# Data Mining Usages in Knowledge Management: A Review Study

Brij Kishore[1], Dr. Raj Kumar[2,] Reshu Grover[3,] Anurag Maloo[4]
[1]Ph.D Scholar, Suresh Gyan Vihar University Jaipur
[2]Assistant Professor, Deptt. Of CSE, JIET, Jind, Email: rajshira@gmail.com
[3] Assistant Professor LIET Alwar, Email reshugroer3@gmail.com
[4] Assistant Professor Sangam University Bhilwara, Email maloo.anurag@gmail.com

***Abstract***— *Data mining is one of the most important steps of the knowledge discovery in databases process and is considered as significant subfield in knowledge management. Research in data mining continues growing in business and in learning organization over coming decades. This review paper explores the applications of data mining techniques which have been developed to support knowledge management process. The journal articles indexed in ScienceDirect Database from 2007 to 2012 are analyzed and classified. The discussion on the findings is divided into 4 topics: (i) knowledge resource; (ii) knowledge types and/or knowledge datasets; (iii) data mining tasks; and (iv) data mining techniques and applications used in knowledge management. The article first briefly describes the definition of data mining and data mining functionality. Then the knowledge management rationale and major knowledge management tools integrated in knowledge management cycle are described. Finally, the applications of data mining techniques in the process of knowledge management are summarized and discussed.*

***KEYWORDS:*** *Data mining; Data mining applications; Knowledge management*

## 1. INTRODUCTION

A data warehouse is a "subject-oriented, integrated,time varying, non-volatile collection of data that is used primarily in organizational decision making. ( Inmon,W.H.,1992) Typically, the data warehouse is maintained separately from the organization's operational databases There are many reasons for doing this. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases. Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions. It serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions. The data can be stored in many different types of databases. One data base architecture that has recently emerged is the "data warehouse", a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision-making. Data warehouse technology includes data cleansing, dataintegration and online Analytical processing. OLAP standsfor analysis techniques with functionalities such assummarization, consolidation and aggregation, as well asthe ability to view information from different angles.Ten years ago, Data Warehousing was largely unknown.Today, many companies are receiving considerable businessvalue from their warehousing efforts. First American
Corporation (FAC), a regional bank located in the Southeast, lost $60 million in 1990 and was operating under letters of agreement with regulators. A new senior management team developed a customer intimacy strategy attract, maintain, and enhance their customer base; create profitable new product and service offerings; and redesign their distribution channels to increase profitability and better meet customers' needs. Data warehousing helped FAC to become a profitable, innovative leader in the financial services industry.
Since the introduction of the data warehouse concept in the late 1980ies (e.g. Devlin/Murphy 1988), data warehouse systems are now an established component of information systems landscape in most companies. Due to high failure rates of data warehouse projects, several procedure models for building data warehouse systems were published considering their special requirements (e.g. Inmon 1996, Kimball 1996, Gardner 1998, Simon 1998). However, these methodologies were mainly focused on technical issues, like architectural concepts and data modeling. But according to studies about critical success factors of data warehouse projects organizational, political and cultural factors are at least as important as technical ones (Frolick/Lindsey 2003, Hwang et al. 2002, Finnegan/Sammon 1999). In addition, most development methodologies are lacking concepts to ensure long-term evolution and establishment of data warehouse systems (O'Donnell et al. 2002), which are both primarily organizational challenges (Meyer/Winter 2001). Up to now only few authors have adopted a mainly organizational-driven view on data warehouse systems. Kachur (2000) describes activities and organizational structures for data warehouse management. With a data warehouse at the heart of the strategy. Using warehouse data, FAC was able to determine the profitability of all of their clients and products; develop programs to

## 2. DATA MINING

### 2.1 Definition of Data Mining

Data mining is an essential step in the knowledge discovery in databases (KDD) process that produces useful patterns or models from data [7]. The terms of KDD and data mining are different. KDD refers to the overall process of discovering useful knowledge from data. Data mining refers to discover new patterns from a wealth of data in databases by focusing on the algorithms to extract useful knowledge [7].

