Implementation of The Apriori Algorithm Method to Analyze the Pattern of Motorcycle Sales at PT. Alpha Scorpio

Oloan Sihombing[1], Veronica T. D. Simamora[2*], Vivian P. Leo[3], Petrus Kristian P. Marpaung[4]

[1,2,3,4]Faculty of Technology and Computer Science, Prima Indonesia University
Jl. Sampul No.4, Indonesia

e-mail: [2*]veronicatdsimamora99@gmail.com

Abstrak - Penelitian ini membahas bagaimana membuat pola pada proses perjualan sepeda motor, dimana setiap bulannya penjualan setiap jenis sepeda motor pasti berbeda-beda. Sehingga setiap pegawai mengalami kewalahan dalam mengelompokkan jenis-jenis barang yang akan di rekomendasikan dalam penambahan stok. Tujuannya adalah agar permintaan konsumen tetap terpenuhi setiap bulannya. Berdasarkan penelitian yang telah dilakukan, penelitian ini menyimpulkan: hasil penelitian dengan batas dukungan minimum 30% dan kepercayaan minimum 70%, membentuk 20 aturan setiap tahun. Salah satu aturan yang dibentuk adalah jika yang dijual adalah sepeda motor 2X4, maka tipe Yamaha A 90 dijual dengan nilai support 45% dan nilai confidence 74,46%.

Kata Kunci: Pola Persediaan, Metode Apriori, Penjualan

Abstract - This study discusses how to make a pattern in the motorcycle sales process, where every month the sales of each type of motorcycle must be different. So that every employee is overwhelmed in classifying the types of items that will be recommended in adding stock. The goal is to keep consumer demand met every month. Based on the research that has been done, this study concludes: the results of research with a minimum support limit of 30% and a minimum confidence of 70%, form 20 rules every year. One of the rules established is that if a 2X4 motorcycle is sold, the Yamaha A 90 type is sold with a support value of 45% and a confidence value of 74.46%.

Keywords: Inventory Pattern, Apriori Method, Sale

INTRODUCTION

Considering that currently in every company, school and any institution, the number of community services is increasing, not decreasing anymore. So with that, the loyalty of the system which was previously only done with a manual system has now led to faster and more accurate information. However, the high need for information to gain new knowledge is sometimes not matched by the presentation of adequate information.

Motorcycle sales are currently increasing because everyone is currently on a work trip or anywhere, motorbikes are one of the most practical forms of transportation to take anywhere. Because of that, every motorcycle sales forum has problems in determining the pattern of supply of goods because the difference is that every month there are many types of motorcycles that will be sold.

Data mining is part of the KDD (Knowledge Discovery in Databases) process which consists of several stages such as data selection, pre-processing, transformation, data mining, and evaluation of results (Pane, 2013). Data mining is a process that employs one or more computer learning techniques (machine learning) to analyze and extract knowledge automatically (Putria, 2018).

According to Kursini dan Taufiq (2009:149) The Apriori algorithm is a type of association rule in data mining. In addition to a prior, which are included in this group are the Generalized Rule Induction method and the Hasd Based Algorithm. Data mining is used to explore added value in the form of unknown information manually from a database. Information is obtained by extracting and recognizing important or interesting patterns from the data contained in the database (Sulindawaty, Muhammad Zarlis, 2018), (Yonata Laia et al., 2019). In accordance with the background of this problem is to regulate the pattern of motorcycle sales, in this research will use the a priori method in regulating the sales pattern. Hopefully this system can help the company in knowing sales every month. The a priori algorithm is one type of association rule in data mining that explains the association of several attributes, often called affinity analysis or

market basket analysis. Association analysis or association rule mining is a data mining method to find rules for a combination of Vulandari items, (R.T. 2017). At the association analysis stage, researchers are interested in getting an efficient algorithm, which is high frequency pattern analysis (frequent pattern mining). The level of importance of an association can be known by two benchmarks, namely: support and confidence. Support (support value) is the percentage of the combination of these items in the database, while confidence (certainty value) is the strength of the relationship between items in the association rules.

The basic methodology of association analysis is divided into two stages (Wijayanti, 2017).

