Medicine Product Recommendation System using Apriori Algorithm and Fp-Growth Algorithm

[1]Sukenda, [2]Ari Purno Wahyu, [3]Sunjana

Abstract---As one of the pharmacy products is a company that sells Medicine and website-based health products, this pharmacy is able to produce sales data every day which continues to grow and is not considered to be able to maximize the utilization of the data. Sales data is only stored without further analysis, so an application is needed to analyze the market basket of transaction data on medicine product sales using data mining as a data analysis technique that can help pharmacies, so pharmacy owners obtain knowledge in the form of sales patterns in certain month period. Data mining applications are built using linear sequential processes with the PHP programming language and MySQL database. The algorithm used as the main process of market basket analysis to find out the stock of goods is a priori algorithm by using minimum support, minimum confidence, and month period of the sales transaction to find the association rules. Data mining applications produce association rules between items purchased items, namely consumers make transactions to purchase medicine products simultaneously with min support of 50% and confidence of 80%. Thus, if there is a consumer buying medicine product, then there is a possibility that 80% of consumers buy similar products.

Keywords---Apriori, Medicine,FP-Growth

I. INTRODUCTION

The need for information in the current era of globalization is increasingly important with regard to the purpose of information that is to produce something more useful and meaningful for making a decision quickly, precisely and accurately. The need for fast, precise and accurate information presentation is increasingly needed by various parties and developing company companies. Examples of information systems used medicine product are decision support systems for the sale of goods. Goods sales decision support system is a system that is used to produce useful information from the number of transaction data and stock of sales proceeds in which there is hidden information, to be processed into a decision and a knowledge. Every day, sales transaction data will always increase. The more stored transaction data causes large data storage and only becomes an archive. This data is then processed so that the sales report is produced, so that the inventory (stock of goods) and demand can be fulfilled in a balanced manner so that there is no depreciation of the value of goods.

II. BASIC THEORY

In this study data association analysis with a priori algorithm for stock decision support systems from the sale of goods using php programming language and inputting data with mysql can be a solution for medicine product for decision making such as knowing the relationship between goods purchased by consumers so that can balance the value of goods and can improve service to consumers then the thesis is made with the title "analysis of product sales

[1]*Informatics Engineering, Widyatama University, Indonesia*

[2]*Informatics Engineering, Widyatama University, Indonesia*

[3]*Informatics Engineering, Widyatama University, Indonesia*

Sukenda@widyatama.ac.id

data to determine the stock of goods in Indonesian alwahida bidder herbs". Apriori is an algorithm that is very well known for conducting frequent itemset searches using the association rule technique. The Apriori algorithm uses knowledge about frequent itemset that has been previously known, to process further information. In the Apriori algorithm to determine possible candidates by paying attention to minimum support. The two main processes carried out in the Apriori algorithm(Tampubolon, Saragih, & Reza, 2013).

Association rule is one method that aims to find patterns that often arise among many transactions, where each transaction consists of several items, so this method will support the medicine product sales recommendation system through the discovery of patterns between items in transactions that occur in HPAI Products. . Association rules (associative rules) or affinity analysis regarding the study of "what is together" for example can be in the form of a supermarket transaction, for example someone can buy a product to buy baby milk also buy bath soap in this case means baby milk in buy along with bath soap, because initially it came from the study of customer transaction databases to determine a product(Aritonang, 2013)

The association rules want to provide this information in the form of if then or if - then. This rule is calculated from data that is probalistic that association analysis is also known as a data mining process. Especially in one stage in association analysis called high frequency pattern analysis, the importance of associative rule patterns can be known by two parameters, namely support and confident, namely the presentation of the combination item. In database and confident, the strong relationship between items in associative rules that meet the minimum support requirements and confident for minimum confidence(Sibagariang & Auliasari, 2013).

The FP-Growth algorithm uses the FP-Tree data structure. The information stored is an FP-Tree node: Item, Parent Index, Support, and Next (Pointer). When finished creating FP-Tree, we cannot just get the frequent itemset contained in the dataset. An itemset combination can be in several different paths. To get a pattern in FP-Growth, the easier step is to find the direction from the end of a path, then we look for the header for the item at the end, then it is made based on each node containing the item the direction of the node path up. This is certainly faster than the up-down because the direct pointer for each node is a pointer to the parent. Explored paths are only the paths that have the node being searched for. So in the FP-Tree structure there is a link from an item to the paths that have the item, so when it takes a search for patterns for a particular item, just search for those paths(Annisa & Ruldeviyani, 2007).

The phase of FP-growth algorithm consists of two phases namely FP-tree formation phase and FP-growth excavation process phase. In the process of searching itemset data frequency with a priori algorithm, the data retrieval process is done by scanning the database then the results are accumulated with each other item and grouped to get the frequency of itemsets that meet the minimum support. Then the value of items that are less than the minimum support will be trimmed and discarded, so that the remaining items or items that meet the criteria will be arranged in descending order and then the database scanning process is done for the second time(Dharmayanti, 2012).

