Wisdom Journal For Studies And Research volume 04 Issue 06(20) 15/11/2024~

ISSN print/ 2769-1926 ISSN online/ 2769-1934

Data Mining In Healthcare Sectors

Lecturer Alaa Hussain Al-Obaidi*
Babylon University
E_mail: ahaa_ma2009@uobabylon.edu.iq

Received :05/10/2024 Accepted : 08/11/2024

Abstract:

Data mining is a new technology that supported with discovering information of a large amount of data. It means how to mine the knowledge from data. Recently, this technical tool has attracted a great deal of attention in the society in general and at information industry in particular. In this paper, the researcher tried to focus on health care sector where there are serious challenges in the application. The use of Data mining in healthcare informatics and challenges will be also examined. This is a descriptive study that examines the concepts, related issues and techniques used in data mining in healthcare sector.

Keywords: Data mining, HealthCare.

Introduction

Data mining is a meaningful term refers to knowledge discovery data. It indicates to finding corrections between the institutions' tasks and the vast amount of data saved in huge databases of their consumers' information. This paper deals with attention to healthcare section, it generates mountains of

* **Corresponding author** : Lecturer Alaa Hussain Al-Obaidi, e-mail : ahaa_ma2009@uobabylon.edu.iq

ISSN print/ 2769-1926 ISSN online/ 2769-1934

administrative information and data about patients, hospitals, bed costs, claims, etc.

Clinical trials, electronic patient records and computer supported disease management will increasingly produce mountains of clinical data. This data is a strategic resource for healthcare institutions .Data mining products are designed to take this one stage further. It brings the property of funding out patterns and hidden correlation within the data repository. Professionals used this data to find out these patterns and put them in a framework useful enough for feeding back and getting more knowledge about any case, but making decisions of healthcare related with those professionals, not with the information system experts (Baylis, 2009). Till this time, there is not any clear definition related with this field.

The American Medical Informatics Association defined health Informatics as "all aspects of understanding and promoting the effective organization, analysis, management, and use of information in health care" (Masic, 2014). Similarly, the Canada's Health Informatics Association definition of Health Informatics is "Intersection of clinical, IM/IT and management practices to achieve better health". These are both broad definitions that cover a wide range of technologies, from developing electronic patient record data warehouses to installing wireless networks in hospitals. A more specific definition is provided by the National Library of Medicine, which defines Health Informatics as "the field of information science concerned with the analysis and dissemination of medical data through the application of computers to various aspects of health care and medicine". Note that here, Health Informatics is limited to "analysis and dissemination of medical data" (Fayyad U., 1996), and would not cover pure IT practices such as installing a network in a hospital. Zaiane provides an even more specific definition, which divides Health Informatics into four subfields:

> ***"Health Informatics is the computerization of health information to support and optimize (1) administration of health services; (2) clinical care; (3) medical research; and (4) training. It is the application of computing and communication technologies to optimize health information processing by collection, storage, effective retrieval (in due***

ISSN print/ 2769-1926 ISSN online/ 2769-1934

time and place), analysis and decision support for administrators, clinicians, researchers, and educators of medicine."

The successful data mining expert system depends not only on efficient algorithms, but also on who will use these systems and trying to get their advantages by how getting reports from these data.

This article contains two main sections. The first refers to data mining terms with its scientific related definitions, its task and its architecture. The second refers to data mining management application with detailed example from healthcare sector.

2. Data Mining

Data mining is an essential step in the knowledge discovery of databases (KDD) process that produces useful patterns or models from data (Figure 1) (Silwattananusarn, 2012). The term 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 data wealth of databases by focusing on the algorithms to extract useful knowledge.

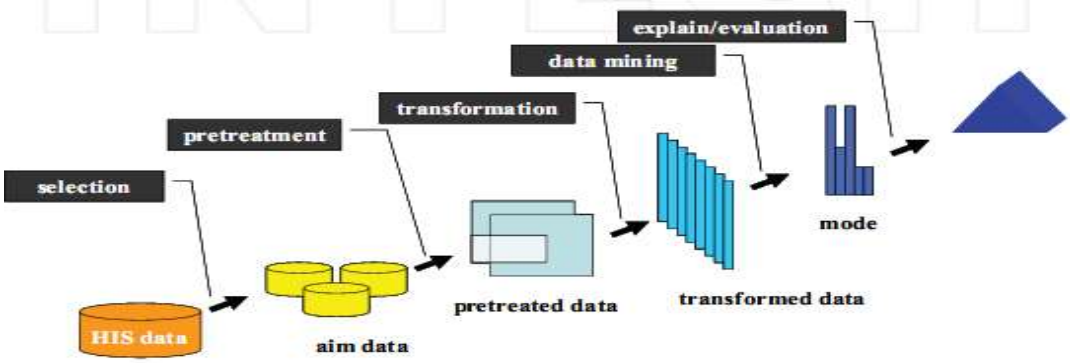


Figure (1) Data Mining and the KDD Process

Figure (1) refers to KDD process which \consists of iterative sequence methods as mentioned as follows (Fayyad U., 1996) (Han J., 2012)

