# APPLICATION OF DATA MINING USING ASSOCIATION RULE METHOD ON SALES DATA IN KOPI CILIK CAFE

Heri Wijana[1], Alif Finandhita[2]

[1,2] Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipati Ukur 112-116 Bandung, 40132
E-mail : heri.wijana@email.unikom.ac.id[1], alif.finandhita@email.unikom.ac.id[2]

## ABSTRACT

Kopi Cilik Cafe is a culinary company located on Jalan Raya Laswi no 607 Ciparay, Bandung Regency. The number of rivals out there makes the cafe sales tend to decrease and make the profit of the cafe also decreased in the last few months which makes the cafe lose, so that the manager's thoughts to implement a new strategy in cafe sales. The manager wants to make a promotion in the form of packaging a product in its sales because it can attract consumer interest from a more affordable price. However, to choose a product that will be used as a package that is in accordance with the interests of consumers requires analyzing the sales transaction data in advance to see what menus are often sold together as an initial basis for package formation. One technique that can be applied is to do association rule mining on cafe sales data. The algorithm used in this ARM is the CT-Pro algorithm. This study uses 57 sales transaction data to be tested with minimum support 2 and confidence of 20% to 60% can produce 4 to 13 product packages. The results of the research on the product packaging system that was built in this study were in accordance with the wishes of the Kopi Cilik cafe manager.

**Kata kunci :** Data mining, ARM, association rule, CT-Pro, market basket analysis.

## 1. INTRODUCTION

Kopi Cilik Cafe is a culinary company located on Jalan Raya Laswi no 607 Ciparay, Bandung Regency which sells a variety of foods and drinks, especially coffee. The location of this cafe is quite strategic because of its location close to government offices, health centers, water park, main square, and schools so that this cafe has the potential to be visited every day and generate many transactions every day.

Based on the results of direct interviews with Rinrin Yuliani as the manager of Kopi Kopi Cilik, the number of rivals out there has made cafe sales tend to decline in the last few months resulting in declining profits generated by the cafe and can cause losses to the cafe, so managers are thinking to implement a new strategy. in cafe sales. The manager wants to make a promotion in the form of packaging a product in its sales, because sales in the form of a package will provide new choices to customers at a more affordable price. That way, cafe sales will be more leverage in an effort to attract consumer interest.

However, to choose a product that will be used as a package that suits the interests of consumers requires analyzing the sales transaction data in advance to see what menus are often sold together as the initial foundation in the formation of the package. Currently, managers cannot implement sales in the form of packages because there is no system that can create packages from several menus based on menu combinations that are often sold simultaneously.

One technique to overcome this problem is with data mining, because data mining can search for the desired pattern on large data [1]. In this problem, data mining can find patterns in sales data that will be used as decision support in determining menu combinations on packages by examining the entire sales data. There are many methods in data mining, but this research focuses on the data association function, the Association Rules. Association Rules is a data mining method that can look for interconnections from associations between items in data [2], so that from the use of the association rules method, it can be seen which menus often appear together in product purchase transactions. Then, the pattern formed will be considered in determining the combination of menus that will be made as a package. The association rules algorithm used in this study is the CT-Pro algorithm. CT-Pro algorithm is an algorithm similar to the FP-Growth algorithm [3], the difference is when making a tree where FP-Growth makes FP-Tree while CT-Pro makes Compressed FP-Tree (CFP-Tree) with the number of nodes can be half times the FP-Tree [4]. Therefore CT-Pro algorithm has better performance compared to FP-Growth. When compared with the OpportuneProject algorithm, H-mine and Apriori are still better [4]. CT-PRO also performs better than LCM and kDCI, which are known as the two best algorithms in FIMI Repository 2003 [4].

