

Shopping Cart Analysis System in Product Layout Management with Apriori Algorithm

Komang Kurniawan Widiartha^{*1}, Desak Putu Diah Kumala Dewi²

^{1,2}Teknik Informatika, STMIK STIKOM Indonesia, Denpasar, Indonesia

e-mail: ^{*1}komang.kurniawan@stiki-indonesia.ac.id, ²desakdiah@stiki-indonesia.ac.id

Abstract

Coco Mart which is one of the retail network brands engaged in household needs is a company from PT. Bali Pawiwahan (Coco Group). As time goes on, transaction data continues to multiply and cannot be utilized by managerial parties to process transaction data that continues to grow every time. From transaction data, there is information or knowledge that can be extracted. The results of the research conducted, the construction of a shopping basket analysis system to process sales transaction data over the past four years, starting from 2013 with the aim of knowing consumer spending patterns. So that the Shopping Cart Analysis System can be a decisions support in business strategies to regulate the layout of goods and promote product combinations that are often purchased by Coco Mart Jimbaran customers.

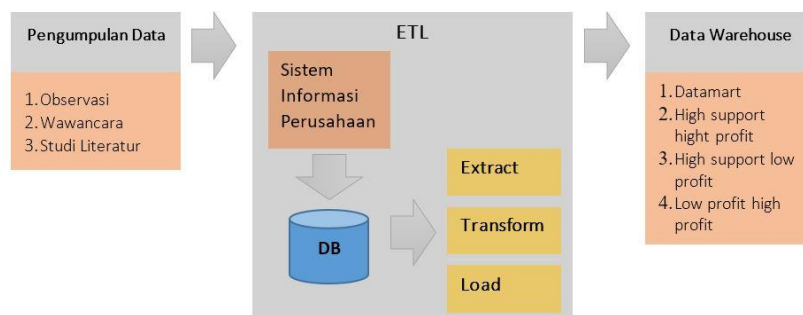
Keywords—Data Mining, Apriori Algorithm, Shopping Cart, Itemset, Profit

1. INTRODUCTION

Layout in product marketing at companies engaged in the sale of goods is one of the company's strategies in marketing that can determine the operational efficiency of the company, increasing profits in the long run. The layout of product marketing is related to the combination of products placed on the goods shelf so that there is a connection between the displayed goods and the consumers' needs. The right layout has an impact on the smooth flow of production up to the process of consumers buying products and has various strategic implications for the culprit. Besides that, the layout strategy brings competition between companies in terms of capacity sufficiency, a smooth process in the company to service to consumers. Various layout strategies are implemented by the company through the marketing division to create a more effective product layout. Starting from studying consumer shopping patterns to analyzing shopping receipts, marketing analyzes how the product layout should be. However, the amount of data with the support of existing information systems and support applications such as MS Excel have not been effective in raising information about how an effective layout strategy to be implemented. Information technology is needed that helps the company to be able to take product layout decisions by considering profits in a combination of these products. Based on the Association Rules theory that buyers who buy a certain item will tend to buy other goods in one transaction, with this information can help the managerial with the policy to place items that are often bought by buyers close together [1]. To help the process of analysis and decision making correctly need to use shopping cart analysis techniques. Shopping basket analysis technique is a technique that adapts the science of data mining. This technique is used to design a strategy for selling and marketing goods through the process of finding associations or relationships between data items from a relational database using Apriori algorithms.

2. RESEARCH METHODS

The data collection method used is observation and interview. In representation, the authors observe that the minimarket transaction data is processed by the Point of Sales system to determine the number of sales in a certain period. There is no stage of sales transaction analysis to find out consumer spending patterns as a basis for managing goods layout. This is also supported by interview data collection techniques, the authors conducted interviews with marketing managers with the results of the minimarket managerial wanted to optimize sales with a strategy to determine consumer spending patterns. By knowing the pattern of consumer spending, managerial can adjust the position of the shelf to make it easier for consumers to buy certain groups of goods with the theoretical basis is that consumers buy one item, it will tend to buy other groups of goods. With this strategy, managerial can sell one group of goods with certain promos (Cross Selling). The overall methodology is illustrated in the research methodology chart in Figure 1.

