ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 157- 167 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

DATA MINING - A DOMAIN SPECIFIC ANALYTICAL TOOL FOR DECISION MAKING

Ms. Somanjoli Mohapatra[1], Dr. D Ramesh[2], Mr. Chinmaya Dash[3], Mr. Prakash Chandra Behera[4]

[1]Assistant Professor, Department of Computer Science, St. Claret College,Bangalore, Karnataka, India
somanjoli.mohapatra@gmail.com
[2]Professor & HOD, Department of MCA, SSIT, Tumkur, Karnataka, India
rameshd_ssit@yahoo.com
[3]Assistant Professor, Department of Computer Science, St. Claret College, Bangalore, Karnataka, India
da.chinmaya@gmail.com
[4]Assistant Professor, Department of Computer Science, St. Claret College, Bangalore, Karnataka, India
prakasbehera@gmail.com

ABSTRACT

In 21st century, human beings use huge in day-to-day transaction in various fields.These data may be in the form of documents, graphical formats,the video or records.To analyse these data and hence forth taking effective managerial decision, the techniques- data warehouse and data mining are used. To analyze, manage and make a decision of such type of huge amount of data we need techniques called data mining. Data Mining can be defined as the process of extracting interesting, interpretable, useful and novel information from data. It has been used by businesses, scientists and governments to sift through volumes of data like airline passenger records, census data and the supermarket scanner data. Data mining is also known as Knowledge Discovery in Databases (KDD) which is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. This paper focuses on applications of data mining.

Keywords: Data mining life cycle, Data mining Methods, KDD, Visualization of the data mining model

1. INTRODUCTION

In information technology the discovery of useful information from large collections of data industry towards field of "Data mining" is due to the perception of "*we are data rich but information poor*". There is huge volume of data but we hardly able to turn them in to useful information and knowledge for managerial decision making in business. To generate information it requires massive collection of data. It may be in different formats like audio/video, numbers, text, figures and Hypertext formats. To take complete advantage of data; the data retrieval is simply not enough, it requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data. With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, to develop powerful tool for analysis and interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. The only answer to all above is 'Data Mining'. Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, helps organizations to make proactive knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by prospective tools typical of decision support systems. Data mining tools find and produce hidden patterns, information that experts may miss because it lies outside their expectations.

2. DATA MINING TASK

The data mining tasks are classified as:

2.1 Exploratory Data Analysis:

This data mining task will serve the two purposes which are interactive and visual to customer.

- without the knowledge for what the customer is searching
- it analyses the data

2.2 DescriptiveModelling:

It describes overall probability distribution of the data, partitioning of the p-dimensional space into groups and models describing the relationships between the variables.

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

2.3 Predictive Modelling:

This model permits the value of one variable to be predicted from the known values of other variables.

2.4 Discovering Patterns and Rules:

This task is primarily used to find the hidden pattern as well as to discover the pattern in the cluster. In a cluster a number of patterns of different size and clusters are available.This can be accomplished by using rule induction and different algorithm called clustering algorithm.

2.5 Retrieval by Content:

The primary objective of this task is to find the data sets of frequently used for audio/video as well as images.

3. TYPES OF DATA MINING SYSTEM

Data mining systems can be categorized as follows:

3.1 Classification of data mining systems according to the type of data source mined:

In an organization huge amount of data are available where we need to classify those data according to its type(may be audio/video, text format).

3.2 Classification of data mining systems according to the data model:

There are number of data mining models (Relational data model, Object Model, Object Oriented data Model, Hierarchical data Model) are available. According to these data model the data mining system classify the data into various models.

3.3 Classification of data mining systems according to the kind of knowledge discovered:

This classification based on the knowledge discovered or data mining functionalities, such as characterization, discrimination, association, classification, clustering, etc. Some systems tend to be comprehensive systems offering several data mining functionalities together.

3.4 Classification of data mining systems according to mining techniques used:

This classification is according to the data analysis approach used such as machine learning, neural networks, genetic algorithms, statistics, visualization, database oriented or data warehouse-oriented, etc. The classification can also take into account the degree of user interaction involved in the data mining process such as query-driven systems, interactive exploratory systems, or autonomous systems. A comprehensive system would provide a wide variety of data mining techniques to fit different situations and options.

4. DATA MINING LIFE CYCLE

The life cycle of a data mining consists of six phases.

