International Journal of Public Health Science (IJPHS)
Vol. 1, No. 2, December 2012, pp. 69~78
ISSN: 2252-8806

Prediction and Decision Making in Health Care using Data Mining

Boris Milovic*, Milan Milovic**
*Faculty of Economics Subotica, Novi Sad University, Serbia
**General Hospital Vrbas, Serbia

Article Info

Article history:

Received Jul 29, 2012
Revised Oct 21, 2012
Accepted Nov 8, 2012

Keyword:

Data mining
Medicine
Information system
Healthcare

ABSTRACT

Tendency for data mining application in healthcare today is great, because healthcare sector is rich with information, and data mining is becoming a necessity. Healthcare organizations produce and collect large volumes of information on daily basis. Use of information technologies allows automatization of processes for extraction of data that help to get interesting knowledge and regularities, which means the elimination of manual tasks and easier extraction of data directly from electronic records, transferring onto secure electronic system of medical records which will save lives and reduce the cost of the healthcare services, as well and early discovery of contagious diseases with the advanced collection of data. Data mining can enable healthcare organizations to predict trends in the patient conditions and their behaviors, which is accomplished by data analysis from different perspectives and discovering connections and relations from seemingly unrelated information. Raw data from healthcare organizations are voluminous and heterogeneous. They need to be collected and stored in the organized forms, and their integration enables forming of hospital information system. Healthcare data mining provides countless possibilities for hidden pattern investigation from these data sets. These patterns can be used by physicians to determine diagnoses, prognoses and treatments for patients in healthcare organizations.

Corresponding Author:

Boris Milovic,
Faculty of Economics Subotica
Novi Sad University
24000 Subotica, Serbia.
Email: boris.milovic@gmail.com

1. INTRODUCTION

Healthcare organizations today are capable of generating and collecting a large amounts of data. This increase in volume of data requires automatic way for these data to be extracted when needed. With the use of data mining techniques it is possible to extract interesting and useful knowledge and regularities. Knowledge acquired in this manner, can be used in appropriate area to improve work efficiency and enhance quality of decision making process. Above stated points that there is a great need for new generation of computer theories and tools to help people with extracting useful information from constantly growing volume of digital data [1]. Information technologies are being increasingly implemented in healthcare organizations in order to respond to the needs of doctors in their daily decision making activities. Data mining tools can be very useful to control limitations of people such as subjectivity or error due to fatigue, and to provide indications for the decision-making processes [2]. The essence of data mining is in the identification of relations, patterns and models that provide support for predictions and of decision making process for diagnoses and treatment planning. These models can be called predictive, and they are being integrated into information systems of hospitals as a models for decision making, reducing the subjectivity and decision making time. In addition, the use of information technology in healthcare enables comprehensive management of medical knowledge and its secure exchange between recipients and providers of healthcare services [3]. Widespread use of information technology enables the elimination of manual tasks of data extraction from charts or filling of specialized questionnaires, extraction of data directly from electronic records, transfer on secure electronic system of medical records that will save lives and reduce the cost of health care, early detection of infectious diseases with advanced collection of data etc. Retrieval of information with the help of computers can help the quality of decision making and avoiding human errors. When there is a large volume of data that needs to be classified, decision making by people is usually poor [4]. Data mining represents the process of analyzing raw data with the help of computer and extraction of their meaning. It is frequently defined as discovering previously unknown and potentially useful information from large volume of (unstructured) data [5]. Thanks to this technique, it is possible to predict trends and customer behavior and thus provide the organization's business success. This is accomplished by data analysis from various perspectives and finding the connections and relations between mutually unconnected information. In the process of data mining previously unknown trends and patterns from a database of historical information are being discovered and that information is being converted into significant business solutions [6].

