Hybrid-Dimension Association Rules for Diseases Track Record Analysis at Dr. Soetomo General Hospital

Silvia Rostianingsih, Gregorius Satia Budhi, Ni Wayan Yessy Dwijayanti
Informatics Department
Petra Christian University
Surabaya, Indonesia
silvia@peter.petra.ac.id, greg@peter.petra.ac.id

***Abstract*—Dr. Soetomo General Hospital already has a Hospital Information System which has been computerized for data storage of each recapitulated patient's disease. Since data recapitulated the patient's disease is increasing, Dr. Soetomo Hospital needed an application which can provide information for decision makers. One application that can help in decision making is data mining. Data mining with hybrid-dimension association rules method where the method to analyze the relationship between disease with patient identity. This method is made using apriori algorithm. Hybrid dimension association rules is a multidimensional association rule that allow the repetition of the predicate on each rules. This method is very suitable to describe the rules of the relationship with the patient's disease and patient's identity. Data that has been prepared is processed by the algorithm to generate frequent itemset so that will produce hybrid dimension association rules and the rules displayed by form of tables and graphs. The implementation of the algorithm is using Java Netbeans 6.5 software and Oracle 10g. By using the output of this application is in the form of association rules and charts, decision makers can know the relationship between the patient's disease with the patient's identity such as gender, status, domicile, education and occupation with other diseases.**

Keywords-Data Mining; Apriori; Hybrid Dimension Association Rules

I. INTRODUCTION

Dr. Soetomo is a top class general hospital in eastern Indonesia. Dr. Soetomo has a Hospital Information System (SIRS) which is computerized by using the Oracle database and Oracle Developer for its application. One of the outputs of the system is a recapitulation of the patient's disease. Dr. Soetomo requires software that is used as a tool that can provide useful knowledge for doctors and medical students to analyze the track record of disease to enable them to take further action.

Since the increasing number of patients with a variety of different illnesses, Dr. Soetomo need to know what steps can be taken in dealing with these problems. Therefore, the presence of historical data can provide information about patient's historical diseases. From the existing data, data mining can create a relationship between a disease with some attributes which are owned by the patient such as sex, area of residence, and work with other diseases in patients by using Hybrid-Dimension Association Rules.

II. HYBRID DIMENSION ASSOCIATION RULES

Multidimensional association rules is a rules that involve more than one dimension or predicate (such as rules that connect between what the customer purchased (buys) by customer age (age)). Purchases data from database and related information are stored in a relational database. The stored data are essentially multidimensional. (Jiawei & Kamber, 2001, p.251).

Hybrid-dimension association rule association rules is composed of two or more dimension or predicate and its predicate can happen over and over again in a single rule that is known as repeated predicated. Example of Hybrid-dimension association rules, are:

Age (X, "20 ... 29") ^ Buys (X, "laptop ")
=> Buys (X, "b / w printer")
[Support = 2%, Confidence = 60%]

Rule above consists of two predicates of age and buys, with repetition of the predicate buys. These relationships describe that 2% of customers are aged 20 to 29 years. There is a 60% chance that the buyer in this age while buy a laptop also purchase a printer.

The attribute database can be an absolute / definite or quantitative. Absolute attribute is an attribute that has no value (eg: work, brands, and colors). Absolute attribute is also called the nominal attribute; the value is the name of something. Quantitative attribute is an attribute that has value (eg: age, income, prices).

III. DATA MINING PROCESS

The sequences of the implementation system are:

A. Preprocessing Process

Process of transform the transaction data into a ready data-mining. The changed data are disease data with all the patient's attributes. Input: transaction date, tables name, patients' tables, diagnosis tables, and icd_x (disease coding) tables. Output: create table confirmation, conversion confirmation, conversion table results.

B. Frequent Itemsets Generate Process

The process of searching for the couple items those often appear simultaneously in the whole transaction. User is asked to input the minimum support and patient's attributes. Data on the selected table is filtered then sought his support,

one by one, the items that support is less than the minimum support at delete and the item that exceeds the minimum support is stored in an array then the array is still a 1-itemsets are processed in the join process. Join the process repeated until no more itemset that can be combined. Flowchart generate frequent itemsets can be seen in Figure 3.10.

Input: patients' attribute, frequent itemset generate, minimum support, conversion table results. Output: generate frequent itemset process confirmation, frequent itemset data.

C. Process Generate Association Rules:

Process of form an association rules. Input: frequent itemset table, association rules generate, the minimum confidence. Output: rules process confirmation, data association rules information.

IV. IMPLEMENTATION

The results from the application of the relationship or association between the data of patients with the disease described by using a hybrid method dimension association rules can be seen at Figure 2. There are two or more dimensions or predicate and predicate repetition as shown in rule No. 77.

Status (B) ^ disease (volume depletion) => disease (diarrhea and gastroenteritis of presumed infectious origin).

In these rules there are two predicate that is the status and disease, where there is repetition of the predicate disease. These relationships using a sample of patients, 6% of them have the status of unmarried and 100% of patients with this status if you suffer from volume depletion also suffered from diarrhea and gastroenteritis of presumed infectious origin. In other words, if the patient has an unmarried status and suffer from volume depletion, then 6% possibility of the patient to suffer from diarrhea and gastroenteritis of presumed infectious origin.

From Table I and Figure 1 can be seen that the more attributes of patients are used, the longer time process required. Large number of transactions does not affect the processing time because it depends on the number of itemset that contain elements of the disease.

