



Application of Data Mining for Product Purchase Pattern Analysis with Frequent Pattern Growth (FP-Growth) Algorithm on Sales Transaction Data

Nuke L Chusna^{1,*}, Herwanto², Indah Permata Sari³

^{1,2}Lecture of Universitas Krisnadwipayana, Jatiwaringin, Jakarta Timur

³Student of Universitas Krisnadwipayana, Jatiwaringin, Jakarta Timur

nukelchusna@unkris.ac.id^{1,*}, herwanto_p@yahoo.com², indah.199719@gmail.com

ARTICLE INFO

Article history:

Received: Feb 20,
2021 Revised: March 24,
2021 Accepted: April 27,
2021

Keywords:

*Pet Shop, Saller, Data
mining, frequent itemset,
FP-Growth*

ABSTRACT

The rapid growth of pet products, especially in Indonesia, makes the business competition in this field more even strict. In these conditions, a good data processing technique is needed, one of them is called the data mining technique. The algorithm used in this study is FP-Growth to generate frequent itemset that will later be used in the decision-making process that can result in a choice. This study took an object of sales transaction data from a pet shop named Ciwo Pet Shop, and the transaction data is processed by using the FP-Growth algorithm. From the results of the tests, it is found that a rule that has the best confidence value is 100% with a combination of pattern products fragrant cat sand 25 liters and equilibria kitten 7.5 kg + 750 gr, It will most certainly buy Whiskas cat food from 85 gr.

1. Introduction

In line with the growth of information technology, Data Mining is one of the rapidly growing fields due to the huge need for added value from large-scale databases which is accumulating more and more. In general, data mining can be defined as a series of processes to explore added value in the form of knowledge that has not been known manually from a data set. (Rahayu, Pane, Frimanda, & Kunci, 2018 : p.14). Data mining is the process of extracting patterns and models from data sets (data sets), where models and roles have an effective role in decision making and data mining based on data quality (Bhaya, 2017)

Ciwo Pet shop is a place for the sale, care, care, and purchase of goods for pets. Ciwo Pet shop is located in the city of Bekasi. Ciwo Pet shop always strives to create satisfaction for consumers by providing quality products, superior service and a pleasant shopping atmosphere. As competition increases, special strategies are needed to increase customer satisfaction and increase sales.

At Ciwo Pet Shop, there are many sales and purchases of pet products. By observing the number of consumers who go in and out of the store and buy the goods they need as well as the various kinds of goods sold there, it will be very difficult to determine the condition of the goods sold in the store including selling or not selling well or not selling well. By observing this, the author wants to do research on what goods are most often purchased by consumers at the same time. Therefore, the Ciwo pet shop requires a system to process data that can produce sales data for the most frequently purchased or most sold products so that these results can be used as a reference to increase the stock of products that run out and reduce some products that are rarely purchased by consumers.

At Ciwo Pet shop every transaction is recorded in the system and stored in the system database. From this data, new information can be generated in the form of a recommendation system for related goods that are often purchased by consumers. One method that can solve this problem is the FP-Growth association rule, this method can produce a combination of items in the purchase transaction. (Fitriana, Kustanto, & Vlandari, 2018 : p.21)

The problem of the object of research, namely Ciwo pet shop, has difficulty in finding product purchasing patterns. This is because the number of customer purchase transactions is quite varied.

The purpose of the study was to determine the results of the confidence value generated by changing the support value several times. And to find out the pattern of purchases made simultaneously which results in the rules of buying patterns with a confidence value of one.

This study discusses how to sell at Ciwo pet shop by knowing the most desirable products and consumer buying patterns in the form of habits of a product that is purchased simultaneously using data mining. The existence of data that accumulates and amounts to a lot is the main reason for using the FP-Growth algorithm for decision making or strategic sales policies. Researchers also want to compare differences in consumer purchases between other branches. The difference in location in the selection of branches will of course have different patterns of consumer purchases, for this reason this needs to be analyzed further so that it can produce maximum results and profits.