2. PROCESS OF KNOWLEDGE DISCOVERY AND DATA MINING

Knowledge discovery (KDD) is a process that allows automatic scanning of high-volume data in order to find useful patterns that can be considered as knowledge about the data. Once the discovered knowledge is presented, the evaluation measures can be improved, mining can be further "refined", new data can be selected or further transformed, or new data sources can be integrated in order to obtain different, the corresponding results [7]. This is the process of converting low level information into high level knowledge. Therefore, KDD is a non-trivial extraction of implicit, previously unknown and potentially useful information from data that is located in databases. Although data mining and KDD are often treated as equivalent, in essence, data mining is an important step in the KDD process. The process of knowledge discovery involves the use of the database along with any selection, preprocessing, subsampling and transformation; by applying data mining methods for enumerating the models from it; evaluating the products of data mining to identify subsets of enumerated models that represent knowledge. Data mining component of the knowledge discovery process refers to the algorithmic means by which models are extracted and enumerated from the data [1].

With the application of data mining tools in spreadsheet of the program that analyzes data to identify patterns and relations, user profiling and development of business strategies can be started [8]. Most data mining software include online analytical processing, traditional statistical methods, such as cluster analysis, discriminant analysis and regression analysis, and non-traditional statistical analysis such as neural networks, decision trees, link analysis and association analysis. This wide range of techniques is not surprising in light of the fact that data mining is derived from three different disciplines, database management, statistics and computer science, including the use of artificial intelligence and machine learning [9].

Because of all this data mining process is inextricably linked to computers. With the help of special software, a big computer systems analyze data from different angles, find a hypothesis, experiment with them and learn from previous experiences. One should always bear in mind that the software is just a tool that is still required the presence of human experts to give the final decision in the fields in which data mining is being applied. But in the first stage of processing computer systems are indispensable for their speed and lack of prejudice. Unlike humans, which would let the obvious connection between the two data missed because it is beyond their expectations, such an error cannot happen to a computer. Also a human can be a victim of the conditionality with previous experience, which can be both positive and negative, but in any case impossible to avoid.

It can be argued that data mining represents finding out the legality of the information. The technology of data mining is closely associated with data warehousing and intertwined with the system for database management. Data mining involves the process of searching a large amount of previously unknown information, which it later used to make important business decisions. The key phrase here is "unknown data", which means that the information is cluttered with massive amounts of operational data which, when analyzed, provide relevant information to organizational decision makers. Datasets are generally large, complex, heterogeneous and hierarchical, and they vary in quality. Preprocessing and transformation of data are needed even before data mining and discovery can be applied. Sometimes data features are not optimal for data mining and analytical processing. The challenge here is to convert data into the appropriate form before the learning and data mining can start [10].

Daily intake of information that large organizations are storing in their databases, are measured in terabytes. We can say only that one terabyte can store enough text that is equal to about two million books. Sources of information are different (internal, external, analytical), information can be attributive or numeric, it can relate to factors that affect an organization's operations, internal procedures, the organization's service users (consumers), business competition, business environment. However, such a raw data, inadequately structured, with different formats, do not have a great utility value. It is necessary to prepare and analyze it, and based on that, get the information that can ensure organizational business success [5].

The choice of parameters, actions, components for the design of data mining applications is often made based on previous data, information or knowledge. A number of models and algorithms have been developed for autonomous prediction based on data corresponding to different features [11]. Data mining consists of applying data analysis and discovering algorithms that, under acceptable computational efficiency limitations, produce a specific enumeration of patterns (or models) over data. It should be noted that the model space is often infinite, and enumeration of the model involves some form of search in this area. Practical computer restrictions are placing severe limits on the subspace which can be explored by data mining algorithms [1].