4.1 Business Understanding:

This phase focuses on understanding the objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

4.1.1 Determine business objectives
The first objective of the data analyst is to understand thoroughly, from a business perspective, what the client really wants to accomplish. The analyst's goal is to uncover important factors, at the beginning, that can influence the outcome of the project.
4.1.2 Assess situation
This task involves more detailed fact-finding about all of the resources, constraints, assumptions and other factors that should be considered in determining the data analysis goaland project plan.
4.1.3 Determine data mining goals
A *business goal* states objectives in business terms. A *data mining goal* states project objectives in technical terms.
4.1.4 Produce project plan
Describe the intended plan for achieving the data mining goals and thereby achieving the business goals. The plan should specify the anticipated set of steps to be performed during the rest of the project including an initial selection of tools and techniques.

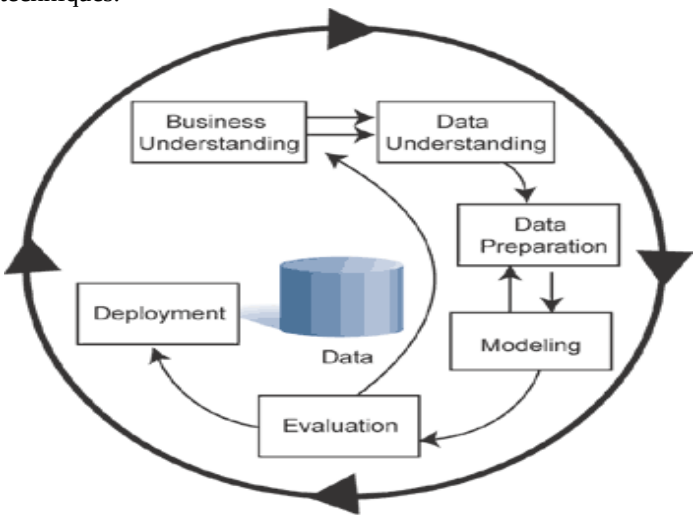

FIGURE 1 Data-Mining Process Model

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

4.2 Data Understanding:

It starts with an initial data collection, to get familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

4.2.1 Collect initial data
Acquire within the project the data listed in the project resources. This initial collection includes data loading if necessary for data understanding. This effort may lead to initial data preparation steps.

4.2.2 Describe data
Examine properties of acquired data and report on the results.

4.2.3 Explore data
This task tackles the data mining questions that can be addressed using querying, visualization and reporting. These analyses may address the data mining goals directly.

4.2.4 Verify data quality
Examine the quality of the data.

4.3 Data Preparation:

This phase collects different data sets and constructs the varieties of the activities based on the initial raw data.

4.3.1 Select data
Decide on the data to be used for analysis. Criteria include relevance to the data mining goals, quality and technical constraints such as limits on data volume or data types.

4.3.2 Clean data
Data cleaning may involve selection of clean subsets of the data, the insertion of suitable defaults or more ambitious techniques such as replacing the dirty data with derived values, or building separate models for those entities that possess dirty data.

4.3.3 Construct data
This task includes constructive data preparation operations such as the production of derived attributes, entire new records, or transformed values for existing attributes.

4.3.4 Integrate data
Two methods used for integrating data are merging data and generating aggregate values. In these methods information is combined from multiple tables or other information sources to create new records or values.

4.3.5 Format data
Formatting transformations refer to primarily syntactic modifications made to the data that do not change its meaning, but might be required by the modeling tool.

4.4 Modelling:

In this phase, various modelling techniques are selected and applied and their parameters are calibrated to optimal values.

