International Journal of Interpreting Enigma Engineers (IJIEE)

https://ejournal.svgacademy.org/index.php/ijiee/

OPEN ACCESS

STUDY AND ANALYSIS OF MARKET BASKET ANALYSIS USING APIRORI ALGORITHM

Gorak Tanusha

Original Article

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women

*Corresponding Author's Email: tanushaa15@gmail.com

Abstract

One data mining technique that has been increasingly popular across several industries, especially in retail and e-commerce, is market basket analysis, or MBA. This technique examines transactional data to find patterns and relationships between the products that are frequently purchased. An MBA is essential for understanding consumer behavior, which helps companies place products more effectively, develop more effective marketing campaigns, and run their operations more efficiently overall. The MBA is still a vital tool for organizations to stay abreast of changing customer preferences, stimulate rapid decision-making, and extract relevant information from the massive amounts of transactional data they are dealing with. MBA is used in various industries outside of retail, including supply chain management, healthcare, and internet platforms.

Keywords: Support, Confidence, Lift, Association, Apirori Algorithm.

Introduction

A data mining approach called market basket analysis (MBA) has revolutionized how businesses understand and profit from customer purchasing patterns. The constantly expanding retail and e-commerce industries generate enormous volumes of transactional data. MBA offers a powerful way to uncover these hidden patterns and linkages. The fundamental idea of an MBA is to analyze customer interactions to identify product sets that are frequently purchased together. Businesses are known for these connections and make informed judgments that can enhance customer experiences, maximize product placements, and advance strategic marketing initiatives.

One of the most crucial aspects of the MBA is data mining, which is essential for drawing insightful conclusions from complicated information. Data mining algorithms like the Apriori algorithm have been used to effectively examine transactional data to identify recurring itemsets or groupings of related products that are frequently purchased together. Businesses may sort through large datasets using this algorithmic approach, which reveals trends that could be difficult to identify with more conventional techniques [1].

From the study, we address how data mining approaches help create association rules that characterize the connections between various entities. These principles give businesses a better grasp of the strength and importance of the detected correlations by providing quantifiable measurements like lift, confidence, and support [3]. Retailers may make well-informed decisions based on evidence derived from data when they use lift to assess the strength of correlations, confidence to quantify the reliability of rules, and support to estimate the frequency of itemsets.

Account to companies looking to glean insights from transactional data now find the combination of Market Basket Analysis and data mining essential. The importance of data mining in MBA programs grows with the amount of data available. Which allows firms to make strategic decisions, improve operational efficiency, and adjust to changing consumer habits.

Literature Survey

Kutuzova Tatianaa et al. 2018. Describes how information systems that analyze a lot of data have become essential to our daily lives. Retailers now need to examine the behavior of their customers due to the development of online markets and market technologies where analysis can lead to improvements in supplier profitability, service quality, and customer happiness, which will raise interest in further study, feasible to improve the recommendation system's quality by analyzing more data gathered from external, diverse sources [1].

Manpreet Kaura et al. 2016. Explained that Giving merchants the information they need to understand consumer purchasing behavior and make wise selections is the main objective in marketing. Different algorithms are available to execute tasks. The current algorithm only processes static data and cannot change the data over time. However, this technique offers a fresh method for accounting for data changes mining static data [2].

M. Qisman et al. 2021. Stores that sell goods should take advantage of all available data resources, including data. The anticipated is the processing of data will be able to deliver insights that will help marketing plans. Techniques of Data mining, such as Market Basket Analysis utilizing an apriori algorithm, are the data processing approaches that use marketing strategies. Designed using the waterfall method, which begins with a user needs analysis that ends with the UML to construct a process that includes activity, sequence, and use case diagrams. PHP was the development of the Market Basket Analysis application [3].

Sajid Mahmood et al. 2014. Illustrated that, Association rule mining algorithms often aim to extract frequent features (item sets) or features in a transactional database. However, these algorithms overlook large significant item sets that have poor support or occur infrequently. In contrast to the common-data items, these rare item sets can yield Significant negative association rules (NARs) with high confidence that are not visible. Therefore, create a trustworthy decision support system. It is crucial to identify probable negative association rules. [4].

