MARKET BASKET ANALYSIS ON TRANSACTION DATA USING THE APRIORI ALGORITHM

Sri Karnila[1], Akbar Rizkyandi[2], Nurjoko[3], Rio Kurniawan[4]

[1,2,3]Program Studi Sistem Informasi, Institut Informatika dan Bisnis Darmajaya
[4]Program Studi Teknik Informatika, Institut Informatika dan Bisnis Darmajaya
[1,2,3,4]JL. ZA Pagar Alam no 93, Gedong Meneng, Bandar Lampung, Indonesia
Email: srikarnila_dj@darmajaya.ac.id[1], akbar.2.rizkyandi@gmail.com[2],
nurjoko@darmajaya.ac.id[3], riokurniawan@darmajaya.ac.id[4]

Abstract
This research aims to get information about the relationship between sales patterns carried out by CV. Dian Abadi Jaya workshop by using APRIORI algorithms through transaction data sets carried out by customers. The subject of research is a record of shopping cart transactions made by customers, namely vehicle parts sales transactions and vehicle repair service transactions. The data collection techniques used are interviews and documentation. The criteria used in this research are a minimum of frequent itemset of 20 transactions with support criteria of 1,7%, confidence value of 40% and lift ratio value above 1. The results of the research have produced 9 sales pattern relationships with the highest confidence of 100%. The results that have been obtained are expected to help the CV. Dian Abadi Jaya workshop in making a decision for the next sale.

1. INTRODUCTION

The development of information technology is currently happening so rapidly thaat it affects all aspects of life. Technological developments also affect data growth, which allows data to continue to grow every day. There is a large amount of data that is around us but is not useful for human purposes if the data is not processed and analyzed into patterns, models, or knowledge[1]. These problems can be overcome by applying data mining methods, namely the process of discovering new patterns from very large datasets, which includes slice methods from artificial intelligence, machine learning, statistics, and databases[2][3]. One improvement that has been made in data mining is how to browse existing data to build a model, then use the model to recognize other data patterns that are not in the stored database[4]. Data mining can be qualified as a science and technology because it is able to provide something new to solve the problems at hand [5][6].

CV workshop. Dian Abadi Jaya is a company engaged in the repair and sale of spare parts for four-wheeled vehicles, which is located on Jalan Kimaja No. 70 Way Halim, Bandar Lampung. The services offered include tune ups, oil changes, and light services like AC repairs. As a company that has quite a lot of customers every day, the volume of transaction data generated will also increase[2][3]. Based on the results of interviews with the salesperson, CV Dian Abadi Jaya can serve 5 to 8 vehicles every day. The transaction data [7]is processed according to the company's needs, such as expense reporting, income reporting, as well as the number of transactions in each particular month[8][9].
Often, the arrangement of goods that is used in the storage warehouse causes the mechanic to find it difficult to find the needed goods because the location of the goods is not in accordance with the habits of transactions made by the customer. In addition, some inappropriate placement of goods in

the warehouse causes the goods to take a long time to sell[8][7].

Therefore, in this study, a shopping basket analysis or market basket analysis will be carried out, that is a method of analyzing consumer behavior which is generally used as an initial step in seeking knowledge from data transactions when we do not know the specific pattern being sought[6][10] The analysis is carried out to find models of patterns or relationships in the purchase of purchasing goods that are often purchased against other goods simultaneously by using association rules with a priori algorithms[9][11]. It is hoped that it can help the company in compiling the layout of goods to make it easier for employees to find spare parts in the storage warehouse and avoid selling old goods. In addition, it is hoped that it can help the company to provide information about the stock of goods that should be provided based on the habits of transactions made by customers[12][8].

II. LITERATURE

2.1 Data

Data is a compound term which means a fact or part of a fact that contains meaning associated with reality, symbols, pictures, numbers, letters, or symbols that indicate an idea, object, condition, or situation[13] [14]

2.2. Variable

A variable is defined as a characteristic that can assume more than one set of values in a given numerical measure. A variable can also be a name given to an observation, measurement, or calculation from a known set of data[15][16] .Variables are often also called attributes, namely properties, properties, or characteristics of objects whose values can vary from one object to another, one time to another.

2.3. Information

Information is data that has been processed or interpreted for use in the decision-making process. Information processing systems process data into information, or rather, data processing from useless to useful for the recipient.[17][18]

2.4. Data Mining

Data mining the process of analyzing hidden data patterns according to various perspectives so that they become useful information[12]. The information generated can facilitate users' tasks, such as business decision making, weather forecasts, and other information [19].

