The 5th Asia-Pacific Management Research Conference (APMRC) 2024
"Adapting Sustainability Strategies in Business and Management"

Understanding Transactional Data Patterns from Micro Business using Machine Learning Algorithms

Kenny[1*]; Jimmy Tjen[2]
[1] Department of Digital Business, Faculty of Information Technology;
[2]Department of Informatics, Faculty of Information Technology
[1,2]Universitas Widya Dharma Pontianak
[1,2]Pontianak, Indonesia, 78117
[1]21430106_kenny@widyadharma.ac.id; [2]jimmy.tjen@mathmods.eu

Corresponding Author

Abstract— **This study aims to predict sales trends using the Auto Regression (AR) method, supported by the Market Basket Analysis (MBA) in predicting hidden patterns in sales transaction data. The subject of this study is Logic Store, a retail store selling computer and cellphone accessories located in Pontianak, Indonesia. This quantitative research utilizes primary data from the Logic Store sales database. The research data for the AR method consists of 158 data points, representing sales revenue over 3 years, from January 2021 to February 2024, with a weekly time interval. The research data for the MBA method includes 2391 transactions, which were then filtered down to 145 transactions. The data used involves transactions containing more than one type of product. The results show that sales trends generally increased from 2021 to 2023. Sales trend predictions indicate that sales will rise at the beginning of each month, especially during the transition between September, October, and November. The results also suggest potential bundling promotions, such as earphone-earphone bag, True Wireless Stereo (TWS)-earphone, mouse-mousepad, and charger adapter-cable. Future research should consider combining other methods, such as decision trees, moving average, and so on with the methods used in this paper.**

Keywords—Family Business; Machine Learning; Business Intelligence; Marketing Management

I. INTRODUCTION

Competition in the business world is intensifying. Companies are trying to control market share in each region and emerge as leaders. To compete effectively, they must adopt and follow developments in the business world for informed decision-making. Although companies employ various strategies to increase sales, they often do not fully utilize the data they have collected, which can be processed using machine learning. In the business world, machine learning aims to extract valuable insights from big data to enhance decision-making. It serves various purposes, such as accurately predicting future trends, uncovering hidden patterns, segmenting customers, analyzing consumer behavior, optimizing business processes, and minimizing risks, all to achieve success [1].

The Auto Regression (AR) method in machine learning is used to predict time series or time behavior patterns. AR is a subset of the Auto-Regressive Integrated Moving Average (ARIMA) model, which typically consists of pure autoregressive, mixed autoregression and moving average, or pure moving average components [2]. AR is used in forecasting the capital market to predict stock data, for example [3] which combines AR and Moving Average (MA). In addition to predicting stock data, AR is also commonly used to forecast inventory items, such as [4]. In this research, AR model will be used to predict future sales trends based on past data, with Market Basket Analysis (MBA) as a supporting method. MBA identifies relationships between two or more items and uncovers hidden patterns among items frequently purchased together by customers in sales transaction data. These relationships can be used for product structuring, bundling promotions, or product recommendations, for example [5], [6], [7]. Given these conditions, this approach allows for bundling promotions when the predicted trend indicates a decline in sales at a certain time, helping to maintain stable sales. Additionally, it enables the preparation of stock when sales are predicted to increase.

Related works. Several studies have applied these methods in the business world. For example: [8], [9], [10], [11]. In [8], they used data obtained from Instacart to predict the most effective product bundles with the aim that the company can increase its sales potential. They are trying to make more reliable predictions and improve the decision-making process. In [9], they used the Autoregressive Integrated Moving Average (ARIMA) model to predict motorcycle part sales. Forecasting is employed to facilitate decision-making in determining the quantity of goods to procure. In [10], they use data mining techniques, specifically Market Basket Analysis (MBA) with the Apriori algorithm to generate information that supports marketing strategies. In [11], they predicted the sales volume of vegetable and fruit products based on historical sales using the Autoregressive Integrated Moving Average (ARIMA) model, which was enhanced by a recurrent Long Short-Term Memory (LSTM) neural network with parameters tuned and optimized using the Adam algorithm. Previous studies have commonly employed AR for data prediction and MBA for identifying consumer purchasing patterns. However, the combination of these two methods—predicting sales trends supported by MBA — remains underexplored. Similarly, applications in the realm of computer and cellphone accessories retail stores are still uncommon.