### 2.2 Data Mining Tasks

Fayyad et.al. (1996) define six main functions of data mining:

1. *Classification* is finding models that analyze and classify a data item into severalpredefined classes
2. *Regression* is mapping a data item to a real-valued prediction variable
3. *Clustering* is identifying a finite set of categories or clusters to describe the data
4. *Dependency Modeling* (*Association Rule Learning*) is finding a model which describes
significant dependencies between variables
5. *Deviation Detection* (*Anomaly Detection*) is discovering the most significant changes in
the data
6. *Summarization* is finding a compact description for a subset of data

Data mining has two primary objectives of prediction and description. Prediction involves using some variables in data sets in order to predict unknown values of other relevant variables (*e.g. classification, regression, and anomaly detection*) Description involves finding humanunderstandable patterns and trends in the data (*e.g. clustering, association rule learning, and summarization*) [8].

## 3. KNOWLEDGE MANAGEMENT

### 3.1 Definition of Knowledge Management

There are various concepts of knowledge management. In this paper we use the definition of knowledge management by McInerney (2002): "Knowledge management (KM) is an effort to increase useful knowledge within the organization. Ways to do this include encouraging communication, offering opportunities to learn, and promoting the sharing of appropriate knowledge artifacts" This definition emphasizes the interaction aspect of knowledge management and organizational learning. Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge [5]. Each of knowledge units of capture and creation, sharing and dissemination, and acquisition and application can be facilitated by information technology.

### 3.2 Knowledge Management: Capture and Creation Tools

This section provides an overview of a classification of KM technologies as tools and focuses on tools for capture and creation knowledge. Liao (2003) classifies KM technologies using seven categories:

1. KM Framework
2. Knowledge-Based Systems (KBS)
3. Data Mining
4. Information and Communication Technology
5. Artificial Intelligence (AI)/Expert Systems (ES)
6. Database Technology (DT)
7. Modeling

Ruggles et.al. (1997) classify KM technologies as tools that generate knowledge (e.g. data mining), code knowledge, and transfer knowledge. Dalkir (2005) classifies KM tools according to the phase of the KM cycle . We can see that data mining involves in the part of knowledge creation and capture phase.

## 4. THE APPLICATIONS OF DATA MINING IN KNOWLEDGE MANAGEMENT

The reviews of ten articles has discussed on the applications of data mining to organizational
knowledge management for effective capturing, storing and retrieving, and transferring knowledge. We divided the reviewed articles into four main groups: (i) knowledge resource; (ii)
knowledge types and/or knowledge datasets; (iii) data mining tasks; and (iv) data mining
techniques and applications used in KM.

### 4.1 Knowledge Resources

In the study, we divided knowledge resources into eight groups as that which knowledge object to
be stored and manipulated in KM and how data mining aids.

1. Health Care Organization: this domain was a use of the disease knowledge management
system (KMS) of the hospital case study [10]. Data mining tool was used to explore
diseases, operations, and tumors relationships. This tool used to build KMS to support
clinical medicine in order to improve treatment quality [10].
2. Retailing: this was customer knowledge from household customers for product line and
brand extension issues [14]; data mining can help and propose suggestions and solutions
to the firm for product line and brand extensions. This doing by extracting market
knowledge of customers, brands, products, and purchase data to fulfill the customers'
demands behavior [14].
3. Financial/Banking: the domain knowledge covered financial and economic data; data
mining can assist banking institutions making decision support and knowledge sharing
processes to an enterprise bond classification [4].
4. Small and Middle Businesses (food company and food supply chain): there were two
methods and processes to obtain knowledge resources: knowledge seeding-the relative
knowledge to the problems; knowledge cultivating-the process to find the key knowledge
from knowledge seeding [12]. Data mining and knowledge management integrated can
help making better decisions [12]. As Death-On-Arrival (DOA) problem encountered in
food supply chain networks (FSCN), Li et al. (2010) aimed to build Early Warning and
Proactive Control (EW&PC) systems to solve such problems [13]. Knowledge Base was
an important part of EW&PC systems. It contained data analysis by managers and
organizes in an appropriate way for other managers. Data mining methods were helpful
for the EW&PC systems [13].
5. Entrepreneurial Science:the knowledge resource was research assets in a knowledge
institution [3]; there were three types of the research assets: research products,
intellectual capital, and research programs. Data mining facilitated for knowledge extraction and helped guiding managers in determining strategies on knowledge-oriented organization competition [3].
6. Business: data collected from questionnaire, an intensive literature review, and discussions with four KM experts [27]. Data mining can discover hidden patterns
between KM and its performance for better KM implementations[27]. 7. Collaboration and Teamwork:

Worker's log and documents were analyzed each worker's referencing behavior and construct worker's knowledge flow. Data mining techniques can mine and construct group-based knowledge flows (GKFs) prototype for task-based groups [16]. 8. Construction Industry: a large part of this enterprise information was available in the form of textual data formats [24]. This leads to the influence of text mining techniques to handle textual information source for industrial knowledge discovery and management solutions [24].

