ORIGINAL PAPER

A DATA MINING BASED MODEL TO IMPROVE UNIVERSITY MANAGEMENT

CRISTINA OPREA [1], DELIA MIOARA POPESCU[1], ANCA GABRIELA PETRESCU[1], IRINA BARBU[1]

Manuscript received: 04.02.2017; Accepted paper: 30.03.2017;
Published online: 30.06.2017.

Abstract. *The main purpose of university managers is to manage as well as possible the institutional resources, the academic and administrative services. Thus they have to be able to manage and use a large amount of data and information, inside the own university as well as outside of it. The big volume of information increases the difficulty of decision making.*

The paper emphasizes an efficient way to structure the decisional process in the superior school, using a generic architecture that contains data mining technology. Being a process of knowledge discovery, data mining uses data to create models through a predictive mechanism.

Having as purpose the improvement of the decisional process, the model proposed in the current paper uses software instruments of last generation in the field of web development, data bases and interface design. This allows the development, inside universities, of a system that analyses the workflow, being able to identify the deficiencies and assure making good decisions.

Using specific algorithms for statistics and searching techniques, the model extracts information meant to improve the learning process, to correlate the educational offer with the work market, to lead to increase of performance in the process of making decisions.

Keywords: *higher education, university management, decision-making, data mining.*

1. INTRODUCTION

Globalization is a key factor that has profoundly influenced the higher education under the influence of an increasingly integrated world economy. Higher education institutions are called upon to permanently redefine them selves under the influence of both staff's mobility and of the increased availability of the information and ideas. The economic and the cultural globalization have ushered a new era in higher education, the universities being part of a single network that ensures their visibility worldwide. The ubiquity of the changes in technology, economy, knowledge, people, values and ideas has imposed, on the other hand, these institutions control endurance.

At European level, the objective of the Lisbon Strategy of the European Union (2000-2010)-creating a Europe of knowledge-has imposed a comprehensive regional integration, the compatibility among higher education systems throughout Europe. Universities need to be in constant compliance with the external environmental changes, in order to implement change quickly, to make permanent choices that directly affect teachers and students. When being

[1] Valahia University of Târgovişte, Doctoral School of Economics and Humanities, 130105, Târgovişte, România. E-mail: oprea_cris2005@yahoo.com; depopescu@yahoo.com

part of the change process, taking measures involves a rapid assessment of the results by the involved parts. In order to maximize their effectiveness in the global environment, it is essential for the universities to retain a strong sense of identity and purpose on one hand and on the other hand, it is essential to be open and engaged in collaboration with others. Being under the influence of emerging technologies, they must venture into the virtual blend of the educational models used in the global digital economy.

Over the past decades, the Romanian university has been facing major changes, from the transition to the university autonomy and the massive increase in student numbers, to the change of the traditional process of teaching/assessment by introducing online courses. Due to the evolution of information technology that plays an increasingly important role in all fields, the current university system in Romania needs efficient and appropriate systems adjusted to the continuous technological dynamics.

2. LITERATURE REVIEW

The educational system computerization requires the development of new organizational structures, new ways of management and new ways to use the information technology [28]. The recent developments in information technologies and communications have caused a fundamental change both in the needs for knowledge acquisition and their ability to respond, and in the way the study programs are offered in higher education institutions. The concerns regarding the symbiosis of the specialized training and the general training of the students' knowledge acquisition process are favorably reflected on the course of business. These concerns contribute greatly both to the quick assimilation of new achievements in science and technology, as well as to achieving an increased efficiency. But one can not overlook that, during the university administration activity, there are frequent cases when the institutions have achieved low efficiency when the managerial decision-making process has become more complex. The Higher Education Institute seeks more effective technologies to better manage, support the decision-making process and help them to establish new strategies to better manage current processes. One way to effectively address the challenges of improving quality is to provide new knowledge about educational processes to the managerial entities,too. These pieces of knowledge can be extracted from the historical and operational data residing in databases of HE institutions, using the techniques of data mining technology [7].

By integrating OLAP, data mining techniques which can assure multidimensional data analysis, as well as fraud detection and the discovery of the data hidden behind the classical knowledge, new opportunities are provided to the auditors.

The literature abounds with definitions of the concept of data mining, but in essence they express the same thing: a process of significant knowledge discovery from large amounts of data stored in databases, data warehouses or warehouses whose information is used, especially in industry, in the media, among statisticians, data analysts, specialists in database and in the communities dealing with management information systems [13].

