International Journal of Soft Computing 6 (4): 136-142, 2011
ISSN: 1816-9503

Data Mining Applications in Present Scenario: A Review

[1]Harshita Patel and [2]D.S. Rajput
[1]Department of Bio-Informatics, [2]Department of Computer Applications,
M.A.N.I.T., 462051 Bhopal, India

Abstract: The fast technological era facilitated with various computational devices leads to the exponential growth of data, need proper treatment; well known data warehousing and data mining terms are directly imply to produce informational processing of these huge datasets. Apart from storing and updating routine data, data mining provide the extraction of valuable knowledge, actually needed for decision making. From scientific data to organizational details, sports to banking data or library to security concerns every field dealing with large amount of data require data mining tools to dig strategic information for their respective requirements. In this study, researchers will study some data mining application areas.

Key words: Data warehousing, data mining, knowledge extraction, data mining applications, datasets, India

INTRODUCTION

In last two or three decades the name of computer science department in an enterprise change from data processing to management information system then to information system and presently it is known as information technology (Ponniah, 2001). Here the change is not limited to the name of department it become visible in working scenario, amount of data and technology used. As in today's new technical era technologies (Either software or hardware) are fast growing, amount of organizational data is also increasing. So, by the time problem of proper storage and easy access become important. Data warehouses are then used for storage and data mining as a tool of finding hidden information from large amount of data. Data mining is being used in almost enterprises of different fields including both private and public sectors having the need to extract useful information from data repository. These are all types of business, various scientific areas, whole web, sports, education and many more. Business and industries like banking, retail, insurance, medicine, etc. have the need to minimize the cost and maximize the profit including new findings all the time essentially require data mining. In public sector it is needed to detect fraud as well as to get strategies for proper execution of program.

DATA WAREHOUSE

Retrieval of data from any data repository is not enough to be called data mining. User either information technology professional or just an unskilled end-user first has to seek toward a well defined data storage which contains integrated, easily accessible, subject oriented and timely oriented data known as data warehouse. As in organization data is collected from various sources, it first to be clean and transformed before use. A single, complete and consistent store of data obtained from a variety of different sources made available to end users in what they can understand and use in a business context (Barry Devlin). A data warehouse is a subject-oriented, integrated, time-varying and non-volatile collection of data that is used primarily in organizational decision making.

Data warehouse is an informational environment that provides an integrated and total view of the enterprise makes the enterprise's current and historical information available for decision making makes the organizational information consistent and present a flexible and interactive source of strategic information. The data warehouse exist to answer user's questions about the business, the performance of the various operations, the business trends and about what can be done to improve the business. The data warehouse exist to provide business users with direct access to data, to provide a single unified version of the performance indicators, to record the past accurately and to provide the ability to view the data from many different perspectives (Ponniah, 2001).

DATA MINING

Data Mining is the discovery of hidden information found in databases and can be viewed as a step in the knowledge discovery process (Chen *et al*., 1996; Fayyad *et al*., 1996). The rapid development of computer

Corresponding Author: Harshita Patel, Department of Bio-Informatics, M.A.N.I.T., 462051 Bhopal, India

technology, especially increased capacities and decreased costs of storage media has lead to store huge amounts of external and internal information in large databases at low cost. Mining useful information and helpful knowledge from these large databases has thus evolved into an important research area (Chen *et al*., 1996; Margahny and Mitwaly, 2005; Agrawal *et al*., 1993; Wang *et al*., 2002). Data mining is the process of extracting valuable information from large amounts of data (Hand *et al*., 2001). Data mining uses the data itself to uncover relationships and patterns. In doing so, hidden relationships, patterns and interdependencies can be discovered, predictive rules can be generated and interesting hypotheses can be found. These are the advantages of data mining (Hedberg, 1995; Gargano and Ragged, 1999). Description of what data mining does is: Discover useful, previously unknown knowledge by analyzing large and complex data sets (Jensen, 2003). Data mining is one step in a broader knowledge-discovery process.

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 much time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations (http://www.thearling.com/text/dmwhite/dmwhite.htm). Data mining also known as knowledge discovery in databases (Fayyad and Uthurusamy, 1996) is an interdisciplinary area that encompasses techniques from a number of fields including information technology, statistical analyses and mathematic science (Bohen *et al*., 2003).

A major function of data mining is to help analyze, understand or even visualize the huge amount of data stored in databases, data warehouses or other information repositories (Li and Shue, 2004).

