a category could be structurally important without consistently corresponding to an elevated probability of another—an insight inaccessible through traditional association-rule methods [4, 6]. Moreover, it reveals a key limitation of association-rule methods: they cannot condition on additional variables to reveal how relationships change across different contexts. This demonstrates that co-occurrence alone is insufficient to characterize purchasing behaviour without accounting for conditional dependencies.

From a methodological perspective, the findings also emphasize the importance of stability analysis in structure learning [4, 10]. The use of cross-validation and multi-seed aggregation reveals that not all edges are equally reliable, and that some dependencies are sensitive to sample variation. By incorporating edge support as part of the evaluation, the research avoids over-interpreting relationships that may arise from a single optimization run. The approach is particularly relevant for retail transaction data, which are often sparse and context-dependent.

These findings also suggest a broader modelling principle in retail analytics. The most useful BN is not necessarily the one with the highest score or the largest number of nodes, but rather the one that balances structural stability, interpretability, and substantive meaning. In the research, the reduced network achieves the balance by preserving stable local dependencies while eliminating peripheral complexity, making it suitable for both analysis and decision support.

From a managerial perspective, the results provide insights into potential merchandising and promotion strategies [13]. The strong positive conditional association between MIE_INSTANT and SNACK suggests a natural opportunity for cross-selling, such as product bundling or shelf co-location. In contrast, the position of ROKOK in the dependency structure requires more careful interpretation. Although it is structurally central, its conditional association with SNACK does not consistently correspond to an increased probability of SNACK purchases. This highlights an important practical implication: categories that appear influential in raw co-occurrence patterns may not necessarily serve as effective anchors for targeted promotions when conditional relationships are taken into account.

More broadly, the reduced network provides a structured view of product relationships that can inform store layout and category management [13]. By identifying clusters of interacting categories, such as those involving coffee, cigarettes, and convenience foods, the model offers a more systematic basis for organizing product placement compared to traditional association-rule outputs. While the network does not prescribe specific actions, it supports more informed decision-making by clarifying how different categories interact within the same transaction context.

The modelling decision to restrict outgoing edges from SNACK raises the question of how sensitive the findings are to this assumption. To assess this empirically, we compared the constrained network to an unconstrained version where outgoing edges from SNACK were permitted. In the unconstrained network, the edge ROKOK → SNACK remained highly stable with a support of 1.00, confirming that this relationship is genuinely driven by the data. However, the edge MIE_INSTANT → SNACK was replaced by the reverse direction SNACK → MIE_INSTANT, also with a support of 1.00. This reversal is substantively implausible: snack purchases are unlikely to drive instant noodle purchases in a convenience-store setting. Therefore, the constraint does not artificially induce the parent set of SNACK; it merely resolves directional ambiguity among Markov-equivalent structures, selecting the direction that aligns with domain knowledge.

The robustness of the network to the choice of prior was similarly confirmed. When the Equivalent Sample Size (ESS) was varied from 1 to 20, the edge ROKOK → SNACK maintained a support of 1.00 across all values, while MIE_INSTANT → SNACK remained present with support ≥ 0.50 (Table 9). The slight decrease at ESS=1 is expected, as a very weak prior provides minimal regularization, allowing weaker dependencies to fluctuate.

Importantly, the key empirical finding of this study—that MIE_INSTANT and ROKOK are the primary predictors of SNACK—is not an artefact of the modelling decisions. If the constraint were forcing these relationships, both MIE_INSTANT and ROKOK would appear as parents of SNACK in all ten cross-validation folds. Instead, ROKOK appeared in eight folds and MIE_INSTANT in four (Table 3), indicating that their selection is driven by genuine dependence structures in the data rather than by the modelling restrictions. The constraint only guarantees that no edges leave SNACK; it does not dictate which edges enter it, and the prior sensitivity analysis confirms that the key edges persist across reasonable variations in ESS.

