International Journal of Scientific & Engineering Research, Volume 5, Issue 6, June-2014
ISSN 2229-5518

Data Mining and Its Applications: A Review

Kamlesh Kumar[1], Bhavesh Kumar Chauhan[2], J. P. Pandey[3]

ABSTRACT: *As in reference this paper begins with the definition of data mining classification and its application. This paper also includes knowledge resource; knowledge types and/or knowledge datasets; data mining tasks; and 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.*

Index Terms**:** Data mining; Data mining applications; Knowledge management, Knowledge resource, Data Base management, Data mining process, Techniques.

1. INTRODUCTION

In information era, knowledge is becoming a crucial organizational resource that provides competitive advantage and giving rise to knowledge management (KM) initiatives. Many organizations have collected and stored vast amount of data. However, they are unable to discover valuable information hidden in the data by transforming these data into valuable and useful knowledge [1]. Managing knowledge resources can be a challenge. Many organizations are employing information technology in knowledge management to aid creation, sharing, integration, and distribution of knowledge.Data mining is the process of discovering potentially useful, interesting, and previously unknown patterns from a large collection of data [2].Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both [3].Text mining involves the application of techniquesfrom areas such as information retrieval,naturallanguage processing, information extraction anddata mining [4].The aim data mining isto uncover previously unknown, useful knowledge [4].POWERFUL data acquisition systems (such as minicomputers, microprocessors, transducers, and analog-to-digital converters) that collect, analyze, and transfer data are in use in various mid-range and large organizations [5], [6]–[7], [8].Knowledge management is a process of data usage [9]. The basis of data mining is a process of using tools to extract useful knowledge from large datasets; data mining is an essential part of knowledge management [9].Wang & Wang (2008) point that data mining can be useful for KM intwo main manners: (i) to share common knowledge of business intelligence (BI) context among data miners and (ii) to use data mining as a tool to extend human knowledge. Thus, data mining tools could help organizations to discover the hidden knowledge in the enormous amount of data.

As a part of data mining research, this paper focuses on surveying data mining applications in knowledge management through a literature review of articles from 2007 to 2012. The reason for reviewing research article this period is that data mining has emerged in KM research theme since 2006 [10] and it plays important roles as a link between business intelligence and knowledge management [11].

1.Uttarkhand Technical University, Dehradun
2. ABESIT Ghaziabad
3. KNIT Sultanpur

For article filtering, we search for the keyword "data mining" and "knowledge management" in the article title, abstract, and keywords fields on the Science Direct database. We limit the document search to date range published 2007 to 2012 and only document type of "research article" is included. The total number of documents published for this query by year shows in Figure 1.

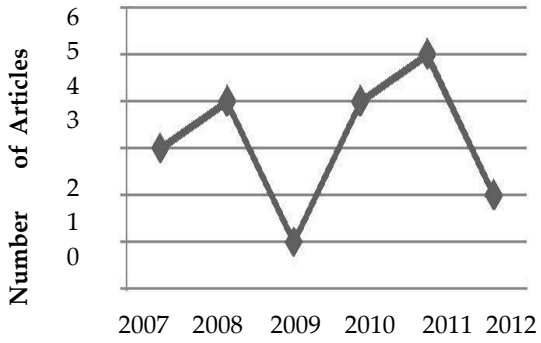

Figure 1 Number of Articles with"*Data Mining*" and "*Knowledge Management*" in the Title, Abstract, or Keyword Fields 2007-2012

The full text of each article is carefully reviewed to eliminate those articles that are not related to application of data mining in KM and not described how data mining could be employed or helped in KM. There are 10 articles related to these selection criteria.

Based on 10 articles on data mining applications for knowledge management, we survey and classify according to the six categories of data mining techniques: classification, regression, clustering, dependency modeling, deviation detection, and summarization.

The purpose of this paper is to review literature related to application of data mining techniques for KM in academic journals between 2007 and 2012. We organize this paper as follows: first, data mining definition and the data mining task primitively used in this study are described; second, the definition of knowledge management and the knowledge capture and creation tools are presented; third, articles about data mining in KM are analyzed and the results of the classification are reported; and last, the conclusions of the study are discussed.

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 (Figure 2) [12]. 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 [12]. The KDD process is defined in [13] as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data." Fayyad *et al.* [13] distinguish between two main categories of data mining (fourth step above): *verification-oriented* and *discovery-oriented*.

