JURNAL INFORMATIKA, Vol.12 No.1 April 2025, Page 10-14
ISSN (Print) 2355-6579 | ISSN (Online) 2528-2247
DOI: https://doi.org/10.31294/inf.v12i1.20835

Data Mining Approach to Improve Minimarket Sales using Association Rule Method

Harlinda Lahuddin[1*], Ramdan Satra[2]

[1,2] Universitas Muslim Indonesia
Jl. Urip Sumoharjo KM. 05 Makassar, Indonesia

Correspondence e-mail: harlinda@umi.ac.id

Submission:	Revision:	Acceptance:	Available Online:
04-01-2024	17-12-2024	06-02-2025	04-03-2025

Abstract

This research aims to provide recommendations for the placement of goods sold by the UMI Faculty of Computer Science mini supermarket. A data mining approach is used to determine the position of sales items between related items. This is done to make it easier for customers to search for items to buy based on the type of item. Another problem is determining the best-selling items and also determining the types of items that will receive promotions. The data mining approach uses association rules with a priori algorithms. Association rule mining is a data analysis technique used to find patterns and relationships in big data. This technique is widely used in business to help optimize marketing and sales strategies. The results of the rule association using an a priori algorithm show that if consumers buy 200 milli of Ultra Milk Slim Chocolate, they also buy 600 milli of LE MINERAL with a support value of 10% and confidence of 60%. This shows that these two items are related when consumers purchase.

Keywords: Sales, Association rule mining, Apriori algorithm

1. Introduction

Retail businesses involve the sale of goods to customers through kiosks, markets, department stores, boutiques, and more. In Indonesia, retail businesses generally consist of traditional and modern retail. Modern retail has evolved from traditional retail, influenced by economic, technological, and lifestyle developments. Modern retailers implement marketing mix strategies to attract customers. These strategies aim to create comfort and improve purchasing decisions, encouraging unplanned purchases. However, implementing marketing strategies without complementary methods often results in unsold products, market segmentation mismatches, and inventory oversupply (Mastuhin, Saputra, and Oyama 2021). To address these challenges, effective strategies such as product layout, identifying best-selling items, and determining discounted items are required. This is expected to achieve sales targets and align purchases with customer needs.

The Faculty of Computer Science at UMI has a mini supermarket that serves as a shopping center for students and lecturers. This mini supermarket is named “Gade-Gade,” a term derived from the Makassar language, which means “small shop.” The problem with “Gade-Gade” is that the product arrangement is not based on the relationship between items frequently purchased together by customers. This makes it difficult for customers to find related products. Another issue is the absence of a system to determine promotional items and best-selling products. Therefore, product arrangement needs to be based on customer interests regarding the items they want to purchase, along with the identification of promotional items and best-sellers. Solution to address this is by applying association rule mining to determine the placement of related products, identify best-sellers, and select promotional items. This can make it easier for customers to find related products and view promotions and best-selling items.

Association rule mining is a data analysis technique used to discover patterns and relationships in large datasets. This technique is widely applied in business to optimize marketing and sales strategies. The implementation of this algorithm includes strategically positioning products in stores or on websites. Proper product placement can increase sales and business profits. By using the association rule technique in data mining, patterns and relationships between products frequently purchased or viewed together—either on websites or in physical stores—can be identified. This information can be used to position related products close to each other, increasing the likelihood of purchase and boosting overall sales. Research on the application of association rule data mining for

product placement continues to be conducted to improve the effectiveness and efficiency of marketing and sales strategies. By applying this technique, businesses can increase profits and enhance customer satisfaction by placing relevant products in the right locations.

The use of association rule mining methods in product sales can help identify complex patterns and relationships while enabling more effective product placement. Some popular association rule algorithms include Apriori, Eclat, and FP-Growth (Shaukat, Zaheer, and Nawaz 2017)(Saxena and Rajpoot 2021). These algorithms can be implemented using programming languages such as Python or R with the help of available libraries or modules.

Research related to the implementation of association rule mining has been widely conducted, particularly in determining the layout of products in supermarkets (retail) (Tarigan 2017) (Hartanto 2015)(Wulansai and Chulkamdi 2022), the relationships between items purchased by customers (Saxena and Rajpoot 2021) (Hartanto 2015)(Wulansai and Chulkamdi 2022)(Safitri 2022)(Wulandari and Mursidah 2019) (Sylwia Wrona 2022), marketing and strategy (Citra Dewi, Irawan, and Sormin 2022)(Hartanto 2015)(Brilliant, Handoko, and Sriyanto 2017)(Wulandari and Mursidah 2019)(Safitri 2022)(Sofyan 2019)(Mastuhin, Saputra, and Oyama 2021), customer behavior (Wijaya and Siddik 2022), customer reviews (Maulidah and Bachtiar 2021) and determining sales promotions (Ripaii, Puspitorini, and Astuti 2022). Based on previous studies, further research is needed to determine product arrangement, identify best-selling items, and establish discounted items in the mini supermarket of the UMI Faculty of Computer Science, “Gade-Gade.”.

