5

Application of Data-Mining in an Enterprise System

By

OGOCHUKWU C. OKEKE
Department of Computer Science,
Anambra State University,
Uli.

IKE J. MGBEAFULIKE
Department of Computer Science
Anambra State University,
Uli.

IFEANYI C. EJIOFOR
Department of Computer Science,
Anambra State University,
Uli.

And

PETER C. EZEANYEJI
Department of Computer Science,
Anambra State University,
Uli.

Abstract

Over recent years, data-mining has been establishing itself as one of the major disciplines in computer science with growing industrial impact. Undoubtedly, research in data- mining will continue and even increase over coming decades. In this article, vision of the future of data- mining is stated. Before an enterprise is developed, a consideration is given to how it will help the stake holders to achieve their strategic business objectives. The focus of this research has been aimed at determining effective strategies for successful enterprise development through the use of data-mining. A framework for transformation that consists of a set of interdependent methodologies, tools and enterprise principles that support holistic enterprise transformation is described in this paper. This research work focuses on the use of data-mining

in the development of enterprise systems. Data-mining helps to predict hidden patterns in large scale enterprise systems, which would help in taking good decision and business process redesign.

Data-mining is the process of analysing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data-mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases

Data-mining techniques usually fall into two categories namely: predictive and descriptive. Predictive data-mining uses historical data to infer something about future events while descriptive data-mining aims to find patterns in the data that provide some information about what the data contains.

Data-mining can be used for several purposes by different people and organizations. The most notable users of data-mining come from commercial, scientific or government backgrounds (Clifton ,2010).

Commercial entities may use the information gathered through data mining techniques to help discover something about their consumers, to help market their products better. Data mining is also used by search engines such as Google to mine web pages for information relating to a specific search query (Clifton ,2010).

Scientific communities may benefit from data-mining by using it to find anomalies, clusters or co-locations to name a few. For example, they could discover a relationship between people getting cancer and the location of a chemical plant (McGrail & Gulski ,2010).

The government could use data-mining techniques to uncover patterns in their data. For example, data-mining is used to find unusual patterns in the stock market in order to detect insider trading. It is also used to detect scams sent by email. It could also be used to find unusual behavior to prevent a terrorist attack. (http:// en .wikipedia. org/wiki/data-mining)

There are many more applications of data-mining, which are continually being expanded. The main requirement for performing data-mining is suitable data (Vladimir , 2010). The purpose of this research work is to determine how data-mining can effectively be used to develop large-scale, integrated application-software packages that use the computational, data storage, and data transmission power of data mining to support business processes, information flows, reporting, and data analytics within and between complex organization Before an enterprise is developed, a consideration is

given to how it will help the stake holders to achieve their strategic business objectives. The purpose of this research has been aimed at determining effective strategies for successful enterprise development through the use of data-mining. If the enterprise system had already been developed, it can be totally transformed with the power of data-mining. A framework for transformation that consists of a set of interdependent methodologies, tools and enterprise principles that support holistic enterprise transformation is described in this paper.

Enterprise Systems

Enterprise Systems (ES) are large-scale, integrated application-software packages that use the computational, data storage, and data transmission power of modern information technology to support business processes, information flows, reporting, and data analyses within and between complex organizations. ES is a special class of enterprise application software—namely, packaged enterprise application software). Computer-based systems built using ES are types of Enterprise Information Systems, or Management Information Systems, which, in turn, are types of information systems citation needed (Wagner,2006).

An Enterprise Information System is generally any kind of computing system that is of "enterprise class". This means typically offering high quality of service, dealing with large volumes of data and capable of supporting some large organization ("an enterprise"). Enterprise Information Systems provide a technology platform that enables organizations to integrate and coordinate their business processes. They provide a single system that is central to the organization and ensure that information can be shared across all functional levels and management hierarchies. Enterprise systems are invaluable in eliminating the problem of information fragmentation caused by multiple information systems in an organization by creating a standard data structure. A typical Enterprise Information System would be housed in one or more data centres, run enterprise software, and could include applications that typically cross organizational borders such as content management systems (Wikipedia.org 2011).

Data Mining Technology in Enterprise System Development

Data-Mining is largely used in several applications such as enterprise systems, understanding consumer research marketing, product analysis, demand and supply analysis, e-commerce, investment trend in stocks & real estates, telecommunications and so on. Data-mining is based on mathematical algorithm and analytical skills to drive the desired results from the huge database collection.