1. **Selection**: that means selecting data relevant to the analysis task from the database

ISSN print/ 2769-1926 ISSN online/ 2769-1934

2. **Preprocessing**: that means to removing noise and inconsistent data or combining multiple data sources.
3. **Transformation**: that means transforming data into appropriate forms to perform data mining
4. **Data mining**: that refers to choosing a data mining algorithm which is appropriate to pattern in the data; or to extracting data patterns
5. **Interpretation/Evaluation**: that refers to interpreting the patterns into knowledge by removing redundant or irrelevant patterns; or to translating the useful patterns into terms that human understandable (Silwattananusarn, 2012).

2.1. Importance of Data Mining

Data mining is to extract information from large amount of a database. There are two main reasons to use data mining as a rapidly increase demands of data. These are:
- Too much data and too little information.
- Actual need to extract useful information from the data and to interpret it.



We are data rich
But
Information poor

The abundance of data, coupled with the need for powerful data analysis tools, has been described as a *data rich but information poor* situation. Data collected in large data repositories become “data tombs” Data archives that are seldom visited (Sharma, 2014).

ISSN print/ 2769-1926 ISSN online/ 2769-1934

2.2. History of Data Mining

The term "Data mining" is a technical term that was first used in the 1990s, but data mining is to set up a field with a long history.Data mining roots are traced back along with three family lines: classical statistics, artificial intelligence, and machine learning, herewith, is a brief explanation of these terms:

Statistics is the foundation of most technologies on which data mining is built, e.g. regression analysis, standard distribution, standard deviation, standard variance, discriminate analysis, cluster analysis, and confidence intervals. All of these are used to study data and data relationships.

Artificial intelligence (AI) which is built upon heuristics as opposed to statistics, attempts to apply human-thought-like processing to statistical problems. Certain AI concepts which were adopted by some high-end commercial products, such as query optimization modules for Relational Database Management Systems (RDBMS).

Machine learning is the union of both statistics and AI. It could be recognized as an evolution of AI, because it blends AI heuristics with advanced statistical analysis. Machine learning attempts to let computer programs learn about the data they study, such that programs make different decisions based on the qualities of the studied data, using statistics for fundamental concepts, and adding more advanced AI heuristics and algorithms to achieve its goals (Sharma, 2014). Data mining, in many ways, is fundamentally the adaptation of machine learning techniques to business applications. Data mining is best described as the union of historical and recent developments in statistics, AI, and machine learning. These techniques are then used together to study data and find previously-hidden trends or patterns within.

2.3. Data Mining Tasks

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

1. Classification is finding models that analyze and classify a data item into several predefined classes.

ISSN print/ 2769-1926 ISSN online/ 2769-1934

2. Regression is mapping a data item to a real-valued prediction variable

3**. Clustering** is identifying a finite set of categories or clusters to describe the data

4. **Dependency Modeling** (Association Rule Learning) is finding a model which describes significant dependencies between variables

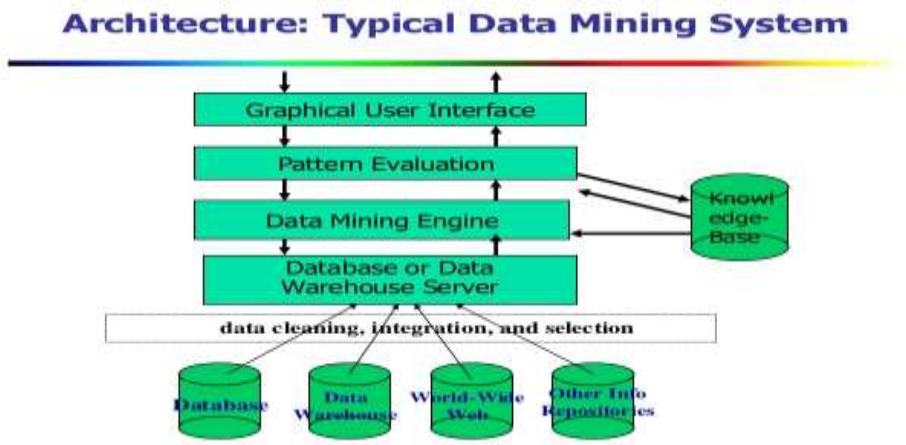
5**. Deviation Detection** (Anomaly Detection) is discovering the most significant changes in the data

6. **Summarization** is finding a compact description for a subset of data

Data mining has two primary objectives of prediction and description. Prediction involves using some variables in data sets in order to predict unknown values of other relevant variables (e.g. classification, regression, and anomaly detection) Description involves finding human understandable patterns and trends in the data (F., 2011).