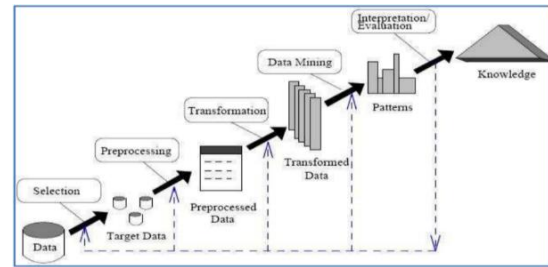### 2.5. Proses *Knowledge Discovery in Database* (KDD)

Figure 1. Knowledge Discovery Process in Databases (KDD)

The process of KDD is described as follows[20]:
1. Goal identification is the initial stage that must be done in understanding what domains or areas will be taken and what goals are to be achieved (targets must be clear).
2. Data Selection is carried out to see the relevance of the data originating from the database and adjusted to the knowledge of the relevant experts in order to achieve appropriate results, and the level of reliability is also good.
3. Data preprocessing this stage includes operations for data cleaning, such as handling data disturbances, as well as handling inconsistent data, data integration can combine data with different data sources, data transformation and data reduction, including feature selection and extraction[21].
4. Data mining, is an important process where the method is used to extract valid data patterns. This step includes the choice of the most suitable DM strategy, such as classification, regression, clustering, or association[22].
5. Interpretation, or evaluation, is a stage for estimating, identifying, and interpreting patterns that are truly important and represent knowledge based on a measure of the degree of importance.
6. Knowledge, is the last stage that can involve knowledge directly, then combine knowledge into other systems for further processing .

2.5. APRIORI Algorithm

The analysis of association rules aims to obtain a set of items or item sets that appear simultaneously in a collection of transactions. An itemset consists of an i-item which is called an i-itemset[13]. The percentage of transactions or combinations consisting of one item set is called the support itemset and confidence (certainty value) is the ratio of the strength of the relationship between items in the association rules[1][23]. Support and confidence are the two basic criteria in the association rules [19].

An association rule is a data mining technique to find associative rules between a combination of items with the aim of finding frequent item sets contained in a set of data[21]. A priori analysis [24]is defined as a process to find all a priori rules that meet the minimum requirements for support

and confidence. [25][11] The initial stage in the a priori algorithm is to find frequent item sets by analyzing high-frequency patterns, namely by looking for combinations that meet the minimum requirements of the support value in the databases. or datasets[12].

 X => Y in DB transactions with confidence c if c% of transactions in DB contain X and also contain Y. In addition, the rule X => Y has support in DB transactions if s% of transactions in DB are X Y. With the rule having a high level of confidence and a high support value, it can be said that the rule is a strong rule. [26]

The steps for calculating the association rules can be written as follows:

1. The following equation is used to find the support value of an item.

$$Support(X) = \frac{the\ number\ of\ transaction\ contains\ X}{total\ transaction\ amount} \quad (1)$$

2. The following equation is used to find the support value of the two items:

$$Support(A=>B) = \frac{the\ number\ of\ transction\ A\ end\ B}{total\ transaction\ amount} (2)$$

3. The equation is used to find the support value of the two items.

$$Confidence = \frac{jumlah\ transaksi\ mengandung\ Anteseden\ dan\ Konsekuen}{jumlah\ transaksi\ mengandung\ Anteseden}$$
(3)

The antecedent is the cause that makes the item consequent. The consequent is a result or conclusion, namely the item that will be purchased after buying the antecedent. If the association rule A = B is obtained, then A is the antecedent, and B is the consequent. So the formula for calculating the confidence above can be simplified to:

$$Confidence(A=>B) = \frac{the\ number\ of\ transaction\ A\ and\ B}{total\ transaction\ amount\ A} \quad (4)$$

The next stage is testing the association rules, namely testing the lift value. The Lft Ratio is a measure to determine the strength of the association rule that has been formed from the support and confidence values in each resulting pattern. The lift ratio accuracy test is commonly used in unsupervised learning calculations [14]. Testing the accuracy of the lift ratio value is written using the following equation:

$$Lift\ (X => Y) = \frac{Support\ X\ dan\ Y}{Support\ X * Support\ Y} \quad (5)$$

The steps of the APRIORI algorithm to find association patterns in the data are as follows.
1. Set k=1 (points to the 1st items).

2. Count all k-itemsets (item sets that have k item sets), to get candidate 1-items.
3. Calculate the support of all candidate items, then sort the item sets based on the calculation of the minimum support, to get frequent items.
4. Combine all item sets of size k to produce candidate itemset k+1 or candidate k-items.
4. Set k=k+1.

