Knowledge Data Discovery (Frequent Pattern Growth): The Association Rules for Evergreen Activities on Computer Monitoring

Fauzan Asrin[1,2], Saide Saide[3,4(✉)], Silvia Ratna[5], and Alex Wenda[6]

[1] Department of Information System, Tanjungpura University, Pontianak, Indonesia
[2] Pasca Sarjana Ilmu Komputer, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
[3] Department of Information System (PRO-Knowledge Research Group), State Islamic University of Sultan Syarif Kasim Riau (UIN SUSKA Riau), Pekanbaru, Indonesia
saide@uin-suska.ac.id
[4] Department of Information Management, National Taiwan University of Science and Technology (NTUST), Taipei, Taiwan
[5] Department of Computer Science, Islamic University of Kalimantan, Banjarmasin, Indonesia
[6] Department of Electrical Engineering, Faculty of Science and Technology, State Islamic University of Sultan Syarif Kasim Riau (UIN SUSKA Riau), Pekanbaru, Indonesia

Abstract. The aim of this research paper is to construct a set of guidelines that can improve the quality and efficiency of knowledge data discovery process by utilizing various types of knowledge domains. In addition, this paper offered the way of how the knowledge domain could be adopted for helping the system developer. The methodologies contain various scenarios of data exploring and the authors used data mining approach. The paper shows evidence of important to adopt data mining methods in the industry sector as well as the advantages and disadvantages. Evergreen human machine interface (HMI) at PT. Chevron Pacific Indonesia (CPI) is kind of activities to maintenance computer equipment. Nowadays, the errors were frequently happened in the accuracy of computer maintenance which has a profound effect on production results. Therefore, this study focuses on the rules of association using the frequent pattern growth algorithm (FP-growth) which is producing knowledge with trust value of 100% and a support value is 95%. The value results of support and confidence are the new approach and knowledge for the management level to decide decisions in the evergreen activities process.

Keywords: Knowledge data in discovery · Association rules · Algorithm FP growth · Evergreen activities · HMI (Human machine interface)

C. Kahraman et al. (Eds.): INFUS 2020, AISC 1197, pp. 807–816, 2021.
https://doi.org/10.1007/978-3-030-51156-2_93

1 Introduction

The process of knowledge discovery in databases (KDD) was traditionally presented as a sequence of operations which applying iteratively, lead from the raw input data to high level, interpretable, and useful knowledge. Knowledge discovery in databases (KDD) is an automatic, exploratory, and modeling of large data repositories [1, 2]. This concept is the organized process of identifying validity, useful, and understandable patterns from large and complex data sets. The major steps in KDD process typically such as selection, preprocessing, transformation, data mining, and interpretation or evaluation [2, 3]. Chevron Pacific Indonesia (CPI) is a company engaged in the exploration of oil and gas [4]. CPI operates in the block area of Rokan Sumatera Riau province. The success of exploration activities is not freelancing work with the role of information technology that supports companies' activities. Information technology (IT) as an instrumental to facilitate the workers in receiving information [3, 4] and monitoring activities of the special oil and gas explorations in the area of Duri city (Riau Province).

Monitoring generally refers to the observation, regulation, control, and reporting of processes, procedures, team work, and persons [6, 7]. Monitoring concerns about employee privacy and activities; therefore, employers must find a balance between monitoring gains and employee activities [8, 9]. One of IT's roles is to carry out evergreen activities. The function of evergreen activities is for HMI personal computer (PC) or HMI treatment. HMI PC is working to facilitate the operator of monitoring directly from control room. HMI PC used in oil exploration field in the area of Rokan block amounted to approximately 180.

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner [10, 11]. The two terms have different concepts but are related to each other [12]. Data mining is part of knowledge discovery in databases process, consisting of stages such as data selection, pre-processing, transformation, data mining, and evaluation of results [13, 14]. Data mining brings together techniques from machine learning, pattern recognition, statistics, databases, and visualizations to handle information retrieval problems from large databases [13, 15]. Data mining is not stand-alone technology. It is an important step in the KDD mainly relating to the extraction and calculation of the patterns of the data being sliced [5, 6].

