Machine Learning in Business: Product Bundling Strategy and Customer Segmentation via Market Basket Analysis Algorithm

[1*]Jacky Felix, [2]Jimmy Tjen
[1]Department of Digital Business, Faculty of Technology Information Technology,
Universitas Widya Dharma Pontianak, Pontianak, Indonesia;
[2]Department of Informatics, Faculty of Technology Information Technology,
Universitas Widya Dharma Pontianak, Pontianak, Indonesia;
Email: [1]21430100_jacky_f@widyadharma.ac.id; [2]jimmy.tjen@mathmods.eu

Abstract

This study employs the Market Basket Analysis (MBA) algorithm to uncover associations between product categories and item names. MBA aims to discern customer purchasing patterns and segmentation across regions. A company selling building materials in Pontianak, West Kalimantan, Indonesia, has never analyzed sales history data to enhance promotional strategies or gain insights into customer segmentation across regions. This study employs an experimental quantitative study with a company in Pontianak as the study object, using primary data such as sales history, customer database, and product database. The study population comprises 12,600 sales transactions, with a sample of 3,462 transactions focusing specifically on the Onda brand from January 2 to December 30, 2023. The results of the MBA algorithm will be evaluated based on support, confidence, and lift values. From the analysis results, associations between product subcategories and names are identified, providing insights for determining bundling or cross-selling strategies based on consumer purchasing patterns, such as combination angle valve JF 11 with basin tap Y 321 C. Customer segmentation based on consumer interests in each region is also obtained, which can inform the implementation of advertising strategies on social media platforms to bolster product sales and raise awareness.

Keywords: Customer Segmentation, Machine Learning, Marketing Management, Strategic Management.

Abstrak

Penelitian ini menggunakan algoritma analisis keranjang belanja untuk mengetahui asosiasi antara kategori produk dan nama barang. Algortima ini bertujuan untuk mengidentifikasi pola pembelian pelanggan dan segmentasi pelanggan di berbagai wilayah. Sebuah perusahaan yang menjual bahan bangunan di Pontianak, Kalimantan Barat, Indonesia, belum pernah menganalisis data riwayat penjualan untuk meningkatkan strategi promosi atau memperoleh wawasan terkait segmentasi pelanggan berdasarkan wilayah. Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan perusahaan yang berlokasi di Pontianak sebagai objek studi, menggunakan data primer berupa riwayat penjualan, basis data pelanggan, dan basis data produk. Populasi penelitian terdiri atas 12.600 transaksi penjualan, dengan sampel sebanyak 3.462 transaksi yang berfokus pada merek Onda dari 2 Januari hingga 30 Desember 2023. Hasil algoritma analisis keranjang belanja akan dievaluasi berdasarkan nilai support, confidence, *dan* lift*. Dari hasil analisis, ditemukan asosiasi antara subkategori produk dan nama barang, yang memberikan wawasan untuk menentukan strategi* bundling *atau* cross-selling *berdasarkan pola pembelian konsumen, seperti kombinasi stop kran JF 11 dengan kran wastafel Y 321 C. Segmentasi pelanggan berdasarkan minat konsumen di setiap wilayah juga diperoleh, yang dapat digunakan untuk mengembangkan strategi periklanan pada platform media sosial guna meningkatkan penjualan produk dan kesadaran merek.*

Kata Kunci: *Manajemen Pemasaran; Manajemen Strategis; Pembelajaran Mesin; Segmentasi Pelanggan.*

A. Introduction

Rapid economic development is intensifying competition in business. As we can see, the emergence of new brands in the market necessitates companies to adapt by

innovating through attractive and targeted promotional strategies. To develop these promotional strategies, data is required to support the marketing strategy process and decision-making. Generally, companies do not optimally use their sales history, often archiving it or only using it to generate monthly sales reports (Efrat et al, 2020). Existing sales history data can be processed using a machine learning algorithm to leverage this information effectively.