Yanyan Zhang et al. 2021. From the association analysis carried out on a sizable volume of data gathered from real-world applications. Data mining creates a knowledge-point association rule table, collects and pre-processes data, and a college English diagnostic practice system. The knowledge-point association rule table was presented by beginning with the current model of the system. The research proposes a diagnostic evaluation methodology for college English that can dynamically diagnose learning problems in learners and accurately assess their current level of learning. Throughout the comparative tests, it's then determined that the diagnostic evaluation methodology used in this research may successfully enhance learners' learning status and knowledge point problems by offering improved practice assistance and test question recommendations [7].

Fulin Li et al. 2023. It shows that data mining technology is a technique for concluding vast amounts of data. This research proposes a broad association rule-based method for equipment quality information mining for complicated equipment drawbacks of traditional association rule mining methods and their lengthy execution times and large-memory use. An enhanced Apriori technique is suitable, and the candidate itemset creation procedure is optimized. Time Complexity and spatial complexity are used as assessment criteria to test the algorithms' performance on five experimental data sets. Comparative tests demonstrate the benefits of the upgraded method. To then further integrate the information and data processing [8].

Mustakim1 et al 2018. Elobarated with an explanation of Berkah Mart, one of the newest minimarkets in Pekanbaru City, and how it started as a retail enterprise to serve the requirements of the neighborhood. Berkah Mart continues to operate under the conventional store management belief that product arrangement on sale shelves determined by sales analysis. The marketing economics idea offers a range of consumer and product management strategies, such as different itemset search algorithms and market bucket analysis methodologies, for company management in Berkah Mart. While existing products evaluated from the perspective of the consumer. This study suggests using the FP-Growth algorithm in conjunction with Market Basket Analysis to plan and organize the availability of commodities [9].

Alfiqra1 et al. 2020. Association Rule is a technique used to create a marketing plan based on consumer purchasing patterns. Association rules the process of identifying relationships or associations between products that are part of a single transaction. There are situations where the Association Rule Market Basket Analysis rule is insufficient to give an analysis due to a significant degree of customer buying pattern variability. The total Variability of Association Rule is an indicator that concentrates on market basket analysis and assumes that customer behavior while making product purchases is very variable [10].

Yusuf Kurnia1 et al. 2019. To survive in a highly competitive market, owners of food and beverage businesses, particularly restaurants, must make the right decisions. They must constantly innovate to meet customer needs by creating products that draw in customers and devising marketing plans that increase sales. O! Fish restaurants can leverage data regarding sales trends to develop highly inventive marketing tactics to increase sales by highlighting products [11].

Angela Hsiang-Ling Chen et al. 2023, It shows trans hows actions of retail have become a crucial component of the economic lifestyle in each nation globally. The trade sector of retail transactions has the potential to grow steadily in the future. The ultimate goal of this study was to produce a targeted recommendation system based on consumer purchase and product-selling patterns. Two variables—period and customer engagement index—a clustering algorithm, like K-means, KNN, and Ward's method, Association rule to identify the pattern for the effect of the relationship on each transaction, and four different types of classifiers to use, validate recommendation systems added to modified RFM analysis. The findings demonstrated that in light of consumer behavior [12].

Rakesh Agrawal et al. To tackle this problem, they introduce two novel algorithms that differ significantly from the existing approaches. These algorithms surpass existing methods that range from minor issues to more than an order of magnitude for big problems and experiments conducted with synthetic and real-life data. Additionally, we demonstrate how Apriori Hybrid, a hybrid algorithm, can be created by combining the best elements of the two suggested methods. Apriori Hybrid scales with the number of transactions to the scale-up trials. Excellent scale-up capabilities are also included in Apriori Hybrid about transaction size and item count [13].

Abdulsalam, S.O. et al. 2014. The Apriori algorithm is used in association rules mining and referred as market basket analysis. To extract useful information from a supermarket's database, Data created from a well-structured transactional database that mirrored the sales pattern of a grocery store and then represented six different products over thirty single transactions. For data, the frequency of products and several association rules derived. These guidelines proved that buying one product will inevitably result in buying another, as seen by the correlation between chocolate and apples. The relationship found is that it will help businesses create advertising and marketing campaigns that will set them apart from the competition [14].

