A Decision Support System for "De Joglo" Restaurant Using Frequent Pattern Tree Data Mining

Andreas Handojo [1], Gregorius Satia Budhi [2], Nadia Andreani Dwiyono [3]

Department of Informatics Engineering, Faculty of Industrial Technology, Petra Christian University
Siwalankerto 121-131, Surabaya 60236, Indonesia
+62-31-2983455

[1] handojo@petra.ac.id, [2] greg@petra.ac.id

1. Introduction

Now days, the technology developments in Indonesia is increasing fast. There's so many companies that use computer technology to obtain information, this is including on the restaurants that use computer technology to recording sales transactions and data storage system. For a small restaurant, such technology may be sufficient for managers to making decisions. But, unlike on the quite large restaurant, which has more variety menu and many different models of customers. This will make managers will have more difficulty to decide anything because the data that must be consider and processed is quite a lot.

So a system that can assist managers on making decisions that needed. Therefore in this research, will be build a computer based system that process sales data using Frequent Pattern tree data mining to create a decision support system for the restaurant manager. For example, if customers order food A, usually customers will order food B. So, the restaurant can increase sales of food A and food B by made it a food package that contains both of the food. But, even though the system can issue a correlation between sales order, expertise from the manager is still needed, for example to predict whether the decision is appropriate to be applied or not.

2. Data Mining

Data mining is a process to discover interesting knowledge from large amounts of data that stored in the database [1]. Technically this analysis in general is dedicated to be able to work on large amount of data, with the goal of data mining that produce decisions and conclusions that guaranteed for accuracy. Data mining is also part of the process of Knowledge Discovery in Databases (KDD). KDD process consists of the following steps (Figure 1):

- Data cleaning, remove noise and inconsistent data.
- Data integration, integrating multiple data sources that can be combined.
- Data selection, selection of data relevant to the analysis task is retrieved from the database.
- Data transformation, the process where's data is transformed or consolidated into a format suitable for processing in data mining.
- Data mining, an essential process in which the method was applied to extract data patterns.
- Pattern evaluation, the process to identify interesting patterns to be represented into the knowledge-based.
- Knowledge presentation, process visualization and knowledge representation techniques are used to provide useful knowledge to the user.

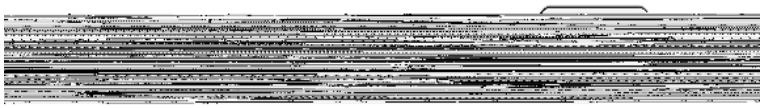

Figure 1 : The Process of Knowledge Discovery in Database [1]

3. Market Basket Analysis

Market Basket Analysis refers to the methodology of the study of consumer habits by finding associations between some of the products offered. For example: finding that the item “A” is usually purchased by the consumer with another item “B” at a specific time. This technique is also called association rule analysis.

The purpose of the market basket is to know what products are purchased simultaneously. Analysis of transaction data to generate a product purchase patterns that often occur on sales. Such information can help sellers to increase sales rate [4]. Suppose in a restaurant, to answer the question "What are the foods that may be ordered at the same visit by customer?" To answer this question market basket analysis can process on the sales transaction data at these restaurants so that the result can be applied to a variety of sales planning in the restaurant, for example in determining the food package [4].

4. Frequent Pattern Tree (FP Tree) [2][3]

FP-Tree is an efficient algorithm for finding patterns in database transactions. FP-Tree is formed by a root (root) are labeled “null”, that’s a set of sub-tree which consists of certain items and a frequent header table. Each node in the FP-tree contains three important information, (1) label item that inform the type of items that represented the node, (2) count that represents the number of transactions that passes through the node, and (3) link that connects node and label item between node-link.

For more details, we can look at this example, we see the transaction database in Table 1 with minimum support = 3.

Table 1 : Data Transaction

Transaction ID	Item Bought	Frequent Item
100	f, a, c, d, g, i, m, p	f, c, a, m, p
200	a, b, c, f, l, m, o	f, c, a, b, m
300	b, f, h, j, o	f, b
400	b, c, k, s, p,	c, b, p
500	a, f, c, e, l, p, m, n	f, c, a, m, p

1. Scan the database to determine the total count of each item. If there is the same item, it can be combined with the same number of transaction. In accordance with a pre-determined minimum support, then get the frequency of occurrence of each item as follows: (f: 4), (c: 4), (a: 3), (b: 3), (m: 3), (p: 3) sorted from the largest number

6. TESTING

For testing, this research will process one year period restaurant sales data and the result can be seen on Table 4.

