THE SCIENTIFIC TEMPER (2026) Vol. 17 (4): 6077-6085
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.18
https://scientifictemper.com/
RESEARCH ARTICLE

Genetic Algorithm-Based Adaptive Pattern Mining for Customer Basket Analysis

T. Javith Hussain[1*], K.N. Abdul Kader Nihal[2]

Abstract

Customer basket analysis aims to understand purchasing patterns and enhance marketing strategies by leveraging data mining techniques. To the best of our knowledge, this study is one of the few that applies low-support adaptive pattern mining in personalization, introducing a robust genetic algorithm for improving static transaction baskets. The proposed method uses GA integrated with sophisticated pattern mining methods that quickly finds complex associations as well as patterns from transaction data. By utilizing GAs, our model has the ability to adapt and evolve to fit this changing landscape of consumer purchasing behaviors. The GA-based method is compared to traditional data mining methods using a large dataset of retail transactions. Further it shows big improvements in both accuracy and speed. The results indicate that the adaptive nature of the proposed method not only helps uncover purchasing patterns but also supports retailers in making informed decisions on inventory control, promotional activities, and personalized marketing. The findings provide a promising insight into how implementing genetic algorithms in combination with pattern mining frameworks could yield more proactive and responsive customer analysis within the industry of retail.

Keywords: Genetic Algorithms, Adaptive Pattern Mining, Customer Basket Analysis, Retail Analytics, Market Basket Analysis and Machine Learning.

Introduction

The growing access to extensive datasets in the retail industry has highlighted the need for sophisticated analytical methods to better understand and interpret consumer behavior. As retailers strive to leverage vast amounts of information generated from consumer transactions, the limitations of traditional methods become apparent (Ishumbaev, 2025). Financial penalties make a huge difference in consumer behavior, particularly within the countries they target. People feel pressure to support home businesses, or to boycott foreign goods that are traceable to the punishing countries as a reaction to penalties, researchers have said. Periods of economic sanctions often intensify nationalistic attitudes, resulting in a stronger inclination toward domestically produced goods rather than imports. (Sezer et al., 2026). This preference is especially noticeable in markets where national alternatives may be preferred over specific brands. As Companies find that the CMSB is set to change in ways driven by Geopolitics, they must market differently. To stabilize their share in the market during challenging times (Massari et al., 2025), companies should strengthen value propositions, create strong national brands and support local production capabilities. Focusing on such aspects as brand resilience, local value chains and patriotic narratives can take consumers a long way in the loyalty and trust domain. Traditional methods of customer basket analysis are inadequate in capturing product combination interactions that can form the basis for cross-selling or inventory management (Matthews et al., 2013). Traditional methods, though valuable in less complex analytical scenarios, tend to rely heavily on fixed structures that cannot change with the adaptive landscape of consumer appetites and buying behaviors. As a result, retail companies find it challenging to design precise marketing

[1]Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, India

[2]Assistant Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, India

***Corresponding Author:** T. Javith Hussain, Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620020, India, E-Mail: tjh@jmc.edu

How to cite this article: Hussain, T.J., Nihal, K.N.A.K. (2026). Genetic Algorithm-Based Adaptive Pattern Mining for Customer Basket Analysis. The Scientific Temper, **17**(4):6077-6085.

Doi: 10.58414/SCIENTIFICTEMPER.2026.17.4.18

Source of support: Nil

Conflict of interest: None.

Received: 08/04/2026 **Accepted:** 19/04/2026 **Published:** 25/04/2026

strategies that connect with their customers and efficiently manage inventory in line with demand.Search heuristics that replicate the basis of natural selection which must be employed to GA-based adaptive pattern mining (Reinhardt et al., 2018). It have been effectively utilized in multiple fields, including optimization and machine learning. This research leverages the fundamental principles of genetic algorithms to provide a framework that dynamically adjusts to evolving consumer patterns, facilitating the identification of significant purchasing habits that static analysis methods may struggle to reveal (Kavitha et al., 2026). Genetic Algorithm make takes use of for mapping relationship among multiple products by the transaction datasets. In the process of selection, crossover and mutation, this algorithm produces patterns that uncover subtle correlations in what customers buy. Unlike conventional methods that rely on fixed rules or thresholds, the adaptive capabilities of genetic algorithms allow systems to evolve alongside shifting consumer preferences, ensuring consistently accurate insights derived from customer behavior data (Zandi et al., 2025). To evaluate the effectiveness of our GA-driven adaptive pattern mining method, It's conducted a comparative analysis against conventional mining techniques using a dataset drawn from thousands of retail transaction records. The proposed strategy's ability to enhance understanding of consumer behavior will be assessed through key performance indicators, including classification accuracy and computational efficiency. The ability to learn and adapt from a constantly changing dataset positions this technique among the most impactful breakthroughs available to modern data-centric enterprises (Athanasios et al., 2026). Consumer decisions are influenced by the credibility and operational resilience of brands. In times of macro-economic adversity, brand resilience defined by the capacity to overcome anti-market pressures like those caused by sanctions is becoming an important qualitative attribute that determines consumer trust and loyalty.