Various data mining functionalities of data mining are characterization, discrimination, association, classification, prediction, clustering, outlier analysis, evolution and deviation analysis and many more.

Characterization: Data characterization is a summarization of general features of objects in a target class and produces what is called characteristic rules. The data relevant to a user-specified class are normally retrieved by a database query and run through a summarization module to extract the essence of the data at different levels of abstractions.

Discrimination: Data discrimination produces what are called discriminant rules and is basically the comparison of the general features of objects between two classes referred to as the target class and the contrasting class.

Association analysis: Association analysis is the discovery of what are commonly called association rules. Association rules, first introduced in 1993 (Agrawal *et al*., 1993) are used to identify relationships among a set of items in a database. These relationships are not based on inherent properties of the data themselves (as with functional dependencies) but rather based on co-occurrence of the data items. It studies the frequency of items occurring together in transactional databases and based on a threshold called support, identifies the frequent item sets. Another threshold, confidence which is the conditional probability than an item appears in a transaction when another item appears is used to pinpoint association rules. Association analysis is commonly used for market basket analysis. This approach is useful when one has an idea of the different associations that are sought. This is because one can find all kinds of correlations in a large dataset. Cunningham and Frank (1999) applied the association rules to the task of detecting subject categories that co-occur in transaction records of books borrowed from a university library. Use of Apriori algorithm, its set of procedure and data structure is and foot step in association rule mining (Park *et al*., 1995; Brin *et al*., 1997; Toivonen, 1996; Savasere *et al*., 1995).

Classification: Classification analysis is the organization of data in given classes. The goal of classification is to construct a classification procedure from a set of cases whose classes are known so that such a procedure can be used to classify new cases (Liu and Kellam, 2003). In other words, classification can be used both to understand existing data and to predict how new cases will behave. For example, Ng *et al*. (2003), Mateos *et al*. (2002) and (Tukey, 1977) used classification to infer the functions of genes.

Also known as supervised classification, the classification uses given class labels to order the objects in the data collection. Classification approaches normally use a training set where all objects are already associated with known class labels. The classification algorithm learns from the training set and builds a model. The model is used to classify new objects.

Prediction: Prediction has attracted considerable attention given the potential implications of successful forecasting in a business context. There are two major

types of predictions: One can either try to predict some unavailable data values or pending trends or predict a class label for some data. The latter is tied to classification. Once a classification model is built based on a training set, the class label of an object can be foreseen based on the attribute values of the object and the attribute values of the classes.

Clustering: Similar to classification, clustering is the organization of data in classes. However, unlike classification in clustering, class labels are unknown and it is up to the clustering algorithm to discover acceptable classes. Clustering is also called unsupervised classification because the classification is not dictated by given class labels. There are many clustering approaches all based on the principle of maximizing the similarity between objects in a same class (Intra-class similarity) and minimizing the similarity between objects of different classes (Inter-class similarity).

In other words clustering, a major Exploratory Data Analysis method (EDA) (Roussinov and Zhao, 2003) is concerned with the division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar among themselves and dissimilar to objects of other groups (Chen *et al*., 1996). This approach has the advantages of uncovering unanticipated trends, correlations or patterns and no assumptions are made about the structure of the data.

Outlier analysis: Outliers are data elements that cannot be grouped in a given class or cluster. Also known as exceptions or surprises, they are often very important to identify. While outliers can be considered noise and discarded in some applications, they can reveal important knowledge in other domains and thus can be very significant and their analysis valuable.

Evolution and deviation analysis: Evolution and deviation analysis pertain to the study of time related data that changes in time. Evolution analysis models evolutionary trends in data which consent to characterizing, comparing, classifying or clustering of time related data. Deviation analysis, on the other hand, considers differences between measured values and expected values and attempts to find the cause of the deviations from the anticipated values (Chen *et al*., 1996; Fayyad and Uthurusamy, 1996; Han and Kamber, 2000; Frawley *et al*., 1991; Platetsky-Shapiro and Frawley, 1991). Other than these basic functionalities we have lot of resources in data mining to actually find or mine something like hierarchical association rules (Savasere *et al*., 1995; Han and Fu, 1995; Fu, 1996), association rules maintenance (Cheung *et al*., 1996, 1997; Thomas *et al*., 1997; Ayan *et al*., 1999), sequential pattern mining (Agrawal and Srikant, 1995; Srikant and Agrawal, 1996), episode mining (Mannila *et al*., 1995) and functional dependency discovery (Huhtala *et al*., 1999, 1998). Various algorithms on each of above functionalities, pattern mining, traversal and many more.