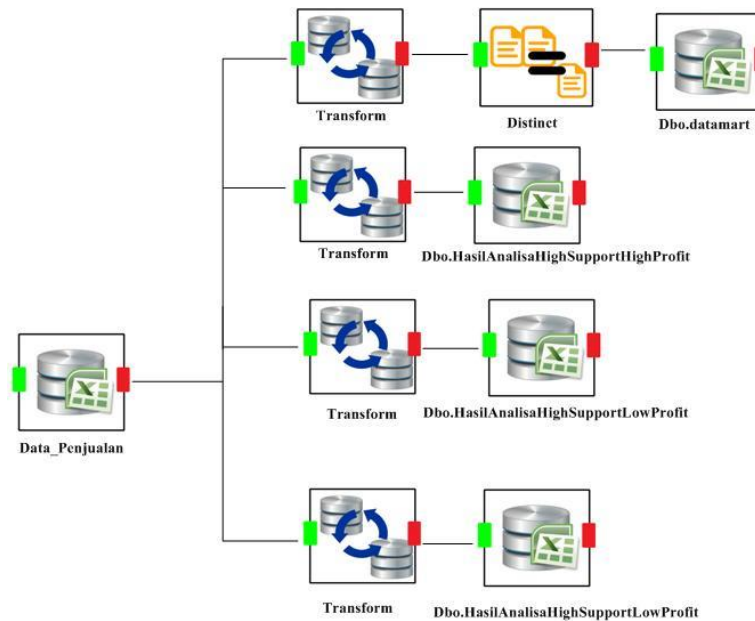


Gambar 1. Research method

2.1 Design of Extract, Transform, Load (ETL) Process

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and detailed knowledge from various large databases. besar [2]. The techniques used in data mining are Association Rules, Clustering, Description, Estimation, Prediction and Classification [3]. The terms data mining and knowledge discovery in databases (KDD) are often used interchangeably to describe the process of extracting hidden information in a large database. Actually the two terms have different concepts, but are related to each other. And one of the stages in the KDD process is *data mining* [4]. The KDD process includes *Data Selection*, *Pre-processing/Cleaning*, *Transformation*, *Data mining* dan *Interpretation/ Evaluation*.

ETL process is carried out to move transactional data from the data source table to the data warehouse. However, because the data retrieved from the source table there are formatting errors such as data redundancy and missing value, an ETL process is needed so that the data entered into the data warehouse complies with the standard star schema that has been designed. The result of the ETL process is the production of data that meets the data warehouse criteria that have been set in the transformation process such as naming fields, character length, data type [5]. The overall ETL process diagram can be seen in Figure 2.



Gambar 2. Diagram Proses ETL Secara keseluruhan.

2.2 Forming Tables in Datawarehouse

Data warehouse is a collection of data that has subject-oriented, integrated, time-variant, and fixed characteristics of data collection in support of the decision making process *management* [6]. This is reinforced by the theory from Han [7] which says that a data warehouse is a repository for information collected from various sources. Data warehouse schemes lie in 2 types of elements, namely facts and dimensions. Facts are used to store a measure and dimensions are used to analyze the measure through aggregation operations. The fact table contains data about a particular topic that is usually associated with numerical (measurable) values known as measure and has 2 or more foreign keys connected to the primary key in the dimension table. Dimension tables contain perspective data about an entity and are defined by a single primary key [8]. There are 3 types of data warehouse schemes, namely: Star schema, Snowflake schema and Fact constellation schema. The following is an example of the ETL process for forming tables in a data warehouse with the Star Schema scheme applied.

a. Tabel Datamart

The results of the ETL process from the Sales Data source table are datamart tables. In this table, transaction data has been extracted at the minimarket with a time span of 2013 to 2017, with the primary key InvoiceNo.

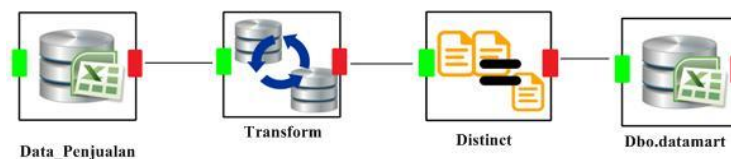


Figure 3. ETL Process Diagram Datamart table

b. High Support High Profit Analysis Results Table

In this table the results of basket analysis data with the sale of products that are often bought and have a high profit value will be stored in the High Support High Profit Analysis Table.

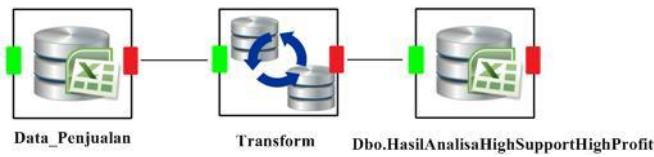


Figure 4. ETL Process Diagram Table of High Support High Profit Analysis Results

c. High Support Low Profit Analysis Results TableProfit

In this table the basket analysis data results with the sale of products that are often purchased but have a low profit value will be stored in the High Profit Low Low Profit Analysis Table.