### 4.2 Knowledge Types

This section described knowledge types in 8 organization domains for data mining collaboration process in the knowledge creation.

- **Health-care System domain,** the dataset composed of three databases: the health-care providers' database; the out-patient health-care statistics database; and the medical status database [11].Another data source was from hospital inpatient medical records [10].
- **Construction Industry domain,** a sample data set was in the form of Post Project Reviews (PPRs) as defining good or bad information [24]. Multiple Key Term Phrasal Knowledge sequences (MKTPKS) formation was generated through applications of text
mining and was used an essential part of the text analysis in the text documents classification[24].
- **Retailing domain:** customer data and the products purchased have been collected and stored in databases to mine whether the customers' purchase habits and behavior affect the product line and brand extensions or not [14]. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012 21
- **Financial domain:** There were two datasets posed in financial domain: (i) to identify bond ratings, knowledge sets contained strings of data, models, parameters and reports for each analytical study; and (ii) to predict rating changes of bonds, cluster data of bond features as well as the model parameters were stored, classified, and applied to rating predictions [4].
- **Small and Middle Businesses (SMBs) domain:** Knowledge types in small and middle businesses in case of Food Company were related to the corporate conditions or goals of the problem among all departments to develop a decision system platform and then
formed the knowledge tree to find relations by human-computer interaction method and optimize the process of decision making[12].To solve food supply chain networks problems, Li et al. (2010) developed EW&PC prototype which composed of major components of: (i) knowledge base, (ii) task classifier and template approaches, (iii) DM methods library with expert system for method selection, (iv) explorer and predictor, and
(v) user interface [13]. This system built decision support models and helped managers to accomplish decision-making.
- **Research Assets domain:** In Cantu & Cellbos (2010) focused on managing knowledge assets by applied aknowledge and information network (KIN) approach. This platform contained three components types of research products, human resources or intellectual
capital, and research programs. The various types of research assets were handled on domain ontologies and databases [3].
- **Business domain:** there were two types of knowledge attributes conducted: condition attributes and decision attribute [27]. Condition attributes included four independent attributes of the KM purpose, the explicit-oriented degree, the tacit-oriented degree, and
the success factor. Decision attribute included one dependent attribute of the KM performance [27].
- **Collaboration and Teamwork domain:** a dataset used from a research laboratory in a research institute. It contained 14 knowledge workers, 424 research documents, and a workers' log as that recorded the time of document accessed and the documents of workers' needed [16]. For the workers' log, it was generated to 2 levels of codified-level knowledge flow and topic-level knowledge flow [16]. The two types of knowledge flow
were determined to describe a worker's needs. To collect the knowledge flow, documents in the datasetwere categorized into eight clusters by data mining clustering approach [16].

### 4.3 Data Mining Techniques/Applications Used in Knowledge Management

Within the context of articles reviewed, applications of data mining have been widely used in various enterprises ranging from public health-care, construction industry, food company, retailing to finance. Each field can be supported by different data mining techniques which generally include classification, clustering, and dependency modeling. We provided a brief description of the four most used data mining techniques including its common tools used and some references as follows [7]: International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012 22

**Classification:** Classification is one of the most common learning in data mining. This task aims
at mapping a data item into one of several predefined classes. Examples of classification methods
used as part of knowledge management include the classifying of the patients from primary
health-care centers to specialists; the combination of the data mining and decision support
approaches in planning of the regional health-care system; and the implementation of
visualization method to facilitate KM and decision making processes [11]. In the financial
company, Cheng, Lu & Sheu (2009) implemented an ontology-based approach of KM and
knowledge sharing in financial knowledge management system (FKMS) and applied the hybrid
SOFM/LVQ classifier of clustering and classification data mining techniques to classify
corporate bonds [4]. For small and middle businesses: food company domain, data mining can
improve decision-making by knowledge cultivating method namely *Extenics* and *Extension data mining (EDM)* [12]. This method was the integration of data mining and knowledge management,
to develop a decision support system platform for better decisions [12]. To solve the death-onarrival
(DOA) in food supply chain networks, corporate manager selected variables that might
have influence on DOA by using *"decision tree"* of data mining method; and used *"neural network"* to monitor potential DOA for prediction [13]. As knowledge assets played an important