Literature Review

Data mining techniques designed to model customer purchasing behavior seek to identify the likelihood that a buyer who acquires one product category will subsequently select a complementary set of items. A collection of commodities in the basket is called item sets and the relevant relations among item sets is enclosed as dependence rules originally used in retail environments, a market basket analysis looks for relationships and associations between items present in business transactions to help identify which items would look good together within a catalogue design.

Cavique (2007) identify a substantial Frequent patterns for market basket analysis. This method utilizes condensed data derived from converting the market basket problem into a maximum-weighted clique problem. Initially, the input dataset is converted into a graph-based structure, after which the maximum-weighted clique problem is addressed by a meta-heuristic method to identify the most frequent patterns.

Lim et al. (2012) integrating a genetic algorithm with association criteria. Initially, they employed correlation to substitute support-confidence in the genetic algorithm, facilitating a dynamic data-driven assessment of support and confidence, specifically utilizing correlation to enhance the extraction of positively correlated association rules. Secondly, they employed correlation as the fitness function to facilitate upward closure in association rules, whereas previously, association rules only supported downward closure. The capacity to facilitate upward closure enables the derivation of highly specific association rules (business model) from less specific association rules and generic association rules.

Arthur et al. (2022) suggested that Contemporary research predominantly focuses on item-user relationships, neglecting intricate heterogeneous connections and compelling item/user data, which leads to sparsity and cold-start issues, hence hindering the effectiveness of Neighborhood-Based Recommendation Systems (NBRS). The basket prediction layer employs an enhanced Particle Swarm Optimization (PSO) method to optimize the network's weights and biases, facilitating the iterative learning of diverse couplings and influential information both inside and between baskets to recommend the subsequent basket. In PSO, its integrate the exploratory Gravitation Search Algorithm to equilibrate exploration and exploitation. This architecture incorporates an Adaptive Response to Particle Adjustment Strategy to enhance exploitation and avert particle entrapment.

Telikani et al. (2020) examine recent studies on evolutionary computation for ARM. We examine the applicability of evolutionary computations across various ARM methodologies, including numerical rules, fuzzy rules, high-utility item sets, class association rules, and rare association rules. Based on their evolutionary mechanisms, ARM algorithms of evolutionary nature are organized into four fundamental groups: those driven by evolution, swarm intelligence, physics-based principles, and hybrid combinations. Additionally, it examines the outstanding issues of evolutionary ARM and its applications, as well as prospective themes for future exploration.

He et al. (2022) introduces an innovative technique for high utility occupancy itemset mining, termed SHO (Suffix-based High-utility Occupancy itemset mining), which takes into account both the quantities and profits of item sets. SHO formulates the algorithm utilizing suffix-based segmentation, generation pruning, and itemset linkage, enabling the effective mining of high utility occupancy item sets from large-scale databases. Initially, the database is segmented into non-overlapping suffix-based partitions and stored in a vertical manner, allowing for the calculation of the support and utility occupancy of item sets within a specific partition rather than traversing the entire database. Additionally, two optimization algorithms and four trimming strategies are introduced to enhance the speed of SHO. The comprehensive studies demonstrate that SHO significantly outperforms the existing state-of-the-art technique, with efficiency improvements of up to three orders of magnitude.

Kumar and Mohbey (2022) delivers a thorough examination of the diverse approaches employed in pattern mining within the context of Big Data. The initial focus is on the challenges inherent to pattern mining methods and their associated technologies, including Apache Hadoop, Apache Spark, and parallel and distributed computing frameworks. The study then proceeds to analyze notable progress in scalable, parallel, and distributed pattern mining from a Big Data standpoint, while pinpointing the obstacles in developing efficient algorithms. A central emphasis is placed on four distinct forms of frequent pattern, namely parallel frequent pattern mining, high utility itemset mining, sequential pattern mining, and frequent pattern mining under data uncertainty. The paper concludes with an exploration of open challenges and future research directions, along with recommendations for advancing current approaches.