Therefore based on the above problems as a solution of this research, a proposed "Application of Data Mining Using the Association Rule Method of Sales Data in Kopi Cilik Cafe" as a recommendation that can assist in making decisions in finding information in the form of buying patterns in sales data and determining menu that will be made in one package. With this implementation, hopefully it can facilitate the Kopi Cilik cafe in carrying out its

business strategy and developing the company towards more advanced ones.

## 2. RESEARCH CONTENT

### 2.1 Research Method

The application method used is from the Cross-Industry Standard Process for Data Mining (CRISP-DM) which was developed in 1996 by analysts such as Daimler Chrysler, SPSS and NCR [5]. The flow in this study can be seen in Figure 1.
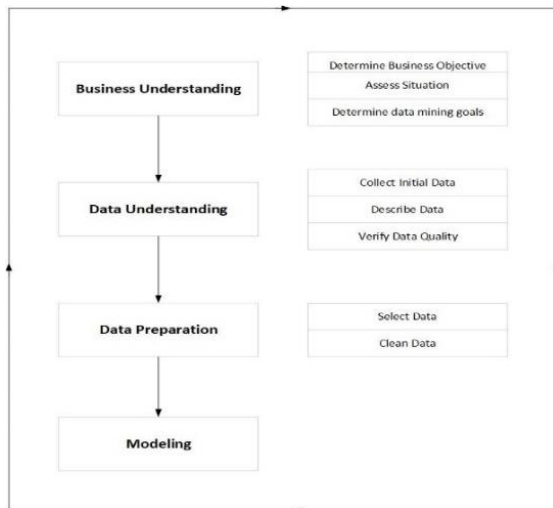


**Figure 1.** Research flow

1. Business understanding

This stage includes understanding the objectives and business needs from a business perspective, and then translating them into data mining problems.

a. Determine Business Objectives

The main purpose of analysis is to truly understand business goals from a business perspective, what the client wants to achieve.

b. Assess Situation

Search for facts that are more about the resources and existing limitations.

c. Determine Data Mining Goals

In this stage a business goal states the goal in view of business terminology.

2. Data understanding

At this stage, data collection is needed and understanding of data will be used as an initial hypothesis to find information in the data.

a. Collect Initial Data

Gather data from project resources that will be used for mining.

b. Describe Data

Explain and check data from noise and report results.

c. Verify Data Quality

This stage is a test of data quality, addressing questions such as: Is the data complete? (Does it cover everything that is needed?) [5].

3. Data preparation

This stage is a data processing process in building a dataset to be processed at the modeling stage. Data processing includes the selection of tables, records, and data attributes, besides that at this stage the data cleaning and transformation process is also carried out [5].

a. Select Data

Determining the data to be analyzed, the criteria must be relevant to the purpose of data mining.

b. Clean Data

This data cleaning is the process of cleaning data from noise to form quality data and will be mining with the chosen technique.

4. Modeling

At this stage of modeling will use the Association Rule method and CT-Pro algorithm. The results of the process will produce associative rules in the form of a combination of menus that can be made as one package, so that it can be known what menus can be used as combinations in packages that will be offered to consumers.

#### 2.1.1. Data

Conceptually, the data is a description of anything that has no meaning before the data is processed by the user [6].

#### 2.1.2. Data mining

Data mining is an activity that includes collecting or using data to find interesting patterns or meaningful relationships in data [7]. The output of data mining can be used for decision making in the future [8]. Data mining methods can be seen from two different approaches, namely descriptive and predictive approaches [9]. Descriptive approach is an approach that describes input data such as association rules and target class concepts, while the predictive approach, which is the approach used to predict, such as classification and clustering [9].

#### 2.1.3. Association rules

Association rules are the process of detecting a collection of attributes that appear together (co-occur) in frequent frequencies and form the rules of the data set [1].

#### 2.1.4. CT-Pro

CT-Pro algorithm is a development algorithm from FP-Growth which is intended to simplify the FP-Tree so that the tree formed is more concise and the processing time is shorter [10].

