

Analysis of Shopping Cart At Drugs Store By Using An Apriori Algorithm

Adi Panca Saputra Iskandar[1], Kheri Arionadi Shobirin[2], and Komang Oka Saputra[3]
[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3] Department of Electrical and Computer Engineering, Udayana University
E-Mail: adipanca@gmail.com

Abstract-- Competition in the business world in the pharmacy industry, demands the Pharmacist to determine a strategy that can increase drug sales. One way is to overcome the problems that occur in drugs store such as problems and sales of drug recommendations to suppliers who still use the prediction system of pharmacists. Data growth is less so utilized, data and information become useless like waste, sales strategy only as a little assumption without information and knowledge in server, consequently sales down and not range and items become expired and stock procurement is not according to market demand. Extracting data or an effort to retrieve valuable knowledge and information in a large database is called data mining or Knowledge Discovery in Database or usually shortened as KDD . One of the most popular algorithm in data mining technic is Apriori Algorithm, while the discovery of “relational combination pattern among itemset used Assosiation Rules”. Data mining has been implemented into the various fields like : business or trade, education and telecommunication. In bussiness for instance, the implementation result of data mining use ‘algorithm Apriori which can give a hand to help the Businessmen make decision on supplies. For example, the necessity of supplies system in a drugstore as one of the meical stuff supplier, and to determine which product as the priority should be supplied to anticipate out of stock of supplies availability in the store, as the results will also affect to the consumer service and daily income. Drugs are essential unit should be supplied and being and essential factor which will impact to the consumer trust to a hospital or another medical service. That is why the availability of medical drugs in drugstores is completely needed to support the succes of distribution to the consumers, so the activity of medical service to consumers run thoroughly. In this case, data mining is seen as able to buildt intelligent business environment as solution for competing increased competition among the drugstores in future.

Index Term--- Data Mining, Apriori Algorithm
I. INTRODUCTION

The development and business competition in the field of economy by utilizing information technology makes the competition increasingly tight and more open in meeting the demands of customers who are also higher. Competition in business can not be separated from the information technology that became a hot topic today. The number of competition in the business world, espe-cially in the pharmacy industry, demanding drug store Pharmacist to determine a strategy that can increase drug sales. One way that can be used to increase drug sales is to overcome the problems that occur in drug store, such as problems of sales patterns and recommendations of drug purchases to suppliers who have been using the prediction system from pharmacists drug store. Determi-nation of drug sales pattern and drug purchase recom-mendation can be done by digging information from drug sales transaction data in Anggita Pharmacy which number is increasing. The determination of this drug selling pattern has several benefits, one of which determines the product that is often purchased along with other products so it can help Apotek Aggita Pharmacist in purchasing drug against supplier to overcome the problem of excess or shortage of drug stock.

To find out what medicines are bought by the consumers, the market basket analysis technique is analyzed from consumer buying habits. Application of the Apriori Algorithm, assisting in forming possible candidate item combinations, then testing whether the combination meets the minimum support and con fi dence parameters which are the threshold values given by the user. Activities and transactions in an organization are generally the same, different are the constituent elements, such as different government organizations with business organizations, as well as educational organizations. Different organizational characteristics will thus result in varied business mod-eling, and this difference is also seen in data and data flow. Of course this situation one time

become an obstacle factor in improving services as the number of transactions and types of items and itemset transactions that occur and stored in a certain time, making it difficult for the pharmacist in analyzing the type of item and itemset which goods are most preferred or not interested consumers. Therefore it is necessary to have a system capable of performing Shopping Cart Analysis At Apotek By Using Apriori Algorithm

II. PURPOSE OF PAPER

Implement Data Mining on the database of drug item sales transactions, Apply the Apriori Algorithm to determine the itemset's high frequency to predict the inventory of goods in the future and Know how to determine the pattern of drug sales with an Apriori Algorithm based on the amount of drug sales transaction data.

III. LITERATUR REVIEW

Data mining is a term used to describe the discovery of knowledge in a database. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases. Based on the definitions that have been submitted, the important things associated with Data Mining are:

1. Data mining is an automated process of existing data.
2. Data to be processed in the form of very large data.
3. The purpose of data mining is to get a relationship or pattern that will probably provide useful indications.

Data mining is divided into several groups based on the tasks that can be done, namely:

1. Description

Sometimes researchers and analysts simply want to try to find data to describe the patterns and trends contained in the data. For example, the polling officers may not be able to determine the information or the fact that who is not professional enough will be little supported in the presidential election. Descriptions of patterns and trends often provide a possible explanation for a pattern or trend.