FP-tree is a tree data structure, which consists of: one header table, one root labeled null, and one set of prefix subtree items as nodes of the root child. Each entry in the header table is frequent itemset, and each record consists of two attributes, namely the item name (item_name) and the node that becomes the head of node-link (Hussain et al., 2019). Whereas each child node consists of item name, count, and node-link information. Where the count indicates the number of transactions represented by the branch containing the node, and the node-link connects the node to the next node in the tree which has the same item name or null if there is no(Putra, et al.).

FP-Growth Algorithm (Frequent Pattern Growth). This algorithm is an extension of the pre-existing a priori algorithm, because every combination of a priori items will scan the database repeatedly causing the amount of time needed to scan the database, especially if the amount of data is large enough, besides that it needs to generate a large candidate for get item combinations from the database. FP-Growth is here to solve the problem, because this algorithm only performs 2 times the database scanning process to determine frequent itemset and also does not require generating candidates as in a priori. The FP-Growth algorithm is the development of a priori algorithm so that the shortcomings of a priori algorithms are fixed in the FP-Growth algorithm by calculating the pattern of data that almost appears (Itemset frequency) and FP-tree data structure by using a data tree structure so that the FP dialgorithm data search -growth faster. FP-Growth uses a different approach from the paradigm that has often been used, namely the a priori paradigm. A priori algorithms scan the database repeatedly to determine the frequent itemset, while the FP-Growth algorithm requires only two database scans(samuel & David, 2012).
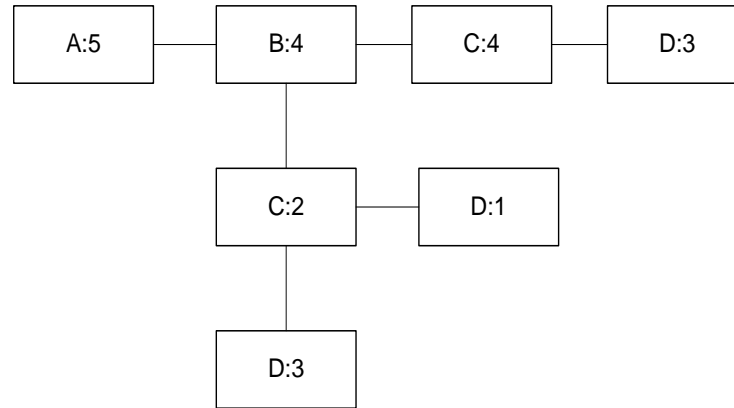
III. RESEARCH METHODS

Figure 1.1:*Struktur pada patren Tree (Tamassia, T, & Roberto, 2008)*

After processing FP-Tree into patterns, it is necessary to process the accumulation of found patterns given that the same pattern can be found in different paths for that

- This study uses data from the sales of PT. HPAI product sales as much as 5000 transaction data in the range of sales years from 2015 to 2016, with the provisions of the minutesup amounting to 1 to 4, and Minconf by 50% to 80%. Furthermore, from the results of the application execution, a rule is found or in this study it is called a sales product recommendation. The rule found in the application varies greatly from the itemset length. The length of these items will determine the number of recommendations given. If the length of itemset reaches 6-itemset, that means there are medicine Products recommended

- *Rapid miner*
  By using rapid miner the author tries to determine the pattern of association pattern rules on the datasheet and stock on the goods sale transaction. The first step in the calculation of the Priori Algorithm to find association rules is to write down the transaction data that occurred

IV. IMPLEMENTASI SISTEM

Used to find 2 itemset items that meet the full combination of support and minimum Analysis of medicine product sales transaction data Determine the pattern of the number of product sales transactions that will be calculated by association techniques. Making tabular data patterns Calculation of transactions from product sales to determine the number or frequency each product sold so that it will determine the calculation of goods product 01 and item 02 also called mutual exclusive. Determination of data 2 itemset and 4 itemset of product An itemet is a set of items that are in product 01, and product 02 is itemset containing product 02, and calculation of 3 itemset is product 01.02, 03

Table 1.1:*Apriori Tests*

Name Product	Support	Confi	Laplace	Gain
Acitral tab	0.25	1.0	1.0	-0.25
Aceton 60ml	0.25	1.0	1.0	-0.25
Absolute	0.25	1.0	1.0	-0.25
Abbocath 20	0.25	1.0	1.0	-0.25
Acitral tab, aceton 60ml	0.25	1.0	1.0	-0.25
Acitral tab, absolute	0.25	1.0	1.0	-0.25
Acitral tab, abbocath 20	0.25	1.0	1.0	-0.25
Aceton 60ml, absolute	0.25	1.0	1.0	-0.25
Aceton 60ml, abbocath 20	0.25	1.0	1.0	-0.25
Absolute, abbocath 20	0.25	1.0	1.0	-0.25

Association Rules
[ACITRAL TAB] --> [ACETON 60ML] (confidence: 1.000)
[ACETON 60ML] --> [ACITRAL TAB] (confidence: 1.000)
[ACITRAL TAB] --> [ABSOLUTE] (confidence: 1.000)
[ABSOLUTE] --> [ACITRAL TAB] (confidence: 1.000)