3. Data Mining Architecture

The architecture of a typical data mining system may have the following major components:



Database, data warehouse, World Wide Web, or other information repository One or set of databases, data warehouses, spreadsheets, or other kinds of information repositories, data cleaning and data integration techniques may be performed on the data.

Database or data warehouse server is responsible for fetching the relevant data, based on the user's data mining request can be decouples/loose coupled/tightly coupled with the database layer knowledge base the domain

ISSN print/ 2769-1926 ISSN online/ 2769-1934

knowledge that is used to lead the search or estimate the interestingness of resulting patterns interestingness constraints or thresholds, metadata, concept hierarchies, etc.

Data mining engine is a potential to the data mining system and basically consists of a set of functional modules for tasks such as; characterization, association and correlation analysis, classification, prediction, cluster analysis, outlier analysis, and evolution analysis.

The query languages (DMQL) based on mining primitives to access the data pattern evaluation module interact with the data mining modules so as to focus the search toward interesting patterns may use interestingness thresholds to filter out discovered patterns may be integrated with the mining module user interface communicates between users and the data mining system that allows the user to interact with the system by specifying a data mining query or task, providing information to help focus the search, and performing exploratory data mining based on the intermediate data mining results allows the user to browse database and data warehouse schemas or data structures, evaluate mined patterns, and visualize the patterns in different forms (Mahamune, 2014).

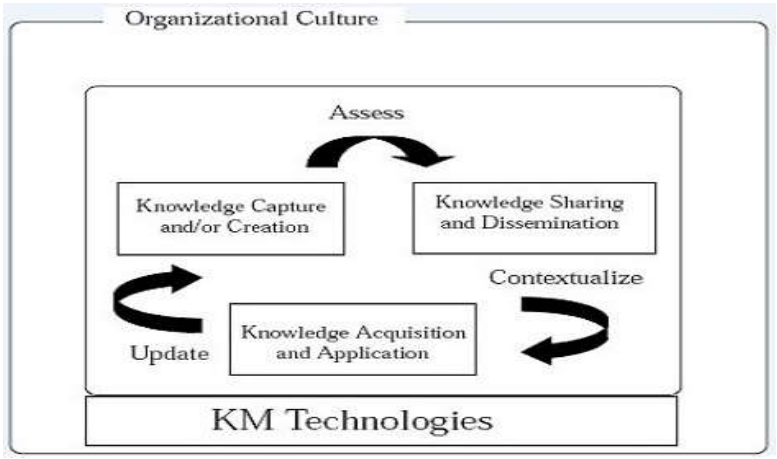
4. Knowledge Management

4.1. Definition of Knowledge Management

There are different conceptions of knowledge management. In this article, we use the definition of knowledge management by McInerney (2002): Knowledge management (KM) is an effort to increase useful knowledge within the organization. There are ways to do this include encouraging communication, offering opportunities to learn, and promoting the sharing of appropriate knowledge artifacts. This definition emphasizes the interaction aspect of knowledge management and organizational learning.

Knowledge management process focuses on knowledge flows and the process of creation, sharing, and distributing knowledge (Figure 3]. Each of knowledge units of capture and creation, sharing and dissemination, acquisition and application can be facilitated by information technology.

ISSN print/ 2769-1926 ISSN online/ 2769-1934



KM Technologies Integrated KM Cycle

Technologies play an important role in KM, technologies stand to be a necessary tool for KM usage. Thus, KM requires technologies to facilitate communication, collaboration, and content for better knowledge capture, sharing, dissemination, and application (Silwattananusarn, 2012).

4.2. Knowledge Resources

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

1. Health Care Organization: it is a use of the disease knowledge management system (KMS) of the hospital case study. Data mining tool was used to explore diseases, operations, and tumors relationships. This tool used to build KMS for assisting clinical medicine in order to develop treatment quality.

2. Retailing: it means customer knowledge from household customers for product line and brand extension issues; data mining can help and propose suggestions and solutions to the firm for product line and brand extensions. It is done by extracting market knowledge of customers, brands, products, and purchase data to achieve the customers' demands behavior.

3. Financial/Banking: the domain knowledge includes financial and economic data; data mining assists banking institutions making decision support and knowledge sharing processes to an enterprise bond classification.

ISSN print/ 2769-1926 ISSN online/ 2769-1934

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. Data mining and knowledge management integrated can help making better decisions. 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. 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 system.

5. Entrepreneurial Science: the knowledge resource is research assets in a knowledge Institution; there are 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.

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

7. Collaboration and Teamwork: Workers log and documents were analyzed each workers referencing behavior and construct workers knowledge flow. Data mining techniques can mine and construct group-based knowledge flows (GKFs) prototype for task-based groups.