Knowledge discovery techniques are considered analytical tools that can improve the quality of managerial decisions. Al-Twijri and Noamanb propose a data mining model that supports higher education institutions in the decision-making process at the strategic level by detecting the best way for the admission of students to university [16]. Shahira, Hussain and Rashida provides an overview of techniques for exploiting the data that has been used to predict the students' performance, focusing themselves on how the algorithms for prediction can be used to identify the most important attributes in a set of data [26].

Sușnea develops a conceptual architecture model for an integrated IT system (IDSS) to help decision makers in the military academy take the right decisions at the right time [4]. The IDSS architecture includes three subsystems: the data management system (DMS) that provides the data needed to develop IDSS models, the models management subsystem (MMS) - a software package based on the data provided by DMS and the use of data extraction techniques, generates models and the user interface (UI) that facilitates communication between IDSS and users (Figure 1).

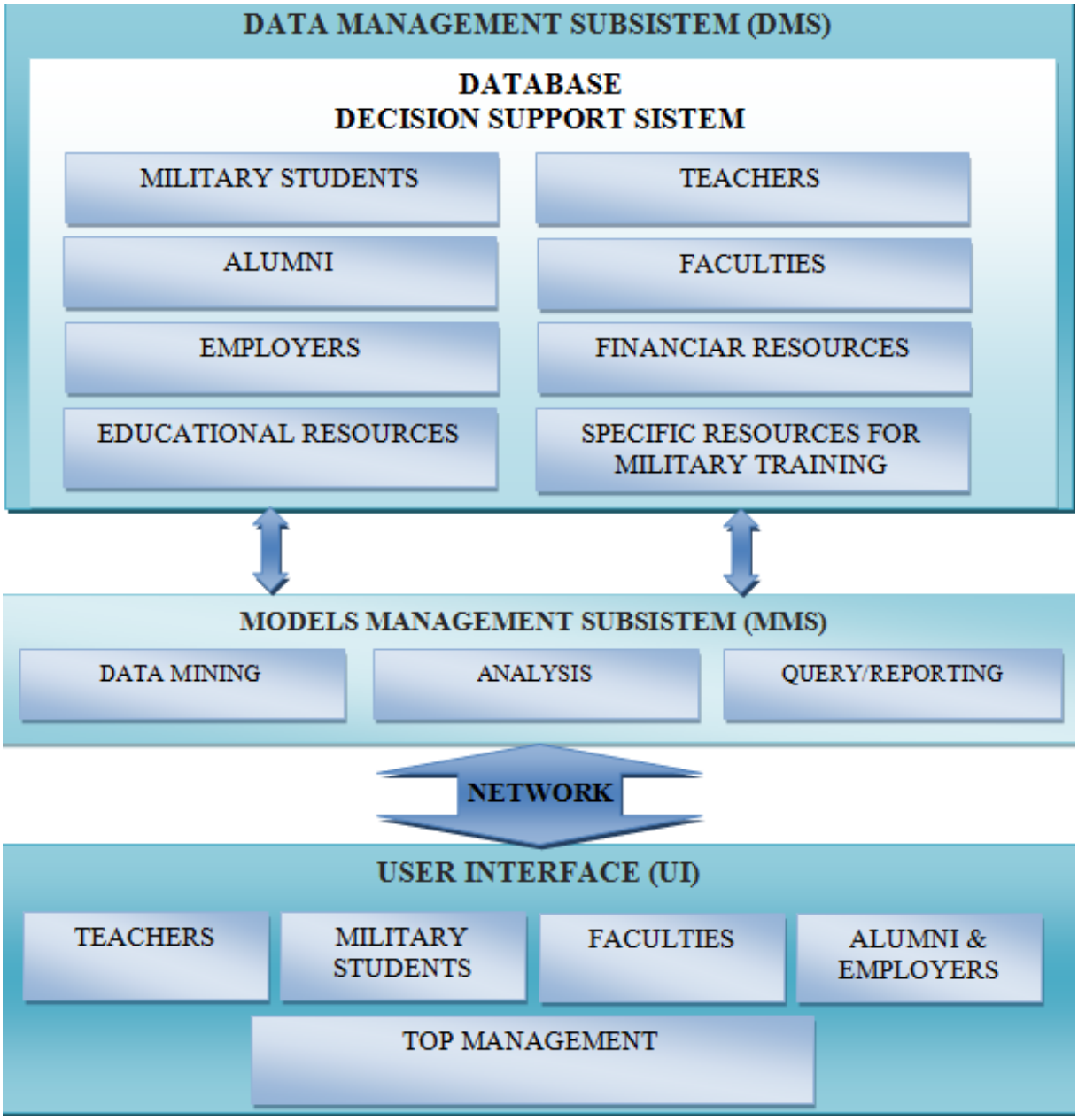

Figure 1. IDSS Architecture [28].

