

IMPLEMENTATION OF DATA MINING TO ANALYZE ONLINE SHOP SALES DATA WIMASK_ID

Yohanni Syahra¹, Firahmi Rizky², Hevlie Winda Nazry S³, Rina Mahyuni⁴, Putri Maharani Sebayang⁵

^{1,2,3,4,5}Universitas Muhammadiyah Sumatera Utara

^{*1}*email: yohannisyahra@umsu.ac.id*

Abstract: Wimask_id is a shop that sells various brands of skin care products and online souvenirs, the online shopping method that is currently popular with the public makes the sales data owned by online business actors increasingly piling up, including the Wimask_id online shop. The pile of data is unfortunate if it is not used as a source of information for the analysis process, but the problem is how the Wimask_id online shop can analyze the pile of data. To overcome this problem, the database in the form of sales data owned by the Wimask_id online shop can be analyzed, to obtain and explore consumer needs patterns through data mining information using the FP-Growth algorithm, to find a set of information that often arises in a large collection of information in optimizing product sales. In this study, the implementation of data mining using the FP-Growth algorithm is proposed. The results of this data mining analysis will help Wimask_id in determining the number and types of products that consumers are interested in, so that it can increase revenue and provide product offerings that suit consumer needs.

Keywords: Data Mining, FP-Growth, Sales, Online Shop.

Introduction

Trading industry the more develop with data transaction which increasing in number all the time. This development requires continuity existing companies to be able to compete with existing companies develop And to obtain profit [1]. Can't denied that the use of information technology has become a necessity considering the rapid development and competition business retail on moment this for give service marketing which best [2]. Wimask_id is shop on line which get involved in field sale product maintenance skin and craft hand through media social, wimask_id sell various type brand product maintenance skin and some handicrafts. Wimask_id has abundant transaction data, however the data only used as an archive. In fact, the sales data that this abundance can be used as an object of analysis to determine patterns. consumer needs, which can be useful in organizing products. Which most lots sold and each other relate [3]. Therefore, in analyzing sales data on the Wimask_Id online shop, a series of processes are needed that can produce conclusions. Data mining or also called Knowledge Discovery in Database (KDD) is the process of obtaining information through collection and use activities past data to find patterns or relationships in large-scale data. large [4]. Techniques that can be used to determine product sales patterns Wimask_id is an association technique, where this technique can process data to knowing the relationship between the products ordered [5]. Therefore, the technique that is used will used on work scientific this is association with use algorithm FP-Growth [6].

Literature Review

Data Mining

Data mining is a series process for dig information beneficial or pattern important and interesting which can used for take a conclusion from data which the amount big, with involves one or more computer learning techniques to analyze and extract knowledge in a way automatic or a series process for digging up added value from a collection of data in the form of knowledge that has been this can't known with technique manual [7] . Based on the definition that has been explained previously, there are things important related with data mining as following (Nurkholis & Anjumi, 2021) :

1. Data mining is a process automatic to data which already there is.
2. Data which will processed is data with amount which is very big.
3. Objective data mining is for get connection or pattern which possible will give indication which beneficial.

Knowledge Discovery in Database (KDD) is name other from data mining although In fact, these two terms have different concepts, but both are interconnected with each other and one of the stages the KDD process is data mining which become the core of the KDD process [3] . there is a number of process from stage operate KDD that is as following [9] :

1. Data Selection

Data selection from a set of operational data needs to be done when the KDD mining stage begins. The selected data used for the data mining process is stored in a separate file in the operational database.

2. Election Data

Process preprocessing is a process which covering between other delete or remove duplicate data that is not used, inspect data that is considered inconsistent, and correcting errors in the data, like error authorship (typography).

3. Transformation

On phase this done change a form data Which Not yet have a number of entity which clear to in data which ready for done process data mining.

4. Data Mining

On process this which done is implementation algorithm or method re search knowledge from data which produced in the process transformation.

5. Interpretation/Evaluation

On phase which most end this process which done is process to form an output or result that is easy to understand and in the form of a information which beneficial.

Sales

Sales is a normal economic activity where business owners receive the planned proceeds or benefits from the sale or acquisition cost which issued. In sale there is activity purchase a goods or services from one party to another in exchange for money from one party. party [10] . Philip Kotler give the definition of sales as an activity aimed at getting buyers, influencing and providing guidance so that buyers can adjust their needs to the products offered and make purchases. agreement regarding a price that is profitable for both parties [11] . Strategy sale is series activity which organized for achieve goals with a creative and innovative mind wet, to face tendencies that occur in companies and outside companies influential to time front company [12] .