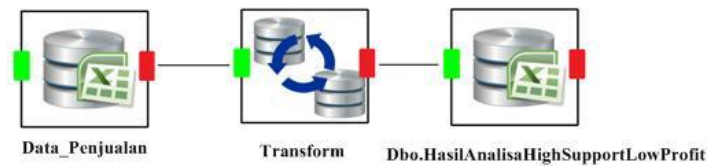


Figure 5. ETL Process Diagram Table of High Support Low Profit Analysis Results

d. Low Profit High Profit Analysis Results Table

In this table the results of basket analysis data with the sale of products that are rarely purchased but have a high profit value will be stored in the Analyst Low Support High Profittable.

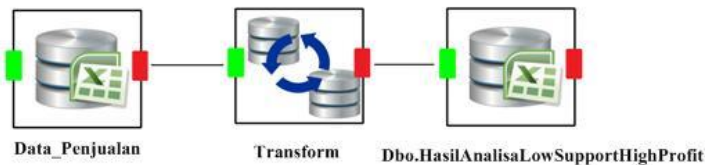


Figure 6. Diagram of ETL Process Table of Results of Analysis of Low Support High Profit Analysis

2.3 Data Processing With Apriori Algorithm

In determining the association rules the first thing to do is to find a high frequency pattern [9]. In this high frequency pattern search utility variable is added. Next is the application of a priori algorithm with the addition of utility variables. By setting the minimum support value = 50% and the utility variable profit = 10,000.

Table 1. Transaction Table

No. Transaksi	Number of Items sold on Transaction		
	Item A	Item B	Item C
1	1	2	0
2	2	1	0
3	2	1	1
4	0	1	2
5	2	0	1
6	1	0	0
7	0	1	0
8	1	0	0
9	1	1	1
10	1	1	0

Table 2. Item Benefits

Nama Item	Profit (IDR)
Item A	5000
Item B	7000
Item C	2000

First, a 1-itemset candidate is developed by calculating the support value and the profit value for each item.

$$\begin{aligned} \text{Support (A)} &= 8/10 \times 100\% = 80\% \\ \text{Profit (A)} &= 8 \times 5000 = 40.000 \end{aligned}$$

Then calculate the value of support and profit for items B and C. The results of the calculation of the value of support and confidence can be seen in table 3.

Table 3. Value *Support* and *Profit*

Item	<i>Support</i>	<i>Profit</i> (IDR)
A	80%	40.000
B	70%	49.000
C	40%	8.000

Then pruning is done for candidates who do not meet the minimum support and minimum profit.

Table 4. Reduction of Candidates

Item	<i>Support</i>	<i>Profit</i> (IDR)
A	80%	40.000
B	70%	49.000
C	40%	8.000

Tabel 5. K-Itemset collection

Item	<i>Support</i>	<i>Profit</i> (IDR)
A	80%	40.000
B	70%	49.000

In Table 5, it is necessary to build a set or combination of items with 2 items, and so on so that the next set of items cannot be raised.

$$\begin{aligned} \text{Support (AB)} &= 5/10 \times 100\% = 50\% \\ \text{Profit (AB,T1)} &= 1 \times 5000 + 2 \times 7000 = 21.000 \\ \text{Profit (AB)} &= \text{Profit(AB,T1)} + \text{Profit(AB,T2)} + \text{Profit(AB,T3)} + \text{Profit (AB,T9)} + \\ &\quad \text{Profit(AB,T10)} \\ &= (1 \times 5000 + 2 \times 7000) + (2 \times 5000 + 1 \times 7000) + (2 \times 5000 + 1 \times 7000) + (1 \times 5000 + \\ &\quad 1 \times 7000) + (1 \times 5000 + 1 \times 7000) \\ &= 21.000 + 17.000 + 17.000 + 12.000 + 12.000 \\ &= 79.000 \end{aligned}$$

Table 6. 2-Itemset collection

Item	<i>Support</i> (%)	<i>Profit</i> (IDR)
A,B	50%	79.000

No combination can be formed for the next k-itemset, the process stops, the frequent high pattern found in the item "A, B". at this time can create association rules that meet minimum requirements by calculating confidence association rules A-> B. Where to find Confidence is the number of transactions A and B amounted to 5 transactions divided by the

number of transactions containing A amounted to 8, then the value of Confidence A and B is $\frac{5}{8} \times 100\% = 62.5\%$.

