

Increasing Cross-Selling Retail Sales Using the Apriori Algorithm on A Groceries Dataset

Agung Nugroho¹, Ananto Tri Sasongko²

^{1,2}Informatic Engineering, Universitas Pelita Bangsa, Indonesia

Abstract

A company's sustainability is highly dependent on the role of consumers in conducting transactions. As it turns out, consumers have diverse behaviors and characteristics, so the ability to analyze consumer patterns or habits in transactions is needed. A similar situation also occurs in the retail business, which faces several problems in the sales process, such as stock scarcity, unsold products, and differences in product popularity and consumer demand. Therefore, it is necessary to analyze consumer habits in transactions. The association method with the Apriori algorithm has proven effective in analyzing consumer transaction habits in the retail business. The grocery dataset calculations show an average percentage of support values of 5% and a confidence value of 3%. Applying the association method with the Apriori algorithm can assist retail businesses in designing products and determining strategic steps to increase sales, such as providing discounts or promotions on certain products.

Keywords: Consumer Transaction Habits, Data Analysis, Apriori Algorithm, Retail Business

INTRODUCTION

A company's sustainability is highly dependent on the role of consumers in conducting transactions. Consumers have various behaviors and characteristics, so company owners need to be able to analyze consumer patterns or habits in transactions. Recognizing the importance of this analysis, this study aims to apply the association method using the Apriori algorithm to increase the effectiveness of cross-selling on sales data in the retail business (Albab & Hidayatullah, 2022; Alma et al., 2020; Hermaliani et al., 2020; Maryani et al., 2022; Shely et al., 2021).

The retail business faces several problems in the sales process, including stock shortages, unsold products, and differences in product popularity and consumer demand. This indicates the importance of understanding the habits and patterns of consumer transactions in the retail business. In facing this challenge, the Apriori algorithm has proven effective in analyzing consumer buying patterns and providing cross-selling recommendations that can increase sales (Albab & Hidayatullah, 2022; Mahmud & Hartanto, 2020).

Several previous studies have explored the application of the Apriori algorithm in the context of retail sales. A study by Priyanto (2022) shows that the Apriori algorithm has succeeded in increasing retail cross-selling (Priyanto & Arifa, 2022). The study applied the Apriori algorithm to retail sales data and identified significant buying patterns. The results show an increase in additional product sales by implementing the cross-selling recommendations generated by the Apriori algorithm.

Research by Adiwihardja et al. (2019) also highlighted the importance of the Apriori algorithm in optimizing sales strategies in a retail business. They use the Apriori algorithm to identify relationships between frequently purchased products (Adiwihardja et al., 2019). By using cross-selling recommendations from the Apriori algorithm, retail businesses can direct sales strategies more effectively and increase cross-product sales.

In addition, research by Manopo et al. (2021) revealed the benefits of implementing the Apriori algorithm in increasing retail cross-selling (Manoppo et al., 2021). This study analyzes consumer transaction habits and uses the Apriori algorithm to identify significant buying patterns. The results show that retail centers can design more effective promotional strategies by utilizing Apriori's algorithm analysis, such as providing discounts on products often purchased together.

Based on the literature review conducted, applying the association method with the Apriori algorithm to sales data in the retail business has the potential to increase the effectiveness of cross-selling and sales strategy. This study aims to apply this method and provide recommendations for retail businesses in designing products and determining strategic steps to increase sales.

RESEARCH METHOD

The Apriori algorithm is an algorithm used in association analysis in data mining. The main goal of this algorithm is to find significant association patterns in transactional datasets or datasets consisting of items that occur together.

The basic principle of the Apriori algorithm is to use a bottom-up approach to find frequent item sets in a dataset. The Apriori algorithm uses the concept of association rules, where association rules are statements of the form "If A, then B," indicating the relationship between items or itemsets in a dataset.