In the current era of big data, with data volumes reaching a zettabyte (one trillion gigabytes), problems and data are growing exponentially while the number of data analysts and data scientists is growing linearly. Therefore, reliable computational techniques are needed that can analyze data quickly and accurately. Unfortunately, most of the designers and developers of data mining-based systems have difficulty in determining the most suitable techniques and methods for the problems they face (Nur Khormarudin, 2016).

Heru Sulistiono in his research entitled "Study of Application of C4.5 Algorithm, Neural Network and Naïve Bayes for Classification of Students with Problems in Registration". The purpose of this study is a comparative analysis of three data mining classification algorithms, namely C4.5, Neural network and Naïve Bayes so that the most accurate algorithm can be found to predict students who have problems in registration. The results of this study are the Naïve Bayes algorithm has the highest level of accuracy, so it is good for classifying students who have problems in registration with a percentage of 93.58%, (Salmu & Solichin, 2017).

In educational organizations, an information system can be used to obtain information that supports every activity in making a decision. Data mining algorithm C4.5 can be used to develop a system that has the ability to see student graduation patterns, which can then be used as a strategy in the lecture process. The C4.5 algorithm is a decision tree classification algorithm that is widely used because it has major advantages over other algorithms. The advantages of the C4.5 algorithm are that it can produce a decision tree that is easy to interpret, has an acceptable level of accuracy, is efficient in handling discrete type attributes and can handle discrete and numeric type attributes. In constructing the tree, the C4.5 algorithm reads all training data samples from storage and loads them into memory. (Kamagi & Hansun, 2014).

Data mining clustering process that applies the K-means clustering algorithm on human infectious diseases in Maja-lengka Regency for 2014 data. other groups so that the data in one cluster/group has

a small degree of variation. To calculate the distance of all data to each cluster center point, the Euclidean distance theory is used (Bastian et al., n.d.).

(Jaroji, Danuri, 2016) discussed K-Means todetermine the prospective recipients of Bidik Misi in Polbeng. The K-Means algorithm is able to group prospective Bidik Misi scholarship recipient into 4 clusters, in which cluster 0 means giving recommendations with consideration, cluster 1 providing very feasible recommendations, cluster 2 providing appropriate recommendations and cluster 3 providing less feasible recommendations (Jaroji, Danuri, 2016).

(Sulistino, 2015) discusses the study of the application of the C4.5 Algorithm, Neural Network and Naïve Bayes for the classification of students with problem in registration. In this study, analyze three data mining classification algorithms, namely C4.5, Neural Network and Naïve Bayes to find out the most accurate algorithm for predicting students who have problems withregistration (Sulistino, 2015).

Amrin discusses Data Mining with the Apriori Algorithm for determining the association rules for buying fertilizer patterns. In this study, perform a combination test to determine the minimum support and minimung confidence parameters which are the threshold values using the Apriori Algorithm (Amrin Amrin, 2017).

(Rerung, 2018) discussed about in a business, efforts are needed to maximize profits, including by doing promotions. There are many ways that can be done to promote products such as online by utilizing social media Facebook and sites that provide advertisements. However, to obtain maximum results, it is necessary to calculate how likely it is that customers will be interested in the products offered (Rerung, 2018).

(Yanto & Kesuma, 2017) discussed Data Mining is the process of analyzing data to find a pattern of hidden data collection. by utilizing the traffic data library, can dig up information about what books are often borrowed by visitors and interrelationships of each loan to be able to do the preparation and layout of the book (Yanto & Kesuma, 2017).

(Santoso, 2017) discussed Data mining is needed to extract information from large amounts of data so that it can be used as useful new information. Apriori algorithm technique to find information on the pattern of interrelationships between books that are often borrowed by library visitors at the same time. So that it can provide recommendations for the preparation of books in accordance with the level of support and confidence possessed by each interrelated book (Santoso, 2017).

RESEARCH METHODOLOGY

The research framework is a temporary explanation of a phenomenon that is the object of our problem.