Data-mining has great importance in today’s highly competitive business environment. A new concept of Business Intelligence data mining has evolved now, and is widely used by leading corporate houses to stay ahead of their competitors.

Business Intelligence (BI) can help in providing latest information and used for competition analysis, market research, economic trends, consumer behaviour, industry research, geographical information analysis and so on (Kantardzic, 2003). Business Intelligence data-mining helps in decision-making.

It's applications are widely used in development of enterprise systems direct marketing, health industry, e-commerce, customer relationship management (CRM), Fast Moving Consumer Goods (FMCG) industry, telecommunication industry and financial sector. Data-mining is available in various forms like text mining, web mining, audio & video data mining, pictorial data-mining, relational databases, and social networks data mining (Andrea , 2010).

Advantages of Data-mining

Marking/Retailing: Data-mining can aid direct marketers by providing them with useful and accurate trends about their customers' purchasing behaviour (Clifton ,2010). Based on these trends, marketers can direct their marketing attentions to their customers with more precision. For example, marketers of a software company may advertise about their new software to consumers who have a lot of software purchasing history. In addition, data mining may also help marketers in predicting which products their customers may be interested in buying. Through this prediction, marketers can surprise their customers and make the customer's shopping experience becomes a pleasant on. Retail stores can also benefit from data mining in similar ways. For example, through the trends provided by data mining, the store managers can arrange shelves, stock certain items, or provide a certain discount that will attract their customers (Clifton, 2010).

Banking/Crediting: Data-mining can assist financial institutions in areas such as credit reporting and loan information. For example, by examining previous customers with similar attributes, a bank can estimate the level of risk associated with each given loan. In addition, data mining can also assist credit card issuers in detecting potentially fraudulent credit card transaction. Although the data mining technique is not a 100% accurate in its prediction about fraudulent charges, it does help the credit card issuers reduce their losses (Clifton ,2010).

Law Enforcement: Data-mining can aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviours (Clifton ,2010).

Researchers: It can assist researchers by speeding up their data analysing process thus, allowing them more time to work on other projects. Enterprises are faced with a challenge when rapid growth and aging technology threatened to stand in the way of

their goals. Instead of enabling operational improvements, the existing systems were impeding progress. Existing technology was causing missed deliveries and creating a large number of back orders. Inventory control was poor, and the planning was inaccurate. With some customers expecting shipment in as long as nine months and others expecting shipment in as little as nine days or even less, more sophisticated and accurate planning was critical. Customer satisfaction was at risk and internal morale was slipping. If appropriate data mining techniques are not used to develop the enterprise, the established reputation for high-quality, high-performance products was also at risk within a short period of their existence (Davenport, 1998).

Objectives of Data-mining

Some objectives are associated with data mining like controlling of Growing Data Volume. The main reason for necessity of automated computer systems for intelligent data analysis is the enormous volume of existing and newly appearing data that require processing. The amount of data accumulated each day by various business, scientific, and governmental organizations around the world is daunting. According to information from GTE research centre, only scientific organizations store each day about 1 TB (terabyte!) of new information. And it is well known that academic world is by far not the leading supplier of new data. It becomes impossible for human analysts to cope with such overwhelming amounts of data (Superby ,2006).

Human analysis limitation, two problems that surface when human analysts process data are the inadequacy of the human brain when searching for complex multifactor dependencies in data and the lack of objectiveness in such an analysis. A human expert is always a hostage of the previous experience of investigating other systems. Sometimes, this helps, sometimes this hurts, but it is almost impossible to get rid of this fact.

Effecting Low Cost of Machine Learning

One additional benefit of using automated data mining systems is that this process has a much lower cost than hiring an army of highly trained (and paid) professional statisticians(Wagner ,2006). While data-mining does not eliminate human participation in solving the task completely, it significantly simplifies the job and allows an analyst who is not a professional in statistics and programming to manage the process of extracting knowledge from data. Businesses employing data-mining may see a return on investment, but also they recognize that the number of predictive models can quickly become very large. Rather than one model to predict how many customers will churn, a business could build a separate model for each region and customer type. Then instead of sending an offer to all people that are likely to churn, it may only want to send offers to customers. Finally, it may want to determine which customers are going to be profitable over a window of time and only send the offers to those that are

likely to be profitable. In order to maintain this quantity of models, they need to manage model versions and move to automated data-mining (Monk & Wagner,2006).