role in knowledge economies, Cantu & Ceballos (2010) employed data mining agents for extracting useful patterns to assist decision makers in generating benefits from the knowledge assets and used a knowledge information network (KIN) platform for managing the knowledge assets[3].In the business organizations with a large volume of works, such companies wanted to better understand what the hidden patterns between the KM and its performance, using the combination of data mining techniques: Bayesian Network (BN) classifier and Rough Set Theory (RST) in their business could help companies producing the KM to be performed effectively and achieve higher efficacy resulted [27]. Common tools used for classification are decision trees, neural network, Bayesian network and rough set theory.

**Clustering:** This involved seeking to identify a finite set of categories and grouping together objects that are similar to each other and dissimilar to the objects belonging to other clusters.This technique has been applied in many fields, for example:

- **Healthcare:** clustering categories and attributes used in analyzing the similarities between community health centers [11].
- **Retailing:** clustering the segmentation for possible product line and brand extension to identify market to customer clusters [14];
- **Financial/Banking:** identifying groups of corporate bond clusters according to the industry and a specific segment within an industry; then tuning cluster data for each industry as a template for predicting rating changes [4].
- **Construction Industry:** clustering textual data to discover groups of similar access patterns [24].
- **Collaboration and Teamwork:** identifying groups of workers with similar task-related information needs based on the similarities of workers' knowledge flow [16].

Common tools used for clustering include k-means, principal component analysis, the Kolmogorov-Smirnov test and the quantile range test and polar ordination.

**Dependency Modeling:** This concerned with finding a model that describes significant relationships between attribute sets. For example, it is widely used in healthcare to develop clinical pathway guidelines and provide an evidence-based medicine platform [10]. In medical International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012 23 records management, it is helpful for clinical decision making [10]. It could give better results in knowledge refinement through a use of this technique on the construction industry dataset [24]; this technique used to mine customer knowledge from household customers [14]. Common tools for dependency modeling are *Apriori* association rules and sequential pattern analysis. As above, we can see that data mining techniques and applications in literature reveal different solutions to different KM problems in practice.

## 5. CONCLUSIONS

In organization, knowledge is an important resource. Management of knowledge resources has become a strong demand for development. Discovering the useful knowledge has also significant approach for management and decision making. As data mining is a main part of KM, this paper has identified ten articles related to data mining applications in KM, published between 2007 and 2012. This aims to give a research summary on the application of data mining in the KM technologies domain. The results presented in this paper have some assumptions:

- On the basis of the publication rates, research on the application of data mining in KM will increase in the future and cover the interest in different areas
- The classification of data mining tasks is usually the employed model in organization for description and prediction. However, we will see the hybridization techniques e.g. association rule and clustering; classification and clustering etc. in order to solve different KM problems. This trend will give rising in the future.
- In the context of healthcare, one article used the visualization technique as a supplement to other data mining tasks. This visualization system could enhance and lead to better performance in decision making.
- KM is an interdisciplinary research area. Thus, in the future, KM development may need integration with different technologies and demand more methodologies to solve KM problems.
- KM applications development tends to support expert decision making and will be the application of a problem-oriented domain.

In this paper, we have shown that data mining can be integrated into KM framework and enhanced the KM process with better knowledge. It is clear that the data mining techniques will have a major impact on the practice of KM, and will present significance challenges for future knowledge and information systems research.

## References

[1] An, X. & Wang, W. (2010). Knowledge management technologies and applications: A literature review. IEEE, 138-141. doi:10.1109/ICAMS.2010.5553046

[2] Berson, A., Smith, S.J. &Thearling, K. (1999). Building Data Mining Applications for CRM. New York: McGraw-Hill.

[3] Cantú, F.J. & Ceballos, H.G. (2010). A multiagent knowledge and information network approach for managing research assets. Expert Systems with Applications, 37(7), 5272-5284. doi:10.1016/j.eswa.2010.01.012

[4] Cheng, H., Lu, Y. & Sheu, C. (2009). An ontology-based business intelligence application in a financial knowledge management system.Expert Systems with Applications, 36, 3614–3622. Doi:10.1016/j.eswa.2008.02.047

International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012 24

[5] Dalkir, K. (2005). Knowledge Management in Theory and Practice. Boston: Butterworth-Heinemann.