2. Estimates

Estimates are almost identical to the classification, unless the target variable estimates are more numerical than the direction of the category. The model is built using a complete record that provides the value of the target variable as a prediction. Next, in the next review the estimated value of the target variable is based on the predicted variable value. For example, an estimate of systolic blood pressure in hospital patients will be based on patient age, sex, body weight index, and blood sodium level. The relationship between systolic blood pressure and predicted variable values in the learning process will result in estimation models. The resulting estimation model can be used for other new cases.

3. Prediction.

The prediction is almost the same as the classification and estimation, except that in the predictive value of the will there will be in the future. Examples of business and research predictions are: a. Predicted rice price in three months to come. B. Predict the percentage increase in traffic accidents next year if the lower limit of speed is raised. Some methods and techniques used in the classification and estimation can also be used (for the right circumstances) for prediction.

4. Classification

In classification, there are target categories of variables. For example, the classification of income can be separated into three categories, namely high income, medium income, and low income. Other examples of classification in business and research are: a. Determining whether a credit card transaction is a fraudulent transaction. B. Estimating whether a mortgage application by a customer is a good or bad credit. C. Diagnosing a patient's illness to get into which category of illness.

5. Clustering

Clustering is the grouping of records, observations, or observing and forming classes of objects that have similarities. The cluster is a collection of records that resemble one another and have an incompatibility with records in other clusters. The clustering is different from the classification of the absence of the target variable in the clustering. Clustering does not attempt to classify, estimate, or predict the value of the target variable. However, the clustering algorithm tries to divide the entire data into homogeneous groups, where the similarity in a group will be maximized, while the similarity with the records in the other group will be of minimal value. Examples of business and research diversification are: a. Get consumer groups to target marketing from one product to a company that does not have large enlarged funds. B. For the purpose of accounting audit, which is to separate the financial behavior in good and suspicious. C. Perform clustering of the expression of the gene, to attain the behavioral resemblance of a large number of genes.

6. Association

Association The task of association in data mining is to find attributes that appear at a time. In the business world more commonly called shopping cart analysis. Examples of associations in business and research are: a. Examining the number of subscribers from mobile telecommunication companies that are expected to provide a positive response to the offer of service upgrades provided. B. Determine items in supermarkets purchased simultaneously and which are never purchased simultaneously.

Apriori algorithm includes the type of association rules in data mining. The rules that state the association between several attributes are often called affinity analysis or market basket analysis. Association analysis or association rule mining is a data mining technique for finding the rules of a combination of items. One of the stages of association analysis

that attracts many researchers to produce an efficient algorithm is the analysis of frequent pattern mining. The importance of an association can be identified by two benchmarks, namely: support and confidence. Support is the percentage of combinations of items in the database, whereas confidence is a strong inter-item relationship in association rules.

IV. METODOLOGI

First, Selecting and clearing the data to be analyzed, then looking for all types of items that are in the list of sales transactions, then look for the number of each item that existed in every sale of drug store.

Second, Of the items that have been selected sales are formed a combination of two items. With the support specified then selected is some data two items (the formation of a combination of two items). And so on until the combination of the maximum limit of transaction items. To perform the data processing of drug transactions it is necessary to describe the work stages that can be done on a priori algorithm like Figure 1 below :

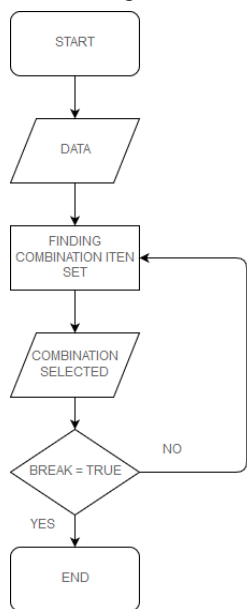


Figure 1. Flowchart Apriori

Before searching the pattern of the transaction data, first look for all items of drugs in the transaction as in table 1 as well as determine the support per item type of drugs - this drug is also looking for a combination of items that meet the minimum requirements of the support value in the database.

The value formula supports an item :

$$Support(A) =$$

$$\frac{Jumlah\ transaksi\ yang\ mengandung\ A}{\sum Transaksi}$$

$$Support(A, B) =$$

$$\frac{\sum Transaksi\ yang\ mengandung\ A\ dan\ B}{\sum Transaksi}$$

The value formula supports two item

In the system we use some table from transaction the other table of the system can be shown on Table 1.