Table 4: Testing Results for Process Time

No	Transaction Amount	Minimum Support (Frequent Item Sets)	Amount of Rules	Process Time
1	66	10 % (7)	362	1 m 44 s
		15 % (10)	174	53 s
2	178	10 % (18)	486	3 m 30 s
		15 % (27)	58	41 s
3	271	10 % (28)	366	3 m 59 s
		15 % (41)	44	1 m 2 s
4	827	10 % (83)	146	9 m 38 s
		15 % (125)	26	3 m 58 s
5	4221	10 % (423)	66	28 m 13 s
		15 % (634)	31	9 m 12 s
6	8454	10 % (846)	46	7 h 43 m 21 s
		15 % (1269)	14	2 h 58 m 48 s

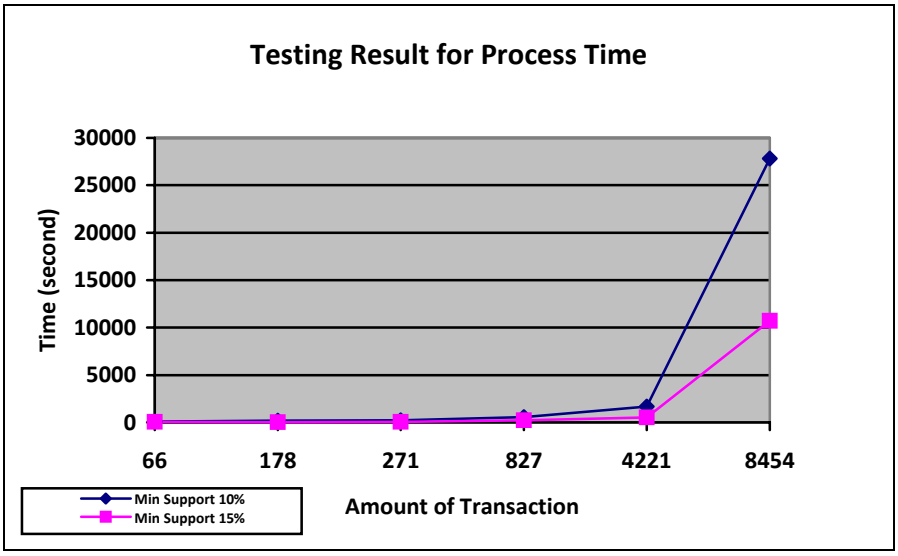

Figure 9. Results Comparison for Process Time

From Figure 9, we can see that the more amounts of transaction data, the more time is required to run the mining process. Meanwhile, the smaller the minimum support, then the more time is required in carrying out the process.

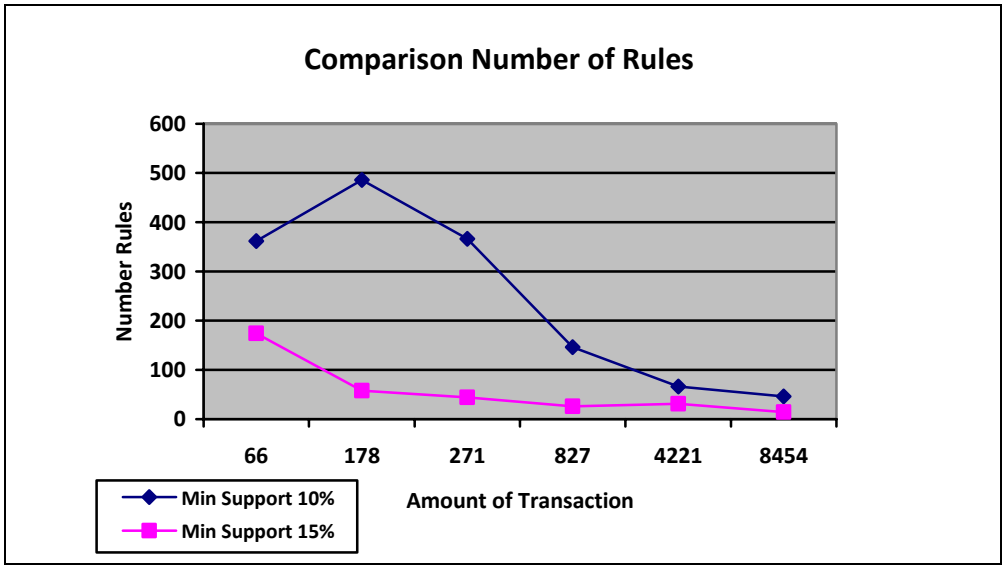

Figure 10 : Results Comparison of Number of Generated Rule

From Figure 10, we can see that number of rules that generated not influenced by number of the transaction. The greater the minimum support, the fewer number of rules generated.

7. Conclusion

From the testing of this application, we have conclusion as follows:

- The smaller the minimum support that user define, then the more frequent item sets generated so the more time is required in carrying out the process.
- Number of Rules that generated not influenced by number of the transaction, but by number of frequent item sets that generated

References

[1] Han, Jiawei., Kamber, Micheline. (2001). Data mining concepts and techniques. Morgan Kaufmann.

[2] Han, J., Pei, J., Yin, Yiwen. (2004). Mining frequent patterns without candidate generation. Simon Fraser University. Netherlands: Kluwer Academic.

[3] He, Z., Xu, X., Deng, S. (2004). A FP-Tree based approach for mining all strongly correlated pairs without candidate generation. USA: World Scientific and Engineering Academy and Society.

[4] Olson, David dan Yong Shi. (2007). Introduction to bussines data mining. New York: McGraw-Hill.