Implementation of Apriori Algorithm for Data Mining on Sales Transaction Data

Hana Bernika Sabila*
Department of Information Technology
Universitas Riau
Pekanbaru, Indonesia
hana.bernika1436@student.unri.ac.id

Feri Candra
Department of Information Technology
Universitas Riau
Pekanbaru, Indonesia
feri.candra@lecturer.unri.ac.id

Corresponding author: Hana Bernika Sabila, hana.bernika1436@student.unri.ac.id

Abstract— Angkasa Mart Store is currently experiencing a decline in sales for specific products, leading to the implementation of the Apriori Algorithm to create bundled offerings that combine less popular items with top-selling products, aiming to revitalize sales and promote the underperforming inventory. Following the CRISP-DM methodology, the study analyzes sales transaction data from June to July in 2022, covering 65,892 purchased items, to extract ten critical association rules essential for devising bundled packages. The study's findings propose two strategies for package composition: first, the development of bundled packages comprising strongly related products, identified through comprehensive data analysis and positively received by the store; second, the introduction of 21 bundled packages consisting of products with a relatively weaker relationship, effectively expanding consumer choices and encouraging additional purchases within the store. By implementing the Apriori Algorithm and adhering to the CRISP-DM methodology, this study effectively formulates bundled product packages for Angkasa Mart Store, addressing the challenge of declining sales and contributing to an overall improvement in business performance and customer satisfaction.

Keywords—Apriori Algorithm, Association Rule, Bundle, Data Mining

I. INTRODUCTION

In today's digitally advanced era, retail companies face the challenge of increasing product sales amidst rapid technological advancements and a competitive trading landscape. Effectively managing sales data and identifying patterns that can enhance business performance has become complex due to the growing volume and complexity of the data. Manual analysis of this data is inefficient and ineffective. Data mining technique is a method of extracting potentially useful information from data that was previously unknown [1]. One such technique is Market Basket Analysis, which involves identifying groups of products frequently purchased together [2]. This technique provides valuable insights into customer behavior and popular product combinations, aiding in strategic decision-making for retailers. This study focuses on the application of the Apriori Algorithm, a widely used data mining algorithm, to discover popular product combinations for a retail store named Angkasa Mart. The goal is to address the challenge of increasing sales for less popular products by implementing a bundling strategy. Bundling is a collection of different products or services sold together as a single package [3], it can involves offering these less popular products together with popular ones as a package, potentially increasing their appeal to customers [4]. The Apriori Algorithm is known for its ability to efficiently mine association rules from large datasets. By analyzing transactional data from Angkasa Mart, the algorithm will identify frequent itemset and generate meaningful insights into popular product combinations. This information can help the store make informed decisions regarding inventory management, pricing strategies, and marketing campaigns.

Several research papers have utilized the Apriori Algorithm for bundling strategies, indicating its relevance in various contexts. In the study by Eky Angliara, titled "Development of Bundling Module Using Association Rule Apriori Algorithm in the Sales Information System of Adien Mart as a Promotion for Less Popular Products," the focus is on enhancing sales of less popular products through bundling strategies [5]. Similarly, Muhamat Maariful and Wiji Setyaningsih, in their article "Decision Support System for Determining Sales Bundle of Products using the Apriori Method," emphasize the development of a decision support system for identifying product matches to optimize profitability [6]. Iqbal Nur Muhammad, Mochammad Fakhrul Islam, and Aryo Nugroho's research highlights the identification of effective association rules for promotional bundling [7], while Anni Fauiyyah's work explores product bundling strategies, particularly emphasizing the key role of shampoos in bundled products [8]. Additionally, Sri Kuswayati and Djajasukma Tjahyadi's study underscores the importance of sales strategies in minimizing promotion costs and accelerating sales turnover, ultimately yielding association rules crucial for enhancing overall sales and revenue [9].

II. METHODOLOGY

A. Apriori Algorithm

There exist multiple models that implement association technique systems, one of which is the Apriori Algorithm. The distinguishing feature of the Apriori Algorithm is its use of a collection of itemset, or frequent itemset, by utilizing subsets as a constituent part of the frequent itemset. According from the book [7], the primary elements of the Apriori Algorithm system are as follows:

- Search for frequent itemset (a set of itemset that meets the minimum support value) from the transaction database.
- Remove itemset that do not meet the minimum support level.
- Create association rules from the itemset that meet the minimum confidence value in the database.

B. Apriori Testing Success Measure Metrics

The success measures of association techniques differ from other techniques such as classification techniques, which use metrics such as Recall, Precision, and F-Score [10]. To evaluate the association rules, the metrics commonly used are support and confidence. These metrics are also considered as interestingness measures used to derive benefits. Apart from these two metrics, several other metrics can be used to determine the success measure that suits each user's needs [11].