Table 7. Asosiatif rule

Item	Support (%)	Confidence (%)	Profit (IDR)
A,B	50%	62,5%	79.000

3. RESULTS AND DISCUSSION

The analysis step is to understand the user's needs for the new system, which includes five main components, input, output, process, performance, and control. The system requirements checklist acts as a benchmark for measuring the success of applications built. The system requirements checklist it is necessary to understand the potential users. The following system requirements checklist :

Table 8. Requirement checklist

component	information
<i>Input</i>	Transaction Period Data
	Minimum value of Support
	Minimum Confidence value
	The minimum value of sales profit
<i>Output</i>	Relationship between related items
	Value of support for the connection between products
	Confidence value of the connection between products
	The Value of Profit for each related item
<i>Process</i>	Search for relationships between products with a priori algorithm
	Calculate the support value of each item
	Calculates the confidence value of each item
	Calculates the profit value of each item
<i>Perfomence</i>	The system can determine the relationship between goods based on the value of Support and profit
<i>Control</i>	The system provides security facilities
	Users who only have access rights can manage data

3.1 Perancangan Data Flow Diagram

Data flow diagrams are divided into context diagrams, zero diagrams and level one diagrams. DFD is used as a depiction of data flow in a system [10]. The context diagram of the Shopping Cart Analysis System in Product Layout Management in Minimarket with Apriori Algorithm is illustrated in Figure 7.

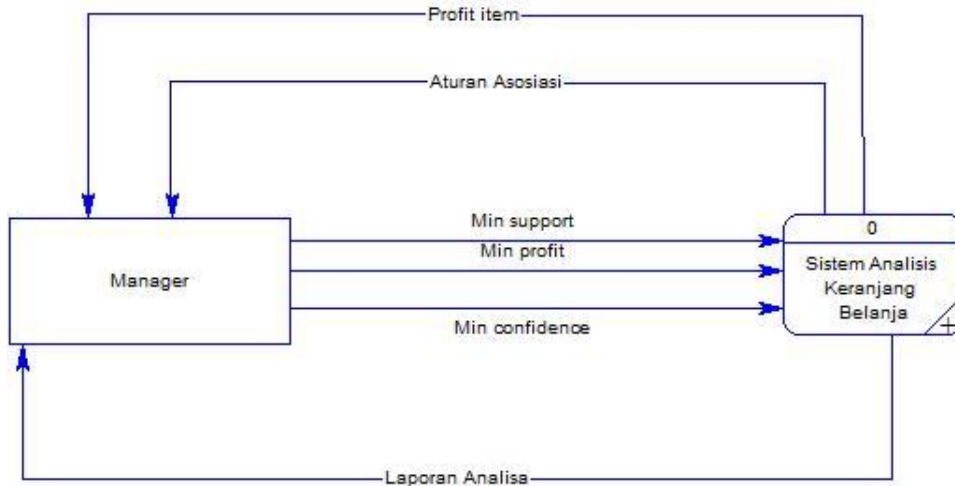


Figure 7. Context Diagram

From the context diagram, it is decomposed into a zero diagram. A diagram explains the overall process in the system. The zero diagrams illustrated in Fig 8.

