

Machine Learning Application in Small and Micro Business: Uncover Consumer Purchase Behavior using Market Basket Analysis

Felix Aristio^{1,*}; Jimmy Tjen²

¹Department of Digital Business, Faculty of Information Technology,

²Department of Informatics, Faculty of Information Technology

^{1,2}Universitas Widya Dharma Pontianak

Jl. Hos Cokroaminoto No.445, Darat Sekip, Kec. Pontianak Kota, Pontianak, West Kalimantan, Indonesia, 78243

*1felixaristio@gmail.com; 2jimmy.tjen@mathmods.eu

**Corresponding Author*

Abstract— This paper analyzes consumer purchasing patterns in the reseller business of skincare, cosmetics, and health products named Bee Market ID to determine the right marketing strategy based on product sales history data. This study uses the Market Basket Analysis method to identify combinations of products frequently purchased by consumers. The secondary data for this research consists of the sales history of Bee Market ID in August 2023, which is readily exportable from the Shopee e-commerce platform. The consumer purchasing pattern at Bee Market ID, based on Market Basket Analysis, shows that when consumers buy the 'Geamoore Addicted Series Perfume 15ml', they also purchase 'Zudaifu Ointment Cream 15g'. This pattern seems logical as part of a marketing strategy, as individuals who use body care products to enhance their appearance are likely to also seek to enhance their overall appearance with perfume. This pattern can be used for various marketing strategies such as product bundling, product recommendations, and gifts for loyal consumers. While this strategy can boost business growth, there is still untapped data beyond sales history that business owners can leverage. Further research can increase the data for more accurate analysis or employ other methods to enhance business development even further.

Keywords— *Machine Learning; Marketing Management; Market Basket Analysis; Consumer Behavior*

I. INTRODUCTION

Every day, thousands of transactions are recorded in various fields such as marketing, finance, and healthcare. [1]. Although this data is available, sales history is typically only stored as reports, rendering historical data unused [2]. Utilizing sales history data to determine marketing strategies and decision-making is necessary to survive and thrive in highly competitive business environments. Competitive advantage will be gained by those who can extract the knowledge hidden in the data generated by these systems and use it to optimize their marketing decision-making [3]. To extract hidden knowledge within large datasets, data mining processes need to be conducted.

Association rule induction is a powerful method for so-called market basket analysis [4]. Association rule mining is useful for discovering interesting relationships hidden in large datasets [5]. Market Basket Analysis is one of the main techniques of data mining that focuses on the “purchase behavior” of consumers. Despite differences in consumer behavior when meeting their needs, consumers share similarities, such as the desire to maximize satisfaction when consuming a needed product or service [6]. This analysis aims to identify interesting patterns by considering the products bought together by consumers [1]. Support, Confidence, and Lift are the metrics used for evaluating the quality of association rules produced by market basket analysis [6]. This analysis will generate item sets between two product items that have met the measurements using market basket analysis.

Several recent related works on market basket analysis. First, in [7] provides a comprehensive overview of association rule mining techniques and their application to transactional data. The authors explain that there are five evaluation metrics commonly used in association rule mining, such as support, confidence, lift, conviction, and interest. Reference [8] applied market basket analysis to 14,963 transaction records in a minimarket. They utilized the Apriori algorithm to generate recommendations for arranging

items in a display cabinet in the store. Reference [1] applied market basket analysis in the healthcare industry. They identified all the frequent diseases occurring in an area by analyzing a medical dataset. The result provided details of commonly occurring diseases, allowing health authorities to take preventive measures against frequently occurring diseases. Previous research has used market basket analysis for various purposes, such as organizing product layouts in minimarket [8], movie business trend prediction [10], identifying commonly occurring diseases [1], and designing arcade facilities [9]. However, there is still a lack of application of market basket analysis for small and micro business in designing marketing strategies, particularly in leveraging product combinations to increase business revenue.