In this paper, the authors focus how to demonstrate the improvement of the entire KDD process by using knowledge domains in several phases. Problems that occur at Chevron Pacific Indonesia is the difficulty of CPI to explore the rules of knowledge and see what tendencies exist in evergreen activities carried out on PC HMI. Thus, in this study, we tested whether the Association Rule Algorithm FP-Growth method can be used to solve problems in evergreen activity data. The results of this study are expected to help leaders at PT. Chevron Pacific Indonesia about how to develop computer maintenance policies more appropriate. This study is structured as follows: introduction, literature review, methods, results and analysis, and conclusion.

2 Literature Review

Knowledge discovery in databases refers to the use of methodologies from machine learning, pattern recognition, statistics, and other fields to extract knowledge from large data, where the knowledge is not explicitly available as part of the database structure which are includes concepts, concept interrelations, classifications, decision rules, and other patterns of interest [2, 7–9]. The knowledge discovery procedure is consisting of nine steps (see Fig. 1). Thus, it is required to understand the process and the different needs and possibilities in each step.

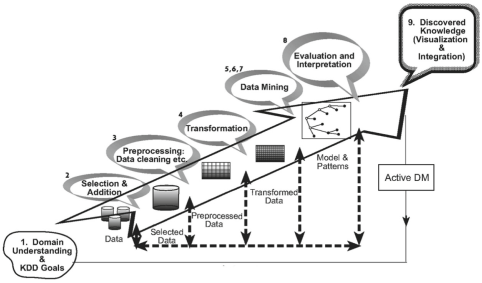

Fig. 1. The procedure of knowledge discovery in databases [1]

(1) Developing understanding. This is the initial preparatory step to sets the scene for understanding what should be done [16]. (2) Selecting and creating a data set. This process is very important because the data mining learns and discovers from the available data. (3) Pre-processing and cleansing. At this step, data reliability is enhanced. (4) Data transformation. Methods here include dimension reduction, such as feature selection, and extraction, and attribute transformation [14]. (5) Choosing the appropriate data mining task. Prediction is often referred to as supervised data mining while descriptive includes the unsupervised and visualization [17]. (6) Choosing the data mining algorithm. This stage includes selecting the specific method for searching patterns and multiple inducers [3, 16]. (7) Employing the algorithm. In this phase needs to employ the algorithm several times until a satisfied result is obtained. (8) Evaluation. In this step, evaluating and interpreting the mined patterns (rules, reliability etc.). (9) Using the discovered knowledge. The knowledge becomes active in the sense that we may make changes to the system and measure the effects [5, 18].

3 Methods

This research methodology describes step by step the process of this research. The methodology was conducted with systematically which can be used as a guideline about how to conduct the research. (1) Problem analysis. A problem analysis investigates a situation/problem in order to allow the authors to understand more fully the problem and to recommend practical solutions for solving it. (2) Determining research objectives. At this stage, the targets are should be achieved, especially to overcome the existing problems. (3) Reviewing references. The search strategy is aimed at finding references from the best scientific papers such as SSCI/SCI, Elsevier Sciencedirect, IEEE explore and other related links [7, 8]. (4) Collecting evergreen activity data. The more data is collected, the better rules are being able to solve the problem. Then, to test whether the data result is able to produce the output as planned at the time of inserting the evergreen data that has been transformed at the first stage. (5) Implement FP growth algorithm. After that, based on the results, it will provide and generate the rules for leadership and management level of the company about how to create a strategy for the evergreen activities further. FP-Growth can be divided into three main parts such as conditional pattern base, conditional FP-tree, and frequent itemset [11, 19, 20]. (6) The test results stage. At this stage, the authors start to develop the testing process.