N. Isa1 et al. 2018. That is typically involved in retail, particularly transaction mining or basket data mining. It demonstrated that this method works well for figuring out the preferences and purchasing habits of the customers. Trends in business today have undergone a significant shift in tandem with technological advancements. A rise in precision in corporate processes is necessary to adapt to changes in client demand. This essay suggests applying an MBA to a small-

to medium-sized enterprise, using Corm Café as a case study. Utilized has been daily transaction data extracted from customer order sheets [15].

Pathan Afrin F et al. 2019. One data mining technique for determining the relationship between an item and another is association rules. To identify the available frequent data appearance on the item set, use specific algorithms, such as the Apriori algorithm, in which value of each datum created to develop new knowledge. This study compared market basket analysis using the apriori algorithm with market basket analysis not using an algorithm for rule creation to produce information [16].

Xiaohua Li et al. 2022. As artificial intelligence and big data technology spread throughout the medical field, the issues with traditional pharmacy management such as its high cost and inefficiency—became increasingly apparent. Therefore, the purpose of the work was to propose the design and development of the intelligent pharmacy's dispensing process and equipment using data mining technology. First, overview the state of association rule and data mining techniques and discuss their potential applications in related fields. Second, the Intelligent pharmacy dispensing the data standards and integration platform created. The interactive interface was developed using web service technology and connected to the pharmacy's Intelligent gadget. Ultimately, an association rule mining-based Intelligent pharmacy management system was constructed [17].

Proposed Methodology

A. Market Basket Analysis:
A data mining approach called market basket analysis (MBA) is used in e-commerce and retail to examine consumer purchasing patterns. The main objective is to find patterns and correlations among the products customers frequently purchase together. Understanding consumer preferences, maximizing product positioning, and making wise business decisions are all aided by this approach.

As an illustration

The retailer may choose to arrange Coffee and Sugar near to one another, provide promotions on these items together, or adjust their inventory in light of the observed relationships if Market Basket Analysis finds a rule along the lines of "If a customer buys Coffee, there's an 80% chance they will also buy Sugar."
Generally, Market Basket Analysis assists companies in determining which goods are frequently bought in tandem, enabling them to make well-informed decisions that will boost sales, improve consumer pleasure, and improve the overall shopping experience.

B. Association Rule:
Simple definitions of association rules are associations between elements in a dataset, typically obtained from transactional data. These rules highlight patterns like "if X, then Y," which show relationships or linkages between several objects. Association rules are used in market basket analysis when companies try to know which products are usually purchased together.
Here's a more straightforward summary of the essential elements:

Precursor and Effect:
The antecedent refers to the "if" portion of the rule. It is a representation of the existing item or group of elements.
The item or items most likely to be connected to the antecedent indicated by the "then" portion, which is the consequent.

Assistance:
Support quantifies the frequency with which the rule's items occur together in the dataset. High support denotes a more widespread correlation.

Self-assurance:
The probability that the existence of the antecedent is esstential for the existence of the consequent is measured by confidence. It serves as a gauge for the rule's dependability or robustness.

Elevate:
When the antecedent is present in a transaction, lift compares the likelihood of discovering the consequent for likelihood.

In layman's words:
Assume you examine consumer purchases and discover the following association rule: "There is a 70% chance that a customer who purchases chocolate will also purchase ice cream (antecedent)."

Interpretation :
Support: 20% of all transactions include chocolates and ice cream.
Confidence: 70% of ice cream transactions also involve Chocolate.
Lift: When a customer purchases Ice Cream, their chance of purchasing Chocolate rises by 70% relative to their overall chance of purchasing.

Essentially, association rules give organizations important information about the relationships between items and can be used to guide decisions about client preferences, marketing, and product placements.