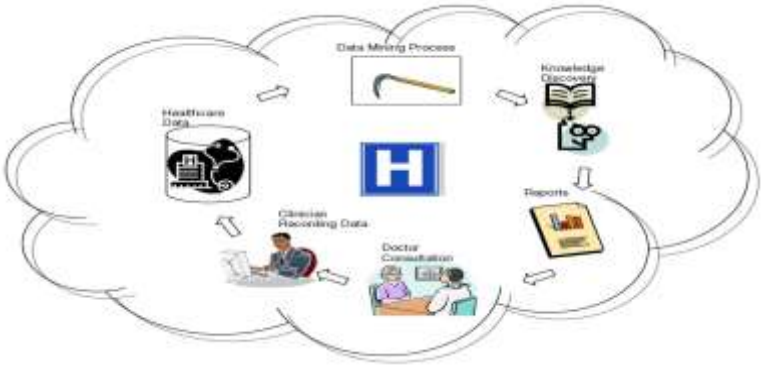
8. Construction Industry: a large part of this enterprise information is available in a form of textual data formats. This leads to the influence of text mining techniques to handle textual information source for industrial knowledge discovery and management solutions (Mahamune, 2014).

5. **Data mining Aids Health Care**

- Healthcare industry daily generates large amounts of complex and vast data about patients, hospitals resources, disease diagnosis, electronic patient records, medical devices etc.
- Large amounts of data become the key resources to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. Data mining determines a set of tools and techniques that could be applied to this processed data for discovering

ISSN print/ 2769-1926 ISSN online/ 2769-1934

hidden patterns and supporting healthcare professionals as an extra source of knowledge for making decisions. Professionals of health care are the only persons who make decisions.



5.1. Health Informatics Divisions

Logically, health informatics can be divided into four main subfields:
1. Administration of health services
2. Clinical care
3. Medical research
4. Training.

The following subsections present an overview of each division and how data mining is, or can be, applied to extend and improve each one.

5.1.1. Administration of Health Services

Administrators of health care organizations make hundreds of critical decisions on daily basis. As in any administrative position, the quality of these decisions directly depends on the quality of the information that the decisions are based on. For example, the administrators in a hospital need to decide on the amount of supplies and number of staff and free beds required for an upcoming month. To make this decision, the administrators require an accurate prediction of the number of patients to expect during the coming month, and an approximation of how long each patient will remain in the hospital. As another example, the federal and provincial health administrators need to decide whether a disease outbreak is in progress, and if so, what preventive measures will be most effective against it. To make these

ISSN print/ 2769-1926 ISSN online/ 2769-1934

decisions, the administration requires a system that can accurately predict a disease outbreak, and also model the cost and benefit of different preventive measures. The following case study illustrates the applications of data mining techniques on epidemic detection.

5.1.2. Clinical Care

Making diagnostic decisions can only be done by physicians and nurse practitioners, those who also advice treatment recommendations based on different factors such history, medical imaging, lab results and other text or multimedia records of patients.

Health informatics allows doctors to have faster access to more relevant information, and thus make more optimal decisions. For example, a centralized patient record database will allow a physician in a local clinic to have access to all the relevant medical records of the patient, anywhere in the country. Furthermore, applying data mining techniques on the centralized database will give doctors analytical and predictive tools that go beyond what is apparent from the surface of the data. For instance, a new practitioner can query for all the decisions that previous practitioners have made on a similar case. Similarly, a predictive model can advise doctors whether a certain case would be better treated as an outpatient or an inpatient.

5.1.3. Medical Research

Most current successful applications of data mining in Health Informatics are in the subfield of medical research. The reason behind that is most of the current health related data are stored in small datasets scattered through various clinics, hospitals, and research centers. However, most applications of data mining in clinical and administrative decision support systems require homogeneous and centralized data warehouses (see section 3). On the other hand, data mining methods can still be successfully applied on small and scattered datasets, and help researchers extract insightful patterns, cause and effect relationships, and predictive scoring systems from currently available data. The following subsections introduce examples of data mining tools applied on small datasets for medical research.

5.1.4. Education and Training

The fourth subfield of health informatics is related to educating new healthcare professionals and retraining and keeping the current staff up-to-

ISSN print/ 2769-1926 ISSN online/ 2769-1934

date with recent advances in technology. The education and training subfield of Health Informatics can be viewed as an instance of the rapidly growing field of e-learning. An increasing interest in applying data mining techniques to e-learning has emerged in recent years, and some of the early applications show promising results. Data mining techniques can benefit all three groups of people who are in contact with a learning system: students, educators, and administrators (Han J., 2012). Data mining techniques can monitor the success of students at various learning tasks, and recommend relevant resources, materials, and learning paths to achieve a more successful learning experience. For educators, data mining techniques can provide objective feedback of the structure and the content of a course, discover the learning patterns of the students, and cluster learners into smaller groups that have similar educational habits and needs. Administrators benefit from data mining techniques by learning about the behavior of their users, so they can optimize the servers, distribute network traffic, and learn about the overall effectiveness of the offered educational programs. The following two case studies present an overview of a relatively new Health Informatics e-learning tool called HOMER, and a data mining technique to find relevant articles for a particular gene (N, 2012).