[ACITRAL TAB] --> [ABBOCATH 20] (confidence: 1.000)
[ABBOCATH 20] --> [ACITRAL TAB] (confidence: 1.000)
[ABBOCATH 18] --> [ACITRAL TAB] (confidence: 1.000)
[ACETON 60ML] --> [ABSOLUTE] (confidence: 1.000)
[ABSOLUTE] --> [ACETON 60ML] (confidence: 1.000)
[ACETON 60ML] --> [ABBOCATH 20] (confidence: 1.000)
[ABBOCATH 20] --> [ACETON 60ML] (confidence: 1.000)
[ABBOCATH 18] --> [ACETON 60ML] (confidence: 1.000)
[ABSOLUTE] --> [ABBOCATH 20] (confidence: 1.000)
[ABBOCATH 20] --> [ABSOLUTE] (confidence: 1.000)
[ABBOCATH 18] --> [ABSOLUTE] (confidence: 1.000)
[ABBOCATH 18] --> [ABBOCATH 20] (confidence: 1.000

within a certain period. Examples of transaction data are as in the following table:

$$(\ Total\ Transaction\ \ ACETON\ 60ML, *\ \ ACITRAL\ TAB\ )/(Transaction\ ACITRAL\ TABHoot)$$

V. CONCLUSION

Recommendation results using a priori algorithm and FP Growth able to be implemented on a system, apriori reading system sales data on the database,apriori able to display sales data and product recommendations health medicine, the number of databases will make a computerized system become long and should always be combined with the FP-Growth algorithm, this algorithm has a process and speed scanning is faster because it has a FP Tree structure in the database so the data scanning process does not need to be done twice. Product recommendation data can be used by store owners or a company to register transaction data, and use the data to consider the data to be sold next, or transaction data is used to create a list of stock data and can also be used as a media promotion of health goods and products.

BIBLIOGRAPHY

[1]   Annisa, & Ruldeviyani, Y. (2007). Pancarian Target Event Rules dengan metode Pattren Growth. Jakarta: Fakultas ilmu Komputer,Universitas Indonesia.
[2]   Aritonang. (2013). Perancangan basis data terdistribusi untuk barang dan perelatan pada balai riset dan standardisasi industri palembang. Program studi teknik informatika Fakultas ilmu komputer Universitas binadarma.
[3]   Dharmayanti, D. (2012). Model Sistem Pendukung Keputusan Dalam Penentuan Matakuliah Pilihan Di Jurusan. Bandung: Teknik Informatika Unikom.
[4]   Hussain, H.I., Kamarudin, F., Thaker, H.M.T. & Salem, M.A. (2019) Artificial Neural Network to Model Managerial Timing Decision: Non-Linear Evidence of Deviation from Target Leverage, *International Journal of Computational Intelligence Systems*, (forthcoming).
[5]   Putra, Andriana, Kadek, Drs.Mardj, MUflikhah, & Lailil. (n.d.). penerapan Metode Association Rule Dengan Algoritma. jurusan Ilmu Komputer Program Teknologi Informasi Univ.Brawijaya.
[6]   Samuel, & David. (2012). penerapan stuktur fp-tree dan algoritma fp-growth dalam optimasi penentuan frequent itemset. Bandung: Program Studi Teknik Informatika, Sekolah Teknik Elektro & informatika Institut Teknologi Bandung.
[7]   Sibagariang, D., & Auliasari, K. (2013). Analisa Pola Data Hasil Pembangunan Kabupaten Malang Menggunakan Metode Association Rule. malang: Universitas Islam Negeri (UIN) Fakultas Sains dan Teknologi,.
[8]   Tamassia, T, a. G., & Roberto, G. (2008). Teaching Data Structure Design Patterns. Baltimore: Dept. of Comp. Sci. Dept. of Comp. Sci. Brown Univ. Johns Hopkins Univ.
[9]   Tampubolon, K., Saragih, H., & Reza, B. (2013). Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan. Informasi dan Teknologi Ilmiah (INTI), 1(ISSN : 2339-210X),
[10]   Babu, N.V.N., Murali, G., Bhati, S.M. Casson fluid performance on natural convective dissipative couette flow past an infinite vertically inclined plate filled in porous medium with heat transfer, mhd and hall current effects(2018) International Journal of Pharmaceutical Research, 10 (4), pp. 809-819. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062416271&partnerID=40&md5=fca24568447156b565c741990ab45592
[11]   Paras virani , rajanit sojitra, hasumati raj, vineet jain (2014) a review on irbesartan co administered with atorvastatin for the treatment of cardiac risk. Journal of Critical Reviews, 1 (1), 25-28.
[12]   Patel DM, Jani RH, Patel CN. "Ufasomes: A Vesicular Drug Delivery." Systematic Reviews in Pharmacy 2.2 (2011), 72-78. Print. doi:10.4103/0975-8453.86290