FP-Growth Algorithm

Frequent Pattern Growth (FP-Growth) is an algorithm from one of the method in data mining, namely Association Rule, which is a procedure for look for connection between Items in one dataset [6] . FP-Growth is an alternative algorithm that can be used to find the most frequent data sets. appear (frequent itemset) in a gathering data [13] . Algorithm FP-

Growth is development from algorithm A Priori, on algorithm FP-Growth use draft tree construction, commonly called FP-Tree, in frequent itemset search no using generate candidate like what which done algorithm A Priori. With using draft the algorithm FP-Growth become more fast than algorithm A Priori [14] . Frequent itemset mining using algorithms FP-Growth is done by generating a tree data structure (FP-Tree) [15] . FP-Tree is a compressed data storage structure, FP-Tree is built by mapping each transaction data into each path particular in FP-Tree. Because in each mapped transaction, there may be there are transactions that have the same items, then the path is possible to overwrite each other. The more transaction data that has items that the same, then the compression process with the FP-Tree data structure is more effective [3].

Method

Research Stages

There are two techniques in data collection in this study, namely observation and interviews as follows:

a. Observation

Before conducting further research, observation activities have been carried out to find out what problems and obstacles occur in relation to determining sales patterns in the Wimask_id online shop.

b. Interview

The data collection method involves conducting direct Q&A with sources from the research object to obtain the required information.

c. Literature review

This study uses 15 articles as reference materials. It is expected that these references can solve problems related to finding consumer needs patterns in sales data.

Results and Discussion

Implementation of FP-Growth Algorithm

Data analysis in the case of analyzing sales data on the Wimask_Id online shop, is carried out based on the association rule technique using the FP-Growth algorithm with several iterations or steps. The following are the stages of implementing the FP-Growth algorithm

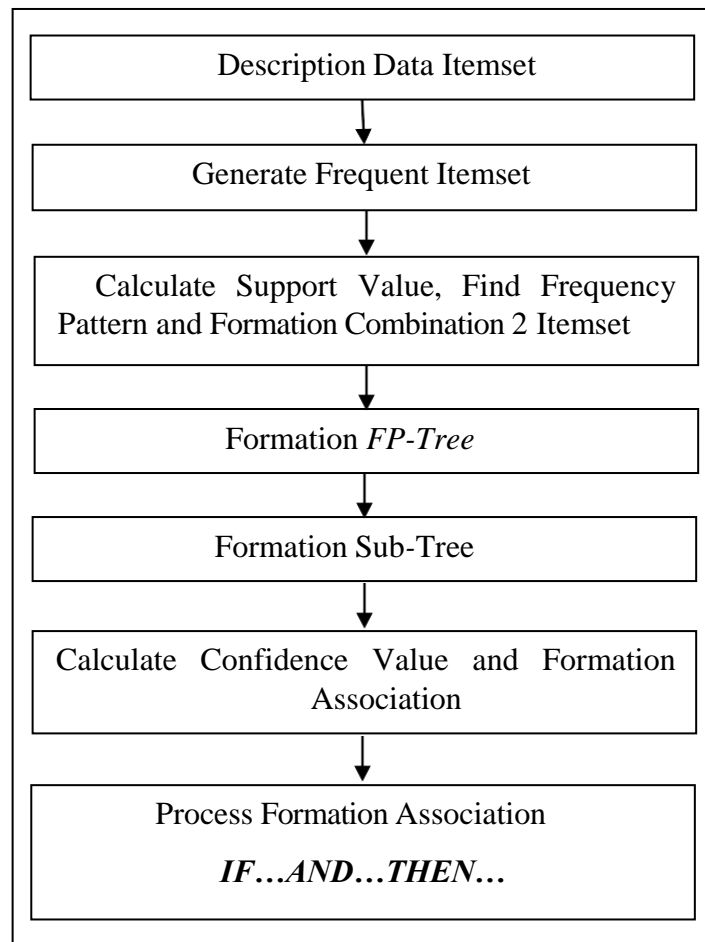


Figure 1. FP-Growth Algorithm Framework

Problem Analysis

In analyzing sales data on the Wimask_Id online shop, transaction data is needed, so that from the transaction data analysis process, consumer needs patterns can be identified which are useful for determining the best sales strategy.

