Frequent pattern growth algorithm for maximizing display items

Asyahri Hadi Nasyuha[1], Jalius Jama[2], Rijal Abdullah[3], Yohanni Syahra[4], Zulfi Azhar[5], Juniar Hutagalung[6], Buyung Solihin Hasugian[7]
[1,4,6]Information System, STMIK Triguna Dharma, Medan, Indonesia
[2,3]Faculty of Engineering, Universitas Negeri Padang, Indonesia
[5]Information System, STMIK Royal Kisaran, Indonesia
[7]Information System, Universitas Dharmawangsa, Medan, Indonesia

Article Info

Article history:

Received Mar 29, 2020
Revised Aug 9, 2020
Accepted Aug 29, 2020

Keywords:

Data mining
Display items
Frequent pattern growth

ABSTRACT

Products are goods that are available and provided in stores for sale. Products provided in stores must be arranged properly to order to attract the attention of consumers to buy. Products arranged in a store will depend on the type of store. The product arrangement at a retail store will be different from the product arrangement at a clothing store. Store display will reflect a picture that is in the store so consumers know the types of products sold by product arrangement. An attractive arrangement will stimulate the desire of consumers to buy. In data mining there are several types of methods by use including prediction, association, classification and estimation. In the prediction method there are several techniques including the frequent pattern growth (FP-growth) method. FP-growth algorithm is the development of the apriori algorithm. So, the shortcomings of the apriori algorithm are corrected by the FP-growth algorithm. FP-growth is one alternative algorithm that can be used to determine the set of data that most often appears (frequent itemset) in a data set. Results of research on the application of the FP-growth algorithm to maximizing the display of goods. It is hoped that this research can be used to adjust the product layout according to the level of frequency the product is sought by the customer so that the customer has no difficulty finding the product they want.

Corresponding Author:

Asyahri Hadi Nasyuha
Information System
STMIK Triguna Dharma
A. H Nasution St. No. 73 Medan Johor, Medan, Indonesia
Email: asyahrihadi@gmail.com

1. INTRODUCTION

Products are goods that are available and provided in stores for sale [1]. Products provided in stores must be arranged properly to attract the attention of consumers to buy. Products arranged in a store will depend on the type of store [2]. The arrangement of the product is one thing that is not less important, because this is the first impression of the visitor of the store, therefore merchandise displayed in the storeroom must be arranged so that it looks neat, harmonious and attractive to everyone, especially prospective buyers, for the arrangement of goods needed special expertise, the arrangement of goods should be changed at any time so it is not boring and adapted to the situation. Data mining is mining or discovering new information by looking for certain patterns or rules from a large amount of data that is expected to overcome these conditions [3]. Data mining is a branch of science from artificial intelligence [4]. In data mining there are several types of methods by the use including prediction, association, classification and estimation [5]. In the prediction method there are several techniques including the frequent pattern growth (FP-growth) method. FP-growth algorithm is the development of the apriori algorithm [6]. So, the shortcomings of the apriori algorithm are corrected by the FP-growth algorithm [7]. FP-growth is one alternative algorithm that can be used to determine the set of data that most often appears (frequent itemset) in a data set [8].

2. RESEARCH METHODS

2.1. Data mining

The term data mining has several views, such as knowledge discovery or pattern recognition [9]. Both of these terms have their respective accuracy. The term knowledge discovery is appropriate because the main purpose of data mining is to get knowledge that is still hidden in chunks of data [10]. The term pattern recognition is also appropriate for use because the knowledge to be extracted does indeed take the form of patterns that may also still need to be extracted from inside the chunk of data being faced.

2.2. Phase of data mining

Knowledge discovery in database (KDD) is the process of determining information that serves to determine the patterns contained in data [11]. This information is contained in a large database that was previously unknown and potentially useful. Data mining is one step in a series of KDD iterative processes [12]. The stages of the KDD process consist of [13-15]: data selection, pre-processing and cleaning data, transformation, data mining, and interpretation/evaluation. Data mining is one of the KDD series of knowledge. KDD is the process of determining useful information and patterns in data. Data mining is one step of an iterative KDD process. KDD stages process as shown in the following Figure 1.