The main steps in the Apriori algorithm are as follows and depicted in Figure 1:

Step 1:

- Generates a 1-item candidate itemset.
- Counts the number of occurrences of each item in the dataset.
- Stores 1-item itemsets that meet the minimum support requirements (min_support).

Step 2:

- Generates a k-item candidate itemset.
- Combining the frequent itemsets in the previous step produces k-item candidate itemsets.
- Check whether the candidate itemsets are a subset of all transactions in the dataset.
- Store k-item itemsets that meet the minimum support requirements (min_support).

Step 3:

- Repeat step 2 until no new frequent itemsets are found.

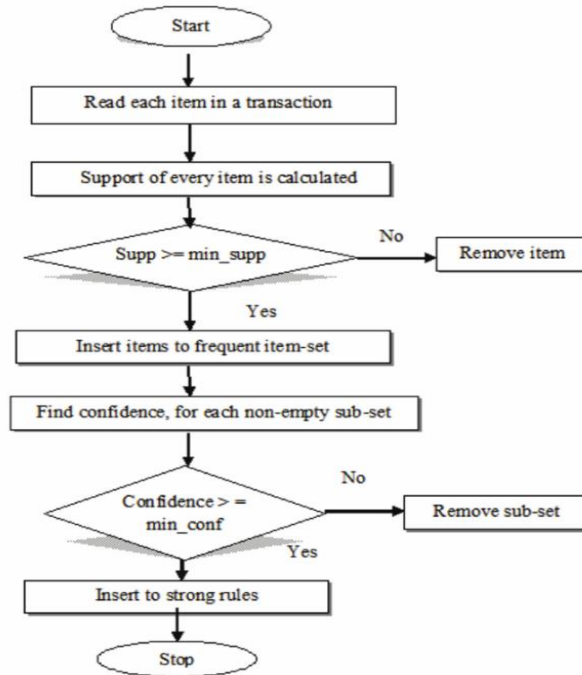


Figure 1. Apriori Algorithm Flowchart

After the frequent itemsets are found, the Apriori algorithm can be used to generate significant association rules. Significant association rules are rules that have a level of support and level of confidence above the specified minimum threshold.

In data science, the Apriori algorithm is used to identify hidden association patterns in transactional datasets. This can help one of them in shopping cart analysis applications, namely looking for purchase patterns that often appear together to support marketing strategies, product recommendations, and customer grouping.

Data Source

The dataset is a public dataset from GitHub with the URL <https://github.com/stedy/Machine-Learning-with-R-datasets/blob/master/groceries.csv>, which is this research's precious public data source. This dataset contains information about sales transactions of daily necessities at a retail store. This dataset provides a realistic representation of consumer buying patterns of daily necessities. The data includes various types of products such as food, beverages, dairy products, and other household products. Thus, this dataset can provide a better understanding of consumer behavior and their purchasing preferences. With this dataset, researchers can perform an in-depth analysis of buying patterns. The Apriori algorithm can be applied to this dataset to identify the relationship between frequently purchased products. This will provide valuable insight into consumer habits and cross-selling opportunities that can be used to increase sales. An example of a grocery dataset is shown in Figure 2 below. Meanwhile, Figure 3 shows a graph of the grocery dataset based on the purchase frequency.

Experimental Design

The experiment was carried out in several stages and used the Python program as a tool to run the a priori algorithm with a public dataset of sales transactions in the retail business. The following are the stages of carrying out the experiment:

