Journal of Information Technology Ampera
Vol. 4, No. 3, 2023 e-ISSN: 2774-2121
https://journal-computing.org/index.php/journal-ita/index

A Pattern Analysis on Goods Purchase Relationship for Minimarket's Customers by Using Association Rule - Market Basket Analysis Method (AR-MBA)

Agus Hindarto Wibowo[1], Marulan Andivas [2], Dimaz Harits[3] ,
Hasyrani Windyatri[4], Davin Satria Nugraha[5]

[1]Industrial Technology, Vocational Education Program, Institut Sains & Teknologi AKPRIND, Yogyakarta, Indonesia
[2,3]Industrial Engineering Department, Universitas Balikpapan, Balikpapan, 76114, Indonesia
[4]Industrial Engineering Department, Faculty of Engineering, Universitas Pelita Bangsa, Cikarang, Indonesia
[5]Industrial Engineering Department, Faculty of Industrial Technology, Institut Sains & Teknologi AKPRIND, Yogyakarta, Indonesia
Email: [1]bagushind@akprind.ac.id, [2]andivas@uniba-bpn.ac.id, [3]dimaz.harits@uniba-bpn.ac.id, [4]Hasyrani@pelitabangsa.ac.id, [5]pindapinn19@gmail.com

Abstract

Nowadays, technology has been rapidly developed, while data has become the most valuable component to be processed to produce useful information. Technology is very helpful for analyzing data clearly or in more detail. The implementation of this technology can be found in real governmental, social, and business activities, for example in business activity, it is indicated by the number of minimarkets spread across Indonesia. Thus, it makes business competition highly increased. Therefore, it is necessary to conduct a study by utilizing the existing data to compete. This study used the Association Rule-Market Basket Analysis method to determine customers' interest patterns when shopping. The results of this study indicated that there were 2 rules that showed the highest confidence value, such as 63% (food and beverages) and 58% (cigarettes and drinks). Regarding these results, the minimarket can determine the next steps that should be conducted, such as setting up the layout and so on.

Keywords: AR-MBA, Association Rule, Data Mining, Market Basket Analysis

1. INTRODUCTION

Technological developments in the industrial revolution 4.0 are increased rapidly, such as technology development in the agricultural sector, information, and in household life. Meanwhile, when viewed from information technology, there is a computing sector that has many uses or functions. One of them is the processing and collection of very large data, called as Big Data [1] or sometimes called Data

Mining, Data Mining is a data collection technique which is obtained from several sources, and then from the data [2], it will be converted into very useful information by using various methods. A very large collection of data and its processing can help someone to do a more detailed and clear data analysis [3]. For example, in business world, business activities are similar with competition in particular, which is a minimarket or retail business [4].

There are a lot of minimarkets which began spreading their branches in Indonesia, and make the owners of the business thinking about how to attract and maintain consumers' interest in their minimarkets. One strategy that can be implemented is by determining the pattern of customers' purchase interest on the product, as a solution to compete with the others. This study combined two methods that aimed to look for the patterns on the impacts of goods purchased by customers in a minimarket, by using Association Rule-Market Basket Analysis. This method is commonly used in various applications in business such as layout design, catalog design, marketing strategies, business orientation development, and so on [5]. For example, in a layout design, an analysis is conducted in order to identify which items that have the highest purchase intensity. Visualization in the layout arrangement can affect customers' interest in choosing certain items. Therefore, the layout should be reorganized if the owner wants to equalize the intensity of goods purchase. Based on the previous explanations.

The purpose of this research is to identify customer purchasing patterns, such as what products are often purchased at the same time or which products are frequently purchased after purchasing a particular product. This information can help the minimarket understand the preferences and needs of the customer. Another purpose is to understand the relationship between the products. This can help minimarket arrange their products on the shelves so that the products that are often bought at the same time are placed close to each other. It can improve customer comfort and encourage cross-sales, and identifying products that have strong links can help in designing more effective sales and promotion strategies. For example, offer discounts or promotions simultaneously for products that are often purchased at the same time.