Market Basket Analysis (MBA) aims to identify associations between product categories and items frequently purchased together, e.g., (Patwary et al, 2021; Kabasakal, 2020, Musalem et al, 2018). Association rules derived from MBA can inform the creation of discount offers and bundled product strategies aimed at boosting company sales (Gehlot & Singh, 2022). Transaction data collected for this purpose are typically evaluated using metrics such as support, confidence, and lift values to determine the strength of associations between product categories and individual items. Considering these circumstances, a consumer behavior analysis was conducted to develop bundling product strategies, perform customer segmentation analysis to understand consumer preferences, and devise advertising strategies on social media platforms to enhance sales. Such insights can help managers devise effective product bundling and cross-selling strategies (Maheswari & Sujatha, 2023; Pillai & Jolhe, 2021). While previous studies have used the MBA method to discern customer purchasing patterns and optimize product display arrangements, there has been a lack of analysis concerning customer segmentation to understand product preferences across regions.

Customer segmentation aims to divide large groups into smaller segments based on specific characteristics, preferences, and purchasing profiles, thereby enabling the development and targeting of effective marketing strategies. This facilitates enhanced product personalization and tailored services for different customer groups (Li et al, 2021). This approach provides detailed insights, allowing segmentation by region and identification of relative purchasing patterns within each segment. Such information is invaluable for marketing manager in making informed decisions (Griva et al, 2024). Based on this study, a company selling building materials located in Pontianak, West Borneo, Indonesia caters to a market segment in the region, offering various products from renowned brands such as Pralon, Onda, Nikko Steel, and others. However, the company has yet to leverage data analysis to establish connections between products and customer segmentation based on sales history.

Typically, promotional activities conducted by the company rely solely on brand promotions provided by suppliers to boost sales, without implementing product bundling strategies derived from data analysis. The company's sales can still be increased by creating promotions or additional sales programs, such as bundling or cross-selling well-performing products with those that don't sell as well, targeted at their customers. In addition, detailed customer segmentation by region has not been thoroughly conducted. So far, the marketing strategy has targeted all regions in West Kalimantan without further analyzing consumer interests in each area. This consumer segmentation aims to provide a clearer understanding of the product preferences of the company's consumers in each region, thereby supporting the creation of new, targeted promotions for marketing to customers. Therefore, we are taking the initiative to process and analyze the data, primarily focusing on support values, has been taken to address this gap.

Related Works. In (Pillai & Jolhe, 2021), the authors applied the MBA algorithm to determine the arrangement of goods, design sales promotions, and offer discounts for different customer segments, aiming to improve customer satisfaction and sales. Their research measures support, confidence, and lift to assess the associations between products, providing valuable insights for cross-selling, up-selling, and new

product integration. In (Nagaraj, 2021), the authors use MBA to study purchasing patterns for a better understanding of item association purchases. In (Neupane et al, 2023), the authors utilize historical sales data to predict the most effective product bundles through the MBA algorithm, helping the company boost its sales. Their research aims to offer numerous benefits to buyers by aligning the perfect strategy with bundled packages at discounted prices, thereby helping them save money. In (Hananto & Arizona, 2020), the authors use MBA to obtain information about product relationships purchased by customers to design promotions such as bundling or product recommendations for segmented customers. They also employ other methods, such as the K-Means algorithm and RFM patterns, to analyze customer profiles.

Contributions. All these studies have shown that the MBA algorithm can effectively identify relationships between products frequently purchased together by assessing support, confidence, and lift values. This analysis helps determine the strength of association between products or categories, aiming to develop strategies for product bundling and cross-selling. Considering the potential of the MBA method, our research proposes a new application: using MBA for customer segmentation to personalize product offerings. Specifically, this paper contributes two main aspects:

1. Introducing a novel approach using the MBA method for customer segmentation.
2. Assisting the company in developing a more targeted marketing strategy through effective bundling and cross-selling, supported by social media marketing, to boost sales.

By understanding these algorithms, companies can leverage sales data analysis to gain insights into consumer interests and customer segmentation. This information enables the development of innovative and precise marketing strategies, particularly in understanding and catering to customer preferences.

This paper consists of four main parts and is organized as follows: Section I is the introduction. In Section II, we discuss the literature review. In Section III, we proposed methodology, the MBA algorithm, and the dataset considered in this research. Section IV presents and discusses the results of product associations and customer segmentation. Finally, Section V concludes the findings based on the information.