To assess whether the zero-cell issue in the CPT of SABUN_PIRING affects the main conclusions of this study, we performed a sensitivity analysis using Laplace smoothing (adding a pseudo-count of 1 to each cell). Under Laplace smoothing, the previously zero probability would become a small non-zero value. The re-estimated conditional probabilities for SNACK given its parents (MIE_INSTANT and ROKOK) differed by less than 0.01 from the MLE estimates (Table 10). Furthermore, the predictive performance remained highly stable: the LOIO accuracy changed only marginally from 0.6659 (MLE) to 0.6553 (Laplace), and the ROC-AUC remained virtually unchanged from 0.5235 to 0.5219. Importantly, SABUN_PIRING is not part of the Markov blanket of SNACK and does not appear as a parent, child, or spouse of SNACK in the final reduced network. Consequently, the zero-cell issue in its CPT does not affect any of the conditional probability estimates involving SNACK, MIE_INSTANT, or ROKOK, which constitute the core empirical findings of this study.

The conclusions regarding the context-dependent relationship between instant noodles, cigarettes, and snack purchases are therefore robust to the choice between MLE and smoothed estimates, as well as to the choice of prior and directional constraint. Future work with larger datasets or with prediction-oriented applications may benefit from full Bayesian estimation of CPTs using Dirichlet priors to provide posterior distributions over the conditional probabilities rather than point estimates.

These limitations suggest several directions for future research. Expanding the dataset to include multiple outlets would allow for the identification of more generalizable patterns. Incorporating temporal information could enable the analysis of dynamic purchasing behaviour, while the inclusion of additional variables such as price and customer segmentation could enrich interpretation. The current research provides a foundational probabilistic framework that can be extended in these directions to support more comprehensive retail analytics.

4. Conclusion

The study presents a stability-aware Bayesian Network framework for modelling dependency relationships among product categories in retail transaction data, with SNACK as the focal category. The results demonstrate that a consensus-based structure learning approach—incorporating repeated stochastic search, edge-support filtering, and Markov blanket reduction—is effective in identifying a stable and interpretable dependency structure from sparse basket data. Empirically, SNACK is conditionally dependent on MIE_INSTANT and ROKOK, but with distinct patterns of association: the probability of SNACK is substantially higher when instant noodles are present, while cigarettes, despite their structural centrality in the network, do not consistently correspond to an elevated probability of SNACK.

While the individual techniques employed in this study are well established in the Bayesian Network literature, the primary contribution lies in their integration into a coherent, stability-aware Bayesian Network framework specifically tailored for sparse retail transaction data. This framework addresses a practical challenge frequently encountered in market basket analysis: how to obtain a stable and interpretable dependency structure from transaction data characterized by high dimensionality relative to sample size. By systematically combining repeated stochastic searches with consensus-based edge selection and structure simplification, the framework ensures that the resulting network captures only those dependencies that are consistently supported by the data. The robustness of these findings is further supported by sensitivity analyses across prior specifications (ESS 1–20), directional constraints (constrained vs. unconstrained), and parameter estimation methods (MLE vs. Laplace smoothing), confirming that the core dependencies are stable and not artifacts of model choices. Furthermore, this study demonstrates that such a stability-aware approach yields insights—such as the context-dependent relationship between instant noodles, cigarettes, and snacks—that are not accessible through traditional association-rule methods. This highlights the value of robust probabilistic modelling in retail analytics, where the goal is not merely to identify co-occurring products but to understand the conditional structure of consumer purchasing behaviour.

The findings highlight that co-occurrence, structural importance, and conditional dependence represent different aspects of product relationships that are better captured through probabilistic modelling, thereby addressing the limitations of traditional association-rule approaches. From a practical perspective, the model provides actionable insights for retail decision-making, particularly in product placement and cross-selling strategies. From a methodological perspective, the study demonstrates the importance of stability-aware structure learning in producing reliable and interpretable Bayesian Network models for sparse transaction data.