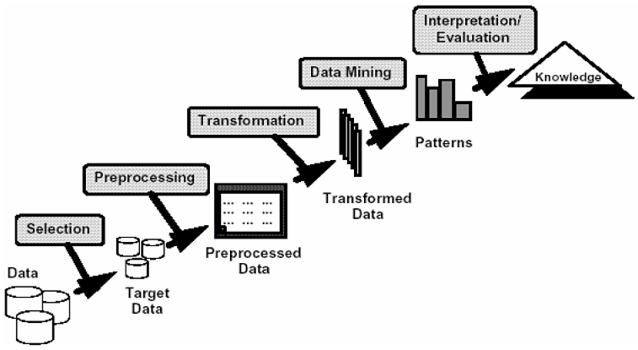

Figure 1. KDD

2.3. Data mining techniques

Some data mining techniques are as follows [16-18]:

- Classification: Assigning a new data record to one of some pre-defined categories (or classes).
- Regression: Predict the value of a given continuous variable based on the value of another variable, assuming a linear or nonlinear dependency model. This technique is widely studied in statistics, the field of artificial neural networks (neural networks).
- Clustering: Partition data set into several sub-sets or groups in such a way so that elements of a particular group have shared property sets together, with a high level of similarity in one group and a low level of similarity between groups. Also called 'unsupervised lerning'.
- Association rules: Detects a collection of attributes that appear together (co-occur) on a frequent frequency, and form a number of rules of the collection.
- Search for sequential patterns (sequence mining): Look for a number of events that generally occur together. If given a set of objects, with each object associated with the time of occurrence, then get a pattern that predicts strong sequential dependencies between different events.

2.4. FP-growth

FP-growth algorithm is the development of apriori algorithm [19]. So, the shortcomings of the apriori algorithm are corrected by the FP-growth algorithm. FP-growth is one alternative algorithm that can be used to determine the set of data that most often appears (frequent itemset) in a data set [20]. Apriori algorithm requires generating candidates to get frequent itemset. However, the FP-growth algorithm generates candidates not done because FP-growth uses the concept of tree development in searching for frequent items [21]. This is what causes the FP-growth algorithm to be faster than the apriori algorithm [22].

The characteristic of FP-growth algorithm is the data structure used is a tree called FP-tree [23]. By using FP-tree, the FP-growth algorithm can directly extract frequent Itemset from FP-tree [24]. Extracting frequent itemset using the FP-growth algorithm will be done by generating a data tree structure or called FP-tree. After the FP-tree development stage of a set of transaction data, the FP-growth algorithm will be applied to look for significant frequent itemset [25].

The FP-growth algorithm is divided into three main steps [26]: 1) Conditional pattern base; 2) Conditional FP-tree; 3) Frequent itemset. The form of the FP-growth algorithm is as follows:

Input: FP-tree tree
Output: Rt A complete set of frequent patterns
Method: FP-growth (tree, null)
Procedure: FP-growth (tree, α)
{
01: If Tree contains single path P;
02: Then for each combination (notated β) of the nodes in the do path
03: Generate the Build pattern β α with support from nodes in the do β path
04: Else for each a1 in the header of the do tree
{
05: generate pattern
06: wake up β = a1 α with support = a1 support
07: If tree β

3. ANALYSIS AND RESULT

3.1. Problem analysis

Problem analysis is a process that involves a survey of the current system and analysis of user needs. After conducting interviews and reviewing documents. The next stage is to describe the system that will be designed in the form of flow, as well as unified modeling language (UML). From the results of this description can be seen what is needed for the development of the system so that the system is designed to run perfectly as desired.

3.2. System algorithm

FP-growth is a frequent itemset search algorithm that is obtained from the FP-tree by exploring the tree from bottom to top. FP-growth is the development of apriori algorithm. This algorithm determines the frequent itemset that ends in a particular suffix by using the devide and conquer method to break the problem into smaller subproblems.

In the formation of FP-growth, namely through the following algorithm:

- Formation of FP-tree.
  a. Determine transaction data.
  b. Count the amount per item.
  c. Determine items that meet the minimum support value≥ 20%.
  d. Determine transaction data that contains minimum support.
- Look for frequent itemset from FP-tree.
- Determine the association rules of the minimum support value and the expected value of confidence.

