Jurnal Mantik, 7 (3) (2023), ISSN 2685-4236 (Online)

IOCSCIENCE

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Analysis of vegetable purchasing patterns in supermarkets using association rule

Kasmawaru[1], Imran Djafar[2], Muh Rafli R[3], Rachmat[4], Muh Yusuf[5]

[1,2] Engineering, Informatic Engineering, University Dipa Makassar, Makassar, Indonesia

[3] Information System, University Dipa Makassar, Makassar, Indonesia

[4,5] Engineering, Informatic Engineering, University Pejuang Republik Indonesia, Makassar, Indonesia

ARTICLE INFO

Article history:

Received Nov 02, 2023
Revised Nov 18 , 2023
Accepted Nov 20, 2023

Keywords:

Association rule mining;
Association analysis;
Supermarket;
Support with Confidence.

ABSTRACT

One data mining technique for identifying associative rules between a set of elements is association analysis, often known as association rule mining. Understanding the likelihood that a consumer will purchase bread and milk together is an example of an associative rule from examining purchases made in a supermarket. Supermarket operators can use this information to plan their product placement or create marketing campaigns that use discount coupons for specific product pairings. The use of association analysis to examine the contents of supermarket shopping baskets helped make it well-known. Another name for association analysis is market basket analysis. Perform a multiplication of the numerous rules acquired by Support and Confidence, where the latter should be at least 80%.

Corresponding Author:

Kasmawaru,
Computer Information System,
University Dipa Makassar,
Perintis Kemerdekaan Km. 9, Makassar, Sulawesi Selatan, 90241-90245, Indonesia.
Email: kasmawaru@undipa.ac.id

1. INTRODUCTION

Over the past few decades, the landscape of supermarket retailing has undergone significant transformations, influenced by changing consumer preferences and advancements in data analytics (Niemeier et al., 2013; Purohit & Jain, 2020). Supermarkets have evolved from simple point-of-sale establishments to data-rich environments where every transaction is recorded, providing a wealth of information for analysis. This evolution has created opportunities for supermarkets to leverage data-driven approaches to enhance customer satisfaction and operational efficiency (Hokka, 2023; Pereira & Frazzon, 2021).

Within the expansive assortment of products offered by supermarkets, vegetables constitute a significant category due to their essential role in a balanced diet. Understanding the intricacies of consumer behavior concerning vegetable purchases is crucial for supermarkets seeking to tailor their offerings to meet customer demands effectively. By unraveling patterns in vegetable purchasing, supermarkets can optimize their inventory, strategically place products, and develop targeted marketing initiatives. This knowledge not only improves the shopping experience for customers but also has implications for waste reduction, as supermarkets can better manage perishable goods based on demand patterns (Basuki & Cahyani, n.d.; Meliza & Lubis, 2022).

Recent advancements in data analytics and machine learning techniques have empowered retailers to extract meaningful insights from vast datasets. Association rule mining, a technique rooted in data mining, has proven to be particularly effective in uncovering hidden relationships within transactional data. This research builds upon these technological advancements to delve into the specifics of vegetable purchasing patterns, aiming to contribute practical insights that supermarkets can use to refine their business strategies (Bharadiya, 2023; Mishra & Tyagi, 2022; Weber & Schütte, 2019).

In an increasingly competitive retail environment, supermarkets are compelled to stay ahead by continuously adapting to changing consumer behaviors. Research on vegetable purchasing patterns serves as a strategic tool for supermarkets to remain agile, aligning their product offerings with customer expectations. As the industry continues to embrace data-driven decision-making, this study positions itself at the intersection of technology and retail, offering supermarkets actionable intelligence to navigate the complexities of vegetable retailing in the modern marketplace (Heller, 2006; Jiménez-Guerrero et al., 2018).

The modern retail landscape is characterized by a diverse array of vegetables available to consumers in supermarkets. Understanding the purchasing patterns of these vegetables is crucial for supermarkets aiming to enhance customer experience and optimize their operations. This research aims to investigate the associations and co-occurrences of vegetable purchases in supermarkets using association rule mining techniques. By analyzing transactional data, the study seeks to identify patterns where certain vegetables are frequently bought together, providing insights that can inform inventory management, marketing strategies, and the overall organization of the supermarket. This research addresses the need for supermarkets to adapt to consumer preferences, improve product placement, and implement targeted marketing initiatives to meet the evolving demands of their customer base (Bag et al., 2021; Cameron et al., 2016; Kumar, 2008).