B. Literature Review and Hypothesis Development

1. Machine Learning

   Machine learning is a part of artificial intelligence (AI) that enables computer systems to learn from historical data and improve their predictions without being explicitly programmed. Instead of following fixed rules, these systems recognize patterns in past data to make predictions on new or unseen inputs. This approach helps solve complex problems by allowing the system to adapt over time. In general, machine learning algorithms are divided into two main types: supervised learning and unsupervised learning. (Pani et al, 2021).

   As a subset of artificial intelligence (AI), machine learning enables software applications to generate increasingly accurate predictions by learning from historical data, eliminating the need for manual programming. Its adaptive nature allows models to improve performance over time as more data becomes available.

2. Market Basket Analysis

   Market Basket Analysis (MBA) is a data mining technique used to identify purchasing patterns within retail environments. It focuses on analyzing combinations of products that are bought together in a single transaction, thereby uncovering recurring trends in consumer behavior and highlighting items that are frequently co-purchased (Vijayalakshmi & Selvan, 2023).

Market Basket Analysis (MBA) is a method used to find patterns in products that are often bought together in a single transaction. It is commonly used in grocery stores and other retail businesses, like restaurants, to understand customer buying habits. By analyzing transaction data that shows which items are purchased together, businesses can discover product relationships and make better marketing or sales decisions (Olson & Lauhoff, 2019).

By identifying associations between related products, MBA enables businesses to gain insights into customer buying behavior. These insights are instrumental in shaping effective marketing strategies, particularly in the areas of consumer segmentation and cross-selling, by targeting product groupings that are often purchased together.

3. Association Rule (Support, Lift, Confidence)

In association rules, there are 3 main components: support, confidence, and lift values. Support measures the frequency of products being purchased together from the total transactions. Confidence indicates the possibility that a customer will purchase product B after purchasing product A (Suryadi & Islami, 2022). Confidence is expressed as (item set A) ⇒ (item set B), with A being the antecedent and B the consequent. Confidence indicates the likelihood of the consequent occurring given the antecedent. For consequents that appear frequently, the specific items in the antecedent are irrelevant. Lift demonstrates the strength of association rules (Efrat et al, 2020). In general, a lift value greater than one suggests that the rule is useful. It indicates that the occurrence of A increases the likelihood of B happening in the same transaction. The equations for support, confidence, and lift values are as follows (Suryadi & Islami, 2022):

$$\text{Support (A)} = \frac{\text{Transactions that contain A}}{\text{Total Transactions}} \quad (1)$$

$$\text{Confidence } (A \rightarrow B) = P(A|B) = \frac{\text{Total transactions containing A and B}}{\text{Total transactions containing A}} \quad (2)$$

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

The association rule algorithm identifies attributes that frequently appear together to determine product items consumers may purchase together (Suryadi & Islami, 2022). Association rule algorithms are widely used for market basket analysis to analyze market basket data sets from sales transactions. This technique is useful for finding relationships between objects in a particular data set (Alawadh & Barnawi, 2022).

4. Decision Tree

Decision trees are widely utilized algorithms in data mining for both classification and regression tasks. Structurally represented as hierarchical trees, they serve as predictive models that map observations to outcomes based on a series of decision rules derived from the data. The primary objective of a decision tree is to construct a model that can accurately estimate the value of a target variable by recursively partitioning the dataset into subsets based on feature values, enabling an interpretable and straightforward decision-making process (Sumathi et al, 2022).

According to (Nadiah et al, 2022), a decision tree is composed of three distinct types of nodes: the root node, branch (or internal) nodes, and leaf nodes. The root and branch nodes are associated with specific attribute names used to split the data, while the leaf nodes represent the classification outcomes or decision labels assigned to the final partitions.

A decision tree is a machine learning algorithm that assesses attributes in a tree structure for classification and regression so as to produce a decision model to predict the value of a targeted variable through a simple decision tree. A decision tree produces layer-by-layer results that are labeled as attribute names for root nodes and branch nodes.

5. Customer Segmentation

Customer segmentation is the process of grouping customers into segments based on similar characteristics. In the retail industry, customer segmentation is an important strategy to better understand and meet customer needs. Customer segmentation is done by analyzing collected customer data, including demographic data, purchase data, and behavioral data. According to, customer segmentation is the process of dividing customers into groups that have similar characteristics based on certain variables such as demographics, geography, behavior, or preferences.