1. Implementation of Apriori Algorithm in Python:
 - Python is used as a programming language for implementing the Apriori algorithm.
 - The Panda's library is used for data manipulation, and the mlxtend library implements the Apriori algorithm.
2. Association Analysis with Apriori Algorithm:
 - The Apriori function from the mlxtend library is used to analyze associations in purchasing data.
 - Parameters such as support and confidence values are set according to research needs.
3. Interpretation and Analysis of Results:
 - Analysis of the association results with attention to the value of support, confidence, lift, and leverage.
 - The relationship between item pairs in purchases is discussed based on the analysis results.
4. Result Visualization:
 - Visualization libraries such as matplotlib are used to create graphical or diagrammatic visualizations that depict association analysis results.

```
citrus fruit,semi-finished bread,margarine,ready soups
tropical fruit,yogurt,coffee
whole milk
pip fruit,yogurt,cream cheese ,meat spreads
other vegetables,whole milk,condensed milk,long life bakery product
whole milk,butter,yogurt,ice,abrasive cleaner
rolls/buns
other vegetables,UHT-milk,rolls/buns,bottled beer,liquor (appetizer)
pot plants
whole milk,cereals
tropical fruit,other vegetables,white bread,bottled water,chocolate
citrus fruit,tropical fruit,whole milk,butter,curd,yogurt,flour,bottled water,dishes
beef
frankfurter,rolls/buns,soda
chicken,tropical fruit
butter,sugar,fruit/vegetable juice,newspapers
fruit/vegetable juice
packaged fruit/vegetables
chocolate
specialty bar
```

Figure 2. Sample Groceries Data

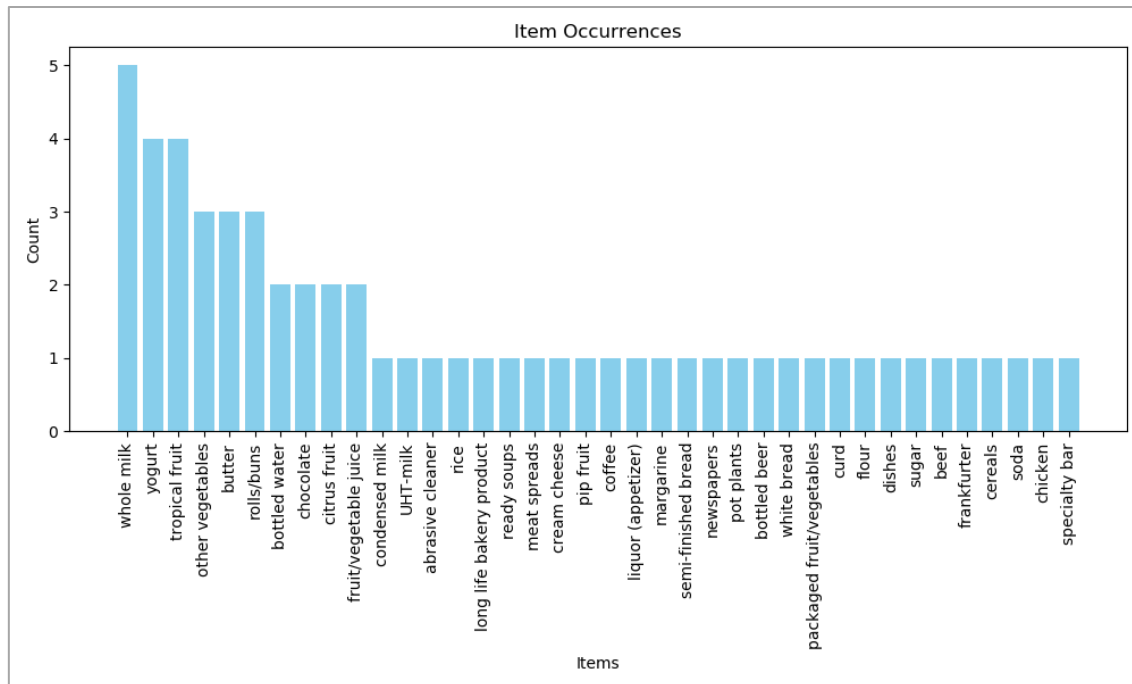


Figure 3. Frequently Purchased Items

RESULTS AND DISCUSSIONS

The analysis results show a relationship between several item pairs in purchases, including "whole milk" and "other vegetables", "rolls/buns," and "soda", and "rolls/buns" and "whole milk". This analysis provides information about each association's level of support, confidence, lift, and leverage. These results are shown in Table 1 and Figure 4. Meanwhile, Figure 5 shows how high the support for each item is.

