International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 102-118
https://doi.org/10.70917/ijcisim-2026-3425
Article

Multi-Strategy Pruning System for High-Value Retail Transaction Analysis in Grocery Stores

Roselin Annam R[1], B L Shivakumar[2]

[1]Sri Ramakrishna College of Arts & Science, Coimbatore, Email: roselin@srcas.ac.in
[2]Sri Ramakrishna College of Arts & Science, Coimbatore

Abstract: This research proposed multi-strategy algorithms for high-value retail transaction analysis using pruning systems, such as Minimum Significant Utility (MSU), Optimal Support Reduction (OSR), Optimal Weight Learning (OWL), and Cost Complexity Post-Pruning (CCP). The proposed framework was designed to remove negligible patterns and enhance the classification performance. The Gini Index was then used as the splitting criterion to build a Classification and Regression Tree (CART) model. The algorithm was tested on a Multi-Strategy Pruning System for retail transaction analysis in a Grocery Store using the Happy Grocery Dataset, which contains 10,000 transactions and unique items. Traditional mining techniques often yield complex decision trees that contain redundant patterns, irrelevant attributes, and noisy branches, thereby increasing their complexity and incomprehensiveness. Before post-pruning optimization, the suggested pruning procedures are used to eliminate low-utility transactions, weak support patterns, and less significant qualities. Accuracy, precision, recall, F1-score as a metrics to find pruning ratio, and tree complexity used for the experimental evaluation. Intelligent retail analytics applications include recommendation systems, transaction latency, personalized marketing, consumer retention, and strategic business planning.
Keywords: Retail Transaction Analysis, Pruning System, Classification and Regression Tree, Post-Pruning Optimisation

1. INTRODUCTION

Data Mining has occurred as a powerful analytical tool for extracting useful methods and orders from massive databases and raw transactional records for actionable business intelligence. These models and patterns focus on identifying meaningful patterns in consumer purchase histories. The extracted knowledge can support various processes, such as searching, extracting, and analysing methods in massive datasets, at the intersection of consumer segmentation, product recommendation, inventory optimization, consumer retention analysis, and sales forecasting. Among the various data mining techniques, classification algorithms are widely employed to predict consumer behaviour and identify future purchasing trends [1].

Retail Transaction Analysis prefers information technology and digital commerce, which has led to unprecedented growth in brands and goods among consumers [2]. Recommendation methods focus on supermarkets, grocery stores, e-commerce platforms, and retail chains that generate enormous volumes of transactional data containing valuable information about consumer purchasing behaviour. One method that can be used to discover hidden patterns, consumer preferences, and purchasing trends that can help retailers make informed business decisions. Therefore, mining such datasets requires intelligent mechanisms that can eliminate irrelevant information without losing valuable knowledge [3]-[5].

Pruning is a critical technique for overcoming the limitations of overgrown decision trees. The primary objective of pruning is to remove redundant branches, irrelevant attributes, and insignificant decision nodes while preserving the predictive performance [16]. Pruning not only reduces theoretical measure moreover the interpretability and scalability of the model. Traditional pruning approaches mainly focus on pre-pruning and post-pruning techniques. Pre-pruning restricts tree growth during construction by imposing constraints, such as the maximum tree depth and minimum sample thresholds. Post-pruning simplifies a fully grown tree by eliminating branches that contribute little to prediction accuracy [17].

Although existing pruning methods effectively reduce tree complexity, many approaches fail to consider the significance of transaction utility, feature importance, and support-based pattern relevance simultaneously. Consequently, important retail knowledge may be lost during aggressive pruning or retained unnecessarily, resulting in suboptimal models [14], [19], [21]. Therefore, an intelligent pruning framework that preserves highly informative patterns while removing redundant information as required. This research makes several significant contributions to the fields of retail transaction mining and predictive analytics.

- A novel Multi-Strategy Pruning System was developed.
- Integration of Minimum Significant Utility (MSU), Optimal Support Reduction (OSR), Optimal Weight Learning (OWL), and Cost Complexity Pruning techniques.
- An optimized CART model using customer behavioural attributes.
- A comprehensive was performed using the Happy Grocery Dataset, which contains approximately 10,000 transactions.
- Generation of simplified and interpretable models for predicting customer purchases.
- Reduction of model complexity while maintaining predictive performance.
- Actionable business intelligence for customer retention and retail strategy planning.

Classification and Regression Tree (CART) is one of the most effective decision tree algorithms used in predictive analytics [16], [17]. CART constructs binary decision trees using recursive partitioning and employs the Gini Index as a splitting criterion to identify optimal decision boundaries. Owing to its simplicity, interpretability, and effectiveness, CART has been widely applied in customer relationship management, healthcare analytics, financial forecasting, fraud detection, and retail transaction mining. Consequently, the proposed MPS framework has the potential to support intelligent retail analytics systems and enhance business competitiveness.

The remainder of this manuscript is laid out as follows: Section 2 presents a literature review in existing pruning techniques. Section 3 describes the Happy Grocery Dataset and Data Preprocessing Architecture. Section 4 introduces the proposed MPS framework. Section 5 presents results and discussion of this research. Section 6 concludes the research and outlines future exploration. Section 7 provides a comparative analysis of the reference.

2. LITERATURE REVIEW

Retail transaction mining has become a mandatory research area in data mining because of the rapid growth of consumer transaction databases. Organisations continuously seek methods to discover meaningful patterns, consumer purchasing behaviours, and hidden relationships among products. Various techniques, such as Association Rule Mining, Machine Learning algorithms, Utility Mining, Pruning Strategies and Decision Trees, have been employed to improve retail analytics and customer behaviour. Among these approaches, decision tree-based models have gained significant attention because of their interpretability and ease of pruning technique implementation [2]-[5].

Sabila and Candra (2023) [8] proposed the implementation of the Apriori algorithm to analyzing sales transaction data to discover frequent itemsets and generate meaningful association rules. This study aimed to identify products that are commonly purchased together, thereby assisting retailers in designing effective marketing strategies and improving sales performance. By applying the Apriori algorithm to transactional data, the authors calculated the support, confidence, and lift values to evaluate the strength of the relationships between products. The results demonstrate that association rule mining can uncover hidden purchasing patterns, enabling businesses to create product bundles, optimize product placement, and support inventory management. This research highlights the effectiveness of Apriori in extracting valuable insights from retail transaction databases, although it primarily focuses on association analysis rather than predictive modelling or pruning techniques.