Table 1. Transaction Data

Date	Name Goods
02-Jan-23	Expo Cleora, Cleo Intensive, Fancy
01-Feb-23	Bouquet Flower, Body Serum Orange, Body Serum Apple, Body Serum Raspberry
02-Feb-23	Bouquet Flower, Sleeping Mask, Nelpamara
05-Feb-23	Sleeping Mask, Bento Cake, Kojic Plankton
08-Feb-23	Body Serum orange, Body Serum Apple, Body Serum Raspberry
08-Feb-23	Body Serum orange, Body Serum Apple, Bento Cake
12-Feb-23	Body Serum Apple, Body Serum Orange, Body Serum Raspberry
15-Feb-23	Kojic Plankton, Body Serum Orange, Tamanu
02-Mar-23	Fancy, Sheet Mask, Paintbrush
02-Mar-23	Moist Origin, Sunscreen Madam Gie, Face Wash Mugworth, Sheet Mask
05-Mar-23	Serum Mayra, Kojic Plankton, Sunscreen Azarine
05-Mar-23	Kojic Plankton, Moist Ceramon, Sunscreen Azarine
06-Mar-23	Fachwash Origin, Moist Ceramon, Sunscreen Azarine
07-Mar-23	Face Mask Camille, Sheet Mask, Blink Soap
11-Mar-23	Serum Mayra, Sunscreen Azarine, Dorskin, Sheet Mask
12-Mar-23	Herborist Mix, Brylee Blackheads, Moist Origin

14-Mar-23	Moist Hey, Expo Cleora, Jelly Tint Implora
22-Mar-23	Sunscreen Skin Aqua, Sheet Mask, Maskom Chocolate, Lip Tint Dazzle Me
01-Apr-23	Sunscreen Madam Hello, Salsa Hansjebb, Lip Crayon Implora
01-Apr-23	2 in 1 Lashes and Bro, Moist Ceramon, Sunscreen Madam Gie
01-Apr-23	Maskom Charcoal, Kojic Plankton, Face Wash Origin
01-Apr-23	Jelly Tint Implora, Lip Serum Xiuxiu, Sunscreen Azarine
01-April-23	Kojic Plankton, Blinkz Soap, Salsa Hansjebb, Nalpamara
01-April-23	Salsa Hansjebb, Kojic Plankton, Lip Tint Implora
01-April-23	Sunscreen Madam Hey, Maskon Charcoal, Expo Cleo
03-April-23	Lip Serum XiuXiu, Salsa Hansjebb, Sheet Mask
03-April-23	Salsa Hansjebb, Kojic Plankton, Implora Lip Tint
03-April-23	Peeling Gel Girlpal, Animate Blue, Sunscreen Madam Gie
15-April-23	Animate Pink, Lip Crayon Implora, Jelly Tint Implora
19-April-23	Marina, Sunscreen Madam Gie, Scarlet 7x Ceramide

Itemset Data Description

To simplify the fp-growth process, products are first coded into the following product codes:

Table 2. Wimask_Id Online Shop Product Data

No	Code Product	1 Itemset	Frequent
1	KP	Kojic Plankton	8
2	SA	Sunscreen Azarine	6
3	SM	Sheet mask	6
4	SMG	Sunscreen Madan Gie	6
5	BSJ	Bodey Serum orange	5
6	BSA	Bodey Serum Apple	4
7	SH	Salsa Hansjebb	4
8	BSR	Body Serum Raspberry	3
9	E.C.	Cleora Expo	3
10	JTI	Jelly Tint Implora	3

Next, the product code will be changed in the transaction data to make the FP-Growth process easier.