Data mining consists of various methods. Different methods serve different purposes, each method has its advantages and disadvantages. Data mining tasks can be divided into descriptive and predictive [12]. While descriptive tasks have a goal on finding a human interpreted forms and associations, after reviewing the data and the whole construction of the model, prediction tasks tend to predict an outcome of interest. Although the goals of description and prediction tasks may overlap, the main difference is that the predictive ones require that data include a special variable of response. The response can be categorical or numeric, further classifying data mining as classification and regression. Predictive tasks make it possible to predict the value of a variable based on other existing information. Descriptive tasks, on the other hand, combine the data in a certain way. The main predictive and descriptive data mining tasks can be classified as following: [13]:

- **Classification and Regression** - identification of new templates with predefined objectives; These tasks are predictive and they include the creation of models to predict target, or dependent variable from the set of explained or independent variables. Classification is the process of finding a function that allows the classification of data in one of several classes. For classification tasks, the target variable usually has a small number of discrete values, while with the regression tasks, target variable is continuous.
- **Association rule –** association rule analysis is descriptive data mining task which includes determining patterns, or associations, between elements in data sets. Associations are represented in the form of rules, or implications.
- **Cluster analysis** – descriptive data mining task with the goal to group similar objects in the same cluster and different ones in the different clusters. Process of grouping determines groups of data that are similar, but different than other data. In this process variables are identified by which the best grouping is being realized.
- **Text mining –** most of the available data is in the form of unstructured or partially structured text, and it is different from conventional data that are completely structured. Text is unstructured if there is no previously determined format, or structure in data. Text is partially structured if there is a structure linked with the parts of data. While text mining tasks usually fall under classification, clustering and association rule data mining categories, it is the best to observe them separately, because unstructured text demands a specific consideration. In particular, method for representation of textual data is critical.
- **Link analysis** – Form of network analysis that examines the associations between objects. Link classification provides category of an object, not just based on its features, but also on connections in which it takes part, and features of objects connected with certain path [14]. Example of link analysis in medicine is task of predicting disease type based on people’s characteristics or predicting age of people based on disease they are infected with and based on age of people they have been in contact with. Link analysis can be used in order to understand where do patients go to receive the healthcare treatment and to identify the components or resources in service that must be addressed. This is a data mining form that includes population tracking during their movement from point to point in the system. This analysis requires population segmentation so the analysis can focus on percentage of the population [15]. In order for the link analysis to be possible, all the patient’s information must be stored in databases (personal information, dates and time of visits, doctors that treated the patient, doctors that gave referrals, patient’s previous diseases)

Upon completion of the information analysis, all results are displayed in a clear manner, usually in the form of tables or diagrams that may be two dimensional or three dimensional. Programs even allow the user to change any of the variables, and the effect of its change is shown in real time on the diagram.

3. APPLICATION OF DATA MINING IN HEALTHCARE

Modern era has brought significant changes, and information technologies have found wide application in the areas of human activities, as well as in the healthcare. Development and implementation of new information technologies that allow global networking, give modern medicine the epithet of "informatical medicine". Information technologies increasingly provide the help in system approach of solving medical problems [16]. Disposition of the right information enables the preparation of accurate reports, for example, usage of hospital capacities, or number of occupied beds. At the same time it is easier to monitor treatment and to check the information exchange. Use of information technologies enables change of the healthcare system - how to improve public health, the healthcare of the system users, reduce costs, save time and money.

Healthcare abounds various information which causes the necessity of data mining application. One of the first applications in the area of data mining for healthcare was KEFIR (Key Findings Reporter), that was automatically analyzing changes in all relevant variables, extracting the important ones, and adding an expert recommendations on what actions need to be taken in response to these changes. It is well known that healthcare is a complex area where new knowledge is being accumulated daily in a growing rate. Big part of this knowledge is in the form of paperwork, resulting from a studies conducted on data and information collected from the patient's healthcare records. There is a big tendency today to make this information available in electronic form, converting information to knowledge, which is not an easy thing to do [17]. With the growth of costs in healthcare organizations and the growing necessity to control all the expenses, suitable analysis of medical information has become an issue of the utmost importance. All healthcare institutions need an expert analysis of their medical data, the project that is time consuming and expensive [18]. There is a great potential for data mining application in healthcare. Healthcare institutions are very oriented on use of patient's information. Ability to use a data in databases in order to extract useful information for quality health care is a key of success of healthcare institutions [4]. Healthcare information systems contain a large volumes of information that include information on patients, data from laboratories that are constantly growing. With the use of data mining methods, useful patterns of information can be found in this data, that will later be used for further research and report evaluation. A very important issue is how to classify large volumes of data. Automatic classification is done based on the similarities that are present in data. This type of classification is useful only if the conclusion, that is drawn, is acceptable for the doctor or the end user. Data mining provides support for identification of reliable relations between treatment and outcome.