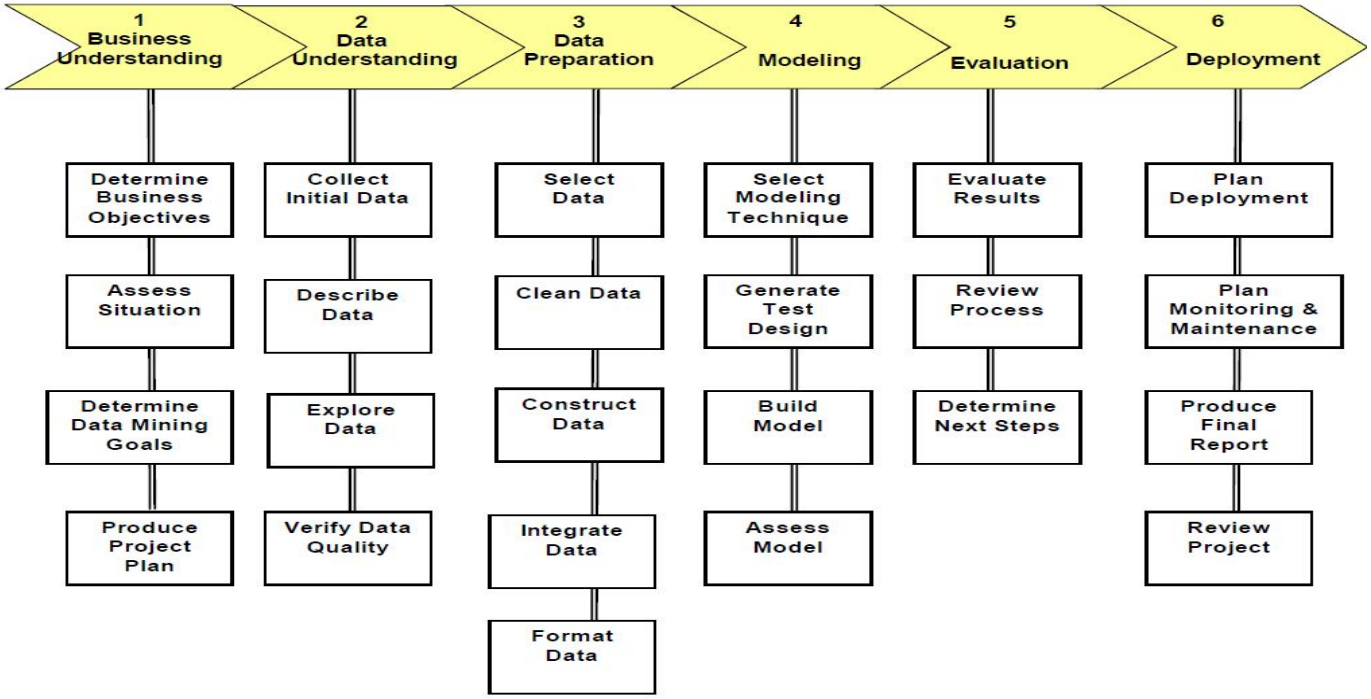

FIGURE 2: Details of Data Mining Process Model

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 157- 167 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

4.4.1 Select modeling technique
As the first step in modeling, select the actual modeling technique to be used. If a tool was selected in business understanding (Phase 1), this task refers to selecting the specific modeling technique, e.g., building decision trees or generating a neural network.
4.4.2 Generate test design
Prior to building a model, a procedure needs to be defined to test the model's quality and validity. If the test design specifies that the dataset should be separated into training and test sets, the model is built on the training set and its quality estimated on the test set.
4.4.3 Build model
The purpose of building models is to use the predictions to make more informed business decisions. The most important goal when building a model is stability, which means that the model should make predictions that will hold true when it's applied to yet unseen data.
4.4.4 Assess model
The model should now be assessed to ensure that it meets the data mining success criteria and passes the desired test criteria. This step is a purely technical assessment based on the outcome of the modeling tasks.

4.5 Evaluation:

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

4.5.1 Evaluate results
Previous evaluation steps dealt with factors such as the accuracy and generality of the model. This step assesses the degree to which the model meets the business objectives and seeks to determine if there is some business reason why this chosen model is deficient.
4.5.2 Review process
At this point the resultant model appears to be satisfactory and appears to satisfy business needsat this stage of Data Mining, the Review Process takes on the form of a Quality Assurance.
4.5.3 Determine next steps
According to the assessment results and the process review, the analyst decides how to proceed at this stage. The analyst needs to decide whether
• to finish the project and move on to deployment (Phase 6)
• to initiate further iterations or
• to set up new data mining projects.

4.6 Deployment:

This phase is used to increase knowledge and further the knowledge will be organized and presented in a way that the customer can use it. The deployment phase can be generating a report or as complex as implementing a repeatable data mining process across the enterprise.

4.6.1 Plan deployment
To deploy the data mining result(s) into the business, this task takes the evaluation results and develops a strategy for deployment. If a general procedure was identified to create the relevant model(s), this procedure is documented here for later deployment.
4.6.2 Plan monitoring and maintenance
Monitoring and maintenance are important issues if the data mining result becomes part of the day-to-day business and its environment. To monitor the deployment of the data mining result(s), the project needs a detailed plan on the monitoring process.
4.6.3 Produce final report
At the end of the project, the project leader and the team write up a final report. Depending on the deployment plan, this report may be only a summary of the project and its experiences or it may be a final and comprehensive presentation of the data mining result(s).
4.6.4 Review project
Assess what went right and what went wrong, what was done well and what needs to be improved.