The Intelligent Discovery Assistants [14] (IDA), helps users in applying validknowledge discovery processes. According to [15], although data mining is at the core of the KDD process, it is just one stepin the overall KDD process, and it usually takes about 15 to 25% of the overall effort.It is the exploration of large datasets to extract hidden and previously unknown patterns, relationships and knowledge that are difficult to detect with traditional statistical methods [16-20].Data Mining is about solving problems by analyzing dataalready present in databases [21].

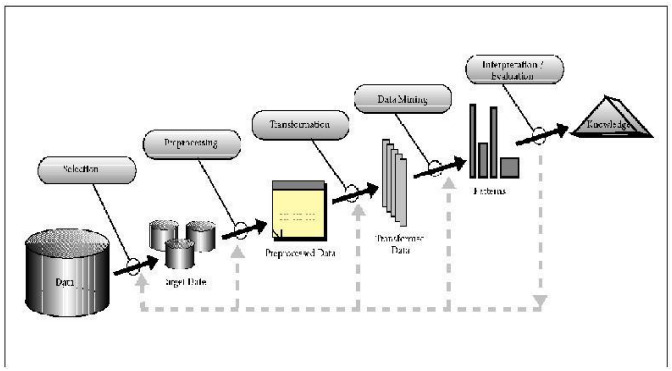

Figure 2 Data Mining and the KDD Process

(Source: Fayyad, et.al., 1996)

Based on figure 2, KDD process consists of iterative sequence methods as follows [22, 23]:

1. *Selection*: Selecting data relevant to the analysis task from the database
2. *Preprocessing*: Removing noise and inconsistent data; combining multiple data sources
3. *Transformation*: Transforming data into appropriate forms to perform data mining
4. *Data mining*: Choosing a data mining algorithm which is appropriate to pattern in thedata; Extracting data patterns
5. *Interpretation/Evaluation* : Interpreting the patterns into knowledge by removingredundant or irrelevant patterns; Translating the useful patterns into terms that human-understandable

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
5. *Dependency Modeling* (*Association Rule Learning*) is finding a model which describessignificant dependencies between variables
6. *Deviation Detection* (*Anomaly Detection*) is discovering the most

significant changes inthe data

7. *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 human-understandable patterns and trends in the data (*e.g. clustering, association rule learning, and summarization*) [24].

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.Vance (1997) defines information as data interpreted into ameaningful framework whereas knowledge is information that has been authenticated andthought to be true. Maglitta (1996) suggests that data is raw numbers and facts, information is processed data, and knowledge is "information made actionable."[25].The importance of KM today for individuals, for communities of practice, and for organizations are described togetherwith the emerging KM roles and responsibilities needed to ensure successful KM implementations[26].
'…..The processes that governs the creation, dissemination, and utilization of knowledge…' (Newman, 1992). [27].
'…. Managing the organization's knowledge of employees so others may be more effective organizational performance…' (O'Leary, 1998). [27].

Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge (Figure 3) [28]. Each of knowledge units of capture and creation, sharing and dissemination, and acquisition and application can be facilitated by information technology.Knowledge management practices measured throughinformation technology, organizationand knowledge positively affect organizational performance[29].

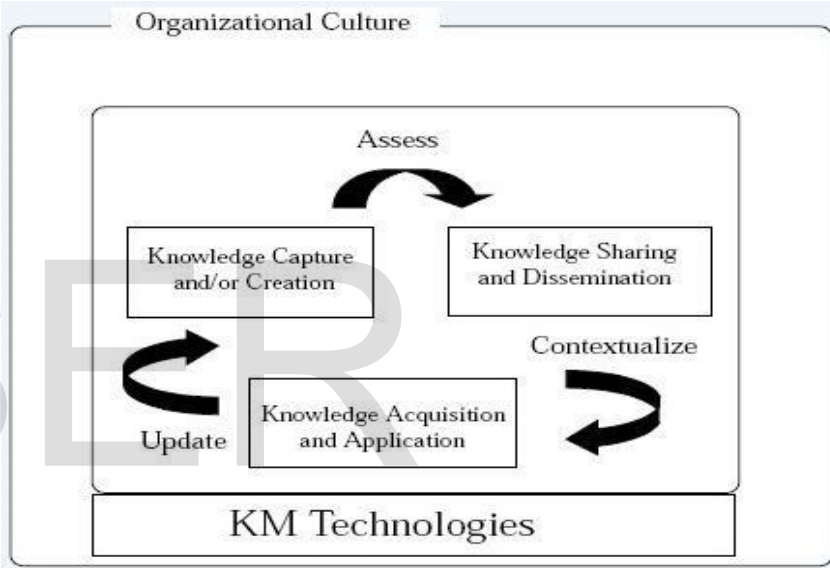

Figure 3 KM Technologies Integrated KM Cycle (Source from Dalkir, .,2005).