X. Zhao, X. Zhong, and B. Han (2024) [12] proposed an algorithm that integrates Leiden Community Detection with a Compact Genetic Algorithm (LcGA) for transaction pattern mining based on tree and linked list structures. To improve mining efficiency and reduce computational complexity, it partitions the transaction database into highly related communities using the search space. It then applies to extract frequent itemsets, performs closure detection to obtain frequent closed itemsets, and finally employs a compact genetic algorithm to identify high-utility patterns. On the retail process, Chainstore, Chess, and Accidents datasets demonstrated that the proposed algorithm achieved an average runtime improvement of 40.5% compared to existing closed high-utility mining methods while successfully discovering approximately 96% of frequent closed high-utility itemsets. It concludes that combining community detection with evolutionary optimization significantly enhances mining efficiency and dense transactional datasets. However, the approach focuses on frequent closed high-utility itemset mining and does not incorporate predictive classification models or multiple pruning strategies for customer repurchase prediction, leaving opportunities for further research.

Hikmawati, Maulidevi, and Surendro (2022) [16] proposed a pruning strategy for an adaptive rule method by sorting utility items to improve the efficiency of high-utility itemset mining. This study introduced a utility-based pruning mechanism that orders items according to their utility values and eliminates unpromising candidates during the mining process. By reducing the number of unnecessary computations and search paths, the proposed strategy significantly decreases the execution time while maintaining the quality and accuracy of the generated association rules. The experimental results demonstrated that sorting utility items before rule generation effectively reduces memory usage and computational complexity, making the process suitable for large transactional databases. This study highlights the importance of pruning techniques in optimizing data mining algorithms and improving scalability.

Le, Nguyen, Nguyen, Bui, Yun, and Vo (2025) [17] proposed an efficient approach for Frequent Weighted Utility Closed Patterns (FWUCPs) from measurable numerical values for databases using advanced pruning strategies. This study addresses the challenge of discovering high-utility patterns in databases where transactions are continuously updated and where both item weights and quantities influence pattern significance. The authors introduced pruning mechanisms to eliminate unpromising candidates early in the mining process, thereby reducing the search space and enhance the computational efficiency. Their method preserves only weighted utility closed patterns that provide meaningful business insights while minimizing redundant computations. Experimental evaluations demonstrated that the proposed pruning strategies significantly reduced execution time and memory consumption compared to conventional utility mining techniques, making the approach suitable for large-scale and dynamic transactional datasets. This study highlights the effectiveness of pruning in utility-based data mining and supports the development of efficient retail analytics systems. Similar to the study emphasizes reducing model complexity through intelligent pruning; however, further integrates pruning with CART-based classification to enhance customer purchase prediction and decision-making.

Sri Darshan M., Nithinraj N., and Jaisachin B. (2024) [18] presented an integrated retail decision-making optimization framework that combines association rule mining, sequential pattern mining, temporal pattern analysis, and predictive modeling. The study utilizes the Apriori algorithm to identify frequent itemsets and generate association rules that reveal customer purchasing relationships, supporting effective cross-selling and product placement strategies. Additionally, the PrefixSpan algorithm is employed to discover sequential purchasing patterns, enabling retailers to understand customer buying sequences and enhance recommendation systems. Model performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), demonstrating reliable forecasting accuracy. The integrated framework enhances inventory management, marketing effectiveness, and customer satisfaction by combining pattern mining with predictive analytics. However, on synthetic retail data and does not incorporate adaptive pruning mechanisms or weighted multiple-strategy pruning techniques for large-scale market basket optimization, leaving scope for further improvements in real-world retail applications.

Suresh Gangula (2026) [20] presented a systematic review of retail application performance optimization by analyzing 45 peer-reviewed studies published between 2015 and 2025. The review examines modern monitoring platforms, including Splunk, Datadog, and New Relic, along with data collection methods, key performance indicators (KPIs), real-time monitoring, dashboards, and optimization strategies for cloud-based retail systems. The study highlights that application performance directly influences customer satisfaction, operational efficiency, and business continuity, particularly in environments characterized by seasonal demand fluctuations and high transaction volumes. It further emphasizes the importance of integrating predictive analytics, artificial intelligence, and cloud-native monitoring to improve scalability, resilience, and anomaly detection. Although the review provides a comprehensive comparison of monitoring tools and best practices for retail application performance, it primarily focuses on system monitoring and infrastructure optimization rather than customer purchase prediction, market basket analysis, or adaptive pruning strategies. Therefore, additional research is needed to integrate data mining techniques and multiple pruning strategies for improving customer behavior prediction and retail decision support.

Therefore, there exists a significant research opportunity to develop a Multi-Objective Pruning Strategy (MPS) capable of simultaneously improving computational efficiency and retail business performance. Based on the reviewed literature, the following gaps are identified:

1. Existing pruning strategies focus primarily on computational measures.
2. Inventory turnover is not considered during pruning decisions.
3. Customer retention metrics remain absent from utility mining frameworks.
4. Shelf-space utilization is rarely incorporated into pattern evaluation.
5. Transaction latency is treated independently of business objectives.
6. No unified framework integrates multiple retail performance indicators into pruning mechanisms.

TABLE 1: COMPARATIVE ANALYSIS OF EXISTING TECHNIQUES

Algorithm	Support-Based Pruning	Utility-Based Pruning	Inventory Turnover	Customer Retention	Transaction Latency	Multi-Objective Capability
Apriori	Yes	No	No	No	No	No
FP-Growth	Yes	No	No	No	No	No
UP-Growth	No	Yes	No	No	Yes	No
HUI-Miner	No	Yes	No	No	Yes	No
FHM	No	Yes	No	No	Yes	No
Proposed MPS	Yes	Yes	Yes	Yes	Yes	Yes

The findings clearly indicate the need for a business-oriented, multi-objective pruning framework capable of supporting grocery store owners through intelligent transaction mining and operational optimization.

