Implementation of Association Rule-Market Basket Analysis in Determining Product Bundling Strategy: Case Study of Retail Businesses in Indonesia

Zakka Ugih Rizqi
Department of Industrial Engineering
Islamic University of Indonesia
Yogyakarta, Indonesia
ugihzakka@gmail.com

Abstract

The increasing retail industry in Indonesia makes competitiveness among retail businesses even higher. Therefore, a good marketing strategy is important to do. Retail Z is a retailer that sells handphone (HP) and electronic accessories, which want to increase sales with limited capital. One of strategy that can be done is product bundling. However, a product bundling strategy that is incompatible with customer needs will be in vain. Therefore, a technique is needed to determine which product will be bundled. By using one of the data mining methods namely Association Rule-Market Basket Analysis it can be determined which products are often purchased simultaneously by taking into account the parameters of support and confidence. Results obtained with the help of RapidMiner software, department 5 and department 2 are departments whose products are often purchased simultaneously. This is indicated by Lift Ratio 1.029, Confidence 0.852 and Support 0.397. So that product bundling can be done by giving discounts to buyers who make product purchases in department 5 and in department 2 simultaneously.

Keywords
Association Rule, Product Bundling, Data Mining, Market Basket Analysis, Retail

1. Introduction

Everyone wants to buy their needs in a place that is easily accessible and provides the desired product. Retail which is one of the links in the distribution of goods is made to simplify the purchasing process. With retail, customers do not have to come all the way to the production site so that the purchasing process is difficult to do.

The current development of the retail world is so rapid that it has an impact on the increasing competition for market share among retail businesses. The Indonesian Retail Entrepreneurs Association (APRINDO) (2016) revealed that the growth of the retail industry in Indonesia reached 10 percent in 2016. The current retail industry is controlled by large organizations, including Department Stores, Chain Organizations, Warehouses, Supermarkets, Catalogs Store, Outlet and Online Store (Bernadetta 2012).

Retail Z is handphone (HP) and electronic accessories retailer located in Yogyakarta. The increasing number of similar retailers creates a separate threat for Retail Z. A marketing strategy is needed that can understand consumer behavior so that consumers will remain loyal to buy the items they need in Retail Z. One of strategy that can be done is by product bundling that can attract customers to buy HP & electronic accessories in place.

Product bundling that are not done properly will only be in vain, because bundled products are not in accordance with the customer's wishes. Therefore, the determination of bundled products will be based on the association relationship between items in Retail Z. By bundling the product, it is expected that customers will be interested in buying products in Retail Z so that they can increase their total sales. Because, the bundling strategy has a significant effect on consumer buying interest (Buananda & Ariyanti 2018).

Along with the development of the retail business, currently the retail business is no longer managed traditionally, but in a modern way so that it becomes an innovative, dynamic and competitive business (Ghosh, 1992). In Retail Z, the transaction process is carried out using a recording machine with barcode technology and has been commonly used in retail. With this machine, retailers can store transaction data in a transaction database so that any information that contains dates and items purchased can be recorded. This data is referred to as basket data, while studies on basket data are called Market Basket Analysis (MBA).

In this study, the MBA method which is one of the applications of association rules will be used to find association rules among large sets of item data in the Z Retail transaction database with the help of RapidMiner software. If the association rules are implemented in the transaction database, this will be very useful in determining marketing strategies, one of which is by bundling the product.

2. Methods

2.1 Data Collection Method

Primary data collection is done by direct interview to retail business ownersrelated to products sold along with departments in Retail Z. In addition, a survey is also conducted directly to Retail Z to obtain data on purchase transactions that will be processed using the method of Association Rule-Market Basket Analysis (AR-MBA) with the help of RapidMiner software so that the rules will be proposed for Retail Z.

2.2 Association Rule Method

Association rules are procedures to find relationships that exist between items in a data set (Han & Kamber 2001). While association in the perspective of data mining is an activity to determine what attributes are often obtained simultaneously. The association pattern is also one of the very interesting functionalities in terms of calling data (Kumar & Wahidabanu 2007). Association rules will be processed through two stages, first looking for combinations that most often occur from a set of items, both defining conditions and results (Ulmer 2002).