Asrul Abdullah (2018) discusses Product Package Recommendations to Increase Sales Using the FP-Growth Method. The results of the study show a combination of items that are often purchased, namely coffee, tea, milk with a support value of 30% and confidence of 70%. Irsyad Djamiludin, and Agus Nursikuwagus (2017) discussed product purchase patterns based on historical data using the Apriori algorithm. The results of this study showed that the highest Frequent item set were sponge cake meses and sponge cake cheese with a confidence value of 84.62%. Sigit Kurniawan, Windu Gata, and Hari Wiyana (2018) regarding the pattern of purchasing cosmetic products. The results of the tests carried out obtained a rule that has the best confidence value of 89% with the rule for every purchase of white rice and white langsung facial foam mask products.

2. Theoretical Basis Data Mining

Data mining is the process of looking for interesting patterns or information in selected data using certain techniques or methods. Techniques, methods, or algorithms in data mining vary widely. The selection of the right method or algorithm really depends on the objectives and the overall Knowledge Discovery in Database (KDD) process.(Mardi, 2017 : p.214). Data Mining is a term used to explore knowledge hidden in databases, semi-automated processes using statistical, mathematical, artificial intelligence and machine learning techniques.(Hartama et al., 2019).

Market Basket Analysis

Market basket analysis is an association in data mining that can find attributes that appear at one time. This process analyzes customer buying patterns by finding relationships between different items that consumers place in the shopping basket. The results that have been obtained can later be used by retail companies such as shops or supermarkets to develop marketing strategies by looking at which items are likely to be purchased simultaneously by consumers.(Rahmawati, Nasution, & Amijaya, 2017 : p.2).

Frequent Pattern Growth Algorithm (FP-Growth)

The FP-Growth algorithm is a development of the Apriori algorithm. Frequent Pattern Growth Algorithm (FP-Growth) is an alternative algorithm that can be used to determine the most frequently occurring data set (frequent itemset) in a data set. The FP-Growth algorithm uses the concept of tree construction, which is commonly called FP-Tree, in searching for frequent itemsets instead of using generate candidates as is done in order to provide maximum results.(Amelia & Utomo, 2019 : p. 417).

Association Rules

Association Rules is a technique or procedure for finding relationships between products in a database. Its use is to determine the extent to which the size of the data output limit is desired by the user. Where the support value shows the proportion of transactions in the database containing elements A and B, while the confidence value indicates the percentage size of the accuracy of a rule containing elements A and B.(Megayasa, 2016 : p.207).

$$\text{Support } (A) = \frac{\sum \text{Transaksi Mengandung } A}{\sum \text{Total Transaksi}}$$

$$\text{Support } (A, B) = \frac{\text{Jumlah Transaksi Mengandung } A \text{ dan } B}{\text{Total Transaksi}}$$

$$\text{Confidence } P(B|A) = \frac{\sum \text{Transaksi Mengandung } A \text{ dan } B}{\sum \text{Total Transaksi } A}$$

3. Method

The research stages in determining the pattern of purchasing joint products that are sold together at Ciwo Pet Shop use the Cross-Industry Standard Process for Data Mining (CRISP-DM) method to carry out the analysis process as a problem solving strategy of the business or research unit.