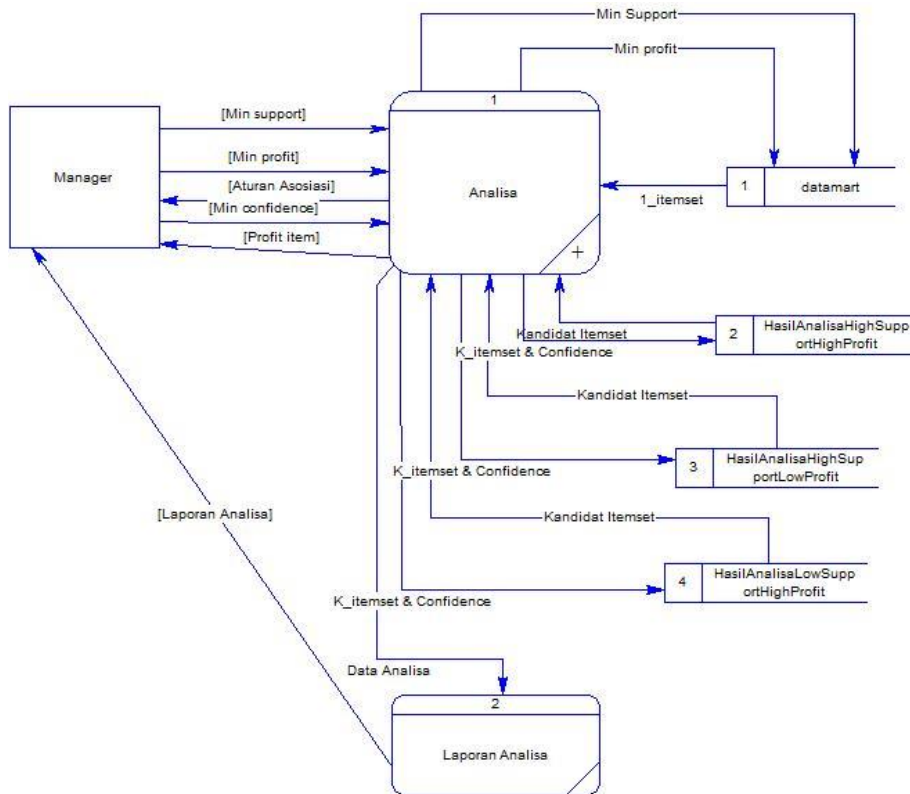


Figure 8. Zero Chart

In the zero diagrams, there are two processes among them analysis and report. The analysis process decomposed into the level 1 diagram of the analysis process illustrated in Figure 9.

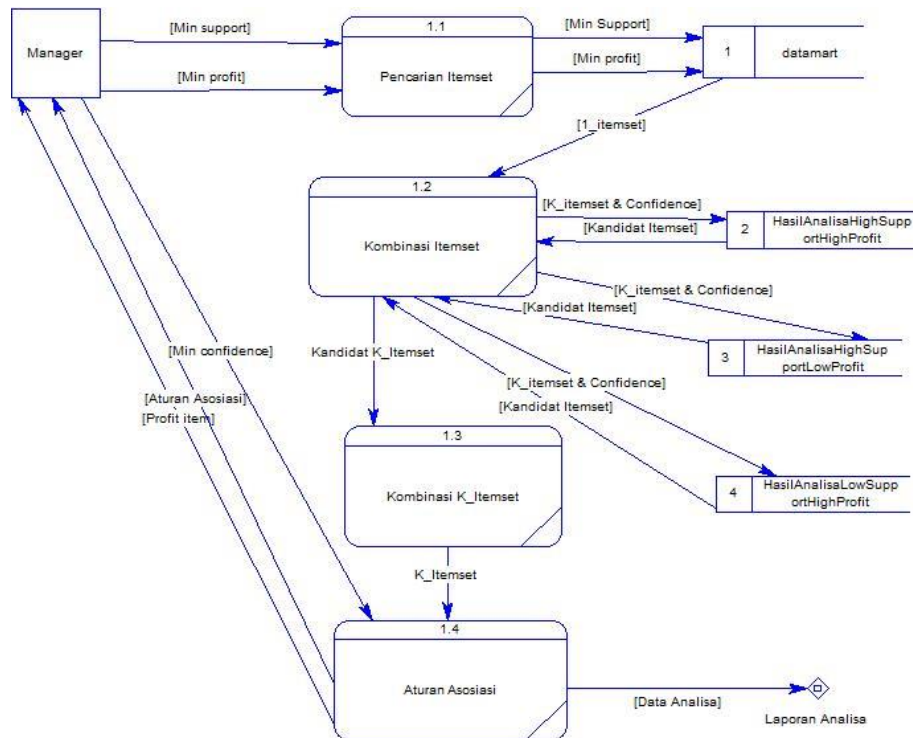


Figure 9. Level 1 Diagram Analysis Process

The purpose of the information that will explore in the data warehouse is "The combination of what products consumers buy and the value of profits". Based on the statement submitted, from the database of transactional records owned by Coco Mart Jimbaran, 4 (four) tables can be raised, as follows:

- a) Tb_Datamart
- b) Tb_HasilAnalisaHighSupportHighProfit
- c) Tb_HasilAnalisaHighSupportLowProfit
- d) Tb_HasilAnalisaLowSupportHighProfit

3.2 Database Design

To be used as a basis for calculations, the table must have at least 1 (one) field that functions as a measure. In connection with the information to be extracted, the measure calculated in the table is the "InvoiceNo" field to calculate the number of transactions in 4 years. Based on the results of determining these tables, the relationships between the tables designed in the form of a Star Schema shown in the figure 10.