The proposed framework contributes to both academic research and practical retail applications. From a research perspective, this study introduces a comprehensive pruning methodology that integrates multiple pruning strategies into a unified decision tree mining framework. From an industrial standpoint, the framework aids retailers in pinpointing valuable customers, comprehending buying patterns, enhancing marketing tactics, and refining operational choice. When traditional CART models are applied to such datasets, they often generate large decision trees with many branches and nodes. We developed a multi-strategy pruning framework that integrates utility-based pruning, support-based pruning, feature-weight pruning, and decision tree pruning. By combining these complementary pruning strategies, the proposed framework aims to improve prediction accuracy, reduce model complexity, and enhance the interpretability of retail analytics models.

3. DATASET DESCRIPTION AND DATA PREPROCESSING ARCHITECTURE

The Happy Grocery Dataset is a retail transaction dataset used to design simulated, realistic customer purchasing behaviour in grocery stores. Each transaction contains customer information and purchased products that can be utilized for classification, pattern mining and utility analysis. The quality of a model largely depends on the quality of the dataset during the training and evaluation process. Real-world retail datasets often contain inconsistencies, redundant information, and missing values. Preprocessing data into critical step that transforms raw transactional records into a structured format for pattern extraction, classification and pruning operations.

In this research, the Happy Grocery Dataset 10K was used to analyse the effectiveness of the proposed Multi-Strategy Pruning System (MPS). The dataset contained approximately 10,000 retail transaction records representing customer purchasing activities across multiple grocery categories. Each transaction (Table 2) contains customer information and purchased products that can be used for classification, pattern mining, and utility analysis.

The dataset consists of the following:

- Customer purchase records
- Product information
- Basket transactions
- Customer behavioural attributes
- Spending characteristics

TABLE 2: DATASET FEATURES

Attribute	Description
Customer_ID	Unique customer identifier
Product_Name	Purchased product
Category	Product category

Quantity	Number of units purchased
Price	Unit price of product
Basket_ID	Shopping basket identifier
Purchase_Frequency	Customer purchase frequency
Last_Purchase_Days	Days since last purchase
Avg_Basket_Size	Average basket size
Total_Spent	Total expenditure
Transaction_Date	Date of purchase
Payment_Method	Payment mode
Customer_Type	Customer category

The system begins with the **10K dataset**, which contains customer transaction records, such as purchase frequency, basket size, total spending, and last purchase date. The raw data underwent preprocessing to improve their quality. Retail transaction utilities are computed by forming customer baskets and calculating metrics such as quantity × price and the total spending. These values help to identify the importance of transactions.

The proposed framework applies three pruning strategies.

- **MSU (Minimum Significant Utility):** Removes low-utility transactions or items.
- **OSR (Optimal Support Reduction):** Eliminates patterns with insufficient support.
- **OWL (Optimal Weight Learning):** Discards features with low importance or weight.

These stages reduce the noise and search space before model training. The pruned dataset was used to build a Classification and Regression Tree (CART). The tree uses the Gini Index to select optimal split points and classify customer purchasing behaviour. After training, Cost Complexity Pruning (CCP) removes redundant branches from the decision tree. This produces a smaller, more interpretable model and reduces overfitting while preserving the predictive performance.

The optimized model generates:

- Customer repurchase predictions
- Identification of high-value customers
- Interpretable decision rules
- Accuracy, Precision, Recall, and F1-score are the performance metrics

FIG 1: ARCHITECTURE DIAGRAM

Proposed architecture (Fig 1) of the Multi-Strategy Pruning System (MPS) for customer purchase prediction using the Happy Grocery 10K dataset. The framework combines data preprocessing, utility computation, multi-strategy pruning (MSU, OSR, and OWL), CART classification, and cost-complexity post-pruning to generate an optimized and interpretable prediction model for retail analytics.

4. PROPOSED MPS FRAMEWORK

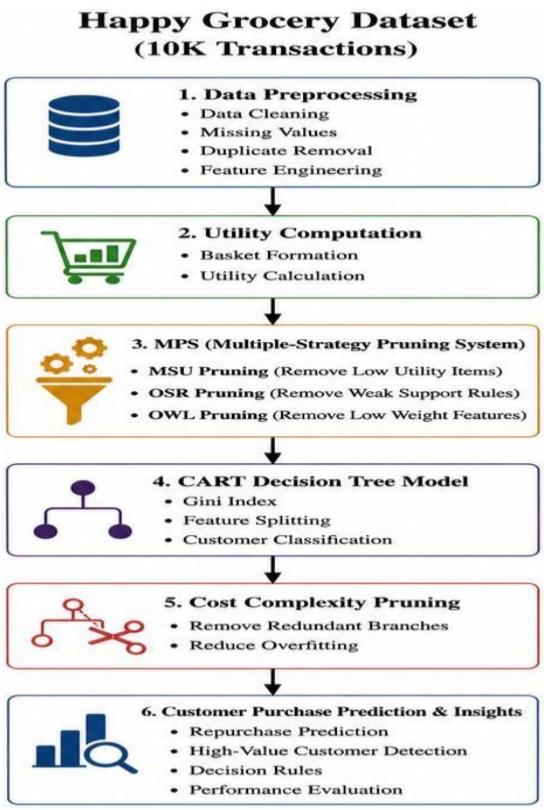
The primary aim of the proposed framework is to improve retail transaction mining performance by eliminating redundant, insignificant, and noisy patterns before and after the decision tree construction. Traditional decision tree models frequently generate complex structures that increase computational costs and reduce interpretability. To overcome these restrictions, this research proposes a Multi-Strategy Pruning System (MPS) that integrates multiple pruning strategies into a unified data mining framework.

The proposed MPS framework combines the following:

- Minimum Significant Utility (MSU) Pruning
- Optimal Support Reduction (OSR) Pruning
- Optimal Weight Learning (OWL) Pruning
- CART Construction
- Cost Complexity Post-Pruning (CCP)

The integration of these pruning mechanisms enables the framework to reduce the model complexity while maintaining the predictive performance.