In medical research, data mining begins with the hypothesis and results are adjusted accordingly. This is different from standard data mining practice, which simply begins with a set of data without obvious hypothesis [19]. While the traditional data mining is focused on patterns and trends in data sets, data mining in healthcare is more focused on minority that is not in accordance with patterns and trends. The fact that standard data mining is more focused on describing and not explaining the patterns and trends, is the one thing that deepens the difference between standard and healthcare data mining. Healthcare needs these explanations since the small difference can stand between life and death of a patient.

Analytical techniques used in data mining, in most cases have long been known mathematical techniques and algorithms. Although data mining is a young technology, the process of data analysis is nothing new. The thing that linked these techniques and large databases is a cheaper storage space and processing power. Here are some of the techniques of data mining, which are successfully used in healthcare, such as artificial neural networks, decision trees, 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 [20]. 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 [21]. The nodes are branching based on if-then condition. Tree view is a clear and easy to understand, a 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 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 [20]. 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. A 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. [22].

Nearest neighbor method is a technique that is 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 [23] 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.

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 [24]. 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 these 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.

4. ADVANTAGES OF DATA MINING APPLICATION IN HEALTHCARE

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 [19]. 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 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 [25].

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 [12]: 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 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 [2].

Using association rules involves finding all the rules, or at least part of key subsets of rules that is characteristic of certain information as a 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 [26].

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; and (4) These tasks can be supported by data mining if the existing knowledge in domain is seriously considered in data mining process [24].

5. THE OBSTACLES FOR DATA MINING IN HEALTHCARE

One of the biggest problems in data mining in medicine is that the raw medical data is voluminous, and heterogeneous [27]. These data can be gathered from various sources such as from conversations with patients, laboratory results, review and interpretation of doctors. All these components can have a major impact on diagnosis, prognosis and treatment of the patient, and should not be ignored. The scope and complexity of medical data is one of the barriers to successful data mining. Missing, incorrect, inconsistent or non-standard data such as pieces of information saved in different formats from different data sources create a major obstacle to successful data mining. It is very difficult for people to process gigabytes of records, although working with images is relatively easy, because doctors are being able to recognize patterns, to accept the basic trends in the data, and formulate a rational decisions. Stored information becomes less useful if they are not available in easily apprehensible format. The role of visualization techniques is increasing in this, as the picture are easiest for people to understand, and can provide plenty of information in a snapshot of the results.

Doctor’s interpretations of images, signals, or other clinical data are written in unstructured free language, so it is very difficult to perform data mining of such data. Even specialists in the same area cannot agree on common terms that indicate the status of the patient. Not only that different names are being used to describe the same disease, but also tasks are getting even more complicated by using different grammatical structures to explain the relations between medical entities. Also, another obstacle is that almost all diagnoses and treatments in medicine are imprecise and subjected to error rates. Here the analysis of specificity and sensitivity are being used for the measurement of these errors. One of the unique characteristics of medical data mining is that the basic data structures in medicine are poorly mathematically characterized in comparison with other areas of physics science, because the conceptual structure of medicine consists of a description in words and pictures, with very few formal restrictions in the dictionary, image composition, or permissible relations between the basic concept.

Within the issue of knowledge integrity assessment, two biggest challenges are: (1) How to develop efficient algorithms for comparing content of two knowledge versions (before and after). This challenge demands development of efficient algorithms and data structures for evaluation of knowledge integrity in the data set; and (2) How to develop algorithms for evaluating the influence of particular data modifications on statistical importance of individual patterns that are collected with the help of common classes of data mining algorithm. Algorithms that measure the influence that modifications of data values have on discovered statistical importance of patterns are being developed, although it would be impossible to develop a universal measure for all data mining algorithms [28].