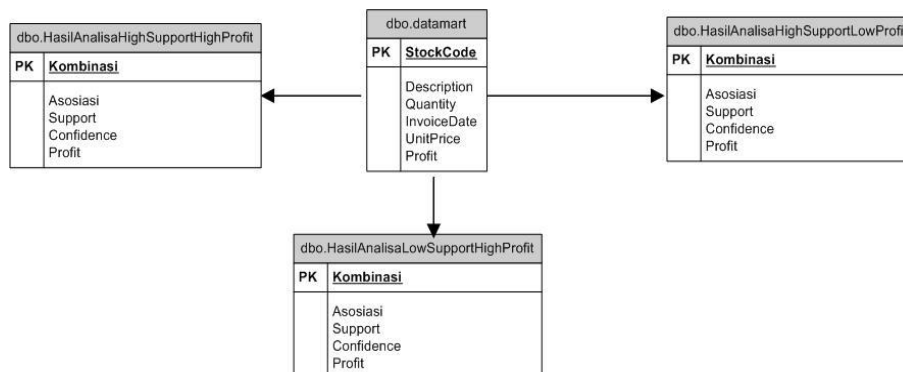


Figure 10. Rancangan Star Schema Database

To facilitate the user when the program is Execute, it will immediately display the Analysis Form. In the Analysis Form, the user must enter inputs in the form of Minimum Support, Minimum Confidence, Minimum Profit, and Sales Year Period. After the results of the analysis will immediately be grouped based on Support and Profit in each Tab. Following Is Display Analysis Form in the Shopping Cart Analysis System contained in Figure 11

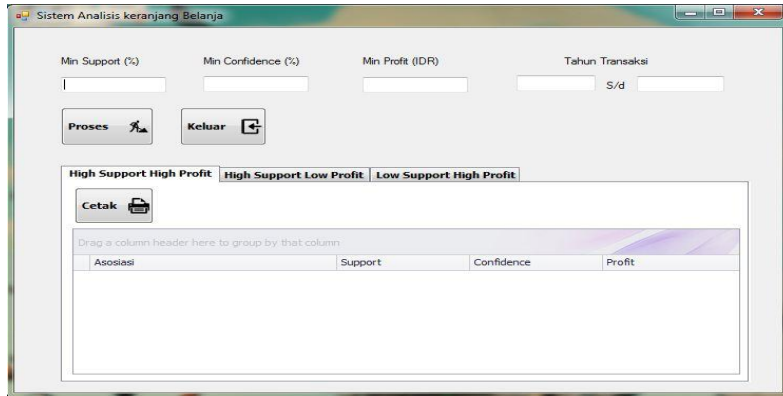


Figure 11. Shopping Cart Analysis System Analysis Form

When finished analyzing the data grid, the association will appear between the items. Support value which is the percentage value of the number of transactions containing a combination of articles on all transactions. The confidence value is the certainty of the strong relationship between item combinations, and the profit value is the profit of selling the combination items over four years.

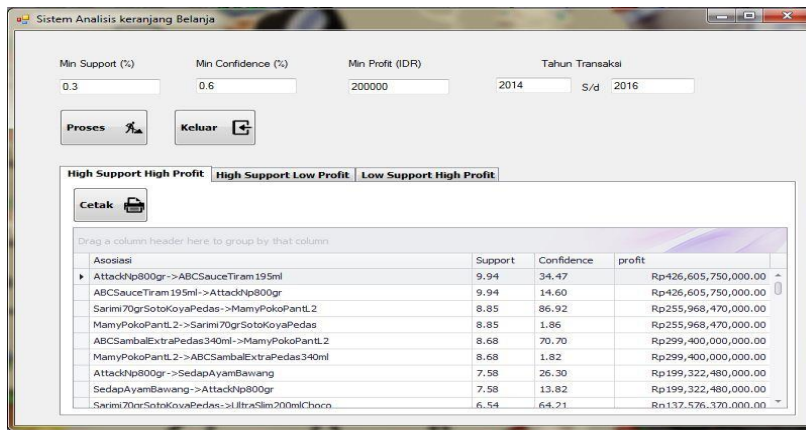


Figure 12. Results of the Analysis Process *High Support High profit*.

Then the other analysis results obtained by the System are Low Support High Profit which is a combination of itemset that has a small support value or which is rarely purchased simultaneously but has a significant profit value.