Future extensions of this work could incorporate promotional campaign data and customer demographic profiles as additional conditioning variables in the Bayesian Network. For instance, the conditional probability of SNACK given MIE_INSTANT and ROKOK could be further stratified by whether a promotion was active for any of these categories at the time of purchase, revealing whether the observed dependency patterns are organic or promotion-driven. Similarly, incorporating demographic attributes such as age group or shopping time could uncover whether the context-dependent relationship between instant noodles, cigarettes, and snacks varies across consumer segments, thereby enabling more targeted cross-selling and promotional strategies tailored to specific customer profiles.

CRediT Authorship Contribution Statement

Muhammad Sahid: Data Curation, Software, Conceptualization, Methodology, Supervision, Project Administration, Funding Acquisition, Writing–Original Draft & Visualization.
I Gede Nyoman Mindra Jaya: Supervision, Methodology, Project Administration, Editing.
Restu Arisanti: Supervision, Methodology, Project Administration, Editing.

Declaration of Generative AI and AI-assisted Technologies

No generative AI or AI-assisted technologies were used during the preparation of this manuscript.

Declaration of Competing Interest

The authors declare no competing interest.

Funding and Acknowledgments

This research received no external funding.

Data Availability

The data supporting this article are derived from a master's thesis completed at Universitas Padjadjaran, Indonesia. The data were collected through manual recording of customer transaction receipts and contain potentially sensitive commercial information. Therefore, the data are not publicly available. Access requests may be directed to the corresponding author upon reasonable request and with permission from the relevant retail partners.

References

[1] A. Holzapfel, H. Kuhn, and M. G. Sternbeck. "Product allocation to different types of distribution center in retail logistics networks". In: *Eur. J. Oper. Res.* 264.3 (2018), pp. 948–966. DOI: 10.1016/j.ejor.2016.09.013.

[2] A. U. Khasanah and M. R. Q. Baihaqie. "Analysis of consumer characteristics on retail business with clustering analysis method and association rule for selling improvement strategy recommendations". In: *OPSI* 17.1 (2024), p. 249. DOI: 10.31315/opsi.v17i1.11411.

[3] G. F. Cooper and E. Herskovits. "A Bayesian method for the induction of probabilistic networks from data". In: *Mach. Learn.* 9.4 (1992), pp. 309–347. DOI: 10.1007/BF00994110.

[4] J. Zeng and Z.-Q. Liu. "Probabilistic Graphical Models". In: *Advances in Intelligent Systems and Computing.* 2015, pp. 9–16. DOI: 10.1007/978-3-662-44690-4_2.

[5] I. G. N. M. Jaya, B. Handoko, Y. Andriyana, A. Chadidjah, F. Kristiani, and M. Antikasari. "Multivariate Bayesian Semiparametric Regression Model for Forecasting and Mapping HIV and TB Risks in West Java, Indonesia". In: *Mathematics* 11.17 (2023), p. 3641. DOI: 10.3390/math11173641.

[6] K. Gupta, K. Shah, and A. A. Kadam. "E-Commerce Market Basket Analysis using Apriori Algorithm". In: *INTERANTIONAL J. Sci. Res. Eng. Manag.* 7.10 (2023), pp. 1–11. DOI: 10.55041/IJSREM25922.

[7] H. Hruschka. "Multi-category purchase incidences with marketing cross effects". In: *Rev. Manag. Sci.* 11.2 (2017), pp. 443–469. DOI: 10.1007/s11846-016-0193-0.

[8] M. Scutari. "Learning Bayesian Networks with the bnlearn R Package". In: *J. Stat. Softw.* 35.3 (2010). DOI: 10.18637/jss.v035.i03.

[9] N. K. Kitson, A. C. Constantinou, Z. Guo, Y. Liu, and K. Chobtham. "A survey of Bayesian Network structure learning". In: *Artif. Intell. Rev.* 56.8 (2023), pp. 8721–8814. DOI: 10.1007/s10462-022-10351-w.