3.3. FP-tree formation

FP-tree is a data storage structure that is utilized. FP-tree is built by mapping each transaction data into every specific path in the FP-tree, because in every mapped transaction, there may be transactions that have the same item, so the paths are possible to overwrite each other [8]. At this stage several steps will be carried out as follows:

- Pre-Processing the transaction database. To get test results, several transaction items will be tested and will be used as a transaction database. Transaction data are listed in Table 1.
- Calculating the amount per item, in a number of the data Table 1 will be scanned, so that it is known the amount per item of all transactions. Here are the quantities per item:

$$Support\ \mathrm{A} = \frac{\text{Number of Transaction A}}{\text{Total Transactioni}} = \frac{5}{20} = 0.25$$

From Table 1, the frequency and support values of each item are ranked from the highest based on the formula. The following data is attached in Table 2. Based on the support count value of 20%, the items used are items that have a frequency of ≥ 20%, which can be seen in Table 3. After finding the minimum support, the transaction data is moved or arranged to meet the minimum support. The data is sorted based on the highest frequency value as shown in Table 4.

Table 1. Data that consists of transactions in each preparation

No	Transactions
1	Silverqueen, Chitato, Pucuk Harum
2	Gulaku, Bimoli, Pop Mie Kari
3	Sari Roti, Oreo Vanila, Aqua
4	Citra Hand And Body Lotion, Eskulin Cologne Gel, Baby Powder
5	Daia Det Putih, Soklin Liquit, Life Buoy Sabun Mandi
6	Silverqueen, Chitato, Pucuk Harum
7	Pop Mie Kari, Aqua
8	Citra Hand And Body Lotion, Baby Powder, Daia Det Putih
9	Silverqueen, Sari Roti, Oreo Vanila
10	Gulaku, Bimoli, Pop Mie Kari
11	Aqua, Chitato, Sari Roti
12	Citra Hand And Body Lotion, Eskulin Cologne Gel, Baby Powder
13	Daia Det Putih, Soklin Liquit, Life Buoy Sabun Mandi
14	Silverqueen, Chitato, Pucuk Harum
15	Chitato, Pucuk Harum, Sari Roti
16	Gulaku, Daia Det Putih, Soklin Liquit
17	Life Buoy Sabun Mandi, Eskulin Cologne Gel, Pucuk Harum
18	Aqua, Sari Roti, Chitato
19	Silverqueen, Oreo Vanila, Pucuk Harum
20	Pop Mie Kari, Citra Hand And Body Lotion, Life Buoy Sabun Mandi

Table 2. The frequency and support of each item is ranked from the highest

No	Items	Frequency	Support	Support 100 %
1	Silverqueen	5	0.25	25%
2	Chitato	6	0.30	30%
3	Pucuk Harum	5	0.25	25%
4	Gulaku	3	0.15	15%
5	Bimoli	2	0.10	10%
6	Pop Mie Kari	4	0.20	20%
7	Sari Roti	5	0.25	25%
8	Oreo Vanila	3	0.15	15%
9	Aqua	3	0.15	15%
10	Citra Hand And Body Lotion	4	0.20	20%
11	Eskulin Cologne Gel	3	0.15	15%
12	Baby Powder	3	0.15	15%
13	Daia Det Putih	4	0.20	20%
14	Soklin Liquit	2	0.10	10%
15	Life Buoy Sabun Mandi	3	0.15	15%

Table 3. Items meet the minimum support ≥ 20%

No	Item	Frequency	Support	Support 100%
1	Chitato	6	0.30	30%
2	Silverqueen	5	0.25	25%
3	Sari Roti	5	0.25	25%
4	Pucuk Harum	5	0.25	25%
5	Citra Hand And Body Lotion	4	0.20	20%
6	Daia Det Putih	4	0.20	20%
7	Pop Mie Kari	4	0.20	20%