### 2.2. Discussion

1. Business understanding
a. Determine Business Objectives

The purpose of the Kopi Cilik cafe business is to serve consumers by marketing existing products to meet consumer demand with a variety of strategies and innovations so that consumers feel satisfied. One marketing strategy to be implemented is to utilize package sales. By applying data mining to cafe sales

data, it can produce knowledge in the form of linkages between menus that are often bought by consumers which can be used as a basis for determining what menus can be used as a combination of packages to be offered.

b. Assess Situation
- Resource
  1) Hardware Resource
  a. There are transaction support tools, namely thermal printers and cash drawers.
  b. Computers with the following specifications in the table 1.

**Table 1.** Hardware Spesification

| Hardware | Spesification |
|---|---|
| Processor | Intel Core i3 3.2 Ghz |
| Harddisk | 500GB |
| Memory | 4GB |

2) Data Resource
a. Existing data resources are stored in the Kopi Cilik café database.
b. The data used in this study are sales data for October 2018.
c. c. The sales data used are export results from the Point of Sale (POS) application in the Kopi Cilik cafe with MS Excel (xlsx) format.

c. Determine Data Mining Goals
The purpose of doing this data mining is to look for patterns of consumer purchases on data as an initial basis to determine what menus can be used as a package that can be offered to consumers.

2. Data understanding
a. Collect Initial Data
Preliminary data obtained at the Kopi Cilik cafe can be seen in the figure 2.


**Figure 2.** Sales transaction data

b. Describe Data
The following are the attributes of the data in the table 2.

**Table 2.** Attribute data

| Information | | |
|---|---|---|
| Description | This data is the result of export from the sales transaction data database at Kopi Cilik cafe | |
| Attribut | Tanggal | Purchase date |

| | | |
|---|---|
| Bill | Purchase number |
| No_Meja | Table number |
| Kode_Menu | The code for each menu purchased |
| Nama_Menu | The name of the menu purchased |
| Kategori | Categories from the purchased menu |
| Harga | Price of the menu purchased |
| Qty | The amount from the menu purchased |
| Total | The total price of the menu purchased |
| Total Bayar | The total price that consumers must pay |

c. Verify Data Quality
The following are the facts in the existing data:
- Data on each attribute is complete (no attributes or values are lost)..
- Jika ada atribut ataupun nilai yang hilang maka akan dianggap noise.
- In this study, sales transaction data that only has one item at a time are considered noise, because this study looks for associations between items, whereas data that only has one item in one transaction has no relationship with other items.

To eliminate noise in the sales transaction data, the data cleaning process will be carried out at a later stage in the data preparation stage.

3. Data preparation
a. Select Data
The attributes selected in this study are only Bill and Kode_Menu, Bill is used to distinguish one transaction from another transaction, and the Nama_Menu attribute is used to find out what menus are bought in one transaction.

b. Clean Data
Data cleaning is done by removing noise on the data. In this study noise means transaction data that has a single data.

4. Modeling
a. Find the frequent itemset
1. Count the number of occurrences of each menu
The following in table 3 is the number of occurrences of each menu in the transaction.

**Table 3.** The appearance of each menu

| Nama_Menu | Frequent |
|---|---|
| Ice Tea | 24 |
| Ayam Bakar | 12 |
| Espresso | 1 |
| Thai Lychee Tea | 8 |
| Nasi Goreng Spesial | 9 |
| Geprek Mozzarela | 1 |
| Katsudon | 9 |

| Nama_Menu | Frequent |
|---|---|
| GepreK Merah | 1 |
| Hot Crispy Chicken | 8 |
| Milky Caramelly | 1 |
| Latte | 5 |
| Milkshake Strawberry | 1 |
| Cappuccino | 1 |
| Geprek Hejo | 5 |
| Ice Cream | 4 |
| Mie Goreng Gila | 1 |
| Pisang Keju | 1 |
| Milkshake Chocolate | 4 |
| Roti Bakar | 1 |
| Greentea Caramelly | 1 |

2. Selection by determining the minimum value of support

In this case the menu limit that appears must be >= 2 times. So for the menu whose frequency does not meet the minimum support limit will be eliminated or can be seen in the following table 4.