Customer segmentation is a process of grouping or dividing customers into certain groups based on similar characteristics. The purpose of this customer segmentation is to better understand and meet customer needs such as aspects of demographics, geography, behavior, and customer preferences.

C. Research Methods

This study adopts an experimental quantitative approach to examine the relationship between product bundling and customer segmentation on sales history. Primary data were collected and analyzed from sales history data, customer database, and product database. The population for this study comprised 12,600 sales transactions, with a sample of 3,462 transactions specifically focusing on the Onda brand from January 2 to December 30, 2023. A company selling building materials located in Pontianak, West Borneo, Indonesia, served as the study object. Sales revenue is considered the dependent variable, whereas customer purchasing patterns and segmentation are treated as independent variables.

Initially, data were collected from sales history and integrated with customer and product databases to create a new table comprising columns for the day, month, customer name, region, product name, product category, and subcategory. Following this, the data is processed and transformed into product patterns for analysis. The purpose of data processing is to identify customer purchasing patterns by analyzing data patterns based on subcategories and product names that are most appealing to customers, as well as the combinations suitable for bundling or cross-selling strategies. This involves assessing support values to determine the best-selling subcategory and product names, and confidence and lift values to establish association rules and assess data validity. Minimum support and confidence values are determined based on the average value of all data and other factors reflecting the reality of best-selling products.

To offer insights into customer segmentation based on regional interests, an analysis is conducted using the MBA method. This analysis considers the support values of product categories and subcategories across the top 8 regions with the highest number of customers in West Borneo. The support values obtained from this analysis will be evaluated against a predetermined minimum support threshold to identify the categories or subcategories in demand within each region. The resulting data will

provide valuable information for the company to develop and execute targeted marketing strategies.

D. Result and Discussion

In this section, data processing will be performed using the MBA algorithm. Minimum values for support and confidence were set based on average values from all subcategories and product names. To determine customer interests and purchasing patterns in West Borneo, the analysis focuses on Onda product subcategories and item names using sales history data from January 2 to December 30, 2023.

This analysis resulted in several combinations of subcategories that are suitable for bundling. Table 1 shows the results of product subcategories combinations that are suitable for bundling with a minimum subcategory confidence value of 25% by reviewing the overall average value. The analysis revealed several suitable product combinations for bundling. For instance, the combination of the sink tap and basin tap has a confidence value of 41%, hand shower and jet shower (40% confidence), and angle valve and jet shower (38% confidence).

Table 1. List of product category combinations with confidence ≥ 25 percent

Sub Category 1	Sub Category 2	Confidence	Lift
Sink Tap	Basin Tap	41%	2,49
Hand Shower	Jet Shower	40%	3,11
Angle Valve	Basin Tap	38%	2,37
Angle Valve	Jet Shower	32%	2,50
Meter	Wall Tap	32%	1,90
Sink Tap	Angle Valve	26%	3,71
Angle Valve	Hand Shower	25%	2,77
Drainer	Jet Shower	25%	1.32

From the list above, further analysis was conducted to obtain a list of product associations often purchased together. Based on products with minimum support values of 1%, specific product combinations are identified to facilitate cross-selling strategies as shown in Table 2.

Table 2. List of possible combinations from the company

Item Name 1	Item Name 2	Confidence	Lift
Angle Valve JF 11	Basin Tap Y 321 C	53%	8,91
Angle Valve JF 11	Basin Tap V 688 CA	47%	5,17
Hand Shower SO 244 PC	Jet Shower S 75 WCS	39%	2,50
Hand Shower SO FRES	Jet Shower S 88 CCS	30%	4,00
Wall Tap BCS	Meter MT M-TUL (5 MTR)	27%	1,98
Angle Valve JF FLOW	Jet Shower S 75 WCS	27%	1,77
Angle Valve JF 11	Jet Shower S 75 WCS	27%	1,73
Sink Tap V 629 TLG	Basin Tap Y 321 C	23%	3,90

To determine customer segmentation based on customer interest in West Borneo, an examination was conducted across eight cities/districts with the highest customer volume at the company. Categories and subcategories of interest in each region were reviewed, and the following data were obtained, establishing a minimum support value of 50% for categories and 20% for subcategory products. Table 3 represents customer interest in the product sub-categories in each region.