Methodology

The proposed methodology in this research is comprised of a Genetic Algorithm-Based Adaptive Pattern Miner (GA-APM) that plays a major role in the process of pattern discovery. The proposed algorithm describes the product combinations solutions as candidate solutions and evolves these candidates using evolutionary operators (selection, crossover, and mutation). Through this optimization-driven process, quality and non-trivial product associations areachieved without predefined thresholds. Despite the propagation effort of channel representation on generic settings, to manage large and dynamic data, this methodology uses an efficient search strategy where you augment potential patterns by refining existing ones rather than generating new candidates.

Figure 1 illustrates the workflow architecture of the proposed GA-APM model, which is designed to efficiently extract meaningful customer basket patterns from large-scale retail transaction datasets. The process begins with the retail transaction dataset, where data preprocessing is performed to remove inconsistent records and integrate multiple relational files into a unified dataset. The preprocessed transactions are then transformed into a binary basket representation, where each row denotes a transaction and each column indicates the presence or absence of a product. During the population initialization stage, random candidate basket patterns are generated for evolutionary search. The GA-APM algorithm then evolves these candidates through successive generations. In each generation, fitness is evaluated using support, confidence, lift, and novelty, followed by the application of genetic operators such as selection, crossover, and mutation. The optimized patterns are finally assessed using performance metrics including Precision, Recall, F1-Score, MAP, and AUC. These optimized patterns provide actionable retail insights for product placement, personalized recommendations, and marketing strategies.

Dataset Description

In the current study, the Instacart Market Basket Analysis Dataset is utilized, which is a large-scale real-world grocery transaction dataset widely used in retail analytics and recommendation system research. The data set contains nearly 3.4 million customer orders made by over 200,000 customers, with approximately 50,000 products and is organized into aisles and departments. Its relational structure facilitates analysis of the customer purchasing behavior through an integration of order-level, product level and category-level information. The dataset used in this research is key to reconstructing customer shopping baskets and discovering product co-occurrence patterns necessary for association rule mining. The merged table contains order history, prior purchases, product metadata and other products that are then combined to build a seamless transaction file among those files. The integrative ensemble helps in fast encoding of basket information and further aids to deploy the proposed GA-APM for mining reasonable, non-trivial association among products. Table 1 provides an overview of the attributes and their contribution to this study.

This merged dataset allows for a comprehensive and scalable analysis of customer purchasing behavior. The use of both transactional and categorical attributes captures complex relationships between products but also improves the effectiveness of the evolutionary algorithm proposed to obtain hidden basket patterns.

Data Preprocessing

The data preparation of raw transaction is essential to get accurate, consistent and computationally efficient data before applying the proposed evolutionary algorithm. The first step to be done here is merging them into one structure since the dataset consists of multiple relational files. In this step, common identifiers like order_id and product_id merge transactions against their full product and category information. Typically, operations of this merging can be represented as a join operation.

$$T = Orders \bowtie Order\ Products \bowtie Products \quad \text{Eq. (1)}$$

Where T represents the integrated transaction dataset used for analysis. The next step involves cleaning the data to

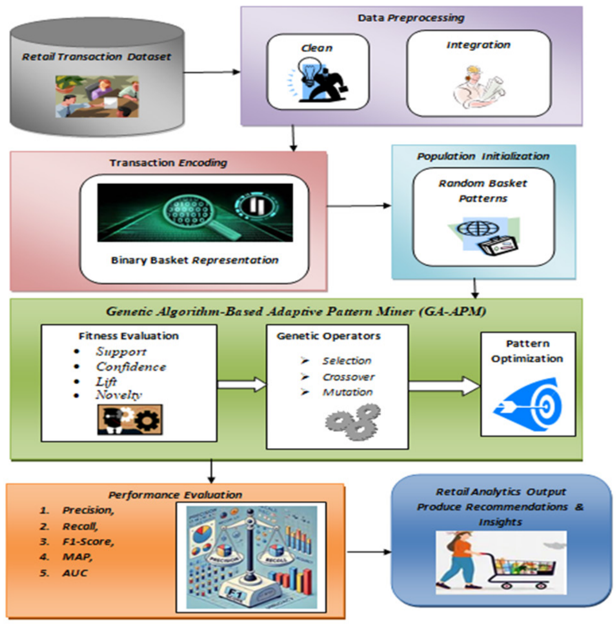

Figure 1: Workflow Architecture of the Proposed GA-APM Algorithm

remove inconsistencies that may hinder pattern discovery. Let N be the total number of records and Nd be duplicate or/spurious data entries, so the size of our cleaned dataset becomes:

