Data Mining and Business Intelligence

R. Hema¹ and Neha Malik²
¹,²Department of Information Technology, Dronacharya College of Engineering, Gurgaon
¹hemaramji@gmail.com and ²nehag78@gmail.com

ABSTRACT
Data mining is a combination of database and artificial intelligence technologies. Although the AI field has taken a major drive in the last decade; this new emerging field has shown that AI can add major contributions to existing fields in computer science. In fact, many experts believe that data mining is the third hottest field in the industry behind the Internet, and data warehousing.

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. Relationships that were thought to have not existed, or ones that give a more insightful view of the data. For example, a computer-generated graph may not give the user any insight, however data mining can find trends in the same data that shows the user more precisely what is going on. Using trends that the end-user would have never thought to query the computer about. Without adding any more data, data mining gives a huge increase in the value added by the database. It allows both technical and non-technical users get better answers, allowing them to make a much more informed decision, saving their companies millions of dollars.

Business intelligence (BI) is about creating value for our organizations based on data or, more precisely, facts. While it seems like another buzzword to describe what successful entrepreneurs have been doing for years, if not centuries, that is, using business common sense. From a modern business-value perspective, corporations use BI to enhance decision-making capabilities for managerial processes (e.g., planning, budgeting, controlling, assessing, measuring, and monitoring) and to ensure critical information is exploited in a timely manner. And computer systems are the tools that help us do that better, faster, and with more reliability.

This paper concludes with a case study of Business Intelligence.

KEYWORDS
Data, Information, Knowledge, Data Warehouse, OLAP, Data Mining, Business Intelligence, Stock Return Volatility, Decision Support, Information Availability

1. INTRODUCTION
Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

Business Intelligence is the art of gaining a business advantage from data by answering fundamental questions, such as how various customers rank, how business is doing now and if continued the current path, what clinical trials should be continued and which should stop having money dumped into! With a strong BI, companies can support decisions with more than just a gut feeling. Creating a fact-based "decisioning" framework via a strong computer system provides confidence in any decisions made.

2. DATA, INFORMATION AND KNOWLEDGE
An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

2.1 DATA
Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

- operational or transactional data such as, sales, cost, inventory, payroll, and accounting
- nonoperational data, such as industry sales, forecast data, and macro economic data
- meta data - data about the data itself, such as logical database design or data dictionary definitions
2.2 INFORMATION
The patterns, associations, or relationships among all this data can provide information. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

2.3 KNOWLEDGE
Information can be converted into knowledge about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behavior. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

3. DATA WAREHOUSE
Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into data warehouses. Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing, like data mining, is a relatively new term although the concept itself has been around for years. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports data mining.

4. WHAT CAN DATA MINING DO?
Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data.

With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments.

For example, Blockbuster Entertainment mines its video rental history database to recommend rentals to individual customers. American Express can suggest products to its cardholders based on analysis of their monthly expenditures.

5. HOW DOES DATA MINING WORK?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- **Classes**: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- **Clusters**: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- **Associations**: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- **Sequential patterns**: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

6. DATA MINING ARCHITECTURE

Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization, in areas such as promotional campaign management, fraud detection, new product rollout, and so on. Figure 1 illustrates an architecture for advanced analysis in a large data warehouse. The ideal starting point is a data warehouse containing a combination of internal data tracking all customer contact
coupled with external market data about competitor activity. Background information on potential customers also provides an excellent basis for prospecting. This warehouse can be implemented in a variety of relational database systems: Sybase, Oracle, Redbrick, and so on, and should be optimized for flexible and fast data access.

Fig. 1: Integrated Data Mining Architecture

An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user business model to be applied when navigating the data warehouse. The multidimensional structures allow the user to analyze the data as they want to view their business – summarizing by product line, region, and other key perspectives of their business. The Data Mining Server must be integrated with the data warehouse and the OLAP server to embed ROI-focused business analysis directly into this infrastructure. An advanced, process-centric metadata template defines the data mining objectives for specific business issues like campaign management, prospecting, and promotion optimization. Integration with the data warehouse enables operational decisions to be directly implemented and tracked. As the warehouse grows with new decisions and results, the organization can continually mine the best practices and apply them to future decisions.