Data-mining can also be helpful to human-resources departments in identifying the characteristics of their most successful employees. Information obtained, such as universities attended by highly successful employees, can help HR focus recruiting efforts accordingly. Additionally, Strategic Enterprise Management applications help a company translate corporate-level goals, such as profit and margin share targets, into operational decisions, such as production plans and workforce levels**.**

Another example of data-mining, often called the market basket analysis, relates to its use in retail sales. If a clothing store records the purchases of customers, a data-mining system could identify those customers who favour silk shirts over cotton ones. Although some explanations of relationships may be difficult, taking advantage of it is easier. The example deals with association rules within transaction-based data. Not all data are transaction based and logical or inexact rules may also be present within a database.

Market basket analysis has also been used to identify the purchase patterns of the Alpha consumer. Alpha Consumers are people that play a key role in connecting with the concept behind a product, then adopting that product, and finally validating it for the rest of society. Analysing the data collected on this type of user has allowed companies to predict future buying trends and forecast supply demands (Zhu & Ian , 2007).

Disadvantages of Enterprise Systems

Privacy Issues: Personal privacy has always been a major concern in this country. In recent years, with the widespread use of Internet, the concerns about privacy have increased tremendously. Because of the privacy issues, some people do not shop on Internet. They are afraid that somebody may have access to their personal information and then use that information in an unethical way; thus causing they harm.

Although it is against the law to sell or trade personal information between different organizations, selling personal information have occurred. For example, according to Washing Post, in 1998, CVS had sold their patient's prescription purchases to a different company. In addition, American Express also sold their customers' credit card purchases to another company. What CVS and American Express did clearly violate privacy law because they were selling personal information without the consent of their customers? The selling of personal information may also bring harm to these customers because you do not know what the other companies are planning to do with the personal information that they have purchased**.**

Security Issues: Although companies have a lot of personal information about us available online, they do not have sufficient security systems in place to protect that information. For example, recently the Ford Motor credit company had to inform 13,000 of the consumers that their personal information including Social Security number, address, account number and payment history were accessed by hackers who broke into a database belonging to the Experian credit reporting agency (Chen ,2006). This incidence illustrated that companies are not willing to disclose and share your personal information, but they are not taking care of the information properly. With so much personal information available, identity theft could become a real problem.

Misuse of information/inaccurate information: Trends obtain through data mining intended to be used for marketing purpose or for some other ethical purposes, may be misused. Unethical businesses or people may use the information obtained through data mining to take advantage of vulnerable people or discriminated against a certain group of people. In addition, data mining technique is not a 100 percent accurate; thus mistakes do happen which can have serious consequence (Chen ,2006).

Applications of Data-Mining in the Development of an Enterprise System

1 Measurement – program that creates a hierarchy of Performance metrics and Benchmarking (a standard test to measure the performance of computer hardware or software) that informs business leaders about progress towards business (http://en.wikipedia.org/wiki/data-mining).

A performance metric is a measure of an organization's activities and performance. Performance metrics should support a range of stakeholder needs from customers, shareholders to employees. While traditionally many metrics are financed based, inwardly focusing on the performance of the organization, metrics may also focus on the performance against customer requirements and value. In project management, performance metrics are used to assess the health of the project and consist of the measuring of six criteria: time, cost, resources, scope, quality, and actions http://en.wikipedia.org/wiki/data-mining.

Developing Performance Metrics Usually Follows a Process of

1. Establishing critical processes/customer requirements,
2. Developing measures,
3. Establishing targets which the results can be scored against.

A criticism of performance metrics is that when the value of information is computed using mathematical methods, it shows that even performance metrics professionals choose measures that have little value http://en.wikipedia.org/wiki/data-mining. This is referred to as the "measurement inversion". For example, metrics seem

to emphasize what organizations find immediately measurable — even if those are low value — and tend to ignore high value measurements simply because they seem harder to measure (whether they are or not).

To correct for the measurement inversion other methods, like applied information economics, introduce the "value of information analysis" step in the process so that metrics focus on high-value measures. Organizations where this has been applied find that they define completely different metrics than they otherwise would have and, often, fewer metrics. There are a variety of ways in which organizations may react to results. This may be to trigger specific activity relating to performance (i.e., an improvement plan) or to use the data merely for statistical information. Often closely tied in with outputs, performance metrics should usually encourage improvement, effectiveness and appropriate levels of control.

Performance metrics are often linked in with corporate strategy and are often derived in order to measure performance against a critical success factor.