Table 1. Master Drug

kode_barang	nama_barang
B01	Amoxicilin
B02	Asamefenamat
B03	Kaditric
B04	Sangobion
B05	Redoxon
B06	Nonflamin
B07	Cefadroxil

Table 1 is a master obat, where each transaction will have a relationship on the master table of this drug

Table 2. Purchasing Transaction

Transaction	Sell Of Item
1	B03, B05, B04
2	B01, B06, B04
3	B04, B07, B02, B06
4	B05, B04, B07, B02
5	B02, B01, B03
6	B06, B01, B02, B07
7	B07, B04

Table 2 shows the transaction data amounted to 7 data with each - each transaction more than 1 item of medicine or more. After that do the next stage of the apriori algorithm is to determine the value of Φ and arrange transactions in Table 3.

Table 3. Formula Summary Transaction

Transactio n	b0 1	b0 2	b0 3	b0 4	b0 5	b0 6	b0 7
1	0	0	1	1	1	0	0
2	1	0	0	1	0	1	0
3	0	1	0	1	0	1	1
4	0	1	0	1	1	0	1
5	1	1	1	0	0	0	0
6	1	1	0	0	0	1	1
7	0	0	0	1	0	0	1
8	0	0	1	0	1	0	1
9	1	1	0	0	0	1	0
10	0	1	0	1	0	0	0
11	0	1	1	0	1	1	0
12	1	1	0	0	0	1	0
13	1	1	0	1	0	1	0
14	0	1	1	1	1	0	1
sum()	6	10	5	8	5	7	6

Transaction	B01	B02	f
1	0	0	S
2	1	0	S
3	0	1	S
4	0	1	S
5	1	1	P
6	1	1	P
7	0	0	S
8	0	0	S
9	1	1	P
10	0	1	S
11	0	1	S
12	1	1	P
13	1	1	P
14	0	1	S
Total P			5

Next item 1 in combination with item 3, the result of the combination can be seen in Table 5.

Table 5. Combination Item 1 and Item 3

Transaction	B01	B02	f
1	0	1	S
2	1	0	S
3	0	0	S
4	0	0	S
5	1	1	P
6	1	0	S
7	0	0	S
8	0	1	S
9	1	0	S
10	0	0	S
11	0	1	S
12	1	0	S
13	1	0	S
14	0	1	S
Total P			1

This process will continue until all items have been combined and get the frequency of each expected combination.

The number of item sets frequency must be greater than the frequency of the set item then:

For k = 3 Form a table to form a set of 3 elements. S means no

Terms – Terms of Apriori to know :

1. I = The set of data being analyzed
2. D = The set of all transactions
3. Proper Subset = The Association of Pure
4. Item Set = The set of item / items in the I
5. K-item set = The number of elements in the set
6. Item Set Frekuensi = The number of transactions on I which contains the number of items of a particular set.
7. Φ = Frekuensi item set
8. Fk = The set of all items set frekuensi

Furthermore, existing transaction data will be combined with each item of goods available to know the frequency of items sold together:

If $\Phi = 4$ Then :

- a. For k = 1 All items the number of items setnya frequency then:

$$F_1 = \{\{amoxilin\}, \{Asamefenamat\}, \{Kaditic\}, \{Sangobion\}, \{Redoxon\}, \{Nonflamin\}, \{Cefadroxil\}\}$$

- b. For k = 2 Form table to form the set of 2 elements. S means no items are sold together, P means items are sold together.

Next item 1 in combination with item 2, the result of the combination can be seen in Table 4.

Table 4. Combination Item 1 and Item 2

items are sold together, P means items are sold together.

Next up is a combination of 3 items where item 1 is combined with item 2, and item 3 as shown in Table 6.

Table 6. Combination Item 1, Item 2 and Item 3

Transaksi	B01	B02	B03	f
1	0	0	1	S
2	1	0	0	S
3	0	1	0	S
4	0	1	0	S
5	1	1	1	P
6	1	1	0	S
7	0	0	0	S
8	0	0	1	S
9	1	1	0	S
10	0	1	0	S
11	0	1	1	S
12	1	1	0	S
13	1	1	0	S
14	0	1	1	S
	Σ		1	

Next item 1 in combination with item 3 and item 6, the result of the combination can be seen in Table 7.

Table 7. Combination Item 1, Item 2 and Item 6

Transaksi	B01	B02	B06	f
1	0	0	0	S
2	1	0	1	S
3	0	1	1	S
4	0	1	0	S
5	1	1	0	S
6	1	1	1	P
7	0	0	0	S
8	0	0	0	S
9	1	1	1	P
10	0	1	0	S
11	0	1	1	S
12	1	1	1	P
13	1	1	1	P
14	0	1	0	S
			Σ	4

This process will continue until all items have been combined and get the frequency of each expected combination.