$$N_{clean} = N - N_d \quad \text{Eq. (2)}$$

In short: a 2 step equation that ensures only real and meaningful transactions are considered. After the cleaning step, we use each customer order to create a transaction basket, which consists of all products purchased with one particular order. A transaction can be described by the formula:

$$T_i = \{p1, p2, p3, \ldots\ldots\ldots, pn\} \quad \text{Eq. (3)}$$

Where T_i represents the *i-th* transaction, and p_j represents the *j-th* product contained in the transaction. This representation is critical for detecting co-occurrence relationships of products in the same basket. In order to make it easy to work with for the evolutionary algorithm, transaction data is translated into binary matrix format. In this matrix, each row represents a transaction and each column represents a product. Equation (4) represents the binary encoding of product presence or absence in each transaction.

$$x_{ij} = \begin{cases} 1, & \text{if product } j \text{ is present in transaction } i \\ 0, & \text{otherwise} \end{cases} \quad \text{Eq. (4)}$$

Binary encoding also makes it easier to manipulate the product combinations, and enables chromosomes in our genetic algorithm to be represented as binary vectors. Lastly, dimensionality reduction is performed for computational complexity. This minimum support is used to remove products with a very low occurrence, as such retail datasets often contain a large number of low-purchase items. The support of a product p is computed as the following equation 5:

$$Support(p) = \frac{\text{Number of transactions containing } p}{\text{Total number of transactions}} \quad \text{Eq. (5)}$$

Only products satisfying Support(p)≥ min_sup are retained for further analysis. This step reduces the size of the dataset while preserving the most relevant and frequent product relationships. Through these preprocessing steps, the dataset is transformed into a structured and optimized format, enabling the proposed GA-APM to efficiently discover meaningful customer basket patterns.

It keeps only those products to further analysis which satisfy the condition: Support(p)≥ min_sup. This process allows to minimize the dataset keeping most of the useful and repeated product relationships. These preprocessing steps ensure that the dataset is converted into a structured and optimized format which will allow to efficiently utilize the proposed GA-APM in finding meaningful customer basket patterns.

Table1: Instacart Market Basket Analysis Dataset Description

Dataset File	*Key Attributes*	*Description*	*Contribution to Study*
orders.csv	order_id, user_id, order_number, order_dow	Contains customer order sequence and timing information	Enables reconstruction of customer purchase history and behavior patterns
order_products_prior.csv	order_id, product_id, add_to_cart_order, reordered	Records previously purchased products	Provides historical transaction data for pattern discovery
order_products_train.csv	order_id, product_id, reordered	Contains labeled product purchases	Supports validation and evaluation of discovered patterns
products.csv	product_id, product_name, aisle_id, department_id	Contains product details and category mapping	Helps identify relationships between product categories
aisles.csv	aisle_id, aisle	Defines product grouping within aisles	Supports fine-grained analysis of product associations
departments.csv	department_id, department	Represents higher-level product classification	Enables broader pattern analysis across departments

Genetic Algorithm-Based Adaptive Pattern Miner (GA-APM)

The objective of Customer Basket Analysis is to discover significant relations between products that are purchased together often in transactional data. However, the problem is more complicated as retail companies have millions of products (the high dimensionality) but only few transactions for other consumer purchases (the sparsity of completed transactions), and they also sell both frequent combinations and rare but expensive items. Classical association rule mining approaches are governed by fixed minimum support and confidence criteria, which tend to produce a large volume of redundant and trivial rules on one hand, while eliminating potentially significant rules that exhibit lower frequency patterns on the other.To improve on these limitations, the proposed GA-APM defines customer basket analysis as an optimization venture, with the objective of identifying high-quality configurations of product interactions based on multiple balancing evaluation criteria.

It starts with the creation of a population of candidate patterns. Let P= {C1,C2,C3,………. Cm } each chromosome Ci is represented with a binary vector of n bits, where n is number of products. Each chromosome comprises a defined set of genes responsible for carrying hereditary information, each taking a value in {1, 0}, representing whether a product appeared (1) or not (0) in the pattern. In this method, each candidate basket pattern is represented as a chromosome C_i, encoded as a binary vector.

$$C_i = \left(g1, g2, g3, \ldots\ldots\ldots, gn\right), \qquad g_j \in \{0,1\} \qquad \text{Eq. (6)}$$

Where equation (6) gj=1 indicates the presence of product j in the pattern By representing the data structure in this way, the algorithm can efficiently traverse a combination of products through a very large search space without generating all possible item sets. Proportion of combined product measure is a tough challenge in customer basket analysis. To this, the algorithm computes the statistical measures for every chromosome.