In the association between "whole milk" and "other vegetables", even though the confidence is relatively low, negative leverage indicates a negative effect. This indicates that purchasing "whole milk" and "other vegetables" tends to occur separately rather than together. Similar results were also found in the association between "other vegetables" and "whole milk". Purchases of these two items also tend to occur separately. However, in the association between "rolls/buns" and "soda", a strong relationship was found with relatively high confidence, a lift greater than 1, and positive leverage. This suggests that buying "rolls/buns" and "soda" tends to occur together with a significant positive effect. Similar results were found for the association between "soda" and "rolls/buns". Purchases of these two items also tend to occur together. Meanwhile, in the association between "rolls/buns" and "whole milk" and between "whole milk" and "rolls/buns", a weak relationship was found. Low confidence, less than 1 lift, and negative leverage indicate that the purchases of these two items tend to occur separately. This analysis provides insight into the relationship between the items in the dataset and the extent to which the items are related in purchase. This information can be used for product placement planning, sales strategies, and promotions to increase overall sales.

Table 1. Association Rule Results

no	antecedents	consequences	support	confidence	elevator	leverage
0	(whole milk)	(other vegetables)	0.056773	0.202128	0.985127	-0.000857
1	(other vegetables)	(whole milk)	0.056773	0.276699	0.985127	-0.000857
2	(soda)	(rolls/buns)	0.043825	0.230366	1.223746	0.008013
3	(rolls/buns)	(soda)	0.043825	0.232804	1.223746	0.008013
4	(whole milk)	(rolls/buns)	0.041833	0.148936	0.791174	-0.011041
5	(rolls/buns)	(whole milk)	0.041833	0.222222	0.791174	-0.011041

