

Analysis of Data Mining as a Key Enabling Technology for Business Intelligence

Vijay Pratap Singh¹, Ashish Kumar², Kajal Sawant³

Assistant Professor¹, Students^{2,3}

*Department of Computer Science and Engineering,
Prestige Institute of Engineering Management & Research*

Corresponding Author's Email: - ashishkumar23@gmail.com²

Abstract

The technique of discovering or "mining" new knowledge from huge datasets and applying it to decision making is known as data mining (DM). Companies have recently used a variety of data mining approaches to better understand their consumers and their performance, as well as to address complicated business challenges. Data mining is a method of developing business intelligence (BI) using data collected, organised, and stored by an organisation. The goal of this study is to examine data mining as a technology that enables business intelligence. The study's data collection technique was document analysis, while the data analysis method was qualitative.

Keywords: *Frequent Item Mining, Data Mining, Market Basket Analysis, Parallel Execution*

INTRODUCTION

Data mining is the examination of (typically enormous) observational data sets in order to discover previously unknown associations and summarise the data in unique ways that are both intelligible and beneficial to the data owner (Hand et al. 2001). Data mining is a relatively new and multidisciplinary

branch of computer science. It is at the crossroads of several fields, including artificial intelligence, machine learning, statistics, and database systems. Knowledge extraction, pattern analysis, data archaeology, information harvesting, pattern searching, and data dredging are just a few of the terms related to data mining. The name "data mining" is

relatively new, but the concepts underlying it are not. Many data mining approaches are based on classical statistical analysis and artificial intelligence work from the early 1980s. Pattern extraction from data has been done for ages. Bayes' theorem (1700s) and regression analysis were early tools for spotting patterns in data (1800s).

Due to the ever-changing requirements and severe rivalry from customers globally, data mining has suddenly captured the attention of the industry. The huge drop in the cost of data storage and processing hardware and software has made data mining more widespread among people in businesses and other organisations.

DATA MINING CONCEPTS

Data mining is a multi-step technique for extracting knowledge from huge databases. Data mining technologies discover patterns in data and the correlations and rules that go with them. By detecting associations between data sets or databases, the acquired data may be put into a prediction or classification model. These models and guidelines are then used to guide decision making and examine the consequences of such actions.

Data mining is just the next stage in the data analysis process. Rather than

receiving queries on conventional or user-specified associations, data mining takes a step further by discovering relevant correlations in data.

Data mining is made up of five basic components:

1. Extract, transform and load transaction data on the data warehouse system.
2. Store and manage data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyze data through application software.
5. Display the data in a useful format such as a graph or table.

Data in Data Mining

Data is often defined as a collection of facts gained from experiences, observations, or experiment findings. Data can be a collection of numbers, text, pictures, or other variables.

Table 1: Types of Data

	Level of measurement	Examples
Categorical	Nominal (no inherent order in categories)	Eye colour, ethnicity, diagnosis
	Ordinal (categories have inherent order)	Job grade, age groups
	Binary (2 categories)	Gender
Quantitative	Discrete (usually whole numbers)	Size of household (ratio)
	Continuous (can, in theory, take any value in a range)	Temperature °C/°F (no absolute zero), height, age

Data is regarded as the most basic level of abstraction from which information and knowledge may be gained. Data can be classified as categorical or quantitative at the greatest level of abstraction. Categorical data can be further classified into nominal and ordinal data.

How Data Mining works

Based on user queries, data mining software evaluates correlations and patterns in recorded data. In general, data mining seeks four sorts of relationships:

Data mining may be used to detect associations. In market-basket analysis, associations often seek out co-occurring groupings where beer and diapers are running concurrently.

1. **Predictions:** Stored data is used to anticipate the character of future occurrences based on past events, such

as projecting the absolute temperature on a specific day.

2. **Clusters:** Data objects are grouped together based on logical relationships or user choices. Data may be mined to discover market categories or consumer endowments, for example.

3. **Sequential patterns:** Data mining is used to evaluate behavioural patterns and trends. An outdoor equipment store, for example, can predict the possibility of a bag that can be acquired based on a consumer's sleeping bag and hiking boots.