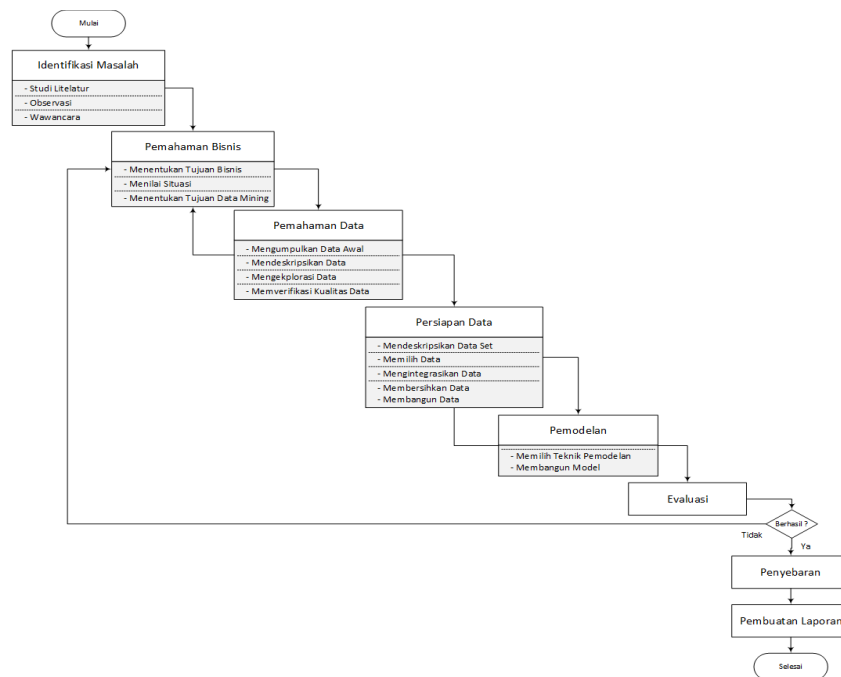


Fig 1. Research Stage Flowchart

1. Identification of problems
At the problem identification stage there are several methods applied which aim to make the analysis more focused to identify problems.
2. Business Understanding
Understanding the goals and needs and then translating the knowledge into problem definition in its use for data mining processing.
3. Understanding Data (Data Understanding)
Starting with data collection then proceed to get an understanding of the data, what is needed for how the data will be processed and processed later.
4. Data Preparation
Includes all activities in building a dataset from start to finish, including selecting records, data attributes, and cleaning up unnecessary parts.
5. Modeling (Modelling)
At this stage, the selection and application of modeling techniques that are adjusted to obtain optimal results is carried out.
6. Evaluation (Evaluation)
An evaluation of the effectiveness of the model used is carried out whether the model used has met the objectives set in the previous phase.

7. Deployment

Knowledge and information that has been obtained will be presented in a special form that is more easily understood by users later.

8. Report Generation

Make a research report in accordance with the results that have been obtained from the initial process of data collection to drawing conclusions.

4. RESULTS AND DISCUSSION

Data collection

Data collection techniques were carried out by direct research at Ciwo pet shop. This study uses data from Ciwo Pet shop sales transactions in the 2020 period period, namely, starting from January 2020 to March 2020, Then the total amount of data used for research is taken as much as 201 transaction data. as in Figure 2:

Tanggal	Invoice	ID_Pembeli	Nama_Pembeli	Kota	ID_Produk	Nama_Produk	Qty	Price	Total
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	1	MAKANAN KUCING WHISKAS SACHET	1	7000	7000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	27	PASIR KUCING GUMPAL WANGI MARKO	1	100.000	100.000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	2	PASIR KUCING WANGI GUMPAL 1KG S	1	5000	5000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	63	MAKANAN KUCING BOLT CAT TUNA 1K	1	156000	156000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	136	CAT FOOD MAKANAN KUCING OMEGA	1	420000	420000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	147	REPACK NUTRICAT NUTRI CATT 1KG M	1	24000	24000
01/01/2020	INV/20200101/XX/I/414746044	145	Tysha Amanda	Kota Bekasi	159	MAKANAN KUCING BASAH FANCY FEAS	3	25000	75000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	2	PASIR KUCING WANGI GUMPAL 1KG S	1	5000	5000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	3	OBAT KUTU TETES DETICK	1	17500	17500
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	5	REPACK PROPLAN KITTEN 500 GR	1	43000	43000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	6	REPACK ORI CAT FOOD ORICAT 1KG SE	1	71000	71000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	18	MAKANAN KUCING NATURE PROTECT	1	57000	57000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	27	PASIR KUCING GUMPAL WANGI MARKO	1	100.000	100.000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	39	ROYAL CANIN ADULT BEAUTY HEALTH	2	70000	140000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	41	BIO GROOM SHAMPO KITTEN KUDDLY	3	70000	210000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	63	MAKANAN KUCING BOLT CAT TUNA 1K	2	156000	312000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	122	EQUILIBRIO PERSIAN LONG HAIR 500G	1	70000	70000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	136	CAT FOOD MAKANAN KUCING OMEGA	1	420000	420000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	138	PASIR KUCING WANGI GUMPAL CUB N	1	71000	71000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	141	DECOMPOVAR ANTI BAKTERI KANDAN	1	71000	71000
01/01/2020	INV/20200122/XX/I/424045785	60	Fitri Ferlani Sari	Jakarta Timur	147	REPACK NUTRICAT NUTRI CATT 1KG M	1	24000	24000
02/01/2020	INV/20200122/XX/I/423486538	132	Saras Putri Meirane	Jakarta Timur	2	PASIR KUCING WANGI GUMPAL 1KG S	1	5000	5000

Fig 2. Sales Transaction Data

Figure 2 shows the itemset format that will be used in the data processing process using RapidMiner.