**Table 4.** Menu that meets the minimum support

| Nama_menu | Frequent |
|---|---|
| Ice Tea | 24 |
| Ayam Bakar | 12 |
| Nasi Goreng Spesial | 9 |
| Katsudon | 9 |
| Thai Lychee Tea | 8 |
| Hot Crispy Chicken | 8 |
| Geprek Hejo | 5 |
| Latte | 5 |
| Ice Cream | 4 |
| Milkshake Chocolate | 4 |

3. Building Global Item Table

After eliminating items that do not meet the minimum support next is to provide a Global Id item on the menu that meets the minimum support by giving ascending numbering from the largest to the smallest occurrence frequency. This process is to facilitate the mapping process. Global item table can be seen in table 5.

**Table 5.** Global item table

| Nama_Menu | Frequent | Id Global Item |
|---|---|---|
| Ice Tea | 24 | 1 |
| Ayam Bakar | 12 | 2 |
| Katsudon | 9 | 3 |

| Nama_Menu | Frequent | Id Global Item |
|---|---|---|
| Nasi Goreng Spesial | 9 | 4 |
| Hot Crispy Chicken | 8 | 5 |
| Thai Lychee Tea | 8 | 6 |
| Latte | 5 | 7 |
| Geprek Hejo | 5 | 8 |
| Ice Cream | 4 | 9 |
| Milkshake Chocolate | 4 | 10 |

4. Mapping Data

Next is to replace the Nama_Menu with the index given to the transaction.

**Table 6.** Mapping data

| Bill | Nama_Menu |
|---|---|
| 181200004 | 8 |
| 181200004 | 1 |
| 181200004 | 6 |
| 181200005 | 4 |
| 181200005 | 3 |
| 181200005 | 1 |
| 181200006 | 2 |
| 181200006 | 1 |
| 181200007 | 5 |
| 181200007 | 2 |
| 181200007 | 1 |
| 181200010 | 2 |
| 181200010 | 1 |
| 181200012 | 3 |
| 181200012 | 1 |
| 181200013 | 4 |
| 181200013 | 10 |
| 181200015 | 3 |
| 181200015 | 1 |
| …………….. | … |
| 181200055 | 5 |

5. Building Global CFP-Tree

After data mapping is complete then the next step is to build Compressed Fp-Tree and enter each transaction data into the tree, CFP-Tree has a root that represents the Nama_Menu index with the highest occurrence ranking.
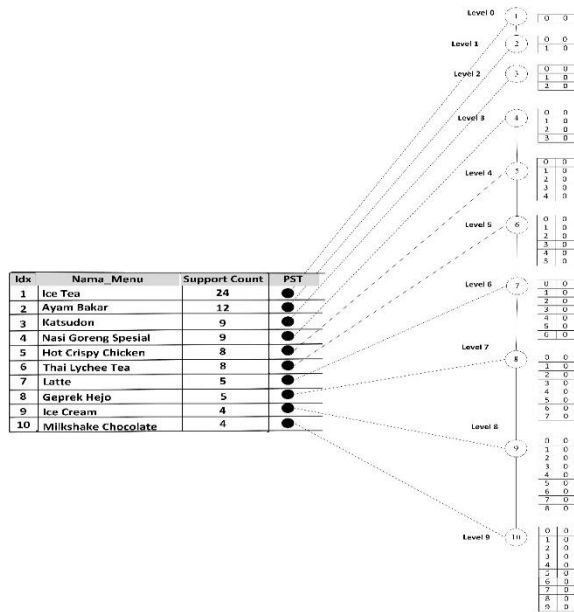
**Figure 3.** Leftmost Branch tree

In the figure above, it can be seen that each level item has two fields, namely parent-id and frequency or count. If the figure above is configured with two boxes to the right of the index on each level.