Uncovering vegetable purchasing patterns through association rule mining has the potential to revolutionize supermarket operations and customer engagement. The findings of this research can guide supermarkets in optimizing product placement, creating effective bundled promotions, and tailoring marketing campaigns to align with observed consumer behavior. By implementing strategies informed by these insights, supermarkets can enhance the overall shopping experience, increase customer satisfaction, and potentially boost sales (Gomez et al., 2004; Terblanche, 2018). Moreover, the study contributes to the broader field of retail analytics by demonstrating the practical applications of association rule mining in understanding and responding to consumer preferences in the specific context of vegetable purchases.

One kind of association rule used in data mining is the A Priori algorithm(Antonie, n.d.). This group also includes hash-based algorithms and generalized rule induction techniques, aside from a priori(Siekmann et al., n.d.). Market basket analysis and affinity analysis are terms used to describe rules that express the relationship between multiple variables(Abdullah et al., 2011). One data mining technique for identifying associative rules between a set of elements is association analysis, often known as association rule mining(Husain & Hidayati, 2021). Understanding the likelihood that a consumer will purchase bread and milk together is an example of an associative rule from examining purchases made in a supermarket(Sousa et al., 2021). Supermarket operators can use this information to plan their product placement or create marketing campaigns that use discount coupons for specific product pairings(Gufroni, 2018). The use of association analysis to examine the contents of supermarket shopping baskets helped make it well-known(Chen, 2012). Another name for association analysis is market basket analysis. Another name for association analysis is a data mining approach that serves as the foundation for many other data mining techniques(Al-maolegi & Arkok, 2014). Specifically, a phase of association analysis known as high frequency pattern mining has drawn the interest of numerous academics in order to develop effective algorithms. Whether a rule of association(Deshpande & Thakare, 2010).

2. RESEARCH METHOD

a. High frequency pattern analysis

High frequency pattern analysis (frequent pattern analysis) is a technique in data mining that is used to identify patterns that often appear or occur simultaneously in a collection of transaction data or other datasets(Kaur & Madan, 2015). This technique is often used in the context of Association Rule Mining to find relationships between items or attributes in a dataset(Sarno et al., 2015).

High frequency patterns include items or attributes that frequently appear together in the data(Apriori & Kunci, n.d.). For example, if you have data on customer purchases in a store, then high-frequency pattern analysis can help you identify product combinations that customers often buy together(Kasieczka et al., 2021).

Explanation of high frequency pattern analysis:

1. Pattern Identification: This technique will identify patterns of combinations of items or attributes that appear frequently in the data, such as "if customers buy product A, then they are also likely to buy product B."
2. Frequency of Occurrence: These patterns are discovered based on the frequency of their occurrence in the dataset. Patterns that have a frequency of occurrence above a certain threshold are considered high-frequency patterns.
3. Determining Associations: By identifying high frequency patterns, you can determine associations or relationships between items or attributes in the data. For example, you may find that when customers buy carrots, they tend to also buy broccoli.

Business Benefits: High-frequency pattern analysis has various business benefits, such as improving product recommendations, optimizing product placement in stores, and designing more effective marketing strategies(Ramageri, n.d.). It can also help in better decision making in a variety of industries, including retail, e-commerce, healthcare, and many others. In other words, high-frequency pattern analysis helps in extracting insights from data that can be used to improve business strategies, marketing and decision making.

$$Support\ (A) = \frac{Number\ of\ Transactions\ Contains\ A}{Total\ Transactions} \quad (1)$$

while the support value of 2 items is obtained from the following formula:

$$Support\ (A \cap B) = \frac{Number\ of\ Transactions\ Contains\ A\ dan\ B}{Total\ Transactions} \quad (2)$$

b. Formation of associative rules

Associative rule formation, or Association Rule Mining, is a process in data analysis that is used to identify relationships or associations between items or attributes in a dataset(Supriyamenon & Rajarajeswari, 2017). This technique is generally used in the context of transaction data or datasets containing information about customer purchases, but can also be applied in a variety of other domains. The following is an explanation of the formation of associative rules:

1. Transaction Data : The formation of associative rules is usually applied to transaction data, where each transaction is a collection of items or attributes obtained from an event or activity. Examples of transactions may include purchases at a store, ordering food at a restaurant, or selling products in e-commerce.
2. Purpose: The main purpose of establishing associative rules is to identify patterns or associations that exist between items or attributes in the data. For example, we want to know whether there is a relationship between product A and product B in customer purchases.
3. Method: The Association Rule Mining technique uses methods such as the Apriori, FP-Growth, or Eclat algorithms to extract associative rules from data. These algorithms calculate the frequency of occurrence of item combinations and use metrics such as support, confidence, and lift to assess the quality of the rules.
4. Rule Selection: Once the associative rules are generated, you can select rules that have relevant support and confidence values for your analysis purposes. For example, you might only be interested in rules that have confidence above a certain threshold.
5. Business Benefits: Establishment of associative rules can provide useful insights for business decision making. It can be used to improve product recommendations,

optimize product placement in stores, design more effective marketing strategies, and identify customer trends or habits.