The support of a pattern X is defined as:

$$Support(X) = \frac{|\{T_i|\ X \subseteq T_i\}|}{|T|} \qquad \text{Eq. (7)}$$

Which indicates how often a product pair appears in the transaction set. But using support in isolation may overlook less common yet significant relationships. Hence, equation 8 confidence is added to evaluate the solidity of relations:

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \qquad \text{Eq. (8)}$$

This guarantees that the chosen patterns represent uniform purchasing behavior. Another problem with basket analysis is finding real relationships rather than just coincident co-occurrences. Equally, is used in equation 9 to solve this.

$$Lift(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X).Support(Y)} \qquad \text{Eq. (9)}$$

A lift value greater than one indicates a strong dependency between products being bought together, helping prevent false associations. Another important challenge in retail analytics is the discovery of hidden or rare patterns that are often ignored by frequency-based measures. To address this issue, a novelty component is introduced in Equation (10). A novelty component is introduced with equation 10 to encourage the discovery of such patterns:

$$Novelty(X) = \frac{1}{1 + Support(X)} \qquad \text{Eq. (10)}$$

This term gives more weight to infrequent but interesting itemsets, thus overcoming a major limitation of threshold-based methods. These metrics are combined into a single fitness function:

$$F(C_i) = w1.Support + w2.Confidence + w3.Lift + w4.Novelty \qquad \text{Eq.(11)}$$

Where Eq. (11),w1,w2,w3,w4 control each attribute contribution. This formulation gives the algorithm room to trade-off between frequency, reliability, correlation strength, and uniqueness when evaluating patterns. Evolutionary processes start with the creation of an initial set of candidate patterns. Using the fitness function, each chromosomes is evaluated, where patterns with higher fitness values are picked over others as preferences during reproduction. A chromosome will be selected with a probability according to equation (12):

$$P(C_i) = \frac{F(C_i)}{\sum_{j=1}^{m} F(C_i)} \qquad \text{Eq. (12)}$$

This genetic algorithm is a probabilistic mechanism that encourages high patterns while sustaining diversity in the population. Crossover is performed by exchanging segments of parent chromosomes to try new combinations of products:

$$Offspring_1 = \left(g_1^A \ldots\ldots\ldots g_k^A, g_{k|1}^B, \ldots\ldots\ldots, g_n^B\right) \qquad \text{Eq.(13)}$$

$$Offspring_2 = \left(g_1^B \ldots\ldots\ldots g_k^B, g_{k|1}^A, \ldots\ldots\ldots, g_n^A\right) \qquad \text{Eq.(14)}$$

Equation 13,14: General Operation to Generate New Candidate Patterns Then, mutation is applied to provide random variations, which is important for sparse and diversified customer transactions:]

$$g_j^{'} = \begin{cases} 1 - g_j, & if\ rand() < p_m \\ g_j & otherwise \end{cases} \qquad \text{Eq. (15)}$$

Specifically, using mutation, the algorithm can identify hidden products combinations and better explore the search space. The population is then updated by replacing lower fitness chromosomes with new offspring. This process of iteratively refining the population continues until a termination criterion is satisfied, such as achieving a set maximum number of generations or when there is no significant difference between the best fitness values:

$$| F_{best}^{(t)} - F_{best}^{(t-1)} | < \epsilon \quad \text{Eq. (16)}$$

All these considerations make GA-APM suitable solution for customer basket analysis challenges such as: candidate redundancy, having a better boolean form in context of their frequencies and also solving the challenge of large search space. Using a multi-objective fitness function, which is based on four criteria highlighting interesting associations are not frequent with a strong correlation but also reliable and potentially novel relationships, provides unique insights into item associations that can be very beneficial for retail recommendation systems and decision-making processes.

Algorithm: Genetic Algorithm-Based Adaptive Pattern Miner (GA-APM)

Input

T (Transaction dataset), N (Population Size), G(Maximum number of generations), pc(Crossover Probability), pm(Mutation Probability), min_sup(Minimum support threshold).