Table 3. List of subcategories that customers interest in West Kalimantan

Regions	Categories	Sub Categories
Pontianak	Tap Shower Plumbing Accessories	Wall Tap, Basin Tap, Sink Tap, Angle Valve Jet Shower, Hand Shower PVC Valve, Brass Valve, PE-X Pipe Fittings, PE-X Pipe Seal Tape, Hose, Floater, Drainers, Meters
Kubu Raya	Tap Plumbing Accessories	Basin Tap, Wall Tap Brass Valve, PVC Valve Seal Tape, Meters
Singkawang	Tap Shower Plumbing Accessories	Wall Tap, Basin Tap, Sink Tap, Angle Valve Jet Shower, Hand Shower PVC Valve Hose, Floater, Drainer
Sanggau	Tap Plumbing Accessories	Basin Tap PE-X Pipe Fittings Hose
Sintang	Tap Plumbing	Wall Tap PVC Valve
Nanga Pinoh	Tap Shower Plumbing Accessories	Basin Tap, Sink Tap Jet Shower, Hand Shower Drainer, Hose, Floater PVC Valve
Sambas	Tap Shower	Basin Tap Jet Shower
Mempawah	Tap Shower Plumbing Accessories	Basin Tap Jet Shower, Hand Shower PVC Valve, Brass Valve Seal Tape, Meters

Table 3 represents the detailed information about customer interest in 8 cities/districts in West Kalimantan. Based on the provided information, customer interests in each region are delineated based on their purchasing characteristics during transactions spanning from January 2 to December 30, 2023. This data can serve as a valuable resource for the company in devising marketing strategies, enabling the determination of sales promotions tailored to each region's interests. In addition, it can serve as a reference for implementing advertising strategies on social media platforms to bolster product sales and raise awareness, focusing on products that are already in high demand or areas with growth potential. The combination of product bundling or cross-selling with customer segmentation can be executed similarly. For instance, promotions involving hand showers (S 244 PC, SO FRES) and jet showers (S 75 WCS, S 88 CCS) can be conducted in four regions (Pontianak, Singkawang, Mempawah, and Nanga Pinoh) with specified minimum order quantities and discounts.

E. Conclusion

In this study, the Market Basket Analysis (MBA) method serves as a crucial tool for uncovering customer purchase patterns and interests in specific regions for the products. These insights are invaluable for devising effective marketing strategies aimed at boosting sales. The analysis highlights combinations such as jet showers and

hand showers, and jet showers and angle valves, which are prime candidates for bundling or cross-selling. Upon closer examination of the item names, products like hand shower angle valve JF 11 with basin tap Y 321 C, hand shower SO 244 PC with jet shower S 75 WCS, and hand shower SO FRES with jet shower S 88 CCS emerge as priorities for bundling. These selections are based on support values for each subcategory and item name, as per predetermined criteria. The company also can combine the association between product subcategories for product bundling such as the combination between sink taps and basin taps, hand showers and jet showers, angle valves and basin taps, and many others. Furthermore, customer segmentation analysis provides essential information to guide the company in tailoring marketing strategies and sales promotions in each region according to their specific interests. This data also serves as a reference point when executing advertising strategies on social media platforms, facilitating the promotion of products that are already in demand or need further awareness. This new marketing strategy can attract the attention of consumers by focusing on their interests in each region, allowing for more targeted and accurate promotions based on a review of in-demand products or categories.

REFERENCES

Alawadh, K. S., & Barnawi, A. (2022). A Survey on Methods and Applications of Intelligent Market Basket Analysis Based on Association Rule. *Journal on Big Data, 4*(1). doi:10.32604/jbd.2022.021744

Efrat, A. R., Gernowo, R., & Farikhin. (2020). Consumer Purchase Patterns Based on Market Basket Analysis Using Apriori Algorithms. *The 9th International Seminar on New Paradigm and Innovation of Natural Sciences.* Central Java: IOP Publishing Ltd. doi:10.1088/1742-6596/1524/1/012109