This research framework was prepared based on a literature review and relevant or related research results. This research framework is an argument for us in formulating a hypothesis. In formulating a hypothesis, the research framework argument uses deductive logic (for quantitative methods) by using scientific knowledge as the basic premise. The research framework can be seen in Figure 1.
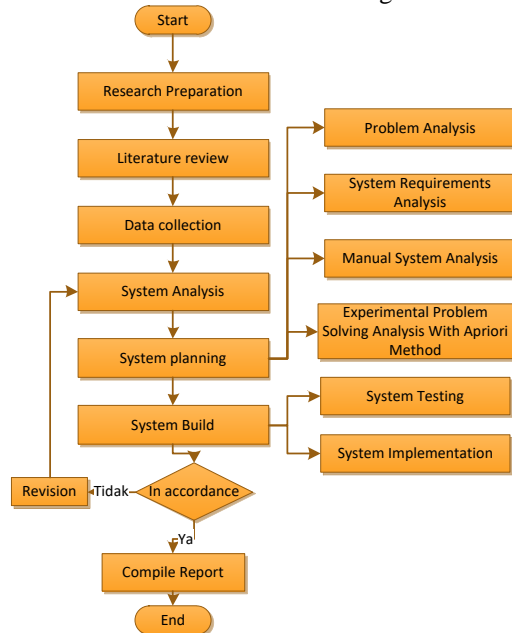
Source: (Sihombing, et.al, 2022)
Figure 1. Kerangka Penelitian

In the figure above, which explains where this research will begin, as seen in the picture above, it will begin with data collection, to become research material. Then literature study in order to find sources from books, journals and various other sources. Then after the data has been obtained, it enters the analysis stage, then after the analysis enters the system design stage according to the results that have been analyzed. After that, the mask is in the making of the system according to the design that has been made.

RESULTS AND DISCUSSION

Table 1. Sales Data in 2021-2022

Data Number	Name	Budget	Brand
00001	A	13000000	Yamaha 1
00002	B	13500000	Yamaha 2
00003	C	15000000	Yamaha 3
00004	D	14500000	Yamaha 4
00005	E	14000000	Yamaha 5
00006	F	15500000	Yamaha 6
00007	G	16000000	Yamaha 7
00008	H	15750000	Yamaha 8
00009	I	16500000	Yamaha 9
00010	J	15000000	Yamaha 10

Source: (Sihombing, et.al, 2022)

Table 2. Inventory Transaction Pattern

Transaction	Items
1	A, C, D
2	F,C,E
3	G,H,I
4	H,D,F
5	B,C,G
6	H,D,F
7	I,C,D
8	H,B,C
9	B,C,E
10	E,H,G
11	B,C,I
12	I,H,J
13	J,I,E
14	G,H,E
15	B,A,C
16	C,A,B
17	B,I,E
18	J,E,A
19	A,I,B
20	C,E,D

Source: (Sihombing, et.al, 2022)

1. Itemset Formation

The following is a solution based on the data provided in table 3. The process of forming C1 or called 1 itemset with minimum support = 25% with the following formula:

$$Support \ A = \frac{Transaction \ amount \ contains \ A}{Transaction}$$

Table 3. Support of each item

Itemset	Amount	Support
a	12 / 20 %	60%
f	8 / 20 %	40%
c	15 / 20 %	75%
i	12 / 20 %	60%
d	12 / 20%	60%
f	8 / 20 %	40%

Source: (Sihombing, et.al, 2022)

2. Combination of 2 Itemset

The process of forming C2 or called 2 itemset with Minimum support = 50%. It can be solved by the following formula:

$Support \ (A,B) = P(A \cap B)$

$$Support \ (A,B) = \frac{The \ number \ of \ transactions \ contains \ A \ and \ B}{Total \ transaction}$$

Table 4. Candidate 2-itemset

Itemset	Amount	*Support*
A, C, D	4	20%
F,C,E	12	**60%**
G,H,I	8	40%
H,D,F	4	20%
B,C,G	0	0%
H,D,F	4	20%
I,C,D	6	10%
H,B,C	12	20%
B,C,E	4	20%
E,H,G	12	20%
B,C,I	8	40%
I,H,J	4	20%
J,I,E	8	40%
G,H,E	4	20%
B,A,C	8	40%

Source: (Sihombing, et.al, 2022)

3. Association Formation rules

After all high-frequency patterns are found, the next step is to look for association rules that meet the minimum requirements for confidence by calculating the confidence of the associative rule A B. Minimum confidence = 60%

The confidence value from rule A $\longrightarrow$ B is obtained by the following formula:

$Confidence =$

$P(B|A) = \dfrac{\text{The number of transactions contains A and B}}{\text{Transaction amount contains A}}$

Table 5. Association Rules

Rule	Confidence	
Jika sepedamotor a maka sepedamotor b	12/15 * 80	**80%**

Source: (Sihombing, et.al, 2022)

Based on the association rules above, it can be seen that the motorcycle inventory pattern in the implementation of the association method using the a priori algorithm in the motorcycle sales pattern analysis application at PT. Alpha Scorpio.