Repeat steps 2-5 until no larger itemset can be formed

III. RESEARCH METHODS

This research is carried out with several steps starting from the collection and a priori stages, which are described as follows:

3.1. Method of collecting data
a) Observation Method
   Conduct a review and direct observation of the company's CV. Dian Abadi Jaya by collecting data and information related to the problem under study.

b) Interview method
   Conducting interview sessions with the authorities related to transaction data at the CV company. Dian Abadi Jaya is like the manager and cashier to obtain accurate and correct transaction data.

c) A literature Review
   At this stage, the researcher conducts a literature review by collecting data from several related reference books, as well as other sources that support the implementation of the research conducted. The books and references used are related to data mining theory, a priori algorithms, data search, and android programming.

3.2. Metode Data Mining
   This study looks for the relationship of two or more items on sales transaction data in CV. Dian Abadi Jaya. Not all attribute data contained in transaction data will be searched for relationships, only a few attributes that match the search for patterns of purchasing spare parts and types of repair services provided by the company[5]. The stages that will be carried out in the data mining process are as follows:

a) Data Source
   Research material or data is the object of research. In the research conducted, the data used is transaction data on sales of spare parts and vehicle repair services. Data obtained from the CV workshop by conducting observations and interviews with the company, Dian Abadi Jaya Bandar Lampung the transaction data used is sales transaction data for 3 months, from September 2021

to November 2021, as many as 2720 transaction records.

b) Data Selection

The data selection process is the selection of relevant data for research. In the initial data, there are many attributes that are not used in the data mining process. Therefore, several initial attributes were selected, namely No_Transaction, Item Name, and Service_Name. The following is an explanation of the selected attributes in table 1.

Table 1. Selected Attributes in Sales Transaction Data

No	Attribute name	Explanation
1	Transaction Number	This attribute is used as an attribute that connects the Purchase_Goods table and the Purchase_Services table.
2	Item name	The item name attribute is used to find out the name of the item purchased on each sales transaction number. This attribute is obtained from the Purchase_Goods table.
3	Name service	The item name attribute is used to find out the name of the service performed on each sales transaction number. This attribute is obtained from the Purchase_Services table.

c) Preprocessing Data

The initial data obtained in the study is raw data that is still irregular, so it must be processed first so that it can be used in the data mining process[11][27]. The data cleaning process includes the cleaning of duplicate, irrelevant, and missing data. This is because data that is not duplicate, not missing value, and not redundant is the initial requirement in data mining[28].

In this process, irrelevant data is removed so that the final result is 1160 sales transaction records that are ready to be processed in the data mining process. After cleaning the data, the next step is to change the shape or transformation of the data by calculating the number of transactions for each type of item and the type of repair into tabular data on all transactions so that it can be processed with data mining applications. The results of a series of processes that have been carried out produce data that is ready to be processed on 1160 transaction records with the criteria of 22 spare parts data attributes and 13 vehicle repair service data

attributes, for a total of 35 attributes, which are described as follows.

Tabel 2 Attributes to be Formed Association Pattern

No	Attribute name
1	Air Accu
2	Baut
3	Boklam K2 Philips
4	C.Cleaner Megacool
5	Castrol Magnatec
6	F Oli Avz/Xn/Agya
7	F Oli Inv 90915-Yzz2
8	F Oli Yaris/Vios
9	Fastron 10/40
10	Full Penetrat Megacool
11	Kampas Rem
12	Karet Stabil
13	Lem Tribon New
14	M.Rem Petro Kcl
15	Meditran Sc
16	Nitrogen
17	Prima Xp 10/40
18	Prima Xp 20/50
19	Rad Coolant Ktb Hijau
20	Shell Hx5
21	Shell Hx6
22	Shell Hx7 5w/40
23	Jasa Balance
24	Jasa Ganti Busi
25	Jasa Ganti F Oli
26	Jasa Ganti F Solar/Bensin
27	Jasa Ganti Kampas Rem
28	Jasa Ganti Oli Bs
29	Jasa Ganti Vanbelt
30	Jasa Kuras Radiator
31	Jasa Oli Matic
32	Jasa Service
33	Jasa Scan
34	Jasa Spooring
35	Jasa Tune Up

The table explains, each transaction[29] will be transformed into tabular data with the criteria that if in one transaction there is one type of item or repair service performed, it is calculated as 1, and if in one transaction there is not one item purchased, it is calculated as 0. Data that is already in tabular form is then stored in Excel format. (*.CSV) so that it is ready to be processed by data mining applications.

