

TPT 1201

RESEARCH METHODOLOGY
IN
COMPUTER SCIENCE

<ASSIGNMENT 1>

PREPARE BY

Student ID	Student Name
1121115725	TAN CHONG RAEN

LECTURE SECTION : TC01
TUTORIAL SECTION : TT02
LECTURER: *DR. POO KUAN HOONG*

DATA MINING AND DATA WAREHOUSE

Lect. Shital P. Bora

Department of Computer Science and Application,
Sardar Patel College,
Chandrapur (India)
Sbora01@gmail.com

August 14, 2015

Abstract

In this paper, the author, *Shital P.Bora*, described how data mining techniques and tools can find the meaningful and valuable information from huge amount of data and convert the information into the knowledge of someone or something behavior. Moreover, P.Bora declared that data mining would help the end-users to make a better decision to avoid unnecessarily cost, or find the trends that shows the end-user more precisely insight instead of graphics generated by computer. In additional, P.Bora also explained the steps for completed a data mining project.

Contents

1	Problem solved	1
2	Related work / Prior research	1
3	Methodology	2
4	Claimed contributions	2
5	Conclusion	3
6	Possible extension / Future work	3
	References	3

1 Problem solved

Data Mining techniques and methods help business organizations to reduce their cost and time for discovery of patterns of data. Besides that, Data Mining also provides a competitive advantage that the accurately prediction of customer response to change in the companys business rules.

2 Related work / Prior research

Data mining process consists four different steps, which are data selection, data transformation, data mining and result interpretation. In this paper, P.Bora cited from *Reality Check for Data Mining* to briefly describe the process of data mining. Furthermore, P.Bora also summarized the two approaches of data mining, which are *Verification Driven Data Mining* (VDDM) and *Discovery Driven Data Mining* (DDDM), in the cited article. Besides that, the author of the cited article, *Evangelos Simoudis* did explanation of two approaches is more details and specified than this paper. Moreover, E.Simoudis stated that verification driven has three operations are associated with data mining, which included multidimensional analysis, query and reporting and statistical analysis. In additional, E.Simoudis also stated that discovery driven has four operations are associated with data mining, which are predictive modeling, deviation detection, link analysis and database segmentation. Not only that, E.Simoudis has showed that the four operations of discovery driven are supported by numerous data mining techniques. For examples, supervised induction techniques support predictive modeling, statistical techniques support deviation detection, and link analysis is supported by sequence discovery and association discovery. (Simoudis, 1996)

3 Methodology

Data Mining consists two approaches: *verification driven*, or *discovery driven*. Verification driven is to verify a hypothesis that provide by a user. This driven will study and find the data that is support the patterns, or the associations. Moreover, verification driven is the most common applied driven for the Data Mining. In the other hand, discovery driven is more complicated and more depend on structure search of data for patterns or associations and analyze them for value. There are four different ways to perform discovery driven, which are link-analysis technique, predictive modeling including neural nets, database segmentation, and deviation detection. Therefore, produce a data mining project is not that easily. It involves a difference of feedback loops if the result do not achieve the optimal values. The worst case is the user has to refine the earlier step repeatedly, or even reproduce the entire process. To produce a data mining project included six phases which are *Problem Definition*, *Data Exploration*, *Data Preparation*, *Modeling*, *Evaluation*, and *Deployment*. During problem definition phase, the experts have to define the objectives and requirements for solve the relevant business problems. After that, domain experts will start to explore data with data analysis tools at data exploration phase. While domain experts finished exploration of data then they will format the data because some mining functions only accept a certain of format but the meaning of data would not change. In additional, they will also create some new derived variables during data preparation phase. Next is the modeling phase, data mining experts and domain experts will work together to build a high quality model. Since they need to perform different mining functions with parameters to achieve the optimal values. So, they will also doing the evaluation during the modeling phase. Once they are satisfied the mining result, data mining experts will export the result into database tables or other application for example *Microsoft Excel* in the deployment phase.

4 Claimed contributions

Shital P. Bora showed that using data mining to find the meaningful information from a massive of data is more efficiency and accurately. Other than that, P.Bora provide a step by step guidelines for develop a data mining project. Furthermore, P.Bora also provide some real-world examples to prove that the Data Mining has been widely applied in the business world since it is a method to obtain competitive advantage.

5 Conclusion

In conclusion, data mining technique and method is the most accurately analysis data tools since it allow user to analyze data from multidimensional. Not only that, it can categorize and summarize the relationships identified automatically to provide more useful information that help user to make a better decision. From the article, it showed that to produce a data mining project need to pass six different steps or process. The experts must be understand the problems and study the data before they process to another four process. Moreover, during fourth step of data mining process, which is modeling, they will apply a lot of different mining functions that to ensure the outcome is more accurately. Therefore, we can confirm that data mining technique is the most accurately analysis data tools. However, all the data mining methods was developed by human so there is some false patterns or associations may be found. Furthermore, data mining more widely applied and more advance. The importance of internet and data warehousing also will be more significantly.

6 Possible extension / Future work

In this paper, we have learned the structures and functions of the data mining and the process of the data mining. Furthermore, we have know that data mining can be apply to any others area if that area has a massive of data. Beside that, we found that the growth of data mining is raising because the demand of internet and data warehouse is growing constantly. And the reason of demand of internet increasing is a lot of people are using social media to communicate with each other. Therefore, we can apply data mining to social media and using data mining technique and method to discover more useful and valuable information. However, there might have some challenges for extend data mining to social media. (Zafarani, Abbasi, & Liu, 2014)

References

- Simoudis, E. (1996, October). Reality check for data mining. , 26-33. doi: 10.1109/64.539014
- Zafarani, R., Abbasi, M. A., & Liu, H. (2014, April). Social media mining. , 15-16. Retrieved from <http://dmml.asu.edu/smm/SMM.pdf>

Paper Title	DATA MINING AND WARE HOUSING
Author(s)	Lect. Shital P. Bora
Abstract/Summary	
Problem Solved	
Claimed Contributions	
Related work	
Methodology	
Conclusions	
What did you learn(algorithm / experiments details)? Possible extension / Future work	
References	