Gehlot, A., & Singh, R. (2022). Execution of Market Basket Analysis and Recommendation Systems in Physical Retail Stores to Advance Sales Revenues. *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, 517-522. doi:10.1109/IIHC55949.2022.10060559

Griva, A., Zampou, E., Stavrou, V., Papakiriakopoulos, D., & Doudikis, G. (2024). A Two-Stage Business Analytics Approach to Perform Behavioural and Geographic Customer Segmentation Using E-Commerce Delivery Data. *Journal of Decision Systems*, 1-29. doi:10.1080/12460125.2022.2151071

Hananto, A., & Arizona, P. (2020). Customer profiling and market basket analysis using k-means algorithm and association rule mining: Evidence from Indonesia e-commerce company. In *Advances in Business, Management, and Entrepreneurship* (p. 6). CRC Press.

Iswahyudi, M. S., Budaya, I., Purwoko, P., Riswanto, A., & Anggia Ayu Lestari, E. W. (2023). *Manajemen Pemasaran : Strategi dan Praktek yang efektif.* Jambi: PT. Sonpedia Publishing Indonesia.

Kabasakal, İ. (2020). Understanding Shopping Behaviors With Category- and Brand-Level Market Basket Analysis. *Journal of IGI Global*, 26. doi:10.4018/978-1-7998-0035-4.ch012

Li, Y., Chu, X., Tian, D., Feng, J., & Mu, W. (2021). Customer Segmentation Using K-Means Clustering and The Adaptive Particle Swarm Optimization Algorithm. *Applied Soft Computing, 113*.

Maheswari, B. U., & Sujatha, R. (2023). Marketing Analytics. In B. U. Maheswari, & R. Sujatha, *Understanding Consumer Behavior Using Market Basket Analysis* (p. 29). Apple Academic Press. doi:10.1201/9781003300632-7

Musalem, A., Aburto, L., & Bosch, M. (2018). Market Basket Analysis Insights to Support Category Management. *European Journal of Marketing, 52*, 1550-1573. doi:10.1108/EJM-06-2017-0367

Nadiah, Soim, S., & Sholihin. (2022). Implementation of Decision Tree Algorithm Machine Learning in Detecting Covid-19 Virus Patients Using Public Datasets. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, 33-43.

Nagaraj, S. (2021). Market Basket Analysis: An Effective Data-Mining Technique for Anticipating Consumer Purchase Behavior. In *Big Data Analytics* (p. 13). New York: Auerbach Publications.

Neupane, A., Dhakal, B., Aryal, B., & Kandel, N. (2023). Identifying Product Bundle from Market Basket Analysis. *International Journal on Engineering Technology, 1*(1), 129–138. doi:https://doi.org/10.3126/injet.v1i1.60935

Olson, D. L., & Lauhoff, G. (2019). *Descriptive Data Mining.* Singapore: Springer.

Pani, S. K., Tripathy, S., Jandieri, G., Kundu, S., & Butt, T. A. (2021). *Applications of Machine Learning in Big-Data Analytics and Cloud Computing.* Gisturp: River Publishers.

Patwary, A. H., Eshan, M. T., Debnath, P., & Sattar, A. (2021). Market Basket Analysis Approach to Machine Learning. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-9. doi:10.1109/ICCCNT51525.2021.9580175

Pillai, R. A., & Jolhe, D. A. (2021). Market Basket Analysis: Case Study of a Supermarket. *Lecture Notes in Mechanical Engineering ((LNME))* (pp. 727–734). Singapore: Springer. doi:10.1007/978-981-15-3639-7_87

Sumathi, S., Rajappa, S., Kumar, L. A., & Paneerselvam, S. (2022). *Machine Learning for Decision Sciences with Case Studies in Python.* Boca Raton: CRC Press.

Suryadi, A., & Islami, M. C. (2022). Analysis of Data Mining at Supermarket X in Surabaya Using Market Basket Analysis to Determine Consumer Buying Patterns. *3rd International Conference Eco-Innovation in Science, Engineering, and Tech-nology* (pp. 28-32). NST Proceedings. doi:10.11594/nstp.2022.2705

Vijayalakshmi, V., & Selvan, S. (2023). *Advanced Data Mining.* India: SK Research Group of Companies.