Osmanbegovic, Edin; Suljic, Mirza

Article

Data Mining Approach for Predicting Student Performance

Economic Review: Journal of Economics and Business

Provided in Cooperation with:
Faculty of Economics, University of Tuzla

Suggested Citation: Osmanbegovic, Edin; Suljic, Mirza (2012) : Data Mining Approach for Predicting Student Performance, Economic Review: Journal of Economics and Business, ISSN 1512-8962, University of Tuzla, Faculty of Economics, Tuzla, Vol. 10, Iss. 1, pp. 3-12

This Version is available at:
http://hdl.handle.net/10419/193806

DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE

Edin Osmanbegović *, Mirza Suljić **

ABSTRACT

Although data mining has been successfully implemented in the business world for some time now, its use in higher education is still relatively new, i.e. its use is intended for identification and extraction of new and potentially valuable knowledge from the data. Using data mining the aim was to develop a model which can derive the conclusion on students' academic success.

Different methods and techniques of data mining were compared during the prediction of students' success, applying the data collected from the surveys conducted during the summer semester at the University of Tuzla, the Faculty of Economics, academic year 2010-2011, among first year students and the data taken during the enrollment. The success was evaluated with the passing grade at the exam. The impact of students' socio-demographic variables, achieved results from high school and from the entrance exam, and attitudes towards studying which can have an affect on success, were all investigated. In future investigations, with identifying and evaulating variables associated with process of studying, and with the sample increase, it would be possible to produce a model which would stand as a foundation for the development of decision support system in higher education.

Keywords: data mining, classification, prediction, student success, higher education

JEL classification: L86

1. INTRODUCTION

For higher education institutions whose goal is to contribute to the improvement of quality of higher education, the success of creation of human capital is the subject of a continuous analysis. Therefore, the prediction of students' success is crucial for higher education institutions, because the quality of teaching process is the ability to meet students' needs. In this sense important data and information are gathered on a regular basis, and they are considered at the appropriate authorities, and standards in order to maintain the quality are set. The quality of higher education institutions implies providing the services, which most likely meet the needs of students, academic staff, and other participants in the education system. The participants in the educational process, by fulfilling their obligations through appropriate activities, create an enormous amount of data which needs to be collected and then integrated and utilized. By converting this data into knowledge, the gratification of all participants is attained: students, professors, administration, supporting administration, and social community.

All participants in the educational process could benefit by applying data mining on the data from the higher education system (Figure 1.1). Since data mining represents the computational data process from different perspectives, with the goal of extracting implicit and interesting samples (Witten and Frank, 2000), trends and information from the data, it can greatly help every participant in

* *University of Tuzla, Faculty of Economics, edin.osmanbegovic@untz.ba*

** *University of Tuzla, Faculty of Economics, mirza.suljic@untz.ba*

the educational process in order to improve the understanding of the teaching process, and it centers on discovering, detecting and explaining educational phenomenon’s (El-Halees, 2008).

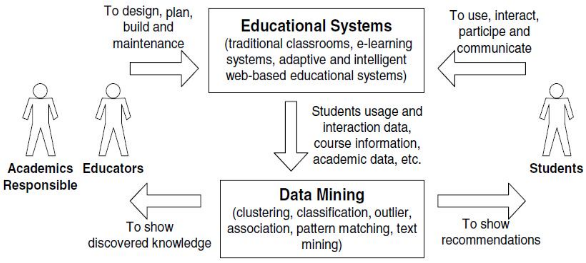

Figure 1.1. The cycle of applying data mining in educational systems

Source: Romero and Ventura, 2007, pp. 136

So with data mining techniques, the cycle is built in educational system which consists of forming hypotheses, testing and training, i.e. its utilization can be directed to the various acts of the educational process in accordance with specific needs (Romero and Ventura, 2007, pp. 136):

- of students,
- professors and
- administration and supporting administration.

Thus, application of data mining in educational systems can be directed to support the specific needs of each of the participants in the educational process. The student is required to recommend additional activities, teaching materials and tasks that would favour and improve his/her learning. Professors would have the feedback, possibilities to classify students into groups based on their need for guidance and monitoring, to find the most made mistakes, find the effective actions, etc. Administration and administrative staff will receive the parameters that will improve system performance (Romero and Ventura, 2007, pp. 136).

In recent years there has been an increased interest in using data mining for educational purposes. Data mining represents promising areas of researches in education, and it has specific requirements which other fields lack. A very comprehensive review of data mining in education from 1995 to 2005 is published in 2007 by Romero and Ventura. One of the educational problems that are solved with data mining is the prediction of students' academic performances, whose goal is to predict an unknown variable (outcome, grades or scores) that describes students. The estimation of students' performances includes monitoring and guiding students through the teaching process and assessment. Assessments, as the main procedure for the measurement of studying outcomes, indicate the level of students' performance, which is expressed qualitatively and quantitatively. Therefore, exams play an important role in any student’s lives, determining their future.