5. VISUALIZING DATA MINING MODEL

The main objective of data visualization is the overall idea about the data mining model.In data mining most of the times we are retrieving the data from the repositories which are in the hidden form. So visualization of the data mining model helps us to provide levels of understanding and trust. The data mining models are of two types: Predictive and Descriptive.

5.1 Predictive Model:

It makes prediction about unknown data values by using the known values. Ex. Classification, Regression, Time series analysis, Prediction etc. Many of the data mining applications are aimed to predict the future state of the data. Prediction is the process of analysing the current and past states of the attribute and prediction of its future state.

Classification is a technique of mapping the target data to the predefined groups or classes, this is a supervise learning because the classes are predefined before the examination of the target data. The regression involves the learning of function that map data item to real valued prediction variable. In the time series analysis the value of an attribute is examined as it varies over time. In time series analysis is used for many statistical techniques which will analyse the time-series data such as auto regression methods etc.

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

5.2 Descriptive model:

It identifies the patterns or relationships in data and explores the properties of the data examined. Ex. Clustering, Summarization, Association rule, Sequence discovery etc.

The term clustering means analysing the different data objects without consulting a known class levels. It is also referred to as unsupervised learning or segmentation. It is the partitioning or segmentation of the data in to groups or clusters. The clusters are defined by studying the behaviour of the data by the domain experts. The term segmentation is a process of partitioning of database into disjoint grouping of similar tuples. Summarization is the technique of presenting the summarize information from the data. The association rule finds the association between the different attributes. Association rule mining is a two-step process: Finding all frequent item sets, Generating strong association rules from the frequent item sets. Sequence discovery is a process of finding the sequence patterns in data. This sequence can be used to understand the trend.

6. METHODS OF DATA MINING

Data mining methods are broadly classified as: On-Line Analytical Processing,(OLAP), Classification, Clustering, Association Rule Mining, Temporal Data Mining, Time Series Analysis, Spatial Mining, Web Mining etc. These methods use different types of algorithms and data. The data source can be data warehouse, database, flat file or text file. The algorithms may be Statistical Algorithms, Decision Tree based, Nearest Neighbour, Neural Network based, Genetic Algorithms based, Ruled based, Support Vector Machine etc.

6.1 Classification:

It is learning a function that maps (classifies) a data item into one of several predefined classes

6.2 Regression:

It is learning a function that maps a data item to a real-valued prediction variable.

6.3 Clustering:

It is a common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data.

6.4 Summarization:

It involves methods for finding a compact description for a subset of data.

6.5 Dependency modelling:

It consists of finding a model that describes significant dependencies between variables. Dependency models exist at two levels: (1) the structural level of the model specifies which variables are locally dependent on each other and (2) the quantitative level of the model specifies the strengths of the dependencies using some numeric scale.

6.6 Change and deviation detection:
It focuses on discovering the most significant changes in the data from previously measured or normative values.

6.7 Decision trees and rules:

It use univariate splits have a simple representational form, making the inferred model relatively easy for the user to comprehend. However, the restriction to a particular tree or rule representation can significantly restrict the functional form (and, thus, the approximation power) of the model.

6.8 Nonlinear Regression and Classification Methods:

These methods consist of a family of techniques for prediction that fit linear and nonlinear combinations of functions to combinations of the input variables.

6.9 Probabilistic Graphic Dependency Models:

Graphic models specify probabilistic dependencies using a graph structure. The model specifies which variables are directly dependent on each other. Typically, these models are used with categorical or discrete-valued variables, but extensions to special cases, such as Gaussian densities, for real-valued variables are also possible.

6.10 Relational Learning Models:

Relational learning (also known as inductive logic programming) uses the more flexible pattern language of first-order logic. A relational learner can easily find formulas such as X = Y. Most research to date on model-evaluation methods for relational learning is logical in nature.

Generally the data mining algorithms are fully dependent of the two factors these are

(i) Which type of data sets are using
(ii) What type of requirements are needed for user

Knowledge discovery (KD) process involves pre-processing data, choosing a data-mining algorithm, and post processing the mining results. The Intelligent Discovery Assistants (IDA), helps users in applying valid knowledge discovery processes.

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

7. DATA MINING ELEMENTS

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table

8. COMPONENT OF DATA MINING ALGORITHM

There are three primary components in data-mining algorithm:

Model representation, Model evaluation, and Search.