Output: *Best set of optimized basket patterns*

Begin

1. Preprocess dataset T and convert into binary transaction matrix.
2. Initialize population P with N Chromosomes

Each Chromosome Ci represents a product combination (binary vector)

3. For each chromosome Ci in P do

Compute Support(Ci)

Compute Confidence(Ci)

Compute Lift(Ci)

Compute Novelty(Ci)

Evaluate Fitness:

$F(C_i) = w1.Support + w2.Confidence + w3.Lift + w4.Novelty$.

End For

4. Set generation count t=1
5. While (T<=G) do

a. Selection: Parent chromosomes from P based on fitness

$$P(C_i) = \frac{F(C_i)}{\sum_{j=1}^{m} F(C_i)}.$$

b. ossover: For each pair of selected parents

If rand()<pc then

Perform single-point crossover

Generate offspring

End If

End for

c. Mutation: For each offspring chromosome

For each gene gj

If rand() < pm then

Flip gene value(0->1 or 1->0)

End If

End For

End For

d. Fitness Evaluation: Compute Support, Confidence, Lift, Novelty

Calculate Fitness for all offspring

e. Population Update: Replace low-fitness chromosomes in P with new offspring

f. Check Convergence:

If |F_best(t)-F_best(t-1)|<ò then .

reak

End If

t= t +1

End While

6. Select best chromosomes with highest fitness values
7. Return optimized basket patterns

End

This algorithm describes the GA-APM that consists of transforming the transaction dataset into binary input and initializing a set of candidate product combinations as chromosomes. Then each chromosome is evaluated to providing a multi-objective fitness function, including the support, confidence, lift and novelty to measure the quality of basket patterns. According to their fitness values, good patterns will be selected and evolved through genetic operation like crossover and mutation to produce new combinations of products and enhance the variety in search space. The population is repeatedly updated, with inferior solutions being replaced by better offspring until a termination condition (usually some convergence criteria or maximum generations) is met. At the last edge, we get the best basket patterns that are scored as meaningful and qualitative product pairs to apple store basket exploration.

Results and Discussion

The proposed GA-APM algorithm was evaluated using the Instacart Market Basket Analysis Dataset to measure its effectiveness in discovering meaningful customer purchasing patterns. The experimental results demonstrate that the proposed model achieves superior performance in terms of Precision, Recall, F1-Score, MAP, and AUC when compared with conventional approaches such as Apriori and FP-Growth.

Pattern Discovery Performance

It shows in experiments that the GA-APM algorithm is able to discover frequent and hidden dependencies among products based on the transaction data. Instead of relying on strict frequency thresholds defined like the traditional approaches, our evolutionary search mechanism allowed for a wider space exploration which led to the finding of

Table 2: Performance of Proposed GA-APM

Metric	*Performance analysis (%)*
Precision	92.68
Recall	89.57
F1-Score	90.94
Mean Average Precision (MAP)	88.63
AUC	94.52

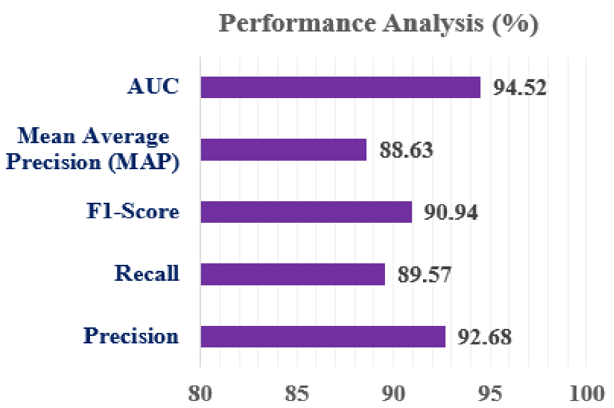

Figure 2: Performance analysis of the Proposed GA-APM Algorithm

non-trivial high impact basket patterns. This algorithm was able to catch the co-occurrence relationships between products, including rarer but meaningful combinations that greatly contribute to recommending quality. This capability is especially crucial for retail analytics, wherein infrequent associations are often the key to cross-selling opportunities. Founded patterns can be implemented as-is in recommender systems to provide better product recommendations and allow a more pleasant experience for the customers. Table 2 shows the performance of proposed GA-APM algorithm.

The performance table shows values for proposed model and the high-performance output of customer basket analysis. Here, accuracy of 92.68% indicates majority of the product relationships predicted are true and relevant. The recall value of 89.57% tells us that the model is capable to capture most, if not all but 10% or so of actual meaningful product relationships in the dataset. An F1-Score of 90.94%, which serves an analogy for the most balanced metric for this combined model performance, speaks to no favored score between precision and recall.

