

Implementation of Market Basket Analysis based on Overall Variability of Association Rule (OCVR) on Product Marketing Strategy

Alfiqra¹, A U Khasanah¹

¹ Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Islam Indonesia, Jalan Kaliurang km. 14,5 Yogyakarta 55584, Indonesia

E-mail: 14522289@students.uui.ac.id, annisa.uswatun@uui.ac.id

Abstract. Marketing strategy is an important thing that must be developed by retail. A method that can be used to develop a marketing strategy based on customer buying pattern is Association Rule (AR). AR is the process of finding association relationships between products that occur in one transaction. The application of AR to analyze customer buying patterns is referred to as Market Basket Analysis (MBA). Rule obtained from ARMBA is sometimes not enough to provide an analysis when the variability of customer buying pattern is high. Overall Variability of Association Rule (OCVR) is an indicator that focuses on analyzing market basket which assumes high variability in customer behavior in buying products. This study used customer transaction data of a retail in Yogyakarta. The data consisted of 57784 transactions in a month involving 41248 items. This study produced rules for each period (weeks), then the rules were used for further analysis using OCVR. 59 rules produced on the 1st period, 48 rules on the 2nd period, 54 rules on the 3rd period, and 58 rule on the 4th period. From the rules obtained there were 17 rules have OCVR value smaller than 30%, thus these rules can be used to make marketing strategies. Product bundling and shelves product arrangement based on obtained rules were proposed as marketing strategies to promote product sales.

1. Introduction

Retail businesses in Yogyakarta are growing rapidly. Based on data from the Yogyakarta Central Bureau of Statistics, the growth rate of Gross Regional Domestic Product (GRDP) in the third quarter of 2017 for the business sector in accommodation, food and beverage providers including retail is 62% [1]. These business sectors are the three largest business sectors that influence economic growth in Yogyakarta. This increasing number forced businesses to find suitable marketing strategies to promote sales and survive in the competition. One marketing strategy that can attract customers is sales promotion. Sales promotion has a big influence on the desire of customers to buy product suddenly without any plan or usually called as impulse buying [2]. Sales promotions are usually simply carried out to introduce a new product or to promote less salable products.

Another way that can be done to design an attractive sales promotion is by analyzing consumer buying patterns. Strategy for decision making and understanding consumer spending behavior is a



challenge for an organization in maintaining its position in market competition [3]. The objective of this analysis is to find out what kind of products that usually bought by customers in one transaction time and it is usually called as Market Basket Analysis (MBA). The purpose of market basket analysis is to get a selling strategy with up-selling and cross-selling [4]. One method that commonly used to conduct this kind of analysis is Association Rule (AR). AR is a very popular data mining techniques to find important rule of association relationship between product in transaction data. AR was firstly proposed for marketing, but now this method has widely used in other fields, such as bioinformatics, nuclear science, pharmacoepidemiology, and geophysics [5]. In MBA, AR can be expressed as “A customer who buys product X1 and X2 will also buy product Y with probability c%” [6]. To implement AR for MBA, transaction data in a certain retail or supermarket is needed and sometimes it involves very big transaction data.

Supermarket X is a popular retail business in Yogyakarta, which already has several branches. This supermarket provides daily needs, fashion, stationery, toys, cosmetics, etc. The supermarket has already made some promotion strategies to attract customers, but the management still did not use a specific method to design the promotion strategy. Supermarket X have already implemented information system to manage their business process, including to store their transaction data. Unfortunately, the data has not been optimally analyzed and utilized. In this case, AR is very useful to be implemented in analyzing the transaction data and help the management in designing promotion strategy which is suitable for the customer buying pattern.

One branch of supermarket X have more than 1000 transaction in a day and almost 60000 transactions in a month, Figure 1 represents the transaction data of Supermarket X in March (only for one branch). As shown in Figure 1, the supermarket has fluctuated transaction number. With this kind of condition, sometimes it is difficult to find out the pattern and to make suitable promotion strategy. In this study, the customer buying pattern of the supermarket will be revealed using Overall Variability of Association Rule (OCVR).

