CHAPTER 9

APPLICATIONS OF DATA MINING

This section will be used to discuss some of the typical applications of data mining organizations. The uses or applications of data mining will be grouped into main sections.

General Data Mining Applications

This section highlights some of the applications of data mining in organizations in general typical business functions that these data mining applications fall under are:

(i) **E-commerce and E-business Functions**

Data mining can help organizations to identify the information that would be most suitable to put on the Web by conducting an analysis of competitors and potential customers, and determining the expertise of the organization (Thuraisingham, 2003:276) [23] In order for any organization to get onto the World Wide Web, it needs to decide what information to put on its web site. Data mining can also be used to assist organizations waning to conduct e-business with other organizations by providing information with regard to selecting the best partners, identifying competitors and determining the best pricing policies for its products of services. Summarises the e-commerce and e-business applications of data mining as being the following:
• Selecting partner organization.
• Analyzing customer profiles.
• Determining which products to market online.
• Identifying web pages that are viewed together.
• Assessment of the similarity of web page contents.
• Categorization of web pages based on content.
• One of the most useful applications of data mining in an e-commerce environment can be found on the Amazon.com web site.

(ii) Business Intelligence Functions
In order to be able to run a business effectively, management needs intelligence relating to competitors, customer, partners and employees, as well as intelligence relating to market conditions, future trends, government policies and more lists the applications of data mining in business intelligence as follows:

• Analyzing customer profiles.
• Determining you won business strategies.

(iii) Customer Relationship Management Functions
The CRM is a key application of business intelligence and deals with mining information about customers from public as well as private databases in order to build customer profiles.
Typical application of data mining in customer relationship management are therefore:

- Building and analyzing customer profiles.
- Developing customer-specific products.

(iv) **Marketing and Sales Functions**

Marketing and sales were two of the early business functions that drove the development of the data mining. The following are some of the data mining application relevant to these functions:

- Performing targeted marketing.
- Market segmentation.
- Online sales support.
- Analyzing or predicting customer reaction to promotions.
- Marketing basket analysis.

Data mining has several applications in the manufacturing and planning functions of organizations. Data mining has several applications that are relevant to education and training. Developing courses and schedules based on the needs identified by data mining.

(v) **Financial Services Industry**

The major benefit offered by data mining to the financial services industry related to the industry’s dominant problem category of trying to predict the future.
The financial services industry can benefit a great deal from data mining, especially with regard to the major role that Business Intelligence and Customer Relationship Management play in this industry. Typical data mining applications in the financial services industry include the following:

- Analyzing stock market quotes.
- Portfolio management.
- Identifying suspicious transactions.
- Risk management

Customers needs and adapt their business process in the value chain to effectively meet those needs. The insurance industry is totally dependent on the ability to convert raw data into intelligence about customers, markets, competitors and the business environment.

(vii) Customer Relationship Management

Typical insurance companies have huge customer bases, varied product lines, may distribution channels and market spreads across geographies. The following applications of data mining that can assist them in doing this:

Customer profitability

The insight gained from customer profitability analysis can also lead to product customization and new product development.
Customer Lifetime Value

Data mining tools can assist in modeling this by taking a large number of actors relating to the customer into consideration. Some of the more general marketing and sales based data mining applications, such as customers segmentation, attrition analysis, market basket analysis, target marketing, campaign analysis and cross selling can also be used with great effect to improve an insurance company’s customer relationship management efforts.

- Channel Analysis.
- Loss Analysis
- Claims Analysis.

E-commerce

Business-to-consumer and business-to-business commerce conducted by way of the Internet or other electronic networks. In addition to standard forms for business-to-business transactions, e-commerce encompasses much wider activity - for example, the deployment of secure private networks (intranets) for sharing information within a company, as well as selective extensions of a company’s intranet to collaborating business networks (extranets).
Electronic Commerce

Electronic Commerce, or e-commerce, is the conduct of business by electronic means. In the twenty-first century, e-commerce referred more specifically to transactions between businesses (B2B e-commerce) and between businesses and consumers (B2C e-commerce) through the use of computer communication, particularly the Internet. This form of electronic commerce began in 1968, when what was called Electronic Data Interchange permitted companies to carry out electronic transactions. The business-to-business form of e-commerce fared better in 2000 and 2001, although a faltering economy lowered expectations. B2B e-commerce evolved with the development of the Internet. Other market-creating firms focus on specific products. Businesses remained reluctant to guarantee strict privacy protection because selling information about customers was a valuable part of the e-commerce business. E-commerce, commerce conducted over the Internet, most often via the World Wide Web. E-commerce can apply to purchases made through the Web or to business-to-business activities such as inventory transfers. A customer can order items from a vendor's Web site, paying with a credit card (the customer enters account information via the computer) or with a previously established "cyber cash" account. Personal and account information is kept confidential through the use of "secured transactions" that use encryption technology (see data encryption).
Activities

In the dot com era, it came to include activities more precisely termed "Web commerce" -- the purchase of goods and services over the World Wide Web, usually with secure connections (HTTPS, a special server protocol that encrypts confidential ordering data for customer protection) with e-shopping carts and with electronic payment services, like credit card payment authorizations.