APPLICATION AREAS OF DATA MINING

A wide range of organizations have deployed successful applications of data mining. The applications of data mining includes: banking industries, credit card companies, financial industries, telecommunications industries, health care industry, transportation companies, turnover management, food-service menu analysis, fraud detection, package goods companies, government policy settings, hiring profiles, medical management, process control, quality control, store management, student recruiting and retention, warranty analysis, etc. (Setty *et al*., 2010).

Use of software forms in place of manual documents or digitalization of many things is very beneficial as the E-library or digital library provide the facility of number of world class books at a place with negligible physical storage space. Researchers know very well the important of books and study materials but because of very large quantity, finding of essential ones is arise need of data mining. The advent of electronic resources and their increased use in libraries has brought about significant changes in library (Jadhav and Kumbargoudar, 2007). The digital library collects, stores, preserves and retrieves the digital data. The data and information are available in the different formats. These formats include text, images, video, audio, picture, maps, etc. which makes digital library as a suitable platform for data mining application. Data mining techniques can be applied on sports to find game strategies. The data mining techniques are used to determine the best or the most optimal group to represent a team in a team sport in a season or tour (Chodavarapu). In the sports world the vast amounts of statistics are collected for each player, team, game and season. Data mining can be used by sports organizations in the form of statistical analysis, pattern discovery as well as outcome prediction. Patterns in the data are often helpful in the forecast of future events. Data mining can be used for scouting, prediction of performance, selection of players, coaching and training and for the strategy planning (Solieman, 2006).

Data mining techniques are also applying in terrorist-activities detection on web. Data mining is one technique that has significant potential for use in countering

terrorism (De Rosa, 2004). As facility of web is available to all, terrorist are fast accessing the web for exchanging information and increase the quantity of new members and supporters (Lemos, 2002; Inmon, 1996). Because of this law enforcement agencies are putting major efforts to gather information from the Web about terror-related activities. It is a possible and positive belief that the detection of terrorists on the Web will be helpful to prevent further terrorist attacks (Kelley, 2002). But terrorist frequently change IP addresses and URLs so, it becomes difficult to monitor or trace their sites. The geographical locations of Web servers hosting those sites are also not fixed in order to prevent successful eavesdropping. Monitoring all ISPs traffic (Ingram, 2001) is a possible solution to the problem. Elovici *et al.* (2004) suggested a methodology using computer security (Intrusion Detection systems), information retrieval (The vector-space model) and data mining (Cluster analysis). Market analysis is important to every business to keep control over the changing market circumstances. Data mining tools with their in-depth analysis of business data help organizations in market analysis in terms of determining market shares and business potential, customer profile analysis, customer profitability analysis, market-basket analysis, determining methods to maximize distributor's efficiency and analyzing web markets. Even for electronic form of business (E-commerce), the integration of data mining improve the results and guide the user in generating knowledge and making correct business decisions (Ansari *et al*., 2001).

Another important domain of data mining is web mining can be defined simply as the application of data mining techniques to web data. With the explosive growth of information sources available on the world wide web, it has become increasingly necessary for users to utilize automated tools in order to find, extract, filter and evaluate the desired information and resources (Bin and Zhijing, 2003). Web mining is mainly for obtaining useful information and knowledge resource from a large number of web pages and it can be regarded as the data mining continuing to use in the web which can draw automatically, standardization and analyzing explaining the data (Maedche, 2000). Web mining is based on the network that is different from the traditional knowledge discovery in database (KDD) (Chakrabarti, 2002).

Technological advances are enabling scientists to collect vast amounts of data in fields such as medicine, remote sensing, astronomy and high-energy physics. These data arise not only from experiments and observations but also from computer simulations of complex phenomena. They are often complex with both spatial and temporal components. As a result, it has become impractical to manually explore, analyze and understand the data. In medical science there is large scope for application of data mining. Diagnosis of diesis, health care, patient profiling and history generation, etc. are the few examples. Computer-aided methods could assist medical staff and improve the accuracy of detection (Antonie *et al*., 2001).