This conceptual model allows users to access data from multiple sources and chooses a different level of data aggregation (top management, students, and teachers). Discovered models allow recipients to analyze data on academic and military perspectives.

3. RESEARCH METHODOLOGY

The proposed educational data mining model (E.D.M. model) stores and processes data from an institute of higher education in an efficient manner, in order both to improve the teaching process or to predict the students' performance and behavior and for administrative purposes. According to the literature, the data mining process includes the following activities[10]:

- *business understanding* that requires a clear definition of the objectives to be pursued in accordance with the purpose of research;
- *data understanding* that aims the extraction of the relevant data for the analysis from the available massive data;
- *data cleaning* dealing with data cleaning and with the preparation of the activities that are necessary to ensure the correct results;
- *data transformation* that converts the data into a two-dimensional table and removes the unwanted fields, so that the results are valid;
- *data mining,* extracting patterns from data, is to analyse data through a set of appropriate algorithms to discover meaningful patterns and rules and producing predictive models;
- *data valid*ation that requires the proper interpretation of the results of data mining and aims to select those models that are valid and useful in future decisions in different areas.

The purpose of exploring data from a university is [20]:

- Classify the students according to their performance;
- Estimatethe students' performance (the credits obtained, their marks/grades);
- Identification ofthe patterns regarding the students' progress in the future (MA, Ph.D programs);
- Notification and emphasize the factors that have contributed to high performance learning;
- Estimate the degree of correlation between the specialization and the professional route chosen.
- Identify the students who tend to benefit from their preferment opportunities or to abandon their studies;
- Identify the profile of students who are tempted to transfer themselves to another university;
- Identify the profile students who obtain most loans;
- Identify the courses that are preferred by students;
- Identify the courses that are normally requested together;
- Identify the specialization that is preferred by students.

Setting the variables whose data are to be analysed is one of the main activities leading up the modelling activity, due to the fact that the variables may be interrelated or redundant. In this phase, the necessary raw data and attributes are collected according to the goals. In accordance with the objectives set, the raw data are related both to students/graduates, teachers, and the administration. Thus, the following characteristics have been identified (Figure 2):

- Demographic and social characteristics of teachers;
- Characteristics of teachers' education;
- Geographical characteristics of students / MA students;
- Demographic and social characteristics of students / MA students;
- Characteristics of students'/MA's education;

- Students'/MA's economic characteristics
- Administrative and financial characteristics of the university administration.

According to the studies [3, 5, 6, 16, 23, 26], the results achieved by students is the best performance indicator of the education system. Therefore, the attributes such as "the final college graduation grade", "the license exam grade", "the number of credits accumulated", "the job obtained after graduation" can be used as purpose variables in the analysis conducted with the E.D.M. system assistance.