The ever growing dependence of modern industries on electronically enabled business processes gave impetus to the growth and development of supporting systems, including backend systems, applications and middleware. Examples are broadband and fibre-optic networks, supply-chain management software, customer relationship management software, inventory control systems and financial accounting software.

Web development

In many cases, an e-commerce company will survive not only based on its product, but by having a competent management team, good post-sales services, well-organized business structure, network infrastructure and a secured, well designed website. Following factors will make business of companies succeed in e-commerce:
Business failure is as much a reality in e-commerce as in any other form of business.

A good management team armed with information technology strategy. Providing an easy and secured way for customers to effect transactions. Providing reliability and security. Constructing a commercially sound business model.

Streamlining business processes, possibly through re-engineering and information technologies.

**Customer experience**

A successful e-commerce organization must also provide an enjoyable and rewarding experience to its customers. **Providing value to customers:** Vendors can achieve this by offering a product or product-line that attracts potential customers at a competitive price, as in non-Electronic commerce.

**Product suitability**

Virtual marketers can sell some non-digital products and services successfully. Consumers have accepted the e-commerce business model less readily than its proponents originally expected.
Concerns about security

The problem of access to web commerce, mainly for poor households and for developing countries. Electronic commerce relates to the usage of electronic communication networks to conduct business transactions. Increasing levels of integration of e-commerce systems into business has led to an increasing level of reliance on these systems. Inter organizational systems and globally distributed data means that ensuring the availability, integrity and confidentiality of the information these systems process is of paramount importance.

Advantages of E-Commerce

E-commerce possesses a variety of attributes that have made it attractive to businesses and to their customers. Implementing e-commerce systems has enabled organizations and their customers to conduct business in ways previously not possible.

E-commerce allows businesses to have market presence without physical presence, and thus less capital is required. Performing transactions electronically decreases the cost of creating, process, distributing, storing and retrieving paper-based information. E-commerce may incite business process re-engineering, as e-commerce systems change the way business processes interact. The open nature of e-commerce allows businesses to select from a greater variety of business
partners - the opportunity for businesses to build network-based ad hoc partnerships (Wang et al. 2000). For e-commerce to work, incentives for customers must exist as well. Greater information accessibility is enabled via e-commerce. Information that may have taken days to obtain is now available to customers immediately from a single computer terminal, enhancing the level of customer service. Furthermore, because the information is exchanged in electronic format, customers are able to integrate transaction processing into their own e-commerce systems, thereby automating the procurement process and reducing costs.

The advantages of paperless transactions are therefore applicable to customers.

Online marketplaces allow customers to place requests for tenders, and businesses to bid for them.

**E-Commerce Risks and Weaknesses**

As businesses and customers become increasingly reliant on their e-commerce systems, they should be aware of the risks they have increased exposure to. E-commerce must be regarded as a safe and practical way to do business, by both businesses and customers, to succeed. Udo's (2001) survey concludes that security forms a major barrier for the spread e-commerce with regards to consumers: "Security concern is one
of the main reasons Web users give for not purchasing over the Web."

For businesses, security is a major issue. Other security breaches, such as the Code Red worm, and various hacking attacks which have disclosed databases of customer credit card details (eg: Leyden 2000) all lower customer and business confidence in e-commerce. A central unit of e-commerce is the transaction. Transactions form the lifeblood of e-commerce in which information and money are traded amongst businesses and consumers. E-commerce has changed the face of most business functions in competitive enterprises. In general, e-commerce and ebusiness (henceforth referred to as e-commerce) have enabled on-line transactions. Also, generating large-scale real-time data has never been easier. With data pertaining to various views of business transactions being readily available, it is only apposite to seek the services of data mining to make (business) sense out of these data sets.

Data mining (DM) has as its dominant goal, the generation of non-obvious yet useful information for decision makers from very large databases.

**Ease of interfacing with legacy systems:**

It is commonplace to find large organizations run on several legacy systems that generate huge volumes of data. It must now be noted that e-commerce data, being the result of
on-line transactions, do satisfy all the above proper criteria for data mining. For instance, mining the web logs certainly enhances web server architecture-related decisions. Improved web server availability results in faster transactions, thus increasing the revenue. Data mining in e-commerce mostly relies on the controller for generating the data to mine on. In summary, it is little surprise that e-commerce is the killer application for data mining.