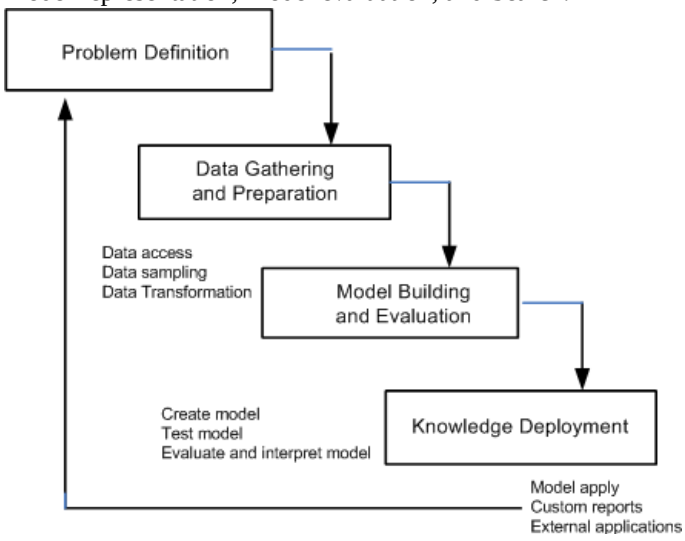

FIGURE 3: Data Mining Model

8.1 Model representation:

It is the language used to describe discoverable patterns. If the representation is too limited, then no amount of training time or examples can produce an accurate model for the data. It is important that a data analyst fully comprehend the representational assumptions that might be inherent in a particular method and also algorithm designer clearly state which representational assumptions are being made by a particular algorithm.

8.2 Model evaluation:

It criteria are quantitative statements of how well a particular pattern (a model and its parameters) meets the goals of the KDD process. For example, predictive models are often judged by the empirical prediction accuracy on some test set where descriptive models can be evaluated along the dimensions of predictive accuracy, novelty, utility, and understandability of the fitted model.

8.3 Search method:

It consists of two components: (1) parameter search and (2) model search. Once the model representation and the model-evaluation criteria are fixed, then the data-mining problem has been reduced to purely an optimization task: Find the parameters and models from the selected family that optimize the evaluation criteria. In parameter search, the algorithm must search for the parameters that optimize the model-evaluation criteria given observed data and a fixed model representation.

9. KNOWLEDGE DISCOVERY FROM DATABASE(KDD)

There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. The main KDD application areas are marketing, finance or investment, fraud detection, manufacturing, telecommunications, and Internet agents etc. The term data mining has mostly been used by statisticians, data analysts, and the management information systems (MIS) communities. It has also gained popularity in the database field.

So, KDD is a process of mapping low-level data into other forms that might be more compact, more abstract, or more useful. KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data. The unifying goal is extracting high-level knowledge from low-level data in the context of large data sets. KDD focuses on the overall process of knowledge discovery from data, including how the data are stored and accessed, how algorithms can be scaled to massive data sets and still run efficiently, how results can be interpreted and visualized, and how the overall man-machine interaction can usefully be modeled and supported.

KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Here, data are a set of facts, and pattern is an expression in some language describing a subset of the data or a model applicable to the subset. The term process implies that KDD comprises many steps, which involve data preparation, searchfor patterns, knowledge evaluation, and refinement, all repeated in multiple iterations. The process consists nine steps:

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

9.1 Develop an application:

Develop an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the customer's viewpoint.

9.2 Create a target data set:

Selecting a data set, or focusing on a subset of variables or data samples, on which discovery is to be performed.

9.3 Data cleaning and pre-processing:

Basic operations include removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields, and accounting for time-sequence information and known changes.

9.4 Data reduction and projection:

Finding useful features to represent the data depending on the goal of the task. With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found.

9.5 Goal of KDD process:

Match the goals of the KDD process (step 1) to a particular data-mining method. For example, summarization, classification, regression, clustering, and so on.

9.6 Exploratory analysis and model and hypothesis selection:

Choosing the datamining algorithm(s) and selecting method(s) to be used for searching for data patterns. This process includes deciding which models and parameters might be appropriate and matching a particular data-mining method with the overall criteria of the KDD process

9.7 Data mining:

Searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression, and clustering. The user can significantly aid the data-mining method by correctly performing the preceding steps.

9.8 Pattern interpretation:

Interpreting mined patterns, possibly returning to any of steps 1 through 7 for further iteration. This step can also involve visualization of the extracted patterns and models or visualization of the data given the extracted models.

9.9 Working on the discovered knowledge:

Using the knowledge directly, incorporating the knowledge into another system for further action, or simply documenting it and reporting it to interested parties. This process also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge.