- Enter the transaction into the tree starting from the item that appears from the first transaction number that has been mapped. Count the number of occurrences of the item that appears according to the level or with the parent id.
- If the node already has children, then the item that appears next is the same as the child, then add the number of occurrences as much as 1 at the same level or parent.
- If the node already has children, but the item that appears next is different from the child of the node, then create a new node and add the number of occurrences of 1 at the same level or parent id, connect the new node to its parent with a line (nodelink).

i.  CFP-Tree bill 181200004

In bill 181200004 there are index 1, 6, 8. The first index to appear is 1 where index 1 is at level 0. The number of occurrences in index 1 will increase 1, create a new node with index 6 and index 8 as the offspring of index 1, add 1 to the number of occurrences or count.
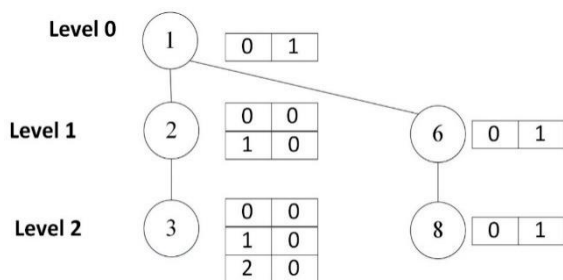


**Figure 4.** CFP-Tree bill 181200004

ii.  CFP-Tree bill 181200005

In bill 181200005 there are index 1, 3, 4. The first index to appear is 1 where index 1 is at level 0. The number of occurrences in index 1 will increase 1, creating a new node with index 3 and index 4 as the offspring of index 1. Increase the number of occurrences or count 1.
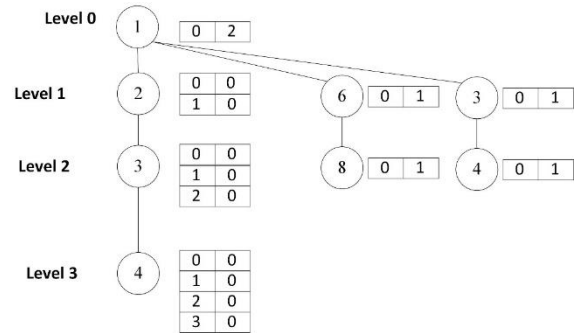


**Figure 5.** CFP-Tree bill 181200005

iii.  CFP-Tree bill 181200006

On bill 181200006 there are index 1, 2. The first index that appears is 1 where index 1 is at level 0. The number of occurrences in index 1 will increase by 1, the number of occurrences of index 2 at level 0 or parent-id 0 also increases by 1
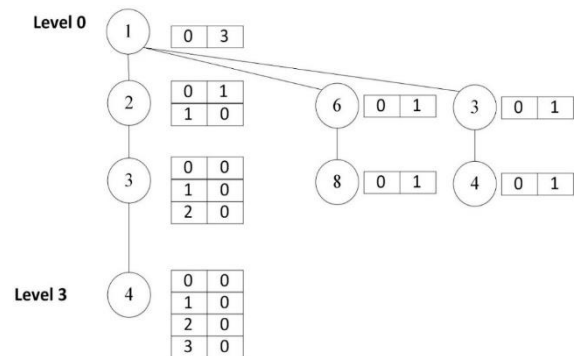