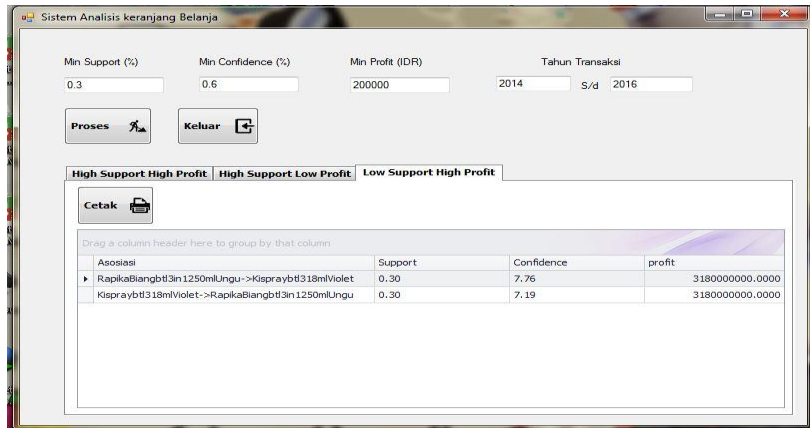



Figure 13. Analysis results *Low Support High Profit*

After conducting the analysis process, the user can print reports according to the results of the analysis. With reports, the managerial party can find out the association of purchases of goods and profit combinations of these associations made by consumers during the last four years.

 **Laporan Hasil Analisa High Support High Profit**

Asosiasi	Support	Confidence	profit
Ultra Slim200mlStoberi->AttackNp800gr	10.45	13.52	Rp 464.307.000.000
AttackNp800gr->Ultra Slim200mlStoberi	10.45	36.25	Rp 464.307.000.000
Ultra Slim200mlStoberi->SedapAyamBawang	11.67	15.10	Rp 308.534.770.000
SedapAyamBawang->Ultra Slim200mlStoberi	11.67	21.28	Rp 308.534.770.000
Ultra Slim200mlStoberi->TaroNet10grPotato	2.52	3.26	Rp 70.636.460.000
TaroNet10grPotato->Ultra Slim200mlStoberi	2.52	62.86	Rp 70.636.460.000
Ultra Slim200mlStoberi->RapikaBiangbtI3in 1250mlUngu	1.85	2.39	Rp 47.760.000.000
RapikaBiangbtI3in 1250mlUngu->Ultra Slim200mlStoberi	1.85	48.59	Rp 47.760.000.000
Ultra Slim200mlStoberi->ABCSauceTiram195ml	15.88	20.54	Rp 620.534.500.000
ABCSauceTiram195ml->Ultra Slim200mlStoberi	15.88	23.34	Rp 620.534.500.000
Ultra Slim200mlStoberi->Sarimi70grSoto KoyaPedas	5.49	7.10	Rp 149.199.170.000
Sarimi70grSoto KoyaPedas->Ultra Slim200mlStoberi	5.49	53.91	Rp 149.199.170.000
Ultra Slim200mlStoberi->Rinso Molto Ultra800gr	21.43	27.72	Rp 3.233.404.000.000

Figure 14. Analysis Report *High Support High Profit*

Analysis report is important for managerial parties to support future decision making.

 **Laporan Hasil Analisa Low Support High Profit**

Asosiasi	Support	Confidence	profit
RapikaBiangbtI3in 1250mlUngu->KispraybtI318mlViolet	0.30	7.76	Rp 3.180.000.000
KispraybtI318mlViolet->RapikaBiangbtI3in 1250mlUngu	0.30	7.19	Rp 3.180.000.000

Figure 15. Analysis Report *Low Support High Profit*

3.3 Association Rule Results

Some other association rules that are formed can be seen in Table 9.

Table 9. *Output Association*

No	Association	Support (%)	Confidence (%)	Profit (IDR)
1	MoltoUltra800gr -> Mamy Poko Pant L2	60,8	81	512.195.000
2	ABCSambalManis Pedas -> SunMarieBisc	55,4	75	294.675.000
3	UltraSlim200ml -> SelamatWfr60gr	51	75	134.772.00
4	ABCSausTiram195ml -> GaruaKatom200gr	48	72	133.234.000