This design represents a fundamental shift from conventional decision support systems. Rather than simply delivering data to the end user through query and reporting software, the Advanced Analysis Server applies users’ business models directly to the warehouse and returns a proactive analysis of the most relevant information. These results enhance the metadata in the OLAP Server by providing a dynamic metadata layer that represents a distilled view of the data. Reporting, visualization, and other analysis tools can then be applied to plan future actions and confirm the impact of those plans.

7. WHAT IS BUSINESS INTELLIGENCE?

Business intelligence (BI) refers to computer-based techniques used in spotting, digging-out, and analyzing business data, such as sales revenue by products and/or departments, or by associated costs and incomes.

BI technologies provide historical, current, and predictive views of business operations. Common functions of business intelligence technologies are reporting, online analytical processing, analytics, data mining, business performance management, benchmarking, text mining, and predictive analytics.

8. BI PRODUCT LANDSCAPE

Business intelligence is not necessarily about tools and technologies; rather it is a strategy of combining data from various sources with methodologies that make those facts solidify in a cohesive manner.

The data part of this strategy is data warehousing (see Figure 1). Once the data is sourced, scrubbed, enriched, conformed, and finally housed in “access-ready” formats BI tools can make the data sing and dance.

Fig. 2: Data Warehouse Architecture

9. CASE STUDY

The case examines the need for Business Intelligence (BI) systems to support organizational processes. It throws light on the advantages of tapping information stored in corporate databases for managing business processes. The case explores the need for creating a Business Intelligence (BI) environment and a framework for managing information. It discusses the use of Decision Support Systems (DSS) for improving the efficiency of operations of insurance firms. The case also examines the use of Online Analytical Processing Tools (OLAP) by insurance organizations.

Stanza Insurance, a leading insurer realized the need for leveraging its greatest asset - data. It was essential for Stanza to convert raw data into information so that users could access and use it easily. Stanza offered life insurance, health insurance and annuity products.

The claims settlement process in Stanza was quite tedious and time consuming. The time taken to settle a claim was much more than the industry average. In a recent review meeting of the claims department, it was decided that all activities in the claims settlement process would be restructured.

James Lewis, the Vice President was assigned the responsibility of designing the new process. Lewis approached Peter D'Souza (D'Souza), a consultant with Precision Consultancy to obtain his expert advice on process redesigning.
He told D’Souza, "Our claims settlement time is much more than the industry average and we want to reduce it significantly."

D’Souza replied, "The problem can be solved if you implement a Business Intelligence (BI) solution and decision support system. It will surely speed up claims processing activities."...

### 10. CONCLUSION

Data Mining is the extraction of hidden predictive information from large databases. This is a new powerful new technology with great potential to help companies focus on the most important information in data warehousing. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. "The automated, prospective analyses offered by data mining move beyond the analyzes of past events provided by retrospective tools typical of decision support systems" (Pilot).

Data mining is important to large systems because it finds things in large data repositories that you did not know existed. "A simple metaphor would be finding two needles in a haystack that match. The haystack is the database, the individual lengths of the hay represent your data fields, and the needles represent data fields with a relationship worth more to you than all the hay put together" (Newquist).

Whether or not a company has business intelligence capabilities can mean the difference between real success and mediocre performance. More and more business owners are now realizing the important role business intelligence plays in the success of their business. The company who can exploit its own data and information to gain insight and make smarter decisions will have a clear competitive advantage.

As business intelligence capabilities move to center stage as a top management priority, companies will need flexible IT solutions that will meet their business intelligence needs not only in the short-term but as their company grows and their need for more complex business intelligence capabilities increases.

### 11. FUTURE SCOPE

There is a concept called Business Intelligence 2.0 (BI 2.0) or the next generation of BI. The focus is on empowering users to freely access and share the information most relevant to their needs. There are, of course, a number of interpretations of what BI 2.0 is (or should be) but we (ThotWave) believe that BI should not only do what we’ve been promised to date (ubiquitous information at your fingertips) but also just in time (aka real-time BI) and BI should necessarily involve a collaboration component making BI part of the fact-based decisioning system where facts come from tacit and explicit data sources.

The goal of every entrepreneur should be to achieve beyond ordinary (business) men through foreknowledge, as Sun Tzu put it. A corporation’s foreknowledge comes from a comprehensive look at all the available data. This can only be accomplished through a strong technological strategy built around reliable tools.

Gaining an understanding of BI will move any company from simply delivering reports to enabling interactive analysis and information sharing. When BI is integrated into everyday business processes, corporations stay focused on what to achieve, not only what has been accomplished.

### REFERENCES