Data mining tools have been used in telecommunication industry and the primary applications include telecommunication marketing, fraud detection, network fault isolation and prediction. The industry faces several drawbacks due to enormous size and sequential nature of data. Research can be directed to improve existing methods and develop new possibilities to overcome these challenges (Carbonara *et al*., 1997).

Bioinformatics is an interdisciplinary field. Bioinformatics can be considered as a mean to use information technology applied to the management and analysis of biological data. This has implications in diverse areas, ranging from AI and robotics to genome annotation, sequence analysis, analysis of gene expression, analysis of protein expression, analysis of mutations of cancer, protein structure prediction, comparative genomics, modeling biological systems, high throughput image analysis and protein-protein docking, etc. Data mining is employed for extraction of useful knowledge from large amount of data which can be used for bioinformatics. Phylogenetic analysis of gene expression data using clustering, gene expression profile for cancer gene, pattern discovery for protein modeling, etc. are some of the examples of synergy between bioinformatics and data mining approaches (Raza, 2010).

Application of data mining in banking operations has revolutionized the entire banking industry. Banks are using data mining applications in segmenting customer data, determining customer preferences, detecting fraud, cross selling banking services and retention of customers. To keep pace with changing customer preferences, data mining tools become a necessity for banks.

In additions to these data mining application domains many other fields are using data mining including text classification for example Arabic classification (Khreisat, 2006), E-learning (Luis *et al*., 2003), pharmaceutical industry, retail sector, language engineering, distance-learning, web-based education, anomaly detection in the network, software maintenance image processing, bio-technology, etc.

CONCLUSION

Researchers are presently looking the problems and their obtainable solutions; data mining provide us

convenient ways to reach the goal of getting meaningful information. Data mining tools are available for the applications to improve the standard, to achieve various objectives such as how to works at low cost, increase the quality of data used and to effectively use the stored data. Data mining is used to improve the performance of existing system in various domains. As the world of knowledge is increasing explosively, extracting necessitate become more important. In the consequence new data mining techniques are arriving and facilitating different application areas.

REFERENCES

Agrawal, R. and R. Srikant, 1995. Mining sequential patterns. Proceedings of the 11th International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan, pp: 3-14.

Agrawal, R., T. Imielinksi and A. Swami, 1993. Database Mining: A Performance Perspective. IEEE Transactions on Knowledge and Data Engineering, USA.

Ansari, S., R. Kohavi, L. Mason and Z. Zheng, 2001. Integrating E-commerce and data mining: Architecture and challenges. Proceedings of the IEEE International Conference on Data Mining, November 29-December 2, 2001, San Jose, CA, USA., pp: 27-34.

Antonie, M.L., O.R. Zaiane and A. Coman, 2001. Application of data mining techniques for medical image classification. Proceedings of the 2nd International Workshop on Multimedia Data Mining (MDM/KDD 2001) in Conjunction with ACM SIGKDD Conference, August 26, 2001, San Francisco, CA, USA.

Ayan, N.F., A.U. Tansel and M.E. Arkm, 1999. An Efficient Algorithm to Update Large Itemsets with Early Pruning. The Institute of Electrical Engineering, London, Pages: 304.

Bin, W. and L. Zhijing, 2003. Web mining research. Proceedings of 5th International Conference on Computational Intelligence and Multimedia Applications, September 27-30, 2003, IEEE Computer Society, pp: 84-89.

Bohen, S.P., O.G. Troyanskaya, O. Alter, R. Warnke, D. Botstein, P.O. Brown and R. Levy, 2003. Variation in gene expression patterns in follicular lymphoma and the response to rituximab. Proc. Natl. Acad. Sci. USA., 100: 1926-1930.

Brin, S., R. Motwani, J.D. Ullman and S. Tsur, 1997. Dynamic itemset counting and implication rules for market basket data. Proc. 1997 ACM SIGMOID Int. Conf. Manage. Data, 26: 255-264.

Carbonara, L., H. Roberts and B. Egan, 1997. Data mining in the telecommunications industry. Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge of Discovery (PKDD'97), Springer-Verlag, London, UK., pp: 486-491.

Chakrabarti, S., 2002. Mining the Web: Discovering Knowledge from Hypertext Data. 1st Edn., Morgan Kaufmann Publishers, San Francisco, USA., ISBN-13: 978-1558607545, pp: 344.

Chen, M.S., J. Han and P.S. Yu, 1996. Data mining: An overview from a database perspective. IEEE Trans. Knowledge Data Eng., 8: 866-883.