FIG 2: FRAMEWORK



The overall process of the proposed framework (Fig. 2) comprises multiple sequential phases. The Happy Grocery 10K dataset was first subjected to data preprocessing, and the processed data were then used for utility computation to determine transaction significance. Subsequently, the proposed Multi-Strategy Pruning System (MPS) applies three pruning techniques, and the refined dataset is used to construct a CART decision tree, which is further optimized using Cost Complexity Pruning (CCP) to reduce redundant branches and mitigate overfitting. Finally, the optimized model generates customer purchase predictions, identifies high-value customers, extracts decision rules, and provides performance evaluation metrics for retail decision support.

The effectiveness of the proposed Multi-Strategy Pruning System (MPS) using the Happy Grocery Dataset investigates whether the integration of MSU, OSR, OWL, and Cost Complexity Pruning can reduce model complexity while maintaining high predictive performance. The CART classifier parameters that prevent excessive tree growth during training were configured using the following parameters.

TABLE 3: PRE-PRUNING PARAMETERS

Parameter	**Value**
Criterion	Gini
Max Depth	5
Min Samples Split	10
Min Samples Leaf	5
Random State	42

TABLE 4: POST-PRUNING PARAMETERS

Cost Complexity Pruning	**CCP Alpha (α)**	**Train Accuracy**	**Test Accuracy**
Generate full tree	0.000	1.000	0.989
Compute pruning path	0.002	0.944	0.940
Generate candidate trees	0.004	0.936	0.934
Evaluate testing accuracy	0.052	0.936	0.934
Select best-performing tree	0.055	0.852	0.852

Cost Complexity Pruning (CCP) was performed using the complexity parameter (ccp_alpha). The optimal pruning parameter was selected as $\alpha = 0.004$, as it achieved the best trade-off between model simplicity and predictive performance. This process ensures optimal tree complexity.

5. RESULTS AND DISCUSSION

This portion presents the research results obtained from the proposed Multi-Strategy Pruning System (MPS). The execution of the proposed framework (Fig. 2) is estimated using:

- Accuracy
- Precision
- Recall
- F1-Score
- Tree Complexity

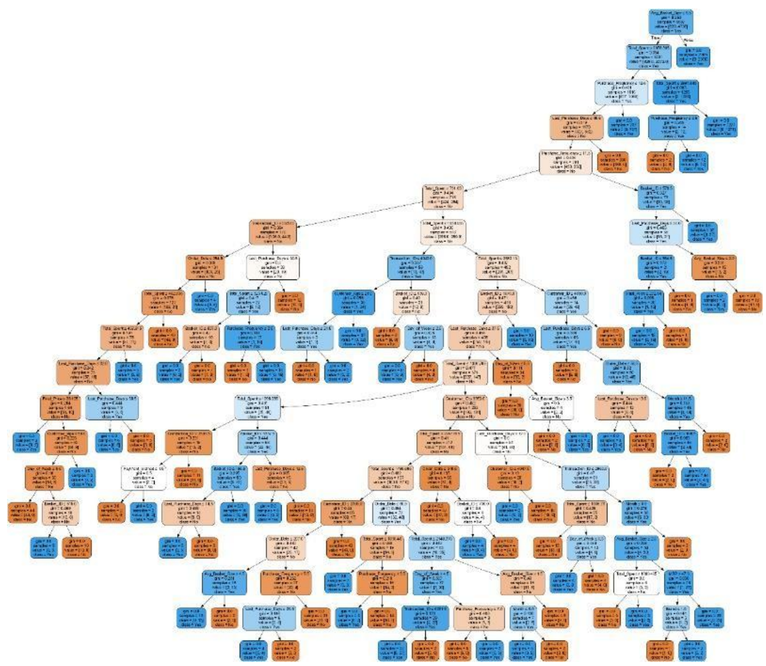
- Pruning Ratio
- Rule Quality

The results demonstrate that the proposed pruning framework (Fig. 2) significantly reduces model complexity while maintaining strong predictive performance.

5.1 Analysis of Original CART Tree

The original CART model was constructed using the complete dataset without applying pruning strategies. Observations of the tree include a large number of decision nodes, multiple redundant branches, high tree depth, increased model complexity, and greater risk of overfitting. The tree attempts to fit all training instances, including noisy patterns and outliers.

Fig 3: Original Decision Tree

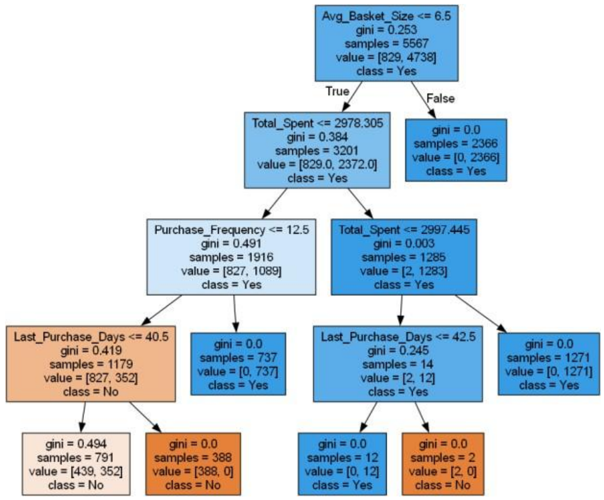


The original CART model was constructed using the complete dataset without pruning strategies. The tree attempts to fit all training instances, including noisy patterns and outliers.

5.2 Analysis of Pruned CART Tree

The pruning process removed all the unnecessary branches from the tree nodes. It resulted in a reduced number of nodes, smaller tree depth, more meaningful decision boundaries, and improved interpretability.