C. The main steps of the Apriori algorithm:
1. Starting Point:
Find every single unique item (singleton) in the dataset to begin with.
2. Produce Sets of Candidate 1 Items:
Make an inventory of every single item as a candidate 1-itemset.
3. Examine and Determine Support:
Determine the support (frequency) of each 1-item set by scanning the dataset.
If a prospective 1-itemset's support level falls below a predetermined minimum, discard it.
4. Create candidate sets of k items (k > 1):
To create candidate k-item sets, use the often occurring (k-1)-item sets from the preceding phase.
Check if the union of these item sets is a legitimate candidate k-itemset by combining them with the identical first (k-2) items.
5. Examine and Determine Support for Potential K-Item Sets:
To find the support for each candidate k-item set, scan the dataset.
Throw away candidate k-item sets that don't meet the required level of support.
Till there is no more frequent item, repeat steps 4-5.
Until no more frequently occurring item sets occur, keep creating and scanning candidate item sets of progressively larger size.
6. Make Association Regulations:
Create association rules from the collected frequent item sets by taking every potential pairing of antecedents and consequents.
Determine the confidence of every rule.
7. Rule Selection for Confidence:
Rules that don't satisfy a minimum confidence criterion have been discarded.
The fundamental tenet of the Apriori method is that all of an itemsets subsets must likewise be common if it is. The technique effectively finds frequent item sets and association rules by iteratively finding and eliminating infrequent item sets.

D. Application and analysis of Apriori algorithm:
According to the steps of the Apirori algorithm given above, we now look at steps based on an example. Let's continue with the example, incorporating a minimum support value and calculating support, lift, and confidence values. For simplicity, let's set the minimum support threshold to 40%.

Consider the dataset:

Items	Transaction
1.	Bread, Milk
2.	Bread, Butter
3.	Milk, Butter
4.	Bread, Milk, Egg
5.	Bread, Butter

Let's go through the Apriori algorithm steps:
Step 1: Initialization
Initial candidate 1-itemsets: {Bread}, {Milk}, {Butter}, {Egg}
Step 2: Support Calculation
Count support for each candidate 1-itemset.

Candidate Itemset	Support
{Bread}	4/5 = 80%
{Milk}	3/5 = 60%
{Butter}	3/5 = 60%
{Egg}	1/5 = 20%

Discard items with support below 40%.

Frequent Itemset	Support
{Bread}	80%
{Milk}	60%
{Butter}	60%

Step 3: Generate 2-itemset Candidates
{Bread, Milk}, {Bread, Butter}, {Milk, Butter}
Step 4: Support Calculation
Count support for each 2-itemset candidate.

Candidate Itemset	Support
{Bread, Milk}	2/5=40%
{Bread, Butter}	3/5=60%
{Milk, Butter}	2/5=40%

Discard item sets with support below 40%.

Frequent Itemset	Support
{Bread, Butter}	60%

Step 5: Generate 3-itemset Candidates
{Bread, Butter} is the only frequent 2-itemset, so generate candidates using it: {Bread, Butter, Milk}
Step 6: Support Calculation
Count support for the new 3-itemset candidate.

Candidate Itemset	Support
{Bread, Butter, Milk}	2/5=40%

Discard item sets with support below 40%.

Frequent Itemset	Support
{Bread, Butter}	60%

Step 7: Association Rule Generation

Generate association rules from the frequent item sets.

Association Rule	Support	Confidence	Lift
{Bread=>{Butter}	60%	75%	1.25

The result includes the frequent item sets and association rules that meet the specified minimum support threshold. The support, confidence, and lift values provide insights into the strength and significance of the associations between items. In this example, the rule {Bread} => {Butter} has a confidence of 75% and a lift of 1.25, indicating a positive association.