2. Analytics – program that builds quantitative processes for a business to arrive at optimal decisions and to perform Business Knowledge Discovery. Frequently involves: data mining, process mining, statistical analysis, Predictive analytics, Predictive modelling, and Business process modelling http:// en. wikipedia. org/ wiki/ data-mining**.**

Analytics is the application of computer technology, operational research, and statistics to solve problems in business and industry. Analytics is carried out within an information system: while, in the past, statistics and mathematics could be studied without computers and software, analytics has evolved from the application of computers to the analysis of data and this takes place within an information system or software environment. Mathematics underpins the algorithms used in analytics - the science of analytics is concerned with extracting useful properties of data using computable functions , and typically will involve extracting properties from large data bases. Analytics therefore bridges the disciplines of computer science, statistics, and mathematics.

A simple definition of analytics is "the science of analysis". A practical definition, however, would be that analytics is the process of obtaining an optimal or realistic decision based on existing data (http://wikipedia.org/wiki/data-mining). Business managers may choose to make decisions based on past experiences or rules of thumb, or there might be other qualitative aspects to decision making; but unless there are data involved in the process, it would not be considered analytics.

Common applications of analytics include the study of business data using statistical analysis in order to discover and understand historical patterns with an eye to

predicting and improving business performance in the future. Also, some people use the term to denote the use of mathematics in business. Others hold that the field of analytics includes the use of Operations Research, Statistics and Probability. However, it would be erroneous to limit the field of analytics to only statistics and mathematics.
Analytics closely resembles statistical analysis and data mining, but tends to be based on modelling involving extensive computation. Some fields within the area of analytics are enterprise decision management, marketing analytics, predictive science, strategy science, credit risk analysis and fraud analytics.

3. Reporting/Enterprise Reporting – program that builds infrastructure for Strategic Reporting to serve the Strategic management of a business, NOT Operational Reporting. Frequently involves: Data visualization, Executive information system, OLAP (Davenport ,1998).
A report is a talkative work (usually of writing, speech, television, or film) made with the specific intention of relaying information or recounting certain events in a widely presentable form.

Written reports are documents which present focused, salient content to a specific audience. Reports are often used to display the result of an experiment, investigation, or inquiry. The audience may be public or private, an individual or the public in general. Reports are used in government, business, education, science, and other fields.

Reports often use persuasive elements, such as graphics, images, voice, or specialized vocabulary in order to persuade that specific audience to undertake an action. One of the most common formats for presenting reports is IMRAD: Introduction, Methods, Results and Discussion. This structure is standard for the genre because it mirrors the traditional publication of scientific research and summons the ethos and credibility of that discipline. Reports are not required to follow this pattern, and may use alternative patterns like the problem-solution format.

Additional elements often used to persuade readers include: headings to indicate topics, to more complex formats including charts, tables, figures, and pictures, tables of contents, abstracts, summaries, appendices, footnotes, hyperlinks, and references.
Some examples of reports are: scientific reports, recommendation reports, white papers, annual reports, auditor's reports, workplace reports, census reports, trip reports, progress reports, investigative reports, budget reports, policy reports, demographic reports, credit reports, appraisal reports, inspection reports, military reports, bound reports, etc.

Enterprise/Client reporting: With the dramatic expansion of information technology, and the desire for increased competitiveness in corporations, there has been an increase in the use of computing power to produce unified reports which join different views of the enterprise in one place. Termed Enterprise Reporting, this process involves querying data sources with different logical models to produce a human readable report—for example; a computer user has to query the Human Resources databases and the Capital Improvements databases to show how efficiently space is being used across an entire corporation.

Enterprise Reporting is a fundamental part of the larger movement towards improved business intelligence and knowledge management. Often implementation involves extract, transform, and load (ETL) procedures in coordination with a data warehouse and then using one or more reporting tools. While reports can be distributed in print form or via email, they are typically accessed via a corporate intranet.

4. Collaboration/Collaboration Platform – program that gets different areas (both inside and outside the business) to work together through Data sharing and Electronic Data Interchange.

An emerging category of computer software, collaboration platforms are unified electronic platforms that support synchronous and asynchronous communication through a variety of devices and channels. Collaboration platforms offer a set of software components and services that enable individuals to find each other and the information they need and to be able to communicate and work together to achieve common business goals. The core elements of a collaboration platform are messaging (email, calendaring and scheduling, and contacts), team collaboration (file synchronization, ideas and notes in a wiki, task management, full-text search), and real-time collaboration and communication (e.g., presence, instant messaging, web conferencing, application or desktop sharing, voice, audio and video conferencing), and Social Computing tools (blog, wiki, tagging, RSS, shared bookmarks).