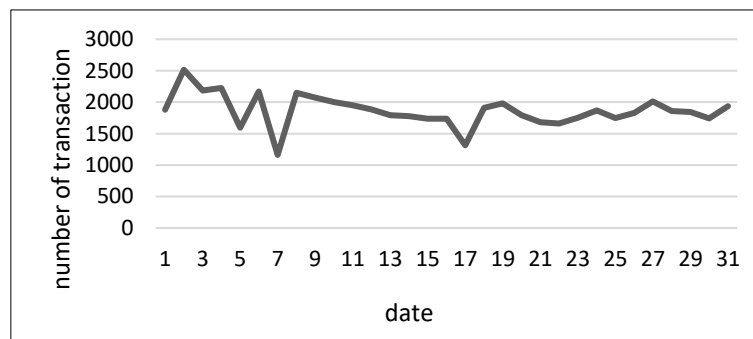


Figure 1. Number of transactions in March

OCVR is a new indicator in the AR-MBA application that was first proposed by Papavasileiou and Tsadiras in 2011. This indicator is applied in an MBA with the assumption that customers have high variability of shopping habit. Indicator of variability is related to the changes in customer buying habits in a certain period. This analysis can increase the efficiency of the rules generated from AR to create marketing strategies to promote sales [7], the main purpose of this study is to find the association rules of the product based on OCVR and to create marketing strategy which is suitable with the obtained rule.

2. Literature Review

2.1. Association Rule Mining

Association rule is one of the popular data mining techniques that can be used to find associative relationship from a group of data set. Association rules are formed by analyzing frequent patterns and by using parameters of support and confidence. Tan et al. [8] explain that support determines how often

a rule applies to a given data set. A low support rule is uninteresting from a business perspective and sometimes it will be eliminated. While confidence determines how frequently items in Y appear in transactions that contain X, the higher the confidence the more likely it is for Y to present in transactions that contain X.

$$\text{Support, } s(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

$$\text{Confidence, } c(X \Rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

One parameter that also used to determine important rule in AR is Lift Ratio. Lift Ratio is a parameter that is used to see whether the rule gained from AR are strong or not. Lift Ratio measures the possibility of X and Y occurring together divided by the possibility of X and Y occurring if they are independent events. The formula of ratio is as followed.

$$\text{Lift Ratio} = \frac{\text{Support}(X \cap Y)}{\text{Support}(X) \cdot \text{Support}(Y)} \quad (3)$$

In data mining, association rules are useful for analyzing and predicting customer behavior and play an important role in the analysis of shopping basket data, product clustering, catalog design, and store plan layout. Not only on product analysis, Association Rule Mining also used for detect operational problems in heating, ventilation and air conditioning (HVAC) systems of buildings [9]. Association rule also used in strategic management fields to determinate of effective management strategies [10]. Several algorithms are applied to the association rule technique. Based on the research conducted by Dhanalakahmi and Porkodi [11], several algorithms that have been used by researchers include Apriori, Eclat, and FP-growth Algorithms. Sandhu et al. [12] found a new method in developing the association rule algorithm based on the assumption of profit and quantity.

The apriori algorithm was used in this study to produce a set of association rules from the database. Apriori algorithm is an algorithm that is useful for finding frequent itemset for the Boolean association rule. The name of the apriori algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset mining search [13]. The original Apriori algorithm is dependent on the number of items, the number of transactions, and the market basket size [14]. The study proposed an efficient approach based on weighting factors and utilities for effective data mining from association rules with high utility. Kaur and Kang [15] in his study explained that some algorithms in existing AR only worked on static data and cannot capture real-time changes on data. The algorithm proposed in this study did not only work on static data but also captured changes in dataset. Chen et al. [16] applied MBA in multiple-store environment. The purposed method was proved to be more computationally efficient compared with the traditional method when implemented in supermarket which have more than one store, varying in sizes, over time product changes and when using longer period dataset.

One of the goals of MBA is to determine and prediction customer's behaviour based on expenditure patters from previous clients [17]. MBA is not only applied in a supermarket but also in other fields of study. There are to compare with other methods to propose a product network analysis, a network-based analysis to analyze a network leveled relation between products [18]. MBA also used for product inventory prediction and combine with other method, there is Artificial Neural Network (ANN) Back propagation [19]. MBA could apply in hospital field. Market Basket Analysis could increase revenue by enabling hotels to determine the most attractive additional products and services (beyond the room type) to offer new and repeat hotel guests [20].

