The Sixth International Conference on e-Learning (eLearning-2015), 24-25 September 2015, Belgrade, Serbia

DATA MINING INFLUENCE ON E-LEARNING

DUŠAN PEROVIĆ

University of Nis, Faculty of Economics, bokaperovic@yahoo.com

Abstract: *Data analysis has improved many social aspects. Modern business is unthinkable without a proper use of data analysis tools. Also many education and security systems rely on this technique for solving their problems. With data mining there are bigger possibilities to improve on many ways, but there are still some obstacles that can slow down the whole process. Data mining has typical use in IT sector, but as the years have passed other sides have appeared and started to use data mining. The aim of this paper is show how data mining helped profit and non-profit organizations, but also to whole society to make better decisions and find a better direction for their everyday activities. Also, there will be mention some limitations and challenges about data mining use in modern world.*

Keywords: *data mining, making decisions, society, organization, effective business*

1.INTRODUCTION

One of the biggest challanges for every organization is to make proper decisions that will help them to improve in every single direction. To make a proper decision there must be some data that will be organized into databases, so later they could be used for everyday activities. Many people are usually confused with the amount of data and they have problem with managing those data.

Inside every organization knowledge management has a crucial role in all business and non-business processes. All data must be well organized, so everybody can know where to find a proper data that can be later used. Sometimes there is a need for upgrading older data, because they can be used in future, although their value is not significant as in past. Individuals use databases for discovering data that can improve their knowledge and skills. Use of modern communication technology such as internet browsers and social networks can help individuals to find needed data, but there is a still problem of finding right data. In the "ocean of data" question is what data can be useful for further actions?

Development of different professional searching tools can make every search more easier. Developing new search techniques is important in a modern world where huge movement of data takes places. Implementing rightful knowledgement strategy requiers well IT and management education so all activities can be done without bigger problems. Many new methods of learning rely on a good database, so search techniques are vital part of learning process, especially when IT technology is used for discovering new knowledge and skills.

2.DATA MINING IN EDUCATION

Every educational system has intention to offer best knowledge to people. With quality knowledge many people can do their work tasks much easier, but basic knowledge can also be base for further education. Lifetime education is something that is very represented all around the world and with every new knowledge and skill everyday work activities are no problem for knowledge-seekers.

Educational systems have changed as a consenquence of IT technology development. Computers, tablets, mobile phones changed some basic habits and helped people to find needed information on a much easier way. Access to the Internet, certain device and basic computer knowledge are required for all users. With changing some basic habits, amount of data increased. Databases are full of extraordinary data that can be used, but some of data need permanent change or to be deleted from database. Knowledge management developed few techniques or tools for better handling of data. Among those techniques is data mining.

Data mining is defined as discovering hidden knowledges, samples and rules inside huge databases [7]. For further analysis data mining usually relies on results from use of statistics, expert systems, machine learning, artificial intelligence and database managment. The main principal of data mining is extraction of previously unused data that can have significant use in future. Data mining is usually used for analyzing available data, which were not part of original research, but still they have scientifical value for some processes. Discovering rightful informations is going through some phases. First, we have preparation and selection of data. Here internal and external databases provide data for discovering relations between them. Second, when relations and values are determined, there is a need for upgrading database. This includes deleting useless data and transforming current data into precise ones. This is important phase for every research, especially for those with bigger use of statistic. Many statistical institutions usually have problem with enormous databases and there is a need for cleaning some data. Also, updating data time by time is important so older data can have use in future. Third and final step is data coding. With data coding all data can be used for statistics, artifical intellingence or machine learning and results will come very quickly. That way all data mining techniques have full implementation in certain processes.

The aim of data mining use is to discover hidden paths at time series behavior that is linked to different kind of data [4]. Data mining relies on various techniques such as queries, visualisation, case study method, decision trees, neural networks and genetic algorithms.