6. Application of Data Mining in Health Care Sector

Business and marketing organizations may be ahead of healthcare in applying data mining to derive knowledge from data. Successful mining applications have been implemented in the healthcare arena, three of which are described below.

Hospital Infection Control Nosocomial infections affect 2 million patients each year in the United States, and the number of drug-resistant infections has reached unprecedented levels.

Early recognition of outbreaks and emerging resistance requires proactive surveillance. Computer-assisted surveillance research has focused on identifying high-risk patients, expert systems, and possible cases and detecting deviations in the occurrence of predefined events. A surveillance system that uses data mining techniques to identify new and interesting patterns in infection control data has been implemented at the University of Alabama. The system uses association rules on culture and patient care data obtained from the laboratory information management systems and generates

ISSN print/ 2769-1926 ISSN online/ 2769-1934

monthly patterns that are reviewed by an expert in infection control. Developers of the system conclude enhancing infection control with the data mining system is more sensitive than traditional infection control surveillance, and significantly more specific.

Ranking Hospitals Organizations rank hospitals and healthcare plans based on information reported by healthcare providers. There is an assumption of uniform reporting, but research shows room for improvement in uniformity. Data mining techniques have been implemented to examine reporting practices. With the use of International Classification of Diseases, 9th revision, codes (risk factors) and by reconstructing patient profiles, cluster and association analyses can show how risk factors are reported.

Standardized reporting is important because hospitals that underreport risk factors will have lower predications for patient mortality. Even if their success rates are equal to those of other hospitals, their ranking will be lower because they reported a greater difference between predicted and actual mortality. Standardized reporting would also be important for meaningful comparisons across hospitals.

Identifying High-Risk Patients: American Health ways provide diabetes disease Management services to hospitals and health plans designed to enhance the quality and lower the cost of treatment of individuals with diabetes. To augment the company's ability to prospectively identify high-risk patients, American Health ways uses predictive modeling technology.

Extensive patient information is combined and explored to predict the likelihood of short-term health problems and intervene proactively for better short-term and long-term results.

A robust data mining and model-building solution identifies patients who are trending toward a high-risk condition. This information gives nurse care coordinators a head start in identifying high-risk patients so that steps can be taken to improve the patients' quality of healthcare and to prevent health problems in the future (Obenshain, 2004).

6.1. Techniques of Data Mining in Healthcare

In most cases, analytical techniques are used in data mining that for a long have been known as mathematical techniques and algorithms. Although data mining is a young technology, the process of data analysis is not new. The thing that linked these techniques and large databases is a cheaper storage space and processing power.

ISSN print/ 2769-1926 ISSN online/ 2769-1934

Here are some of the techniques of data mining, which are successfully used in healthcare, such as artificial neural networks, decision trees, and genetic algorithms and nearest neighbor method.

- **Artificial neural networks** are analytical techniques that are formed on the basis of superior learning processes in the human brain. As the human brain is capable to, after the learning process, draw assumptions based on previous observations, neural networks are also capable to predict changes and events in the system after the process of learning. Neural networks are groups of connected input/output units where each connection has its own weight. The learning process is performed by balancing the net on the basis of relations that exist between elements in the examples. Based on the importance of cause and effect between certain data, stronger or weaker connections between "neurons" are being formed. Network formed in this manner is ready for the unknown data and it will react based on previously acquired knowledge. Artificial neural networks are ideal for multiprocessor systems, where a large number of operations are performed in parallel.

- **Decision tree** is a graphical representation of the relations that exist between the data in the database. It is used for data classification. The result is displayed as a tree, hence the name of this technique. Decision trees are mainly used in the classification and prediction. It is a simple and a powerful way of representing knowledge. The models obtained from the decision tree are represented as a tree structure. The instances are classified by sorting them down the tree from the root node to some leaf node. The nodes are branching based on if-then condition.

  Tree view is a clear and easy to understand, decision tree algorithms are significantly faster than neural networks and their learning is of shorter duration. Decision tree is a tree where each (non-terminal) node represents a test or decision on the item of information that is listed for consideration. The choice of a particular industry depends on the outcome of the test. In order to classify the data, process is starting from the root node and following the argument down until it reaches the final node, at which time a decision is made. Decision tree

ISSN print/ 2769-1926 ISSN online/ 2769-1934

can also be interpreted as a special form of a rule set, which is characterized by its hierarchical organization of rules.