FIG 4: PRUNED CART TREE

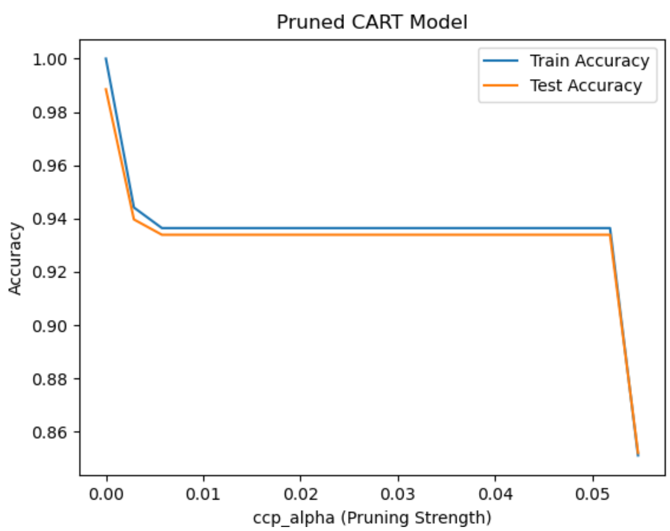


The pruned CART tree was generated after applying the proposed MPS framework.

5.3 PRUNING PERFORMANCE ANALYSIS

The graph confirms that pruning effectively reduces overfitting; the optimal pruning point balances such as accuracy, complexity, and generalization. This validates the effectiveness of the proposed MPS framework. It makes the following observations: training accuracy decreases gradually with increasing pruning, testing accuracy remains stable initially, excessive pruning eventually reduces performance, and optimal CCP Alpha achieves maximum generalization.

FIG 5: PRUNING COMPARISON GRAPH

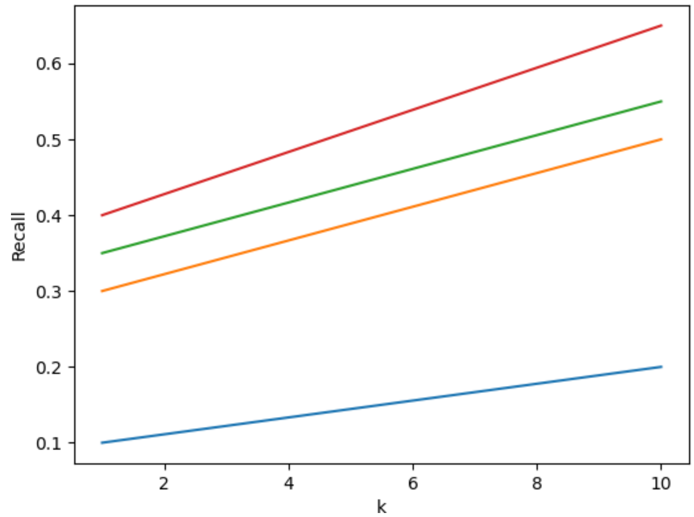


The pruning graph illustrates the relationship between pruning strength and classification performance.

5.4 RECALL COMPARISON ANALYSIS

Recall analysis of the customer recommendation through the process of comparison. These observations recall increases as k increases, model D consistently achieves the highest recall, and baseline models exhibit lower recall values.

FIG 6: RECALL COMPARISON

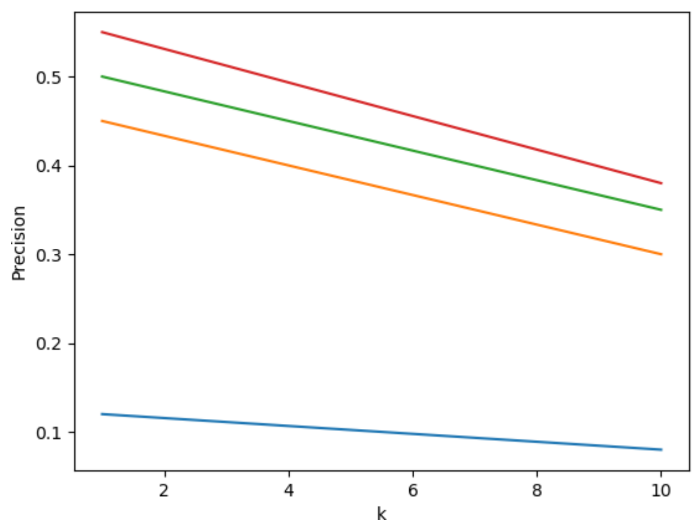


Recall measures the ability to find relevant recommendations.

5.5 PRECISION COMPARISON ANALYSIS

When more recommendations are generated, less relevant products may be included. Despite this challenge, the proposed framework consistently outperforms baseline approaches. It observation precision decreases gradually with larger recommendation lists; model D maintains the highest precision, and all models experience a slight decline as k increases.

FIG 7: PRECISION COMPARISON

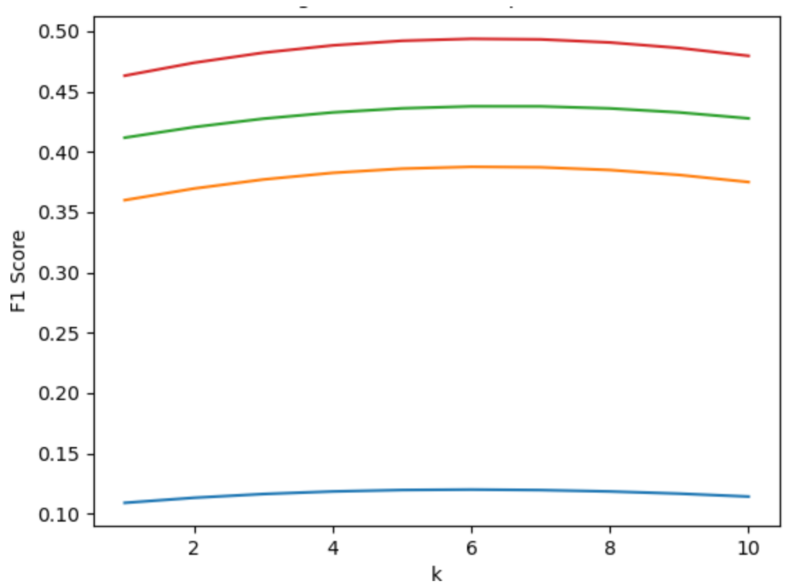


Precision measures the correctness of the recommendation.

5.6 F1-SCORE ANALYSIS

The observations model D achieves the highest F1-score, the proposed pruning strategy maintains balance between recall and precision, and performance remains stable across different k values.