Contributions. In particular, previous studies used MBA to obtain information about effective product bundling for marketing strategies, while autoregression was used to predict sales, either from sales volume or from the revenue received by the company. Given this condition, the researcher attempted to combine the two models because the prediction of sales trends using AR supported by MBA still does not exist. Objects in the form of computer and cellphone accessories in retail stores are also still very rarely used. Researchers are exploring adjustments to apply MBA to predictions generated by the AR model. For instance, during periods of predicted sales decline, cross-selling or bundling strategies can be implemented based on MBA findings. Otherwise, during periods of predicted sales increase, stocking strategies should be adjusted to meet demand.

This paper aims to enhance efficiency and address challenges at Logic Store, a retail business specializing in computer and cell phone accessories, including mouse, keyboards, and earphones. Previously, Logic Store never utilized the data that had been collected over the years. The data collected was only used to see profit margins, determine rough product prices and sales growth from month to month and year to year as well as a recapitulation of product stocks that were still available and those that had run out. These data can be used to obtain more useful information to make the right decisions in developing the business. Sales at Logic Store fluctuate unpredictably. Additionally, there is a disparity in product demand, ranging from high sales to items that remain unsold. This imbalance leads to inventory accumulation and shortages of high-demand items, resulting in potential losses due to product deterioration. To mitigate these challenges, Logic Store will leverage previously unexplored data analysis techniques. The objective is to gain novel insights and facilitate informed decision-making in sales strategy.

The paper is structured as follows: Section I introduces the study. Section II details the data collection methods, research objectives, and data analysis techniques. Section III presents the results and discusses the research findings, including figures, tables, and pertinent information. Section IV summarized the implications of the findings for theory, practice, and policy in the field of business and management, along with potential directions for future research.

II. RESEARCH METHODS

This paper employs Auto Regression (AR) and Market Basket Analysis (MBA) methods to support decision-making. The research utilizes documentation techniques to gather primary data from the sales database of Logic Store, a retail outlet specializing in computer and cellphone accessories located in Pontianak, West Kalimantan, Indonesia. The data is sourced from Logic Store, which as a family business, facilitates easier data acquisition. The familiarity with the store's conditions, including challenges faced, implemented strategies, and various operational limitations, enhances this research. The research is quantitative, with products sold and sales revenue as dependent variables, and customer purchasing trends and sales patterns as independent variables. Initially, data processing involves using spreadsheets to eliminate redundant information. Subsequently, data processing software aids in identifying patterns to expedite and simplify the analysis process.

In this paper, the AR model was constructed, by considering a dataset that comprises sales data spanning 3 years and 2 months, from January 2021 to February 2024, totaling 158 data points. This timeframe was selected because Logic Store began collecting data in 2021, with the latest available data extending to

February 2024. The training dataset constitutes 80% of the total data to enhance result accuracy. An AR model of order 1 was used to indicate that current sales are influenced by sales from the previous week, linking sales variables in period 't' (past data) to current or future sales in period 't+1'. The use of weekly data intervals in this paper aims to enhance dataset processing, thereby achieving higher accuracy in the results. Data accuracy is evaluated using Mean Absolute Percentage Error (MAPE) and Normalized Root Mean Square Error (NRMSE), selected for their capability to accurately assess data validity. Results are deemed accurate if MAPE is less than 15%, and higher NRMSE values indicate better accuracy [12]. The formulas of AR models, MAPE, and NRMSE are as follows [13], [14], [15]:

$$x_t = \sum_{i=1}^{n} (wi * x_{t-i}) + b \tag{1}$$

While $x_{t-i}, \dots, x_{t-n}$ Represent the n values used to forecast $x_t, and\ b$ represents the bias, the weights wi, which multiply each of the n lagged values $xt - i$, are also known as the AR coefficients.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - f_t}{x_t}\right| \times 100 \tag{2}$$