4 Results and Analysis

At this step, there will be several calculations, testing, processing, and analyzing data based on existing data. The data was collected from the department of IT process control network (PCN) of Chevron Pacific Indonesia in 2018 (chapter: Duri city). The data is expected can help during testing and implementation of the association rule with FP growth algorithm. Also, can provide the output (knowledge or information) and show the trend of evergreen activities in Chevron Pacific Indonesia. Then, the data will be grouped, compiled, and transformed into tabular data which will be processed by the software of Rapidminer 9.0. In the following is the process:

4.1 Defining Variables

In determining the association rules, the evergreen data and activities variables were adopted. The variables such as backup variable, update anti virus variable, inventory variable, user variable, technical control variable, event-log variable, and restore test variable.

4.2 Data Transformation

At this stage, the transformed data will be grouped by the variables of the evergreen activity in every existing computing ID. The variables include backup activities, updating anti-virus, inventory, event log, OS user, technical control, and restore test. Grouping the data is the result of the analysis of variables. All of those variables presented at the Table 1 below.

Table 1. Tabular data

Computing ID	Evergreen activities
DRIA13NOE01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
DRIGBSPI04	Backup, Eventlog, Inventory, OS_User, Restore, AV_Update
DRIA13RSOH01	Inventory
DRIA13SOE01	Backup, Eventlog, Inventory, OS_User, Restore, AV_Update
DRIAWT4SWH01	Backup, Eventlog, OS_User, Restore, AV_Update
DRISLNMWTH01	Backup, Eventlog, Inventory, OS_User, Restore, AV_Update
BKSAMPGSH01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
DMIHCTPI01	Backup, Eventlog, OS_User, Restore, AV_Update
DRIHCTPIPEH01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
DRIHCTMONH03	Eventlog, Inventory, OS_User, TC, AV_Update
DMIHCTMTRH05	Eventlog, Inventory, OS_User, Restore, AV_Update
DRIHOSS6E01	Eventlog, Inventory, OS_User, AV_Update
LBOLBOGPH01	Eventlog, OS_User, AV_Update
BKOBNRGSH01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
PGRPGRGPI01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
PGRPGRGSE01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
PGRPGRGSH01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update
BKOBLMGSH01	Backup, Eventlog, Inventory, OS_User, Restore, AV_Update
BKOBLMGSE01	Backup, Eventlog, Inventory, OS_User, Restore, AV_Update
PGRPGRGPH01	Backup, Eventlog, Inventory, OS_User, Restore, TC, AV_Update

4.3 The Item Appears

Before designing the FP growth algorithm and developing the pattern of the association rules, it is required to do a calculation of each variable that appears in the computing ID in order to calculate the value of support and confidence of each evergreen activities variable such as Backup (B)(15), Eventlog (E)(19), Inventory (I)(17), OS User (OU)(19), Restore (R)(16), Technical Control (TC)(11), AV update (AV)(19).

4.4 FP Tree Formation

The establishment of FP tree from the transaction data that transformed into tabular data with a transaction amount of 20 evergreen. Figure 2 below presents FP tree formation.

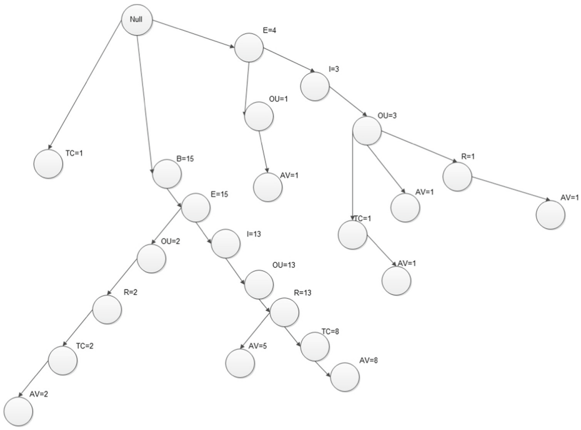

Fig. 2. TID 20

4.5 FP Growth Implementation

Once the development of FP tree from the set of transactions have been done, then proceed with the FP growth stage to find a qualified and frequently of itemset. The FP growth has 3 main steps: conditional pattern base, FP tree conditional generation, and finding the frequent itemset.