Invoice	ID_Pembeli	Nama_Pembeli	Kota	ID_Produk	Nama_Produk	Qty	Price
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	13	ISKIES SEAFOOD SELECTION 400G	1	7000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	28	CING GUMPAL WANGI MARKO	1	89000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	59	FRISKIES DEEP SEA DELIGHT	1	5000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	64	ANAN KUCING BOLT CAT TUNA 1K	1	2E+05
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	135	MAKANAN KUCING CAT CHOISE	1	15000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	137	KUCING OMEGA CAT TUNA 2KG	1	24000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	148	JTRI CATT 1KG MAKANAN KUCI	3	24000
00101/XX/I/41	145	Tysha Amanda	Kota Bekasi	161	SAH FANCY FEAST 85 GR / CAT	2	9850
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	1	NAN KUCING WHISKAS SACHET	1	4.350
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	2	S WANGI GUMPAL 1KG SAMOD	1	17500
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	3	DETICK DETIK DETIC / OBAT KU	1	18000
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	7	OD ORICAT 1KG SEJENIS BOLT	3	17500
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	19	URE PROTECTION KITTEN NATI	1	57000
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	28	CING GUMPAL WANGI MARKO	1	89000
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	40	CANIN ADULT BEAUTY HEALTH	2	70000
00122/XX/I/42	60	Fitri Ferlani Sari	Jakarta Timur	42	ROOM SHAMPO KITTEN KUDDLY	3	70000

Fig 3. Transaction Data Normalization Table

Information:

1. No Invoice is an ID for the sale of goods.

2. Fragrant Clumped Cat Sand 25 Liter, Dog Cat Shampoo 250 ml, Whiskas Cat Food Sachet 85 gr, etc are goods sold to the company.
3. *false* (0) indicates that the item was not purchased by the buyer.
4. *True* (1) indicates that the item was purchased by the buyer.

Application of the FP-Growth Algorithm

Experiments and tests in this study were carried out using the FP-Growth algorithm. This experiment was carried out on a validated dataset based on the results of preprocessing data that had been carried out using RapidMiner tools with the following stages:

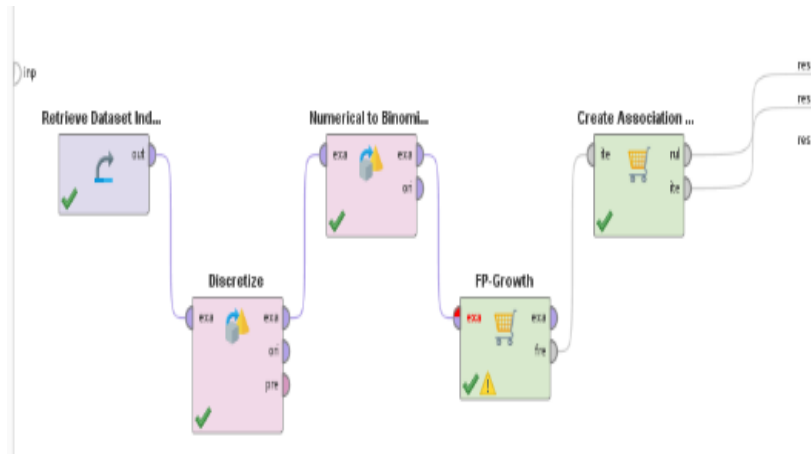


Fig 4. FP-Growth implementation process

Modeling Results with FP-Growth . Algorithm

Table 1 below is the result of data modeling based on the highest support value using the FP-Growth algorithm.