2.3 Market Basket Analysis Method

Market Basket Analysis is an analysis technique of customer habits when shopping by finding associations and correlations between various kinds of items that customers put in their shopping basket (Muzakir & Adha 2016). Market Basket Analysis provides an understanding of merchandise by telling which products are possible to be purchased simultaneously and which products are more approved to be promoted. Because in Market Basket Analysis, what is seen is not only understanding the quantity of items purchased in the basket, but also how the items purchased relate to each other with other items.

2.4 Support, Confidence and Lift Ratio

According to Nofriansyah (2014), an associative rule can be known whether it is important or not with two parameters, namely support (supporting value) which is the percentage combination of items and confidence (certainty value) which explains the strong relationship between items that exist in associative rules. Associative rules are usually expressed in the form of:

$$\{A, B\} \rightarrow \{C\} \ (support = 40\%, confidence = 50\%)$$

Which means that 50% of the total transactions in the database contain items A and B also contain item C. While 40% of all transactions in the database contain all three items.

Lift Ratio is an important parameter besides support and confidence in the association rule. Lift Ratio measures how important a rule is formed based on the value of support and confidence. Lift Ratio is a value that shows the validity of the transaction process and provides information whether product A is purchased together with product B. A transaction is said to be valid if it has a value of Lift Ratio of more than 1, which means that the transaction A and B are actually purchased simultaneously.

3. Results and Discussion

3.1 Data Collection Results

Based on the results of interview with the owner of retailers Z, it is obtained that the department in Retail Z is as shown in Table 1.

Table 1. Division of Department of Retail Z

Departments	Type of Goods
Department 1 (HP Protector)	HP Case, Tempered Glass
Department 2 (HP Support Accessories)	Charger, Powerbank, Earphone
Department 3 (Computer/Laptop Accessories)	Wireless Mouse, Printer Ink, Keyboard, Bluetooth Speaker, USB Fan
Department 4 (Cable & Electronics)	VGA Cable, HDMI Cable, USB Cable, Adapter, CCTV
Department 5 (Memory & Games)	Hard Disk, SD Card, Joystick

3.2 Data Pre-Processing

After the data is collected, pre-processing data is carried out which is a step so that the data matches with desired format. In addition, there is also a cleaning process which includes, among others, recapitulating data, removing duplicate data, checking inconsistent data, and carrying out transformation. After checking, the data obtained all meet the requirements to be processed in data mining so that all data obtained will be used as material for calculation. The data used are survey data of customers who have purchased goods in Retail Z, with 58 data that are clean of noise and missing data. Data from the recapitulation results can be seen in Table 2.

Table 2. Data Recapitulation Results

Transaction	Type of Goods
1	Tempered Glass, HP Case, Charger, Earphone, Powerbank
2	HP Case, USB Cable, Printer Ink, Wireless Mouse, SD Card
3	HP Case, Tempered Glass, Powerbank, Joystick, Earphone
4	HP Case, Tempered Glass, USB Cable, Charger, SD Card
5	HP Case, Tempered Glass, Earphone, Powerbank, SD Card
6	HP Case, SD Card, Tempered Glass, Powerbank, Charger
7	HP Case, Tempered Glass, Charger
8	Tempered Glass, HP Case, Charger, Powerbank, Bluetooth Speaker
9	HP Case, Charger, Earphone, USB Cable, Adaptor
10	HP Case, Earphone, Charger, Wireless Mouse
11	Tempered Glass, USB Cable, Powerbank, Wireless Mouse, SD Card
12	HP Case, Tempered Glass, Charger, USB Cable, USB Fan
13	USB Cable, SD Card, Wireless Mouse, Bluetooth Speaker, Powerbank
14	USB Cable, Charger, HP Case
15	Tempered Glass, HP Case, SD Card, Charger, Powerbank
16	HP Case, Tempered Glass, Bluetooth Speaker, Wireless Mouse, Hard Disk
17	Tempered Glass, HP Case, SD Card, USB Fan
18	Earphone, Charger, Powerbank, HP Case
19	HP Case, Tempered Glass, USB Cable, SD Card, Powerbank
20	HP Case, Charger, SD Card, Powerbank, Bluetooth Speaker
21	Charger, USB Cable, HP Case, Tempered Glass
22	HP Case, Tempered Glass, Hard Disk, Earphone, USB Fan
23	Tempered Glass, HP Case, USB Cable, Earphone, Wireless Mouse
24	Tempered Glass, HP Case, Earphone, Joystick, Wireless Mouse
25	HP Case, Earphone, Tempered Glass, SD Card, Harddsik