**Figure 6.** CFP-Tree bill 181200006

iv.  CFP-Tree bill 181200007

On bill 181200007 there are indexes 1, 2, 5. The first index that appears is 1 where index 1 is at level 0. The number of occurrences in index 1 will increase by 1, the number of occurrences of index 2 at level 0 or parent-id 0 will also increase by 1 Create a new node index 5 connected to index 2 as child of index 1 and increase the number of occurrences 1
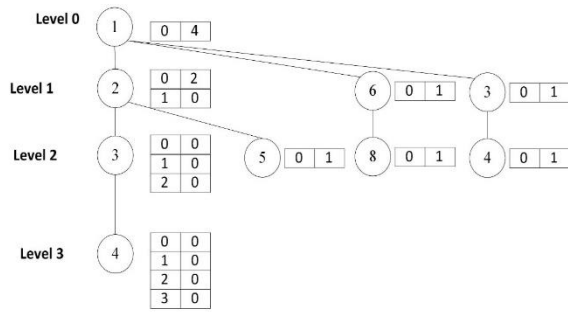
**Figure 7.** CFP-Tree bill 181200007

And the final results of forming the Global CFP-Tree are as follows:
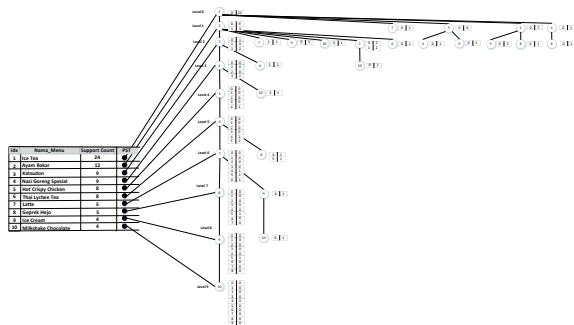


**Figure 8.** CFP-Tree last bill

6. Pembentukan Local CFP-Tree

From the Global CFP-Tree that has been built, the frequent pattern of each index is obtained by tracing the Global CFP-Tree ending in index n bottom-up from the most recent node or from the smallest frequency to the largest frequency or root.

Frequent Pattern:

**Table 7.** Frequent Pattern index

| Index | Frequent Pattern |
|---|---|
| 10 | 1,2 : 1. 1,2,3 : 1. 4 : 1. 7, 9 : 1 |
| 9 | 1,2 : 1. 1,3 : 1. 6,7,8 : 1. 7 : 1 |
| 8 | 1,4 : 1. 1,6 : 1. 5,6 : 1. 6 : 1. 6, 7 : 1 |
| 7 | 1 : 1. 2 : 1. 5,6 : 1. 6 : 1 |
| 6 | 1 : 1. 1,2 : 1. 3 : 2. 5 : 2 |
| 5 | 1,4 : 2. 1,2 : 2. 2 : 2 |
| 4 | 1 : 7. 1,3 : 1 |
| 3 | 1 : 6. 1,2 : 1 |
| 2 | 1 : 9 |

i. Local CFP-Tree index 10
• Nodes that meet the minimum support for frequent pattern index 10 will be entered into the local item table and a new index will be given to

each Nama_menu based on the number of occurrences.

**Table 8.** Local item table index 10

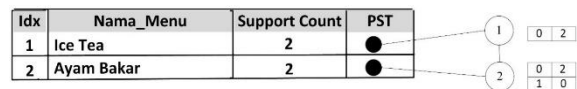| Idx local | Idx gobal | Nama_menu | Count |
|---|---|---|---|
| 1 | 1 | Ice Tea | 2 |
| 2 | 2 | Ayam Bakar | 2 |

• Mapping local item table into tree



**Figure 9.** Local CFP-Tree index 10

• Create a local CFP-Tree Projection where the index n will be root and each index in the Local item table index n will be the child of the index n
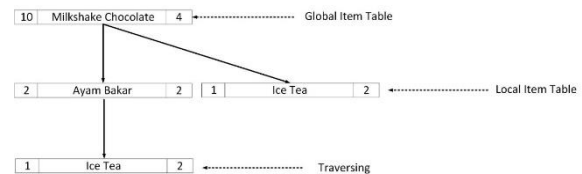


**Figure 10.** Local CFP-Tree Projection index 10

From the Local CFP-Tree projection, it is found that frequent items for Index 10 with Nama_Menu Milkshake Chocolate are as follows: (10,2:2), (10,2,1:2), (10,1:2).

