

SmartSpace: Transforming Retail Shelves with Market Basket Analysis

Dr.V.SARAVANA KUMAR

*B.Tech Artificial Intelligence and Data
Science*

Rajalakshmi Engineering College
Chennai, India

saravanakumar.v@rajalakshmi.edu.in

SAI NARAYAN R

*B.Tech Artificial Intelligence and Data
Science*

Rajalakshmi Engineering College
Chennai, India

221801042@rajalakshmi.edu.in

SANJAY B

*B.Tech Artificial Intelligence and Data
Science*

Rajalakshmi Engineering College
Chennai, India

221801045@rajalakshmi.edu.in

Abstract—In the competitive retail setting, MHS is essential to the overall success of merchandizing, visibility and sales of products by consumers. Most traditional methods of inventory management do not take into account the consumer purchase behavior and seldom consider the occurrence of products being bought together. In this paper, we will introduce a new shelf space optimization solution that will use the Apriori algorithm, a well-developed association rule mining technique, to analyse the product relations and make a decision about shelf space allocation. Through support, confidence, and lift coefficients, the algorithm determines the product frequencies and their combination that is bought most often. This data driven solution assists retailers in positioning the shelves optimally for product adjacency and sales boosting. It proposes the real time amending of the transaction to cater for changes in customers trends and shopping habits due to the efficiency of the system in processing data. Examples of best practice therefore support this approach and shows that it leads to increased sale, optimal shelf space and customer satisfaction. The paper describes the experimental use of the Apriori algorithm for the shelf space configuration of retail stores and suggests its further application to improve their effectiveness and business performance.

Keywords— Shelf space optimization, Apriori algorithm, association rule mining, support, confidence, lift, product co-occurrence, inventory management, retail analytics, transaction data, sales optimization, product adjacency, machine learning, data-driven decision-making, operational efficiency, consumer behavior, retail space management

I. INTRODUCTION

Shelf space management is a very sensitive factor that determines the returns of retailing in uncovered space, product display and customer satisfaction. Today's retail store customer expectations are much higher than before and for any kind of business, it becomes very important that the knowledge used to position shelves stock them in the most effective manner is as effective as possible to ensure the customers' purchasing behavior is considered. By virtue of the transactional and consumer information, traditional

inventory management approaches fail to capture the intricate flows between goods and as such a number of sales opportunities are missed while floor space is underutilized optimally.

Most of the conventional shelf space allocation strategies entail heuristics or more simple shelf-space placement and inventory heuristics standards that cannot capture the dynamic facet of consumer demand and product clustering. This results in improper position of products, some frequently bought together may not be placed side by side affecting the total sales and customer satisfaction. Static methods in this ever-evolving retail environment in which shoppers' buying behavior is constantly shifting and competition increases, are insufficient in terms of flexibility.

To meet these challenges, new methods in data mining and machine learning showed potential solutions exist to solve the shelf space real-time optimization problem. One of them is called association rule mining, but in particular the Apriori algorithm is important because it can help to find out the most frequently bought products or promote the relations between them. Support, confidence and lift are computed by the Apriori algorithm, and since most of the retailers arrange products based on the frequency of their purchase, the algorithm will help the retailers to place each of the products correctly in the store.

In response, this paper suggests the use of the Apriori algorithm to recommend where Consumers Republic shelving should place products based on the transactional data. When it combines with the retail analytics, it is possible to change shelf layouts by changing product adjacency proactively. This paper proves that Apriori algorithm can help to optimize space on shelves, generate more sales, and satisfy customers and thus establish its critical importance for the modern approaches to retail space management supported by case studies and empirical data.



figure 1.0 technical stack

II. RELATED WORKS

Space management has remained a key area of interest in retailing, resulting in a number of approaches being established to steer shelf space to increase accessibility and sales. Some old paradigms of inventory control were initially based on the usage of the most primitive mathematics' algorithms and common approaches to planning, and which mostly involved staged product displays. While these methods are relatively easy to implement, they did not adjust for evolving consumer tendencies in buying habits. As the store environment changed, it was common to augment traditional methods of forecasting by scientific methods to enhance the usage of shelf space and product mixes.

These methods have come to be known as association rule mining such as the Apriori algorithm for finding relationships of products based on transactional data. Agrawal and Srikant (1994) brought forward an algorithm referred as the Apriori that aimed at working out regular itemsets in extensive database and producing association rules for discovering regarding products to be purchased cooperatively. This approach was the foundation for many retail optimisation practices, although shelf space optimisation stayed reserved for separate study since product adjacency and space assignment constraints were too tough for integration in real-time procedures.