2.2. Overall Variability of Association Rule (OCVR)

The values of support, confidence, and lift ratio on the results of AR often vary in each period due to the uncertain customer buying habits. OCVR is the variability index of parameter changes in AR. The calculation of this variability index is based on the concept of standard deviation in statistical analysis. *Variability Index (CV)* is calculated using formulation as followed,

$$CV = \frac{S}{\bar{X}} \quad (4)$$

Where S is standard deviation and \bar{X} is average. This variability index is used to calculate Index Variability Lift (CVL) and Index Variability Confidence (CVC). Thus, the analysis of Overall Variability of Association Rule (OCVR) comes from the use of CVL and CVC and can be formulated as followed

$$OCVR = \frac{CVL+CVC}{2} \quad (5)$$

The results of the OCVR analysis show the degree of variation in AR from one period to the next period. These results can be considered in determining what rules are important in MBA. Also, the rule with a high OCVR value, which indicates high changes in the customer buying habit in every period, can be further. Then, the decision maker can make suitable marketing strategies based on those results [7].

3. Research Method

This study was implemented in a supermarket in Yogyakarta and the data used was one-month transaction data. The data consisted of 57784 transactions in a month involving 41248 items. AR will be implemented for each period within a week, then there were 4 periods in this study. Table 1 shows transaction data for each period.

Table 1. Number of transaction and item for each period

Period	Number of transaction	Number of item
1	13741	17935
2	15370	17141
3	13836	15925
4	14837	17355

Apriori algorithm was used to implemented AR and software R was used to run the algorithm. This study uses the KDD (Knowledge Discovery in Database) process which includes sequential selection, preprocessing, transformation, data mining, and interpretation/evaluation [21]. Before the data was analyzed, pre-processing data should be conducted to prepare the data. The pre-processing data consisted of cleaning, reduction, and integration. In data cleaning, noisy and incomplete data was identified and erased. Data reduction conducted to reduce some unnecessary variables such as date transaction and item code. While data integration conducted to combine several items such as 1 litre and 2 litre cooking oil. Figure 2 shows the research flowchart.

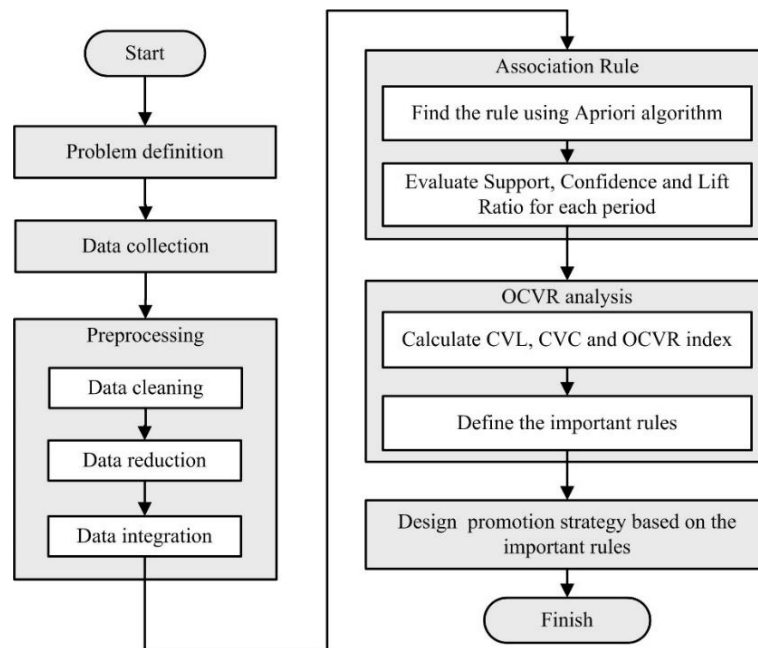


Figure 2. Research flow chart

4. Result and Discussion

Before implanting AR on the transaction data, the most frequent item for each period was analyzed as initial information as shown in Table 2. As shown in the charts the most frequent item for all period is similar. It contains several items such as instant noodles, snacks, milk, soap, etc.

Table 2. The 5 most frequent item

1 st period		2 nd period		3 rd period		4 th period	
Item	Number	Item	Number	Item	Number	Item	Number
Noodle X	1100	Noodle X	1070	Wafer	985	Noodle X	1114
Wafer	833	Wafer	1006	Noodle X	971	Wafer	1091
Milk	745	Milk	791	Milk	693	Milk	741
Noodle Y	465	Noodle Y	437	Biscuit	433	Biscuit	488
Soap	367	Tea in bottle	435	Noodle Y	363	Tea in bottle	381

4.1. Association Rule

Parameter setting for Support and Confidence in this study was done by trying and error. The important rules were determined after the best parameters were identified. The result of the parameter setting is presented in Table 3.