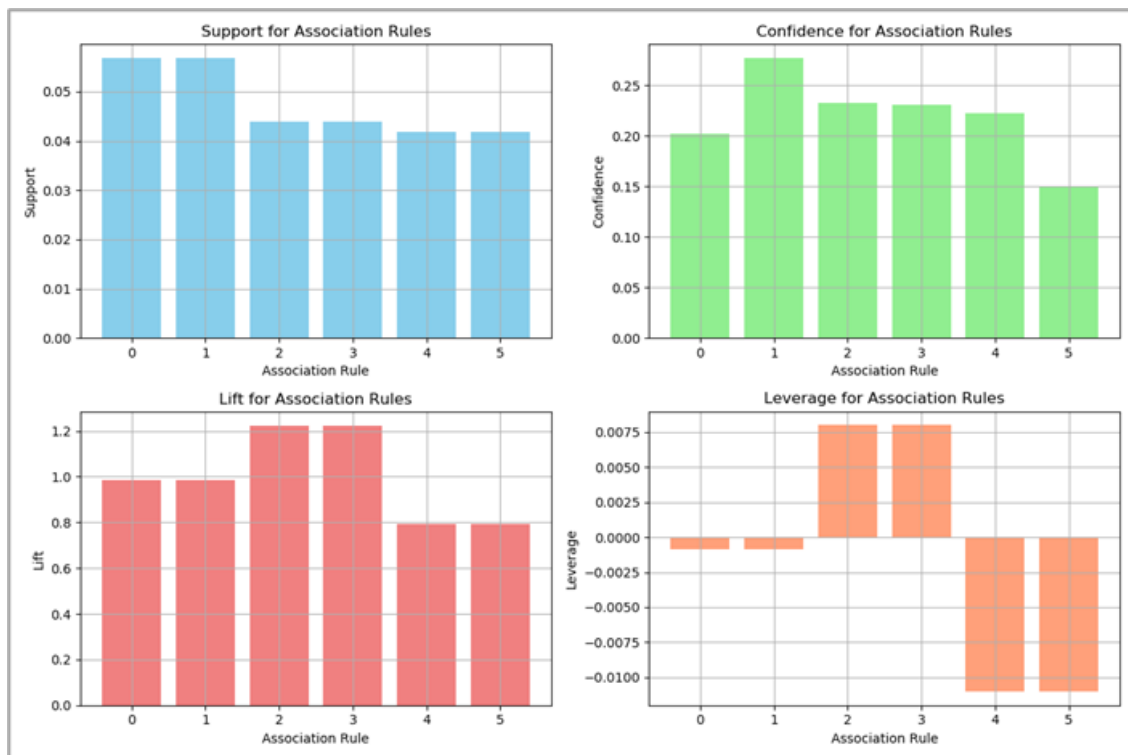


Figure 4. Graph of Association Rule Results

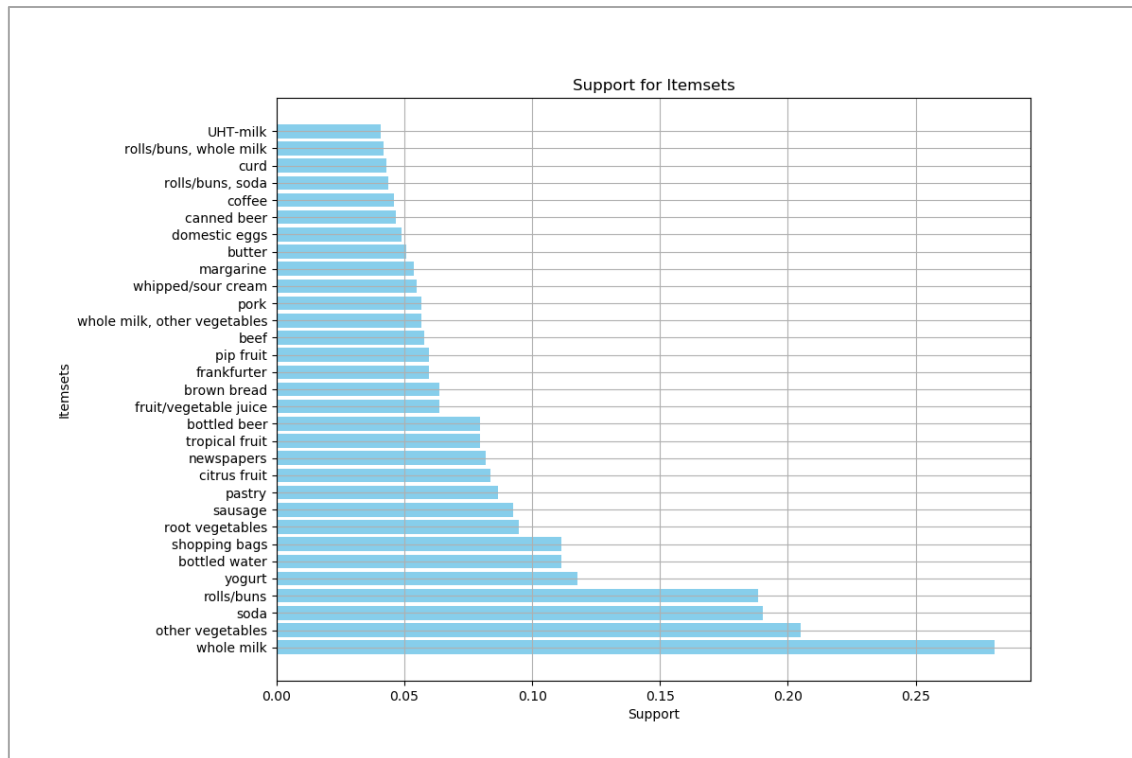


Figure 5. Support Value for Each Item

CONCLUSION

In this study, the Apriori algorithm was applied to analyze associations in purchasing data. Identification of the relationship between certain items such as "whole milk" and "other vegetables", "rolls/bread" and "soda", as well as "soda" and "roll/bread" has been carried out. The analysis results show that some items tend to be purchased together, while others are purchased separately. While there are significant associations, such as between "rolls/bread" and "soda", it should be noted that this analysis does not show a definite cause-and-effect relationship. However, these results provide valuable insights for product placement, sales strategy, and promotions to increase overall sales. The Python programming language and libraries such as pandas are used for data manipulation and mlxtend for implementing Apriori's algorithms. This approach makes it possible to produce association analysis efficiently and accurately. Although it has limitations, including the dataset used and the use of the Apriori algorithm, this research makes an important contribution to understanding buying patterns and forms the basis for further research. This study shows that association analysis using the Apriori algorithm provides valuable insights into understanding purchase patterns. This approach can support better decision-making in business and improve overall efficiency and profitability.

	1st Pelita International Conference Volume 01 No 01 September 2023 E-ISSN: 3026-4235 https://jurnal.pelitabangsa.ac.id/index.php/pic	
---	--	---

REFERENCES

- Adiwiwardja, C., Hardi, N., Widyastuty, W., Information, T., BSI Jakarta, A., Information, S., & Nusa Mandiri Jakarta, S. (2019). Implementation of Cosmetic Sales Data Mining at Zahrani's Store Using the Apriori Algorithm. In *Journal Speed-Sentra Research Engineering and Education* (Vol. 11, Issue 2).