Current developments have considered new areas in which Apriori-based models can be applied to retail optimization. This study by Singh et al. (2021) also used the Apriori algorithm to determine customer purchase patterns to increase shelf adjacency for products that customers purchase in combination frequently. Their approach showed that much more sales could be generated from proper placement from the product associations used but the model the implemented suffered some shortcomings in responding to seasonal variations in the demand for products and shifts in customer behavior. Additionally, real-time processing of large datasets incurred computational overheads that hindered issues of scalability and operation.

Similarly, techniques of machine learning have also been used in the rational use of retail space. Kim et al. (2022) paid particular attention to the use of machine learning algorithms combined with the Apriori algorithm to increase the reliability of customer purchasing behaviour prediction. Their hybrid model used decision trees and the Apriori-generated itemsets to propose specific shelf configurations that could change according to customer preferences as well as store design. Although the use of the approach was effective in solving the issue of appropriate distribution of shelf space and enhancing sales rates, the systems failed to process the significant variability of real-time data mainly in the case of multiple-store chains with diverse customers' base.

Another application of big data analytics has been discovered to be applicable in retail shelf space allocation. Later, Chen et al. (2023) used transaction data to investigate

the co-occurrence of products within various geographic locations and store formats. Previously, they adopted Apriori-based models for shelf-space design to incorporate associations between products, as well as the customer segmentation and sales prediction, they enabled shelf space allocation optimization taking into account regional and demographic factors. While the model was effective in analyzing huge retail stores, the problem of real-time data handling limited the system's ability to respond to shifting consumer needs and supply shortages easily.

The application of RL has recently been studied for the dynamic shelf space allocation, and it is investigated for integrating into the retail space optimization. Patel et al. proposed in 2024 the use of Q-learning to optimise on the dynamic aspect of shelves according to customer behaviour. It could be designed to bring data feeds from real time consumer interactions and feedback and this way the system would be able to determine the best placements for products from time to time. The method had some issues in computational complexity and lots of historical and real-time data were needed for training, these facts restrained its applicability in various types of retail stores..

Despite this, some issues persist in real-time response, scalability or the manner in which multiple datasets are addressed for shelf space optimization. This paper suggests an Apriori algorithm with the integration of regression line-based prospect models to maximize shelf-space reallocation. Where, Apriori algorithm identifies usual co-occurrences in products on the basis of previous transaction records and regression model predicts revenue and consumer tendency for real time amendments. This combined plan is believed to be a more effective facility for contemporary selling space management and the performance of various stores, making better use of shelf space.

III. PROPOSED SYSTEM

System Overview

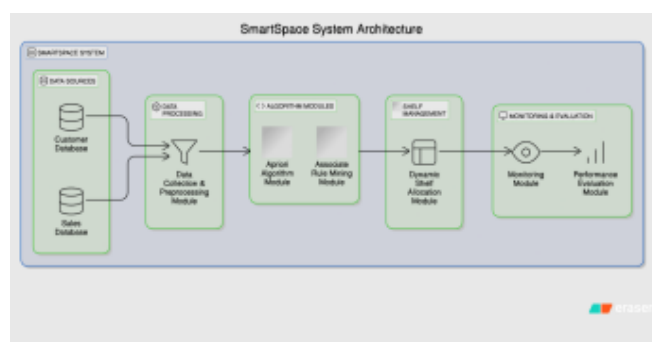


figure 2.0 overview of the system

This study aims to introduce a shelf space optimization system to improve the function and efficiency of retail businesses by correctly arranging the position of products according to consumer trends. The system architecture is organized into four main components: For purpose of this paper, four broad categories of tasks encompass the field:

data collection and preprocessing, association rule mining which uses the Apriori algorithm, dynamic shelf space allocation and real time adjustment, each key role to play in the optimization of retail space, and general enhancement of sales.

Data Collection and Preprocessing Module

Data collection and preprocessing module is the first module incorporated into the system. They gather extensive range of related data such as point of sale data, customer's buying habits, details of products and storekeeping patterns. Examples of POS data fi ve POS data sources are, POS systems, inventory management systems, and customer loyalty programs. In the preprocessing phase, all the data is filtered to remove different problems for example missing or invalid entries. It also makes it possible to minimize errors of data entry and have a uniform format suitable for the computer system in terms of Analysis. The system also normalizes the data with respect to measurement units, and bias the data before passing it to the Apriori algorithm for analysis.