After all high frequency patterns are found, then look for associative rules that meet the minimum requirements for confidence by calculating the confidence of associative rule A ➔B. The confidence value of rule A ➔B is obtained from the following formula:

$$Confidence = P(B|A) = \frac{Number\ of\ Transactions\ Contains\ A\ dan\ B}{Number\ of\ Transactions\ Contains\ A} \quad (3)$$

3. Results and Discussion

Thus, associative rule formation is an important tool in data analysis to gain valuable insights from transaction data and help businesses make better decisions in terms of product, sales, and marketing strategies(Liao et al., 2012).

The percentage of transactions that contain all items in the rule. Support measures the extent to which a rule appears in the dataset. The extent to which item B often appears in transactions that already contain item A. Confidence reflects the extent to which the rules can be relied upon. A measure of the relationship between item A and item B that compares the co-occurrence of A and B with the independent occurrence of both. A lift above 1 indicates a positive correlation.

Table 1. Initial Data

Transaction	Item Purchased
1	Brocoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Brocoli
6	Squash, Asparagus, Beans, Tomatoes
7	Tomatoes, Corn
8	Brocoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Brocoli, Beans, Squash
12	Asparagus, Beans, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green, Peppers, Tomatoes, Beans, Broccoli

In this initial step, it can be seen that the items frequently purchased by customers are as follows:

Table 2. Items purchased

Items Purchased
Asparagus
Beans
Broccoli
Corn
Green Peppres
Squash
Tomatoes

Transaction	Asparagus	Beans	Broccoli	Beans	Green Peppers	Squash	Tomatoes

I = {Asparagus, Beans, Broccoli, Corn, Green peppers, Squash, Tomatoes}

Then calculate the number of purchases for each item. So it can be like the tabulation table below.

Table 3. Purchase Hits for each item

Transaction	Item Purchased
1	Brocoli, Green Peppers, Corn
2	Asparagus, Squash, Corn
3	Corn, Tomatoes, Beans, Squash
4	Green Peppers, Corns, Tomatoes, Beans
5	Beans, Asparagus, Brocoli
6	Squash, Asparagus, Beans, Tomatoes

7	Tomatoes, Corn
8	Brocoli, Tomatoes, Green Peppers
9	Squash, Asparagus, Beans
10	Beans, Corn
11	Green Peppers, Brocoli, Beans, Squash
12	Asparagus, Beans, Squash
13	Squash, Corn, Asparagus, Beans
14	Corn, Green, Peppers, Tomatoes, Beans, Broccoli

The next step is to determine the size Φ

Suppose we determine the price Φ = 4 (7/2 =3,5 => 4) then we can determine the frequency of the item set. Until then $_k F$ we can search. for k=1

Table 4. Determine the size Φ

Transaction	Asparagus	Beans	Broccoli	Corns	Green Peppers	Squash	Tomatoes
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0
11	0	1	1	0	1	1	0
12	1	1	0	0	0	1	0
13	1	1	0	1	0	1	0
14	0	1	1	1	1	0	1
Σ	6	10	5	8	5	7	6

From the table above, the total can be seen Φ for each item that occurs in each transaction for each item that occurs in each transaction For k = 1, If you look at the table, all items have a frequency of item sets greater than Φ for:

$_1F$ = *{{Asparagus}, {Beans}, {Broccoli}, {Corn}, {Green Peppers}, {Squash}, {tomatoes}}*

for k = 2, A table is needed for each pair of items so that we get pair tables for each item as below:

Table 5. Asparagus, Beans

Transaction	Asparagus	Beans	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	1	S
11	0	1	S
12	1	1	P
13	1	1	P
14	0	1	S
Σ			5

Look at the tables for a set that has 2 elements. P means the items are sold together, while S means no items are sold together or no transaction occurs. The blue column indicates the number of P. Σ represents the total frequency of the item set. The total frequency of the item set must be greater than the frequency of the item set (Σ >= Φ), then we can get:

$_2F$ = *{{Asparagus, Beans}, {Asparagus, Squash}, {Beans, Corn}, {Beans, Squash}, {Beans, Tomatoes}, {Broccoli, Green Peppers}, {Corn, Tomatoes}}*

for k = 3, A table is needed for each pair of items so that the pair tables for each item are obtained as below:

Table 6. Asparagus, Beans Dan Broccoli

Transaction	Asparagus	Beans	Broccoli	f
1	0	0	1	S
2	1	0	0	S
3	0	1	0	S
4	0	1	0	S
5	1	1	1	P
6	1	1	0	S
7	0	0	0	S
8	0	0	1	S
9	1	1	0	S
10	0	1	0	S
11	0	1	1	S
12	1	1	0	S
13	1	1	0	S
14	0	1	1	S
Σ				1

Look at the tables for a set that has 3 elements. P means items are sold together, while S means no items are sold together or no transaction occurs. The blue column indicates the number of P. Σ represents the total frequency of the item set. The total frequency of the item set must be greater than the frequency of the item set (Σ >= Φ), then we can get:

After that, calculate Support and Confidence in the following ways:

1. Look for Support Value

$$Support = \frac{\Sigma\ Items\ purchased\ at\ once}{\Sigma\ Total\ number\ of\ Transactions} x100\% \quad (4)$$

$$Support = \frac{4}{14} x100\% = 28.6\%$$

2. Look for Confidence Value

$$Confidence = \frac{\Sigma\ Items\ purchased\ at\ once}{\Sigma\ Number\ of\ Transactions\ in\ the\ Antecedent\ section} x100\% \quad (5)$$

$$Confidence = \frac{4}{14} x100\% = 80\%$$

4. CONCLUSION

In conclusion, the application of association analysis, specifically market basket analysis, has provided invaluable insights into consumer behavior within the context of supermarket shopping. By examining the associative rules derived from transactional data, we have identified meaningful patterns such as the strong likelihood of customers purchasing bread and milk together. The support and confidence metrics have not only quantified the prevalence of these associations but also allowed us to prioritize rules based on their significance. The 80% confidence level set for rule selection ensures a robust foundation for actionable insights, enabling supermarket operators to make informed decisions regarding product placement, inventory management, and targeted marketing strategies. This research underscores the strategic implications of leveraging association analysis in the supermarket industry. The multiplication of Support and Confidence, leading to metrics like Lift, enables supermarkets to prioritize high-impact rules that can shape business strategies. The ability to identify and act upon strong associations in consumer purchasing patterns empowers supermarkets to enhance the overall shopping experience, optimize resource allocation, and stay competitive in a dynamic market. As technology continues to advance, the integration of data-driven decision-making, as demonstrated through association analysis, is pivotal for supermarkets looking to adapt to evolving consumer preferences and maintain a responsive and customer-centric approach in their operations. For future research development, it is recommended to explore the integration of association analysis with advanced technologies such as machine learning and artificial intelligence. Combining these methods can improve the accuracy of identifying more complex and dynamic consumer patterns, and enable faster adaptation to changes in shopping behaviour. In addition, the research focus can be extended to the e-commerce sector and the use of digital platforms, where transactional data is becoming larger and more diverse. Exploring the social impact and ethics of association analysis in the context of consumer privacy is also a crucial research area. Furthermore, research could focus on developing new algorithms to improve the interpretation and management of association analysis results, providing added value in strategic decision-making for retail companies. Finally, collaboration between academic research and industry can be enhanced to ensure that research findings can be implemented more effectively in everyday business practices.

REFERENCE

Abdullah, Z., Herawan, T., Ahmad, N., & Mat, M. (2011). Mining significant association rules from educational data using critical relative support approach. *Procedia - Social and Behavioral Sciences*, *28*, 97–101. https://doi.org/10.1016/j.sbspro.2011.11.020

Al-maolegi, M., & Arkok, B. (2014). *A N I MPROVED A PRIORI A LGORITHM FOR*. *3*(1), 21–29.

Antonie, L. (n.d.). *I NTRODUCTION TO D ATA M INING WHO AM I ?*

Apriori, M., & Kunci, K. (n.d.). *No Title*.

Bag, S., Gupta, S., & Kumar, S. (2021). Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development. *International Journal of Production Economics*, *231*, 107844.

Basuki, A., & Cahyani, A. D. (n.d.). *Digital Marketing*. Media Nusa Creative (MNC Publishing).

Bharadiya, J. P. (2023). Leveraging machine learning for enhanced business intelligence. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY*, *7*(1), 1–19.