Table 3. Converted Transaction Data

Date	Data Sale Goods
2-Jan-23	EC, CI, FY
1-Feb-23	BB, BSJ, BSA, BSR
2-Feb-23	BB, SLM, NA
5-Feb-23	SLM, BC, KP
8-Feb-23	BSJ, BSA, BSR
8-Feb-23	BSJ, BSA, BC
12-Feb-23	BSA, BSJ, BSR
15-Feb-23	KP, BSJ, TU
2-Mar-23	FY, SM, KS
2-Mar-23	MO, SMG, FWM, SM
5-Mar-23	SENIOR HIGH SCHOOL, KP, SA
5-Mar-23	KP, MOC, SA
6-Mar-23	FWO, MOC, SA
7-Mar-23	FMC, SM, BS
11-Mar-23	SM, SA, DN
12-Mar-23	HM, BBD, MO
14-Mar-23	MOE, EC, JTI
22-Mar-23	SKA, SM, MCT, LDM
1-Apr-23	SMG, SH, LCI
1-Apr-23	LAB, MOC, SMG
1-Apr-23	MC, KP, FWO

1-Apr-23	JTI, LSX, SA
1-Apr-23	KP, BS, SH, NA
1-Apr-23	KP, SENIOR HIGH SCHOOL, SA
1-Apr-23	SH, KP, LI
3-Apr-23	SMG, MC, E.C.
3-Apr-23	LSX, SH, SM
3-Apr-23	PGG, AB, SMG
15-Apr-23	AP, LCI, JTI
19-Apr-23	MA, SMG, SCE

Generate Frequent Itemset

Based on table 3.3, the minimum support will be searched from 30 transaction data using the following formula:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}} \times 100\%$$

support value will be obtained as in table 3.4 as follows:

Table 4. Frequency of occurrence of each item

No	Code Product	1 Itemset	Frequent	Support
1	KP	Kojic Plankton	8	27%
2	SA	Sunscreen Azarine	6	20%
3	SM	Sheet mask	6	20%
4	SMG	Sunscreen Madan Gie	6	20%
5	BSJ	Bodey Serum orange	5	17%
6	BSA	Bodey Serum Apple	4	13%
7	SH	Salsa Hansjebb	4	13%
8	BSR	Body Serum Raspberry	3	10%
9	E.C.	Expo Cleo	3	10%
10	JTI	Jelly Tint Implora	3	10%
11	MOC	Moist Ceramon	3	10%
12	BB	Bouquet Flower	2	7%
13	BC	Bento Cake	2	7%
14	BS	Blink Soap	2	7%

Based on table 4 containing the minimum support, the minimum support value is set at 8% as seen in table 5 below:

Table 5. Items that meet minimum support

No	Code Product	1 Itemset	Frequent	Support
1	KP	Kojic Plankton	8	27%
2	SA	Sunscreen Azarine	6	20%
3	SM	Sheet mask	6	20%
4	SMG	Sunscreen Madan Gie	6	20%
5	BSJ	Bodey Serum orange	5	17%
6	BSA	Bodey Serum Apple	4	13%
7	SH	Salsa Hansjebb	4	13%
8	BSR	Body Serum Raspberry	3	10%
9	E.C.	Expo Cleo	3	10%
10	JTI	Jelly Tint Implora	3	10%
11	MOC	Moist Ceramon	3	10%

Based on table 5, the transaction data that meets the minimum support and will be continued to the FP-Tree formation stage is as follows:

Table 6. Items that meet minimum support

Date	Data Sale Goods
2-Jan	E.C.
1-Feb	BSJ, BSA, BSR
2-Feb	-
5-Feb	KP
8-Feb	BSJ, BSA, BSR
8-Feb	BSJ, BSA
12-Feb	BSJ, BSA, BSR
15-Feb	KP, BSJ
2-Mar	SM
2-Mar	SM, SMG
5-Mar	KP, SA
5-Mar	KP, SA, MOC
6-Mar	SA, MOC
7-Mar	SM
11-Mar	SA, SM
12-Mar	-
14-Mar	EC, JTI
22-Mar	SM
1-Apr	SMG, SH
1-Apr	SMG, MOC
1-Apr	KP
1-Apr	SA, JTI
1-Apr	KP, SH
1-Apr	KP, SA
1-Apr	KP, SH
3-Apr	SMG, E.C.
3-Apr	SM, SH
3-Apr	SMG
15-Apr	JTI
19-Apr	SMG

Formation of FP-Tree

The image below provides an illustration of the formation of an FP-Tree based on products that have met minimum support.