As technologies play an important role in KM, technologies stand to be a necessary tool for KM usage [30]. Thus, KM requires technologies to facilitate communication, collaboration, and content for better knowledge capture, sharing, dissemination, and application.

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. Knowledge management is the system and managerial approach to thegathering, management, use, analysis, sharing, and discovery of knowledgein an organization or a community in order to maximize performance (Chen,2001) [31-34].

Text mining aims to extract useful knowledge from textual data ordocuments (Hearst, 1999; Chen, 2001). Although text mining is oftenconsidered a subfield of data mining, some text mining techniques haveoriginated from other disciplines, such as information retrieval, informationvisualization, computational linguistics, and information science [35-36].

Finally,the results of numerous researches [36-40] show that KM affects organizationalperformance in a positive manner, but this relationship is very difficultto prove.

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 [41]. 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 [41].Data mining applications in health can have tremendous potential and usefulness [42].

2. Retailing: this was customer knowledge from household customers for product line and brand extension issues [43]; 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 [43].

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 [44].

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 [45]. Data mining and knowledge management integrated can help making better decisions [45]. 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 [46]. 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 [46].

5. Entrepreneurial Science: the knowledge resource was research assets in a knowledge institution [47]; 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 [47].

6. Business: data collected from questionnaire, an intensive literature review, and discussions with four KM experts [48]. Data mining can discover hidden patterns between KM and its performance for better KM implementations [48].

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 [49].

8. Construction Industry: a large part of this enterprise information was available in the form of textual data formats [50]. This leads to the influence of text mining techniques to handle textual information source for industrial knowledge discovery and management solutions [50].

9. The data mining is used an emerging trends in the education system [51-54] inthe whole world.

4.2 Knowledge Types:

This section described knowledge types in some organizational domains for data mining collaboration process in the knowledge creation.

- **Health-care System domain,** the dataset composed of three databases: the health-careproviders' database; the out-patient health-care statistics database; and the medical status database [55].Another data source was from hospital inpatient medical records [56].

- **Construction Industry domain,** a sample data set was in the form of Post ProjectReviews (PPRs) as defining good or bad information [57]. 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 [57].

- **Retailing domain:** customer data and the products purchased have been collected andstored in databases to mine whether the customers' purchase habits and behavior affect the product line and brand extensions or not [58].

- **The Intelligence Agencies domain:** The Intelligence Agencies collect and analyze information to investigate terrorist activities.One challenge to law enforcement and intelligent agencies is the difficulty of analyzing largevolume of data involve in criminal and terrorist activities. Now

a days the intelligence agency areusing the sophisticated data mining algorithms which makes it easy, to handle the very large databases databases for organizations. The different data mining techniques are used in crime data mining. [59-63] .Though the organizations have used large data bases but data mining helps usto generate the different types of information in the organization like personal details of thepersons along with, vehicle details.

Internal Revenue Service domain: The data mining technology SAS/EM is used to discover therules those are unknown before and it can improve the quality of products and decrease the cost [64]. The data mining system implemented at the Internal Revenue Service to identify high-incomeindividuals engaged in abusive tax shelters [65] show significantly good results and generates base data for next year.

E-commerce is also the most prospective domain: E-commerce is also the most prospective domain for data mining [66]. It is ideal becausemany of the ingredients required for successful data mining are easily available: data records areplentiful, electronic collection provides reliable data, insight can easily be turned into action, and return on investment can be measured. The integration of e-commerce and data mining significantly improve the results and guide the users in generating knowledge and making correctbusiness decisions.

The Digital Library domain: The data mining application can be used in the field of the Digital Library where the user willfinds or collects, stores and preserves the data which are in the form of digital mode. The adventof electronic resources and their increased use in libraries has brought about significant changes in Library [67]. The data and information are available in the different formats. These formatsinclude Text, Images, Video, Audio, Picture, Maps, etc. therefore digital library is a suitabledomain for application of data mining.