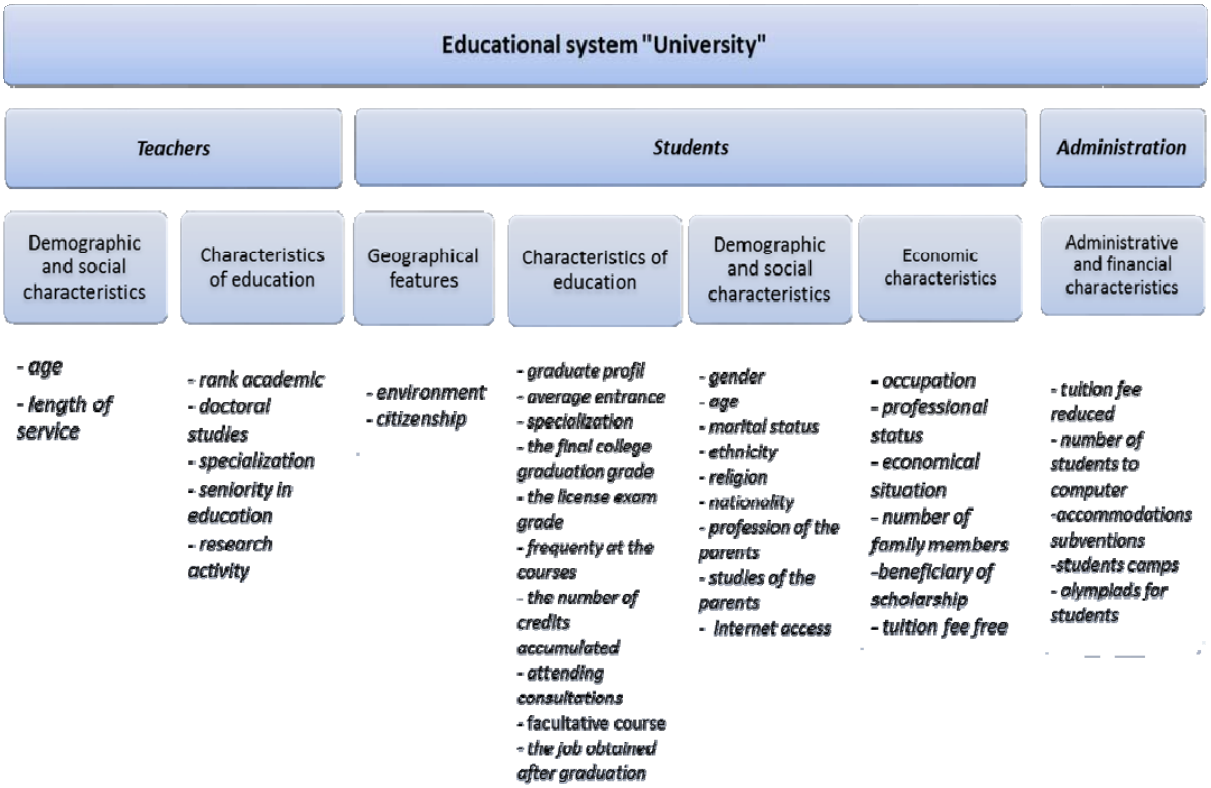

Figure 2. Relevant characteristics for DM analysis.

Weka take over files resulting from SQL database query, but it does not allow the declaration of attribute in terms of stored data, nominal value, numeric or string. That is why it is mandatory that the Weka data entry to be organized as .CSV files (Comma Separated Value) or .ARFF (Attribute-Relation File Format). Therefore it was considered opportune to develop a converter to ensure the transformation of information extracted from the database university .CSV or .ARFF.

The data conversion procedure was performed by using the functions and the functionality present in Microsoft SQL Server 2008 R2 and the functions developed in Microsoft SQL Server 2008 R2 and Embarcadero Delphi XE2 Enterprise. By creating the SpGetDateCSV.sql procedure, the Weka application data export is allowed. Then the functions which convert the nominal values in numerical values are called.

Data cleansing involves removing null values that may cause the distortion of the results. This can be achieved by using unattended filtering data from one application filters provided by WEKA.

For *the application of data mining techniques*, the data mining algorithms can be grouped into the following categories: classification and regression, clustering and association rules.

Classification and regression creates models for predicting the membership sets of classes or values: the decision trees, the Bayesian technique, the neural networks and the K-nearest neighbour.

Techniquely, there can be built several decision trees given an attributes set, but some of them have a better classification precision then others. Thus, there exists a serie of algorithms that can be used to obtain trees with an accuracy as high as possible. The most well known algorithms are: Hunt Algorithm, CART (Classification and Regression Tree), ID3, C4.5, CHAID, SLIQ, SPRINT, QUEST, FACT, THAID.

Hunt algorithm is the base of algorithms that create decision trees. We consider Dt the set of elements in node t, and C={C1, C2,..., Ck} is the set of class labels assigned to node t, so we will have 2 situations:

If Dt contains elements of the same class Ct, then t is a leaf labeled Ct

If Dt contains elements of more then one class, then we choose an attribute test for dividing the Dt set into subsets (nodes). [15]

The process is applied recursively to each node.

A fundamental problem in building a tree is how the attributes are selected for each node of the tree. The goal is to get the most accurate division of a subset of data in a node, so that a higher purity of son-nodes can be achieved.So, choosing attributes for achieving the most accurate classification is based on the purity level of son-nodes. In order to determine the level of impurity we use the following formulas of calculation:

1. Entropy : showes how „desordered" is a set of data:

$$Entropia(S) = -\sum_{i=1}^{c} p_{i/S} \log_2 p_{i/S} \tag{1}$$

where,

S – objects set

Pi/S – weight of elements i in set S

C – number of classes

If entropy is 0 then all objects in S are in the same class. If entropy is 1 there is an equal cardinal number in each class, and if it is between 0 and 1 the number of objects is different.

2. Gini index, used mainly in CART and SPRINT, is based on selecting that partitioning attribute that minimises the impurity of division (2)

$$I_G(S) = 1 - \sum_{j=1}^{c} p_{i/S}^2 \tag{2}$$

The optimum partition of node i is that which assures the smallest value of partition index GINI.