2. Research Methodology

The stages in this research method are shown in Figure 1 which includes:

a. Literature Review
   In this stage, preliminary data is gathered through literature reviews related to the application of association rule mining. Sources include journals, conference proceedings, and scientific literature.
b. Data Collection
   Data collection was conducted through questionnaires and the Point of Sale (POS) application at the UMI Faculty of Computer Science's mini supermarket.
c. Data Analysis
   The analysis focuses on product layout optimization using association rule mining

Figure 1. Stages of research

Apriori Algorithm

The method is used to uncover data trends with the aim of extracting information from datasets and transforming it into an understandable structure (Saxena and Rajpoot 2021). The data mining process involves various techniques and algorithms from statistics, mathematics, and machine learning to manage and analyze data, such as clustering, classification, regression, association, and categorization. Data mining is applied in various fields, including business, science, medicine, social sciences, and others. The data mining process can be seen in Figure 1.

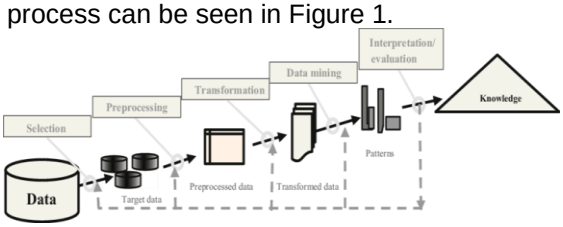

Figure 2. Data Mining Process

The Apriori algorithm is one of the techniques in Data Mining used to find associations between items in a dataset. These associations can be used to identify patterns that frequently occur together in the dataset, such as products that are often bought together in an online store or symptoms of diseases that frequently appear together. This algorithm works by searching for frequent itemsets in the data (Safitri 2022). Below is the formula for the Apriori Algorithm in Data Mining:

1. *Support*:

Support is a metric that measures how often an itemset appears in the dataset, which is crucial in determining how popular a particular combination of items is. It is calculated as the ratio of the number of transactions that contain the itemset to

the total number of transactions in the dataset. This helps identify the most frequent item combinations.

$Support(X)$ = (Number of transactions containing X) / (Total number of transactions)

2. Confidence:

Confidence measures the strength of the association rule. It calculates how often the rule X → Y (if X is bought, Y will also be bought) holds true. A higher confidence value means that if a customer buys X, they are more likely to also buy Y.

$Confidence\ (X \rightarrow Y)$ = (Number of transactions containing both X and Y) / ((Number of transactions containing X)

3. Lift:

Lift measures the strength of a rule by comparing the observed frequency of the combination of items with the expected frequency if the two items were independent. A lift greater than 1 indicates that the items are more likely to be purchased together than by chance, while a lift less than 1 indicates that they are less likely to be purchased together.

$Lift\ (X \rightarrow Y) = Support\ (X \rightarrow Y) / (Support(X) * Support(Y))$

4. Kulczynski:

Kulczynski is a measure of the similarity between two itemsets, X and Y, based on their support values. It is an average of the support values of both X → Y and Y → X. A higher Kulczynski value indicates a stronger similarity between the two itemsets.

$Kulczynski\ (X \rightarrow Y) = 0.5 * [(Support\ (X \rightarrow Y) / Support(X)) + (Support\ (X \rightarrow Y) / Support(Y))]$

5. Jaccard:

Jaccard is a measure of the similarity between two sets, based on the ratio of their intersection to their union. It gives a value between 0 (no overlap) and 1 (identical sets). It is especially useful when comparing itemsets with significant overlap in their purchases.

$\text{Jaccard}\ (X \rightarrow Y) = (X \cap Y) / (X \cup Y)$

In Data Mining, these formulas are used to calculate the values of support, confidence, lift, Kulczynski, and Jaccard for the association rules found using the Apriori Algorithm. These values are used to select the most relevant and significant association rules in the dataset.

3. Result and Discussion

The data collection for this study consists of 100 sales records from Gade-Gade, Faculty of Computer Science, which were used to determine which items are associated with other items in sales transactions. These records provide valuable insights into customer purchasing behavior and are essential for identifying relationships between products. By analyzing these transactions, it is possible to uncover patterns that reveal which products are commonly bought together, providing a deeper understanding of customer preferences and shopping habits.

In addition to analyzing item associations, sales data for one month was used to identify the top 10 best-selling items and their sales frequency. This helps to highlight the most popular products, allowing businesses to optimize their marketing strategies, inventory management, and product placement. Knowing the best-sellers is crucial for focusing promotional efforts and improving the overall sales process.