The number of item sets frequency must be greater than the frequency of the set item then:

For k = 4 Form a table to form a set of 4 elements. S means no items are sold together, P means items are sold together. Next is a combination of 4 items in which item 1 is combined with item 2, item 3 and item 4 as shown in Table 8

Table 8. Combination Item 1, Item 2, Item 3 and Item 4

Transaksi	B01	B02	B03	B04	f
1	0	0	1	1	S
2	1	0	0	1	S
3	0	1	0	1	S
4	0	1	0	1	S
5	1	1	1	0	S
6	1	1	0	0	S
7	0	0	0	1	S
8	0	0	1	0	S
9	1	1	0	0	S
10	0	1	0	1	S
11	0	1	1	0	S
12	1	1	0	0	S
13	1	1	0	1	S
14	0	1	1	1	S
			Σ		0

The number of item set frequencies must be greater than the frequency of the set item then F4 is the empty set $\{\{\}\}$, so F5, F6, F7 are empty sets and the associative rule search stops. Next is the establishment of an association role.

If x Then y

X is antecedent (ss-s) and y is consequent (s)
There are 2 ie F2 and F3, F1 is not included because only one element

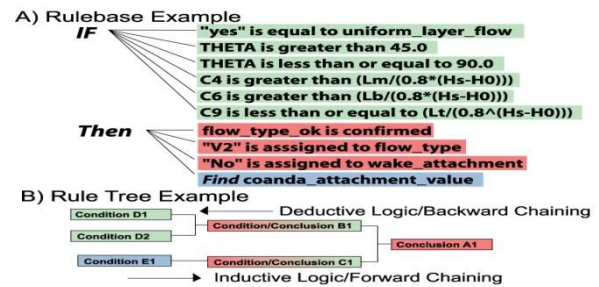


Figure 2. Make Role Asosiasi

Here is the result of making rule from Apriori algorithm implementation.

$F_2 = \{\{Amoxiclin, Asammefanamam\}, \{Amoxiclin, Nonflamin\}, \{Asammefanamam, Sangobion\}, \{Asammefanamam, nonflamin\}, \{Asammefanamam, Cefadroxil\}, \{Kaditic, Redoxon\}, \{Sangobion, Cefadroxil\}\}$

Compilation $\{Amoxiclin, Asammefanamam\}$

IF (ss-s) = *Amoxiclin*

IF (s) = *Asammefanamam*

Then => if *Pengantar Amoxiclin* then *Asammefanamam*

IF (ss-s) = *Asammefanamam*

IF (s) = *Amoxiclin*

Then => if *Asammefanamam* then *Amoxiclin*

Then the rule that has been formed must be validated by the formula as shown in Figure 3 and Figure 4:

$$\text{Confidence} = \frac{\sum \text{Item yang dibeli sekaligus}}{\sum \text{Jumlah Transaksi pada bagian Antecedent}} \times 100\%$$

Figure 3. Presentation Validation Confidence Association

$$\text{Support} = \frac{\sum \text{Item yang dibeli sekaligus}}{\sum \text{Jumlah seluruh Transaksi}} \times 100\%$$

Figure 4. Presentation Validation Support Association

The result of support and confidence validation can be shown in Table 9.

Table 9. Support and Confidence Validation Results

Rule	Support	Confidence
if Amoxicilin then Asamefenamat	$(5/14) * 100\% = 35.7\%$	$(5/6) * 100\% = 83.3\%$
If Asamefenamat then Amoxicilin	$(5/14) * 100\% = 35.7\%$	$(5/10) * 100\% = 50\%$

From the validation results, the highest support and confidence values are the strongest associative. The following a priori algorithm implementation implementation on drug sales, Can be shown in figure 5

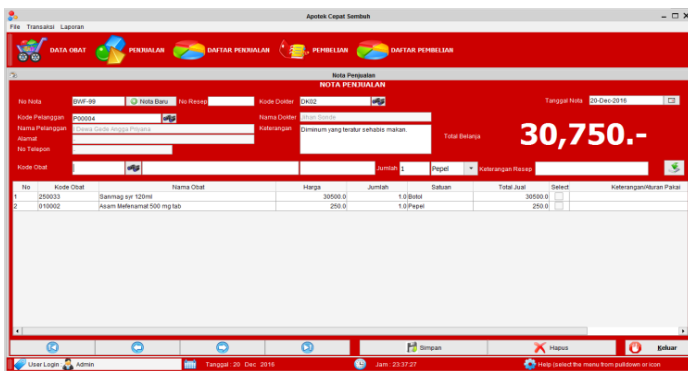


Figure 5. Form Sales With Apriori

In figure 4 it can be seen that the input to be filled is the name of the customer, the name of the doctor if accompanied by a prescription, the name of the drug to be purchased, if the drug to be purchased has been entered then the system will offer a drug that has a minimum support and confidence on the association

rule, The provision of associative buying recommendations can be shown in figure 6.