3. Information gain: used mostly in ID3, C4.5 and C5.0 and is calculated as the difference between the impurity level of father-nods and impurity level of son-nodes. The higher the difference, the better the chosen attribute.

$$Gain(S) = Entropia(S) - \sum_{i=1}^{k} \frac{N(v_j)}{N} Entropia(v_j) \tag{3}$$

where,

Entropy (S) – entropy of father node S

N – number of objects in father node

k – state number of chosen attribute for classification
N(vj) – number of objects that belong to son-node vj
Entropy (vj) – entropy of son-node vj

Less implemented in applications for exploring data, the Bayes technique is based on probabilities technique with the same name (3). [24]

Given B an arbitrary event in ∑ and {A1, A2,…, An} a partition of space Ω. Then:

$$P\{A_i \mid B\} = \frac{P\{B \mid A_i\}P\{A_i\}}{\sum_{i=1}^{n} P\{B \mid A_i\}P\{A_i\}}, \; P\{B\}>0,\; P\{A_i\}>0,\; i=1,2,..,\; n. \tag{3}$$

Bayes is a clasification technique with a predictive and also descriptive potential. It permits analysing the relationship between each independent variable and dependent variable, by calculating a conditionate probability foe each of this relationship. When a new instance needs to be classified, the prediction is realised by combining efects of independent variable in the dependent variable.

Artificial neural networks are particularly interesting because they deliver an efficient modelation method of complex problems that can have hundreds of independent variables inbetween which there exists lots of interactions [8].

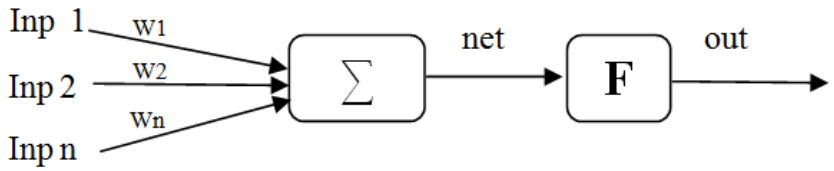

Figure 3. Neural Networks (McCulloch-Pitts Model) [14].

where:

Inp 1, Inp 2,…, Inp n – input values;
W1, W2,…Wn – input wheights.

$$I_{net} = \sum_{i=1} Inp_i W_i \tag{4}$$

$$Out = F(net) \tag{5}$$

The technique is based on two concepts from the artificial intelligence. The artificial neuron represents tha base unit for processing information in a neural calculation. Through analogy woth the biological neuron , it has been difined as a unit that processes infoamtional input and generates output. The artificial neural network is an assembly of artificial neurons, linked by connections. Neural networks are dinamic systems, whose behaviour can be characterised by traking the states at different moments in time. The state of network in a particular moment can be defined by the assembly of neurons activation levels and the intensity of connections between neurons. In addition to these adjustable parameters a network is definied by the following non-variable parameters: connections configuration and type of activation functions.

Nearest Neighbor (K-NN), is a classifing method based on learning through analogy, meaning on searching solutions obtained for similar problems. This method classifies a new object/case based on the closest (k) objects / cases in the neighborhood [12].

The K-closest neighbor algorithm is based on two principal elements:
- Number of closest cases that will be use;
- An indicator to measure the distance between data values.

This method requires more and more calculation equipements as the number of potential neighbors increases for each unknown case. That is why there needed efficient indexing techniques [13].

While appling a decision tree or a neural network to a new case is rapidly realised, the K-NN method imposes a new calculation for each new case. For increasing the calculation speed of K-NN algorithm, frequently, all data is stored in the memory. K-NN method is efficient when there are only a few independent variables and is very useful when there are built models that imply types of data that are not standard [8].

Clustering seeks to identify a set of categories (clusters) for describing the data and divides a database into different groups. The purpose of clustering is to find different groups, whose members are similar components. Clustering algorithms are grouped into five categories: partitioning, hierarchical, evolutionary approach, dense-based, model-based and graph-based algorithms. K-means algorithm is the representative partitioning clustering algorithm.

Association rules pursue the connections between data detection, based on associations or on the sequences discovery. The most popular algorithm for discovering association rules is Apriori. It uses the property of frequent itemset, any subset of an articolset frequently has to be frequent.

The validation of models requires an analysis of the applied algorithms performance.

4. THE DATABASE

In Romanian universities have been identified four sources of data that provides information about candidates, students, master students and graduates. These data sources are:
- Data coming from the admissions committees of university in which are introduced information about the candidates in the entrance examination. Information is entered into the database via a sign-up form to be completed on the basis of a dossier.
- The current database of the university created and maintained by the department of information and communications technology (ICT) of the university. The pieces of information in this database are updated throughout the academic year, being “a mirror” of the activity of each student and master student of the university. The data is collected by the secretarial network.
- The data from the university departments where there are withheld information about the tenured professors and the associates.
- The evidence from the interviews based on questionnaires of three categories: the students, the master students and the alumni of the university.
  The questionnaires are completed online and include questions regarding the financial situation and the students' family , the parents' education , the risk of abandonment, the courses, seminars and consultations attendance, the conditions of the labs, the Internet access, the future plans, the perception on the follow faculty, student services issues/administrative resources and spaces of learning, the curriculum, curricula and courses of study and work obtained after the completion of the study program.