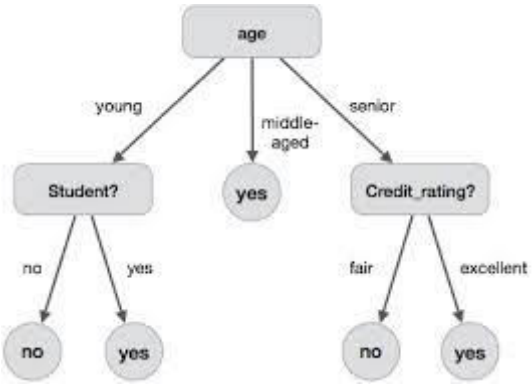

Image 1: Decision tree example

Data mining techniques can be very useful for everyday activities. Many financial companies use different data mining techniques for their operation on financial markets. Stock and credit markets can bring risks to many companies and if they want to count all postive effects from possible business ventures data mining can help them. With right selection and distribution of data, managers can make final decision that will for surely have an impact on whole organization.

In recent years main focus was moved from typical data mining techniques to educational data mining (EDM). Importance of knowledge database have influenced development of EDM. EDM is emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [11]. EDM is interdisciplinary that includes computer science, education and statistics. Mix of these areas gives subareas that are important for data mining process such as computer-based education, machine learning and learning analytics. The last one, learning analytics, improves decision-making data by integrating technical and social side of data. Learning analytics follows path of data with all their changes for better understanding and optimizing learning processes. This way, analysis of learning abilities is much easier and it is way for discovering new or improving current learning methods.

New way of searching and analyzing data represents a collaborative model where people can improve their knowledge and skill with progressive computer techniques. EDM is different to typical data mining because exploatation of multiple levels of meaningful data and arranging better and clear hierarchy in educational data. Also, it can provide software that analyze all data on a different levels such as location level, answer level, school level etc. EDM can turn data to be more feasible. This means full verification of data and improving their ecologycal validation. This allows better management of data with making easier learning processes. With transformation of data base, it will be easy to access data and begin research. EDM does not have limitations such as subject recruitment, scheduling of studies and data entry, since data is already online [1]. Here can some internal factors such as behavior and motivation be added in analysis. With addition of these factors difference between methods of learning can be seen. Most of students and professors have similar methods of learning but with the use of EDM difference can be revealed.

Data mining is widely used in business ventures for prediction of future outcomes. It can carefully examine behaviour of market participants which is important for making cruical decisions. In last few years use of data mining in education increased due to importance of quality reasearch. This is case, especially for higher education where searching methods represents a powerful tool for improving final result of some reserach. EDM can integrate abandoned knowledge into educational system and provide an assistance for decision-making process. With allocation of knowledge resources many processes in education can be more effective. EDM bring better view of students learning experience with providing enough data for academic stuff to improve their own methods of presenting the informations. From a learning perspective, data mining is being used in higher education to [3]:

- to assess students learning performance
- to provide feedback and adapt learning recomandations based on students learning behaviours
- to evaluate learning materials and web-based courses
- to detect atypical students learning behaviours.

There are different users of EDM, so there are many different objectives of its use. Learners or students use EDM to personalize e-learning. That way they can develop their own learning techniques and improve their knowledge and skills. Discussions about some problems will be raised to higher level as they can find and analyze data quickly. Also, this is good for changing working habits into more effective direction. Teachers or instructors use EDM to recieve a feedback about thier students learning results. It will be easier to follow the knowledgement path which is useful for detection of possible problems. That can help teachers to organize students into groups or to arrange proper learning method. Many lectures and working materials can be prepared much better with finding weak spots in learning processes and this leads to active role of teachers in improving student performances. Researchers use EDM for evaluation of their entire work. They usually combine EDM techniques and compare them, because of discovering the best ways of finding and implementing knowledge. With a proper EDM technique researchers can construct effective model of learning that can be used in future work of researchers. For organizations and administartive stuff EDM is one of the best way for organizing all available resources in more effective way. Decision-making process will be enhanced and many organizations can improve their data bases with deeper analysis of currrent data. EDM can be useful for designing future educational program and for improving

many aspects of organization. This is good not just only for educational institutions, which represent non-profit organizations, but also to profit orientied organization. It will help organizations to recognize weak spots in their sturctures so they can strenght them and not having problems in that part of structure anymore. EDM can help profit oriented organizations to analyze market better and to know where to invest money.