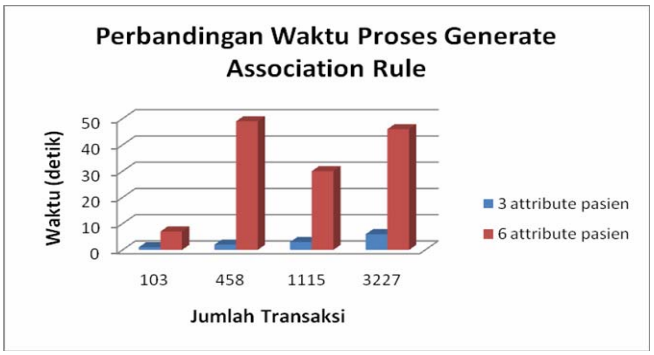

Figure 1. Comparison of generate association rule process

V. CONCLUSION

Mining result is to display the correlation between data (association rules) along with information on support and confidence that can be analyzed. Information provides additional considerations for the user in decision making further. Applications can process data recapitulation of the patients at Dr. Soetomo to find frequent itemset that meets the minimum support and generate Dimension Hybrid Association Rules. If the specified minimum support smaller and patient attributes that are used more and more the more itemset generated and the longer processing time, and vice versa. If the specified minimum confidence that the smaller the more rules are generated, and vice versa.

ACKNOWLEDGMENT

This research was supported by "Penelitian Hibah Bersaing" research grants 110/SP2H/PP/DP2M/IV/2009 and 326/SP2H/PP/DP2M/IV/2010 from Directorate General of Indonesian Higher Education.

REFERENCES

[1] Han, Jiawei, Kamber, Micheline, "Data mining : Concept and techniques", Canada: Simon Fraser University, 2001.

[2] Purnama, I Wayan Jatu Wira, "Market basket analysis mining on Sembilan Minimarket with hybrid-dimension association rules method", unpublished literature, Christian University, Surabaya, 2009.

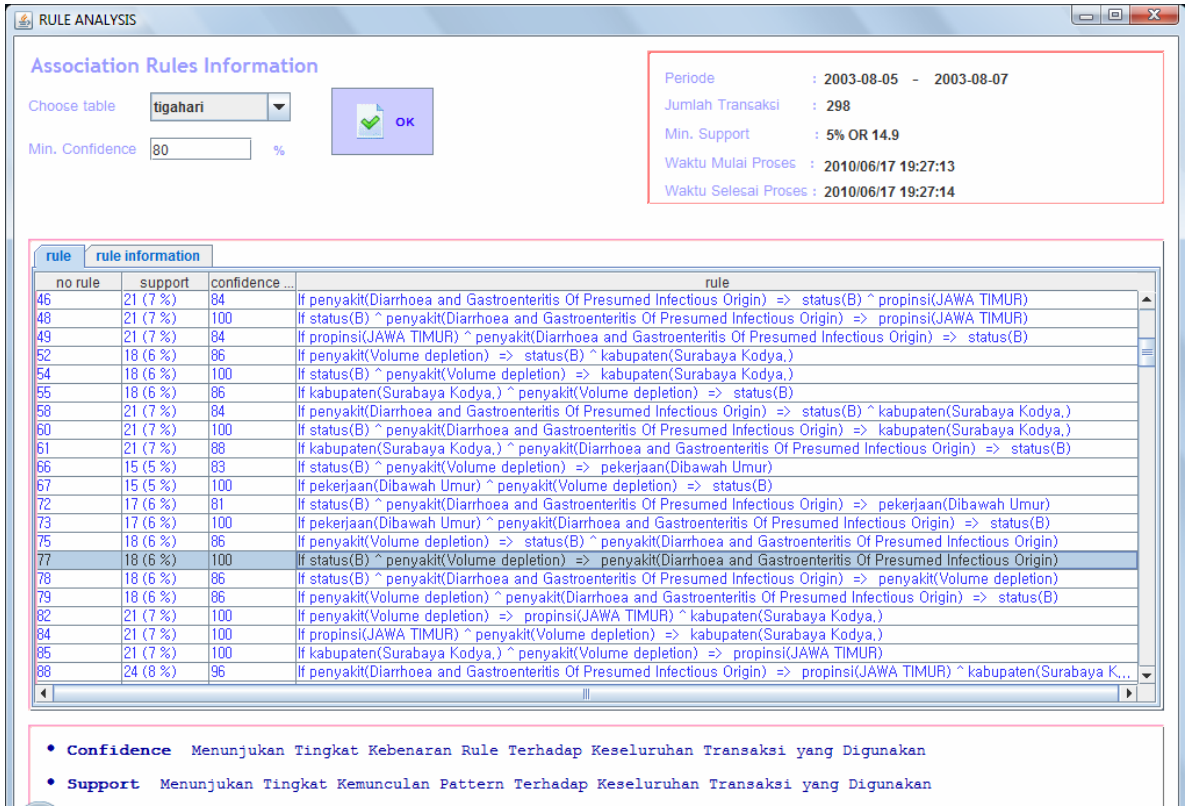

Figure 2. Rule analysis form

TABLE I. GENERATE ASSOCIATION RULES PROCESS EVALUATION

No	Transact-ion Number	Minimum Support	Patients' Attribute	Process Time	Rule Number
1.	103	5%	Gender, provinsi, education	1 detik	14
			Gender, status, provinsi, district, education, occupation	7 detik	94
2.	458	5%	Gender, provinsi, education	2 detik	26
			Gender, status, provinsi, district, education, occupation	49 detik	666
3.	1115	5%	Gender, provinsi, education	3 detik	34
			Gender, status, provinsi, district, education, occupation	30 detik	410
4.	3227	5%	Gender, provinsi, education	6 detik	32
			Gender, status, provinsi, district, education, occupation	46 detik	528