The equation for calculating Support value:

$$Support\ (X) = \frac{Number\ of\ transactions\ containing\ X}{Total\ transaction} \quad (1)$$

The equation for calculating Confidence value:

$$Confidence\ (A{=}{>}B) = \frac{Number\ of\ transactions\ containing\ A\ and\ B}{Number\ of\ transactions\ containing\ A} \quad (2)$$

C. Lift Ratio

The lift ratio is one of the interestingness measures used to determine the strength of association rules in transaction data compared to the occurrences of items in random situations [12]. The evaluation derived from a large lift ratio equals 1, indicating the usefulness of the rule. The higher the lift ratio value, the stronger the association rules are [13]. On the other hand, when the lift ratio is less than 1, it indicates poor cross selling performance which is not an effective strategy to increase sales [14].

The equation for calculating Lift Ratio value:

$$BC = \frac{Number\ of\ transactions\ containing\ B\ (concequent)}{Total\ transaction} \quad (3)$$

$$Lift\ ratio\ (X) = \frac{Confidence\ (A{\Rightarrow}B)}{Benchmark\ confidence\ (A{\Rightarrow}B)} \quad (4)$$

D. FIND THE LESS POPULAR PRODUCT

In this research phase, a list of historical records of less popular products is provided, where the count is equal to 1 or 0, indicating that these products have only been involved in one or no transactions. Additionally, a list of historical purchase records of less popular products, along with association rules, is provided as an overview for recommendations on strategies to increase sales of these less popular products [15].

III. RESULT AND DISCUSSION

A. The Association Rules

TABLE I. TEN STRONG ASSOCIATION RULES WITH MINIMUM CONFIDENCE 0.08

Index	Antecedents	Consequents	Support	Confidence	Lift
1	Indomie Kari Ayam 72 G	Indomie Goreng Special 80 G	0.004639	0.33330	19.44
2	Indomie Goreng Special 80 G	Indomie Kari Ayam 72	0.004639	0.270517	19.44
3	Indomie Soto Medan 70 G	Indomie Goreng Special 80 G	0.002189	0.291667	17.01
4	Indomie Goreng Special 80 G	Indomie Soto Medan 70 G	0.002189	0.127660	17.01
5	Selera Keluarga 80 G	Mekar Jaya 10.000	0.002658	0.240566	11.02
6	Mekar Jaya 10.000 dan	Selera Keluarga 10.000	0.002658	0.121718	11.02
7	Prendjak Celup 25	Gula Pasir	0.003023	0.287129	7.78
8	Gula Pasir	Prendjak Teh Celup 25	0.003023	0.081921	7.78
9	Minyak SIIP 900ml Bantal	Gula Pasir	0.002919	0.093645	2.54
10	Fortune Bantal 1 L	Gula Pasir	0.002450	0.092885	2.52

TABLE II. ASSOCIATION RULES WITH THE CATEGORY

Rules of Product		Category	
Antecedent	Consequent	Antecedent	Consequent
Indomie Kari Ayam 72 G	Indomie Goreng Special 80 G	Instant Food	Instant Food
Indomie Goreng Special 80 G	Indomie Kari Ayam 72	Instant Food	Instant Food
Indomie Soto Medan 70 G	Indomie Goreng Special 80 G	Instant Food	Instant Food
Indomie Goreng Special 80 G	Indomie Soto Medan 70 G	Instant Food	Instant Food
Selera Keluarga 80 G	Mekar Jaya 10.000	Unknown	Snack
Mekar Jaya 10.000 dan	Selera Keluarga 10.000	Snack	Unknown
Prendjak Celup 25	Gula Pasir	Basic Necessities	Breed Drink
Gula Pasir	Prendjak Teh Celup 25	Breed Drink	Basic Necessities
Minyak SIIP 900ml Bantal	Gula Pasir	Unknown	Basic Necessities
Fortune Bantal 1 L	Gula Pasir	Beverage	Basic Necessities

In this study, it obtained 10 association rules based on transaction data. These rules were identified using a minimum support value of 0.002, indicating that they appeared at least 131 times. Additionally, a minimum

confidence value of 0.08 (8%) was used to determine the level of confidence that the consequent is purchased due to the antecedent. Furthermore, a minimum lift value of 1 was employed as a testing parameter to establish the validity of the association rules. This value ensures that the identified rules occur more frequently compared to the products appearing randomly or individually.