Traditional eductional system relies on face-to face contacts between teachers and students organized through lectures, class discussion, individual work or working in groups. In traditional education, teacher follow students performances through their test results and observe their active role on classes. Students marks, attendence, goals and working plans are carefully followed and all data about their performances are stored in databases. These databases contain various data about students, teachers, schedules, literature and access to needed informations. Sometimes it takes more time to access and find rightful data, so many educational institutions need better organization in data base management.

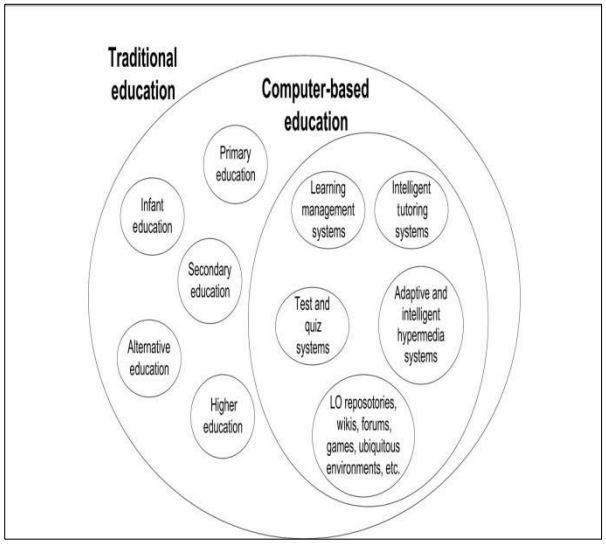

Image 2: Traditional vs computer-based education

Unlike traditional education, computer-based education emphasizes the role of computer in educational system. Through computers students get directives for further learning process. They usually access Internet and with the help of artificial intelligence they can personalize all conditions for finishing work tasks or their research. Learning management systems and intelligent tutoring systems offer students new ways of e-learning as they can have better comunication channels and information sharing is much quicker. Students can improve their working, reading, writing and talking skills with less time than usual. Offer of many interesting courses is huge and it can be customized for every single person, which is good because of different habits that students possess. All what do students need is good will and good access to modern technology, if they want to improve their performances.

3.APPLICATION OF EDM TECHNIQUES

EDM can improve overall quality of education with transforming many traditional methods of learning into powerful tools for future science challanges. Overall improvement in education is a priority task for every country, because many developed economies rely on highly-educated labour, so all processes can go faster and bring welfare to country. Investment in education is also investment in new researches, technologies and ideas that can change many processes in better direction. EDM techniques that have common use are: prediction, clustering, relationship mining, discovery with models, social network analysis and distillation of data for human judgement.

Prediction is used for developing model where combination of data viewed from diffrent aspects can effect single viewed data. Here, is important to watch how predictors variables are effecting predicted variable. For example, if we want to follow some student scores on a test, then we have to assume possible factors that had influenced on final scores. There are a lot of possible factors and because of that whole research must be very carefully done. Prediction requires big degree of data reliabilty, so true value of data must be determined. This method can be used for giving valuable information about construction process. Any factors that can change final results are not important in the beginning, but in reality every score is determined by internal or external factors, so they have to be included sometime during the process. For example, prediction of students final performances is obvious result of their activity, time spend reading, learning educational materials from lectures, and discussion with other collegues or teachers. Prediction can be also used in process of recognizing any risks for acomplishing the final performances. Student behavior is the central theme of prediction method. It is important for educational institutions to follow work of their students and see their reaction on any changes. Behaviour is very sensitive category and because of that, data have to be fully analyzed and then other steps in this process can be made.There are three types of prediction: classification, regression and density estimation. Classification is used when predicted variable is categorical variable. It has common use in predicting class of objects whose class label is unknown. Classification relies on use of decision trees, which can help many organizations to arrange data so they can be used for proper reason in future. Regression is used for prediction of continuous variables that can change faster in future. For this type of prediction neural networks and linear regression are used. Density estimation is used when predicted value is a probability density function. The density estimation problem is to construct a density function of the distribution based on the data drawn from it. By knowing the density distribution of a data set, we can have an idea of the distribution in the data set. Kernel density estimation is usually used for finding dense area in data set very quickly so medians can be calculated with no problem.