Transaction	Type of Goods
26	Earphone, Charger, USB Cable, USB Fan, Powerbank
27	HP Case, Tempered Glass, Charger, USB Cable, Powerbank
28	SD Card, HP Case, Tempered Glass, Charger, Powerbank
29	Earphone, Printer Ink, HP Case, Powerbank, Charger
30	Earphone, HP Case, Charger, Tempered Glass, Powerbank
31	SD Card, Charger, USB Cable, Keyboard, Adaptor
32	Printer Ink, Charger, Earphone, Tempered Glass, Bluetooth Speaker
33	HP Case, Tempered Glass, Earohone, SD Card, Printer Ink
34	Printer Ink, Charger, Powerbankm HP Case, Hard Disk
35	HP Case, Charger, Tempered Glass, Powerbank, SD Card
36	HP Case, USB Cable, Wireless Mouse
37	HP Case, Tempered Glass, Powerbank, USB Cable, Wireless Mouse
38	HP Case, USB Cable, Charger, Printer Ink, Powerbank
39	HP Case, Tempered Glass, Charger
40	HP Case, Charger, Powerbank
41	Tempered Glass, HP Case, Charger, Bluetooth Speaker, Powerbank
42	Tempered Glass, HP Case, Adaptor
43	HP Case, Tempered Glass, Adaptor
44	Tempered Glass, HP Case, Adaptor
45	HP Case, Tempered Glass, Adaptor
46	Earphone, Hard Disk, HDMI Cable
47	Earphone, Wireless Mouse, Hard Disk, HDMI Cable
48	Wireless Mouse, Earphone, HDMI Cable, Bluetooth Speaker
49	Powerbank, Hard Disk, HDMI Cable
50	Bluetooth Speaker, VGA Cable, Earphone, HDMI Cable, Printer Ink
51	Earphone, Hard Disk, Bluetooth Speaker
52	Charger, Adaptor, USB Cable
53	Wireless Mouse, Powerbank, HP Case
54	Keyboard, Joystick, SD Card
55	VGA Cable, HDMI Cable, USB Fan
56	Tempered Glass, HP Case, Bluetooth Speaker
57	Powerbank, Charger, HP Case
58	HP Case, Tempered Glass, Powerbank, SD Card, Wireless Mouse

The data that has been recapitulated is then transformed based on existing departments so that data can be processed using the RapidMiner application. The transformation data can be seen in Table 3.

Table 3. Data Transformation Results

Transaction	Departement 1	Departement 2	Departement 3	Departement 4	Departement 5
1	1	1	0	0	0
2	1	0	1	1	1
3	1	1	0	0	1
4	1	1	0	1	1
5	1	1	0	0	1
6	1	1	0	0	1
7	1	1	0	0	0
8	1	1	1	0	0
9	1	1	0	1	0
10	1	1	1	0	0
11	1	1	1	1	1

Transaction	Departement 1	Departement 2	Departement 3	Departement 4	Departement 5
12	1	1	0	1	0
13	0	1	1	1	1
14	1	1	0	1	0
15	1	1	0	0	1
16	1	0	1	0	1
17	1	0	1	0	1
18	1	1	0	0	0
19	1	1	0	1	1
20	1	1	1	0	1
21	1	1	0	1	0
22	1	1	1	0	1
23	1	1	1	1	0
24	1	1	1	0	1
25	1	1	0	0	1
26	0	1	1	1	0
27	1	1	0	1	0
28	1	1	0	0	1
29	1	1	1	0	0
30	1	1	0	0	0
31	0	1	1	1	1
32	1	1	1	0	0
33	1	1	1	0	1
34	1	1	1	0	1
35	1	1	0	0	1
36	1	0	1	1	0
37	1	1	1	1	0
38	1	1	1	1	0
39	1	1	0	0	0
40	1	1	0	0	0
41	1	1	1	0	0
42	1	0	0	1	0
43	1	0	0	1	0
44	1	0	0	1	0
45	1	1	0	1	1
46	0	1	0	1	1
47	0	1	1	1	1
48	0	1	1	1	0
49	0	1	0	1	1
50	0	1	1	1	0
51	0	1	1	0	1
52	0	1	0	1	0
53	1	1	1	0	0
54	0	0	1	0	1
55	0	0	1	1	0
56	1	0	1	0	0
57	1	1	0	0	0
58	1	1	1	0	1