The data collected were unified into a relational database that contains 37 tables organized in 8 package: CANDIDATES, STUDENTS, MASTER STUDENTS, TEACHERS, NOMENCLATURES, QUESTIONNAIRES, LOGIN and SETTINGS.

The information about the candidates considered admitted at the entrance exam are copied in STUDENTS tables respectively MASTER TABLES to be supplemented by other data: grades, the number of failed exams, the number of credits etc. The TEACHERS package contains materials about the university teachers divided into 5 tables. The QUESTIONNAIRE package contains questions online questionnaires completed by students, master students and graduates.

5. RESULTS AND DISCUSSIONS

In order to design the educational data mining model, there has been created a client-server architecture on three levels developed and managed as independent modules (Figure 4). The use of such an architecture offers several advantages: both the sharing and reusing of the created components and services, the data protection and security are easier to obtain and maintainand there is the possibility of implementing the components and services on a server to keep up with the changes.

The architecture presentation level is the user interface and is the topmost level. This model enables the communication between the users and the system by using web-browser-type client applications. Most browsers are Microsoft Internet Explorer, Mozilla Firefox, Apple Safari, Google Chrome, Mozilla Camino, Opera Software Opera browser DS Nintendo and Flock.

The Logic level controls the functionality of the application and the logic rules of application. This level includes the application servers, the web server, the application logic and the server pages. The application servers provide communication services, security and persistence. They receive applications from the web level, they handle applications by querying the database and they send the reply to the web.

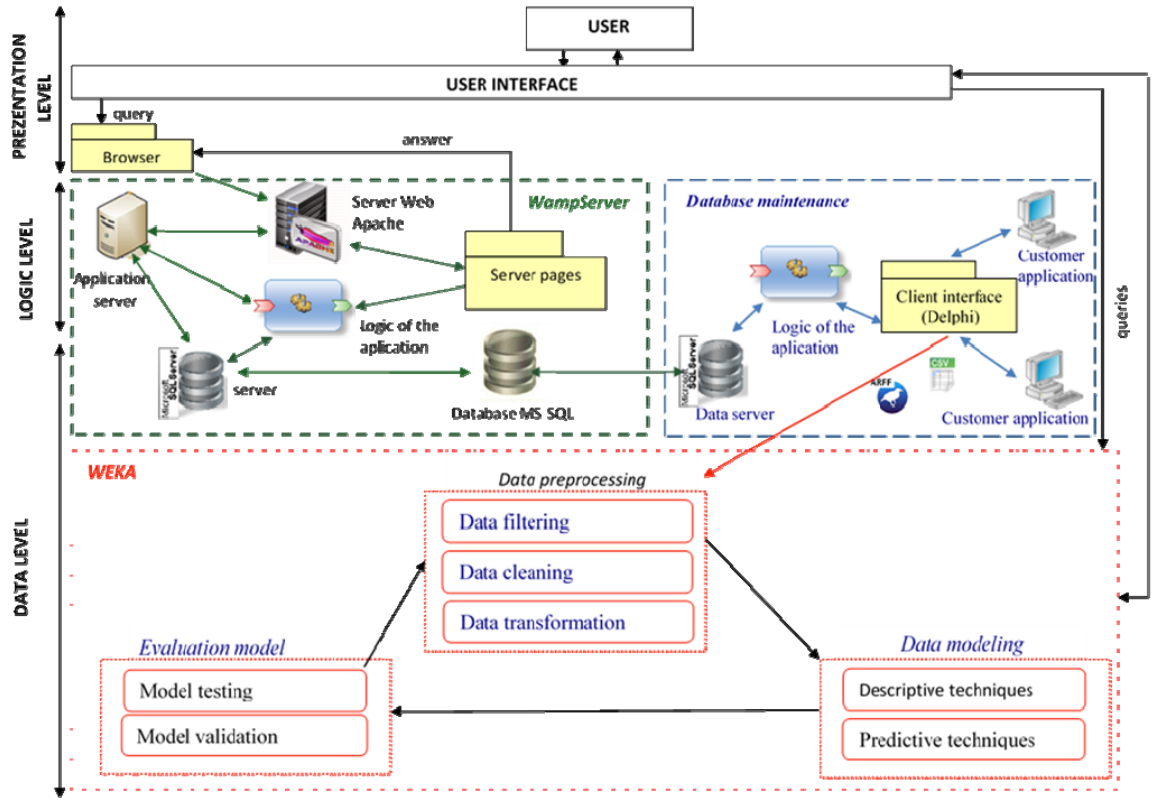

Figure 4. The DM Educational System Architecture.