3. Business Intelligence

Company intelligence refers to computer-based methods for analysing business data, such as sales revenue or related costs and revenues for goods and/or segments. Analytics, data mining, business performance management, benchmarking,

text mining, and predictive analytics are common functions of business intelligence tools. Business intelligence solutions are in high demand at a time when most information technology (IT) products are in short supply (Whiting, 2003). While the phrase "business intelligence" is relatively new, a computer-based business intelligence system was first introduced forty years ago. Real-time data warehousing, data mining, automated differentiation and exclusion detection, proactive alert with automated recipient decision, seamless follow-through workflow, automated learning and refinement, geographic information systems, and data visualisation are all important components of BI (Langseth and Vivatrat, 2003).

LINKS BETWEEN DM, BI, AND KM

Data mining connects corporate intelligence with knowledge management (KM). Data mining is a powerful business intelligence method for discovering new knowledge. Because it requires human knowledge, data mining is a knowledge management process (Brahman et al., 1996). This approach to DM readily links BI and KM. In terms of employing data mining to enhance human knowledge, it is a tool for knowledge management. For example, given a sales database, DM

might reveal previously undisclosed customer purchasing trends to the data miner.

5. Data Mining Applications

Data mining has been shown to be extremely effective and beneficial in a variety of industries, as evidenced by the examples below. A few examples of how data mining is used in financial data analysis include loan payment prediction and customer credit policy analysis; customer categorization and clustering for targeted marketing, detection of money laundering and other financial crimes; and so on. Some examples of data mining in the retail industry include the design and construction of data warehouses based on data mining benefits; multidimensional analysis of sales, customers, products, time, and region; analysis of the effectiveness of sales campaigns; analysis of customer loyalty; and so on. Data mining applications in the telecommunications industry include multidimensional telecommunication data analysis; mobile telecommunication services, the use of visualisation tools in telecommunication data analysis; and so on. Data Mining is also used in Biological Data Analysis, for example, semantic integration of heterogeneous, distributed genomic and proteomic databases;

alignment, indexing, similarity search, and comparative analysis of multiple nucleotide or protein sequences; discovery of structural patterns and analysis of genetic networks and protein pathways; visualisation tools in genetic data analysis, and so on. Aside from the aforementioned applications, data mining is commonly utilised in scientific applications, such as data collection for scientific research.

DATA MINING PROCESS

A broad method is generally followed in order to carry out data mining projects in a systematic manner. In reality, numerous standardised data mining procedures are employed, including CRISP-DM, SEMMA, the KDD process, domain-specific methodology, and so on. In the mid-1990s, the CRISP-DM (Cross Industry Standard Process for Data Mining) identified six key processes in the data mining process. The procedure is as follows:

Step 1: Business Understanding—understand the project's goals and demands from a business standpoint, and then translate this information into a definition of the data mining problem and the initial strategy developed to meet the objectives.

Step 2: Data Understanding—Begin with gathering data, then learn about it, discover data quality concerns, look for insights in the data first, or uncover intriguing subsets to test theories about hidden information.

Step 3: Data Preparation: All operations necessary to generate the final data set (data that will be input into the modelling tool) from the original raw data are included in this step. Table, case, and attribute selection, as well as data processing and cleansing for modelling tools, are all tasks.

Step 4: Model Construction: Choose and use a range of modelling methodologies, then calibrate the instrument's parameters to the proper values. In general, numerous approaches exist for the same data mining issue type. Some techniques have data requirements that must be satisfied. As a result, it is frequently required to return to the data preparation phase.

Step 5: Testing and Evaluation- Evaluate the model as a whole, as well as the procedures used to construct it, to ensure that it fits the business objectives. Determine whether there is a substantial business problem that is not being addressed adequately. The application of

data mining results is determined at the end of this step.

Step 6: Deployment entails managing and presenting the findings of data mining. Deployment might be as easy as creating a report or as sophisticated as putting in place a recurring data mining process.