d) Mining process

Is a process where the method used will be calculated to extract valid data patterns. In this study[5], a data mining process[30] will be carried out with association rules and a priori algorithms[21] to find item association relationships in CV companies. Dian Abadi Jaya.

e) Pattern or model

After the mining process has been carried out, a pattern or model that fits the problem will be obtained.

f) Interpretation or evaluation

It is a stage for estimating, identifying, and interpreting patterns that are truly important and represent knowledge based on measures of degree of importance. The evaluation uses the lift calculation as a reference for the strength of the relationship between the items.

g) Knowledge

This is the last stage that can involve knowledge directly. It then combines knowledge into other systems for further processing. Knowledge of applying the pattern "If you buy x then buy y".

2.3 Testing tools and materials

a) Testing tools

The tools used in this study are divided into two categories, namely hardware and supporting software. The following hardware is used in the study:

- Prosessor *Ryzen* 5 2400G,
- Memori 20GB,
- Internet Connectivity
- *Storage* 250Gb SSD dan 2TB HDD,
- *Keyboard* dan *Mouse*

The supporting software in this study is as the follows:

- Operating system, *Windows* 10 64bit.
- *Microsoft Excel* 2016
- IntelliJ IDEA 2021.3.2
- *Bluestack* 4 with *android* oreo.

b) Materials

The material used in this study is sales transaction data obtained from interviews and observations with related parties in the CV company. Dian Abadi Jaya. The data used is 1160 sales transaction data records for three months starting from September 2021 until November 2021.

IV. RESULTS

The results of the research conducted using a simulation of the minimum frequent item value of 20 transactions for the entire dataset so that by using the support formula, the values obtained are:

$$Minimum\ Support = \frac{\text{Minimal } Frequent\ Itemset}{\text{Total Transaksi}}$$

$$= \frac{20}{1160} x\ 100\%$$
$$= 1,7\%$$

From the calculation results, the support value is 1.7%, which will be the minimum support value in this study. After the support value is determined, the next step is to find the itemset according to the minimum support value.

a) 1-*itemset candidate search*

The candidate 1 itemset can be calculated by the following equation (1) formula :

$$support\ \text{A} = \frac{transaction\ amount\ contains\ \text{A}}{\text{total transactions}}$$

Tabel 3. 1-itemset candidate calculation result

Atributte candidate	Amount	Support
Air Accu	31	2,7%
C.Cleaner Megacool	45	3,9%
Castrol Magnatec	67	5,8%
F Oli Avz/Xn/Agya	64	5,5%
F Oli Inv 90915-Yzz2	24	2,1%
Fastron 10/40	24	2,1%
Full Penetrat Megacool	20	1,7%
Prima Xp 20/50	29	2,5%
Shell Hx5	57	4,9%
Shell Hx6	58	5,0%
Jasa Balance	23	2,0%
Jasa Ganti F Oli	271	23,4%
Jasa Ganti F Solar/Bensin	27	2,3%
Jasa Ganti Kampas Rem	55	4,7%
Jasa Ganti Oli Bs	164	14,1%
Jasa Ganti Vanbelt	20	1,7%
Jasa Service	326	28,1%
Jasa Scan	36	3,1%
Jasa Spooring	132	11,4%
Jasa Tune Up	104	9,0%

From table 3, The results obtained on the attributes that meet the minimum value of 15 frequent itemset or minimum support of 1.7% are 20 attributes.

b) 2-*Itemset Candidate Search*

At the stage of forming the 2-itemset combination, the attributes used are based on the attributes selected in table 3, and the combination of these attributes will be calculated in total with the overall transaction data for repairing and selling vehicle parts. If in one transaction record there are two combinations sold, it is counted as 1, and if in one transaction it is not sold, it is counted as 0.