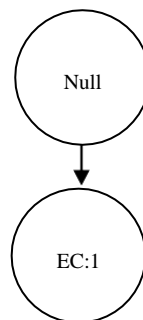


Figure 2. Results Formation FP-Growth Reading TID T1

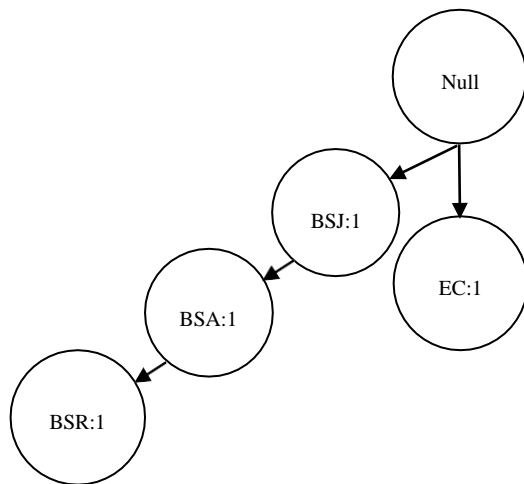


Figure 3. Results Formation FP-Growth Reading TID T2

Sub Formation Tree

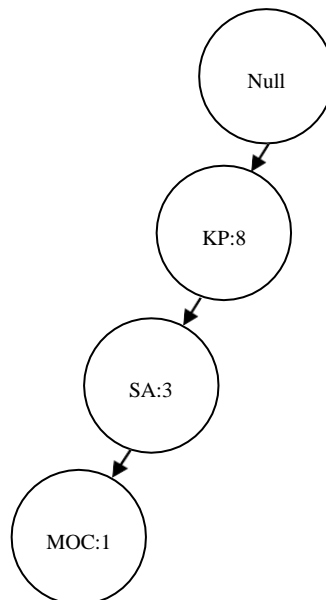


Figure 4. Tracks which contain knot MOC

- a. Conditional Pattern Base
MOC = {KP, SA : 1}
- b. Conditional FP-Tree
MOC = {KP : 8, SA : 3}
- c. Conditional Frequency Itemsets
MOC = {KP, MOC : 8}, {SA, MOC : 3}, {KP, SA, MOC : 1}

Association Rules

After find sub set which sufficient condition, furthermore will obtained frequency value according to sub set. Following this is table frequent pattern.

Table 7. Frequent Pattern

No	Frequent Items set		Frequency
	1	KP	
2	SA	MOC	1
3	SA	JTI	1
4	SMG	E.C.	1
5	BSJ	BSR	3
6	BSA	BSR	2
7	KP	SH	2
8	BSJ	BSA	3
9	KP	BSJ	1
10	SM	SMG	1
11	SA	SM	1
12	KP	SA	3

At this stage, it is used to determine the t support and confidence values using the following formula:

$$\text{Support } (A \cap B) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}} \times 100\%$$

$$\text{Confidence}(A \rightarrow B) = P(A|B) \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Transaksi Mengandung A}} 100\%$$

By using the formula above, researchers can determine the support and confidence values for each path.

Table 8. Mark Support and Confidence

Item A	Item B	Amount Transaction	Support	Transaction Item A	Confidence
KP	MOC	1	3%	8	13%
MOC	KP	1	3%	1	33%
SA	MOC	1	3%	6	17%
MOC	SA	1	3%	3	33%
SA	JTI	1	3%	6	17%
JTI	SA	1	3%	3	33%
SMG	E.C.	1	3%	6	17%
E.C.	SMG	1	3%	3	33%
BSJ	BSR	3	10%	5	60%
BSR	BSJ	3	10%	3	100%
BSA	BSR	2	7%	4	50%
BSR	BSA	2	7%	3	67%
KP	SH	2	7%	8	25%
SH	KP	2	7%	4	50%
BSJ	BSA	3	10%	5	60%
BSA	BSJ	3	10%	4	75%

KP	BSJ	1	3%	8	13%
BSJ	KP	1	3%	5	20%
SM	SMG	1	3%	6	17%
SMG	SMG	1	3%	6	17%
SA	SM	1	3%	6	17%
SM	SA	1	3%	6	17%
KP	SA	3	3%	8	38%
SA	KP	3	3%	6	50%

After knowing the support and confidence values, the next step is to determine the minimum confidence value = 35%. The following is a table of the association rule results that meet 35%.