[6] Dawei, J. (2011). The Application of Date Mining in Knowledge Management.2011 International Conference on Management of e-Commerce and e-Government, IEEE Computer Society, 7-9. doi: 10.1109/ICMeCG.2011.58

[7] Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.AI Magazine, 17(3), 37-54.
[8] Gorunescu, F. (2011). Data Mining: Concepts, Models, and Techniques. India: Springer.
[9] Han, J. &Kamber, M. (2012). Data Mining: Concepts and Techniques. 3rd.ed. Boston: Morgan Kaufmann Publishers.
[10] Hwang, H.G., Chang, I.C., Chen, F.J. & Wu, S.Y. (2008). Investigation of the application of KMS for diseases classifications: A study in a Taiwanese hospital. Expert Systems with Applications, 34(1), 725-733. doi:10.1016/j.eswa.2006.10.018
[11] Lavrac, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M. &Kobler, A. (2007).Data mining and visualization for decision support and modeling of public health-care resources.Journal of Biomedical Informatics, 40, 438-447. doi:10.1016/j.jbi.2006.10.003
[12] Li, X., Zhu, Z. & Pan, X. (2010). Knowledge cultivating for intelligent decision making in small & middle businesses.Procedia Computer Science, 1(1), 2479-2488. doi:10.1016/j.procs.2010.04.280
[13] Li, Y., Kramer, M.R., Beulens, A.J.M., Van Der Vorst, J.G.A.J. (2010). A framework for early warning and proactive control systems in food supply chain networks. Computers in Industry, 61, 852–862. Doi:101.016/j.compind.2010.07.010
[14] Liao, S.H., Chen, C.M., Wu, C.H. (2008). Mining customer knowledge for product line and brand extension in retailing.Expert Systems with Applications, 34(3), 1763-1776. doi:10.1016/j.eswa.2007.01.036
[15] Liao, S. (2003). Knowledge management technologies and applications-literature review from 1995 to 2002. Expert Systems with Applications, 25, 155-164. doi:10.1016/S0957-4174(03)00043-5
[16] Liu, D.R. & Lai, C.H. (2011). Mining group-based knowledge flows for sharing task knowledge. Decision Support Systems,50(2), 370-386. doi:10.1016/j.dss.2010.09.004
[17] Lee, M.R. & Chen, T.T. (2011). Revealing research themes and trends in knowledge management: From 1995 to 2010. Knowledge-Based Systems.doi:10.1016/j.knosys.2011.11.016
[18] McInerney, C.R. & Koenig, M.E. (2011). Knowledge Management (KM) Processes in Organizations: Theoretical Foundations and Practice. USA: Morgan & Claypool Publishers. doi:10.2200/S00323ED1V01Y201012ICR018
[19] McInerney, C. (2002). Knowledge Management and the Dynamic Nature of Knowledge.Journal of the American Society for Information Science and Technology, 53(12), 1009-1018. doi:10.1002/asi.10109
[20] Ngai, E., Xiu, L. &Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36, 2592-2602. doi:10.1016/j.eswa.2008.02.021
[21] Ruggles, R.L. (ed.). (1997). Knowledge Management Tools. Boston: Butterworth-Heinemann.
[22] Sher, P.J. & Lee, V.C. (2004). Information technology as a facilitator for enhancing dynamic capabilities through knowledge management.Information & Management, 41, 933-945. doi:10.1016/j.im.2003.06.004
[23] Tseng, S.M. (2008). The effects of information technology on knowledge management systems.Expert Systems with Applications, 35, 150-160. doi:10.1016/j.eswa.2007.06.011
[24] Ur-Rahman, N. & Harding, J.A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications, 39, 4729-4739. doi:10.1016/j.eswa.2011.09.124
[25] Wang, F. & Fan, H. (2008). Investigation on Technology Systems for Knowledge Management.IEEE, 1-4. doi:10.1109/WiCom.2008.2716
[26] Wang, H. & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. Industrial Management & Data Systems, 108(5), 622-634.
[27] Wu, W., Lee, Y.T., Tseng, M.L. & Chiang, Y.H. (2010). Data mining for exploring hidden patterns between KM and its performance.Knowledge-Based Systems, 23, 397-401. doi:10.1016/j.knosys.2010.01.014