In this research, we consider a dataset from Bee Market ID which operates as an online reseller business, utilizing the Shopee e-commerce platform to sell skincare, cosmetics, and health products. The business processes, from product stocking to packaging and product shipment, are conducted in Pontianak, West Borneo, Indonesia. The reach of the Shopee e-commerce platform is limited to Southeast Asian countries only. Thus, Bee Market ID's consumer base is confined to Southeast Asia, with most orders originating from Indonesia. Until now, the business has not utilized its data to make informed marketing strategy decisions; marketing efforts have been solely based on the owner's intuition, resulting in ineffective marketing. The reason for choosing Bee Market ID as a case study is its suitability for market basket analysis. Beauty product often consist of multiple types, prompting consumers to purchase several items in a single transaction. Additionally, the available data is comprehensive and well-structured. The analysis utilizes sales history data from the beginning of August 2023 to the end of August 2023 provided by the Shopee e-commerce platform. The results of market basket analysis can be utilized for various marketing strategies such as product bundling, product recommendations, and appropriate product bonuses.

The remaining part of the paper is organized as follows: In Section II, we discuss the research methods and data used. In Section III, the results and discussion are detailed. In Section IV, the conclusions of the results are presented. Lastly, Section V contains the references.

II. RESEARCH METHODS

This research employs quantitative research methodology, focusing on the analysis of numerical data. The dataset utilized in this research comprises secondary data sourced from the sales history of the Bee Market ID business, readily exportable from the Shopee e-commerce platform. The population for this study consists of the sales history data of the Bee Market ID business, while the sample specifically encompasses the sales records from the beginning of August 2023 to the end of August 2023, amounting to a total of 1,520 entries.

The purchase pattern analysis in this study is conducted using market basket analysis. Market basket analysis is carried out using the association rule algorithm along with training data to generate knowledge. In processing the collected data, the researcher utilizes a spreadsheet to preprocess the data, removing unnecessary information and adjusting the format for analysis. Additionally, the researcher employs programming platforms to analyze consumer purchase patterns.

In association rules, the importance of a rule is usually measured by two numbers: Its support, which is the percentage of transactions that the rule can be applied to, and its confidence, which is the number of cases in which the rule is correct relative to the number of cases in which it is applicable [4]. The formula for support and confidence values is as follows [10]:

$$\text{Support}(A \rightarrow B) = P(A \cap B) \quad (1)$$

where $\text{support}(A \rightarrow B)$ denotes how frequently the itemset A occurs in the data, given that itemset B also occurs, and $P(A \cap B)$ indicates the probability of A and B occurring together, meaning the likelihood of a transaction containing both A and B.

$$\text{Confidence}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)} \quad (2)$$

where $\text{confidence}(A \rightarrow B)$ denotes how often itemset B occurs in transactions that contain A, $P(A \cap B)$ the support of both A and B, and $P(A)$ represents the probability of A occurring out of the total transactions.

The selection of item sets in association rules does not only consider the highest confidence percentage; it could also be influenced by a high support percentage. To evaluate a rule, the lift ratio can be utilized. The lift ratio represents the ratio between the support value of the rule with the antecedent and consequent support value. The higher the lift ratio value, the more intriguing the rule, or it can be referred to

as a strong rule [11]. An association rule can be deemed valid if the lift ratio exceeds 1 [12]. The lift ratio value can be calculated using the formula [11]:

$$\text{Lift}(A \rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A) \times \text{Support}(B)} = \frac{\text{Confidence}(A \rightarrow B)}{\text{Support}(B)} \quad (3)$$

where $\text{lift}(A \rightarrow B)$ denotes how much more likely B is to occur with A than would be expected by chance. The rule can be considered as showing negative correlation if the lift value is less than 1, and conversely, if the lift value is greater than 1, it indicates positive correlation [13].

III. RESULT AND DISCUSSION

In this paper, we used the market basket analysis method to identify consumer purchasing patterns. In the case study of the Bee Market ID business, product data needs to be categorized before analysis due to the large number of products. The market basket analysis of the sales history data for Bee Market ID in August 2023 will yield support, confidence, and lift ratio for each category.

Based on the results of the support category analysis, the minimum support threshold is set at 15%, as the level of support begins to vary significantly above this threshold. The categories that meet this criterion are body care, face care, foot care, fragrance, and makeup. According to the support calculation results, the probability of consumers purchasing face care products is the highest at 65%. Therefore, it can be concluded that consumers of Bee Market ID prioritize facial care over other types of body care products.