2. METHODS

2.1. CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM is a standard methodology used in data mining processes. This methodology provides guidelines and a structured framework to guide data analysis teams at every stage of the data-mining process, from business

understanding to solution implementation [6]. Association rule mining's primary goal is to look at sets of objects that commonly appear together in the provided dataset. With the use of data machine learning mining algorithms, this methodology, which is goal-oriented and mature, is still commonly used in data mining projects [7]. CRISP-DM is known as a life cycle approach that consists of six major stages[8] [9]:

a. Business Understanding:
   This phase focuses on understanding the business objectives and problems that are to be solved. The data mining team works with business stakeholders to formulate relevant questions and define the purpose of data analysis. In this study, the business objectives are to know customer purchasing patterns, product optimization, and sales and promotion strategy formulation.
b. Data Understanding:
   At this stage, the team understands the data available for analysis. This involves data collection, data quality checking, and an understanding of the basic characteristics of the datasets to be used. This study uses 100 spending data transactions of product purchases carried out by customers of the minimarket while shopping.
c. Data Preparation:
   Pre-processing the data is done at this phase to make sure it is prepared for analysis. Data integration from several sources, data transformation, and data cleaning are some of the tasks involved.
d. Modeling:
   The creation of descriptive or predictive models utilizing the proper data mining techniques is the main goal of this stage. Here, model creation, model assessment, and model performance modification are some of the most employed strategies. The major functions of data mining estimation, prediction, classification, clustering, and association have been grouped by the earlier references [10]. This study uses the association model (AR-MBA) and the FP-Growth approach.
e. Evaluation:
   To make sure the created model satisfies the original business goals, evaluate the model. Evaluation entails gauging the model's effectiveness and determining whether the resultant solution can be put into practice.
f. Deployment:
   Using data mining models or solutions in a production environment is what this phase entails. It involves making sure that the anticipated advantages of data analysis can be achieved and incorporating solutions into company processes.

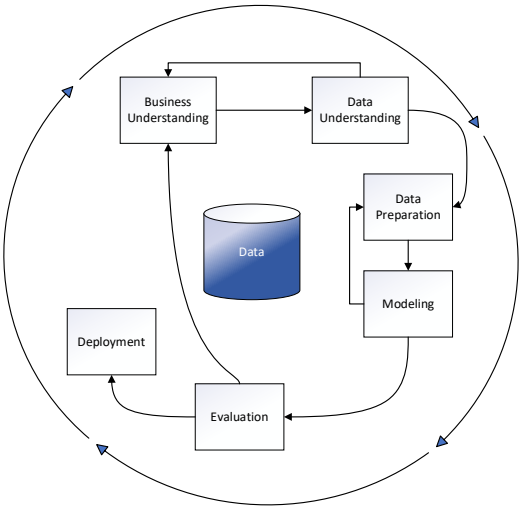

Figure 1. The CRISP-DM process model of data mining

2.2.Association Rule

Association rules aim to find certain rules that associate one data with another data. In order to find the association rule for a data set, the first step is to look for the *frequent itemset* (a group of items that often appear together). After the frequent itemset patterns are found, then the associative rules are supposed to adjust the specified conditions. One algorithm that can be used to find association rules is Apriori algorithm. One of the features in Apriori algorithm is: if an itemset is included in a large itemset, then all subsets of the itemset are also included in a large itemset. The main objective of association rule mining is to examine groups of items that frequently occur together in the given dataset [11].

Association Rule is a recommendation method based on the value of Confidence and Support as references. Confidence value is the intensity value obtained when the product is purchased together. Meanwhile, Support is the intensity value obtained when a product is purchased together with other products in the whole transaction. Since this method utilized data mining, the recommendations can be prepared in advance. Association rule is one of the main techniques in data mining and becomes the most common form used in finding patterns from a data set [12].

$$S = \frac{\Sigma(Ta+Tc)}{\Sigma(T)} \quad (1)$$

$$Support = P(A \cap B) = \frac{The\ Number\ of\ Transaction\ that\ Contains\ A \cap B}{Total\ transaction} \quad (2)$$

Where: S = Support, $\sum(Ta + Tc)$ = the number of transaction that contains *antecendent* and *consequen.* $\sum(T)$ = the number of transaction.