**A review of data-mining methods**

Given a truly massive amount of data, the challenge in data mining is to unearth hidden relationships among various attributes of data and between several snapshots of data over a period of time.

**Role of statistics in data mining**

Extracting causal information from data is often one of the principal goals of data mining and more generally of statistical inference. Statisticians have done aggregate data analyses on data for decades; thus DM has actually existed from the time large scale statistical modeling has been made possible.

The regression methods may be consider analogous to the association rules in data mining. We assume here that the database (here, the web logs) has transactions recorded on a
per-customer basis. • Use and reveal uncertainty and not hide it; some data-mining approaches ignore the causal relations due to lack of sufficient data.

**The role of AI in data mining**

Apart from learning relationships as above, neural networks are also useful in clustering data sets. Secondly, neural networks perform well with missing or incomplete data. Artificial intelligence based methods using neural networks are used in clustering and classification methods of data mining.

**The role of database research in data mining**

Data mining can certainly not be run on the transaction databases in their native state. It is to be observed that data warehouses are essentially snapshots of transactional data aggregated along various dimensions (including time, geographies, demographics, products etc.) In order to run data mining algorithms, it is common practice to use the data available in the data warehouse rather than by running real time scripts to fetch transactional data. Real-time data is not relevant for tactical decision making, which is where data mining is used. Data warehousing is nevertheless fraught with technological challenges.

Customers buy various products and each transaction records the products bought by the customer. com.
Concerning e-commerce itself, it may be noted that data warehousing may or may not be required depending on the application.

**DM in customer profiling**

Acquiring new customers, delighting and retaining existing customers, and predicting buyer behavior will improve the availability of products and services and hence the profits. At the most basic level, the information available in web log files can illuminate what prospective customers are seeking from a site. Rule-mining based systems could be used to propose such alternatives to the customers.

**DM in recommendation systems**

The event prediction system is based on association rule-mining and clustering algorithms. The PENS system is used to actively help an e-commerce service provider to forecast the demand of product categories better. Data mining has also been applied in detecting how customers may respond to promotional offers made by a credit card e-commerce company. A method to build customer profiles in e-commerce settings, based on product hierarchy for more effective personalization. They divide each customer profile into three parts: basic profile learned from customer demographic data; preference profile learned from behavioral data, and rule profile mainly referring to association rules.
DM in web personalization

A comprehensive overview of the personalization process based on web usage mining. The goal of this paper is to show how pattern discovery techniques such as dustedng, association rule-mining, and sequential pattern discovery, performed on web usage data, can be leveraged effectively as an integrated part of a web personalization system. The author observes that the log data collected automatically by the Web and application servers represent the fine-grained navigational behavior of visitors.

A Data to be captured by web logs

Depending on the goals of the analysis, e-commerce data need to be transformed and aggregated at different levels of abstraction. e-Commerce data are also further classified as usage data, content data, structure data, and user data. Usage data contain details of user sessions and page views. Structure data represent the designer's view of the content organization within the site. Structure data for a site are normally captured by an automatically generated site map which represents the hyperlink structure of the site. The operational database(s) for the site may include additional user profile information. Once the data types are clear, data preparation is easily achieved by processes such as data cleansing, page view identification, user identification, session
identification, the inference of missing references due to caching, and transaction (episode) identification. In most web usage mmmg tasks, the focus is generally on anonymous user navigational activity where the primary sources of data are server logs. This transaction-page view matrix can then be used to perform various data mining tasks.

**DM and buyer behavior in e-commerce**

For a successful e-commerce site, reducing user-perceived latency is the second most important quality after good site-navigation quality. The core of their approach involves extracting knowledge from integrated data of purchase and path traversal patterns of past users (obtainable from web server logs) to predict the purchase and traversal behaviour of future users.

Web sites are often used to establish a company's image, to promote and sell goods and to provide customer support. The success of a web site affects and reflects directly the success of the company in the electronic market.

**Enabling data collection in e-commerce**

It may be observed that there are various ways of procuring data relevant to ecommerce DM. Web server log files, web server plug-ins (instrumentation), TCP/IP packet sniffing, application server instrumentation are the primary
means of collecting data. It is quite common to expend about 80% of any DM effort in e-commerce in data filtering.

**Analyzing web transactions**

Once the data are collected via any of the above mentioned mechanisms, data analysis could follow suit. This could be done along session level attributes, customer attributes, product attributes and abstract attributes. In web-usage analysis, they have been proposed as the underlying modeling machinery for web pre-fetching applications or to minimize system latencies.

A new approach called on-line analytical mining for web data. Their approach consists of data capture, web house construction, and pattern discovery and pattern evaluation.