Candidate 2 itemset can be calculated by equation (2) below.

$$Support\ \text{A} => \text{B}$$
$$= \frac{transaction\ amount\ contains\ \text{A and B}}{\text{transaction number}}$$

Based on the results obtained for the 2-itemset combination that meets the minimum value of 20 frequent itemset or minimum support of 1.7%, the results obtained are 24 combinations as follows:

Tabel 4. 2-*itemset* candidate calculation result

Atributte candidate	Amount	Support
C.Cleaner Megacool, Jasa Service	21	1,8%
C.Cleaner Megacool, Jasa Tune Up	20	1,7%
Castrol Magnatec, Jasa Ganti F Oli	29	2,5%
F Oli Avz/Xn/Agya, Jasa Ganti F Oli	63	5,4%
F Oli Inv 90915-Yzz2, Jasa Ganti F Oli	24	2,1%
Shell Hx5, Jasa Ganti F Oli	27	2,3%
Shell Hx6, Jasa Ganti F Oli	27	2,3%

Jasa Balance, Jasa Spooring	20	1,7%
Jasa Ganti F Oli, Castrol Magnatec	29	2,5%
Jasa Ganti F Oli, F Oli Avz/Xn/Agya	63	5,4%
Jasa Ganti F Oli, F Oli Inv 90915-Yzz2	24	2,1%
Jasa Ganti F Oli, Shell Hx5	27	2,3%
Jasa Ganti F Oli, Shell Hx6	27	2,3%
Jasa Ganti F Oli, Jasa Ganti Oli Bs	108	9,3%
Jasa Ganti F Oli, Jasa Service	22	1,9%
Jasa Ganti F Oli, Jasa Tune Up	36	3,1%
Jasa Ganti Oli Bs, Jasa Ganti F OliJasa	108	9,3%
Service, C.Cleaner Megacool	21	1,8%
Jasa Service, Jasa Ganti F Oli	22	1,9%
Jasa Service, Jasa Spooring	41	3,5%
Jasa Spooring, Jasa Balance	20	1,7%
Jasa Spooring, Jasa Service	41	3,5%
Jasa Tune Up, C.Cleaner Megacool	20	1,7%
Jasa Tune Up, Jasa Ganti F Oli	36	3,1%

c) Establishment of Association Rules

After obtaining a combination that meets the minimum support value, the next step will be the formation of association rules by calculating the confidence value for each 2-itemset combination candidate. In this study, a minimum confidence value of 40% will be set, which means that all patterns that have a confidence value of less than 40% will be trimmed.

d) Associate candidates to be searched for can be calculated by equation (3) below:

$$= \frac{Confidence \ A => B}{transaction \ amount \ contains \ A \ end \ B} \times 100\%$$
$$\frac{}{transaction \ amount \ contains \ A}$$

The results obtained in the 2-item combination that meet the minimum value of 40% confidence are 9 patterns. After obtaining a pattern that meets the minimum support and minimum confidence values, the next step will be to calculate the strength of each rule pattern that has been generated. A strong rule is a rule that has a lift ratio value of more than or equal to 1 (lift >= 1).

The lift ratio value can be calculated by equation (4) as follows

$$Lift \ (A => B) = \frac{Support \ A \ and \ B}{Support \ A \ x \ Support \ B} \quad (4)$$

Based on the results of the calculations, there are 9 patterns that have a value above 1, so that it can be stated that the pattern has met the lift ratio criteria that have been set.

e) **Knowledge Representation**

After carrying out the various stages contained in the association pattern search stage using the a priori algorithm with the criteria for calculating the minimum support value of 1.7%, the minimum confidence of 40% and the lift ratio of more than 1, we have obtained a pattern that meets the criteria of 9 combination patterns. itemset with APRIORI calculation

The following is an explanation of each purchase pattern that has been obtained as follows:

1. **If you buy C. Cleaner Megacool, then buy Service.**
   The pattern above means that if a customer buys C. Cleaner Megacool, it will have the possibility of buying the service at the same time with a confidence level of 47%. The frequency of sales made in the purchase pattern was 21 transactions, with C.Cleaner Megacool as the antecedent and Services as the consequent.

2. **If you buy *C. Cleaner Megacool*, then buy *Jasa Tune up*.**
   The pattern above means that customers who buy C. Cleaner Megacool will have the possibility of buying Tune Up Services simultaneously with a confidence level of 44%. The frequency of sales made in the purchase pattern is 20 transactions with the criteria of C. Cleaner Megacool as an antecedent of Tune Up Services as a consequence.

3. **If you buy *Castrol Magnatec*, then buy *service F Oli***
   The above pattern means that if a customer buys Castrol Magnatec, it will have the possibility of buying an F Oil Change Service at the same time with a confidence level of 43%. The frequency of sales made in the purchase pattern is 29 transactions with the criteria of Castrol Magnatec as an antecedent and Oil Change Service as a consequence.

4. **If you buy *Oli Avz/Xn/Agya*, you buy *service F Oli***
   The above pattern has a meaning, namely that if a customer buys Bs Oil Change Services, it will have the possibility of buying Services simultaneously with a 95% confidence level. The frequency of sales made in the purchase pattern is 37 transactions with the criteria of Bs Oil Change Service as the antecedent and F Oil Change Service as the consequent.