$$C = \frac{\sum(Ta+Tc)}{\sum(Ta)} \quad (3)$$

$$Confidence = P(B/A) = \frac{Support\ (A \cap B)}{P\ (A)} \quad (4)$$

Where: C = Confidence, $\sum(Ta + Tc)$ = the number of transaction that contains *antecendent* and *consequen.* $\sum(T)$ = the number of transaction that contains *antecendent.*

2.3.Market Basket Analysis

Market Basket Analysis can be applied in various fields, for example marketing [13]. Market Basket Analysis is an analysis of consumer shopping habits, by finding associations between various items that consumers put in their shopping baskets. This analysis is commonly applied in analyzing supermarkets. Shopping basket can inform about what items that are purchased simultaneously by consumers who eventually go to the cashier to pay, and stored in the computerized system. At a glance,there is no special thing happens in the transaction. However, from the supermarket side, actually it has acquired a knowledge that is rarely known in the previous, such as the pattern of goods purchased by consumers, which can be seen from a series of items purchased by consumers.

In several cases, the pattern of items bought simultaneously by consumers is easy to guess, for example milk is bought together with bread. However, there might be an item purchase pattern that is out of our thought, for example purchasing cooking oil with detergent. Perhaps this pattern has never been thought since cooking oil and detergent have no relationship, both are distingushed as supplementary goods and substitute goods. This might have never been thought and another case cannot be anticipated, for example when the detergent is out of stock. This is one of the benefits that can be obtained from Market Basket Analysis. On the other words, the purpose of Market Basket Analysis is to design sales or marketing strategies by utilizing sales transaction data that is already available in the company.

2.3. Data Preporocessing

The first step conducted after obtaining the transaction data is to classify each variable or each transaction into the appropriate department. In this study, there are 10 (ten) departments, such as drinks, food, toiletries, stationery, raw food, medicine, baby equipments, cigarettes, candies, and household equipments. Table 1. shows the classification of the departments based on the type of goods purchased.

Table 1. The classification of goods and its department

No	Type of goods purchased	Department
1	Drinks	dept1
2	Food	dept2
3	Toiletries	dept3
4	Stationary	dept4
5	Raw Food	dept5
6	Medicine	dept6
7	Baby Equipments	dept7
8	Cigarette	dept8
9	Candies	dept9
10	Household Equipments	dept10

Next, after each type of goods purchased is classified into each department, then it is changed to the type of goods purchased into each department for each transaction. Table 2. shows the examples of transactions from 1 to 100 and types of goods that have been classified with their respective departments.

Table 2. Change items into department types

Transaction	Kind of Department	
Transaction 1	dept1	dept2
Transaction 2	dept1	dept2
Transaction 3	dept1	dept8
Transaction 4	dept2	dept1
Transaction 5	dept1	dept2
...	...	...
Transaction 100	dept2	dept10

2.4. Data Transformation

Data transformation is a process of changing the type of data, for example changing numerical data into categorical data, or changing from several existing variables, when a new composite of variable is created. Data transformation is a process to change the type of the data in order to ready to be analyzed. In table 3.3, it shows the examples of variable changes into numerical form, from transaction 1 to 5. Meanwhile, number 1 shows the transaction of the goods purchased within the specified department, as well as number 0 which means that the transaction is not in purchasing goods from the department.

Table 3. Data Transformation

Transaction	dept1	dept2	dept3	dept4	dept5	dept6	dept7	dept8	dept9	dept10
Transaction 1	1	1	0	0	0	0	0	0	0	0
Transaction 2	1	1	0	0	0	0	0	0	0	0
Transaction 3	1	0	0	0	0	0	0	1	0	0
Transaction 4	1	1	0	0	0	0	0	0	0	0
Transaction 5	1	1	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...
Transaction 100	0	1	0	0	0	0	0	0	0	1

2.5 Flow chart

This research starts from the first step, which is a preliminary study and then continues with a literature study. Then , the problem of this study is identified and formed a research model. In forming the research model, research objects and data collection techniques are determined. Furthermore, data is collected from XYZ customers' shopping receipt, in Kaliurang Street Km 8.3 Yogyakarta. After the data has been collected, then it is processed and analyzed into the final step to draw conclusions and provide recommendations for the company and other researchers.