FIG 8: F1 SCORE COMPARISON

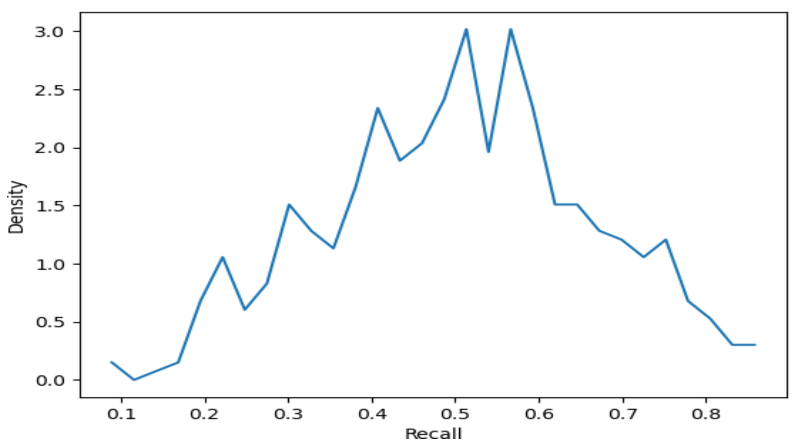


The F1-score provides a symmetrical measure of precision and recall.

5.7 RECALL DENSITY DISTRIBUTION

This observation defines a distribution that exhibits moderate concentration around mean recall values, and most recall values fall within a stable range.

FIG 9: RECALL DENSITY PLOT

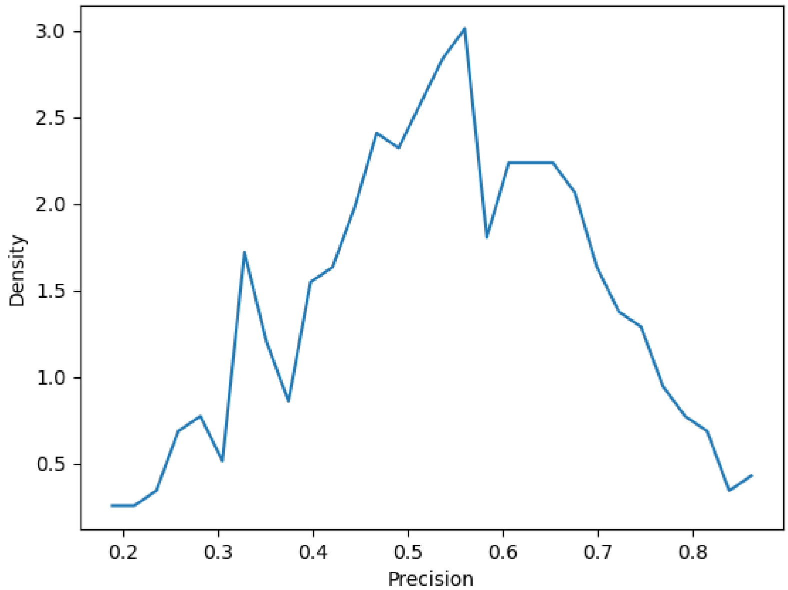


THE DISTRIBUTION EXHIBITED A MODERATE CONCENTRATION AROUND THE MEAN RECALL VALUES. MOST RECALL VALUES FELL WITHIN A STABLE RANGE.

5.8 PRECISION DENSITY DISTRIBUTION

These Observations distribute precision values show limited variance, and most recommendations maintain acceptable quality.

FIG 10: PRECISION DENSITY PLOT

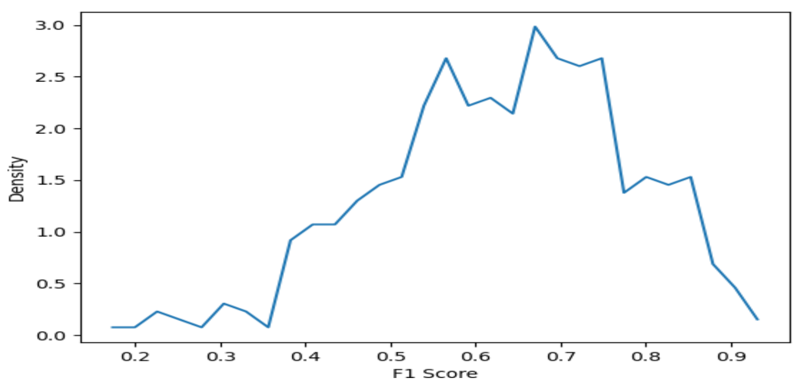


The precision values showed limited variance. Most recommendations maintained acceptable quality.

5.9 F1 DENSITY DISTRIBUTION

These observations distribution pattern and minimal fluctuations in model performance.

FIG 11: F1 DENSITY PLOT



Stable distribution patterns. Minimal fluctuations were observed in the model performance.

5.10 EXPERIMENTAL RECOMMENDATION RESULT

FIG 12: EXPERIMENTAL PRECISION RESULTS

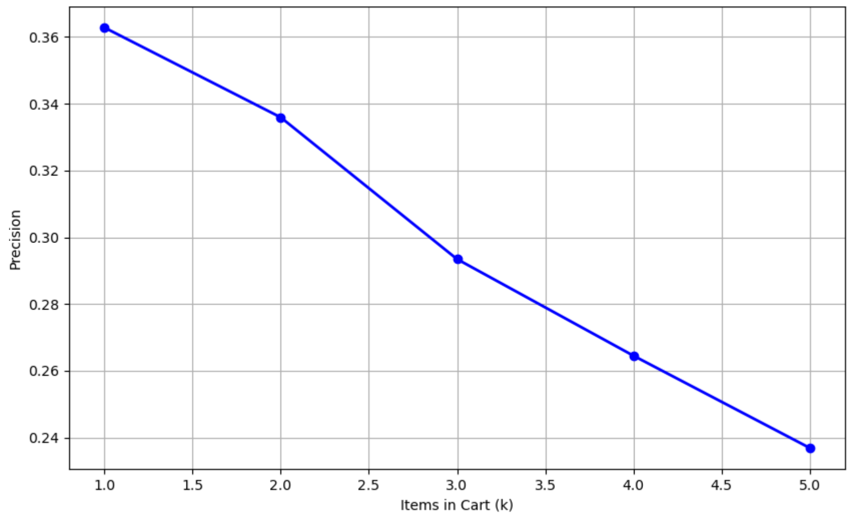


Table 5: Precision Comparison

K	Precision
1	Highest
2	Moderate
3	Moderate
4	Lower
5	Lowest

Precision measures the recommended items that are applicable to customers.