5. **If you buy *F Oli Inv 90915-Yzz2*, then buy service *Ganti F Oli***
   The above pattern means that if a customer buys F Oli Inv 90915-Yzz2, it will have the possibility of buying F Oil Change Services at the same time with a confidence level of 100%. The frequency of sales made in the purchase pattern is 24 transactions with the criteria of F Oli Inv 90915-Yzz2 as the antecedent and F Oil Change Service as the consequent.

6. **If you buy *Shell Hx5*, then buy service *F Oli***
   The above pattern means that if a customer buys Shell Hx5, it will have the possibility of buying an F Oil Change Service at the same time with a confidence level of 47%. The frequency of sales made in the purchase pattern is 27 transactions with Shell Hx5 criteria as the antecedent and F Oil Change Service as the consequence.

7. **If you buy *Shell Hx6*, then buy service *F Oli***

    The above pattern means that if a customer buys Shell Hx6, it will have the possibility of buying an F Oil Change Service at the same time with a confidence level of 47%. The frequency of sales made in the purchase pattern is 27 transactions with Shell Hx6 criteria as the antecedent and F Oil Change Service as the consequence.
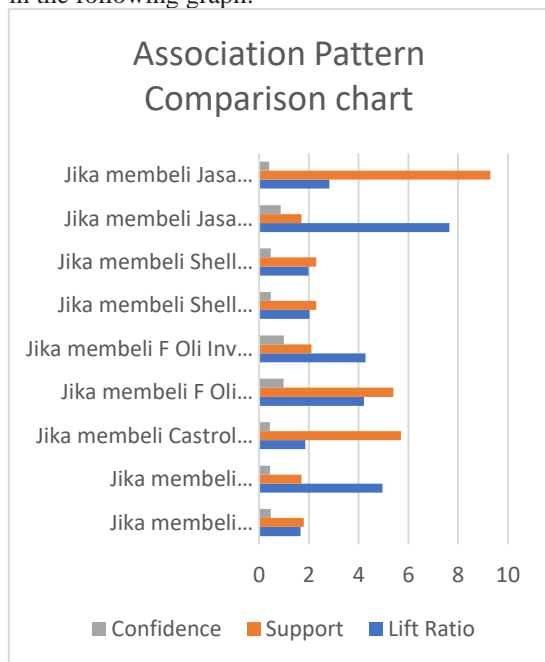
8. **If you buy service *balance*, then buy service *spooring*.**

    The pattern above has a meaning, namely that if a customer buys balance services, it will have the possibility of buying spooring services simultaneously with a confidence level of 87%. The frequency of sales made in the purchase pattern is 20 transactions with the criteria of Balance Services as an antecedent and Spooring Services as a consequent.

9. **If you buy service *F Oli*, you buy service *Oli Bs***

    The pattern above has a meaning, namely that if a customer buys an F Oil Change Service, it will have the possibility of buying a Bs Oil Change Service at the same time with a confidence level of 40%. The frequency of sales made in the purchase pattern is 108 transactions with the criteria of Oil Change Service as an antecedent and Bs Oil Change Service as a consequence.

    The results of the correlation between the values of support, confidence, and lift ratio are described in the following graph:

Figure :3 : Association Pattern Comparison Chart

    Figure 3 shows a graph of the values for each pattern based on table 7, namely

From figure 3:
1. "If you buy C. Cleaner Megacool, then buy services" has a confidence value of 47% with a support of 1.8% and a lift ratio value of 1.660532.
2. If you buy C. Cleaner Megacool, then you buy Tune Up services, which has a confidence value of 44% with a support of 1.7% and a lift ratio value of 4.957265.
3. If you buy Castrol Magnatec, then buy "F Oil Change services it has a confidence value of 43% with a support of 1.7% and a lift ratio value of 1.8527.
4. If you buy Avz/Xn/Agya F Oli, then buy F Oil Change Services with has a confidence value of 98% with a support of 5.4% and a lift ratio value of 4.213561.
5. If you buy F Oli Inv 90915-Yzz2, then you buy F Oli Change services. has a confidence value of 100% with a support of 2.1% and a lift ratio value of 4.280443.
6. If you buy Shell Hx5, then you buy "Oil Change Service" has a confidence value of 47% with a support of 2.3% and a lift ratio value of 2.027578.
7. If you buy Shell Hx6, you buy "Oil Change Service, which has a confidence value of 47% with a support of 2.3% and a lift ratio value of 1.99262.
8. If you buy Balance Services, you buy Spooring Services, which has a confidence value of 87% with a support of 1.7% and a lift ratio value of 7.641634.
9. If you buy an F Oil change service, then you buy an "Bs Oil Change Service" has a confidence value of 40% with 9.3% support and a lift ratio value of 2.818828.

