DATA WAREHOSUING/MINING OF FINANCIAL AND ECONOMIC DATA: "A COMPREHENSIVE STUDY, DESIGN, DEVELOP, IMPLEMENT AND MAINTAIN OF THE FINANCIAL AND ECONOMIC MANAGEMENT SYSTEM (FEMS)"

By

TEE KIAM KHAI

Project Paper Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Information Technology

OPEN UNIVERSITY MALAYSIA
(2007)

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 Executive Summary

Every day, enormous amounts of information are generated from all sectors, whether it is economic, business, exchange rate, the scientific community, the World Wide Web (WWW), or more of many readily available off-line and online data sources. From all of this, which represents a sizable repository of data and information, it is possible to generate worthwhile and usable knowledge. As a result, the field of Data Mining (DM) has grown in leaps and has shown great potential for the future. The purpose of this project is to clean, centralize and warehouse all the raw data for the analysts or economists. The economists or analysts will identify many of the critical and future trends in the field of DM that helps the Bank to determine or predict the behaviors of the data. This assists in discovering information within the data that queries and reports can't effectively reveal. Raw data by itself however does not provide much information. Data mining, or knowledge discovery, is the computer assisted of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing the Economics Department of Bank Negara Malaysia (BNM) to make proactive knowledge-driven decisions. They scour databases for hidden patterns, finding predictive information that many economists /analysts may miss because it lies outside their expectations.

1.2 Before the Project

Data, data everywhere. Since the establishment of the Bank in 1959, a lot of data were collected by Economics Department (JEK) and stored in various format such as tables in word documents, spreadsheets, Microsoft access databases or even text files. The data is stored either in the diskettes, hard drives or hard printed copies. Recent advances in data collection and data storage technology have enabled the Bank to accumulate large quantities of operational, economic and financial data from their daily activities.

Since the amount of raw data collected in the department is exploding, raw data by itself, however, does not provide much information. The dilemma faces by the department is data rich but information poor. The data collected by JEK is valuable for various purposes. For example, the data collected from the usage of credit cards issued by the financial institutions may be used to understand the pattern of purchasing behaviour and this invaluable information could be challenged to the right industries for further action. Traditional data analysis techniques, which are often based on statistical inference and hypothesis testing, encounter tremendous difficulties in handling large quantities of data. Motivated by the need to overcome the problems of analyzing large-scale data, analysts and economists have turned to data mining for assistance.

1.3 Problems Face by the Economists and Analysts and the Need of Having a Centralized Database System

The Department reveals that the present system of data collection and compilation has the following main weaknesses:

- Significant time spent in re keying and checking of data by different sections of the same department;
- Duplication of work in compiling and updating of data stored by different sections within the same department;
- No sharing of data compiled, except through exchange of diskettes or hard copy reports. Online access of data is impossible;

- Retrieval of information for detailed analysis is cumbersome, as there is no proper indexing of data files; and
- The raw or source data used by different sections may be different as the data is stored in different storage devices or with the owners/authors.

In order to resolve these problems, the Economics Department has recommended for the implementation of a financial and economic management system (FEMS), as part of the measures to reduce time spent on data compilation, to improve information accessibility and availability in order to improve the ability of the department to provide better quality reports, analysis, forecasting and recommending in a shorter time. The problems discussed can be summarized as follows:

- Data Explosion Problem
  - Automated data collection tools and mature database technology lead to tremendous amounts of data accumulated and/or to be analyzed in databases, data warehouses, and other information repositories.
- The Economics Department is drowning in data, but starving for knowledge!
- To resolve the problems, the department has suggested to implement data warehousing and data mining in stages:
  - Data warehousing and on-line analytical processing (OLAP) to filter and clean up existing databases in Phase I; and
  - Mine interesting knowledge (rules, regularities and patterns) from the existing databases in Phase II.

1.4 Project Objectives

The FEMS project is started in view of enhancing the capability of JEK and JASM's staff in performing their analytical activities. At present their capability is partly hindered due to the limitation of the current Time Series Management Tool, DBank, which acts mainly as a data storage tool for the economic and financial data and unable to perform multi-dimensional analysis. The main objectives of the project are as follows:

- To facilitate multi-dimensional analysis and interactive reporting for all levels of users;
- To facilitate consolidation of data from several sources;
- To improve data analysis, estimation and forecasting by having the ability of performing scenario analysis;
- To segregate access to data based on security classification level;
- To keep track of data changes up to the lowest level;
- To generate pre-defined interactive and ad-hoc tables and charts for Annual Report, Quarterly Bulletin, Board Papers, Board Briefing, Travel Books and other management reports from the centralised database and timely submission of tables or reports to the management;
- To minimize the risk of losing important data since the data is centralised as opposed to current arrangement where spreadsheets are kept in individual PCs or with the authors;
- Data extracted from the centralised database will be more reliable and consistent;
- To reduce the searching and waiting time for the authors to generate the tables or reports; and
- To have the ability of searching any time series across the entire databases.