The application logic is represented by the application code. It contains the data maintenance software, the data export software and the procedure that converts the data extracted from the database into a CSV and ARFF. The data conversion procedure was performed by using the functions and the functionalities presented in Microsoft SQL Server 2008 R2 and the functions developed in Microsoft SQL Server 2008 R2 and Embarcadero Delphi XE2 Enterprise.

Embarcadero Delphi XE2 Enterprise is designed for the developing of the applications based on client/server, the cloud applications, the GUI multi-level applications and of the web-oriented database that works with a huge array of servers databases and data sources (My SQL, Microsoft SQL Server, Oracle, DB2, Informix, Sybase ASE etc.).

The web server has to wait and respond to requests from customers. When an application appears, the web server relies on certain components hosted on the server, components which vary according to the type of server. There are used a large number of servers in the web applications. The most used are: Corel, Microsoft Internet Information Services (MIIS), iPlanet Web Server, Roxen Web Server, Zeus Web Server.

Embarcadero Delphi XE2 Enterprise is designed for developing the applications based on client/server, the cloud applications, the GUI multi-level applications and the web-oriented database that works with a variety of database servers and data sources (MySQL, Microsoft SQL Server, Oracle, DB2, Informix, Sybase ASE etc.).

We have used the 2.2.22 version of the Apache Web as a web server when using the application. Apache is developed by an open community of developers under the name Apache Software Foundation. The application is available for a wide variety of operating systems including Unix, FreeBSD, Linux, Solaris, Novell NetWare, Mac OS X, Microsoft Windows and OS/2.

The Apache server is a free software and an open source server. It is also a very complex web server, mainly due to the numerous facilities it offers: reliability (it is constantly tested and improved), it is fast, it provides a low consumption of hardware resources, multitasking, virtual hosting, enhanced security, running on different platforms, easy set up.

The Server Pages were made with PHP, the script version 5.5.10. PHP, an acronym which comes from English. The "Hypertext Preprocessor" is a widely-used scripting language, produced and distributed in Open Source in order to develop the dynamic web applications by integrating PHP code in HTML documents. Its syntax draws the languages C, C ++, Java and Perl and combines one of the most complex features. Unlike scripting languages such as JavaScript, PHP runs on the Web server, not on web browser. Accordingly, PHP can gain access to files, databases and other resources inaccessible to a JavaScript program.

The level data provides the data management and comprises three sub-layers: the data server, the database and the module for extracting the knowledge from the database. The data from this level are maintained independently from the application servers or from the application logic.

The data server manages the stored data, processes transactions and prepares the pieces of information for the decision-making processes in a compatible language with the manner of data storage (SQL). The data access is done by running an application by using the interface functions of the native database with ADO OLE/DB or ODBC driver.

The typically used data servers are: Oracle, DB2, Informix, Microsoft SQL Server, Ingres and My SQL. In making the application we chose Microsoft SQL Server.

The module for extracting the knowledge from the database allows the methods and data mining techniques on the data from the database to extract useful information.

There is now a huge number of tools that can perform data mining analysis. These can be divided into:

- Business tools: SPSS Clementine, SAS E-Miner, MATLAB, Oracle Data Mining, SQL Server, SAP data mining, etc.
- Open source tools: Weka, R, Orange, Rapid Miner, Miner Ant, etc.

During the applied E.D.M. model, we have integrated the Weka tool which is a software product of the University of Waikato, New Zealand. Weka, an acronym for the Waikato Environment for Knowledge Analysis (Waikato Environment Knowledge Analysis) is a collection of machine learning algorithms for data mining language written in Java. We have chosen this data mining instrument because it is an open-source application, under GNU general public license, it is independent and portable, being compatible with almost any platform (Fateh ME Shiri Ahmad, 2013) and has the ability to work with data located both on computers and servers.

By integrating Weka package, EDM model provides the extraction and the selection of the characteristics of the objects stored in the database, the descriptive and predictive modelling of the objects stored in the system, the discovery of the templates and rules.

6. CONCLUSIONS

The proposed model shows the manner in which the new technologies can be used in the higher education system in Romania to improve the efficiency of traditional processes. The model presented can be used in the higher education system to analyse the existing work, to identify gaps and to improve decision-making processes.