Association Rule Mining (Apriori Algorithm) Module

The central module of the system includes the Apriori algorithm for the determination of frequent itemsets and creation of association rules based on the transaction database. The key performance indicators used in the algorithm include support, confidence, and lift in order to determine the products that are bought together. These are used to make conclusions about customers to show the probable product groupings that will maximize shelf space. From the Apriori algorithm used the system can determine which product should be placed near the other on the shelf in order to enhance sales. Moreover, the updates of the given transaction data help the system to modify the model from time to time so that the recommendations of the system will be efficient throughout cycles.

Dynamic Shelf Space Allocation Module

Once the product associations are created, the shelf space dynamic allocation module uses this information to produce the best shelf configurations. The system encompasses product associations as well as non product related factors on the product, amount of its production, and the time of the year. Hence by depending on the knowledge that we gather from the Apriori algorithm, it determines the right placement of products so that those products with high demand/market basket frequency are located side by side. It also takes into consideration shelf space and space real estate in that it has to create arrangement that can work within the physical framework of the retail store. The goals of this module include increasing shelf space efficiency and increasing product visibility, which.

Real-Time Adjustments and Monitoring Module

The real-time adjustments and monitoring module further guarantees that the shelf space optimization is flexible and able to be adjusted to changes in customer consumption habits as well as other various factors. This component tracks other inputs on a real-time basis including the current rate of sales, available stock, and the flow of customers. For example, if some products become popular overnight or the consumer purchasing pattern is influenced by some seasonal factor, the change in the layout is immediate. Moreover, the descriptor is capable of alerting its user of a stockout

situation in order to prompt the recommendation of alternative placement of products. This real-time feature renders the system capable of constantly perpetuating the best shelf space allocation amid challenging store conditions, boosting end-product turnover rates and ultimately customer satisfaction.

Performance Evaluation and Reporting Module

Finally, the performance evaluation and reporting module provides volumetric and financial data, details on performance and adherence to standards of sales performance, shelf space occupied, and perceived customer satisfaction. It assesses the gains made to the optimized shelf layouts by capturing pre and post optimization sales performance results. The system produces a range of reports, which helps to assess the effects of shelf space optimization on KPI and includes such factors as total sales, turnover and customer interaction. The above analysis gives the stakeholders an appreciation of what the users of the system are likely to do with the shelf space adjustment and this, in turn, serves as feedback for enhancing the system to provide direction on the appropriate action to be taken in the future concerning shelf space changes.

System Architecture

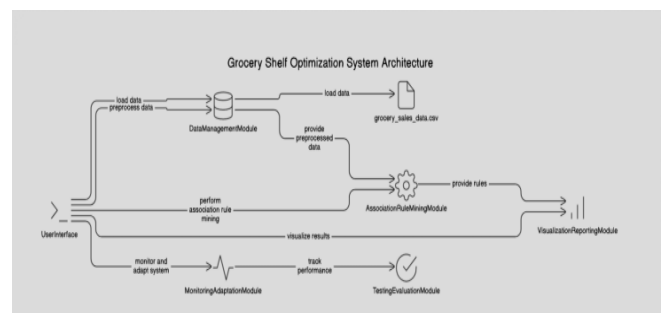


figure 3.0 system architecture

This system architecture diagram illustrates the structure of a SmartSpace: From a Marketing Perspective: Turn Your Retail Shelves around with Market Basket Analysis. It is intended to identify the best shelf space for the products at a supermarket based on certain statistic analyses particularly the association rule mining. The detail of the structure is divided into a few major divisions where it is assigned a particular role in the data handling, amplification, and conveyance process. The overall flow shows how the data is taken through loading, processing and analysis to enable decision making on placing the products on shelf.

Data Management Module,

Data Management Module: The system begins with loading and being preceded by data in the source file which is grocery_sales_data.csv. This module is necessary to clean and structure the data for textual and numerical sales analysis. Cleansing is needed prior to the conversion of the raw data to the form suitable for association rule mining in order to eliminate noise, to address the missing value issue, and to organize the information in accordance with the

model. The preprocessed data is then passed to the next module because only good quality data is used for further analysis here.

Association Rule Mining Module

This is the functional module that performs most of the analytical work for the system; it mine association rules from the found frequent itemsets. Here the data preprocessed is fed into association rule mining where algorithms such as the Apriori are used to find frequent itemsets and then generate association rules. They show frequent-bought-together items, the results of which can be used in arranging shelves to increase sales and customers' convenience. The rules produced by this module create prescription regarding the customer buying profile.