The Table 1 presents the top 10 association rules along with their corresponding values. The association rules are ranked in descending order based on their high lift values, followed by confidence and support values. The highest-ranked association rule is associated with *Indomie Kari Ayam 72 G* followed by Indomie Goreng Special 80 G, with a support value of 0.004639 or equivalent to 305 occurrences in transactions. The confidence level for this rule is 33%, and the lift value is 19.44. On the other hand, the lowest-ranked association rule is between Fortune *Bantal* 1 L and *Gula Pasir*, with a support value of 0.002450 or equivalent to 161 occurrences in transactions. The confidence level for this rule is 0.092885 or 9%, and the lift value is 2.52. The category of each item will serve as a guideline for the next bundled package.

B. Less Popular Product

TABLE III. LESS POPULAR PRODUCT IN JUNE

Index	Name of Product	Price (IDR)	Amount	Category
1	*Chil-Mil 2 Reg 800*	168600	1	Formula Milk
2	*St.Mtc Set Parfum 1st Hitam*	89500	0	Insecticide
3	*Rice of Kuriak Kusuik 10kg*	160000	0	Food
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
3110	*Pro Beras Merah 20 G*	2000	0	Baby Food
3111	*So Nice 1000*	1000	0	Food

The Table III presents a snippet list of less popular products, indicated by a quantity of either 1 or 0, that were purchased in June. The table above also presents a snippet list of less popular products, indicated by a quantity of either 1 or 0, that were purchased in June. There are several transaction records between popular product combinations or association rules with unpopular products.

TABLE IV. LESS POPULAR PRODUCT IN JULY

Index	Name of Product	Price (IDR)	Amount	Category
1	BMT 1 Regular 800	183000	1	Formula Milk
2	Chil School 4 Vanila 800	161600	0	Formula Milk
3	Rice SANIA 10kg	143500	0	Basic Necessities
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
3005	So Nice 1000	9500	0	Food
3006	Max Bio+ 60ml	9500	0	Wash Equipment

TABLE V. TRANSACTION RECORDS BETWEEN POPULAR AND UNPOPULAR ITEMS

Index	Rules of Product	Less Popular Product	Item Category	
			Rules	Less Popular Product
1	Indomie Goreng Special 80 G, Indomie Kari Ayam 72 G	Johnsons Baby Cream 50	Food, Food	Kids Equipment
2	Indomie Goreng Special 80 G, Indomie Kari Ayam 72 G	Koroku Rumput Laut 70 Gr	Food, Food	Food
3	Indomie Goreng Special 80 G, Indomie Kari Ayam 72 G	Kodomo Shampo Orange 200	Food, Food	Kids Shampoo
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
20	Fortune Bantal 1 L, Gula Pasir	Miranda T2 Cocoa 20ml	Drink, Basic Necessities	Breed Drink
21	Fortune Bantal 1 L, Gula Pasir	Biore Lively Refresh 450 Ml P	Drink, Basic Necessities	Body Wash

C. Recommendations Strategies for Bundles

The author's recommendation for this bundled package is based on other authors/researchers and can be summarized into three key points:

- Association rules can be used to create bundled packages with other products, increasing the likelihood of customer purchases.
- Bundled packages can serve as a strategy to boost sales of less popular products by linking them with popular ones.
- Bundled packages with related or relevant products can benefit both the company and unrelated products, even if they are not directly related.

These recommendations have been categorized as good by the supermarket. However, for future research, the researcher suggests that it would be even more beneficial to include expiration dates for each product in order to make the bundled products more efficient and targeted. The first recommendation is to create bundled packages based on association rules or combinations of popular products with relevant less popular products, using a category-based approach.

TABLE VI. FIRST RECOMMENDATION BUNDLE PACKAGE

Rules of Product	Less Popular Product	Bundle Product Recommendations
If Prendjak Teh Celup 25 then Gula Pasir	*Tong Tji Teh Melati 50G Item Category: Breed Drink*	*Tong Tji Teh Melati 50G & Gula Pasir*
If Minyak SIIP 900ml Bantal then Gula Pasir	*CHAMP Chiken Stik 1000G Item Category: Nugget*	*Champ Chiken Stik 1000G & Minyak SIIP 900ml Bantal & Gula Pasir*

The Table VI provides the following recommendations:

- Tong Tji Teh Melati 50G, which is one of the less popular tea brands in the transaction data, can be bundled with association rules that also involve tea, such as the association rule for Prendjak Teh Celup 25 and Gula Pasir (Sugar).
- A bundled package can be created with CHAMP Chicken Stik 1000G and the association rule for Minyak SIIP 900ml Bantal (Pillow) and Gula Pasir (Sugar). This bundled package can be offered at a discounted price instead of offering a price reduction only for CHAMP Chicken Stik 1000G. The aim is to encourage customers to purchase all three items.