Table 4. Transaction data containing minimum support

No	Transactions	No	Transactions
1	Silverqueen, Chitato, Pucuk Harum	11	Chitato, Sari Roti
2	Pop Mie Kari	12	Citra Hand And Body Lotion
3	Sari Roti	13	Daia Det Putih
4	Citra Hand And Body Lotion	14	Silverqueen, Chitato, Pucuk Harum
5	Daia Det Putih	15	Chitato, Pucuk Harum, Sari Roti
6	Silverqueen, Chitato, Pucuk Harum	16	Daia Det Putih
7	Pop Mie Kari	17	Pucuk Harum
8	Citra Hand And Body Lotion, Daia Det Putih	18	Sari Roti, Chitato
9	Silverqueen, Sari Roti	19	Silverqueen, Pucuk Harum
10	Pop Mie Kari	20	Pop Mie Kari, Citra Hand And Body Lotion

3.4. Look for the frequent itemset from FP-tree

After forming the FP-tree, then look for frequent itemset, which is summarized in the Table 5. Based on the 11 frequent itemsets that have been formed above, all will be counted in the next process because they meet the frequent itemsets requirement in producing an association rule that is a minimum of 2 items if opening category, A will open category B, then there are 9 subsets that are eligible to be calculated confidence level his. After forming the FP-tree, then we are looking for the frequent itemset, which is summarized in the Table 6. After obtaining subsets that meet the requirements, then the confidence value is calculated based on a predetermined minimum confidence value of ≥ 20% to measure the validity of the association rules, as in the Table 7.

Table 5. Frequent itemset results

No	Item	Frequent Item Set
1	Pucuk	(Pucuk), (Pucuk, Chitato: 3), (Pucuk, Silverqueen: 5), (Pucuk, Chitato, Silverqueen: 5)
2	Chitato	(Chitato), (Chitato, Silverqueen: 5), (Chitato, Sari Roti: 2)
3	Sari Roti	(Sari Roti), (Sari Roti, Silverqueen: 5), (Sari Roti, Chitato: 2), (Sari Roti, Pucuk: 1), (Sari Roti, Pucuk, Chitato: 2)
4	Citra	(Citra), (Citra, Pop Mie Kari: 4)
5	Daia	(Daia). (Daia, Citra: 3)

Table 6. Results of subsets

No	Suffix	Subset
1	Pucuk	(Pucuk), (Pucuk, Chitato: 3), (Pucuk, Silverqueen: 5)
2	Chitato	(Chitato), (Chitato, Silverqueen: 5), (Chitato, Sari Roti: 2)
3	Sari Roti	(Sari Roti), (Sari Roti, Silverqueen: 5), (Sari Roti, Chitato: 2), (Sari Roti, Pucuk: 1)
4	Citra	(Citra), (Citra, Pop Mie Kari: 4)
5	Daia	(Daia). (Daia, Citra: 3)

Table 7. Frequent pattern results

No	Frequent Pattern	Frequent
1	Pucuk, Chitato	3
2	Pucuk, Silverqueen	5
3	Chitato, Silverqueen	5
4	Chitato, Sari Roti	2
5	Sari Roti, Silverqueen	5
6	Sari Roti, Chitato	2
7	Sari Roti, Pucuk	1
8	Citra, Pop Mie Kari	4
9	Daia, Citra	3

At this stage it is used to determine the value of support and confidence for each itemset using the formula:

$$\textit{Support}\ (A, B) = P\ (A \cap B) = \frac{\text{Number of Transactions containing } A \cap B}{\text{Total Transactions}}$$

$$\textit{Support}\ (\text{Pucuk, Chitato}) = \frac{3}{20} \text{ x } 100 = 15\%$$

$$\textit{Support}\ (\text{Pucuk, Silverqueen}) = \frac{5}{20} \text{ x } 100 = 25\%$$

$$\textit{Support}\ (\text{Chitato, Silverqueen}) = \frac{5}{20} \text{ x } 100 = 25\%$$

$$\textit{Support}\ (\text{Chitato, Sari Roti}) = \frac{2}{20} \text{ x } 100 = 10\%$$

$$\textit{Support}\ (\text{Sari Roti, Silverqueen}) = \frac{5}{20} \text{ x } 100 = 25\%$$