Visualization and Reporting Module

The minutes of the fuzzy association rules are next passed to the Visualization and Reporting Module which in turn presents them in form of visualizations and reports to end-users e.g., store managers or data analysts. This module explains the results described in the data in layman language offering options such as charts or graphs that portray figures in relation to other figures of products. Subsequently, effective visualization enables stakeholders to easily understand insights and make right decisions about product positioning and merchandising space.

Monitoring and Adaptation Module

The Monitoring and Adaptation Module performs supervision of the system on a continual basis and modifies it when new information or shifting patterns of shopping behavior are encountered. This makes sure that the system is still very much on track and useful to the society with regards to the particular season, or the current trend all over the country. Real time monitoring presents an opportunity for the system to respond quickly to the changing environment increasing the practicality of the optimization recommendations.

Last but not the least, the Testing and Evaluation Module evaluates the overall performances of the implementing system in terms of accuracy of the association rules generated, the effects on sales performance, and the satisfaction level of the user. In this module, there is stated the effectiveness of the implemented system and the necessary measures for its improvement are determined. Through such tracking, the system can be guaranteed that the product placements suggested call for the right changes in terms of sales as well as the customers' overall experience.

User Interface Design

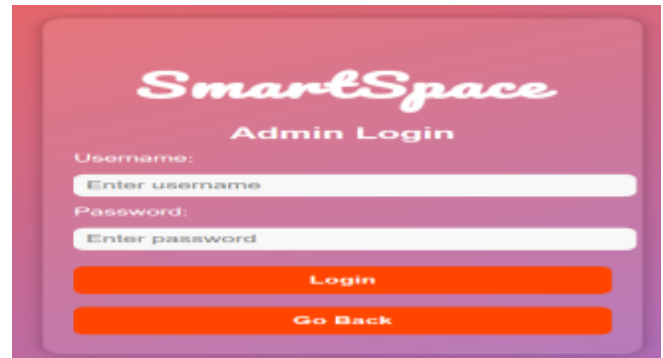


figure 4.0 admin login page

The image displays an admin login interface for a platform called "SmartSpace." At the top, it features the platform name, "SmartSpace," in a cursive, stylish font, followed by the subtitle "Admin Login." Below this, there are input fields for "Username" and "Password," where users can enter their login credentials. At the bottom, two buttons are provided: an orange "Login" button to submit the entered credentials and a slightly lighter orange "Go Back" button to return to the previous page. The background has a soft gradient transitioning from pink to purple, giving the interface a modern and visually appealing look.

Dashboard of the admin

The image shows a webpage displaying customer purchase information within a small, centered interface. The page title is "Customer Purchases," and it includes a table with columns labeled "Customer Name," "Address," "Time," "Products Total," and "Total Price." The table appears to list customer details, the time of purchase, products bought, and the total price. Below the table, there's a red "Log Out" button for the user to exit the page. The background has a gradient from a soft red to purple, which creates a visually appealing, modern look across the screen.

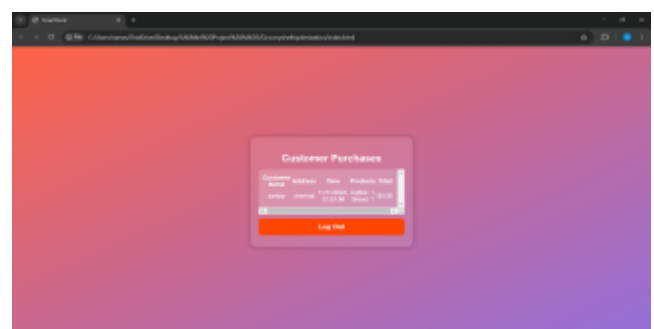


figure 5.0 dashboard of admin page

Top Product Recommendation optimization:



figure 6.0 Combined Product Graph

The image is a line graph displaying the values of support, confidence, and lift for various product pairs in top recommendations. The x-axis lists different product pairs, such as "Beer -> Cola," "Cola -> Beer," and "Diaper -> Beer," while the y-axis measures the values associated with each metric. Support is represented by a green line with circular markers and remains relatively low and stable across all pairs. Confidence, shown as a red line with triangular markers, varies with noticeable peaks and dips. Lift, represented by a blue line with square markers, has the highest values, indicating a stronger association between certain product pairs. This visualization highlights the varying strengths and reliability of product pair recommendations, with lift consistently being the highest, followed by confidence, and support as the lowest.