For the purpose of determining item associations using the Apriori algorithm, preprocessing was conducted to remove variables that do not impact the results. Variables such as transaction ID and transaction amount were excluded because they do not contribute to understanding the relationships between products. This preprocessing step ensures that the analysis focuses on the relevant data, making the results more accurate and meaningful. A sample dataset used for the Apriori algorithm can be seen in Table 1, which showcases the specific items involved in the transactions.

Table 1. Sales Dataset of 100 Records

No	ID Trans	Items
1	29469	TEH PUCUK HARUM 350 ml
2	29470	Ultra Milk Slim cokelat 200ml
3	29471	Marewo Cokelat,Marewo Abon,aoka rasa pandan
4	29472	LE MINERAL 600ml,Marewo Cokelat,Marewo Abon,Ultra Milk Slim cokelat 200ml,TEH PUCUK HARUM 350 ml
…	…	….
….	….	….
94	29562	Momogi Stick Choco 5,Good Day Cappuccino Original 250ml,LE MINERAL 600ml,momogi jagung bakar,sari gandum

No	ID Trans	Items
		coklat 115g,Roma Kelapa Cream Cokelat 40g
95	29563	Good Day Avcdo/D 250 ml,Tora Cafe Iced Milk Late 180ml,LE MINERAL 600ml,Nutriboost orange flavour 300ml,Slai O'lai Biss Srawberry24,aoka rasa pandan
96	29564	Tebs Tea WT/Soda 500ml,LE MINERAL 600ml,Ultra Milk Rs Strw 250ml,aoka nenas,apollo blueberry
97	29565	Momogi Stick Choco 5,Choki-Choki Chococashw 20s,Golda dolce latte 200ml,Chocolatos vanila flavor,Deka jumbo white coffe
98	29566	LE MINERAL 600ml
99	29567	LE MINERAL 600ml
100	29568	Tebs Tea WT/Soda 500ml

Analysis of Item Association Results

The syntax is written using the Python programming language with the help of Google Colab. The analysis begins by processing the sales data using the Apriori algorithm. The initial step involves reading the data using the following syntax:

```
# Load data Google Colab  ( File Excel pada Google Drive) Data Gade2
df= pd.read_excel("Gadde2-100rec-1.xlsx")
df.head()
```

The next step is to remove variables that are not used in the processing of the Apriori algorithm. The syntax for this is as follows:

```
# Membuang kolom No, ID Transaksi data Gade2
data=df.drop(['No','ID Trans'],axis=1)
```

Next, we determine the **fitness value** and **confidence value** for each sales item. The syntax for this is as follows:

```
# Menggunakan fungsi apriori untuk membuat asosiasi
association_rules = apriori(trx, min_support=0.1,
min_confidence=0.20,min_lift=1)
# Membuat list hasil dari algoritma apriori
association_results = association_rules
```

The next step is to display the association rule results for the sales items. The syntax for this is as follows:

```
# Menampilkan hasil asosiasi dari item
pd.set_option('max_colwidth', 1000)
Result=pd.DataFrame(columns=['Rule','Support','Confidence'])
for item in association_results:
    pair = item[2]
    for i in pair:
        items = str([x for x in i[0]])
        if i[3]!=1:
            Result=Result.append({
                'Rule':str([x for x in i[0]])+" -> " +str([x for x in i[1]]),

'Support':str(round(item[1]*100,2))+'%',
                'Confidence':str(round(i[2]*100,2))+'%'
                },ignore_index=True)
Result
```

The results of the Apriori algorithm are as follows:
Support Value = 10%
Confidence Value = 62.5%

Rule:

['Ultra Milk Slim cokelat 200ml'] -> ['LE MINERAL 600ml']

The results of the Apriori algorithm show that with every purchase of 200 ml Ultra Milk Slim Chocolate, customers also purchase 600 ml LE MINERAL. This indicates a strong association between these two items in consumer purchases. The algorithm highlights that these items are frequently bought together, providing valuable insight into customer behavior.

Based on this finding, it is recommended that these products be positioned close together on the shelf. By placing them near each other, customers are more likely to notice both items simultaneously, which can enhance their shopping experience and increase the likelihood of purchasing both products together. Proper product placement can make it easier for customers to find the items they need and encourage unplanned purchases.

It is hoped that by implementing this strategy, the sales of these items at Gade-Gade can be significantly increased. Optimizing product placement based on customer purchasing patterns can help drive higher sales and improve overall store performance. This approach not only benefits the business but also provides a more efficient and satisfying shopping experience for customers.