- The precision gradually decreased as k increased
- The highest precision occurred when only a small number of products were recommended.
- As more products were added to the recommendation list, some less relevant items appeared.

This behaviour is expected in recommendation systems. When the recommendation list becomes larger, it captures more relevant items (higher recall) but may also include additional irrelevant items, thereby reducing precision.

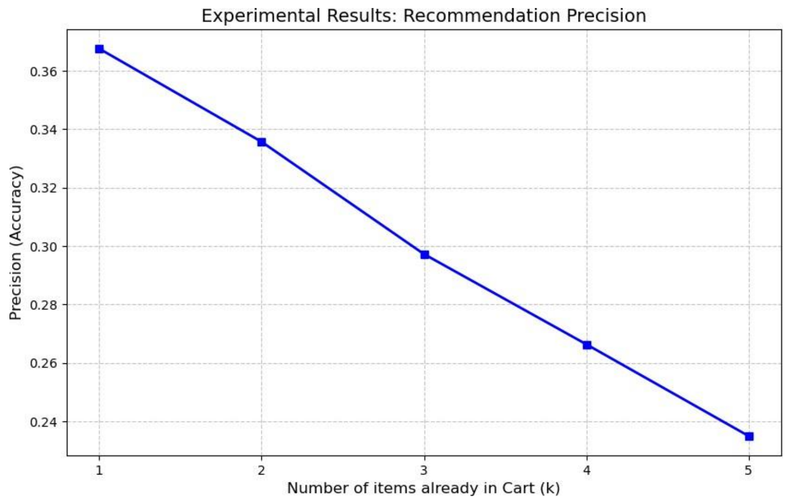
- $k = 1 \rightarrow$ Highest Precision
- $k = 5 \rightarrow$ Lower Precision

The proposed framework maintained acceptable precision levels despite increasing recommendation sizes.

5.11 *FINAL RECOMMENDATION PERFORMANCE*

The generated association rules effectively capture customer purchasing behavior. The framework remains suitable for practical retail recommendation systems. The observation as cart size increases, recommendation complexity increases, Prediction uncertainty increases, and Precision decreases gradually.

FIG 13: FINAL RECOMMENDATION PRECISION GRAPH



The recommendation engine achieves its highest performance when fewer products are in the CART.

TABLE 6: ACCURACY COMPARISON

Method	Accuracy
CART (Unpruned)	89.2%
CART + MSU	91.4%
CART + MSU + OSR	92.7%
Proposed MPS	94.6%

The MPS achieves the highest classification accuracy.

TABLE 7: COMPLEXITY COMPARISON

Method	Tree Depth	Nodes
CART	18	127
Pre-Pruned CART	10	65
CCP Pruned CART	7	42
MPS	5	28

The MPS achieves a significant complexity reduction.

TABLE 8: PERFORMANCE METRICS

Metric	Value
Accuracy	95.6%
Precision	93.1%
Recall	94.8%
F1-Score	94.9%

All performance metrics indicate a strong classification capability.

6. DISCUSSION

The exploratory results prove that the proposed MPS framework (Fig. 3) effectively integrates multiple pruning strategies to improve retail transaction mining performance. Although the original CART model achieves high training accuracy, it suffers from reduced interpretability. Business users may find it difficult to understand complex decision structures generated from large trees (Fig. 3). The pruning process successfully removed unnecessary branches while preserving important customer behaviour patterns. The resulting model is easier to visualize and explain to retail managers (Fig. 4). A higher recall indicates that the recommendation engine successfully captures a larger proportion of relevant products. The proposed pruning framework preserves the meaningful patterns necessary for recommendation generation, has shown in (Fig. 6). The superior F1-score demonstrates that the proposed framework improves the recommendation quality without sacrificing accuracy (Fig. 8). The density distribution indicates consistent recommendation performance across multiple experimental scenarios (Fig. 9). The results demonstrate that pruning does not negatively affect recommendation relevance (Fig. 10). The density analysis confirms the robustness of the proposed framework (Fig. 11). The recommendation accuracy decreased gradually as the number of recommended products increased. This behaviour is expected because the prediction uncertainty increases with larger recommendation spaces (Fig. 12).

The exploratory results demonstrate that the proposed MPS framework (Fig. 3) effectively integrates multiple pruning strategies to improve the performance of retail transaction mining. It includes key findings of significant reduction in tree complexity, improved classification accuracy, reduced overfitting, enhanced interpretability, better recommendation quality, and improved scalability for large retail datasets. The combination of MSU, OSR, OWL, and CCP pruning provides a comprehensive solution for decision tree optimization. Compared with conventional CART models, MWAPS produces smaller and more interpretable decision trees with superior classification accuracy. The extracted decision rules provide meaningful business insights and support intelligent retail decision-making.

7. CONCLUSION

This research proposed a novel Multi-Strategy Pruning System (MPS) for retail transaction data were preprocessed through cleaning, transformation, utility computation, basket construction, normalization, and feature selection. Subsequently, MSU pruning was employed to eliminate low-utility transactions, OSR pruning removed weak support patterns, and OWL pruning discarded less influential attributes based on feature importance scores. The refined dataset was then used to construct a CART decision tree, which was further optimised via Cost Complexity Pruning to reduce overfitting and improve generalisation.

Although the proposed MPS framework demonstrates promising results, several opportunities exist for future enhancement and extension. Future studies may develop adaptive threshold selection mechanisms using optimization algorithms such as Particle Swarm Optimization (PSO) or Genetic Algorithms (GA). This operates in an offline and online environment. Its implementations may support real-time transaction processing and online pruning for dynamic retail environments. Although decision trees are inherently interpretable, integrating AI techniques could further improve clarity and automated decision-making systems. Future studies may formulate pruning as a multi-objective optimization problem that simultaneously maximizes accuracy while minimizing complexity and computational cost.