ii. Local CFP-Tree index 9
• Nodes that meet the minimum support for frequent pattern index 9 will be entered into the local item table and a new index will be given to each Nama_menu based on the number of occurrences

**Table 9.** Local item table index 9

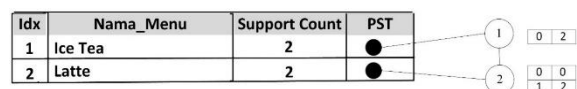| Idx local | Idx gobal | Nama_menu | Count |
|---|---|---|---|
| 1 | 1 | Ice Tea | 2 |
| 2 | 7 | Latte | 2 |

• Mapping local item table into tree



**Figure 11.** Local CFP-Tree index 9

• Create a local CFP-Tree Projection where index 9 will be root and every index in Local item table index 9 will be a child of index 9
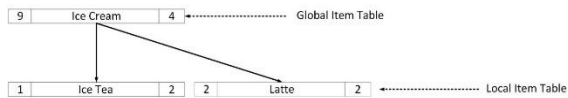
**Figure 12.** Local CFP-Tree Projection index 9

From the Local CFP-Tree projection, it is found that frequent items for Index 9 with Nama_Menu Ice Cream are as follows: (9,2:2), (9,1:2).
b.  Generate Rule

After creating a local cfp-tree a frequent itemset is generated for each menu:

**Table 10.** Frequent itemset

| Nama Menu | Frequent itemset |
|---|---|
| Milkshake Chocolate | • Milkshake Chocolate → Ice Tea<br>• Milkshake Chocolate → Ayam Bakar<br>• Milkshake Chocolate → Ayam Bakar, Ice Tea |
| Ice Cream | • Ice Cream → Ice Tea<br>• Ice Cream → Latte |
| Geprek Hejo | • Geprek Hejo → Thai Lychee Tea<br>• Geprek Hejo → Ice Tea |
| Latte | • Latte → Thai Lychee Tea |
| Thai Lychee Tea | • Thai Lychee Tea → Hot Crispy Chicken<br>• Thai Lychee Tea → Katsudon<br>• Thai Lychee Tea → Ice Tea |
| Hot Crispy Chicken | • Hot Crispy Chicken → Nasi Goreng Spesial<br>• Hot Crispy Chicken → Nasi Goreng Spesial, Ice Tea<br>• Hot Crispy Chicken → Ayam Bakar<br>• Hot Crispy Chicken → Ayam Bakar, Ice Tea<br>• Hot Crispy Chicken → Ice Tea |
| Nasi Goreng Spesial | • Nasi Goreng Spesial → Ice Tea |
| Katsudon | • Katsudon → Ice Tea |
| Ayam Bakar | • Ayam Bakar → Ice Tea |

From the frequent itemset obtained, will be generated to get the appropriate rule. Calculated with the confidence search formula:

$$Confidence(B \rightarrow A) = \frac{Jumlah\ transaksi\ mengandung\ A\ dan\ B}{Jumlah\ transaksi\ mengandung\ A} \quad (1)$$

The results of the generator can be seen in the following table:

**Table 11.** Generate Result

| Frequent itemset | Confidence |
|---|---|
| Milkshake Chocolate → Ice Tea | 2/4*100% = 50% |
| Milkshake Chocolate → Ayam Bakar | 2/4*100% = 50% |
| Milkshake Chocolate → Ayam Bakar, Ice Tea | 2/4*100% = 50% |
| Ice Cream → Ice Tea | 2/4*100% = 50% |
| Ice Cream → Latte | 2/4*100% = 50% |
| Geprek Hejo → Thai Lychee Tea | 4/5*100% = 80% |
| Geprek Hejo → Ice Tea | 2/5*100% = 40% |
| Latte → Thai Lychee Tea | 2/5*100% = 40% |
| Thai Lychee Tea → Hot Crispy Chicken | 2/8*100% = 25% |
| Thai Lychee Tea → Katsudon | 2/8*100% = 25% |
| Thai Lychee Tea → Ice Tea | 2/8*100% = 25% |
| Hot Crispy Chicken → Nasi Goreng Spesial | 2/8*100% = 25% |
| Hot Crispy Chicken → Nasi Goreng Spesial, Ice Tea | 2/8*100% = 25% |
| Hot Crispy Chicken → Ayam Bakar | 4/8*100% = 50% |
| Hot Crispy Chicken → Ayam Bakar, Ice Tea | 2/8*100% = 25% |
| Hot Crispy Chicken → Ice Tea | 4/8*100% = 50% |
| Nasi Goreng Spesial → Ice Tea | 8/9*100% = 88,89% |
| Katsudon → Ice Tea | 7/9*100% = 77,78% |
| Ayam Bakar → Ice Tea | 9/12*100% = 75% |

In taking the determination of the package produced, Kopi Cilik cafe has special limitations:
1.  Minimum Confidence = 60%
2.  Number of packaged menus 2 to 4 menus
3.  The range of package prices is Rp.15,000 to Rp.50,000.
4.  The package price is the total price of the menu combination minus Rp. 2,500

Then obtained the recommended package form:

**Table 12.** The resulting package

| Package | Menu | Price |
|---|---|---|
| Paket 1 | 1. Geprek Hejo<br>2. Thai Lychee Tea | Rp28.500 |

| Package | Menu | Price |
|---|---|---|
| Paket 2 | 1. Nasi Goreng Spesial<br>2. Ice Tea | Rp19.500 |
| Paket 3 | 1. Katsudon<br>2. Ice Tea | Rp17.500 |
| Paket 4 | 1. Ayam Bakar<br>2. Ice Tea | Rp18.500 |

### 2.3. Test Result

The following are the results of the implementation in the discussion above:



**Figure 13.** Test Result

## 3. CLOSING

Based on the results of the implementation and testing that has been done on the Data Mining application system with the association rule method, the following conclusions can be obtained:

1. The development of a data mining system with the association rule method can help the Kopi Cilik Cafe in implementing one of its business strategies, namely making packages in accordance with consumer purchasing patterns.

The suggestions for system development in this thesis research are as follows:

1. In retrieving data from the database, it should be able to retrieve certain period data so that it allows users to choose the desired data.
2. In forming packages, it is best to filter by menu categories, for example packages must be from different menu categories.

## BIBLIOGRAPHY

[1] H. F. Astuti, Data mining, Jakarta: Graha Ilmu, 2013.

[2] M. Brilliant, D. Handoko, and Sriyanto, "Implementation of Data Mining Using Association Rules for Transactional Data Analysis," Prosiding International conference on Information Technology and Business, 2018.

[3] E. Janitra and A. M. Bachtiar, " Penerapan Data Mining Dalam Perekomendasian Setelan Busana Muslim Dengan Metode Association Rules Menggunakan Algoritma CT PRO," KNTIA, 2017.

[4] Y. G. Sucahyo and R.P. Gopalan, "CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure," FIMI, vol. 4, pp. 212-223, 2004.

[5] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: Step-by-Step Data Mining Guide: The CRISP-DM Consortium, 2000.

[6] A. Kadir, Konsep Turunan Basis Data, Yogtakatya: Andi, 2001.

[7] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining: Instructor's Solution Manual. 2006.

[8] B. Santosa, Data Mining Teknik Pengumpulan Data Untuk Keperluan Bisnis, Yogyakarta: Graha Ilmu, 2007.

[9] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., San Francisco: Morgan Kaufmann Publishers, 2006.

[10] Y. G. Sucahyo and R.P. Gopalan, "High performance frequent patterns extraction using compressed FP-tree." Proceedings of SIAM International Workshop on High Performance and Distributed Mining (HPDM), 2004.