Cameron, A. J., Charlton, E., Ngan, W. W., & Sacks, G. (2016). A systematic review of the effectiveness of supermarket-based interventions involving product, promotion, or place on the healthiness of consumer purchases. *Current Nutrition Reports*, *5*, 129–138.

Chen, L. (2012). *The Research of Data Mining Algorithm Based on Association Rules*. 548–551.

Deshpande, S. P., & Thakare, V. M. (2010). *DATA MINING SYSTEM AND APPLICATIONS : A REVIEW*. *1*(1), 32–44.

Gomez, M. I., McLaughlin, E. W., & Wittink, D. R. (2004). Customer satisfaction and retail sales performance: an empirical investigation. *Journal of Retailing*, *80*(4), 265–278.

Gufroni, A. I. (2018). *Associative Analysis Data Mining Pattern Against Traffic Accidents Using Apriori Algorithm*. *5*(2), 91–104.

Heller, A. (2006). *Consumer-centric category management: How to increase profits by managing categories based on consumer needs*. Wiley Online Library.

Hokka, J. (2023). *Controlled Experiments for Data-driven Retail Optimization*.

Husain, T., & Hidayati, N. (2021). *THE OPTIMIZE OF ASSOCIATION RULE METHOD FOR THE BEST BOOK PLACEMENT PATTERNS IN LIBRARY : A MONTHLY TRIAL*. *4*(2), 53–59. https://doi.org/10.31943/teknokom.v4i2.63

Jiménez-Guerrero, J. F., Pérez-Mesa, J. C., de Burgos-Jiménez, J., & Piedra-Munoz, L. (2018). Considering the consumer in the design of a supply chain of perishables. *International Food and Agribusiness Management Review*, *21*(4), 525–542.

Kasieczka, G., Nachman, B., Huang, Z., Cha, Y., Seo, J., & Kim, J. (2021). *Data Mining preparation : Process , Techniques and Major Issues in Data Analysis Data Mining preparation : Process , Techniques and Major Issues in Data Analysis*. https://doi.org/10.1088/1757-899X/1090/1/012053

Kaur, J., & Madan, N. (2015). *Association Rule Mining : A Survey*. *8*(7), 239–242.

Kumar, S. (2008). A study of the supermarket industry and its growing logistics capabilities. *International Journal of Retail & Distribution Management*, *36*(3), 192–211.

Liao, S., Chu, P., & Hsiao, P. (2012). Expert Systems with Applications Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems With Applications*, *39*(12), 11303–11311. https://doi.org/10.1016/j.eswa.2012.02.063

Meliza, J., & Lubis, K. S. (2022). *Manajemen Pemasaran dan Strateginya*. Penerbit P4I.

Mishra, S., & Tyagi, A. K. (2022). The role of machine learning techniques in internet of things-based cloud applications. *Artificial Intelligence-Based Internet of Things Systems*, 105–135.

Niemeier, S., Zocchi, A., & Catena, M. (2013). *Reshaping retail: Why technology is transforming the industry and how to win in the new consumer driven world*. John Wiley & Sons.

Pereira, M. M., & Frazzon, E. M. (2021). A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. *International Journal of Information Management*, *57*, 102165.

Purohit, S., & Jain, A. K. (2020). Evolution of Fuel Retail in India Vis-a-vis Indian Customer: Shift in Consumer Behavior. *International Journal of Management*, *11*(8).

Ramageri, M. (n.d.). *DATA MINING TECHNIQUES AND APPLICATIONS*. *1*(4), 301–305.

Sarno, R., Dewandono, R. D., Ahmad, T., & Naufal, M. F. (2015). *Hybrid Association Rule Learning and Process Mining for Fraud Detection*. *April*.

Siekmann, J., Hartmanis, J., & Leeuwen, J. Van. (n.d.). *Lecture Notes in Artificial Intelligence*.

Sousa, L. R. De, Carvalho, V. O. De, Penteado, B. E., & Jos, F. (2021). *A Systematic Mapping on the Use of Data Mining for the Face-to-Face School Dropout Problem*. *1*(Csedu), 978–989. https://doi.org/10.5220/0010476300360047

Supriyamenon, M., & Rajarajeswari, P. (2017). *A Review on Association Rule Mining Techniques with Respect to their Privacy Preserving Capabilities*. *12*(24), 15484–15488.

Terblanche, N. S. (2018). Revisiting the supermarket in-store customer shopping experience. *Journal of Retailing and Consumer Services*, *40*, 48–59.

Weber, F., & Schütte, R. (2019). A domain-oriented analysis of the impact of machine learning—the case of retailing. *Big Data and Cognitive Computing*, *3*(1), 11.