While n represents the number of data points, t denotes the period, x_t denotes the actual value, and f_t Denotes the forecasted value.

$$NRMSE = \frac{RMSE}{(ymax - ymin)} \tag{3}$$

While y_{max} represents the maximum value of the dependent variable, y_{min} Represents the minimum value of the dependent variable. The RMSE value is calculated using the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4}$$

While $\hat{y}_i$ represents the predicted values, y_i Represents the observed values, and n is the number of data points.

Trend predictions were derived from the output of data processing software are corroborated by MBA. The MBA method utilizes data from December 2023 to February 2024, totaling 2391 transactions, which were subsequently filtered down to 145 transactions. The dataset includes transactions involving multiple types of products. Product categories considered include earphones, True Wireless Stereo (TWS), earphone bags, mouse, mousepads, adapter chargers, cables, finger sleeves, keyboards, and Universal Serial Bus (USB) hubs. These products were selected because they are among the best-selling items at Logic Store.

MBA identifies product relationships based on support, confidence, and lift values. Support indicates how frequently an item appears in transactions, while confidence measures the frequency of products A and B being purchased together. A higher confidence value indicates that the two products are frequently purchased together, suggesting they are suitable for pairing in promotions or bundles. Lift serves as a validity test, indicating the strength of the relationship between product A and product B. A lift value greater than 1 suggests a positive relationship, while a lift value less than 1 indicates that the products are likely purchased independently of each other, occurring by chance [16]. The three elements are obtained from the following formula [17], [18]:

$$\text{Support(A)} = \frac{The\ number\ of\ transaction\ contain\ A}{Total\ transaction\ amount} \tag{5}$$

$$\text{Support(A, B)} = \frac{The\ number\ of\ transaction\ contain\ A\ and\ B}{Total\ transaction\ amount} \tag{6}$$

$$\text{Confidence}\ (\text{A} \rightarrow \text{B}) = \ \text{P}(\text{A}|\text{B}) = \frac{\text{Support}(\text{A}, \text{B})}{\text{Support}(\text{A})} \tag{7}$$

$$Lift(\text{A} \rightarrow \text{B}) = \frac{Confidence(\text{A} \rightarrow \text{B})}{Support\ (\text{B})} \tag{8}$$

Calculations using the formulas for the three elements are performed using software with prepared source code. Transaction data will be transformed into a format readable by the software. Initially, the support for single-item sets is calculated. Subsequently, items below the minimum support are filtered out, and the calculation of confidence and lift continues. Combinations of two similar products, such as between two keyboards, are disregarded even if they meet the confidence and lift criteria, as combining similar products is deemed invalid.

III. RESULT AND DISCUSSION

Section Result. In this section, data processing will be carried out using AR and MBA methods through software. Consider the AR model, figure 1 demonstrates an accuracy of 81.87% and a MAPE of 12.31%, indicating its reliability and precision. The graph illustrates a consistent upward trend from 2021 to 2023, with typical fluctuations observed in the business environment. Despite a decline towards the end of the graph, there remains potential for subsequent increases, as indicated by a sales upturn towards the end of the original data sequence. As a predictive tool, notable increases are observed around the 10th, 15th, and 20th data sequences.