These recommendations suggest bundling less popular products with associated items to increase their appeal and encourage customers to make additional purchases. The response from the store indicates that the bundled packages of relevant products have been categorized as good in terms of success metrics. It is seen as a successful strategy to enhance the sales of less popular products by combining them with relevant items. The second recommendation is to create bundled packages based on association rules or combinations of popular products with unrelated less popular products. This recommendation suggests bundling products from different categories. This recommendation is derived from the transaction history of less popular products with association rules, as mentioned in Table 4 In the transaction history, there are connections between less popular products and association rules. This presents an opportunity for increasing the sales of the less popular products. By creating bundled packages with unrelated but complementary products, it is anticipated that customers will be encouraged to purchase both the popular and less popular items together. This strategy aims to boost the sales of the less popular products and maximize overall sales. The response from the store indicates that the bundled packages of unrelated products have also been categorized as good in terms of success metrics. However, it is important to note that when creating these bundled packages, a plastic separator is needed to prevent any potential spills or leaks from the sanitation category products, such as liquids, from coming into contact with the food category products within the bundled package. This additional measure ensures the safety and integrity of the bundled products.

IV. CONCLUSION

This application of the Apriori Algorithm successfully recommended two bundled package combinations for Angkasa Mart Store, incorporating both related and unrelated less popular products. These recommendations offer practical strategies to improve the store's marketing efforts and drive sales growth. By integrating bundled product offerings that cater to diverse customer preferences, the study highlights the Algorithm's effectiveness in optimizing marketing strategies and enhancing overall business performance.

Based on this research, the author recommends that students or researchers seek information on product expiration dates to provide more targeted recommendations for bundled packages. Including expiration dates in the analysis would enhance the accuracy and efficiency of creating bundled packages. This additional information would help in determining the compatibility and suitability of products within a bundle, ensuring that customers receive fresh and relevant items. Therefore, obtaining and considering expiration dates would improve the precision and effectiveness of bundled package recommendations.

REFERENCES

[1] I. H. Witten, E. Frank, and M. A. Hall, "Data Mining Practical Machine Learning Tools and Techniques", Third Edition. 2011.
[2] M. J. A. Berry and G. S. Linoff, "Data Mining Techniques for Marketing, Sales, and Customer Relationship Management", Second Edition. Indianapolis: Wiley Publishing, 2004.
[3] R. V. Vohra, Prices and Quantities: Fundamentals of Microeconomics. Cambridge University Press, 2020.
[4] Dr. M. A, Marketing Management. Bangalore: Archers and Elevators Publishing House, 2019.
[5] E. Angliara, "Pengembangan Modul Bundling Menggunakan Association Rule Algoritma Apriori Pada Sistem Informasi Penjualan Adien Mart Sebagai Promosi Penjualan Barang Kurang Laris," 2017.
[6] M. Maariful and W. Setyaningsih, "Sistem Pendukung Keputusan Dalam Penentuan Bundling Penjualan Barang dengan Metode Apriori", 2015.
[7] I. N. Muhammad, M. F. Islam, and A. Nugroho, "Prediksi Produk Bundle Pada Promo Dengan Algoritma Apriori Menggunakan Association Rule", Jurnal Ilmu Komputer dan Bisnis, vol. 12, no. 2, pp. 178–188, Nov. 2021.
[8] A. Fauiyyah, "Algoritma Apriori dalam Menentukan Product Bundling" 2019.

[9] S. Kuswayati and D. Tjahyadi, "Market Basket Analysis Menggunakan Algoritma Apriori Untuk Penetapan Strategi Bundling Penjualan Barang", 2011.
[10] C. Zong, R. Xia, and J. Zhang, Text Data Mining. 2021.
[11] W. K. Ng, M. Kitsuregawa, J. Li, and K. C, "Advances in Knowledge Discovery and Data Mining", 10th Pacific-Asia Conference, April 2006.
[12] Y. Vasiliev, Python for Data Science a Hands-On Introduction. 2022.
[13] B. Santoso, "Data Mining: Teknik Pemanfaatan data Untuk Keperluan Bisnis", 2007.
[14] Oracle, "Apriori", Oracle. Nov. 03, 2022.
[15] D. Sarkar, R. Bali, and T. Sharma, "Practical Maching Learning with Python", 2014.

BIOGRAPHIES OF AUTHORS

HANA BERNIKA SABILA was born in Muara Enim, Indonesia and she is a student from Faculty of Informatics Engineering, Universitas Riau Pekanbaru. Her research focus is in the field of data science.

FERI CANDRA obtained a Bachelor's Degree in Electrical Engineering from InstituteSains dan Teknologi Nasional in 1999, a Master's Degree in Electrical Engineering at Universitas Indonesian in 2002, and a Doctoral of Electrical Engineering at Universiti Teknologi Malaysia in 2017. He has been a Lecturer with the Department of Informatics Engineering, Universitas Riau Indonesia, since 2002. His current research interests include signal processing, artificial intelligence, machine learning and data science.