[10] R. J. Lempert, D. G. Groves, S. W. Popper, and S. C. Bankes. "A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios". In: *Manage. Sci.* 52.4 (2006), pp. 514–528. DOI: 10.1287/mnsc.1050.0472.

[11] H. Mamiya, K. Crowell, C. L. Mah, A. Quesnel-Vallée, A. Verma, and D. L. Buckeridge. "Characterizing co-purchased food products with soda, fresh fruits, and fresh vegetables using loyalty card purchasing data in Montréal, Canada, 2015–2017". In: *Int. J. Behav. Nutr. Phys. Act.* 22.1 (2025), p. 19. DOI: 10.1186/s12966-024-01701-8.

[12] M. N. A. Nafi, A. S. Zakaria, N. I. Mohamad Arif, S. N. Mohd Shafie, N. Ab. Aziz, and O. Kairan. "Market Basket Analysis for Sales Transaction in Shopping Stores". In: *Int. J. Acad. Res. Bus. Soc. Sci.* 13.2 (2023). DOI: 10.6007/IJARBSS/v13-i2/15439.

[13] N. H. F. Hartanto and B. Aribowo. "Perancangan Tata Letak Toko Ritel Berdasarkan Pola Belanja Konsumen Dengan Market Basket Analysis (Studi Kasus: Indomaret Sukatani)". In: *J. Al-AZHAR Indones. SERI SAINS DAN Teknol.* 8.2 (2023), p. 119. DOI: 10.36722/sst.v8i2.1375.

[14] V. R. Kalamkar and K. Monkova, eds. *Advances in Mechanical Engineering*. Lecture Notes in Mechanical Engineering. Springer Singapore, 2021. DOI: 10.1007/978-981-15-3639-7.

[15] S. Rana and M. N. I. Mondal. "Seasonal and Multilevel Association Based Approach for Market Basket Analysis in Retail Supermarket". In: *Eur. J. Inf. Technol. Comput. Sci.* 1.4 (2021), pp. 9–15. DOI: 10.24018/compute.2021.1.4.31.

[16] S. A. Rahayu. "Market Basket Analysis Using FP-Growth Algorithm to Design Marketing Strategy by Determining Consumer Purchasing Patterns". In: *J. Appl. Data Sci.* 4.1 (2023), pp. 38–49. DOI: 10.47738/jads.v4i1.83.

[17] N. Nadif and I. Vanany. "Analysis of Demand Disruption on Product Availability in the Retail Industry". In: *PROZIMA (Productivity, Optim. Manuf. Syst. Eng.* 5.1 (2021), pp. 12–20. DOI: 10.21070/prozima.v5i1.1504.

[18] J. Hu and C. Chen. "Exploration of Normal Stochastic Multicriteria Decision-Making and Legal Dilemma in Project Investment". In: *Math. Probl. Eng.* 2021 (2021), pp. 1–10. DOI: 10.1155/2021/1082186.

[19] K. Yu, L. Liu, and J. Li. "A Unified View of Causal and Non-causal Feature Selection". In: *ACM Trans. Knowl. Discov. Data* 15.4 (2021), pp. 1–46. DOI: 10.1145/3436891.

[20] X. Wu, B. Jiang, K. Yu, H. Chen, and C. Miao. "Multi-Label Causal Feature Selection". In: *Proc. AAAI Conf. Artif. Intell.* 34.4 (2020), pp. 6430–6437. DOI: 10.1609/aaai.v34i04.6114.

[21] F. J. Díez, M. Arias, J. Pérez-Martín, and M. Luque. "Teaching Probabilistic Graphical Models with OpenMarkov". In: *Mathematics* 10.19 (2022), p. 3577. DOI: 10.3390/math10193577.

[22] M. R. Islam et al. "Leveraging Machine Learning for Insights and Predictions in Synthetic E-commerce Data in the USA: A Comprehensive Analysis". In: *J. Ecohumanism* 4.2 (2025). DOI: 10.62754/JOE.V4I2.6635.