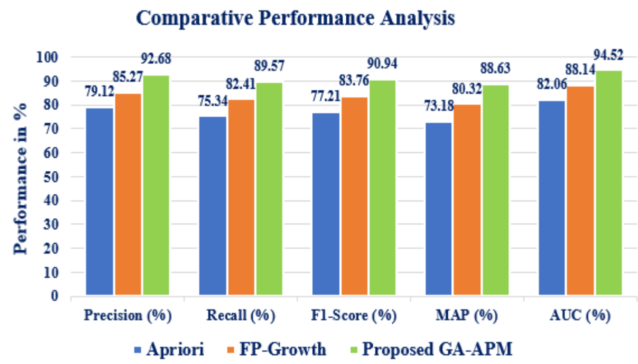

Figure 3: Comparative analysis of the Proposed GA-APM Algorithm

This is important for recommendation systems, since the MAP value of 88.63% shows that our model ranks best product associations at high positions. The area under curve (AUC) value was also 94.52%, providing further support for the model's strong ability to separate relevant from irrelevant pattern types. Figure 2 suggest all these approaches are relatively efficient methods for finding accurate, reliable and meaningful customer pattern.

Comparative Analysis of Algorithms

Proposed method was validated with two benchmark algorithms based on Apriori Algorithm and FP-Growth Algorithm which was used as a compare algorithm to GA-APM.

Table 3 clearly shows that the proposed GA-APM algorithm outperforms traditional methods in all metrics. The accuracy of GA-APM is 92.68% that is markedly greater than Apriori (79.12%) and FP-Growth (85.27%), providing indication about relative goodness of the product associations obtained by our approach Likewise, the recall value of 89.57% showcases the efficiency with this model manages to identify more of the significant basket patterns while Apriori and FP-Growth mere achieve lower recall values of 75.34%, and 82.41% respectively. Further evidencing a strong balance between precision and recall in the proposed solution, as evidenced by F1-Score↑ of 90.94%.

In addition to classification performance, the ranking and discriminative capabilities of the proposed model are also significantly enhanced. The Mean Average Precision (MAP) of 88.63% indicates that GA-APM effectively prioritizes relevant product associations, thereby improving recommendation quality. Furthermore, the AUC value of 94.52% highlights the model's strong ability to distinguish between relevant and

Table 3: Comparative Performance Analysis

Algorithm	*Precision (%)*	*Recall (%)*	*F1-Score (%)*	*MAP (%)*	*AUC (%)*
Apriori	79.12	75.34	77.21	73.18	82.06
FP-Growth	85.27	82.41	83.76	80.32	88.14
Proposed GA-APM	92.68	89.57	90.94	88.63	94.52

irrelevant patterns, outperforming both Apriori (82.06%) and FP-Growth (88.14%). These improvements can be attributed to evolutionary optimization process, which enables efficient exploration of the search space and identification of both frequent and hidden product relationships. Overall, the percentage-based results demonstrate that the proposed GA-APM algorithm provides a more robust and scalable solution for customer basket analysis in large retail datasets.

Metric-wise Performance Analysis

The GA-APM algorithm achieved a precision of 92.68%, indicating that the evolutionary selection mechanism effectively filters out weak and irrelevant product associations. In comparison, Apriori and FP-Growth achieved lower precision values of 79.12% and 85.27%, respectively, largely due to redundant candidate generation and reliance on strict frequency thresholds. The recall of 89.57% demonstrates GA-APM's ability to capture a wider range of meaningful associations, including less frequent but significant product combinations that traditional algorithms often overlook.

The F1-Score of 90.94% reflects a strong balance between precision and recall, confirming that the model consistently identifies relevant patterns while minimizing false associations. Similarly, the Mean Average Precision (MAP) of 88.63% shows that GA-APM ranks relevant product associations higher, which is critical for recommendation systems. The AUC value of 94.52% further highlights model's ability to distinguish between relevant and irrelevant patterns, demonstrating robust classification performance. Collectively, these results confirm that GA-APM outperforms conventional algorithms across all evaluation metrics, ensuring accurate, reliable, and well-ranked basket pattern discovery.

Conclusion

The experimental results validate the effectiveness of the proposed GA-APM algorithm in addressing the challenges of customer basket analysis. Unlike conventional approaches such as Apriori and FP-Growth, GA-APM avoids exhaustive candidate generation and significantly reduces computational complexity while preserving pattern quality. A major advantage of the proposed model is its ability to identify hidden and low-frequency yet meaningful product associations that traditional methods often overlook. By integrating multiple evaluation metrics such as support, confidence, lift, and novelty within a unified fitness function, the algorithm ensures robust and valuable pattern discovery. The evolutionary optimization process further enables adaptability to dynamic and large-scale retail datasets. Overall, GA-APM provides a scalable and intelligent solution for recommendation systems, inventory management, and targeted marketing strategies.