- **Genetic algorithms** are based on the principle of genetic modification, mutation and natural selection. These are algorithmic optimization strategies inspired by the principles observed in natural evolution. The genetic algorithm creates a number of random solutions to the problem. All these solutions may not be good, a group of solutions can be skipped entirely, and it can come down to the overlapping solutions. Poor solutions are discarded, and the good ones retained. Good solutions are then being hybridized, and then the whole process is repeated. Finally, similar to the process of natural selection, only the best solutions remain. So, from the set of potential solutions to the problems that compete with each other, the best solutions are chosen and combined with each other in order to obtain a universal solution from the set of solutions that will become better and better, similar to the process of evolution of organisms. Genetic algorithms are used in data mining to formulate hypotheses about the dependencies between variables in the form of association rules or other internal formalism.

  The disadvantage of this method is that it requires an enormous amount of processing power and it is too slow for trivial issues. Since evolutionary computation is a robust and parallel search algorithm, it can be used in data mining to find interesting knowledge in noisy environment.
- **Nearest neighbor** method is a technique that also used for data classification. Unlike other techniques, there is no learning process to create a model. The data used for learning is in fact a model. When the new data shows up, the algorithm analyzes all the data in the database to find a subset of instances that are the best fit and based on that it is able to predict the outcome. The study conducted on the application of nearest neighbor method on benchmark data set to detect efficiency in the diagnosis of heart diseases, produced the results that application of this method achieved an accuracy of 97.4% which is a higher percentage than any other published study on the same set of data.

ISSN print/ 2769-1926 ISSN online/ 2769-1934

Data mining in healthcare demands close cooperation between managers of quality in healthcare and data mining experts, and it is consisted of analysis driven by data and analysis driven by interest. Analysis driven by interest can be divided into seven key phases:

(1) Acquiring knowledge from area.
(2) Formulation of business questions.
(3) Processing of business questions.
(4) Transformation of business questions into data mining queries.
(5) Execution of data mining queries.
(6) Processing of data mining results.
(7) Transformation of data mining results into answers.

Type of analysis driven by data is used because analysis driven by interest can predict unexpected patterns in data. Association rule is usually used within this analysis. This approach that uses both types of analysis has a good and a bad side, because the users are not thrilled with a large number of findings that are way beyond their field of expectations, and then again, unexpected patterns don't stay unnoticed (Milovic, 2012).

6.2. Advantages of Application

Information technologies in healthcare have enabled the creation of electronic patient records obtained from monitoring of the patient visits. This information includes patient demographics, records on the treatment progress, details of examination, prescribed drugs, previous medical history, lab results, etc. Information system simplifies and automates the workflow of health care institution. Privacy of documentation and ethical use of information about patients is a major obstacle for data mining in medicine. In order for data mining to be more exact, it is necessary to make a considerable amount of documentation. Health records are private information, yet the use of these private documents may help in treating deadly diseases. Before data mining process can begin, healthcare organizations must formulate a clear policy concerning privacy and security of patient records. This policy must be fully implemented in order to ensure patient privacy. Health institutions

ISSN print/ 2769-1926 ISSN online/ 2769-1934

are able to use data mining applications for a variety of areas, such as doctors who use patterns by measuring clinical indicators, quality indicators, customer satisfaction and economic indicators, performance of physicians from multiple perspectives to optimize use of resources, cost efficiency and decision making based on evidence, identifying high-risk patients and intervene proactively, optimize health care, etc.

Integration of data mining in information systems, healthcare institutions reduce subjectivity in decision-making and provide a new useful medical knowledge. Predictive models provide the best knowledge support and experience to healthcare workers. Data mining is using a technique of predictive modeling to determine which diseases and conditions are the leading trends. This requires a review of medical documentation of a healthcare institution and prescription drugs to determine which problems are the most common amongst patients. The problem of prediction in medicine can be divided into two phases: learning phase and the phase of decision making. In the learning phase, a large data set is transformed into a reduced (simplified) data set. Number of features and objects in this new set is much smaller than the original set in several different ways. The rules generated in this phase are used later to make accurate decisions. Newly formed data set is used to make predictions when the new instances with unknown outcomes occur with the predictive algorithm. This algorithm compares the characteristics of a new object with the characteristics of objects represented in the selected data set. If the match is found, the new object gets the outcome which is equal to the corresponding object in the set.

The goal of predictive data mining in medicine is to develop a predictive model that is clear, makes reliable predictions and helps doctors to improve their prognosis, diagnosis and treatment planning procedures. Important questions arise here (Milovic, 2012): if the data and the corresponding predictive characteristics are sufficient to build a predictive model of acceptable performance; what is the relation between attribute and outcome; can it be found an interesting combination, or the relation between attributes; whether the immediate factors can be drawn from the original attributes that can increase the performance of predictive models.