$$\textit{Support}\ (\text{Sari Roti, Chitato}) = \frac{2}{20} \text{ x } 100 = 10\%$$

$$\textit{Support}\ (\text{Sari Roti, Pucuk}) = \frac{1}{20} \text{ x } 100 = 5\%$$

$$\textit{Support}\ (\text{Citra, Pop Mie Kari}) = \frac{4}{20} \text{ x } 100 = 20\%$$

$$\textit{Support}\ (\text{Daia, Citra}) = \frac{3}{20} \text{ x } 100 = 15\%$$

$$Confidence\ (A \rightarrow B) = \frac{\text{Number of Transaction containing } A \cap B}{\text{Total Transaction } A}$$

$$Confidence\ (\text{Pucuk, Chitato}) = \frac{3}{5} \times 100 = 60\%$$

$$Confidence\ (\text{Pucuk, Silverqueen}) = \frac{5}{5} \times 100 = 100\%$$

$$Confidence\ (\text{Chitato, Silverqueen}) = \frac{5}{6} \times 100 = 83\%$$

$$Confidence\ (\text{Chitato, Sari Roti}) = \frac{2}{6} \times 100 = 33\%$$

$$Confidence\ (\text{Sari Roti, Silverqueen}) = \frac{5}{5} \times 100 = 100\%$$

$$Confidence\ (\text{Sari Roti, Chitato}) = \frac{2}{5} \times 100 = 40\%$$

$$Confidence\ (\text{Sari Roti, Pucuk}) = \frac{1}{5} \times 100 = 20\%$$

$$Confidence\ (\text{Citra, Pop Mie Kari}) = \frac{4}{4} \times 100 = 100\%$$

$$Confidence\ (\text{Daia, Citra}) = \frac{3}{4} \times 100 = 75\%$$

3.5. Formation of association rules

Association rule is a method that aims to find patterns that often appear between many transactions, where each transaction consists of several items so that this method will contain a recommendation system through finding patterns between items in frequent transactions. Only combinations greater than or equal to the minimum confidence will be used to form a rule, the rule can be seen in Table 8.

Table 8. Strong association rule

No	Rule	Support %	Confidence %
1	Pucuk, Chitato	15 %	60 %
2	Pucuk, Silverqueen	25 %	100 %
3	Chitato, Silverqueen	25 %	83 %
4	Chitato, Sari Roti	10 %	33 %
5	Sari Roti, Silverqueen	25 %	100 %
6	Sari Roti, Chitato	10 %	40 %
7	Sari Roti, Pucuk	5 %	20 %
8	Citra, Pop Mie Kari	20 %	100 %
9	Daia, Citra	15 %	75 %

4. CONCLUSION

FP-growth is a method that can process transaction data more quickly and accurately. This method can also be used to analyze sales data by determining the types of products and transactions, designing sales data grouping systems. This method can be used to arrange product appearance in order to attract customers and increase sales.

REFERENCE

[1] T. A. Muritala, “International Journal of Advances in Management and Economics”, *International J. Adv. Manag. Econ.*, vol. 1, no. 5, pp. 116-124, 2012. Available: www.managementjournal.info An Empirical Analysis of Capital Structure on Firms Performance in Nigeria.

[2] A. F. Cooper and R. Thakur, “The group of twenty (G20),” *The Group of Twenty (G20)*, pp. 1-194, 2013.

[3] N. Bhatla and K. Jyoti, “An analysis of heart disease prediction using different data mining techniques,” *Int. J. Eng. Res. Tecnol.*, vol. 1, no. 8, pp. 1-4, 2012.

[4] M. Dey and S. S. Rautaray, “Study and Analysis of Data mining Algorithms for Healthcare Decision Support System,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 470-477, 2014.