Acknowledgement

We acknowledge the support of the PG & Research Department of Computer Science, Jamal Mohamed College (Autonomous), Trichy for providing computational resources. This research was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Arthur, J. K., Zhou, C., Osei-Kwakye, J., AppiahMantey, E. and Chen, Y. (2022). A heterogeneous couplings and persuasive user/item information model for next basket recommendation. *Engineering Applications of Artificial Intelligence*, 114. https://doi.org/10.1016/j.engappai.2022.105132

Athanasios, T., Vasilakis, C., Charalampakis, D., Vasilakis, E., Thomaidis, N. and Dounib, G. (2026). Decision support for airfare forecasting using an evolutionary approach: Evidence on search engine data from touristic agency. *Journal of Air Transport Management*, 131. https://doi.org/10.1016/j.jairtraman.2025.102917

Cavique, L. (2007). A scalable algorithm for the market basket analysis. *Journal of Retailing and Consumer Services*, 14(6), 400–407. https://doi.org/10.1016/j.jretconser.2007.02.003

He, J., Han, X., Wang, J. and Zhang, K. (2022). Efficient high-utility occupancy itemset mining algorithm on massive data. *Expert Systems with Applications*, 210. https://doi.org/10.1016/j.eswa.2022.118329

Ishumbaev, K., Zhu, Z., Ileye, Y. and Armah, A. K. (2025). Consumer behavior and market substitution in the automobile retail sectors of Russia and China amidst global economic sanctions and uncertainty. *Journal of Retailing and Consumer Services*, 87. https://doi.org/10.1016/j.jretconser.2025.104441

Kavitha, D., Ashwin Kumar, S., Divya Priya, B. A., Guru Prasadh, M. V., Sri Krishna, S., SidhParakhShiram, V. and Rashid, T. (2026). A fine-grained multi-lingual opinion mining method on social media texts using multi-scale fused features-based adaptive residual convolutional LSTM with attention mechanism. *Data & Knowledge Engineering*, 161. https://doi.org/10.1016/j.datak.2025.102524

Kumar, S. and Mohbey, K. K. (2022). A review on big data-based parallel and distributed approaches of pattern mining. *Journal of King Saud University - Computer and Information Sciences*, 34(5), 1639–1662. https://doi.org/10.1016/j.jksuci.2019.09.006

Lim, A. H. L., Lee, C.-S. and Raman, M. (2012). Hybrid genetic algorithm and association rules for mining workflow best practices. *Expert Systems with Applications*, 39(12), 10544–10551. https://doi.org/10.1016/j.eswa.2012.02.183

Massari, G. F., Nacchiero, R. and Giannoccaro, I. (2025). Transformative supply chains: The enabling role of digital technologies. *International Journal of Production Economics*, 283. https://doi.org/10.1016/j.ijpe.2025.109562

Matthews, S. G., Gongora, M. A., Hopgood, A. A. and Ahmadi, S. (2013). Web usage mining with evolutionary extraction of temporal fuzzy association rules. *Knowledge-Based Systems*, 54, 66–72. https://doi.org/10.1016/j.knosys.2013.09.003

Reinhardt, R., Gurtner, S. and Griffin, A. (2018). Towards an adaptive framework of low-end innovation capability – A systematic review and multiple case study analysis. *Long*

Range Planning, 51(5), 770–796. https://doi.org/10.1016/j.lrp.2018.01.004

Sezer, P., Basar, D. and Ozturk, S. (2026). Financial inclusion and household debt: Evidence from low-income and lower-middle-income countries. *Borsa Istanbul Review*, 26(1). https://doi.org/10.1016/j.bir.2025.10.029

Telikani, A., Gandomi, A. H. and Shahbahrami, A. (2020). A survey of evolutionary computation for association rule mining. *Information Sciences*, 524, 318–352. https://doi.org/10.1016/j.ins.2020.02.073

Zandi, P., SoleimanEkhtiari, N. and Sorooshian, S. (2025). From taste to price: Comparative analysis of customer satisfaction factors in three food products. *Journal of Retailing and Consumer Services*, 87. https://doi.org/10.1016/j.jretconser.2025.104378