A very important application of data mining is for biomedical signal processing expressed by internal regulations and responses to the stimulus conditions, whenever there is a lack of detailed knowledge about the interactions between different subsystems, and when the standard analysis

ISSN print/ 2769-1926 ISSN online/ 2769-1934

techniques are ineffective, as it is often the case with non-linear associations. Data mining provides the link between knowledge of continuous data, such as biomedical signals collected from patients in intensive care units, and it develops an intelligent monitoring system that sends reminders, warnings and alarms for the pre-selected critical conditions.

Using association rules involves finding all the rules, or at least part of key subsets of rules that is characteristic of certain information as consequences or as a antecedent. This type of problem is very interesting for health professionals who are searching for the relations between diseases and lifestyles or demographics or between survival rates and treatment. The tasks of association are used to help strengthen the arguments regarding whether to engage or eliminate certain rules in the knowledge model.

Tasks of the managers that manage quality of the healthcare services can be described as optimization of clinical processes in terms of medical and administrative quality as well as the cost/benefit relation. Key questions of the process of healthcare quality management are quality of data, standards, plans, and treatments. Data mining can be used by quality managers to solve the following tasks:

(1) Discovering new hypothesis for indexes of quality for data, standards, plans and treatments.

(2) Checking if the given indexes of quality for data, standards, plans and treatments are still valid.

(3) Improving, strengthening and adjusting of quality indexes for data, standards, plans and treatments.

(4) These tasks can be supported by data mining if the existing knowledge in domain is seriously considered in data mining process (Milovic, 2012).

6.3. Challenges of Application

The application of data mining, knowledge discovery and machine learning techniques to medical and health data is challenging and intriguing. The datasets usually are very large, complex, heterogeneous, and hierarchical and vary in quality. Data preprocessing and transformation are required even before mining and discovery can be applied. Sometimes the characteristics of the data may not be optimal for mining or analytic processing. The challenge

ISSN print/ 2769-1926 ISSN online/ 2769-1934

here is to convert the data into appropriate form before any leaning or mining can begin. There are a number of issues that must be addressed before any data mining can occur. In the following, we overview some of the challenges that face the data mining process on medical databases.

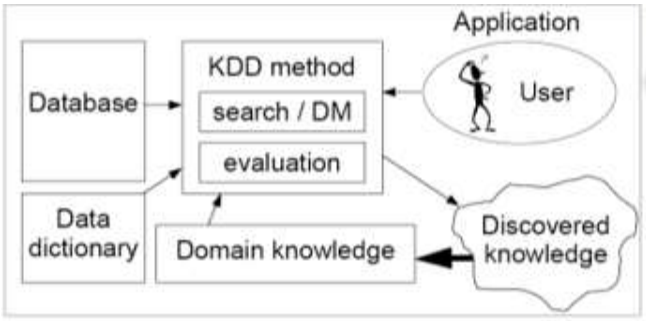
High volume of data Due to the high volume of the medical databases, current data mining tools may require extraction of a sample from the database. Another scheme is to select some attributes from the database. In both approaches, domain knowledge can be used to eliminate irrelevant records or attributes in reducing the size of the database.

Update Medical databases are updated constantly by adding new results for lab tests and new ECG signals for patients. Subsequently, any data mining technique should be able to incrementally update the discovered knowledge.

Inconsistent data representation Inconsistencies due to data entry errors are common problems. Inconsistencies due to data representation can exist if more than one model for expressing a specific meaning exits (e.g., the location of disease for Colitis, one application may enter (sigmoid, or rectum, etc.) and another may enter (measurements such as 20 cm, 30 cm, etc.)). Additionally, the data type does not always reflect the true data type. For example, a column with numerical data type can represent a nominal or ordinal variable encoded with numbers instead of a continuous variable. This plays an important role during statistical analysis (mean and variance).

Previous (background) knowledge as concept hierarchy, domain expert knowledge, previously mined knowledge and rule template must be taken in account in data preparation, modeling and model evaluation phases. Background knowledge can be expressed in different formats: examples may be found in the areas of decision rules, Bayesian models, fuzzy sets and concept hierarchies. Figure 3 shows that background knowledge must be taken in to account in the KDD process (Ashktorab, 2009).

ISSN print/ 2769-1926 ISSN online/ 2769-1934



Background knowledge role in KDD process

Poor Integration Health data is fragmented and distributed between hospitals, insurance companies and government departments. This poses a substantial challenge for data integration and data mining in terms of the confidence that can be placed in the result and the semantics of a derived rule. One can use common data dictionary and standards to integrate data from heterogeneous systems. The emergence of XML as a data standard is gaining wider acceptance and hence making integration fairly easy in the near future.

Number of variables The computational complexity is not linear for certain data mining techniques. In such cases, the time required may become infeasible as the number of variables grow. Techniques such as principle component analysis, available in the XLMiner data mining tool, can be used to reduce the dimensionality (number of variables) of the dataset but retain most of the original variability in the data. In addition, domain knowledge can be used to eliminate the irrelevant attributes from data mining consideration.