The prediction in engineering applications: The prediction in engineering applications was treated effectively by a data miningapproach [68]. The prediction problems like the cost estimation problem in engineering, theproblem of engineering design that involves decisions where parameters, actions, components, and so on are selected. Data mining technique is used for the variety of the parameters in the field ofengineering applications like prior data.

Financial domain: There were two datasets posed in financial domain: (i) to identifybond 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 [69].

- **Small and Middle Businesses (SMBs) domain:** Knowledge types in small and middlebusinesses 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[70].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 [71]. This system built decision support models and helped managers to accomplish decision-making.

- **Research Assets domain:** In Cantu

& Cellbos (2010) focused on managing knowledgeassets by applied acknowledge 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 [72].

- **Business domain:** there were two types of knowledge attributes conducted: conditionattributes and decision attribute [73]. 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 [73].

Collaboration and Teamwork domain: a dataset used from a research laboratory in aresearch 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 [74]. For the workers' log, it was generated to 2 levels of codified-level knowledge flow and topic-level knowledge flow [74]. 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 [74].

Data Mining Life Cycle: The life cycle of a data mining project consists of six phases [75-76]. The sequence of the phases is not rigid. Moving back and forth between different phases is always required. It depends onthe outcome of each phase. The main phases are:

Business Understanding: This phase focuses on understanding the project objectives and requirements from a businessperspective, then converting this knowledge into a data mining problem definition and apreliminary plan designed to achieve the objectives.

Data Understanding: It starts with an initial data collection, to get familiar with the data, to identify data qualityproblems, to discover first insights into the data or to detect interesting subsets to form hypotheses forhidden information.

Data Preparation: In this stage, it collects all the different data sets and constructs the varieties of the activitiesbasing on the initial raw data.

Modeling: In this phase, various modeling techniques are selected and applied and their parameters arecalibrated to optimal values.

Evaluation: In this stage the model is thoroughly evaluated and reviewed. The steps executed to construct themodel to be certain it properly achieves the business objectives. At the end of this phase, adecision on the use of the data mining results should be reached.

Deployment: The purpose of the model is to increase knowledge of the data, the knowledgegained will needto be organized and presented in a way that the customer can use it.Thedeployment phase can beas simple as generating a report or as complex as implementing arepeatable data mining processacross the enterprise.

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 [77]:

Classification: Classification [78] is one of the most common learning in data mining. This task aimsat 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 [79]. 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 [80]. For small and middle businesses: food company domain, data mining can improve decision-making by knowledge cultivating method namely *Extenics* and *Extension datamining (EDM)* [81]. This method was the integration of data mining and knowledge management,to develop a decision support system platform for better decisions [81]. To solve the death-on-arrival (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 *"neuralnetwork"* to monitor potential DOA for prediction [82]. As knowledge assets played an importantrole 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[83].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 [84]. Common tools used for classification are decision trees, neural network, Bayesian network and rough set theory.

The data mining is used emerging trends in the education system [85-87] inthe whole world: In Indian culture most of the parents are uneducated .The main aim of in Indian government is thequality education not for quantity. But the day by day the education systems are changed and inthe 21st century a huge number of universalities are established by the order of UGC. As thenumbers of universities are established side by side, each and every day a millennium of studentsare enrolls across the country. With huge number of higher education aspirants, we believe thatdata mining technology can help bridging knowledge gap in higher educational systems. Thehidden patterns, associations, and anomalies that are discovered by data mining techniques fromeducational data can improve decision making processes in higher educational systems. Thisimprovement can bring advantages such as maximizing educational system efficiency, decreasingstudent's drop-out rate, and increasing student's promotion rate, increasing student's retention ratein, increasing student's transition rate, increasing educational improvement ratio, increasingstudent's success, increasing student's learning outcome, and reducing the cost of systemprocesses. In this current era we are using the KDD and the data mining tools for extracting theknowledge this knowledge can be used for improving the quality of education .The decisions treeclassification is used in this type of applications [88].