Minaei-Bidgolim, et al. (2003) was among the first authors who classified students by using genetic algorithms to predict their final grade. Using the regression methods, Kotsiantis and Pintelas (2005) predicted a student’s marks (pass and fail classes). Superby, Vandamme and Meskens (2006) predicted a student’s academic success (classified into low, medium, and high risk classes) using different data mining methods (decision trees and neural network). Al-Radaideh, Al-Shawakfa and Al-Najjar (2006) applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan. Romero et al. (2008) compared different methods of data mining in order to predict final assessment based on the data obtained from the system of e- learning. Zekić-Sušac, Frajman-Jakšić and Drvenkar (2009) created a model for predicting students' performance using neural networks and classification trees decision-making, and with the analysis of factors which influence students' success. Kumar and Vijayalakshmi

(2011) using the decision tree predicted the result of the final exam to help professors identify students who needed help, in order to improve their performance and pass the exam.

The success of studying at higher educational institutions in Bosnia and Herzegovina until now has only been investigated for the purpose of finding the average grades, length of study and similar indicators, while factors affecting student achievement results in a particular course have not been sufficiently investigated. In this paper different techniques of data mining suitable for classification have been compared: Bayesian classifier, neural networks and decision trees. Neural networks have in many areas shown success in solving problems of prediction, approximation, function, classification and pattern recognition. Their accuracy was compared with decision trees and with the Bayesian classifier. This work is based on the survey conducted on students of the Faculty of Economics, in Tuzla, academic year 2010-2011, in which, aside from the demographic data, the data about their past success and success in college have been collected. This analysis was conducted after the training and testing of the algorithms, making it possible to draw conclusions on possible predictors of students' success.

2. DATA DESCRIPTION

The data for the model were collected through a questionnaire survey conducted during the summer semester at the Faculty of Economics in Tuzla, academic year 2010-2011, among the first year students. After eliminating incomplete data, the sample comprised 257 students who were at the time of researches present at the practice classes. The model of students' success was created, where success as the output variable is measured with the success in the course "Business Informatics". As input to the model 12 variables are used, whose names and coding is shown in Table 2.1.

Table 2.1. Student related variables

Br.	Variable	Coding	Br.	Variable	Coding
1.	Gender (S)	A – male B – female	2.	Family (BCD)	Numeric value
3.	Distance (UAS)	Numeric value	4.	High School (VSS)	A – Grammar School B – High school for economics C – Rest
5.	GPA (PO)	Numeric value	6.	Entrance exam (URK)	Numeric value
7.	Scholarships (SS)	A – Not B – Sometimes C – Yes	8.	Time (VRI)	A – less then 1 hour B – from 1 to 2 hours C – from 2 to 3 hours D – from 3 to 4 hours E – from 4 to 5 hours
9.	Materials (MAT)	A – book, B – the notes of other students, C – notebook from the lectures, D – notes edited or made by student E – all that is available to student	10.	the Internet (INT)	A – Yes B – No
11.	Grade importance (VO)	A – Not important at all, B – not important C – Somewhat important, D – Important, E – Very important	12.	Earnings (MPD)	A – less then 500 KM B – from 500 to 1000 KM C – from 1000 to 1500 KM D – from 1500 to 2000 KM E – over 2000 KM

Distribution of the final students' grades in the course ''Business Informatics'' is shown in Figure 2.1.

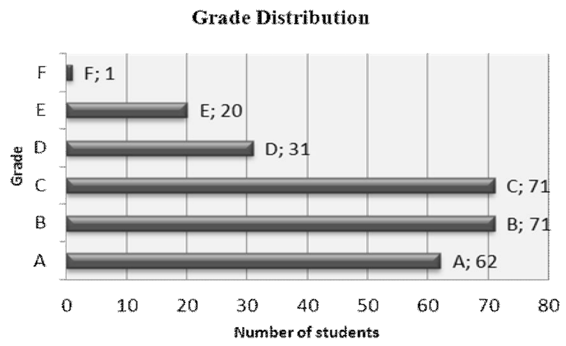

Figure 2.1. Distribution of grades in the course ''Business Informatics''

Output variable - students' assessments in the academic year can be grouped in several ways, 2 of which are:

- Through the six classes coded in the way: labels are the same as students' final grades, as shown in Table 2.2.
- Through two classes coded in this way: category A- failed, category B- passed, as shown in Table 2.3.

Table 2.2. Six class labels regarding students'final grade

Class	Grade	Student	Percentage
1	A	1	0,39%
2	B	20	7,78%
3	C	31	12,06%
4	D	71	27,63%
5	E	72	28,02%
6	F	62	24,12%

Table 2.3. Two class labels regarding as students'final grade

Class	Grade	Student	Percentage
1	A	62	24,12%
2	B	195	75,88%

It is evident that the prediction error rate will be much higher in the first case due to different distribution of grades through classes; hence the advantage is given to the second case of this study.

3. DATA MINING METHOD

Data mining is a computational method of processing data which is successfully applied in many areas that aim to obtain useful knowledge from the data (Klosgen and Zytkow, 2002). Data mining techniques are used to build a model according to which the unknown data will try to identify the new information. Regardless of origin, all data mining techniques show one common feature: automated discovery of new relationships and dependencies of attributes in the observed data. If the goal of the analysis is the categorization of data by class, then that is the new information on classes to which data belongs. In doing so, the algorithms are divided into two basic groups:

- unsupervised algorithms and
- supervised algorithms.

