Volume: 4, Issue: 7
Page:46-51
2020

ISSN 2520-4750 (Online) & ISSN 2521-3040 (Print)

International Journal of Science and Business

Reviewing Data Mining as an enabling technology for BI

Mustafa Nizamul Aziz & A.K.M. Monzurul Islam

Abstract:
Data mining (DM) is the process of discovering or "mining" new knowledge from large databases and applying it to decision making. In recent times, companies have been using a wide range of data mining techniques to better understand their customers and their performance and solve complex business problems. Data mining is a way to develop Business Intelligence (BI) from data that an organization collects, organizes, and stores. The purpose of this paper is to review Data Mining as an enabling technology for Business Intelligence. Documents analysis was used as the data collection method and a qualitative approach was used for data analysis in the study.

IJSB
Accepted 24 June 2020
Published 26June 2020
DOI: 10.5281/zenodo.3908987

Keywords: Data, Information, Knowledge, Data Mining, Business Intelligence, Decision Support.

About Author (s)

Mustafa Nizamul Aziz (Corresponding Author), Senior Lecturer, East West University, Dhaka, Bangladesh.
A.K.M. Monzurul Islam, Independent University, Bangladesh.

1. Introduction

"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" (Hand et al. 2001). Data mining is a young and interdisciplinary field of computer science. It is at the intersection of many disciplines, including artificial intelligence, machine learning, statistics and database systems. Many other names that are associated with data mining include *knowledge extraction, pattern analysis, data archaeology, information harvesting, pattern searching, and data dredging.* The term *data mining* is relatively new, but not the ideas behind it. Many of the methods used in data mining are rooted in traditional statistical analysis and artificial intelligence work from the early 1980s. The extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). Data mining has suddenly attracted the attention of the business due to the ever-changing requirements and intense competition from consumers worldwide. A significant reduction in the cost of hardware and software for data storage and processing has made data mining very common to humans in business and other organizations.

2. Data Mining concepts

Data mining is a multi-layered process that gathers knowledge from large databases. Data mining tools find patterns from data and find associations and rules associated with them. The data collected can be applied to a prediction or classification model by identifying relationships among data records or databases. Those models and rules then guide decision making and assess the impacts of those decisions. Data mining is really just the next step in the process of analyzing data. Instead of getting queries on standard or user specified relationships, data mining goes a step farther by finding meaningful relationships in data. Data mining consists of five major elements:

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.

2.1 Data in Data Mining

Data usually refers to a collection of facts obtained as experiences, observations, or results of experiments. Data can be a measure of numbers, words, images, and similar variables.

Table 1: Types of Data

Type 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 are viewed as the lowest level of abstraction from which information and knowledge are derived. At the highest level of abstraction, one can classify data as categorical or numeric. The categorical data can be subdivided into nominal or ordinal data.

2.2 How Data Mining works

Data mining software analyzes relationships and patterns in stored data based on user queries. Generally, four types of relationships are sought in data mining:

a) Associations: Data can be mined to identify associations. Associations typically seek out co-occurring groups where beer and diapers are running concurrently in market-basket analysis.
b) Predictions: Stored data is used to estimate the nature of future events of certain events based on past events such as predicting the absolute temperature of a particular day.
c) Clusters: Data items are classified according to logical relationships or user preferences. For example, data can be mined to identify market segments or consumer endowments.
d) Sequential patterns: Data is mined to assess behavioral patterns and trends. For example, an outdoor equipment retailer can estimate the likelihood of a bag, which can be purchased based on a consumer's sleeping bag and hiking boots.

3. Business Intelligence

Business intelligence refers to computer-based methods such as sales revenue or related expenses and revenues for products and/or segments used to analyze business data. Typical functions of business intelligence technologies include analytics, data mining, business performance management, benchmarking, text mining and predictive analytics. The demand for business intelligence applications is also increasing at a time when demand for most information technology (IT) products is slick (Whiting, 2003). While the term business intelligence is fairly new, a computer-based business intelligence system appeared forty years ago. Important components of BI (Langseth and Vivatrat, 2003) are: 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, data visualization.

4. Links between DM, BI, and KM

Data mining is the link between business intelligence and knowledge management (KM). Data mining is an effective business intelligence tool for knowledge discovery. The data mining process is a knowledge management process because it involves human knowledge (Brahman et al., 1996). This view of DM naturally connects BI to KM. Data mining is a tool for knowledge management in terms of using data mining as a tool to expand human knowledge. For example, given the sales database, DM can expose consumers' buying patterns previously unknown to the data miner.

5. Data Mining Applications

Data mining has proven to be very successful and helpful in many fields, some of which are illustrated by the following representative examples. Few typical cases of using data mining in *Financial Data Analysis* are loan payment prediction and customer credit policy analysis, classification and clustering of customers for targeted marketing, detection of money laundering and other financial crimes, etc. A few examples of data mining in the *Retail Industry* are design and construction of data warehouses based on the benefits of data mining; multidimensional analysis of sales, customers, products, time, and region; analysis of the effectiveness of sales campaigns; analysis of customer loyalty, etc. Uses of Data Mining for the *Telecommunication Industry* include multidimensional analysis of telecommunication data, mobile telecommunication services, use of visualization tools in telecommunication data analysis, etc. Data Mining is also used for *Biological Data Analysis*, such as semantic integration of heterogeneous, distributed genomic and proteomic databases; alignment, indexing, similarity search, and comparative analysis of multiple nucleotide/protein sequences; discovery of structural patterns and analysis of genetic networks and protein pathways; visualization tools in genetic data analysis, etc. Besides these stated above, Data mining is frequently used in *Scientific Applications*, such as in *Data collection* part for scientific research.