Conclusion

The Apriori algorithm-enabled Market Basket Analysis (MBA) has emerged as a component of contemporary data-driven decision-making, particularly in retail and E-commerce. Businesses use Apriori-based MBA to obtain deeper insights into client behavior and purchasing patterns in today's environment, where data is available. The repetitive structure of an algorithm provides a detailed understanding of product relationships by efficiently mining transactional data to uncover frequently occurring itemsets and association rules. This knowledge is crucial for streamlining commercial operations, including targeted marketing campaigns, inventory control, and product positioning. Apriori's quantitative measures, such as lift, confidence, and support, let firms evaluate the dependability and strength of the correlation found. An MBA with Apriori enables firms to thrive in today's dynamic market, where consumer preferences are changing to stay agile. Modifying tactics in response to real-time information about shifting consumer behavior and industry trends promotes continuous improvement. The algorithm is well-suited for the vast data, where firms deal with enormous volumes of transactional information due to its capacity to handle large datasets. Successful retail strategies continue to be shaped by an Apriori-based MBA as firms work to deliver seamless and tailored consumer experiences. Its uses go beyond conventional retail to several sectors, such as supply chain management, online platforms, and healthcare, where identifying patterns in data can help make better decisions. In summary, the combination of MBA and the Apriori algorithm offers a strategic advantage for companies negotiating the intricacies of the contemporary marketplace, not merely a tool for transaction analysis.

References

1. Kutuzova Tatianaa et al, Market Basket Analysis of Heterogeneous Data Sources for Recommendation System Improvement,2017.

2. Insta cart Market Basket Analysis, 2017. [Online].available: https://www.kaggle.com/c/instacart-market-basket-analysis.

3. Manpreet Kaura et al, Market Basket Analysis: Apply Association Rule Mining to Determine Changing Trends in Market Data,2016.

4. M Qisman et al, using transaction data, market basket analysis with the Apriori algorithm reveals consumer purchasing patterns (case study of Mizan computer retail outlets),2021.

5. Sajid Mahmood et al, Frequent and Infrequent Item Sets for Mining Negative and Positive Association Rules from Text",2014, Volume 2014, Article ID 973750.

6. [http://csn.cancer.org/], [http://www.carepages.com/forums/cancer].

7. Yanyan Zhang, The English Diagnostic Practice Systems Intelligent Knowledge Point Association Analysis Algorithm Realization,2021, Volume 2021, Article ID 5545866.

8. Fulin Li et al, An Enhanced Apriori algorithm-based Information Mining method for Equipment Quality", Volume 2023, Article ID 2155590.

9. Mustakim et al, The Study conducted a Market Basket Analysis at Berkah Mart in Pekanbaru Riau using Apriori and FP-Growth to analyze consumer Expenditure Patterns

10. Alfiqra et al, Application of Market Basket Analysis on Product Marketing Strategy based on Overall Variability of Association Rule.

11. Yusuf Kurnia et al, Analyze the use of the apriori algorithm in Data mining market basket analysis to determine the sales pattern (item association) at the O! Fish restaurant.

12. Angela Hsiang-Ling Chen et al, Improving Retail Transactions: An Association Rules Mining and Modified RFM Analysis Based Data-Driven Suggestion,2023.

13. Rakesh Agrawal et al, Fast Algorithms for Mining Association Rules.

14. S.O. Abdulsalam et al, A Method for Data Mining in Market Basket Transactions using the Association Rules,2014.
15. N. Isa1 et al, Examination of Market Basket on Consumer Buying Patterns at Corm Cafe.

16. Afrin F Pathan,Snehal Palande et al, A Study on the Market Basket Analysis and the Association Mining.

17. Xiaohua Li et al, Data mining involvement in improving the effectiveness of many intelligent devices allocation in an Intelligent pharmacy.

18. Tungana Bhavya et al, A Study of Machine Learning based Affective Disorders Detection using Multi Class Classification, Intelligent Engineering Applications and Applied Sciences for Sustainability IGI Global, 2023.

19. P. R. Anisha et al, A clever deep feature-based prediction algorithm for metabolic syndrome in sleep disorders, Springer Multimedia Tools and Applications, 2023 https://doi.org/10.1007/s11042-023-17296-4.

20. C. Kishor Kumar Reddy et al, An intelligent optimized cyclone intensity prediction framework using satellite images, Springer Earth Science Informatics, 2023 https://doi.org/10.1007/s12145-023-00983-z.

21. Anisha P R et al, Intelligent Systems and Machine Learning for Industry: Advancements, Challenges and Practices, CRC Press, Taylor & Francis, 2022.