Table 9. Results Association Rules

Item A	Item B	Amount Transaction	Support	Transaction Item A	Confidence
BSR	BSJ	3	10%	3	100%
BSA	BSJ	3	10%	4	75%
BSR	BSA	2	7%	3	67%
BSJ	BSR	3	10%	5	60%
BSJ	BSA	3	10%	5	60%
BSA	BSR	2	7%	4	50%
SH	KP	2	7%	4	50%
SA	KP	3	3%	6	50%
KP	SA	3	3%	8	38%

From the previous series of stages, the product attachment pattern can be identified through the following association rule results:

1. IF you buy Raspberry Body Serum, THEN you will also buy Body Serum Oranges with value support as big as 10% AND mark Confidence as big as 100%.
2. IF you buy Apple Body Serum, THEN you will also buy Orange Body Serum with value support as big as 10% AND mark Confidence of 75%.
3. IF you buy Raspberry Body Serum, THEN you will also buy Body Serum Apple with value support as big as 7% AND mark Confidence of 67%.
4. IF you buy Orange Body Serum, THEN you will also buy Body Serum Raspberry with value support as big as 10% AND mark Confidence of 60%.
5. IF you buy Orange Body Serum, THEN you will also buy Apple Body Serum with value support as big as 10% AND mark Confidence of 60%.
6. IF you buy Body Serum Apple, THEN you will also buy Body Serum Raspberry with support value as big as 7% AND Confidence value as big as 50%.
7. IF you buy Salsa Hansjebb, THEN you will also buy Kojic Plankton with support value as big as 7% AND mark Confidence of 50%.
8. IF you buy Azarine Sunscreen, THEN you will also buy Kojic Acid with support value as big as 3% AND mark Confidence of 50%.
9. IF you buy Kojick Plankton THEN you will also buy Azarine Sunscreen with support value as big as 3% AND mark Confidence of 38%.

System Implementation

The result of the interface display is an image of the display results of all forms contained in the system. Below is a display of the Data Mining Implementation system analyzing Wimask_Id Online Shop Sales Data with the FP-Growth Algorithm.

1. Form View Login

The login form is used to secure the system from irresponsible users before entering the main menu.

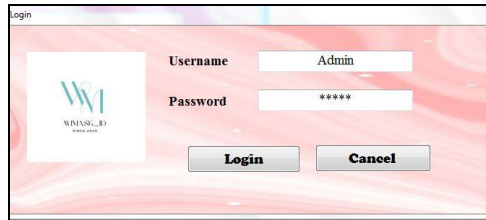


Figure 5. Login Form

2. Main Menu Form View

The main menu form is used as a link for the Goods Data Form, Transaction Data Form, FP-Growth Process Form and Report Form.



Figure 6. Main Menu Form

3. Product Data Form View

Product Data Form is used to manage product data in the Wimask_id online shop. In this form, users can input new Product Data or delete and change Product Data.



Figure 7. Product Data Form

4. Transaction Data Form View

Transaction Data Form is used to manage Transaction data in Wimask_id. The following is a display of the Transaction Data Form.

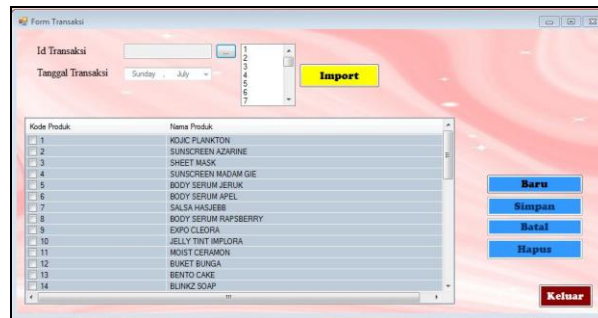


Figure 8. Transaction Data Form View

5. FP-Growth Process Form View

FP-Growth Process Form is used to analyze product sales patterns in the Wimask_id online shop using the FP-Growth algorithm based on transaction and product data that has been inputted.

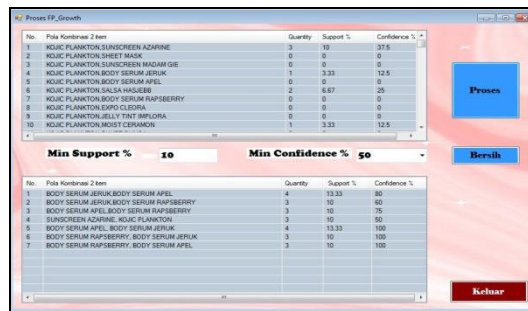


Figure 9. FP-Growth Process Form View

6. Report Form View

The Report Process Form is used to display the results of the FP-Growth algorithm that produces a product combination pattern. The following is a display of the Report form.