Data mining in healthcare can be limited in data access, since the raw inputs for data mining frequently exist in different settings and systems, like administrations, clinics, laboratories etc. Therefore, data must be collected and integrated before data mining can take place. Building of data warehouse before data mining begins can be a very expensive and time consuming process. Healthcare organizations that develop data mining must use big investment resources, especially time, effort and money [9]. Data mining project can fail from numerous reasons, like lack of managerial support, inadequate data mining expertise etc.

6. CONCLUSION

Data mining has great importance for area of medicine, and it represents comprehensive process that demands thorough understanding of needs of the healthcare organizations. Knowledge gained with the use of techniques of data mining can be used to make successful decisions that will improve success of healthcare organization and health of the patients. Data mining requires appropriate technology and analytical techniques, as well as systems for reporting and tracking which can enable measuring of results. Data mining, once started, represents continuous cycle of knowledge discovery. For organizations, it presents one of the key things that help create a good business strategy. Today, there has been many efforts with the goal of successful application of data mining in the healthcare institutions. Primary potential of this technique lies in the possibility for research of hidden patterns in data sets in healthcare domain. These patterns can be used for clinical diagnosis. However, available raw medical data are widely distributed, different and voluminous by nature. These data must be collected and stored in data warehouses in organized forms, and they can be integrated in order to form hospital information system. Data mining technology provides customer oriented approach towards new and hidden patterns in data, from which the knowledge is being generated, the knowledge that can help in providing of medical and other services to the patients. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments. With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data.

REFERENCES

[1] Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence , 37-54.

[2] Candelieri, A., Dolce, G., Riganello, F., & Sannita, W. G. (2011). Data Mining in Neurology. In Knowledge-Oriented Applications in Data Mining (pp. 261-276). InTech.

[3] Bushinak, H., AbdelGaber, S., & AlSharif, F. K. (2011). Recognizing The Electronic Medical Record Data From Unstructured Medical Data Using Visual Text Mining TechniquesProf. Hussain Bushinak. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 6 , 25-35.

[4] Eapen, A. G. (2004). Application of Data mining in Medical Applications. Ontario, Canada, 2004: University of Waterloo.

[5] Milovic, B. (2011). Usage of Data Mining in Making Business Decision. YU Info 2012 & ICIST 2012, (pp. 153-157).

[6] boirefillergroup.com. (2010). Data Mining Methodology. Retrieved 06 12, 2012, from Boire-Filler Group: http://www.boirefillergroup.com/methodology.php

[7] Zaïane, O. R. (1999). Principles of Knowledge Discovery in Databases. Department of Computing Science, University of Alberta.

[8] Milovic, B. (2012). Kompleksnost implementacije koncepta CRM. INFOTEH-JAHORINA Vol. 11, (pp. 528-532).

[9] Koh, H. C., & Tan, G. (2005). Data Mining Applications in Healthcare. Journal of Healthcare Information Management -Vol. 19, No. 2 , 64-72.

[10] Hosseinkhah, F., Ashktorab, H., Veen, R., & Ojaboni, M. O. (2009). Challenges in Data Mining on Medical Databases.