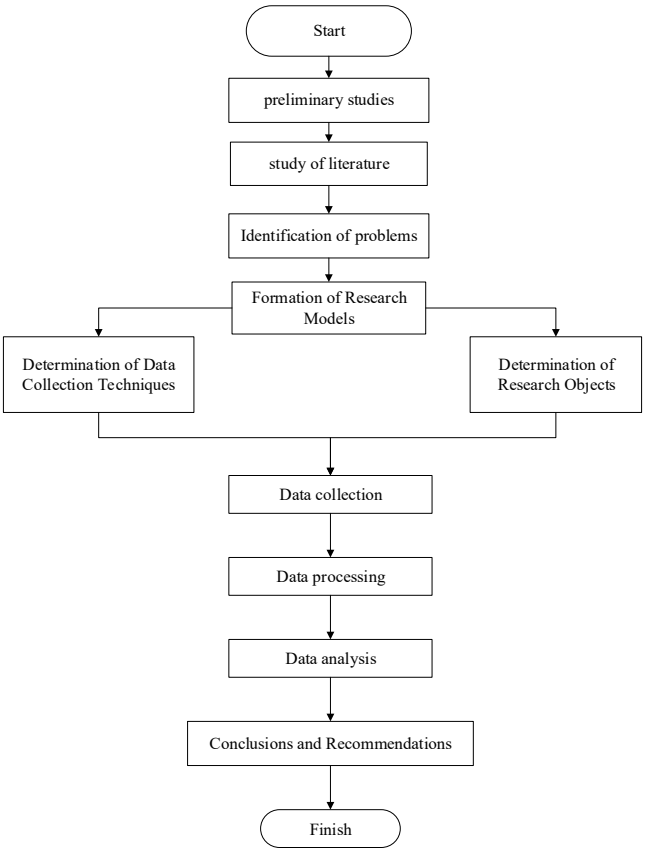

Figure 2. Research flow chart

3. RESULT AND DISCUSSION

The objects of this research were customers of XYZ. The method used to collect the transaction data was by finding and analyzing the shopping receipt of XYZ's customers by using several applications, such as Microsoft Excel and Rapidminer. As an example, the result obtained using the formula is to find out the value of Support by looking at the number of transactions that contain department 8 (A) and department 1 (B) which is 10 transactions, so if entered into the formula, it is as follows:

$$Support(A \cap B) = \frac{The\ Number\ of\ Transaction\ that\ Contains\ (A \cap B)}{Total\ transaction}$$

$$Support(A \cap B) = \frac{10}{100} = 0{,}1 = 10\%$$

Then to find out the Confidence value is using the previously known Support value, so it can be written as follows:

$$Confidence = \frac{Support\ (A \cap B)}{P\ (A)}$$

$$Confidence = \frac{0{,}1}{(17/100)} = 0{,}58 = 58\%$$

In this study, the primary data were obtained from XYZ's customer shopping receipts. Table 4. shows the results of the calculations by using Rapidminer software.

Table 4. The results of Association Rule-Market Basket Analysis

Rule	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
1	dept8	dept1	0.1	0.58	0.94	-0.24	-0.007	0.93	0.89
2	dept2	dept1	0.35	0.63	0.87	-0.75	0.0034	1.01	1.0175
3	dept5	dept2	0.080	0.571	0.947	-0.20	0.003	1.04	1.05
4	dept1	dept2	0.350	0.556	0.83	-0.91	0.003	1.01	1.012

Association Rules
[dept9] --> [dept1] (confidence: 0.500)
[dept9] --> [dept2] (confidence: 0.500)
[dept1, dept5] --> [dept2] (confidence: 0.500)
[dept1, dept9] --> [dept2] (confidence: 0.500)
[dept2, dept9] --> [dept1] (confidence: 0.500)
[dept1] --> [dept2] (confidence: 0.556)
[dept5] --> [dept2] (confidence: 0.571)
[dept8] --> [dept1] (confidence: 0.588)
[dept2] --> [dept1] (confidence: 0.636)