In a B2B e-commerce setting, it is very likely that vendors, customers and application service providers (ASP) (usually the middlemen) have varying DM requirements. The proposed distributed data mining system is intended for the ASP to provide generic data mining services to its subscribers. Heterogeneity implies that the system can mine data from heterogeneous and distributed locations. Maintaining security implies that in some instances, the user might be mining highly sensitive data that should not leave the owner's site.
**Distributed spatial data mining**

In various e-commerce domains involving spatial data (real estate, environmental planning, precision agriculture), participating businesses may increase their economic returns using knowledge extracted from spatial databases. However, in practice, spatial data is often inherently distributed at multiple sites. In the proposed system, a centralized server collects proprietary site-specific spatial data from subscribed businesses as well as relevant data from public and commercial sources and integrates knowledge in order to provide valuable management information to subscribed customers. Spatial data mining software interfaces this database to extract interesting and novel knowledge from data. Specific objectives include a better understanding of spatial data, discovering relationships between spatial and non-spatial data, construction of spatial knowledge-bases, query optimization and data reorganization in spatial databases. Through a knowledge discovery (KDD) process, To propose learning algorithms that perform data modeling using data sets from different fields in possibly different regions and years.

**DM applied to retail e-commerce**

Kohavi et al (2004) have attempted a practical implementation of data mining in retail e-commerce data. (5)
Auditing of data procured for mining, from data warehouses, is mandatory. (6) Mining data at the right level of granularity is essential.

**Future Securing Web Services**

Web services allows applications (e.g. automated business transactions, stock trading and order-tracking systems) to communicate with each other within organizations, across enterprises, and across the Internet in a loosely-coupled, platform- and programming language-independent manner. Several key standards have formed the foundation for Web services: XML (Extensible Markup Language), WSDL (Web Services Definition Language), SOAP (Simple Object Access Protocol), and UDDI (Universal Description, Discovery, and Integration).

**What are Web Services?**

According to W3C, a Web service is defined as: "A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. A group of Web services interacting together in this manner defines a particular Web service application in a Service-Oriented Architecture (SOA)."
Web services exhibit the following definitive characteristics:

Web services communicate using platform-, hardware-independent and language-neutral Web protocols. These Web protocols ensure easy integration over the network & loosely coupling between applications.

A Web service provides an interface that can be called from another program. A Web service is registered and can be located through a Web Service Registry. The registry enables service consumers to find services that match their needs.

Advantages of Web Services

Flexibility - Web services allow loose-coupling, which means that interactions between service applications may not break even there is a change. Agility and Productivity - Rapid application assembly tools allow integration for new business opportunities or trying new business ideas.

Leverage Existing Investments - Web services provide existing or legacy software applications with service interfaces without changing the original applications, allowing them to fully operate in the service environment. This adapts existing applications to changing business conditions and customer needs.
**Web Services Model**

**A typical Web services model consists of three entities:**

1. Service providers who create Web services and publish them to the outside world by registering the services with service brokers.

2. Service brokers who maintain a registry of Published services.

3. Service requesters who find required services by searching the service broker's registry. Requesters then bind their applications to the service provider to use particular services.

4. As more and more business functions are exposed as Web services, the number of participants in a Web services environment will be larger than what we have seen in other environments.

5. Confidentiality - the prevention of unauthorized disclosure of data.

**Web Services Security Schemes**

Web services security language can be defined into two types: computer security and communications security.
Computer security is a node-oriented security focus and it is essentially access control within a computer system. A requirements rule expresses the necessary security-relevant preparations for the use of a service, or security measures needed after the service execution. Non-repudiation is critical for business Web services.

- **XML digital signature** - XML digital signature provides authentication, data integrity and non-repudiation. It is to develop XML syntax for representing digital signatures over any data type.

- **XML encryption** - Its goal is to develop XML syntax for representing encrypted data and to establish procedures for encrypting and decrypting such data.

- **XKMS (XML Key Management Specification)** - XKMS consists of two parts: XKISS (XML Key Information Service Specification) and XKRSS (XML Key Registration Service Specification).

- **WS- Security (Web Services Security)** - It defines a set of SOAP header extensions for end-to-end SOAP messaging Security. It supports message integrity and confidentiality by allowing communicating partners to exchange signed and encrypted messages in a Web services environment.

- **ebXML Message Service** - The ebXML initiative is a set of next-generation XML-based standards enabling electronic business transactions via the Internet.
Identity Management

If there is lack of a well-integrated and interoperable identity management architecture, this makes managing Web properties, applications, identities, and policies non-scalable, and effectively prohibits the interaction of identities across applications or Web services. There are two identity management architectures, centralized model and federated model. Furthermore, organizations can maintain their own customer data while sharing identity data with partners based on their business objectives and customer preferences.