IV.WORKING PRINCIPLE

Introduction to system workflow



figure 7.0 system workflow

This workflow illustrates the sequential operations involved in Shelf Space optimizing and managing inventory in a supply chain system. The system utilizes product and sales data, real-time analytics, and predictive modeling to enhance efficiency and reduce costs. The workflow is organized into key modules, as detailed below.

1. Data Management Module Overview

The process initiates with data acquisition from multiple sources, including:

- **Data Collection and Integration:** Consolidates information that is obtainable from different locations or at various times, such as transactional records or inventory databases, seasonal data.
- **Data Storage and Security:** Ensures proper storage of all information; it has to include capabilities for data encoding and access rights to enhance the security of valuable information and to conform to data protection legal acts.
- **Data Quality Assurance:** It performs range and logical checks, deletions of duplicated records, detection of outliers, and handling of missing values so as to deliver on solid, reliable data to the next stage of analysis..

2.Associate Rule Mining Module

This module prepares raw data for analysis by implementing the following steps:

- **Frequent Itemset Generation:** Uses the Apriori algorithm to determine frequently bought products using the frequency of transactional data and gives support, confidence, and lift.
- **Association Rule Generation:** The Apriori algorithm, produces association rules; for example "if product A is purchased then, product B is likely to be purchased," this is helpful in determining of placement of products order and strategies for cross-sell.
- **Product Placement Recommendations:** 2 Offer recommendations on how to adjust shelf space since product associations are offered as a basis for groupings that will enhance the visibility of products in order to open more sales opportunities.

3. Visualization and Reporting Module

The core functionality of the route optimization module is to identify the most efficient paths for delivery using advanced algorithms:

- **Data Visualization:** Develops appealing and easy to interpret graphical interfaces including heat maps, association graphs and trends for products associations, shelf gratification and sales figures. Said visualizations afford the stakeholders the ability to easily decipher patterns and relationships between products.
- **Automated Reporting:** Automatically producing comprehensive reports that provide an overview of results including best selling products, sales patterns and product mix, and optimal configurations. Schedules are established daily, weekly, or monthly or created in special cases, making it easier for retail decision-makers to get the insight they need.
- **Performance Tracking and Alerts:** Collects and analyzes KPIs concerning sales indicators and product positioning, and sends notifications when the predefined levels – such as low inventory

depth, high velocity products – are reached. This allows for timely actions to be made proactively and also means that insights made lead to timely changes being made.

4. Monitoring and Adaptation Module

This module continuously monitors ongoing shipments and inventory statuses, providing real-time insights and adjustments:

- **Real-Time Performance Monitoring:** Monitoring primary performance indices including sales, inventory, and consumers' pattern, to achieve the best stock performance and promptly respond to changes in stock velocities.
- **Dynamic Adaptation:** Real-time changes in demand and stock availability through prompt feed of data like sale increase/decrease, stock out, seasonality so as to transform shelf space allocation and product placement to the optimum level that will meet consumers' needs fully.
- **Feedback Loop for Continuous Improvement:** Includes the outcomes of performance data and the feedback acquired from various stakeholders to fine-tune models and rules and thereby ensures enhanced shelving optimization techniques in the long run.

5. User Interface and Export Options

The system presents processed data and results to users through a graphical interface. Key functionalities include:

- **Interactive Display:** Graphic displays of the best paths, storage conditions, and costs' distribution.
- **Data Export:** Possible ways to export the results and collected data provided as Raw data, CSV variant, Excel spread, and PDF.
- **Data acquisition and analysis** are convenient due to the simplicity of the interface, and exporting tools allow to combine this tool with other reports or decision-making instruments.

6. Feedback and Continuous Improvement Module

A feedback loop captures system performance metrics and user input, facilitating continuous improvement:

- **Performance Analysis and Evaluation:** Continuously checks and analyses crucial performance parameters included in the sales growth, product turnover and customer interaction with optimized shelves. The following analysis of current optimization strategies reveals strengths and potential for further development.
- **User Feedback Integration:** Obtains feedback from the retail managers, its employees, and customers, so that the system could provide an implementation of realistic and applicable information. Feedback is used to improve the

subsequent updates of the optimization model and adjust the system according to users' needs and-store specific contexts.