Table 1. Confidence for Bee Market ID Sales History Data

Item 1	Item 2	Confidence (in %)
Bodycare	Facecare	40.91
Facecare	Bodycare	21.69
Bodycare	Footcare	13.64
Footcare	Bodycare	26.09
Bodycare	Fragrance	25.00
Fragrance	Bodycare	35.48
Bodycare	Makeup	25.00
Makeup	Bodycare	20.75
Facecare	Footcare	10.84
Footcare	Facecare	39.13
Facecare	Fragrance	18.07
Fragrance	Facecare	48.39
Facecare	Makeup	38.55
Makeup	Facecare	60.38
Footcare	Fragrance	8.70
Fragrance	Footcare	6.45
Footcare	Makeup	34.78
Makeup	Footcare	15.09
Fragrance	Makeup	32.26
Makeup	Fragrance	18.87

Source: 2024

To determine the probability of both categories being purchased together based on the confidence level, we analyzed the confidence calculation results. The minimum confidence is set at 35%, approximately 55% of the highest confidence value. The qualifying category combinations include purchasing face care after buying body care, purchasing body care after buying fragrance, purchasing face care after buying foot care, purchasing face care after buying fragrance, and purchasing makeup after buying face care. Among these five combinations, the highest likelihood is for purchasing face care after buying body care, with a confidence level of 60.3%. This analysis aligns well with real-life situations and individual habits, as individuals tend to purchase facial care products when buying body care products. However, the confidence of this highest category combination cannot be utilized as a marketing strategy because it does not meet the criteria of a lift greater than 1.

Table 2. Lift for Bee Market ID Sales History Data

Item 1	Item 2	Lift
Bodycare	Facecare	0.625
Fragrance	Bodycare	1.024
Footcare	Facecare	0.598
Fragrance	Facecare	0.740
Facecare	Makeup	0.923

Source: 2024

Association rules are deemed valid when the lift exceeds 1. In the context of the Bee Market ID business sales history data, a valid association rule emerges from the combination of fragrance and body care categories, with a lift of 1.02. To pinpoint the specific product within these valid categories, one must examine the highest support for products in both categories. The following are the support values for the products in these two categories.

Table 3. Support for Products in the Fragrance and Bodycare Categories

Category	Product	Support (in %)
Bodycare	Seom Gum Body Soap 20g	12.88
	Zudaifu Ointment Cream 15g	15.15
	Bedda Lotong Powder 45g	4.54
	Nivea Extra White Care & Protect Serum 180ml	0.75
	Herborist Juice For Skin Body Serum 180ml	1.51
	Bangkal Wood Herbal Scrub 45g	2.27
	Meilibahenling	0.75
	Nivea Extra Bright 10 Super Vitamins & Skin Foods Serum 180ml	0.75
Fragrance	WENIS Room Freshener 50g	3.78
	Geamoore Zodiac Series Perfume Zodiac 15ml	3.03
	Basicclub x Geamoore Perfume Spray 50ml	3.03
	Geamoore Addicted Series Perfume 15ml	12.88
	Camellia Eau De Perfume Spray 22ml	1.51

Source: 2024

Based on market basket analysis, the item sets 'Zudaifu Ointment Cream 15g' and 'Geamoore Addicted Series Perfume 15ml' emerge as the most viable options for a marketing strategy involving product bundling. Bundling these two products seems logical as part of a marketing strategy, individuals who use body care products to enhance their appearance are likely to also seek to enhance their overall appearance with perfume. In addition to these products, other items within the fragrance and body care categories could serve as valuable references for suggesting products and gifts to consumers who purchase items from either of these categories.

IV. CONCLUSION

The market basket analysis results indicate that the most frequent consumer purchasing pattern at Bee Market ID is when consumers buy the 'Geamoore Addicted Series Perfume 15ml,' they also purchase 'Zudaifu Ointment Cream 15g'. This purchasing pattern can serve as a guideline for decision-making in determining the appropriate marketing strategy at Bee Market ID. Marketing strategies that can be derived from the analysis of purchasing patterns include product bundling, product recommendations, and suitable product bonuses.