a. The conditional pattern base step is looking for the minimum support at FP tree according to frequency of the path ending with the smallest support value namely Eventlog (E), OS User (OU), and AV Update (AV).
b. Based on the FP tree conditional from AV process, the resuls show that AV, OU, and E meet minimum support.
c. Finding the frequent itemset. In Table 2, the results show that not all itemset can be calculated because the formula rule is if A then B. Therefore, the itemset can only calculated if they have contained two items at least. In the following is a confidence calculation of 9 subsets i.e. (AV, OU, E):

Table 2. Frequent itemset result

Suffix	Frequent itemset
AV	(AV, (AV, E), (AV, OU), (AV, E, OS), (AV, OS, E), (AV, E, OS))
OU	(OU, (OU, AV), (OU, E), (OU, AV, E))
E	(E, (E, AV), (E, OU), (E, AV, OU), (E, AV, OU))

$$\mathrm{AV} \rightarrow \mathrm{E} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{AV} \rightarrow \mathrm{OU} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{AV} \rightarrow \mathrm{E}^\wedge\mathrm{OU} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{AV}^\wedge\mathrm{E} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{AV}^\wedge\mathrm{OU} \rightarrow \mathrm{E} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{OU} \rightarrow \mathrm{AV} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{OU} \rightarrow \mathrm{E} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{OU} \rightarrow \mathrm{AV}^\wedge\mathrm{E} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{E} \rightarrow \mathrm{AV} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{E} \rightarrow \mathrm{OU} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{E} \rightarrow \mathrm{AV}^\wedge\mathrm{OU} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

$$\mathrm{E}^\wedge\mathrm{OU} \rightarrow \mathrm{AV} = \frac{19}{19} \times 100\% = 100\% \; \textit{Support} \; \frac{19}{20} \times 100\% = 95\%$$

Once the rules of the frequent itemset has been created, the next step is developing knowledge with the association rules. Then, the results of the association rules for each itemset presented as Table 3 below:

Table 3. Association rules

Itemset	Association rules
AV → E	If update is Anti-Virus (AV) then definitely take an Eventlog (E) computer
AV → OU	If update Anti-Virus (AV) then definitely do a review log Operating User (OU)
AV → E^OU	If update is Anti-Virus (AV) then definitely do a review log Operating User (OU) and take Eventlog (E) computer
AV^E → OU	If the Anti-Virus (AV) updates and retrieves the Eventlog (E) of the computer, it will definitely perform an Operating User (OU) review log
AV^OU → E	If the Anti Virus (AV) updates and does the Operating User (OU) review log, it will definitely take an Eventlog (E) computer
OU → AV	If do an Operating User (OU) review log, you will definitely update your Anti Virus (AV)
OU → E	If do an Operating User (OU) review log, you will definitely take the computer's Eventlog (E)
OU → AV^E	If do an Operating User (OU) review log, you will definitely take the computer's Eventlog (E) and update your Anti Virus (AV)
E → AV	If take the Eventlog (E) of the computer, you will definitely update your Anti Virus (AV)
E → OU	If take the Eventlog (E) of the computer, you will definitely log the Operating User (OU)
E → AV^OU	If take an Eventlog (E) computer, you will definitely update your Anti Virus (AV) and do a log review Operating User (OU)
E^OU → AV	If take the Eventlog (E) of your computer and log the Operating User (OU) review, you will definitely update your Anti Virus (AV)

5 Conclusions

Regarding the design analysis and test result of evergreen data with the application of rapid miner 9.0 at Chevron Pacific Indonesia Company (Department of IT Process Control Network), we conclude that the application and implementation of association rules with the FP Growth algorithm showing the trend of evergreen activities both manual calculation and using Rapid Miner 9.0 application. The evergreen activity trends including event-log (E) computers, operating user (OU) reviews, and updating anti-virus (AV) with a support value of 95%. Based on the results of the analysis with the FP growth algorithm and testing with the application of rapid miner Studio 9.0, the association rules on simulation has larger data. Lastly, we have acknowledged that future studies could be trying this approach into manufactures sector.