Table1. Frequent ItemSet Results

Rules	Support	Confidence
164.107 → 1	0.289	1
164.139 → 1	0.493	0.990
164.3 → 1	0.293	0.983
139,135 → 1	0.199	0.975

The calculation of the confidence value in some of the results of the frequent items above shows that the minimum confidence value requirement is 90%, as well as other calculations.

Association Rules Results

The results of calculations using RapidMiner produce three rules. The results obtained from the Association Rules also explain that with a 100% confidence value, the highest confidence value is 1 for every product purchaseID_Produk 164 (25 liters of lumpy scented cat sand) and 107 (equilibrio kitten 7.5 kg + 75 gr), Then I will buy ID_Produk 1 (Whiskas cat food sachet 85 gr).

Table2 Frequent Item Set Results

Rules	Product package	Confidence
164.107 → 1	If the consumer buys 25 liter lumpy scented cat sand and equilibrio kitten 7.5 kg + 75 grThen I will buy whiskas cat food sachet 85 gr	1
164.139 → 1	If the consumer buys 25 liters of lumpy scented cat sand and 20 kg of lumpy cub n kit scented cat sandThen I will buy whiskas cat food sachet 85 gr	0.990
164.3 → 1	If the consumer buys 25 liters of lumpy scented cat sand and flea drops, seconds, ticks / cat dog flea medicationThen I will buy whiskas cat food sachet 85 gr	0.983
139,135 → 1	If the consumer buys fragrant cat sand, lumps of cub n kit 20 kg and cat choize repack - cat choize cat food such as bolt omega jioThen I will buy whiskas cat food sachet 85 gr	0.975

CONCLUSION

From the research that has been done on the object of Ciwo Pet shop, the results of the tests carried out were carried out several times by changing the min value. support, with a value range of 50% to 95%, it can be concluded that the lower the min. Support value used, the higher the confidence value will be. Several rules were obtained using a support value of 95%, the best confidence value was 1 for every purchase of 25-liter lumpy scented cat sand product, Equilibrio kitten 7.5 kg + 75 gr, it is certain that you will buy the whiskas cat food product sachet 85 gr. And several other rules with different confidence values.

REFERENCES

- Amelia, R., & Utomo, D. P. (2019). Analisa Pola Pemesanan Produk Modern Trade Independent Dengan Menerepakan Algoritma Fp. Growth (Studi Kasus: Pt. Adam Dani Lestari). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 416–423. <https://doi.org/10.30865/komik.v3i1.1622>
- Bhaya, W. S. (2017). Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107.
- Fitrina, N., Kustanto, K., & Vlandari, R. T. (2018). Penerapan Algoritma Apriori Pada Sistem Rekomendasi Barang Di Minimarket Batox. *Jurnal Teknologi Informasi Dan Komunikasi (TIKOMSiN)*, 6(2), 21–27. <https://doi.org/10.30646/tikomsin.v6i2.376>
- Hartama, D., Perdana Windarto, A., & Wanto, A. (2019). The Application of Data Mining in Determining Patterns of Interest of High School Graduates. *Journal of Physics: Conference Series*, 1339(1). <https://doi.org/10.1088/1742-6596/1339/1/012042>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Megayasa, I. G. P., Agus, I. K., Aryanto, A., Diputra, I. G. S., Arianta, I. N., & Rusditya, S. (2016). Implementasi Algoritma Apriori untuk Menganalisis Pola Pembelian Konsumen pada Produk SPA. *Senapati*, 206–210.
- Rahayu, W. I., Pane, S. F., Frimanda, L. A., & Kunci, K. (2018). Implementasi Data Mining Dengan Metode K-Means Clustering Untuk Menentukan Iklan Audio Berdasarkan User Behaviors Pada Aplikasi Audio Social Media Svara Di Pt . Zamrud Khatulistiwa Technology. 10(2), 13–19.
- Rahmawati, N., Nasution, Y. N., & Amijaya, F. D. T. (2017). Aplikasi Data Mining Market Basket Analysis untuk Menemukan Pola Pembelian di Toko Metro Utama Balikpapan. *Jurnal EKSPONENSIAL*, 8(1), 1–8.