Iterative data mining is used. After a solution is launched, the data mining process continues. Lessons learnt during the process may raise new business concerns. Data changes may necessitate the development of new models. Subsequent data mining techniques gain from earlier ones' experience. Because the latter stages of CRISP-DM are based on the results of the previous phases, it is critical to pay close attention to the earlier steps to avoid starting the research on the wrong foot. Putting in place a repeatable data mining method

CONCLUSION

Data mining is a strong new technique with enormous promise for extracting hidden information from massive databases. Data mining techniques forecast future trends and behaviours, enabling organisations to make informed decisions. Data mining analyses that are automated and prospective go beyond the analytics

given by specialised decision support systems. Traditionally, data mining techniques can answer commercial concerns that are too time-consuming to handle. Corporate intelligence is the art of extracting business value from data by answering simple questions. Companies can better support choices with robust BI. Using a strong computer system to generate a fact-based choice promotes trust in any judgement.

This article focuses on the issues of data mining and business intelligence, with the primary goal of exploring the general characteristics of these two areas. Basic data mining policies and practises are covered in this section. Data mining applications have been discussed. The connections between data mining, business intelligence, and knowledge management are also briefly discussed. Data mining is regarded as an excellent business intelligence method for knowledge discovery. Using the benefits of both areas effectively will improve the business's performance.

REFERENCES

1. Benoit, G. (2005). Data Mining. Annual Review of Information Science and Technology, Volume 36 (1), pp. 265–310.

2. Bose, I., Mahapatra, R. K., (2001). Business data mining – a machine learning perspective. *Information & Management* (Volume 39), pp. 211-225.
3. Brachman, R.J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G. and Simoudis, E. (1996), "Mining business databases", *Communications of the ACM*, Vol. 39 No. 11, pp. 42-8.
4. Cao, Longbing., Yu, Philip S., Zhang, Chengqi. & Zhang, Huaifeng. (2009). *Data Mining for Business Applications* [electronic resource]. Boston, MA: Springer US.
5. Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
6. Elder IV, John F. & Abbott, Dean W. (1998). *A Comparison of Leading Data Mining Tools*. Fourth International Conference on Knowledge Discovery & Data Mining, August 28, 1998. New York.
7. Hai Wang, Shouhong Wang, (2008) "A knowledge management approach to data mining process for business intelligence", *Industrial Management & Data Systems*, Vol. 108 Iss: 5, pp.622 – 634.
8. Hand, D., Mannila, H. & Smyth, P., (2001). *Principles of Data Mining*, MIT Press.
9. Hema, R., & Malik, Neha (2010). *Data Mining and Business Intelligence*. 4th National Conference; INDIACom-2010, Computing For Nation Development, February 25 – 26, 2010. New Delhi.
10. Karl Rexer, Heather Allen, & Paul Gearan (2010) 2010 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2010.
11. Kudyba, Stephan & Hoptroff, Richard. (2001). *Data mining and business intelligence* [Elektronisk resurs] a guide to productivity. Hershey, Pa.: Idea Group Pub.
12. Langseth, J. and N. Vivatrat (2003) "Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound," *Intelligent Enterprise*, (5)18, pp. 34-41.
13. Negash, S. (2004). *Business Intelligence*, *Communications of the Association for Information Systems* (Volume 13, 2004), pp. 177-195.

14. Newell, Robertson, Scarbrough & Swan (2009) Managing Knowledge Work and Innovation (second edition), Palgrave, Macmillan.
15. Robson, C. (2011) Real World Research A Resource for Users of Social Research Methods in Applied Settings, 3rd ed., Blackwell Publishing.
16. Sørensen, Carsten (1993): Introducing CASE Tools into Software Organizations. Ph.D. Dissertation Department of Mathematics and Computer Science, Aalborg University, [R-93-2011].
17. Turban E., Sharda, R. Delen, S. (2011) Decision Support and Business Intelligence Systems (Ninth edition), pp. 190-240.
18. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases, AI Magazine (1996), pp. 37–53.
19. Whiting, R. (Jan. 13, 2003) “Look Within—Business-Intelligence Tools have a New Mission: Evaluating All Aspects of a Company’s Business”, InformationWeek, pp. 32.