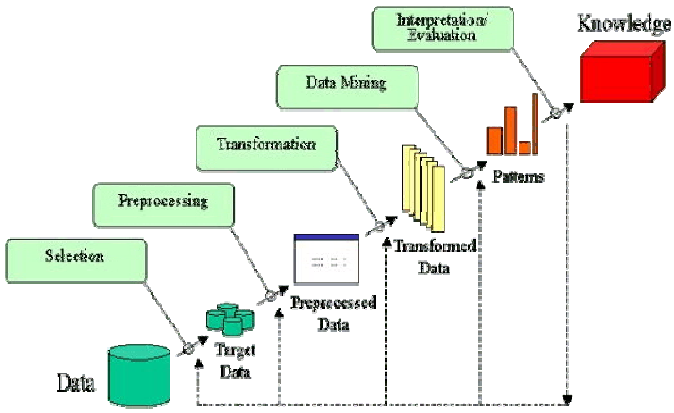

FIGURE 4: KDD Process

10. DATA MINING AND KDD

KDD process is one of mapping of low-level data into other forms that might be more compact, more abstract, or more useful. Data mining is a step in the KDD process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of patterns (or models) over the data.

Online Analytical Processing (OLAP) tool is used to apply data analysis in KDD process. OLAP tools focus on providing multidimensional data analysis, which is superior to SQL in computing summaries and breakdowns along many dimensions. OLAP tools are targeted toward simplifying and supporting interactive data analysis, but the goal of KDD tools is to automate asmuch of the process as possible. Thus, KDD is a step beyond what is currently supported by most standard database systems.

11. RULES OF DATA MINING

- Business objectives are the origin of every data mining solution (**Business Goals Law**)
- Business knowledge is central to every step of the data mining process (**Business Knowledge Law**)
- Data preparation is more than half of every data mining process (**Data Preparation Law**)

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

- The right model for a given application can only be discovered by experiment or There is No Free Lunch for the Data Miner (**NFL-DM**)
- There are always patterns (**Watkins Law**)
- Data mining amplifies perception in the business domain (**Insight Law**)
- Prediction increases information locally by generalisation (**Prediction Law**)
- The value of data mining results is not determined by the accuracy or stability of predictive models (**Value Law**)
- All patterns are subject to change (**Law of Change**)

12. DATA MINING APPLICATIONS

12.1 Data Mining Applications in Healthcare

The success of healthcare data mining hinges on the availability of clean healthcare data. In this respect, it is critical that the healthcare industry look into how data can be better captured, stored, prepared and mined. Possible directions include the standardization of clinical vocabulary and the sharing of data across organizations to enhance the benefits of healthcare data mining applications. As healthcare data are quantitative data, it is necessary to also explore the use of text mining to expand the scope and nature of what healthcare data mining can currently do. This is specially used to mixed all the data and then mining the text.

12.2Data mining for market basket analysis

Data mining technique is used in MBA (*Market Basket Analysis*). When the customer want to buy some products then this technique helps to find the associations between different items that the customer put in their shopping buckets. Here the discovery of such associations that promotes the business technique. In this way the retailers uses the data mining technique so that they can identify that which customers intension. In this way this technique is used for profits of the business and also helps to purchase the related items.

12.3The data mining in education system

With huge number of higher education aspirants, we believe that data mining technology can help bridging knowledge gap in higher educational systems. The hidden patterns, associations, and anomalies that are discovered by data mining techniques from educational data can improve decision making processes in higher educational systems. This improvement can bring advantages such as maximizing educational system efficiency, decreasing student's drop-out rate, and increasing student's promotion rate, increasing student's retention rate in, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes. In this current era we are using the KDD and the data mining tools for extracting the knowledge. This knowledge can be used for improving the quality of education.

12.4Data mining in manufacturing engineering

When we retrieve the data from manufacturing system then the customer use these data for different purposes like to find the errors in the data, to enhance the design methodology, to make the good quality of the data, how best the data can be supported for making the decision. But most of the time, the data can be first analysed then after finding the hidden patterns manufacturing process can be controlled to enhance the quality of the products.

.

12.5Data Mining Applications can be generic or domain specific.

Data mining system can be applied for generic or domain specific**.** The multi agent based data mining application has capability of automatic selection of data mining technique to be applied. The Multi Agent System used at different levels: First, at the level of concept hierarchy definition then at the result level to present the best adapted decision to the user. This decision is stored in knowledge Base to use in a later decision-making. Multi Agent System Tool used for generic data mining system development uses different agents to perform different tasks.