3.3 Analysis of the Output of RapidMiner Results

The results of data transformation will be used to model the AR-MBA using RapidMiner software. RapidMiner is software for processing data using the principle of data mining, RapidMiner extracts patterns from large data sets by combining statistical methods, artificial intelligence and databases (Rahmat et al., 2017). In this AR-MBA method, the FP Growth algorithm is an alternative algorithm that can be used to determine the frequent itemset in a data set (Samuel, 2008). The design of making models using RapidMiner can be seen in Figure 1.
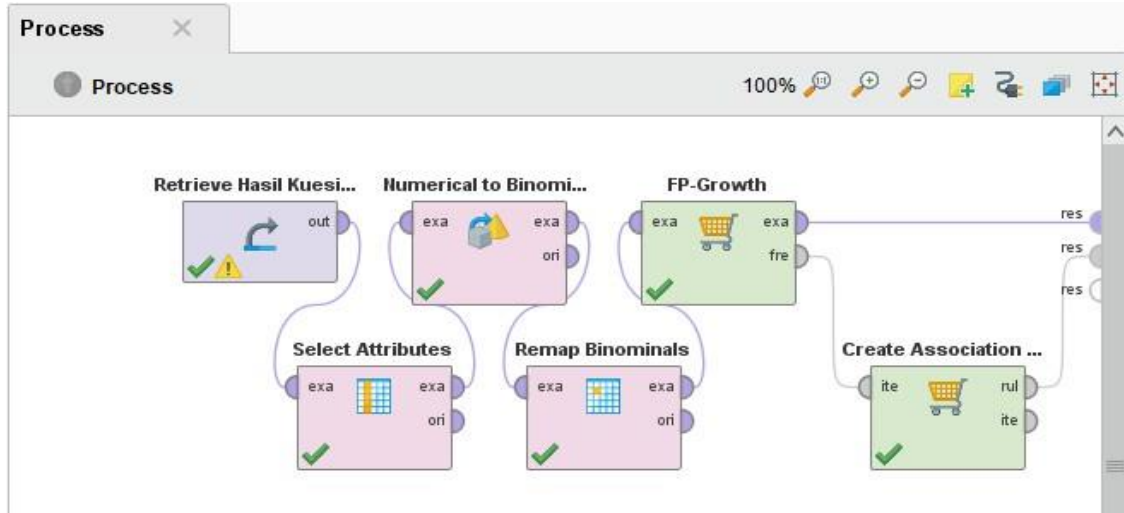

Figure 1. Model design in rapidminer software

By using a minimum Support value of 0.3 or 30% and a minimum value of Confidence 0.6 or 60% (as bigger as more accurate), the output is obtained as shown in Figure 2.

No.	Premises	Conclusion	Support	Confiden... ↑	LaPlace	Gain	p-s	Lift	Convicti...
1	Departemen 5	Departemen 1	0.345	0.741	0.918	-0.586	-0.024	0.934	0.798
2	Departemen 3	Departemen 2	0.397	0.767	0.920	-0.638	-0.032	0.926	0.739
3	Departemen 4	Departemen 2	0.362	0.778	0.929	-0.569	-0.023	0.940	0.776
4	Departemen 2	Departemen 1	0.655	0.792	0.906	-1	-0.001	0.998	0.993
5	Departemen 1	Departemen 2	0.655	0.826	0.923	-0.931	-0.001	0.998	0.991
6	Departemen 5	Departemen 2	0.397	0.852	0.953	-0.534	0.011	1.029	1.164

Figure 2. Output of rapidminer

Based on the output results above, it can be seen that there are 6 rules formed. But a valid rule is only one, namely the sixth rule with the value of Lift Ratio 1.029, Confidence 0.852 and Support 0.397. This means that items in Department 5 (Memory & Games) will be purchased along with items in Department 2 (HP Support Accessories) with a confidence level of 85% and supported by 39% of all data. Therefore, product bundling can be determined by classifying products in departments 5 and 2. Examples of product bundles that can be determined are hard disk products (department 5) and charger products (department 2) as shown in Figure 3.