[11] Kusiak, A., Kernstine, K., Kern, J., McLaughlin, K., & Tseng, T. (2000). Data Mining: Medical and Engineering Case Studies. Industrial Engineering Research 2000 Conference, (pp. 1-7). Cleveland, Ohio.
[12] Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. international journal of medical informatics 77 , 81–97.
[13] Weiss, G. M., & Davison, B. D. (2010). Data Mining. Handbook of Technology Management, H. Bidgoli (Ed.), John Wiley and Sons .
[14] Getoor, L. (2003). Link Mining: A New Data Mining Challenge. SIGKDD Explorations Volume 4, Issue 2 .
[15] Using Link Analysis to Plan the Healthcare System. (2010). Retrieved June 05, 2012, from smartdatacollective.com: http://smartdatacollective.com/vincentg64/27210/using-link-analysis-plan-healthcare-system
[16] Milovic, B. (2011). Informaciono Komunikacione Tehnologije u Zdravstvu i u Razvoju e-health-a. Yu Info 17th conference, (pp. 154-159). Kopaonik.
[17] Ceusters, W. (2001). Medical Natural Language Understanding as a Supporting Technology for Data Mining in Healthcare. KJ Cios (ed.) Medical Data Mining and Knowledge Discovery, Physica-verlag Heidelberg, (pp. 41-67). New York.
[18] Matheus, C. J., Shapiro, G. P., & McNeill, D. (1996). Selecting and Reporting What is Interesting: The KEFIR Application to Healthcare Data. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press , 445-463.
[19] Canlas, R. D. (2009). Data Mining in Healthcare: Current Applications and Issues. Carnegie Mellon University, Australia.
[20] Gupta, S., Kumar, D., & Sharma, A. (2011). Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis. Indian Journal of Computer Science and Engineering (IJCSE) 188-195.
[21] Khan, F. S., Anwer, R. M., Torgersson, O., & Falkman, G. (2008). Data Mining in Oral Medicine Using Decision Trees. World Academy of Science, Engineering and Technology 37, (pp. 225-230).
[22] Ngan, P. S., Wong, M. L., Lam, W., Leung, K. S., & Cheng, J. C. (1999). Medical data mining using evolutionary computation. Artificial Intelligence in Medicine 16, (pp. 73–96).
[23] Shouman, M., Turner, T., & Stocker, R. (2012). Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients. 2012 International Conference on Knowledge Discovery (ICKD 2012) IPCSIT Vol. XX. Singapore: IACSIT Press.
[24] Stühlinger, W., Hogl, O., Stoyan, H., & Müller, M. (2000). Intelligent Data Mining for Medical Quality Management. Proc. Fifth Workshop Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP2000), Workshop Notes of the 14th European Conf. Artificial Intelligence.
[25] Koyuncugil, A. S., & Ozgulbas, N. (2010). Donor Research and Matching System Based on Data Mining in Organ Transplantation. J Med Syst (2010) 34 , 251–259.
[26] Houston, A. L., Chen, H., Hubbard, S. M., Schatz, B. R., Ng, T. D., Sewell, R. R., et al. (1999). Medical Data Mining on the Internet: Research on a Cancer Information System. Artificial Intelligence Review 13 , 437–466.
[27] Cios, K. J., & Moore, G. W. (2002). Uniqueness of Medical Data Mining. To appear in Artificial Intelligence in Medicine journal .
[28] Yang, Q., & Wu, X. (2006). 10 Challenging problems in data mining research. International Journal of Information Technology & Decision Making Vol. 5, No. 4 , 597–604.

BIIOGRAPHIES OF AUTHORS

Boris Milovic. Born 1969 in Vrbas. Finished elementary and high school in his home town. Finished Faculty of Economics in Subotica in 1994. Postgraduate Studies Certificate – MS in Economics in 2005. Phd Studies, University of Novi Sad, Faculty of Economics, Customer Relationship Management 2010.
Works as Commercial Manager and Assistant General Manager for the Investments in Sava Kovacevic ltd Vrbas www.savakovacevic.rs.

Dr. Milan Milovic. Born 1976 in Vrbas. Finished elementary and high school in home town, and then Medical Faculty in Novi Sad. A doctor in the orthopedics and traumatology department of Vrbas General Hospital since February 2004. Secretary of the Executive Board of the SPS OO Vrbas. Provincial Board member and member of the SPS Committee on Health PO SPS in.
Participant in several scientific conferences and courses in the field of orthopedic surgery and traumatology, at home and abroad. Author and co-author of papers in the field of study as a doctor.