12.6A multi-tier data mining system

It consist basic components like user interface, data mining services, data access services and the data. There are three different architectures presented for the data mining system namely one-tire, Two-tire and Three-tire architecture. Generic system required to integrate as many learning algorithms as possible and decides the most appropriate algorithm to use. CORBA (*Common Object Request Broker Architecture*) allows reusability in a feasible way and finally it makes possible to build large and scalable system.

12.7Data mining technique in CRM

Data mining technique used in CRM aims to give a research summary on the application of data mining in the CRM domain and techniques which are most often used.

12.8The Domain Specific Applications

The domain specific applications are focused to use the domain specific data and data mining algorithm that targeted for specific objective. The applications are aimed to generate the specific knowledge. In the different domains the data generating sources generate different type of data. Data can be from a simple text, numbers to more complex audio-video

ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 147- 150 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp32.pdf

data. To mine the patterns and thus knowledge from this data, different types of data mining algorithms are used.

12.9In Medical Science

The use of data mining in health care is the widely used application of data mining. The medical data is complex and difficult to analyse. A REMIND (*Reliable Extraction and Meaningful Inference from Non-structured Data*) system integrates the structured and unstructured clinical data in patient records to automatically create high quality structured clinical data.

12.10Data Mining in the Web Education

Data mining methods are used in the web Education which is used to improve courseware. The relationships are discovered among the usage data picked up during students' sessions. This knowledge is very useful for the teacher or the author of the course, who could decide what modifications will be the most appropriate to improve the effectiveness of the course.

12.11The Intrusion Detection in the Network

The data mining method is used to classify the network traffic normal traffic or abnormal traffic. If any TCP header does not belong to any of the existing TCP header clusters, then it can be considered as anomaly. The data mining methods used to accurately detect malicious executables before they run.

12.12Sports data mining

In the world, a huge number of games are available where each and every day the national and international games are to be scheduled, where a huge number of data are to be maintained. The data mining tools are applied to give the information as and when we required. Data mining tools like WEKA and RAPID MINER are frequently used for sport. In the game sports the data are available in the statistical form where data mining can be used and discover the patterns, these patterns are often used to predict the future forecast. Data mining can be used for scouting, prediction of performance, selection of players, coaching and training and for the strategy planning.

12.13The Intelligence Agencies

The Intelligence Agencies collect and analyse information to investigate terrorist activities. One challenge to law enforcement and intelligent agencies is the difficulty of analysing large volume of data involve in criminal and terrorist activities. Now a day the intelligence agency are using the sophisticated data mining algorithms which makes it easy, to handle the very large databases for organizations. The different data mining techniques are used in crime data mining. In data mining the Clustering techniques are used for the different objects in crime records. Data mining detects and analyses the crime data. The classification technique is also used to detect email spamming and also find person who has given the mail.

12.14The data mining system in Internal Revenue Service

The data mining system implemented at the Internal Revenue Service to identify high-income individuals engaged in abusive tax shelters show significantly good results. The major lines of investigation included visualization of the relationships and data mining to identify and rank possibly abusive tax avoidance transactions. To enhance the quality of product data mining techniques can be used effectively.

12.15E-commerce

E-commerce is also the most prospective domain for data mining because data records are plentiful, 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 correct business decisions. This integration effectively solves several major problems associated with horizontal data mining tools including the enormous effort required in pre-processing of the data before it can be used for mining, and making the results of mining actionable.

12.16The Digital Library Retrieves

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

12.17The prediction in engineering applications

The prediction in engineering applications was treated effectively by a data mining approach. The prediction problems like the cost estimation problem in engineering, the problem 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 of engineering applications like prior data. Once we gather the data then we can generate the different models, algorithms which will predict different characteristic.

13. CONCLUSION

In this paper we briefly reviewed the various data mining applications. This review would be helpful to researchers to

focus on the various issues of data mining. Most of the data mining applications in various fields use the variety of data types range from text to images and stores in variety of databases and data structures. The different methods of data mining are used to extract the patterns and thus the knowledge from this variety databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. Several attempts have been made to design and develop the generic data mining system but no system found completely generic.

Thus, for every domain the domain expert's assistant is mandatory. The domain experts shall be guided by the system to effectively apply their knowledge for the use of data mining systems to generate required knowledge. The domain experts are required to determine the variety of data that should be collected in the specific problem domain, selection of specific data for data mining, cleaning and transformation of data, extracting patterns for knowledge generation and finally interpretation of the patterns and knowledge generation. Most of the domain specific data mining applications show accuracy above 90%. The generic data mining applications are having the limitations. From the study of various data mining applications it is observed that, no application called generic application is 100 % generic. The intelligent interfaces and intelligent agents up to some extent make the application generic but have limitations.