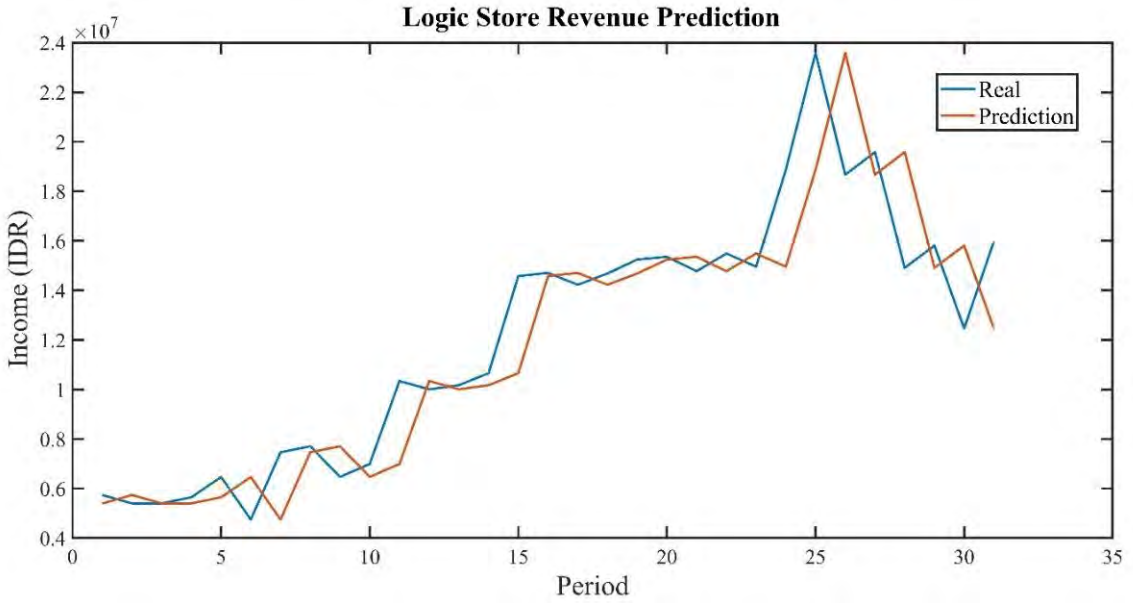

Figure 1. Prediction Income with AR-1

With regards to MBA, the first step involves determining the minimum support. In this study, the minimum support is set at 11%, calculated as half of the average monthly support. The results indicate that out of 10 types of products, one product failed to meet the minimum support in the December 2023 transactions (USB hub), and in February 2024 (finger sleeve), while two product types fell below the minimum support in the January 2024 transactions (keyboard and finger sleeve). Products that do not meet the minimum support for a given month are excluded from the next stage, which involves determining confidence. The minimum confidence level is set at 25% based on the average confidence obtained.

Table 1. Support of 1-itemset for each product type

Number	Product Type	December	January	February
1.	Earphone	30,77%	22,00%	23,21%
2.	True Wireless Stereo (TWS)	28,21%	22,00%	21,43%
3.	Earphone Bag	17,95%	12,00%	19,64%
4.	Mouse	35,90%	44,00%	32,14%
5.	Mousepad	25,64%	52,00%	39,29%
6.	Adaptor	12,82%	12,00%	17,86%
7.	Cable	30,77%	24,00%	28,57%
8.	Finger Sleeve	12,82%	0,00%	3,57%
9.	Keyboard	17,95%	10,00%	17,86%
10.	USB Hub	10,26%	12,00%	19,64%

Table 2. Confidence and Lift of 2-itemset

Product Type 1	Product Type 2	*Confidence*	*Lift*	Description
TWS	*Earphone*	45,45%	1,48	December 2023
TWS	*Keyboard*	28,57	1,01	
TWS	*Earphone Bag*	42,86%	1,51	
Earphone	*Earphone Bag*	71,42%	2,32	
Mousepad	*Keyboard*	42,86%	1,67	
Earphone	Finger Sleeve	60%	1,95	
Adaptor	Cable	100%	3,25	
Cable	Finger Sleeve	40%	1,3	
Mouse	*Mousepad*	70%	1,95	
TWS	*Earphone*	36,36%	1,65	January 2024
Earphone	*Earphone Bag*	50%	2,27	
Cable	TWS	27,27%	1,13	
Cable	*Earphone Bag*	33,33%	1,38	
Mouse	*Mousepad*	72,72%	1,40	
TWS	Adaptor	33,33%	1,51	
Cable	Adaptor	66,67%	2,78	
Mousepad	USB Hub	66,67%	1,28	
TWS	*Earphone*	58,33%	4,3	February 2024
Mouse	*Mousepad*	77,77%	1,97	
Cable	*Earphone bag*	25%	1,27	
Adaptor	Cable	100%	3,5	
Mouse	USB Hub	36,36%	1,13	
Keyboard	USB Hub	30%	1,52	