The implementation of the product bundling strategy in Bee Market ID's business is executed through the e-commerce platform Shopee by introducing a new product consisting of a bundle of 'Geamoore Addicted Series Perfume 15ml' and 'Zudaifu Ointment Cream 15g'. The pricing for the product bundling is set to be lower than purchasing each product separately, aiming to attract consumers. Despite the decrease in income per sale of the product, total sales increased as consumers opted for the bundled package, resulting in an overall revenue increase.

To enhance the Bee Market ID business, it is crucial to implement a diverse range of marketing strategies instead of relying solely on one approach. This entails offering product recommendations and bonuses to loyal consumers to encourage repeat purchases and foster strong consumer relationships. For Bee Market ID, product recommendations are derived from purchasing patterns analyzed through the chat feature on the Shopee e-commerce platform following specific consumer transactions. Additionally, loyal consumers who consistently make repeat purchases, such as those purchasing fragrance products, can receive appropriate bonus items based on the analysis, such as recommendations or offers for body care products.

Applying marketing strategies based on purchasing pattern analysis at Bee Market ID can enhance the business's revenue, number of orders, repeat purchases, and consumer loyalty. While this strategy can boost business growth, there is still untapped data beyond sales history that business owners can leverage. Further research can increase the amount of data for more accurate analysis or employ other methods to enhance business development even further. With these improvements, the business can continue to thrive despite intense competition. However, this research has limitations because market basket analysis identifies static associations between items, which may not capture the dynamics of customer behavior over time, thus making some associations irrelevant.

V. REFERENCES

- [1] A. B. Rao, J. S. Kiran and P. G, "Application of market–basket analysis on healthcare," *International Journal of System Assurance Engineering and Management*, 2023.
- [2] J. P. B. Saputra, S. A. Rahayu and T. Hariguna, "Market Basket Analysis Using FP-Growth Algorithm to Design Marketing Strategy by Determining Consumer Purchasing Patterns," *Journal of Applied Data Sciences*, 2023.
- [3] T. Brijs, B. Goethals, G. Swinnen, K. Vanhoof and G. Wets, "A data mining framework for optimal product selection in retail supermarket data: The generalized PROFSET model," in *Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000.
- [4] C. Borgelt and R. Kruse, *Induction of Association Rules: Apriori Implementation*, Compstat, 2002.
- [5] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," in *Procedia Computer Science*, 2016.
- [6] F. Kurniawan, B. Umayah, J. Hammad, S. M. S. Nugroho and M. Hariadi, "Market Basket Analysis to Identify Customer Behaviours by Way of Transaction Data," *Knowledge Engineering and Data Science*, vol. 1, no. 1, 2017.
- [7] B. Kalaavathi, B. Sridhevasenaathypathy, N. Chinthamu and D. Valluru, "Retail Shop Sales Forecast by Enhanced Feature Extraction with Association Rule Learning," *International Journal on Recent and Innovation Trends in Computing and Communication*, 2023.
- [8] A. Jain, "Association Rule Mining in Transactional Data: Challenges and Opportunities," *International Journal of Mechanical Engineering*, 2021.
- [9] A. H. Priyanto and A. B. Arifa, "IMPLEMENTATION OF MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM IN MINIMARKET," *Jurnal Teknik Informatika (Jutif)*, 2022.
- [10] D. Barman and N. Chowdhury, "Movie Business Trend Prediction using Market Basket Analysis," *International Journal of Computer Applications*, vol. 74, no. 9, 2013.
- [11] S. Halim, T. Octavia and C. Alianto, "Designing facility layout of an amusement arcade using market basket analysis," *Procedia Computer Science*, 2019.
- [12] N. A. Harun, M. Makhtar, A. A. Aziz, Z. A. Zakaria, F. S. Abdullah and J. A. Jusoh, "The application of Apriori algorithm in predicting flood areas," *International Journal on Advanced Science, Engineering and Information Technology*, 2017.
- [13] E. Hikmawati, N. U. Maulidevi and K. Surendro, "Minimum threshold determination method based on dataset characteristics in association rule mining," *Journal of Big Data*, 2021.
- [14] F. N. Hasan, A. S. Aziz and Y. Nofendri, "Utilization of Data Mining on MSMEs using FP-Growth Algorithm for Menu Recommendations," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 2, 2023.

- [15] C. Ju, F. Bao, C. Xu and X. Fu, "A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit," *Discrete Dynamics in Nature and Society*, vol. 2015, 2015.