The domain experts play important role in the different stages of data mining. The decisions at different stages are influenced by the factors like domain and data details, aim of the data mining, and the context parameters. The domain specific applications are aimed to extract specific knowledge. The domain experts by considering the user's requirements and other context parameters guide the system. Therefore it is concluded that the domain specific applications are more specific for data mining.

14. REFERENCES

[1] **Introduction to Data Mining and Knowledge Discovery**, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

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

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

[4] 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".

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "**From Data Mining to Knowledge Discovery in Databases**," AI Magazine, American Association for Artificial Intelligence, 1996.

[6] Tan Pang-Ning, Steinbach, M., Vipin Kumar. "**Introduction to Data Mining**", *Pearson Education, New Delhi, ISBN: 978-81-317-1472-0, 3rd Edition, 2009.*

[7] Bernstein, A. and Provost, F., "**An Intelligent Assistant for the Knowledge Discovery Process**", Working Paper of the Center for Digital Economy Research, New York University and also presented at the IJCAI 2001 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases.

[8] 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.

[9] 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)".

[10] 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

[11] Agrawal, R., and Psaila, G. 1995. **Active Data Mining**. In Proceedings of the First International Conference on Knowledge Discovery and Data Mining(KDD-95), 3–8. Menlo Park, Calif.: American Association for Artificial Intelligence.

[12] Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.;and Verkamo, I. 1996. **Fast Discovery of AssociationRules.** In *Advances in Knowledge Discovery and DataMining*, eds. U. Fayyad, G. Piatetsky-Shapiro, P.Smyth, and R. Uthurusamy, 307–328. Menlo Park Calif.: AAAI Press.

[13] Brachman, R., and Anand, T. 1996. **The Process of Knowledge Discovery in Databases: A Human-Centered Approach**. In *Advances in Knowledge Discoveryand Data Mining*, 37–58, eds. U. Fayyad, G. Piatetsky- Shapiro, P. Smyth, and R. Uthurusamy. Menlo Park, Calif.: AAAI Press.

[14] Berry, M. J., Linoff, G. S. (2000), "**Mastering Data Mining: The Art and Science of Customer Relationship Management**". Wiley Computer Publishing, New York.

[15] Chung, H. M., Gray, P. (1999), "**Special Section: Data Mining**". *Journal of ManagementInformation Systems*, (16:1),11-17.

[16] Colin, S. (2000), "**The CRISP-DM Model: The New Blueprint for Data Mining**", *Journal ofData Warehousing*, (5:4), Fall, 13-22.

[17] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, R (1996). "**The KDD Process for Extracting Useful Knowledge from Volumes of Data**," *Communications of the ACM*, (39:11), pp.27-34.

[18] Fayyad, U., (2001), "**The Digital Physics of Data Mining**", *Communications of the ACM*,March, (44:3), 62-65.

[19] Han, J., Kamber, M. (2001), **Data Mining: Concepts and Techniques**, Morgan-KaufmannAcademic Press, San Francisco.

[20] Hand, D. J. (1998), "**Data Mining: Statistics and More**?",*The American Statistician*, May(52:2), 112-118.

[21] Ranjit, B., Sugumaran, V. (1999), "**Application of Intelligent Agent Technology for Managerial Data Analysis and Mining**", *Database for Advances in Information Systems*, (30:1), 77-94.

[22] Spangler, W. E.; May, J. H., Vargas, L. G. (1999), "**Choosing Data-Mining Methods For Multiple Classification: Representational And Performance Measurement Implications For Decision Support** "*Journal of Management Information Systems*, Summer, 37-62.

[23] White, H.,"**A Reality Check for Data Snooping**" (2000), *Econometrica*, (68:5), September,1097-1126.

[24] Witten, I. H. (2000), **Data mining: practical machine learning tools and techniques with Java implementations**, Morgan Kaufman, San Francisco.

[25] Srivastava, J., Cooley, R., Deshpande, M., Tan, P., "**Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data**", *ACM SIGKDD (Special Interest Group onKnowledge Discovery and Data Mining) Explorations*, January, (1:2)

[26] Kennedy, R. L., Lee, Y. Roy, B. V. Reed, C. D. &Lippman, R. P. (1997). **Solving Data Mining Problems Through Pattern Recognition**. New Jersey: Prentice Hall Professional TechnicalReference.

[27] Kosala, R., Blockeel, H. (2000), "**Web Mining Research: A Survey**", *ACM SIGKDD(Special Interest Group on Knowledge Discovery and Data Mining) Explorations*, June, (2:1), 1-10.