CONCLUSION

Based on the research that has been done and in accordance with the purpose of this study how to find out the sales pattern in the company, this system can analyze sales patterns and can provide recommendations to the company for the type of motorcycle that will increase the addition of inventory. This study concludes: research results with a minimum support limit of 30% and a minimum confidence of 70%, form 20 rules every year. One of the rules established is that if a 2X4 motorcycle is sold, the Yamaha A 90 type is sold with a support value of 45% and a confidence value of 74.46%.

REFERENCES

Amrin Amrin. (2017). Data Mining Dengan Algoritma Apriori untuk Penentuan Aturan Asosiasi Pola Pembelian Pupuk. *Paradigma*, *XIX*(1), 74–79. https://doi.org/https://doi.org/10.31294/p.v19i1.1836

Bastian, A., Sujadi, H., & Febrianto, G. (n.d.). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). 1*, 26–32.

Jaroji, Danuri, F. P. P. (2016). K-MEANS UNTUKMENENTUKAN CALON PENERIMA BEASISWA BIDIKMISI DI POLBENG. *Jurnal Inovtekpolbeng- Seri Informatika*, *1*(1), 8.

Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal ULTIMATICS*, *6*(1), 15–20. https://doi.org/10.31937/ti.v6i1.327

Nur Khormarudin, A. (2016). Teknik Data Mining: Algoritma K-Means Clustering. *Jurnal Ilmu Komputer*, 1–12. https://ilmukomputer.org/category/datamining/

Pane, D. K. (2013). Implementasi Data Mining Pada Penjualan Produk Elektronik Dengan Algoritma Apriori (Studi Kasus : Kreditplus). *Pelita Informatika Budi Darma*, *volume : I*, 25–29.

Putria, N. E. (2018). Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori Pada Terminal Tiket Batam Tour & Travel. *Computer Based Information System Journal (CBIS)*, *6*(1), 29–39. http://ejournal.upbatam.ac.id/index.php/cbis/article/download/643/410

Rerung, R. R. (2018). Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk. *Jurnal Teknologi Rekayasa*, *3*(1), 89. https://doi.org/10.31544/jtera.v3.i1.2018.89-98

Salmu, S., & Solichin, A. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta Prediction of Timeliness Graduation of Students Using Naïve Bayes: A Case Study at Islamic State University Syarif Hidayatullah Jakarta. *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, *April*, 701–709.

Santoso, H. (2017). Data Mining Penyusunan Buku Perpustakaan Daerah Lombok Barat Menggunakan Algoritma Apriori. *Seminar Nasional TIK Dan Ilmu Sosial (SocioTech)*, 25–35.

Sulindawaty, Muhammad Zarlis, N. Z. (2018). *Strategi Peningkatan Penjualan Buku Pada*

PT . Tiga Serangkai Internasional Dengan Metode Clustering. 2, 102–110.

Sulistino, H. (2015). Kajian Penerapan Algoritmac4. 5, Neural Network Dan Naïve Bayes Untuk Klasifikasi Mahasiswa Yang Bermasalah Dalam Registrasi. *Faktor Exacta*, *8*(4), 305–315. http://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/501

Tandian, C., Laia, Y., & Saputra, A. (2019). Penerapan Data Mining Dalam Memprediksi Pemenang Klub Sepak Bola Pada Ajang Liga Champion Dengan Algoritma C.45. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *2*(2), 91–98.

https://doi.org/10.34012/jusikom.v2i2.397

Wijayanti, A. W. (2017). Analisis Hasil Implementasi Data Mining Menggunakan Algoritma Apriori pada Apotek. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *3*(1), 60. https://doi.org/10.26418/jp.v3i1.19534

Yanto, R., & Kesuma, H. Di. (2017). Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *4*(1), 1–10. https://doi.org/10.35957/jatisi.v4i1.83