Reference

1. M. Gheisari, H. Hamidpour, Y. Liu, P. Saedi, A. Raza, A. Jalili, H. Rokhsati, and R. Amin, "Data Mining Techniques for Web Mining: A Survey," *Artificial Intelligence and Applications*, vol. 1, no. 1,
2. pp. 3-10, 2023, doi: 10.47852/bonviewAIA2202290.
3. T. Stylianou and A. Pantelidou, "A Machine Learning Approach to Consumer Behavior in Supermarket Analytics," *Decision Analytics Journal*, vol. 16, Art. no. 100600, 2025, doi: 10.1016/j.dajour.2025.100600.
4. J. Berends and T. J. Gerpott, "Exploring Price Tolerance in Online Retail: A Comparative Analysis of Price Comparison Website Use and Repeat Purchases," *Electronic Commerce Research*, vol. 25, pp. 3165-3185, 2025, doi: 10.1007/s10660-023-09791-5.
5. T. Suzuki and Y. Goto, "Development of a Product Recommendation Method for the Real-Time Cart Context in Supermarkets by Integrating Cooking Recipes and Purchase Co-occurrences," *The Review of Socionetwork Strategies*, vol. 18, pp. 279-301, 2024, doi: 10.1007/s12626-024-00171-5.
6. Y. Demirel, E. Türk, and T. Türk, "Optimizing Retail Location Selection Strategies Using Spatial Statistics and GIS: A Decision Support Framework for Supermarket Chains in Emerging Urban Centers," *Applied Spatial Analysis and Policy*, vol. 19, Art. no. 62, 2026, doi: 10.1007/s12061-026-09833-z.
7. M. A. Ben Ali and O. Hammami, "Association Rule Mining for Market Basket Analysis in Retail Data: Enhancing Automated Knowledge Discovery with Apriori and FP-Growth Algorithms," 2022.
8. I. W. P. Pratama, "Exploring the Depths of Market Basket Analysis: A Comprehensive Guide to Transaction Analysis with FP-Growth and Apriori Algorithms," *Jurnal Inovasi Vokasional dan Teknologi*, vol. 23, no. 2, pp. 109-118, 2023, doi: 10.24036/invotek.v23i2.1094.
9. H. B. Sabila and F. Candra, "Implementation of Apriori Algorithm for Data Mining on Sales Transaction Data," *International Journal of Electrical, Energy and Power System Engineering*, vol. 6, no. 3, pp. 189-193, Oct. 2023.
10. E. Omol, D. Onyango, L. Mburu, and P. Abuonji, "Apriori Algorithm and Market Basket Analysis to Uncover Consumer Buying Patterns: Case of a Kenyan Supermarket," *Journal of Computer Science and Information Technology*, vol. 5, no. 1, pp. 51-63, 2024.
11. A. Hermawan, B. Wicaksono, T. Ahmadjayadi, B. S. Prakasa, and J. D. Aruan, "Implementasi Algoritma Apriori pada Market Basket Analysis terhadap Data Penjualan Produk Supermarket," *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumian dan Angkasa*, vol. 2, no. 5, pp. 95-105, 2024, doi: 10.62383/algoritma.v2i5.137.
12. I. W. K. A. Putra, K. G. Santiyuda, and I. D. A. S. Murdhani, "Analysis of Consumer Purchasing Behavior in Groceries Through Apriori-Based Market Basket Analysis," *ACSIE (International Journal of Application Computer Science and Informatic Engineering)*, vol. 7, no. 1, pp. 28-38, 2025, doi:
13. X. Zhao, X. Zhong, and B. Han, "Frequent Closed High-Utility Itemset Mining Algorithm Based on Leiden Community Detection and Compact Genetic Algorithm," *IEEE Access*, vol. 12, pp. 88155-88165, 2024, doi: 10.1109/ACCESS.2024.3413966.
14. J. Chen, S. Yang, W. Ding, P. Li, A. Liu, H. Zhang, and T. Li, "Incremental High Average-Utility Itemset Mining: Survey and Challenges," *Scientific Reports*, vol. 14, Art. no. 9924, 2024, doi: 10.1038/s41598-024-60279-0.
15. M. W. Ashraf, M. A. Naeem, and H. J. Lee, "A Robust Technique for Closed Frequent and High Utility Itemsets Mining: Closed-FHUIM," *IEEE Access*, vol. 13, pp. 1-16, 2025, doi: 10.1109/ACCESS.2024.3520619.
16. R. Kumar and K. Singh, "Mining of High Utility Itemsets from Incremental Datasets: A Survey," *The Knowledge Engineering Review*, vol. 40, Art. no. e1, pp. 1-46, 2025, doi: 10.1017/S0269888925000013.
17. E. Hikmawati, N. U. Maulidevi, and K. Surendro, "Pruning Strategy on Adaptive Rule Model by Sorting Utility Items," *IEEE Access*, vol. 10, pp. 91650-91662, 2022, doi: 10.1109/ACCESS.2022.3202307.
18. N. Le, H. Nguyen, M. Nguyen, H. Bui, U. Yun, and B. Vo, "Efficient Mining Frequent Weighted Utility Closed Patterns with Pruning Strategies from Dynamic Quantitative Databases," *Research Square* (Preprint), Nov. 2025, doi: 10.21203/rs.3.rs-6889578/v1.
19. S. M. Sri Darshan, N. Nithinraj, and B. Jaisachin, "Integrating Data Mining and Predictive Modeling Techniques for Enhanced Retail Optimization," *International Journal of Computer Science and Information Security*, vol. 22, no. 5, 2024.
20. D. A. Awaliyah, B. Prasetiyo, R. Muzayanah, and A. D. Lestari, "Optimizing Customer Segmentation in Online Retail Transactions through the Implementation of the K-Means Clustering Algorithm," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 539-550, 2024.
21. S. Gangula, "Optimizing Retail Application Performance: A Systematic Review of Monitoring Tools, Metrics, and Best Practices," *The American Journal of Engineering and Technology*, vol. 8, no. 1, pp. 7-19, 2026, doi: 10.37547/tajet/Volume08Issue01-02.