1.5 Financial and Economic Management System (FEMS) – An Business Intelligence (BI) / Online Analytical Processing (OLAP) Project

The FEMS – BI/OLAP is a step forward in using technology to facilitate a more efficient, effective and speedier way of generating information required for the management and publication. It will interface with the existing information system and non-system databases to produce database-linked outputs. An efficient and effective way of generating and disseminating of data or information is essential for the Economics Department; the new system should be able to meet the needs and requirements for the entire department and the users. Data Warehousing follows by Data mining are chosen by the Economics Department for the following reasons:

i. Clean and store all the raw or source data in a common place;
ii. Filter, visualize and interpret the pattern of the data; and
iii. Discover the useful knowledge from large data repositories that can be shared by all the economists and analysts.

For reason (i), it would be the responsibility of database administrator (DBA) while reasons (ii) and (iii) would be the responsibilities of the economists and analysts. The notion of usefulness has different meanings to different analysts or economists. Data mining or knowledge discovery is the computer-assisted process of digging through and analyzing enormous sets of historical economic and financial data, and then extracting the meaning of the data. This data mining tools predict behaviours and future trends, allowing the economists and analysts to be more proactive and make knowledge-driven decisions, as it can reveal previously unknown information about the trends or changes over time.

Data mining is often considered to be an integral part of knowledge discovery in databases (KDD). The process of KDD of the raw data into useful knowledge can be shown in Figure 1. It consists of a series of transformations, including data preprocessing and post processing. Data preprocessing transforms the raw data into a format suitable for subsequent analysis. It also helps to identify subsets of the data that are relevant for a particular data-mining task. As the raw data may be stored in different formats and in different databases, a large amount of time may be spent on data preprocessing which includes data cleaning, metadata repository, key families, naming convention, among others.

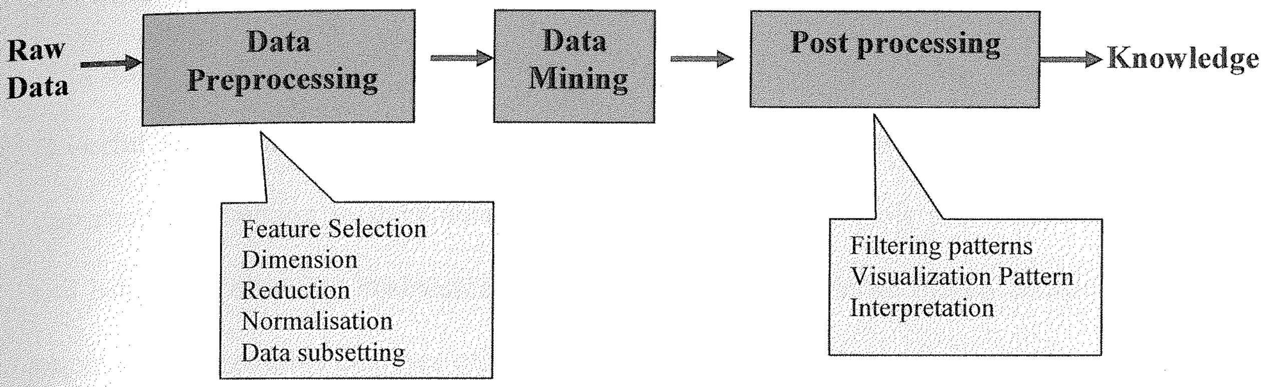

Figure 1: The process of knowledge discovery in databases (KDD)

Post processing encompasses all the operations that are performed to make the data mining results more accessible and easier to interpret. For example, spurious results can be filtered according to a variety of measures. Visualization techniques may also be employed to help analysts explore and understand the data mining results.

REFERENCES

1. Data Warehousing Concepts and Strategies *Stefan M. Neikes Sumit Sircar Bijoy Bordoloi*

2. Data mining by Doug Alexander at www.eco.utexas.edu/~norman/BUS.FOR/course.mat/Alex

3. H. Edelstein, "Technology How to: Mining Data Warehouses", Information Week (January 8, 1996) pg 48-47

4. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

5. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (Paperback) by Michael J. A. Berry, Gordon S. Linoff

6. Data Mining: Opportunities and Challenges, Wang, John (Editor). Hershey, PA, USA: Idea Group Inc., 2003

7. Data Mining in Times Series Databases, Mark Last, Abraham Kandel & Horst Bunke,World Scientific Publishing Company.

8. ERP & Data Warehousing in Organizations: Issues and Challenges, Gerald Grant Carleton University, Canada, Hershey, PA, USA: Idea Group Inc

9. The Research Project: How to Write It. Berry, Ralph. Florence, KY, USA: Routledge

10. How to write a thesis, Rowena Murray, Open University Press, McGraw Hill Education.

11. Discovering Knowledge in Data: An Introduction to Data Mining, Hoboken, NJ, USA: John Wiley & Sons, Incorported.

12. http://www.applic.com

13. Data Sources: IMF's CDs which contain the BOP, GTS, DOT and IFS data

14. Guidelines provided by Financial Thomson in connecting to Datastream Server

15. Data Mining – Introductory and Advanced Topics by Margaret H Dunham,Pearson Education Inc