Clustering has typical use in statistics, but also in data mining. The aim of this technique is to identify groups that are similar in some aspects. It is important to define parameters first and then start identifying and forming data clusters. Cluster analysis divide data into meaningful groups with a stronger or weaker influence between them. Sometimes stronger connection is between clusters than data that belong to that groups. Degree of association is not same to all groups due to data nature. Cluster analysis provides different techniques such as k-means clustering, hierarchial clustering and fuzzy c-means algorithm. These algorithms can start with or without prior hyphoteses and it depends form the research type.

For detecting undesirable student behaviour and for detailed analysis of e-learning use clustering represents very powerful tool. Combined with classification clustering offers better analysis into next five steps [6]:

1. Logging the data;
2. Data preprocessing;
3. Index computation;
4. Metrics calculation;
5. Classification.

The first step is to identify the richness of educational content. The goal is to determine if educational content has high or low enrichment value. This stage results into two courses one with rich and one with poor content. Data preprocessing is doing some correction on data values. Data with missing value and that without good structure must be first improved for further processes. For e-learning process it is important to see what URL adresses have students used so more interesting content would be prepared in future. Here is important identification string of each field and each session. Index computation is used for evaluation process. Indexes such as Sessions, Pages, Unique pages can help the evaluation process. Metrics calculation consists of calculating enrichment and homogeniety values of data. This way we can have classification of higher, average and low educational content which represents basis for educational courses. Usually, courses with higher enrichment and homogeniety value are popular among students, because they offer quality knowledge. At the end, all couses are classified into different groups as a results of their metric results. Many popular online courses are usually marked with a sign so students can recognize that they posses quality.

Relationship mining is also one of the most used EDM techniques. This technique can help in a process of discovering different type of relationships between variables among huge amount of data. This way relationships between data can be observed and see if some data have strong or weak relationship between them. Discovering of variables strenght can give us valuable information for future research. There are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining. Association rule mining is based on IF-THEN rule that exceed a minimum "support" and "confidence" [10]. If one value of variable is found, then other variable will have specific value. For example, if student goes to lectures regulary and if he reads literature then he will pass the exam. IF-THEN rule can be found in statistics regulary, but also can be implemented in social sciences. Correlation mining provides help with linear correlation of variables. Linear correlation can be negative or postive and some statistical and analytic tools such as SPSS or STATA examine linear correlations a lot. Sequential pattern mining can find temporal association between variables. This type of relationship mining is excellent for discovering factors that have influenced on students performance at some moment. For example, student did not pass exam because of less activity on classes or because he was too stressful or did not sleep very well. This way we have closer look at behavioural aspect. Causal data mining is trying to find if one event was caused by other. Also, it can help researchers or teachers to see the original cause of some event. Teachers can examine if good results on exams are result of hard work or something else influenced final grade. Relationship mining is good for discovering behaviour patterns of students and it also recognizes critical factors that can make learning process difficult.

Discovery with models combines other data mining techniques (usually prediction and clustering) to develop a model that can explain some variables. Created model can be used as predicted variable for predicting new variable or to study relationship between created model and additional variables. That way new pattern can be discovered and implemented in future actiivities. Discovery with model can be used for discovering if students can trick educational system like cheating on exams or using forbiden technology. Students behaviour and characteristics are followed very carefully as they use new learning methods.

Social network analysis (SNA) is relatively new EDM technique, but with bigger use of social networks it had found place in EDM methodology. SNA is used for mining to interpret and analyze the structure and relations in collaborative tasks and interactions with communication tools [8]. This technique is trying to find connection between social network users personality and their activities on social network. Every picture, comment and action are involved in examination and with analyzng personal characteristics of user (age, physical characteristics, martial status, occupation) behavioural patterns can be discovered. The only doubt can be reality of new pattern, because many users give false informations, so only true data can be observed.