[5] M. Goyal and R. Vohra, "Applications of Data Mining in Higher Education," *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 113–120, 2012.
[6] B. S. Kumar and K.V.Rukmani, "Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms," *Int. J. Adv. Netw. Appl.*, vol. 1, no. 6, pp. 400-404, 2010.
[7] K. Dharmarajan and M. A. Dorairangaswamy, "Analysis of FP-growth and Apriori algorithms on pattern discovery from weblog data," *2016 IEEE Int. Conf. Adv. Comput. Appl. ICACA 2016*, pp. 170-174, 2016
[8] A. Ikhwan, "A Novelty of Data Mining for FP-Growth Algorithm," *Int. J. Civ. Eng. Technol.*, vol. 9, no. 7, pp. 1660-1669, 2018.
[9] M. S. Chen, J. Han, and P. S. Yu, "Data mining: An overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866-883, 1996.
[10] A. Holzinger and I. Jurisica, "Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8401, pp. 1-18, 2014.
[11] Zulham and Asyahri Hadi Nasyuha, "Application of Data Mining for Wahana Grouping (In Bahasa:Penerapan Data Mining Untuk Pengelompokan Wahana)," *Sains dan Komputer (SAINTIKOM)*, vol. 17, no. 1, pp. 92–104, 2018.
[12] C. Bohm and F. Krebs, "The k-Nearest Neighbour Join: Turbo Charging the KDD Process," *Knowl. Inf. Syst.*, vol. 6, no. 6, pp. 728-749, 2004.
[13] G. W. And, G. J. Williams, and Z. Huang, "A Case Study in Knowledge Acquisition for Insurance Risk Assessment using a KDD Methodology," *Presented at PKAW96, the Pacific Rim Knowledge Acquisition Workshop,* Sydney, Australia, 1996, pp. 117-129.
[14] O. Maimon and L. Rokach, "Data Mining and Knowledge Discovery Handbook," *Data Min. Knowl. Discov. Handb.*, 2010, pp. 1-15.
[15] H. M. Chung and P. Gray, "Special section: Data mining," *J. Manag. Inf. Syst.*, vol. 16, no. 1, pp. 11-16, 1999.
[16] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Comput. Human Behav.*, vol. 73, pp. 247-256, 2017.
[17] A. Bhardwaj, A. Sharma, and V. K. Shrivastava, "Data Mining Techniques and Their Implementation in Blood Bank Sector – A Review," *Int. J. Eng. Res. Appl.*, vol. 2, no. August, pp. 1303-1309, 2012.
[18] K.Srinivas, B. K. Rani, and A.Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *Int. J. Comput. Sci. Eng.*, vol. 02, no. 02, pp. 250-255, 2010.
[19] D. Hunyadi, "Performance comparison of Apriori and FP-Growth algorithms in generating association rules," *Proc. Eur. Comput. Conf. ECC '11*, 2011, pp. 376-381.
[20] H. Li, Y. Wang, D. Zhang, M. Zhang, and E. Y. Chang, "PFP: Parallel FP-growth for query recommendation," *RecSys'08 Proc. 2008 ACM Conf. Recomm. Syst.*, 2008, pp. 107-114.
[21] A. M. Said, P. D. D. Dominic, and A. B. Abdullah, "A comparative study of FP-growth variations," *Int. J. Comput. Sci. Netw. Secur.*, vol. 9, no. 5, pp. 266-272, 2009.
[22] J. Heaton, "Comparing dataset characteristics that favor the Apriori, Eclat or FP-Growth frequent itemset mining algorithms," *Conf. Proc. - IEEE SOUTHEASTCON*, vol. 2016-July, 2016.
[23] Y. G. Sucahyo and R. P. Gopalan, "CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure," *Conference: FIMI '04, Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations*, Brighton, UK, November 1, vol. 4, 2004, pp. 212-223.
[24] K. Malik, N. Raheja, and P. Garg, "Enchanced FP-Growth Algorithm," *Int. J. Comput. Eng. Manag.*, vol. 12, no. April, pp. 54-56, 2011.
[25] M. Wojciechowski, K. Galecki, and K. Gawronek, "Concurrent Processing of Frequent Itemset Queries Using FP-Growth Algorithm," *Proc. 1st ADBIS Work. Data Min. Knowl. Discov. (ADMKD'05), Tallinn, Est.*, no. 1, 2005.
[26] S. H. Ali, "A novel tool (FP-KC) for handle the three main dimensions reduction and association rule mining," *2012 6th Int. Conf. Sci. Electron. Technol. Inf. Telecommun. SETIT 2012*, 2012, pp. 951-961.