V. CONCLUSION

    Based on the research that has been done, it can be concluded that: Frequent itemset was successfully generated through various processes, namely preprocessing data, which includes data cleaning and data transformation, then the process of extracting data information is carried out using data mining algorithms with an a priori algorithms. The process of searching for association information with a priori algorithm on sales transaction data can be used by CV companies. Dian Abadi Jaya, analyzing transaction data and generating frequent item relationships. Data mining with a priori algorithms is not only applied to manual analysis methods, but can also be applied to Android-based mobile applications whose results are close to manual calculations. The association search results use a priori calculations with a minimum frequent itemset value of 15 transactions with a minimum support criterion of 0.08982, a minimum confidence of 40% and shows a lift value above 1. There are 9 patterns of association

obtained, the results of which are in accordance with manual calculations, namely 9 patterns, so it can be said that it is in accordance with what is expected.

Developers can process more data on a scale, including annual data, so that the results of association rules are more accurate. The association search process can be improved by analyzing data for each month's available transactions. Sorting sales data can be upgraded to include association rules for each vehicle.

REFERENCES

[1]   Islamiyah, P. L. Ginting, N. Dengen, and M. Taruk, "Comparison of Priori and FP-Growth Algorithms in Determining Association Rules," 2019, doi: 10.1109/ICEEIE47180.2019.8981438.

[2]   J. Homepage, E. Junianto, and R. Rachman, "IJCIT (Indonesian Journal on Computer and Information Technology) Penerapan Data Mining Metode Apriori dan FP-Tree Pada Penjualan Media Edukasi (Studi Kasus : Oisha Smartkids)," 2020.

[3]   T. Y. Prawira, S. Sunardi, and A. Fadlil, "Market Basket Analysis To Identify Stock Handling Patterns & Item Arrangement Patterns Using Apriori Algorithms," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 6, no. 1, 2020, doi: 10.23917/khif.v6i1.8628.

[4]   D. Nofriansyah, M. Yetri, K. Erwansyah, and _ S., "Penerapan Data Mining Dalam Menganalisa Data Penjualan Untuk Mendapatkan Pola Rekomendasi Penjualan Menggunakan Algoritma Apriori Pada K3 Mart," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 2, 2019, doi: 10.53513/jis.v18i2.157.

[5]   M. Arhami, M. Kom, and S. M. Nasir, *Data Mining-Algoritma dan Implementasi*. 2020.

[6]   Murnawan, A. Sinaga, and U. Nughraha, "Implementation of Apriori Algorithm for determining purchase patterns in one transaction," *Int. J. Eng. Technol.*, vol. 7, no. 4, 2018, doi: 10.14419/ijet.v7i4.33.23560.

[7]   Y. Sutisnawati and M. Reski, "Looking for Transaction Data Pattern Using Apriori Algorithm with Association Rule Method," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 662, no. 2, doi: 10.1088/1757-899X/662/2/022078.

[8]   A. B. Rao, J. S. Kiran, and P. G, "Application of market–basket analysis on healthcare," *Int. J. Syst. Assur. Eng. Manag.*, 2021, doi: 10.1007/s13198-021-01298-2.

[9]   Y. A. Ünvan, "Market basket analysis with association rules," *Commun. Stat. - Theory Methods*, vol. 50, no. 7, 2021, doi: 10.1080/03610926.2020.1716255.

[10]   W. Nengsih, "ANALISA AKURASI PERMODELAN SUPERVISED DAN UNSUPERVISED LEARNING MENGGUNAKAN DATA MINING," *Sebatik*, vol. 23, no. 2, 2019, doi: 10.46984/sebatik.v23i2.771.

[11]   I. Enabled, P. Location, and M. View, "A Study on Market Basket Analysis and Association Mining," *Proc. Natl. Conf. Mach. Learn.*, no. April, 2019.

[12]   A. Hodijah and U. T. Setijohatmo, "Analysis of frequent itemset generation based on trie data structure in Apriori algorithm," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 19, no. 5, 2021, doi: 10.12928/TELKOMNIKA.v19i5.19273.