Data Mining Applications can be generic or domain specific: Data mining system can be applied for generic or domain specific. Some generic data miningapplications cannot take its own these decisions but guide users for selection of data, selection ofdata mining method and for the interpretation of the results. The multi agent based data mining application [89, 90] has capability of automatic selection of data mining technique to be applied. TheMulti Agent System used at different levels [90]: First, at the level of concept hierarchydefinition then at the result level to present the best adapted decision to the user. This decision isstored in knowledge

Base to use in a later decision-making. Multi Agent System Tool used forgeneric data mining system development [91] uses different agents to perform different tasks.

A multi-tier data mining system is proposed to enhance the performance of the data mining process [91]: It has basic components like user interface, data mining services, data access services and thedata. There are three different architectures presented for the data mining system namely one-tire,Two-tire and Three-tier architecture. Generic system required to integrate as many learningalgorithms as possible and decides the most appropriate algorithm to use. CORBA (CommonObject Request Broker Architecture) has features like: Integration of different applications coded inany programming language considerably easy. It allows reusability in a feasible way and finally itmakes possible to build large and scalable system. The data mining system architecture basedon CORBA is given by Object Management Group [91] has all characteristics to accomplish adistributed and object oriented computation. A data-centric focus and automated methodologiesmakes data mining accessible to no experts [92]. The use of high-level interfaces canimplement the automated methodologies that hide the data mining concepts away from theusers. A data-centric design hides away all the details of mining methodology and exposes themthrough high-level tasks that are goal-oriented. These goal-oriented tasks are implementedusing data-centric APIs. This design makes data mining task like other types of queries thatusers perform on the data. In data mining better results could be obtained if large data isavailable. It leads to the merging and linking of local databases. A new data-miningarchitecture based on Internet technology addressed this problem. [93] The context factor playsvital role in the success of data mining. The importance and meaning of same data in the differentcontext is different. A data in one context is very important may not be much important in othercontext. A context-aware data-mining framework filters useful and interesting context factors, and can produce accurate and precise prediction using those factors [94].

Application of Data Mining techniques in CRM: Data mining technique is used in CRM .Now a days it is one of the hot topic to research in theindustry because CRM have attracted both the practitioners and academics. It aims to give aresearch summary on the application of data mining in the CRM domain and techniques whichare most often used. Although this review cannot claim to be exhaustive, it does providereasonable insights and shows the incidence of research on this subject. The results presented inthis paper have several important implications: Research on the application of data mining inCRM will increase significantly in the future based on past publication rates and the increasinginterest in the area. The majority of the reviewed articles relate to customer retention [95].

In language research and language: Engineering much time extra linguistic informationis needed about a text. A linguistic profile that contains large number of linguistic features can begenerated from text file automatically using data mining [96]. This technique found quiteeffective for authorship verification and recognition. A profiling system usingcombination oflexical and syntactic features shows 97% accuracy in selecting correct author for the text. Thelinguistic profiling of text effectively used to control the quality of language and for the automatic language verification. [97] This method verifies automatically the text is of native quality. Theresults show that language verification is indeed possible.

In Medical Science: In medical science there is large scope for application of data mining. Diagnosis of diesis, healthcare, patient profiling and history generation etc. are the few examples. Mammography is themethod used in breast cancer detection. Radiologists face lot of difficulties in detection of tumors

that's why CAM (Computer Aided Methods) could help to the medical staff. So that they canproduce the good quality of the result detection [98]. The neural networks with back-propagationand association rule mining used for tumor classification in mammograms. The data miningeffectively used in the diagnosis of lung abnormality that may be cancerous or benign [99]. Thedata mining algorithms significantly reduce patient's risks and diagnosis costs. Using theprediction algorithms the observed prediction accuracy was 100% for 91.3% cases. The use ofdata mining in health care is the widely used application of data mining. The medical data is complex and difficult to analyze. A REMINDS (Reliable Extraction and Meaningful Inferencefrom Non-structured Data) system [100] integrates the structured and unstructured clinical data inpatient records to automatically create high quality structured clinical data. To adopt the highquality technique, we can mine the existing patient records to support guidelines and givecompliance to improve patient care. [100]

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.

Refertences:

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

2. Hemlata Sahu, Shalini Shrma, A Brief Overview on Data Mining Survey, Issue 3, ISSN 2249- 6343.

3. Kapil Sharma, Sheveta Vashisht, Heena Sharma, Richa Dhiman, Jasreena Kaur Bains,A Hybrid Approach Based On

Association Rule Mining and Rule Induction in Data Mining, ISSN: 2231-2307.