From the results of the data processed by using Rapidminer with a minimum support of 0.1 and a minimum confidence of 0.58, the results of 2 rules are

shown in Table 4. The data above shows the results that rules 1 and 2 are the owners that obtain the greatest confidence value compared to the whole rules, indicates that department 2 is often purchased simultaneously with department 1 and vice versa. This can be seen from the value of support, lift ratio, and the greatest confidence (%) of 35, 1,010 and 63% which shows the number of departments 2 and 1 purchased simultaneously in all transactions. The possibility of purchasing department 2 with 1 is higher than the possibility of purchasing department 1 with 2, as indicated from the value of confidence (%), when the possibility of purchasing department 2 with 1 is greater than the possibility of buying department 1 by 2. Then, department 8 is often bought simultaneously with department 1. This is indicated from the value of the second largest support after the value of support department 2 by 1 obtained, when 10 shows the number of departments 8 with 1 purchased simultaneously in the whole transaction. The ratio and the level of confidence of departments 8 and 1 are purchased simultaneously, as indicated from the value of the lift ratio and confidence of 0.93 and 58%.

4. CONCLUSION

The results revealed that there are 2 rules that have the highest confidence value, mean that both rules are departments that consist of items that are purchased more often. The combination of rule 1 which consists of department 8 (cigarettes) and department 1 (drinks) often bought with the highest value of support, such as lift ratio and confidence (%) of 10, 0.93 and 58% which shows the number of departments 8 and 1 purchased simultaneously in overall transactions, rather than rule 2 which consists of department 2 (food) and 1 (drinks) and often bought with the largest value of support, lift ratio, and confidence (%) of 35, 1.01 and 63% which shows the number of departments 2 and 1 purchased simultaneously in the whole transaction. The knowledge gained in the rules can be used for several applications such as cataloging by combining products that are often bought together and placed it near the catalog, choosing a shopping layout where products that are often bought together should be placed as near as possible, and making shopping coupons by giving products that are rarely bought as a bonus from purchasing products that are often bought.

REFERENCES

[1] S. J. Miah, H. Q. Vu, J. Gammack, and M. McGrath, "A Big Data Analytics Method for Tourist Behaviour Analysis," *Information & Management*, vol. 54, no. 6, pp. 771–785, Sep. 2017, doi: 10.1016/J.IM.2016.11.011.

[2] A. H. Wibowo and T. I. Oesman, "The comparative analysis on the accuracy of k-NN, Naive Bayes, and Decision Tree Algorithms in predicting crimes and criminal actions in Sleman Regency," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1742-6596/1450/1/012076.

[3] R. Agrawal, T. Imieliński, A. Swami, R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93*, New York, New York, USA: ACM Press, 1993, pp. 207–216. doi: 10.1145/170035.170072.

[4] Y.-L. Chen, K. Tang, R.-J. Shen, and Y.-H. Hu, "Market basket analysis in a multiple store environment," *Decis Support Syst*, vol. 40, no. 2, pp. 339–354, Aug. 2005, doi: 10.1016/J.DSS.2004.04.009.

[5] M. A. Valle, G. A. Ruz, and R. Morrás, "Market basket analysis: Complementing association rules with minimum spanning trees," *Expert Syst Appl*, vol. 97, pp. 146–162, May 2018, doi: 10.1016/J.ESWA.2017.12.028.

[6] S. Peker and Ö. Kart, "Transactional data-based customer segmentation applying CRISP-DM methodology: A systematic review," *Journal of Data, Information and Management*, vol. 5, no. 1–2, pp. 1–21, Jun. 2023, doi: 10.1007/s42488-023-00085-x.

[7] J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, "Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 512–517. doi: 10.1016/j.procs.2021.12.278.

[8] F. Martinez-Plumed *et al.*, "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories," *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

[9] V. Kotu and B. Deshpande, *Predictive analytics and data mining : concepts and practice with RapidMiner*. 2015.

[10] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 1st ed. New Jersey: John Wiley & Sons, 2005.

[11] M. Shahin, M. R. Heidari Iman, M. Kaushik, R. Sharma, T. Ghasempouri, and D. Draheim, "Exploring Factors in a Crossroad Dataset Using Cluster-Based Association Rule Mining," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 231–238. doi: 10.1016/j.procs.2022.03.032.

[12] 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, p. 20, Dec. 2017, doi: 10.17977/um018v1i12018p20-25.

[13] M. Kaur and S. Kang, "Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining," *Procedia Comput Sci*, vol. 85, pp. 78–85, Jan. 2016, doi: 10.1016/J.PROCS.2016.05.180.