Table 3. Parameter setting result

Trial	Min. support	Min. confidence	Result
1	0.1	0.1	No rule obtained
2	0.01	0.01	Lots of rules were obtained but cannot be read Example: {} => {INDOMIE}
3	0.001	0.2	Lots of rules were obtained
4	0.001	0.3	Only few numbers of rules were obtained

Based on those results, the minimum threshold for support and confidence that will be used to implement AR using Apriori Algorithm were 0.001 and 0.2. Then, the data for each period that have been preprocessed were put on the software to gain the rules. Table 4 shows the rules obtained for each period.

Table 4. Rules for 1st period

No	Rule	Support	Confidence	LR	Count
1	{Liquid soap} → {Bar Soap}	0.00109	0.62500	74.68	15
2	{Noodle Y} → {Noodle X}	0.00116	0.57143	7.14	16
3	{Milk cleanser} → {Face tonic}	0.00437	0.52174	60.25	60
4	{Noodle Z} → {Noodle X}	0.00138	0.51351	6.41	19
5	{Face tonic} → {Milk cleanser}	0.00437	0.50420	60.25	60
...	→
59	{Margarine} → {Noodle X}	0.00167	0.20000	2.49836	23

For the first rule, if a customer buys liquid soap, then the possibility of the customer also buying bar soap is 62.5% (confidence). 15 transactions contain the transaction from all data in period 1 (support 0.109%). Also, this rule is valid, as seen from the lift value > 1, which is 74.67.

Table 5. Rule for 2nd period

No	Rule	Support	Confidence	LR	Count
1	{Noodle Y} → {Noodle X}	0.00104	0.59259	8.51	16
2	{Face tonic} → {Milk cleanser}	0.00267	0.53247	87.06	41
3	{Noodle A} → {Noodle X}	0.00111	0.50000	7.18	17
4	{Chili sauce} → {Noodle X}	0.00124	0.47500	6.82	19
5	{Milk cleanser} → {Face tonic}	0.00267	0.43617	87.06	41
...	→
48	{Biscuit} → {Wafer}	0.00156	0.20000	3.06	24

Table 6. Rule for 3rd period

No	Rule	Support	Confidence	LR	Count
1	{Noodle B} → {Noodle X}	0.00123	0.53125	7.57	17
2	{Face tonic} → {Milk cleanser}	0.00325	0.46392	61.13	45
3	{Biscuit} → {Wafer}	0.00116	0.45714	6.42	16
4	{Liquid soap} → {Bar Soap}	0.00101	0.45161	71.01	14
5	{Milk cleanser} → {Face tonic}	0.00325	0.42857	61.13	45
...	→
54	{Dish wash} → {Noodle X}	0.00202	0.20000	2.85	28

Table 7. Rule for 4th period

No	Rule	Support	Confidence	LR	Count
1	{Face tonic} → {Milk cleanser}	0.003774	0.56000	72.88	56
2	{Noodle C} → {Noodle X}	0.001348	0.55556	7.40	20
3	{Liquid soap} → {Bar Soap}	0.001146	0.53125	85.68	17
4	{Cat food A} → {Cat food B}	0.001146	0.50000	176.63	17
5	{Milk cleanser} → {Face tonic}	0.003774	0.49123	72.88	56
...	→
58	{Soy sauce} → {Noodle X}	0.001078	0.20779	2.77	16

It can be seen from these results that the rules obtained from each period were quite different. To design a marketing strategy that can be implemented for a month, further analysis was needed. OCVR analysis was applied for the next step to know any important rules that appeared in each period.

4.2. Overall Variability of Association Rule (OCVR) Analysis

To get OCVR values, the rules formed in each period were needed. The first step is to combine rules that always appear in each period, while rules that did not appear in each period were not used. 17 rules always appeared in each period. The confidence and lift ratio value from those rules were used to calculate *Index Variability Lift* (CVL) and *Index Variability Confidence* (CVC) to find OCVR value. Table 8 shows the calculation result of OCVR. OCVR values that show low variability according to Papavasileiou & Tsadiras (2011) [5] are 1% to 30%. OCVR value greater than 30% indicates that the rule is very vulnerable to changes and cannot be used at any time. The result shows that OCVR value or all rules were $1\% < \text{OCVR} < 30\%$, it means all rules can be used to make marketing strategies.