Figure 3. Examples of product bundling that can be applied retail z

4. Conclusions

Based on the research that has been done on Retail Z using the AR-MBA method, it can be concluded that by using a minimum value of Support which is 0.3 or 30% and the minimum value of Confidence 0.6 or 60%, there are 6 rules based on the output of RapidMiner. From the 6 rules obtained, only 1 rule can be said to be valid, namely the sixth rule that describes the associative relationship between items in Department 5 (Memory & Games) will be purchased simultaneously with items in Department 2 (HP Support Accessories) with a confidence level of 85 % and supported by 39% of all data. Therefore, the proposed solution to increase sales with product bundling based on AR-MBA analysis in Retail Z is that the customer gets a discount if buying items included in department 5 (Memory & Games) and department 2 (HP Supporting Accessories) , for example, the customer will get a discount when buying a hard drive with a charger simultaneously.

References

Aprindo, Industri Ritel Tumbuh Sekitar 10 Persen pada 2016, Available: http://bisniskeuangan.kompas.com/read/2016/12/28/172204826/aprindo.industri.ritel.tumbuh.sekitar.10.persen.pada.2016, 2016.

Bernadetta D., A., *Analisis Kepuasan Pelanggan pada Alfamart Tembalang dari Dimensi Pelayanan*, Universitas Diponogoro, Semarang, 2012.

Buananda, M., F., and Ariyanti, M., Pengaruh Strategi Bundling terhadap Minat Beli Konsumen di Jakarta, *Proceeding of Management*, vol. 5, pp. 3259-3265, 2018.

Ghosh, A, *Retailing Management*, The Den Press 2nd Edition, Newyork University, Newyork, 1992.

Han, J., and Kamber, M., *Data Mining: Concepts and Techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2006.

Kumar, A. V., & Wahidabanu, R. S. D., Discovery of Frequent Itemsets: Frequent Item Tree-Based Approach, *Journal ITB J. ICT*, vol. 1c, pp. 42-55, 2007.

Muzakir, A., and Adha, L., Market Basket Analysis (MBA) pada Situs Web E-Commerce Zakiyah Collection, *Jurnal SIMETRIS*, vol. 7, pp. 459-466, 2016.

Nofriansyah, D., *Konsep Data Mining Vs Sistem Pendukung Keputusan*, Deepublish, Yogyakarta, 2014.

Rahmat, B., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun. F. R., Purnamasari, Y. P., and Ransi, N, Implemetasi K-Means Clustering pada RapidMiner untuk Analisis Daerah Rawan Kecelakaan, *Seminar Nasional Riset Kuantitatif Terapan 2017*, vol. 1, pp. 58-62, 2017.

Samuel, D., *Penerapan Stuktur FP-Tree dan Algoritma FPGrowth dalam Optimasi Penentuan Frequent Itemset*, Institut Teknologi Bandung, Bandung, 2008.

Ulmer, D., Mining an Online Auctions DataWarehouse, *the Mid-Atlantic Student Workshop on Programming Languages and Systems*, vol. 19, pp. 8.1-8.10, 2002.

Biography

Zakka Ugih Rizqi is an Industrial Engineering student at the Department of Industrial Engineering, Faculty of Industrial Technology, Islamic University of Indonesia, Yogyakarta, Indonesia. He is an undergraduate student and has a great grade in his university. His research interests in modeling and simulation, optimization, project management, production planning & inventory control, supply chain management and logistics management. He also works as a laboratory assistant of industrial modeling and simulation in Islamic University of Indonesia. He also actives in the competition of industrial engineering and writes many researches published in proceedings and scientific journals.