6. Data Mining Process

In order to systematically carry out data mining projects, a general process is usually followed. There are several standardized data mining processes used in practice, such as *CRISP-DM*, *SEMMA*, *KDD process*, *Domain-specific methodology*, etc. The *CRISP-DM* (Cross Industry Standard Process for Data Mining), proposed in the mid-1990s, has six basic steps in the data mining process. The steps are as follows:

Step 1: Understanding of business- Understand the goals and needs of the project from a business perspective, and then turn this knowledge into a definition of the data mining problem and the initial plan designed to achieve the objectives.
Step 2: Understanding of data- Start by collecting data, then learn the data, identify data quality issues, search for insights first in the data, or find interesting subsets to make hypotheses about hidden information.
Step 3: Preparation of data- Includes all activities required to construct the final data set (data that will be fed into the modeling tool) from the initial raw data. Tasks include table, case, and attribute selection as well as transformation and cleaning of data for modeling tools.
Step 4: Model Building- Select and apply a variety of modeling techniques and calibrate the instruments parameters to the correct values. In general, there are several methods for the same data mining problem type. Some methods have specific requirements in the form of data. Therefore, it is often necessary to step back into the data preparation phase.
Step 5: Testing and Evaluation- Evaluate the model as a whole and review the steps implemented to build the model to ensure it meets the business objectives. Decide if there is some significant business problem that is not considered adequate. At the end of this phase, the usage of data mining results is determined.
Step 6: Deployment- Managing and presenting data mining results. Deployment can be as simple as generating a report or as complex as implementing a repeatable data mining process.

Data mining is iterative. A data mining process continues after a solution is deployed. The lessons learned during the process can trigger new business questions. Changing data can require new models. Subsequent data mining processes benefit from the experiences of previous ones. In *CRISP-DM*, as latter steps are built on the outcome of the former steps, one should pay extra attention to the earlier steps in order not to put the whole study on an incorrect path from the onset.

7. Conclusion

Data mining, *the extraction of hidden information from large databases*, is a powerful new technology with great potential. Data mining tools predict future trends and behaviors, allowing businesses to make knowledge-based decisions. Automated and prospective analysis of data mining moves beyond analytics provided by specific decision support systems. Data mining tools can traditionally answer business questions that are too time consuming to solve. Business Intelligence is the art of getting business benefits from data by answering basic questions. With a strong BI, companies can better support decisions. Creating a fact-based decision through a solid computer system provides confidence in any decision.

This paper mainly focuses on the aspects of data mining and business intelligence and the main objective of the study is to explore the general characteristics of the two domains. Basic data mining policies and techniques are discussed here. Data mining applications have been talked about. The relationship between data mining, business intelligence, and knowledge management is also briefly described. Data mining is considered an effective tool for business intelligence for knowledge discovery. Properly using the advantages of both domains will give the business a better performance.

References

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

Bose, I., Mahapatra, R. K., (2001). Business data mining – a machine learning perspective. Information & Management (Volume 39), pp. 211-225.

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.

Cao, Longbing., Yu, Philip S., Zhang, Chengqi. & Zhang, Huaifeng. (2009). *Data Mining for Business Applications [electronic resource]*. Boston, MA: Springer US.

Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.

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.

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.

Hand, D., Mannila, H. & Smyth, P., (2001). *Principles of Data Mining,* MIT Press.

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.

Karl Rexer, Heather Allen, & Paul Gearan (2010) 2010 Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2010.

Kudyba, Stephan & Hoptroff, Richard. (2001). *Data mining and business intelligence [Elektronisk resurs] a guide to productivity*. Hershey, Pa.: Idea Group Pub.

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.

Negash, S. (2004). Business Intelligence, *Communications of the Association for Information Systems* (Volume 13, 2004), pp. 177-195.

Newell, Robertson, Scarbrough & Swan (2009) *Managing Knowledge Work and Innovation* (second edition), Palgrave, Macmillan.

Robson, C. (2011) *Real World Research* A Resource for Users of Social Research Methods in Applied Settings, 3rd ed., Blackwell Publishing.

Sørensen, Carsten (1993): *Introducing CASE Tools into Software Organizations*. Ph.D.

Dissertation Department of Mathematics and Computer Science, Aalborg University, [R-93-2011].

Turban E., Sharda, R. Delen, S. (2011) *Decision Support and Business Intelligence Systems* (Ninth edition), pp. 190-240.

U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases, AI Magazine (1996), pp. 37–53.

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.

Cite this article:

Mustafa Nizamul Aziz & A.K.M. Monzurul Islam (2020). Reviewing Data Mining as an enabling technology for BI. *International Journal of Science and Business, 4*(7), 46-51. doi: https://doi.org/10.5281/zenodo.3908987

Retrieved from http://ijsab.com/wp-content/uploads/567.pdf

Published by

IJSAB International