Cheung, D.W., J. Han, V.T. Ng and C.Y. Wong, 1996. Maintenance of discovered association rules in large databases: An incremental updating technique. Proceedings of International Conference on Data Engineering, February 26-March 1, 1996, New Orleans, Louisiana, pp: 106-114.

Cheung, D.W., S.D. Lee and B. Kao, 1997. A general incremental technique for maintaining discovered association rules. Proceedings of the 5th International Conference on Database Systems for Advanced Applications (ICDSAA'1997), Melbourne, Australia, pp: 185-194.

Cunningham, S.J. and E. Frank, 1999. Market basket analysis of library circulation data. Proceedings of the 6th International Conference on Neural Information Processing, November 16-20, 1999, IEEE Computer Society, Perth, WA, Australia, Washington, pp: 825-830.

De Rosa, M., 2004. Data Mining and Data Analysis for Counterterrorism. CSIS Press, USA., ISBN: 9780892064434, Pages: 24.

Elovici, Y., A. Kandel, M. Last, B. Shapira and O. Zaafrany, 2004. Using data mining techniques for detecting terror-related activities on the web. J. Inform. Warfare.

Fayyad, U. and R. Uthurusamy, 1996. Data mining and knowledge discovery databases. Commun. Assoc. Comp. Mach., 39: 24-26.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, 1996. From Data Mining to Knowledge Discovery: An Overview: Advances in Knowledge Discovery and Data Mining. AAAI Press, California, USA.

Frawley, W.J., G. Piatetsky-Shapiro and C.J. Matheus, 1991a. Knowledge Discovery in Databases: An Overview. In: Knowledge Discovery in Databases, Piatetsky-Shapiro, G. and W.J. Frawley (Eds.). AAAI/MIT Press, United States, pp: 1-27.

Fu, Y., 1996. Discovery of multiple-level rules from large databases. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.52.2392.

Gargano, M.L. and B.G. Ragged, 1999. Data mining-a powerful information creating tool. Online Comp. Lib. Center Syst. Serv., 15: 81-90.

Han, J. and M. Kamber, 2000. Data Mining: Concepts and Techniques. Morgan-Kaufman Publishers, New York.

Han, J. and Y. Fu, 1995. Discovery of multiple-level association rules from large databases. Proceedings of the 21st international Conference on Very Large Data Bases, September 11-15, 1995, Zurich, Switzerland, pp: 420-431.

Hand, D., H. Mannila and P. Smyth, 2001. Principles of Data Mining. MIT Press, Cambridge, MA.

Hedberg, S.R., 1995. The data gold rush. Byte, 20: 83-88.

Huhtala, Y., J. Karkkainen, P. Pokka and H. Toivonen, 1999. TANE: An efficient algorithm for discovering functional and approximate dependencies. Comput. J., 42: 100-111.

Huhtala, Y., J. Kinen, P. Pokka and H. Toivonen, 1998. Efficient discovery of functional and approximate dependencies using partitions. Proceedings of the 14th international conference on data engineering, (ICDE '98), IEEE Computer Society Washington, DC. USA., pp: 392-401.

Ingram, M., 2001. Internet Privacy Threatened Following Terrorist Attacks on US. http://www.surferbeware.com/privacy/privacy-threatened.htm.

Jadhav, S.R. and P. Kumbargoudar, 2007. Multimedia data mining in digital libraries: Standards and features. Proceedings of the Conference Recent Advances in Information Science and Technology, July 12-13, 2007, Madras Library Association, Kalpakkam Chapter and Scientific information Resource Division, Tamilnadu,India, pp: 54-59.

Jensen, D., 2003. Data Mining in Networks. Presentation at CSIS Data Mining Roundtable, Washington, D.C., July 23, 2003 (slide 10) http://www.thearling.com/text/dmwhite/dmwhite.htm.

Kelley, J., 2002. Terror Groups Hide Behind Web Encryption. USA Today (05/02/2002), URL: http://www.apfn.org/apfn/WTC_why.htm.

Khreisat, L., 2006. Arabic text classification using N-gram frequency statistics a comparative study. Proceedings of the 2006 International Conference on Data Mining, June 26-29, 2006, Las Vegas, Nevada, USA., pp: 78-82.

Lemos, R., 2002. What are the Real Risks of Cyber Terrorism? ZDNet, URL: http://zdnet.com.com/2100-1105-955293.html.