With this software system all the activities involved in the discovery of knowledge are kept together. The advantage of this approach is to have access to all functionalities of SQL Server, Analysis Services through a single software interface which allows: extracting relevant information for the user, the possibility of introducing new attributes, communicating with students, communicating with the user, data filtering, grouping data by different criteria, the interactive changing of the attributes parameters, the presented E.D.M. model is based on large databases that store data in education.

Architecture was achieved by integrating software open source data mining, using the latest technologies in web programming, database and interface design. Thus, the system can be implemented in higher education institutions, providing useful information to both university management and teachers and students.

REFERENCES

[1] Amaral, A., Neave, G., Musselin, C., Maassen, P., *European integration and the governance of higher education and research*, Springer, Dordrecht, 2010.
[2] Shahiria, A.M., Husain, W., Rashid, N.A., *Procedia Computer Science,* **72**, 414, 2015.
[3] Arora, R.K., Dharmendra, B., *International Journal of Computer Science and Mobile Computing*, **3**(1), 428, 2014.
[4] Aziz, A.A., Jusoh, J.A., Hassan, H., Rizhan, W.M., Idris, W., Md Zulkifli, A.P., Yusof, S.A.M., *Journal of Theoretical and Applied Information Technology*, **69**(1), 50, 2014.
[5] Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., Tsolakidis, A., *Procedia - Social and Behavioral Sciences*, 147, 390, 2014.

[6] Dekker, G., Pechenizkiy, M., Vleeshouwers, J., *Predicting Students Drop Out: A Case Study*, Proceedings of the International Conference on Educational Data Mining, 41, 2009.
[7] Delavari, N., Phon-Amnuaisuk, S., *Informatics in Education*, **7**(1), 31, 2008.
[8] Edelstein, H.A., *Introduction to Data Mining and Knowledge Discovery Third Edition*, Two Crows Corporation, 1999.
[9] Fateh, A., Shiri Ahmad Abadi, M.E., *International Journal of Computer Applications*, **63**(10), 14, 2013.
[10] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., *AI Magazine*, **17**(3), 37, 1996.
[11] Gornitzka, Å., Chapter The Lisbon process: A supranational policy perspective in *University dynamics and European integration*, Volume 19, p. 155, Springer, Dordrecht, 2007.
[12] Gorunesu, F., *Data Mining. Concepte, modele şi tehnici*, Editura Albastră, Cluj-Napoca, 2006.
[13] Han, J., Kamber, M., *Data Mining- Concepts and Techiques,* Morgan Kaufmann Publishers, San Francisco, 2001.
[14] Hassoun, M.H., *Fundamentals of Artificial Neural Networks*, The MIT Press, Bambridge, MA, 1995.
[15] Hunt, E.B., Marin, J., Stone, P.T., *Experiments in Induction*, Academic Press, NewYork, 1966
[16] I.Al-Twijri, M., Noamanb, A.Y., *Procedia Computer Science,* **65**, 836, 2015.
[17] Jalaliyoon, N., Therdoost, H., *Procedia - Social and Behavioral Sciences*, **46**, 5682, 2012.
[18] Marginson, S., *Educational Theory*, **56**(2), 205, 2006.
[19] Mohammed, I.T., Amin, Y.N., *Procedia Computer Science,* **65**, 836, 2015.
[20] Oprea, C., Zaharia, M., *Annals of the University of Oradea Fascicle of Management and Technological Engineering*, **X**(XX), 5.222, 2011.
[21] Oprea, C., Zaharia, M., Gogonea, M., *Proceedings of The 14th IBIMA Conference on Global Business Transformation through Innovation and Knowledge Management*, 180, 2010.
[22] Patrascu, A., *Ovidius University Annals-Economic Sciences*, **XVI**(1), 583, 2016.
[23] Romero, C., Ventura, C., *Expert Systems with Applications*, **33**(1), 135, 2007.
[24] Sahami, M., *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 335, 1996.
[25] Sacin, C., Agapito, J., Shafti, L., Ortigosa, A., *Proceedings of Educational Data Mining*, 190, 2009.
[26] Shahiria, A.M., Wahidah H., Rashida, N.A., *Procedia Computer Science,* **72**, 414, 2015.
[27] Stromquist, N.P., Monkman, K., *Globalization and education: Integration and contestation across cultures,* R&L Education, New York, 2014.
[28] Susnea, E., *Procedia - Social and Behavioral Sciences,* **76**, 795, 2013.
[29] Tănăsescu, A., *Economic Insights - Trends and Challenges*, **II**(LXV-1), 124, 2013
[30] Yadav, S.K., Saurabh, P., *International Journal of Computer Applications*, **41**(5), 1, 2012.