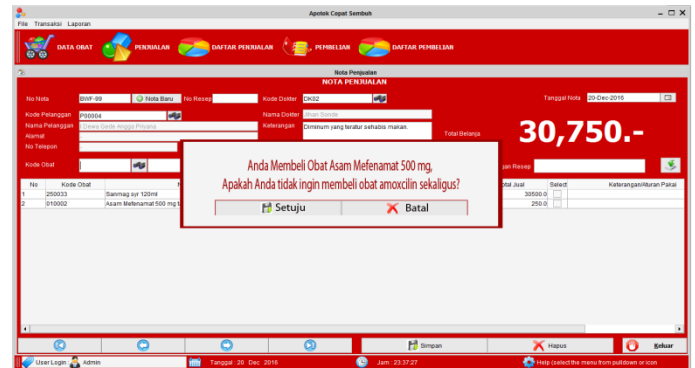


Figure 6. Form Sales With Rekomendasi

In Figure 6 proves that the implementation of data mining in particular by using associative techniques where this stage using a priori algorithm has been successfully tested, so that the end result reaches the desired expectations

V. CONCLUSION AND SUGESTION

The process of determining the pattern of drug purchase can be done by applying data mining with apriori algorithm method. By that method the determination of buying patterns can be done by looking at the result of the tendency of consumers to buy drugs based on itemset combination. Application of Apriori Algorithm in Data Mining technique is very efficient and can accelerate the process of formation of trend of sales item item combination pattern. Can serve as valuable information in making decisions to prepare the stock of drugs that the market needs. Thus reducing the expiration rate on drugs and existing data has been put to good use.

It is hipotesis in subsequent research to implement other than Apriori algorithm so that it can generate how effective algorithm used or can be done research to compare Apriori algorithm with Other Algorithm.

VI. REFERENCES

- [1] Putra, Desak Made Dwi Utami dan subanar. 2012. *Penerapan Algoritma Genetika Untuk Menyelesaikan Permasalahan Perawat dengan Fuzzy Fitness Function.*[Application of Genetic Algorithm for Solving Problems with Fuzzy Nurses Fitness Function] Yogyakarta: Universitas Gajah Mada
- [2] Puspaningrum, Wiga Ayu. Djunaidy, Arif. dan Vinarti, Retno Aulia. 2013. *Penjadwalan Mata Kuliah Menggunakan Algoritma Genetika di Jurusan Sistem Informasi ITS*[Course Scheduling Using Genetic

Algorithm in the Department of Information Systems ITS]. Surabaya: Institut Teknologi Sepuluh November

- [3] Yaqin, M. Ainul dan Lisbiantoro, Totok. 2012. **Optimasi Penjadwalan Perkuliahan Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang Menggunakan Algoritma Genetika Dengan Metode Seleksi Rank[Class Scheduling Optimization Department of Information of the State Islamic University of Maulana Malik Ibrahim Malang Using Genetic Algorithm Selection Method Rank].** Malang: Universitas Islam Negeri Maulana Malik Ibrahim Malang
- [4] Siswono, Teno dan Palgunadi, Sarngadi. 2014. **Analisa Kombinasi Algoritma Genetika dengan Algoritma Palgunadi untuk penjadwalan mata kuliah di Universitas Sebelas Maret[Analysis of Combination of Genetic Algorithm with Palgunadi algorithm for scheduling courses at the University of March].** Surakarta: Universitas Sebelas Maret
- [5] Anamisa, Devie Rosa dan Djunaidy, Arif. 2014. **Penyelesaian Penjadwalan Matakuliah Menggunakan Hibridisasi Algoritma Genetika Dan Algoritma Koloni Semut[Completion Course Scheduling Using Genetic Algorithms Hybridization And Ant Colony Algorithm].** Surabaya: Institut Teknologi Sepuluh November
- [6] Suyanto (2011), **Artificial Intelligence** (Searching, reasoning, planning, learning) Yogyakarta: Informatika