Distillation of Data for Human Judgement consists from descriptive statistics that gives valuable data for decision-making process. Information visualisation and graphics help to explore and understand large amounts of educational data, so this way teachers have better view about students activities. When activities are identified, teachers can recognize current and can predict future behavior of students. Visualisation of data brought new

dimension of solving everyday problems and with the assistance of graphics, curves and tables it is much easier to understand current situation and to start planning future actions.

4.NEW ROLE OF DATA MINING

It is obvious that managing large amount of data is not so easy. Many educational institutions and companies have huge data bases and they are growing every day. There is a need for upgdating data, so they can be used any moment and without any problem. Technical support is a priority, but also many employees must be familiar with data mining so they can have use it for group and individual progress.

Data mining was recognized by many big companies as a chance for increasing profit. Data can impove many aspects of any business organization. With identifying and analyzing right data decision-making process will not be stressful any more. Future plans for production and for trading can be improved and also many products can be personalized for customers. Data about customers are the most important for companies, because they bring profit to company and everything must be specially prepared for customers. Competition on market is very dynamic and identifying right data will be priority for further existance of companies.

In education, data mining has a different role. The aim of data mining use in education is to discover weak spots in learning process and to improve not just them, but every single aspect of learning process. There are many specific objectives in EDM depending on the viewpoint of the final user and problem to resolve such as [2]:

- How to (re)organize classes, or assesment, or the placement of materials based on usage and performance data.
- How to identify those who would benefit from feedback, study advice or other provided help.
- How to decide which kind of help, feedback or advice would be most effective.
- How to help learners in finding and searching useful material, individually or in collaboration with peers.

Data mining in education has a little different role. It can help decision-making process, but also finding and analyzing data that can improve educational facilities. EDM is good for examining learning process and for having closer look to students methods of learning. Analyzing learning process helps teachers to understand all difficulties that students have and that can be a sign for teachers to make changes in thier educational materials and system of teaching. This way pedagogical approach is highlighted and educational institutions and their employees (teachers) have bigger responsibility for further progress of educational program.

Collecting and analyzing data requires changes at organizational structure. First, there must be better information support for all processes in organization, because data are moving very fast and change almost every second. Teachers must be familiar with appropriate information support so they can know what to change in their current work. Data can be collected from administrative stuff, exams, questionnaires, experiments, different kind of observations, scientific magazines, even from a conversation with students. Every part of huge data base can provide crucial help for building effective learning process, so every side can have benefits.

EDM gathers around computer experts, scientists, psychologists, researchers, and many instructors and teachers. With developing and implementing EDM techniques it is easy to discover hidden patterns and relationships in learning process. Behaviour of students can be examined from many aspects and final results should represent directions for changes in educational program. Conventional methods of learning are overcomed by improvement of new technology, especially with development of knowledge management. Educators and institutions should develop a data-driven culture of using data for making instructional decisions and improving instruction [9]. Results from data mining analysis must have wider influence and not to be limeted just for simple research. Creating of an improved model of learning is obligation for both teachers and students. Behaviour observation is complex process, but it can help to find right relationship between students final performance and their learning processes. Many activities are results of internal factors, so it is good to find out something about emotions, way of thinking, habits that students posses.

Data mining in education is focusing on two groups: learners and educators or learning providers. Before starting any kind of course or exam it would be better for educators to know something about thier students (learners). Data base can have information about their gender, age, place of birth, occupations and hobbies, qualifications, but it is good to have data about their behaviour. This includes thier behaviour on some courses in past or how was their communication on previous education levels. Students visions, ambitions, emotions, thoughts should be included in analysis, so appropriate organizational plan would be prepared. For discovering hidden patterns every kind of data is useful, especially data about possible students reactions in some situations. Modern technology can develop models of students behaviour and with the help of simulations researchers can come to an interesting conclusions.

Educational institutions and educators can have benefits from applying data mining techniques. They can observe current educational programs and materials and to follow students performance during a certain period. Data mining is good for analyzing courses from diffrent angles, which allows teachers to see where do they make mistakes and what can they do to improve current courses. Applying data mining techniques will draw attention to areas in which management should invest more, in terms of

degrees, professional certificates, courses and modules [9]. Educational institutions must work on discovering new areas of knowledge that can be implemented through educational program. Many jobs require special qualifications and educational institutions must be among first representatives of knowldge transformations. All important social and scientific facts must be included in process of transforming educational institutions into knowledgement leaders.