[13]   M. Heydari and A. Yousefli, "A new optimization model for market basket analysis with allocation considerations: A genetic algorithm solution approach," *Manag. Mark.*, vol. 12, no. 1, 2017, doi: 10.1515/mmcks-2017-0001.

[14]   Abdul Kadir, "Pengenalan Sistem Informasi Edisi Revisi," *Edisi Revisi*. 2014.

[15]   F. Ayu and N. Permatasari, "PERANCANGAN SISTEM INFORMASI PENGOLAHAN DATA PRAKTEK KERJA LAPANGAN (PKL) PADA DEVISI HUMAS PT. PEGADAIAN," *Intra-Tech*, vol. 2, no. 2, 2018.

[16]   M. A. Christian, N. Nathanael, A. Mauliani, A. Indrawati, L. P. Manik, and Z. Akbar, "Real Market Basket Analysis using Apriori and Frequent Pattern Tree Algorithm," 2021, doi: 10.1145/3489088.3489133.

[17]   A. Nugroho, *Rekayasa Perangkat Lunak Berorientasi Objek dengan Metode USDP (Unified Software Development Process)''*. 2010.

[18]   M. Qisman, R. Rosadi, and A. S. Abdullah, "Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores)," in *Journal of Physics: Conference Series*, 2021, vol. 1722, no. 1, doi: 10.1088/1742-6596/1722/1/012020.

[19]   B. R. B. Purba, N. A. Hasibuan, G. L. Ginting, and S. Suginam, "Implementasi Algoritma Apriori Untuk Mencari Relasi Pada Transaksi Pembelian Alat-Alat Kesehatan," *JURIKOM (Jurnal Ris. Komputer)*, vol. 5, no. 3, 2018.

[20]   M. N. Muhammad Arhami, "Data Minig

Algoritma dan Implementasi," *DATA Min. Algoritm. dan Implementasi*, 2020.

[21]   T. Astuti and B. Puspita, "Analysis of Customer Transaction Data Associations Based on The Apriori Algorithm," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 3, no. 1, 2020, doi: 10.47738/ijiis.v3i1.4.

[22]   N. N. A. Sjarif, N. F. M. Azmi, S. S. Y. And, and D. H. Ten Wong, "A review of market basket analysis on business intelligence and data mining," *Int. J. Bus. Intell. Data Min.*, vol. 18, no. 3, 2021, doi: 10.1504/IJBIDM.2021.114475.

[23]   N. F. FAHRUDIN, "Penerapan Algoritma Apriori untuk Market Basket Analysis," *MIND J.*, vol. 1, no. 2, 2019, doi: 10.26760/mindjournal.v4i1.13-23.

[24]   S. L. Mortale and M. J. Darak, "Clustering and pattern mining of customer transaction data using Apriori algorithm," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, 2019, doi: 10.35940/ijrte.C6424.098319.

[25]   J. Sumpena and N. K. H, "Analisis Prediksi Kelulusan Siswa PKBM Paket C dengan Metoda Algoritma Naïve Bayes," *Tedc*, vol. 13, no. 2, 2019.

[26]   G. & S. D. I. Gunadi, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Penjualan Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (FP-GROWTH) :," *Telemat. Mkom*, vol. 4, no. 1, 2016.

[27]   H. Al Islami, T. Informatika, J. Raya, P. No, and T. Selatan, "Penerapan Metode Asosiasi Datamining Market Basket Analysis Menggunakan Algoritma Apriori Untuk Melakukan Pola Analisis Belanja Konsumen," *J. Artif. Intellegence Innov. Appl.*, vol. 1, no. 1, 2020.

[28]   F. Antho, D. Renaldi, E. -, and Y. -, "PENERAPAN ASSOCIATION RULE DATA MINING UNTUK REKOMENDASI PRODUK KOSMETIK PADA PT. FABIANDO SEJAHTERA MENGGUNAKAN ALGORITMA APRIORI," *ALGOR*, vol. 2, no. 1, 2020, doi: 10.31253/algor.v2i1.437.

[29]   T. Hariguna, U. Hasanah, and N. N. Susanti, "Sales Transaction Data Analysis using Apriori Algorithm to Determine the Layout of the Goods," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 1, no. 1, 2018, doi: 10.47738/ijiis.v1i1.19.

[30]   S. Syafrianto and D. Ayniyah, "IMPLEMENTASI DATA MINING MENGGUNAKAN METODE ALGORITMA APRIORI PADA TOKO DHURROH ELEKTRONIK," *J. Tek.*, vol. 13, no. 2, 2021, doi: 10.30736/jt.v13i2.624.