Section Discussion. Considering the result of the AR model, Figure 1 shows that the business has grown year on year. Further analysis reveals spikes occurring around the transitions from August to September, the beginning of October, and the transition from October to November. Typically, these increases happen early in the month, indicating a need for restocking by month-end if any products are depleted. To mitigate potential declines in sales during these periods, Logic Store can implement promotions such as bundling or cross-selling in the middle and at the end of each month, supported by

MBA. This strategic approach aims to stabilize monthly sales. Furthermore, bundling products helps evenly distribute sales, thereby reducing stock accumulation or shortages.

With regards to the result of MBA, the results indicate that there are 15 product combinations suitable for bundling, each with a confidence value exceeding 25% and a lift value greater than 1, but the result chosen is the best result of all seen from the highest average confidence value. The product combinations with the highest confidence values are earphone-earphone bag, TWS-earphone, mouse-mousepad, and cable-adaptor, boasting average confidence levels of 46.47%, 46.71%, 73.50%, and 89%, respectively. Based on this information, Logic Store plans to implement bundling or cross-selling promotions.

From the consumer's perspective, these four combinations are highly recommended and logical for bundling. Some of these products are already sold as single packages in the market. For example, customers purchasing a mouse often require a mousepad, and those buying a charger cable may also need an adapter. Similarly, buyers of earphones are likely to purchase earphone bags to protect their earphones. Besides these three combinations, other combinations listed in the table remain viable due to their lift value exceeding 1.

From the seller's perspective, these four combinations also make sense. For example, cheap mousepads that don't sell well can be offered for free to customers who buy a mouse. The cost of more expensive mouse can be offset by the savings from these very cheap mousepads. This strategy can also be applied to earphones and earphone bags, allowing the seller to still make a profit while ensuring faster stock rotation. For more expensive items that cannot be given away for free, a discount or special offer can be provided if both items are purchased together. For instance, purchasing TWS and earphones together could result in a 50% discount on the earphones. These promotions will be advertised through social media platforms (Instagram, TikTok) or marketplaces (Shopee, Tokopedia) to enhance customer loyalty.

IV. CONCLUSION

This research focuses on predicting future sales trends using AR techniques and employing MBA techniques for bundling strategies. Based on AR analysis, the trend prediction suggests that sales tend to experience consistent growth, indicating continuous expansion for Logic Store. Sales volume notably increases at the beginning of September, October, and November, necessitating preparation at the start of these months. Logic Store can customize promotions based on predicted sales trends derived from the graphs. MBA analysis indicates that the most recommended products for bundling are earphone-earphone bags, cable-adapters, and mouse-mousepads, supported by the highest confidence levels observed in sales data over three consecutive months. Based on these findings, Logic Store will implement bundling and cross-selling promotions to stimulate consumer interest and maintain sales stability.

This paper has several limitations. The dataset is constrained to a few years of data because Logic Store has been operational for only a short period. Increasing the dataset size would enhance the accuracy of the information derived. Additionally, accuracy can be assessed using alternative error analyses beyond NRMSE and MAPE. The methods employed are also limited and could be further developed in future research. For instance, this paper exclusively utilizes autoregression which could be expanded by incorporating the Autoregressive Integrated Moving Average (ARIMA) method. Future research should strive for a more diverse array of methodologies, such as integrating the MBA method with ARIMA and coupling these with decision tree learning for customer segmentation analysis. Segmentation analysis using decision trees allows researchers to obtain more detailed insights into customer behavior from the collected data. The objective is to enhance store performance through effective marketing strategy implementation.