4. Mrs. Sayantani Ghosh1, Mr. Sudipta Roy2, A tutorial review on Text Mining Algorithms, ISSN: 2278 – 1021.

5. R. Brachman and T. Anand, "The process of knowledge discovery in databases: A human-centered approach," in *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, S. P. Amith, and R. Uthurusamy, Eds. Cambridge, MA: MIT Press, 1996, pp. 37-58.

6. U. Fayyad, "Data mining and knowledge discovery: Making sense out of data," *IEEE Expert*, vol. 11, pp. 20–25, Oct. 1996.

7. M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools," *SIGKDD Explorations*, vol. 1, no. 1, pp.20–33, June 1999.

8. D. Braha, *Data Mining for Design and Manufacturing: Methods and Applications*. Norwell, MA: Kluwer, 2000.

9. 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

10. 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

11. Wang, H. & Wang, S. (2008). A knowledge management approach to data mining process for business intelligence. Industrial Management & Data Systems, 108(5), 622-634.

12. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.AI Magazine, 17(3), 37-54.

13. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: An overview," in *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, S. P. Amith, and R. Uthurusamy, Eds. Cambridge, MA: MIT Press, 1996, pp. 1-36

14. Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data, volume 5, ISSN 1307-6884," Proceedings of World Academy of Science, Engineering and Technology, April 2005.

15. R. Brachman and T. Anand, "The process of knowledge discovery in databases: A human-centered approach," in *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G. Piatetsky-Shapiro, S. P.Amith, and R. Uthurusamy, Eds. Cambridge, MA: MIT Press, 1996, pp. 37-58.

16. Han, j. and M. Kamber, Data Mining Concepts and Techniques. 2006: Morgan Kaufmann Publishers.

17. Lee, I.-N., S.-C. Liao, and M. Embrechts, Data mining techniques applied to medical information. Med. Inform 2000.

18. Obenshain, M.K., Application of Data Mining Techniques to Healthcare Data. Infection Control and Hospital Epidemiology, 2004.

19. Sandhya, J., et al., Classification of Neurodegenerative Disorders Based on Major Risk Factors Employing Machine Learning Techniques, International Journal of Engineering and Technology, 2010. Vol.2, No.4.

20. Thuraisingham, B., A Primer for Understanding and Applying Data Mining. IT Professional IEEE, 2000.

21. Joseph, Zernik, "Data Mining as a Civic Duty – Online Public Prisoners

Registration Systems", International Journal on Social Media: Monitoring, Measurement, Mining, vol. - 1, no.-1, pp. 84-96, September-2010.

22. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.AI Magazine, 17(3), 37-54.

23. Han, J. & Kamber, M. (2012). Data Mining: Concepts and Techniques. 3rd.ed. Boston: Morgan Kaufmann Publishers.

24. Gorunescu, F. (2011). Data Mining: Concepts, Models, and Techniques. India: Springer.

25. Maryam Alavi, The John and Lucy Cook Chair of Information Strategy, Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues.

26. Jonassen , D. H. , K. Beissner , and M. A. Yacci . 1993 .*Structural knowledge: Techniques for conveying, assessing and acquiring structural knowledge.* Hillsdale, NJ : Lawrence Erlbaum Associates .

27. M Monica Sbaffoni, "Knowledge management approaches and definition", Nuclear Knowledge Management Unit, NE Department IAEA.

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

29. Jelena Rašul a1, Vesna Bosilj Vu kšić2, Mojca Indihar Štemb erg er3, THE IMPACT OF KNOWLEDGEMANAGEMENT ON ORGANISATIONALPERFORMANCE.

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

31. Hsinchun Chen1, Sherrilynne S. Fuller2, Carol Friedman3, and William Hersh, "KNOWLEDGE MANAGEMENT, DATAMINING, AND TEXT MINING INMEDICALINFORMATICS.

32. Ahmad, S., Gromiha, M. M., and Sarai, A. (2004). "Analysis and Prediction of DNA-binding Proteins and Their Binding Residues Based on Composition, Sequence, and Structural Information," *Bioinformatics*, 20(4), 477-486.

33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research*, 25(17), 3389-3402.

34. Antani, S., Lee, D. J., Long, L. R., and Thoma, G. R. (2004). "Evaluation of Shape Similarity Measurement Methods for Spine X-ray Images," *Journal of Visual Communication and Image Representation*, 15, 285-302.