4. Conclusion

The association rule results using the Apriori algorithm show that if consumers purchase 200 ml Ultra Milk Slim Chocolate, they also buy 600 ml LE MINERAL with a support value of 10% and confidence of 60%. This indicates that these two items are related when customers make a purchase. The best-selling items, based on sales volume, are LE MINERAL 600 ml, which sold 1,403 times in one month, and Teh Pucuk Harum 350 ml, which sold 297 times in the same period. Additionally, the study found that there are 212 products that require special strategies to increase sales so that the sales volume reaches more than 10 units per month. One way to achieve this is by reducing the price of these 212 products, which can help sell these items before their expiration date. Suggestions for future research include using more sales data and applying classification algorithms to determine the number of best-selling or non-best-selling items.

References

Brilliant, Muhamad, Dwi Handoko, and Sriyanto Sriyanto. 2017. “Implementation of Data Mining Using Association Rules for Transactional Data Analysis.” In *3rd International Conferences on Information Technology and Business (ICITB)*, , 177–80.

Citra Dewi, Lika, Eka Irawan, and Rizky Khairunnisa Sormin. 2022. “Penerapan Algoritma Apriori Terhadap Data Mining Penjualan Di Alfamart Berastagi.” *FATIMAH: Penerapan Teknologi dan Sistem Komputer* 1(1): 1–10. https://ejurnal.pdsi.or.id/index.php/fatimah/index.

Hartanto, Wiwin. 2015. “Metode Data Mining Market Basket Analysis Untuk Menentukan Pola Tata Letak Produk Ritel.” In *Seminar Nasional Pendidikan Transformasi Dan Tantangan Masa Depan Bangsa*,.

Mastuhin, Mastuhin, Nurirwan Saputra, and Sunggito Oyama. 2021. “Implementasi Data Mining Dengan Metode Association Rule Mining Untuk Menentukan Peta Pemasaran.” In *Seri Prosiding Seminar Nasional Dinamika Informatika*, Yogyakarta: Universitas PGRI Yogyakarta Program Studi Informatika. http://prosiding.senadi.upy.ac.id/index.php/senadi/article/view/202.

Maulidah, Aisyatul, and Fitra A. Bachtiar. 2021. “Penerapan Metode Association Rule Mining Untuk Asosiasi Ulasan Terhadap Aspek Tempat Wisata Jawa Timur Park 3.” *Jurnal Teknologi Informasi dan Ilmu Komputer* 8(5): 1029.

Ripaii, Ahmad, Sukma Puspitorini, and Reny Wahyuning Astuti. 2022. “Association Rules Mining Untuk Penentuan Promo Pelanggan Pada PT. Indaco Warna Dunia Depo Jambi.” *Jurnal Teknik Informatika Kaputama (JTIK)* 6(1): 101–7. https://www.jurnal.kaputama.ac.id/index.php/JTIK/article/view/719/524.

Safitri, Pradini Nurul. 2022. “Analisis Strategi Penjualan Produk Menggunakan Association Rule Dengan Algoritma Apriori.” Universitas Islam Indonesia.

Saxena, Akash, and Vikram Rajpoot. 2021. “A Comparative Analysis of Association Rule Mining Algorithms.” In *IOP Conference Series: Materials Science and Engineering*, , 012032.

Shaukat, Kamran, Sana Zaheer, and Iqra Nawaz. 2017. “Association Rule Mining: An Application Perspective.” *International Journal of Computer Science and Innovation* 2015(1): 29–38.

Sofyan. 2019. “Implementation of Data Mining with Association Rule in Decision Making for Product Purchase Correlation Using the Apriori Algorithm.” *Journal of Chemical Information and Modeling* 53(9): 1689–99.

Sylwia Wrona. 2022. “Association Rules - Market Basket Analysis.” *RStudio*: 1–18. https://rpubs.com/smwrona/871575.

Tarigan, Jenda Inganta. 2017. “Penerapan Algoritma Apriori Untuk Menentukan Barang Terlaris (Studi Kasus : PT. Indomarco Prismatama Medan).” *Majalah Ilmiah INTI* 12: 184–85.

Wijaya, Patrick, and Muhammad Siddik. 2022. “Penerapan Association Rule Untuk Menentukan Pola Prilaku Konsumen Dalam Membantu Strategi Penjualan.” *Agustus* 6(3).

Wulandari, Ratih;, and Ichsani Mursidah. 2019. “Pola Pembelian Produk Berdasarkan Association Rule Data Mining.” In *Seminar Nasional Teknologi Informasi Dan Komunikasi STI&K (SeNTIK)*, , 265–70.

Wulansai, Zunita, and Mukh Taofik Chulkamdi. 2022. “Penerapan Algoritma Apriori Untuk Menentukan Tata Letak Menempatkan Barang Dagangan Toko Mekar Sari Di Blitar Zunita.” *Jurnal Generation* 6(1): 45–57.