Collaboration platforms could be proprietary or open source or free software, and used in wider information and communication environments.

5. Knowledge Management

Knowledge Management (KM) comprises a range of strategies and practices used in an organization to identify, create, represent, distribute, and enable adoption of insights and experiences. Such insights and experiences comprise knowledge, either embodied in individuals or embedded in organizational processes or practice. Knowledge Management leads to Learning Management and Regulatory compliance/Compliance

An established discipline since 1991 (Nonaka 1991), KM includes courses taught in the fields of business administration, information systems management, and

library and information sciences (Alavi & Leidner 1999). More recently, other fields have started contributing to KM research. These include information and media, computer science, public health, and public policy.

Many large companies and non-profit organizations have resources dedicated to internal KM efforts, often as a part of their 'business strategy', 'information technology', or 'human resource management' departments (Addicott, McGivern & Ferlie 2006). Several consulting companies also exist that provide strategy and advice regarding KM to these organizations.

Knowledge Management efforts typically focus on organizational objectives such as improved performance, competitive advantage, innovation, the sharing of lessons learned, integration and continuous improvement of the organization. KM efforts overlap with organizational learning and may be distinguished from that by a greater focus on the management of knowledge as a strategic asset and a focus on encouraging the sharing of knowledge.

Future of Data- Mining and Enterprise Systems

People asked questions on the future of data-mining with particular emphasis revolving around whether or not data-mining will occur inside the database or external to the database. Today we have reasons to say that mining will move more inside the database. This has not materialized nearly to the degree expected, though it has progressed especially in the past couple of years with improvements to Oracle Data Miner and Structured Query Language (SQL) Server 2008 Business Intelligence. We are not familiar with the current state of DB2 Data Warehouse Edition and we do not think there has been much work done in recent years on the Teradata Warehouse Miner product.

However, most people who work with data-mining still pull data from a data-mart or warehouse, build models in a standalone application, and then push models and/or scores back up to the warehouse. We think this is going to move more and more into the warehouse either through improved software in the warehouse (Oracle and Microsoft), or, perhaps more likely, through improved interfaces to warehouse functions by standalone data mining software. For example, Clementine from SPSS allows pushback database function to the database itself rather than operating on data that has been pulled from the warehouse. This speeds up basic data processing considerably. The latter is the more likely area of growth in data-mining software and how practitioners use data-mining software.

Enterprise systems will continue to offer high quality service, dealing with large volumes of data and capable of supporting some large organizations**.**

Conclusion

Data-mining can effectively be used to develop enterprise systems which are large-scale, integrated application-software packages that use the computational, data storage, and data transmission. The power of data-mining to support business processes, information flows, reporting, and data analyses within and between complex organizations has been possible. Before an enterprise is developed a consideration is given to how it will help the stake holders of the enterprise to achieve their strategic business objectives. Data-mining can be used in determining effective strategies for successful enterprise development. If the enterprise system had already been developed, it can be totally transformed with the power of data-mining. A framework for transformation that consists of a set of interdependent methodologies, tools and enterprise principles that support holistic enterprise transformation has been exposed.

References

Baker, R. (2007).Educational *Data-mining*. New York :Hershey Press.

Battiti,B.R.&Passerini,A.(2010).Brain-ComputerEvolutionaryMulti-Objective Optimization (BC-EMO).Michigan:Gale Research.

Clifton,C.(2010).*EncyclopediaBritannica:DefinitionofData-mining*.NewYork: Holt, Rinehart and Winston.

Davenport,T.H(1998).*Putting the Enterprise Into The Enterprise System*.USA:Harvard.

Monk,E&Wagner,B(2006).*Concepts In Enterprise Resource Planning*(2nd Ed).Boston: Holt Press.

Fayyad,U,Shapiro,G.P &Smyth,P.(1996).*From Data-mining to Knowledge Discovery in Databases*.New York:McGraw Hill Kogakusha.

Mehmed,K.(2003).*Data-mining:Concepts,Models,and Algorithm*.New York: John Wiley&Sons.

Fountain,T,Dietterich,T&Sudyka,B.(2000).Mining IC Test Data to Optimize VLSI . Testing, in *Proceedings of the sixth ACM SIGKDD International Conference On Knowledge Discovery&Data-mining*.New York:ACM Press.

Zhu,X.&Lan,D(2007).*Knowlegde Discovery and Data-mining:Challenges and Realities*. New York:Hershey Press.