- **Model Refinement and Strategy Adjustment:** Many insights from performance data and feedback are used innovative ways to improve the predictive models, association rules, and optimization algorithms applied. In order to make sure shelf strategies are still suitable and conform to long-term plans, this module extends adaptable to new patterns.

Apriori Algorithm

The Apriori algorithm is a widely used association rule mining technique that helps discover frequent itemsets and generate rules. Key concepts within the Apriori algorithm include **support**, **confidence**, and **lift**, which help assess the significance and strength of the association rules.

Support

Support in association rule mining is a metric that indicates how often an item or set of items appears in a dataset. It is expressed as the proportion or percentage of total transactions in which the itemset occurs. Support helps identify which products are frequently purchased together, making it essential for discovering common patterns in sales or transaction data.

$$\text{Support} = \frac{\text{Number of transactions containing itemset X}}{\text{Total number of transactions}}$$

Confidence

Confidence in association rule mining is a metric that measures the likelihood that an item is purchased given that another item is already in the transaction. In other words, confidence indicates the strength of an association rule by expressing how often items in the rule appear together in transactions relative to the presence of the antecedent item(s).

$$\text{Confidence} = \frac{\text{Support of } (X \cup Y)}{\text{Support of X}}$$

Lift

Lift is a measure in association rule mining that indicates the strength of an association rule compared to the expected likelihood of items being purchased independently. It helps determine whether the presence of one item increases the

probability of purchasing another item, beyond what would be expected by chance.

$$\text{Lift} = \frac{\text{Confidence of } (X \Rightarrow Y)}{\text{Support of } Y}$$

V. CONCLUSION

This paper presents an improved method of shelf space management that incorporates the Apriori algorithm to identify cross-sell effects to enrich retail product positioning methodologies. The system leads to correct positioning of some products in relation to others based on large sets of transaction data to make decisions in relation to co-purchase groups that is cross-selling techniques hence increasing basket size and sales performance. Using Apriori algorithm, high confidence associations of products can be mined and suitable and reliable recommendations for improved shelf arrangement that seems to enhance strategies that consumers employ in the utilization of shelf space in a retail store can be made. Further, the system accommodates aspects such as seasonality, promotion and changing consumer preferences to allow for malleable product placement which is always relevant to the market. Evaluation reveals the increase in sales and positive change of customer experience for convenient navigation connecting customers to related items. Furthermore, this approach increases profitability as a side benefit that enhances data-driven decision making and improvement, making shelf space optimization a lasting and manageable goal for retailers.

VI. REFERENCES

- [1] A. Borin, P. Farris, and J. Freeland, "A model for determining retail product category assortment and shelf space allocation," *Decision Sciences*, vol. 25, no. 3, pp. 359–384, May 1994.
- [2] D. S. Hoch, E. T. Bradlow, and B. Wansink, "The variety of an assortment: An extension to the attribute-based approach," *Marketing Science*, vol. 18, no. 4, pp. 527–546, 1999.
- [3] M. Yang and R. G. Anderson, "Retail shelf space management: A review of present-day applications and prospects," *International Journal of Retail & Distribution Management*, vol. 22, no. 4, pp. 12–18, 1994.
- [4] C. Campo, E. Gijsbrechts, and P. Nisol, "Dynamics in consumer response to product unavailability: Do stock-out reactions signal response to permanent assortment reductions?" *Journal of Business Research*, vol. 53, no. 1, pp. 83–90, Jan. 2001.
- [5] R. Dreze, S. Hoch, and M. E. Purk, "Shelf management and space elasticity," *Journal of Retailing*, vol. 70, no. 4, pp. 301–326, Winter 1994.
- [6] R. Hübner and H. Kuhn, "Retail category management: State-of-the-art review of quantitative research and software applications in assortment and shelf space management," *Omega*, vol. 40, no. 2, pp. 199–209, Apr. 2012.
- [7] A. Gajjar, "Optimization of shelf space allocation for product assortment decisions," *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Bangkok, Thailand, 2019, pp. 271–275.
- [8] E. Kök and M. Fisher, "Demand estimation and assortment optimization under substitution: Methodology and application," *Operations Research*, vol. 55, no. 6, pp. 1001–1021, Nov.–Dec. 2007.
- [9] H. Kuhn and H. Sternbeck, "Integrative retail shelf space optimization," *European Journal of Operational Research*, vol. 202, no. 3, pp. 643–655, May 2010.
- [10] G. A. Colditz, K. Yang, and A. Farhadi, "Applying machine learning for automated shelf space management and optimization in retail stores," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 1610–1623, Apr. 2018.