Table 10. Results of Respondents' Answers

Responden (R)	Question (P)										Total
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
R1	4	4	4	3	3	4	5	4	4	4	39
R2	3	4	3	2	2	3	3	3	4	4	31
R3	4	4	4	3	2	4	4	4	4	4	37
R4	5	4	5	4	3	4	4	4	4	5	42
R5	3	3	4	3	2	3	4	3	3	3	31
R6	3	4	3	3	1	4	3	4	4	4	33
R7	4	5	5	4	2	3	4	3	3	4	37
R8	4	5	4	4	3	4	4	4	4	4	40
R9	4	4	4	4	2	4	3	4	3	3	35
R10	4	4	4	4	4	4	4	4	4	4	40
R11	4	5	5	4	4	4	4	4	4	5	43
R12	4	4	4	4	2	3	3	3	3	3	33
R13	4	5	5	4	1	3	4	3	3	4	36
R14	4	4	4	3	2	4	2	4	4	4	35
R15	4	4	4	4	2	4	3	4	4	4	37
R16	4	4	4	4	3	4	3	4	4	4	38
R17	4	3	4	4	2	4	3	4	4	3	35
R18	5	4	5	5	3	4	5	4	3	4	42
R19	5	4	4	3	2	3	4	3	3	4	35
R20	4	3	4	3	2	3	3	4	4	4	34
R21	4	4	4	3	2	4	4	3	4	4	36
R22	4	3	4	4	3	4	4	4	4	4	38
R23	5	4	4	3	2	4	4	4	4	4	38
R24	4	3	4	4	3	4	3	2	3	4	34
R25	4	4	4	4	2	3	3	4	4	4	36
R26	4	4	3	3	2	4	4	5	5	5	39
R27	4	3	4	4	3	4	4	4	4	4	38
R28	4	4	4	3	2	3	4	4	4	4	36
R29	4	4	4	4	3	4	4	5	5	5	42
R30	4	3	4	4	3	4	4	4	4	4	38
Korelasi	0.56	0.34	0.48	0.53	0.68	0.55	0.57	0.57	0.39	0.67	
tHitung	3,61	1,92	2,97	3,35	4,93	3,51	3,69	3,69	2,25	4,79	
tTabel (95%,28)	2,04										
Question	Valid	Not Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	
Amount Valid	9										

3.4 Validity Results

In testing the validity of the authors conducted a questionnaire against 30 customers of Coco Mart Jimbaran, the purpose of this validity test can reflect current Managerial performance and improve the performance of Coco Mart Jimbaran in the future, which is expected to assist the managerial in making decisions with the Shopping Cart Analysis System. The following is a list of questionnaire questions addressed to Coco Mart Jimbaran customers.

Following Table 10 are the results of a questionnaire from 30 Coco Mart Jimbaran customers whose authors summarized in tabular form.

4. CONCLUSION

Hasil The results of the association rules formed are the results after running the Shopping basket Analysis System. In this test, it provides a minimum transaction limit value (support) of 30% of transactions and a minimum confidence of 70% and a minimum profit of Rp.400,000 in the transaction period from 2013 to 2017, as many as 110 association rules from 23,345 sales transactions to High Support High Profit Analysis results, while no association was formed for the High-Profit Low Profit and Low Support High-Profit analysis results. One of the association rules formed is if you buy Rinso Molto Ultra 800gr, then buy Mamy Poko Pant L2 with a support value = 60.8% from 23,345 transactions and a confidence value = 81% and a profit of Rp.512,195,000 which is a rule with a support value and highest profit.

In testing validity refers to an instrument in carrying out its functions, an instrument is said to be valid if it can be used to measure what is being measured. In this research, the instrument to be measured is the price of Coco Mart Jimbaran, Service Quality and Customer Satisfaction.

REFERENCES

- [1] E. Widiati and K. E. Dewi, "Implementasi Association Rule Terhadap Penyusunan Layout Makanan Dan Penentuan Paket Makanan Hemat Di Rm Roso Echo Dengan Algoritma Apriori, Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)," vol. 3, no. 2, 2014.
- [2] K. K. -Vinay Singh, "Data Mining and Knowledge Management," *Int. Res. J. Eng. Technol.*, vol. 4, no. 3, pp. 1–281, 2017.
- [3] N. Nurdin and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe," *TECHSI - J. Tek. Inform.*, vol. 7, no. 1, Apr. 2015.
- [4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework.," *Int Conf Knowl. Discov. Data Min.*, pp. 82–88, 1996.
- [5] R. Wijaya, "Penerapan Extraction-Transformation-Loading (ETL) Dalam Data Warehouse (Studi Kasus : Departemen Pertanian)," vol. 5, pp. 61–75, 2016.
- [6] K. Khotimah, "Perancangan Dan Implementasi Data Warehouse Untuk Mendukung Sistem Akademik (Studi Kasus Pada Stkip Muhammadiyah Kotabumi)," vol. 02, no. 01, pp. 94–107, 2016.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [8] Y. Sekip, U. Iv, and Y. Indonesia, "Implementasi Big Data Pada Data Transaksi Tiket Elektronik Bus Rapid Transit (BRT)," pp. 370–376, 2017.
- [9] H. N. Wulandari, "Pemanfaatan Algoritma Apriori untuk Perancangan Ulang Tata Letak Barang di Toko Busana," 2014.
- [10] D. S. Le Vie, "Understanding Data Flow Diagrams.," 2014.