Missing/Incomplete data Clinical database systems do not often collect all the data required for analysis or discovery. Some data elements are not collected due to omission, irrelevance, excess risk or inapplicability in a specific clinical context. For some learning methods such as logistic regression, a complete set of data elements may be required. Even when the methods accept missing values, the data that was not collected may have independent information value and should not be ignored. One possible approach for handling the missing data is to substitute missing values with most likely values.

ISSN print/ 2769-1926 ISSN online/ 2769-1934

Noise Medical databases include some noises. Therefore, data mining techniques should be less sensitive to noises.

Amount of results the quantity of output from many data mining methods is unmanageable. Association rule mining has been used in hospital infection control and public surveillance data and in Seplic Shock patient data. Too many rules have been found in both projects. Other problems include trivial and similar patterns observed in drug reaction data and in chronic hepatitis data (Ashktorab, 2009).

7. Conclusion

Data mining has been used prosperously in the favorably perceived areas such as e- business, marketing and retail because of which it is now applicable in knowledge discovery in databases (KDD) in many industrial areas and economy. Data mining is mainly gaining its importance and usage in the areas of medicine and public health. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs, data mining application in healthcare sector is useful for extracting information. The prediction of diseases using Data Mining applications is a challenging task but it drastically reduces the human effort and increases the diagnostic accuracy. Developing efficient data mining tools for an application could reduce the cost and time constraint in terms of human resources and expertise. Exploring knowledge from the medical data is such a risk task as the data found are noisy, irrelevant and massive too.The performance of data mining techniques depends on the type of dataset that we have taken for doing experiment.

8. Contribute

As a first step we have to understand the concepts of data mining and its techniques, algorithms and this is what we present in this research and as a second step we want to use the algorithms of data mining to find the relation between the impact of the financial crisis and the number of beds of psychiatric patients in different countries in Europe as the future work.

Bibliography

1) Ashktorab, H. (2009). Challenges in Data Mining on Medical Databases, Category: Health Information Systems. *Howard University Hospital, USA*.

2) Baylis, P. (2009). Better Healthcare with Data Mining. *Shared Medical Systems Limited*.

3) F., G. (2011). Data Mining: Concepts, Models, and Techniques. *India Springer*.

4) Fayyad U., P.-S. G. (1996). From Data Mining to Knowledge. *AI Magazine*.

5) Han J., K. M. (2012). *Data Mining: Concepts and Techniques.* Boston: Morgan Kaufmann Publishers.

6) Mahamune, M. (2014). Application of Data Mining Techniques in Knowledge Management System. *School of Computational Science*.

7) Masic, I. (2014). Five Periods in Development of Medical Informatics. *Academy of Medical Sciences of Bosnia and Herzegovina*.

8) Milovic, B. (2012). Prediction and Decision Making in Health Care using Data Mining. *International Journal of Public Health Science (IJPHS)*.

9) N, N. (2012). Data Mining in Health Informatics. *International Journal of Public Health Science*.

10) Obenshain, M. K. (2004). Application of Data Mining Techniques to Healthcare Data. *Cambridge University*.

11) Sharma, M. (2014). Data Mining: A Literature Survey , Global Research Institute of Management Technology. *International Journal of Emerging Research in Management Technology*.

12) Silwattananusarn, T. (2012). Data Mining and Its Applications for Knowledge Management. *International Journal of Data Mining*.

Wisdom Journal For Studies And Research volume 04 Issue 06(20) 15/11/2024~

ISSN print/ 2769-1926 ISSN online/ 2769-1934

تنقيب البيانات في مجالات القطاعات الصحية

م.الاء حسين علي العبيدي

جامعة بابل / العراق

E_mail: ahaa_ma2009@uobabylon.edu.iq

ملخص:

تعد عملية استخراج البيانات من التقنيات الحديثة التي تدعم اكتشاف المعلومات من كمية كبيرة من البيانات. وهذا يعني كيفية استخراج معلومات مرتبة من كمية هائلة من البيانات المبعثرة. وفي الآونة الأخيرة، جذبت هذه التقنية قدرًا كبيرًا من الاهتمام في المجتمع بشكل عام وفي صناعة المعلومات بشكل خاص. وفي هذه الورقة البحثية، حاول الباحث التركيز على قطاع الرعاية الصحية حيث توجد تحديات خطيرة في تطبيق هذه التقنية. كما سيتم التطرق الى تقنية استخراج البيانات الخاصة بالقطاع الصحي والصعوبات والتحديات التي تواجهها بسبب كثرة البيانات المتعلقة بهذا القطاع ويمكن اعتبار هذه دراسة وصفية تبحث في المفاهيم والقضايا ذات الصلة والتقنيات المستخدمة في عملية استخراج البيانات في قطاع الرعاية الصحية.

الكلمات المفتاحية: تنقيب البيانات، القطاع الصحي.