Li, S.T. and L.Y. Shue, 2004. Data mining to aid policy making in air pollution management. Expert Syst. Applic., 27: 331-340.

Liu, X. and P. Kellam, 2003. Mining Gene Expression Data. In: Bioinformatics: Genes, Proteins and Computers, Orengo, C., D. Jones and J. Thornton (Eds.). BIOS Scientific Publishers, United States, pp: 229-244.

Luis, R., J. Redol, D. Simoes and N. Horta, 2003. Data warehousing and data mining system applied to E-learning. Proceedings of the 2nd International Conference on Multimedia and Information and Communication Technologies in Education, December 3-6, 2003, Badajoz, Spain.

Maedche, S.S., 2000. Discovering conceptual relations from text. Proceedings of the 14th European Conference on Artificial Intelligence, August 20-25, 2000, Berlin, Germany, 2000, pp: 321-323.

Mannila, H., H. Toivonen and A.I. Verkamo, 1995. Discovering frequent episodes in sequences. Proceedings of the 1st International Conference on knowledge Discovery and Data Mining, August 20-21, 1995, AAAI Montreal, Quebec, Canada, pp: 210-215.

Margahny, M.H. and A.A. Mitwaly, 2005. Fast algorithm for mining association rules. Proceedings of the AIML 2005 Conference, December 19-21, 2005, CICC, Cairo, Egypt.

Mateos, A., J. Dopazo, R. Jansen Y. Tu, M. Gerstein and G. Stolovitzky, 2002. Systematic learning of gene functional classes from DNA array expression data by using multilayer perceptrons. Genome Res., 12: 1703-1715.

Ng, S.K., S.H. Tan and V.S. Sundararajan, 2003. On combining multiple microarray studies for improved functional classification by whole-dataset feature selection. Genome Inform., 14: 44-53.

Park, J.S., M.S. Chen and P.S. Yu, 1995. An effective hash-based algorithm for mining association rules. Proceedings of the 1995 ACM SIG-MOD International Conference on Management of Data, May 22-25, 1995, San Jose, California, pp: 175-186.

Piatetsky-Shapiro, G. and W.J. Frawley, 1991. Knowledge Discovery in Databases. AAAI/MIT Press, Menlo Park, California, ISBN: 0262660709, Pages: 525.

Ponniah, P., 2001. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. A Wiley-Interscience Publication, John Wiley and Sons, Inc., United States, ISBN: 0-471-22162-7, Pages: 481.

Raza, K., 2010. Application of data mining in bioinformatics. Indian J. Comput. Sci. Eng., 1: 114-118.

Roussinov, D. and J.L. Zhao, 2003. Automatic discovery of similarity relationships through Web mining. Decision Supp. Syst., 35: 149-166.

Savasere, A., E. Omieccinski and S. Navathe, 1995. An efficient algorithm for mining association rules in large databases. Proceedings of the 21st International Conference on Very Large Databases, September 11-15, 1995, Zurich, Switzerland, pp: 432-443.

Setty, D.V., T.M. Rangaswamy and K.N. Subramanya, 2010. A review on data mining applications to the performance of stock marketing. Int. J. Comput. Appl., Vol. 3.

Solieman, O.K., 2006. Data mining in sports: A research overview. A Technical Report, MIS Masters Project, August 2006. http://ai.arizona.edu/hchen/chencourse/OsamaDM_in_ Sports.pdf.

Srikant, R. and R. Agrawal, 1996. Mining sequential patterns: Generalizations and performance improvements. Proc. Int. Conf. Extend. Database Technol., 1057: 3-17.

Thomas, S., S. Bodagala, K. Alsabti and S. Ranka, 1997. An efficient algorithm for the incremental updation of association rules in large databases. Proceedings of 3rd International Conference on Knowledge Discovery and Data Mining, August 14-17, 1997, Newport Beach, CA., pp: 263-266.

Toivonen, H., 1996. Sampling large databases for association rules. Proceedings of 22th International Conference on Very Large Databases, Septempber 3-6, 1996, Bombay, India, pp: 134-145.

Tukey J.W., 1977. Exploratory Data Analysis. Addison Wesney, Tukey.

Wang, C.Y., T.P. Hong and S.S. Tseng, 2002. Maintenance of discovered sequential patterns for record deletion. Intell. Data Anal., 6: 399-410.