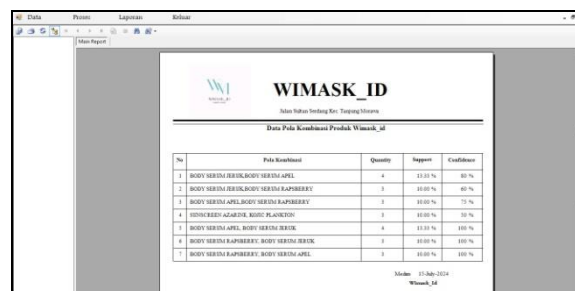


Figure 10. Form Report

Conclusion

Based on the results of the analysis of the problems that occur with the case discussed about analyzing sales data on the Wimask_id online shop. By applying the FP-Growth algorithm to the designed and built system, it can be concluded that, by applying the FP-Growth algorithm, it can help business actors in managing data owned by the company into

useful information to help business actors in providing the best service to consumers and good strategic planning. Transaction data is needed in conducting analysis using the FP-Growth algorithm, the data will be analyzed quickly and efficiently according to the FP-Growth algorithm.

Bibliography

- M. Hutasuhut, MG Suryanata, S. Kusnasari, and MA Lesmana, "Data Mining to Analyze Pesticide Sales Patterns Using the FP-Growth Algorithm," vol. 9, no. 6, pp. 1963–1973, 2022, doi: 10.30865/jurikom.v9i6.5200.
- S. Kurniawan, W. Gata, H. Wiyana, and LF Foam, "FP-GROWTH ALGORITHM ANALYSIS FOR PRODUCT RECOMMENDATION ON RETAIL COSMETIC PRODUCT SALES DATA (CASE STUDY: MT SHOP KELAPA GADING)," vol. 2018, no. Sentika, pp. 23–24, 2018.
- A. Setiawan and G. Anugrah, "Determination of Consumer Purchasing Patterns at Indomaret GKB Gresik using the FP-Growth Method," vol. 2, no. 2, pp. 115–125, 2019.
- G. Shmueli, PC Bruce, ML Stephens, and NR Patel, DATA MINING FOR . 2017.
- S. Suhada, D. Ratag, D. Wintana, and T. Hidayatulloh, "Implementation of FP-Growth Algorithm to Determine Consumer Purchasing Patterns at AHASS Cibadak," vol. 8, no. 2, pp. 118–126, 2020.
- EP Cynthia, "CLASSIFYING OUTLOOK BUSINESS SALES DATA," no. July, 2018, doi: 10.30645/jurasik.v3i0.60.
- A. Nurkholis and KN Anjumi, "Customer Transaction Pattern Analysis Using Apriori Algorithm," vol. 5, no. September, pp. 619–626, 2021.
- R. Aditiya and S. Defit, "Journal of Informatics, Economics, and Business Predicting the Level of Availability of Basic Necessities Using the FP-Growth Algorithm to Increase Sales," vol. 2, 2020, doi: 10.37034/infec.v2i3.44.
- T. Prasetya, JE Yanti, AI Purnamasari, and AR Dikananda, "Transaction Data Analysis on Consumer Purchasing Patterns Using the Apriori Algorithm Method," vol. 6, no. 1, pp. 43–52, 2021.
- R. Takdirillah, "Edumatic: Journal of Informatics Education," vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
- D. Sales et al. , "JOURNAL J – CLICK," no. July 2019, 2021.
- SP Dewi and E. Rahayu, "Application of Data Mining for Predicting Best-Selling Product Sales Using the K-Nearest Neighbor Method," vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- AF Lestari, M. Hafiz, U. Bina, S. Informatika, and K. Kunci, "Application of Apriori Algorithm on Barbar Warehouse Sales Data," pp. 96–105, 2020.
- D. Untari and DE Fajariana, "Marketing Strategy Through Instagram Social Media (Descriptive Study on the @Subur_Batik Account)," vol. 2, no. 2, pp. 271–278, 2018.
- CH Chee, J. Jaafar, IA Aziz, MH Hasan, and W. Yeoh, "Algorithms for frequent itemset mining: a literature review," Artif. Intel. Rev. , vol. 52, no. 4, pp. 2603–2621, 2019, doi: 10.1007/s10462-018-9629-z.