Acknowledgements. The authors would love to say “Alhamdulillah” and thanks to Allah SWT who has given grace and guidance to the authors, to the Dept. Information System, Tanjungpura University, Indonesia and to the PRO-Knowledge Research Group, Indonesia. We also would like to thank the Committees and Reviewers for their critiques and revision recommendations.

References

1. Maimon, O., Rokach, L.: Introduction to knowledge discovery in databases. In: Data Mining and Knowledge Discovery Handbook, pp. 1–17 (2006)
2. Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J.: Knowledge discovery in databases: an overview. Am. Assoc. Artif. Intell. **13**, 57–70 (1992)
3. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: Knowledge discovery and data mining: towards a unifying framework. In: American Association for Artificial Intelligence, pp. 82–88 (1996)
4. Development of H.R.: Executive summary of Chevron Pacific Indonesia (2019)
5. Ding, Z., Saide, S., Siti Astuti, E., Muwardi, D., Najamuddin, N., Jannati, M., Herzavina, H.: An adoption of acceptance model for the multi-purpose system in university library. Econ. Res. Ekonomska Istraživanja **32**, 2393–2403 (2019)
6. Nebeker, D.M., Tatum, B.C.: The effects of computer monitoring, standards, and rewards on work performance, job satisfaction, and stress. J. Appl. Soc. Psychol. **23**, 508–536 (1993)
7. Bhave, D.P.: The invisible eye? Electronic performance monitoring and employee job performance. Pers. Psychol. **67**, 605–635 (2014)
8. Moussa, M.: Monitoring Employee Behavior Through the Use of Technology and Issues of Employee Privacy in America. SAGE Open. vol. 5, pp. 1–13 (2015)
9. Muwardi, D., Saide, S., Indrajit, R.E., Iqbal, M., Astuti, E.S., Herzavina, H.: Intangible resources and institution performance: the concern of intellectual capital, employee performance, job satisfaction, and its impact on organization performance. Int. J. Innov. Manag. **24**, 1–21 (2020)
10. Larose, D.T.: Data mining: Methods and models. Wiley, New Jersey (2006)
11. Hand, D.J.: Principles of data mining. In: Drug Safety, pp. 621–622 (2007)
12. Ben-Dov, M., Feldman, R.: Data mining and knowledge discovery handbook: text mining and information extraction. In: Data Mining and Knowledge Discovery Handbook, pp. 801–831 (2010)
13. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in database. Am. Assoc. Artif. Intell. **17**, 37–54 (1996)
14. Maimon, O., Rokach, L.: Data mining and knowledge discovery handbook. Springers (2010)
15. Sowmya, R., Suneetha, K.R.: Data mining with big data. IEEE Trans. Data Eng. **26**, 97–107 (2014)

16. Padhy, N., Mishra, P., Panigrahi, R.: The survey of data mining applications and feature scope. Int. J. Comput. Sci. Eng. Inf. Technol. (IJCSEIT) **2**, 43–58 (2012)
17. Witten, I.H., Frank, E.: Data mining: practical machine learning tools and techniques with java implementations. In: Data Mining: Practical Machine Learning Tools and Techniques, vol. 31, pp. 76–77 (2002
18. Imielinski, T., Mannila, H.: A database perspective on knowledge discovery. Commun. ACM **39**, 58–64 (1996)
19. Saide, S., Astuti, E.S., Indrajit, R.E., Trialih, R., Diniaty, A., Dewi, F., Herzavina, H.: What we give, we get back: Investigating the dimensions that influence knowledge sharing on profit enterprise in Indonesia. J. Sci. Technol. Policy Manag. **10**, 1047–1062 (2019)
20. Garg, R., Gulia, P.: Comparative study of frequent itemset mining algorithms Apriori and FP growth. Int. J. Comput. Appl. **126**, 8–12 (2015)