Connection between sector of economy and education must be strong and interactive. Many big companies have formed partnership with many universities and scientific institutes so they can overcome many problems that they are facing on markets. Many big companies from IT sector such as Google, Yahoo, Apple have very good collaboration with some universities in process of developing new technologies. Silicon Valley, Research Triangle and Silicon Slopes represent good way of partnership between companies and universities. Educational institutions can help with data mining techniques to change learning methods of employees in many companies and to allocate all needed resources for further progress of company. This way both sides can be pleased. Companies will survive dynamic market changes and increase profit. On the other hand educational institution will build better reputation and they will continue working on discovering of new learning techniques.

Data mining is changing face of many organizational process. With data mining use new space for further search about learning methods is open and it gives enough time for evoluation of current results in data mining and for experimenting new techniques. Complexity of social processes has pointed out need for tools that can help in discovering hidden patterns of human behavior. This means, that with data mining help something new about learning methods can be found.

5.CONCLUSION

Dynamic changes on market influenced many organizations to start changing their organizational culture and philosophy. Data had become intangible assets in every organization and they represent basis for activities of profit and non-profit institutions. For many universities and schools it is important to have developed methods of presenting informations to students, but also these informations must provide high quality knowledge. This means, better communication with students and collecting valuable data about them. These days, educators are trying to get into students heads and figure out their patterns of behavior. It is important to adjust eduactional material and learning methods to students, so they can obtain quality knowledge and be succesful in their future work.

Various data mining techniques can help in process of learning. Their use has been growing in recent years, which shows their significance for research work. EDM can help educators foe self-improving activities. It is good for educators to be familiar with their weak spots and disadvantages so they can work on improving their own learning and work skills. If they work on themselves, quality of lectures and educational materials will be much better and there will be less problems for educators.

Learning analytics became important in education, so developing new techniques is always welcomed by educators and students. By using data mining techniques, educational institutions are becoming winners, because they can attract more students, investments in research work will increase and repution of educational institutions will be recognized by others. Learning process is important part of every human life and life-time learning is something that includes not just learning something in school, but also outside of that educational institution. With achieving more knowledge and skills, people will live better and will not be scared for their future.

LITERATURE

[1] Baker, R., Data mining for education, *International encyclopediaof edition*, 7, 2010., pp. 112-118

[2] Calders, T. Pechenizkiy, M., *Introduction to the special section on educational data mining*, ACM SIGKDD Explar, 13, 2011., pp. 3-6

[3] Castro, F., Vellido, A., Nebot, A., & Mugica, F., Applying data mining techniques to e-learning problems, *Evolution of Teachning and Learning Paradigms in Intelligent Environment*, 2007., pp. 183-221

[4] Han, J., Kamber, M., *Data Mining: Concepts and Techniques,* Second Edition, Morgen Kaufmenn Publishers, San Francisco, 2006.

[5] Hanna, M., Data Mining in the e-learning domain, *Campus-Wide Information Systems*, 21(1), 2004., pp. 29-34

[6] Kaur, M., Krishan, K., Cluster Analysis of Beahavior of E-learners, *International Journal of Soft Computing and Engineering*, 3(2), 2013., pp. 344-346

[7] Milovanović, S., *Upravljanje znanjem*, Ekonomski fakultet, Niš, 2008.

[8] Rabbany, R., Takaffoli, M., Zaiane, O., Analyzing participation of students in online courses using social network analysis techniques, *International conference on Educational Data Mining, Eindhoven*, The Netherlands, 2011., pp. 21-30

[9] Romero, C., Ventura, S., Data Mining in education, *WIREs Data min knowl discov*, 3, 2013., pp.12-27

[10] Scheuer, O., Mclaren, B.M., Educational Data Mining, *Encyclopedia of the science of learning, Springer*, New York 2011., pp. 1075-1079

[11] Web site: www.educationaldatamining.org (date of access 20.07.2015)