35. Baclawski, K., Cigna, J., Kokar, M. W., Mager, P., and Indurkhya, B. (2000). "Knowledge Representation and Indexing Using the Unified Medical Language System," in *Proceedings of the Pacific Symposium on Biocomputing*, 493-504.

36. Ahn, J. H. & Chang, S. G. (2004). Assessing the contribution of knowledge to business performance: the KP3methodology. *Decision Support Systems*, 36 (4), 403–416.

37. Čater, T. &Čater, B. (2009). Tangible Resources as Antecedents of a Company's Competitive Advantage andPerformance. *Journal for East European Management Studies*, 14 (2), 186–209.

38. Lee, K. J. & Yu, K. (2004). Corporate culture and organizational performance. *Journal of Managerial Psychology*,19

(4), 340–359.

39. Lin, Z. (2000). Organizational restructuring and the impact of knowledge transfer. *The Journal of MathematicalSociology*, 24 (2), 129–149.

40. Škerlavaj, M. et al. (2006). Organizational learning culture – the missing link between business processchange and organizational performance. *International Journal of Production Economics*, 346-367.

41. 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

42. HIAN CHYE KOH , School of Business, SIM University, Singapore.

43. 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

44. 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

45. 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

46. 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

47. 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

48. 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.

49. 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

50. 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

51. Educational Data Mining: An Emerging Trends in Education,International Journal of Advanced Research in Computer Science ,ISSN NO- 0976-5697.

52. A comparative Study for Predicting Student's performance , Surjeet Kumar Yadav 1, Brijesh Bharadwaj2, Saurabh Pal3, ,International Journal of Innovative technology and creative engineering (ISSN:2045-711) VOL.1 NO.12 DECEMBER 2011.

53. Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent:An Application to Spatial Data, volume 5, ISSN 1307-6884," Proceedings of World Academy ofScience, Engineering and Technology, April 2005.

54. Educational Data Mining: An Emerging Trends in Education,International Journal of AdvancedResearch in Computer Science, ISSN NO- 0976-5697.

55. 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.

56. 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.

57. 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

58. 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.

59. Elovici, Y., Kandel, A., Last, M., Shapira, B., Zaafrany, O., "Using Data Mining Techniques for Detecting Terror-Related Activities on the Web".WebSite: www.ise.bgu.ac.il/faculty/mlast/papers/JIW_Paper.pdf.

60. Solieman, O. K., "Data Mining in Sports: A Research Overview, A Technical Report, MIS Masters Project, August 2006". Web Site: http://ai.arijona.edu/hchen/chencourse/Osama-DM_in_Sports.pdf.

61. Maciag, T., Hepting, D. H., Slezak, D., Hilderman, R. J., "Mining Associations for Interface Design". Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Volume 4481, pp.109-117, June 26, 2007.

62. Foster, D. P. and Stine, R. A., "Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy". Journal of the American Statistical Association, Alexandria, VA, ETATS-UNIS, vol.99, ISSN 0162-1459, pp. 303-313 January 15, 2004.

63. Kraft, M. R., Desouza, K. C., Androwich, I., "Data Mining in Healthcare Information Systems: Case Study of a Veterans' Administration Spinal Cord Injury Population".7695-1874-5/03, 2002.

64. Elovici, Y., Kandel, A., Last, M., Shapira, B., Zaafrany, O., "Using Data Mining Techniques for Detecting Terror-Related Activities on the Web".WebSite: www.ise.bgu.ac.il/faculty/mlast/papers/JIW_Paper.pdf.

65. DeBarr, D.,Eyler-Walker, Z., "Closing the Gap: Automated Screening of Tax Returns to Identify Egregious Tax Shelters". SIGKDD Explorations Volume 8, Issue 1.

66. Ansari, S., Kohavi, R., Mason, L., and Zheng, Z., "Integrating E-Commerce and Data Mining: Architecture and Challenges" .Proceedings of IEEE International Conference on Data Mining, 2001.

67. Jadhav, S. R., and Kumbargoudar, P., "Multimedia Data Mining in Digital Libraries: Standards and Features READIT 2007, pp 54-59.

68. Kusiak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A., and Tseng, T.L., "Data Mining: Medical And Engineering Case Studies". Proceedings of the Industrial Engineering Research 2000 Conference, Cleveland, Ohio, pp. 1-7,May 21-23, 2000.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, ohn Wiley & Sons, Inc, 2005.

76. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R... "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".

77. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases.AI Magazine, 17(3), 37-54.
\

78. Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006.
\

79. 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

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. Educational Data Mining: An Emerging Trends in Education, International Journal of Advanced Research in Computer Science, ISSN NO- 0976-5697.

86. A comparative Study for Predicting Student's performance, Surjeet Kumar Yadav 1, Brijesh Bharadwaj2, Saurabh Pal3, International Journal of Innovative technology and creative engineering (ISSN: 2045-711) VOL.1 NO.12 DECEMBER 2011.

87. J. A. Harding, M. Shahbaz, Srinivas, A. Kusiak Journal of Manufacturing Science and Engineering NOVEMBER 2006, Vol. 128 / 969.

88. Neelamadhab Padhy1, Dr. Pragnyaban Mishra 2, and Rasmita Panigrahi3,The survey of Data mining Applications And Feature Scope, (IJCSEIT), Vol.2, No.3, June 2012.

89. Rantzau, R. and Schwarz, H., "A Multi-Tier Architecture for High-Performance Data Mining,A Technical Project Report of ESPRIT project, The consortium of CRITIKAL project, Attar Software Ltd. (UK), Gehe AG (Denmark); Lloyds TSB Group (UK), Parallel Applications Centre, University of Southampton (UK), BWI,

University of Stuttgart (Denmark), IPVR, University of Stuttgart (Denmark)".

90. Botia, J. A., Garijo, M. y Velasco, J. R., Skarmeta, A. F., "A Generic Data mining System basic design and implementation guidelines", A Technical Project Report of CYCYTprojectofSpanishGovernment.1998 .WebSite: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.1935.

91. Campos, M. M., Stengard, P. J., Boriana, L. M., "Data-Centric Automated Data Mining",Website:www.oracle.com/technology/products/bi/odm/pdf/automated_data_mining_paper_1205.pdf.

92. Sirgo, J., Lopez, A., Janez, R., Blanco, R., Abajo, N., Tarrio, M., Perez, R., "A Data Mining Engine based on Internet, Emerging Technologies and Factory Automation," ProceedingsETFA'03,IEEEzConference,16-09-2003,WebSite:www.citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.8955.

93. Bianca V. D.,Philippe Boula de Mareüil and Martine Adda-Decker, "Identification of foreign-accented French using data mining techniques, Computer Sciences Laboratory for Mechanics and EngineeringSciences(LIMSI)".Website: www.limsi.fr/Individu/bianca/article/Vieru&Boula&Madda_ParaLing07.pdf.

94. Vajirkar, P., Singh, S., and Lee, Y., "Context-Aware Data Mining Framework for Wireless Medical Application". 381 – 39.

95. QIANG YANG, XINDONG WU, 10CHALLENGING PROBLEMS IN DATA MINING RESEARCH, International Journal of Information Technology & Decision Making, Vol. 5, No. 4 (2006) 597–604.

96. Halteren, H. V., Oostdijk N., "Linguistic profiling of texts for the purpose of language verification, The ILK research group, Tilburg centre for Creative Computing and the Department of Communication and Information Sciences of the Faculty of Humanities,TilburgUniversity,The Netherlands. Website: www.ilk.uvt.nl/~antalb/textmining/LingProfColingDef.pdf.

97. Antonie, M. L., Zaiane, O. R.,Coman, A., "Application of Data Mining Techniques for Medical Image Classification", Proceedings of the Second International Workshop on Multimedia Data Mining MDM/KDD 2001) in conjunction with ACM SIGKDD conference, San Francisco, August 26,2001.

98. Kusiak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A., and Tseng, T.L., "Data Mining: Medical And Engineering Case Studies". Proceedings of the Industrial Engineering Research 2000 Conference, Cleveland, Ohio, pp. 1-7,May 21-23, 2000.

99. Luis, R., Redol, J., Simoes, D., Horta, N., "Data Warehousing and Data Mining System Applied to ELearning, Proceedings of the II International Conference on Multimedia and Information & Communication Technologies in Education, Badajoz, Spain, December 3-6th 2003.

100. Rao, R. B., Krishnan, S. and Niculescu, R. S., "Data Mining for Improved Cardiac Care", SIGKDD Explorations Volume 8, Issue 1.