Table 8. Calculation of Overall Variability of Association Rule (OCVR)

No	Rules	CVC %	CVL %	OCVR %
1	{Noodle E} → {Noodle X}	0.88	7.13	4.01
2	{Green bean drink} → {Milk B}	6.32	3.46	4.89
3	{Dish soap} → {Noodle X}	4.99	5.28	5.13
4	{Small Egg} → {Noodle X}	8.31	4.16	6.24
5	{Noodle D} → {Noodle X}	8.12	5.62	6.87
...
17	{Milk A} → {Milk B}	23.03	21.96	22.5

4.3. Marketing strategy

Based on the OCVR analysis result, the marketing strategies that can be made to promote sales were:

1. Product bundling

Supermarket X can make product bundling with special price, for example for noodle X and noodle E or milk A and milk B, etc.

2. Shelves layout arrangement

The Supermarket can arrange the shelves for products that have association close to each other, for example, egg can be put close to noodle, green bean drink close to milk B, etc.

5. Conclusion

Based on the study that has been done, it can conclude the rules obtained in each period were quite different and it indicated the difference in customer buying patterns in each period. The rules were further analyzed using OCVR analysis and the results indicated that 17 rules have OCVR value smaller than 30% and these rules can be used to make marketing strategies. Product bundling and shelves layout arrangement can be conducted to promote sales in Supermarket X.

References

- [1] Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta 2017 *Pertumbuhan Ekonomi DIY Triwulan III-2017* (Yogyakarta : BPS DIY)
- [2] Arifianti R 2016 Pengaruh Promosi Penjualan Terhadap Impulse Buying Pada Hypermarket di Kota Bandung *Unpad Repository*
- [3] Raorane A A et al 2012 Association Rule – Extracting Knowledge Using Market Basket Analysis *Research Journal of Recent Sciences* pp 19-27
- [4] Videla-Cavieres I F and Ríos S A 2014 Extending market basket analysis with graph mining techniques: A real case *Expert Systems with Applications* vol 41 no 4 pp 1928-1936

- [5] Aguinis H, Forcum L E, and Joo H 2013 Using Market Basket Analysis in Management Research *Journal of Management* pp 1799-1824
- [6] Hipp J, Güntzer U and Nakhaeizadeh G 2000 Algorithms for association rule mining—a general survey and comparison *ACM Explor Newsl* vol 2 pp 58–64
- [7] Papavasileiou V and Tsadiras A 2011 Time Variations of Association Rules in Market Basket Analysis *Artificial Intelligence Applications and Innovations* pp 36-44
- [8] Tan P N et al 2011 *Introduction to Data Mining* (MA: Pearson Education)
- [9] Zhang C et al 2019 An improved association rule mining-based method for revealing operational problems of building heating, ventilation and air conditioning (HVAC) systems *Applied Energy* no 253
- [10] Shi Y et al 2019 Determination of effective management strategies for scenic area emergencies using association rule mining *International Journal of Disaster Risk Reduction* no 39
- [11] Dhanalakshmi P and Porkodi R 2017 A Survey on Different Association Rule Mining Algorithms in Data Mining *IPASJ International Journal of Computer Science* pp 126-133
- [12] Sandhu P S et al 2011 Mining utility-oriented association rules: An efficient approach based on profit and quantity *International Journal of the Physical Sciences* pp 301-307
- [13] Han J et al 2012 *Data Mining : concept and techniques* (Waltham: Morgan Kaufmann)
- [14] Cavique L 2007 A scalable algorithm for the market basket analysis *Journal of Retailing and Consumer Services* vol 14 no 6 pp 400-407
- [15] Kaur M and Kang S 2016 Market Basket Analysis: Identify the changing trends of market data using association rule mining *Procedia Computer Science* pp 78-85
- [16] Chen Y L et al 2005 Market Basket Analysis in a Multiple Store Environment *Decision Support System* pp 339-354
- [17] Tatiana K and Mikhail M 2018 Market basket analysis of heterogeneous data sources for recommendation system improvement *Procedia Computer Science* no 136 pp 246-254
- [18] Kim H K et al 2012 A product network analysis for extending the market basket analysis *Expert Systems with Applications* no 39 pp 7403-7410
- [19] Mansur A and Kuncoro T 2012 Product Inventory Predictions at Small Medium Enterprise Using Market Basket Analysis Approach - Neural Networks *Procedia Economics and Finance* no 4 pp 312-320
- [20] Solnet D et al 2016 An untapped gold mine? Exploring the potential of market basket analysis to grow hotel revenue *International Journal of Hospitality Management* no 56 pp 119-125
- [21] Fayyad U et al 1996 From Data Mining to Knowledge Discovery in Database *AI Magazine* pp 37-54

Reproduced with permission of copyright owner. Further reproduction prohibited without permission.