- Albab, MU, & Hidayatullah, D. (2022). Application of the Apriori Algorithm in Store Inventory Information Systems. *JOURNAL OF MEDIA INFORMATIKA BUDIDARMA* , 6 (3), 1321. <https://doi.org/10.30865/mib.v6i3.4160>
- Alma, E., Utami, E., & Wahyu Wibowo, F. (2020). Implementation of Apriori Algorithms for Product Recommendations at Online Stores Implementation of Apriori Algorithms for Product Recommendations at Online Stores. *Citec Journal* , 7 (1).
- Hermaliani, EH, Kurniawati, L., Haryanti, T., Mutiah, N., Kurniawan, A., & Renhoran, BS (2020). Data Mining Technique to Determine the Pattern of Fruits Sales & Supplies Using Apriori Algorithm. *Journal of Physics: Conference Series* , 1641 (1). <https://doi.org/10.1088/1742-6596/1641/1/012070>
- Mahmud, R., & Hartanto, A. (2020). APPLICATION OF LAPTOP RECOMMENDED DATA MINING USING APRIORI ALGORITHM. *JUISI* , 06 (02).
- Manoppo, DG, Wahyudin, MI, & Winarsih, W. (2021). Akshara.co Shoe Store Marketplace Web Design with a Recommendation System Using Apriori Algorithm Calculations. *JOURNAL OF MEDIA INFORMATION BUDIDARMA* , 5 (3), 1173. <https://doi.org/10.30865/mib.v5i3.3091>
- Maryani, I., Revianti, O., & Muhammad Nur, H. (2022). Implementation of Data Mining on Sales at GOC Cosmetics Stores Using the Apriori Algorithm Method. In *Indonesian Journal on Software Engineering (IJSE)* (Vol. 8, Issue 1). <http://ejournal.bsi.ac.id/ejurnal/index.php/ijse92>
- Priyanto, AH, & Arifa, AB (2022). IMPLEMENTATION OF MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM IN MINIMARKET. *Journal of Informatics Engineering (Jutif)* , 3 (5), 1423-1429. <https://doi.org/10.20884/1.jutif.2022.3.5.606>
- Shely Amalia, F., & Darwis, D. (2021). ANALYSIS OF HANDPHONE AND ELECTRONIC SALES DATA USING APRIORI ALGORITHM (CASE STUDY: CV REY GASENDRA) (Vol. 2, Issue 1).