V. REFERENCES

[1] J. P. Bharadiya, "Machine learning and AI in business intelligence: Trends and opportunities," *International Journal of Computer (IJC),* vol. 48, no. 1, pp. 123-134, 2023.

[2] J. Kaur, K. S. Parmar and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application review," *Environmental Science and Pollution Research,* vol. 30, no. 8, pp. 19617-19641, 2023.

[3] D. Xiao and J. Su, "Research on stock price time series prediction based on deep learning and autoregressive integrated moving average," *Scientific Programming,* vol. 2022, 2022.

[4] D. A. Verano, H. Husnawati and E. Ermatita, "Implementation of autoregressive integrated moving average model to forecast raw material stock in the digital printing industry," *Journal of Information Technology and Computer Science,* vol. 5, no. 1, pp. 13-22, 2020.

[5] R. Dio, A. A. Dermawan and D. A. Putera, "Application of Market Basket Analysis on Beauty Clinic to Increasing Customer's Buying Decision," *Sinkron: jurnal dan penelitian teknik informatika,* vol. 8, no. 3, pp. 1348-1356, 2023.

[6] Y. Kurnia, Y. Isharianto, Y. C. Giap and A. Hermawan, "Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm," in *Journal of Physics: Conference Series*, 2019.

[7] Z. U. Rizqi, "Implementation of Association Rule-Market Basket Analysis in Determining Product Bundling Strategy: Case Study of Retail Businesses in Indonesia," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2019.

[8] A. Neupane, B. Dhakal, B. Aryal and N. Kandel, "Identifying Product Bundle from Market Basket Analysis," *International Journal on Engineering Technology,* vol. 1, no. 1, pp. 129-138, 2023.

[9] M. Maslim and K. Arinanda, "Motorcycle parts sales forecasting using auto-Regressive Integrated moving average model," *International Journal of Computer Theory and Engineering,* vol. 12, no. 1, pp. 28-31, 2020.

[10] M. Qisman, R. Rosadi and A. Abdullah, "Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)," in *Journal of Physics: Conference Series*, Bali, 2021.

[11] Y. Wang and C. Cui, "Forecasting superstore fruit and vegetable sales based on ARIMA-LSTM," *Journal of Education, Humanities and Social Sciences,* vol. 25, pp. 105-110, 2024.

[12] A. M. Khan and M. Osinska, "Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India," *Expert Systems with Applications,* vol. 212, p. 118840, 2023.

[13] L. Ranaldi, M. Gerardi and F. Fallucchi, "Crypto net: using auto-regressive multi-layer artificial neural networks to predict financial time series," *Information,* vol. 13, no. 11, p. 524, 2022.

[14] F. Liantoni and A. Agusti, "Forecasting bitcoin using double exponential smoothing method based on mean absolute percentage error," *JOIV: International Journal on Informatics Visualization,* vol. 4, no. 2, pp. 91-95, 2020.

[15] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions,* vol. 2022, pp. 1-10, 2022.

[16] S. Karnila, A. Rizkyandi, R. Kurniawan and N. Nurjoko, "MARKET BASKET ANALYSIS ON TRANSACTION DATA USING THE APRIORI ALGORITHM," *Jurnal TAM (Technology Acceptance Model),* vol. 13, no. 1, pp. 34-43, 2022.

[17] A. Efrat, R. Gernowo and others, "Consumer purchase patterns based on market basket analysis using apriori algorithms," in *Journal of Physics: Conference Series*, 2020.

[18] A. Zamil, A. Al Adwan and T. Vasista, "Enhancing customer loyalty with market basket analysis using innovative methods: a python implementation approach," *International Journal of Innovation, Creativity and Change,* vol. 14, no. 2, pp. 1351-1368, 2020.