When the mining is ''unsupervised'' or ''undirected'', the output conditions are not explicitly represented in the data set: the task of **unsupervised algorithm** is to discover automatically inherent patterns in the data without the prior information about which class the data could belong, and it does not involve any supervision (Cios, Pedrycz, Swiniarski and Kurgan, 2007). Conversely, in unsupervised learning, no target variable to be learned is identified as such. Instead, the unsupervised learning algorithm searches for patterns and structure among all the variables. The goal of such model is to uncover data patterns in the set of input fields. Sometimes, the model produced by an unsupervised learning algorithm can be used for prediction tasks even though it was not designed for such tasks. A method of clustering and association rules belongs to this group.

Supervised algorithms are those which use data with in advance familiar class to which data belong for building models, and then on the basis of the constructed model predict the class to which unknown data will belong. A method of classification belongs to this group. Methods of data classification represent a process of learning a function that maps the data into one of several predefined classes. To every classification algorithm, that is based on inductive learning, input data set is given, that consists of vectors of attribute values and their corresponding class. The goal of a classification technique is to build a model which makes it possible to classify future data points based on a set of specific characteristics in an automated way. Such systems take a collection of cases as input, each belonging to one of a small number of classes and described by its values for a fixed set of attributes. As output they take a classifier that can accurately predict the class to which a new case belongs. The most common methods of classifications are: decision trees, induction rules or classification rules, probabilistic or Bayesian networks, neural networks and hybrid procedures.

There are many different classifiers in the literature and one cannot choose the best, because they differ mutually in many aspects such as: learning rate, amount of data for training, classification speed, robustness, etc. In this study we investigated the impact of three algorithms for intelligent data analysis: C4.5, Multilayer Perceptron and Naive Bayes (Wu and Kumar, 2009). Classification models are made by using these algorithms whose prediction aim is to predict the class (student's success) to which some new unlabeled sample will belong. The selected three classification techniques are used to discover the most suited way to predict student's success.

Naive Bayes algorithm (NB) is a simple method for classification based on the theory of probability, i.e. the Bayesian theorem (Witten and Frank, 2000). It is called naive because it simplifies problems relying on two important assumptions: it assumes that the prognostic attributes are conditionally independent with familiar classification, and it supposes that there are no hidden attributes that could affect the process of prediction. This classifier represents the promising approach to the probabilistic discovery of knowledge, and it provides a very efficient algorithm for data classification.

Multilayer Perceptron (MLP) algorithm is one of the most widely used and popular neural networks. The network consists of a set of sensory elements that make up the input layer, one or more hidden layers of processing elements, and the output layer of the processing elements (Witten and Frank, 2000). MLP is especially suitable for approximating a classification function (when we are not so much familiar with the relationship between input and output attributes) which sets the example determined by the vector attribute values into one or more classes.

The most commonly, and nowadays probably the most widely used decision tree algorithm is **C4.5**. Professor Ross Quinlan developed a decision tree algorithm known as C4.5 in 1993; it represents the result of research that traces back to the ID3 algorithm (which is also proposed by Ross Quinlan in 1986). C4.5 has additional features such as handling missing values, categorization of continuous attributes, pruning of decision trees, rule derivation, and others. Basic construction of C4.5 algorithms uses a method known as *divide and conquer* to construct a suitable tree from a training set S of cases (Wu and Kumar, 2009):

- If all the cases in S belong to the same class or S is small, the tree is a leaf labelled with the most frequent class in S.

- Otherwise, choose a test based on a single attribute with two or more outcomes. Make this test the root of the tree with one branch for each outcome of the test, partition S into corresponding subsets $S_1, S_2, \ldots$ according to the outcome for each case, and apply the same procedure recursively to each subset.

There are usually many tests that could be chosen in this last step. C4.5 uses two heuristic criteria to rank possible tests: information gain, which minimizes the total entropy of the subsets, and the default gain ratio that divides information gain by the information provided by the test outcomes (Wu and Kumar, 2009).

J48 algorithm is an implementation of C4.5 decision tree algorithm in Weka software tool. Flowchart of decision trees is presented by the tree structure. In every internal node the condition of some attribute is being examined, and every branch of the tree represents an outcome of the study. The branching of the tree ends with leaves that define a class to which examples belong. Decision tree algorithm is a popular procedure today because of its ease of implementation and in particular because of the possibility for the results to be graphically displayed.

To evaluate the robustness of the classifier, the usual methodology is to perform cross validation on the classifier. In this study, a 3-fold cross validation was used: we split data set randomly into 3 subsets of equal size. Two subsets were used for training, one subset for cross validating, and one for measuring the predictive accuracy of the final constructed network. This procedure was performed 3 times so that each subset was tested once. Test results were averaged over 3-fold cross-validation runs. Data splitting was done without sampling stratification. The Weka software toolkit can calculate all these performance metrics after running a specified k-fold cross-validation. The prediction accuracy of the models was compared.

4. EXPERIMENT RESULTS AND DISCUSSIONS

For the purposes of this study WEKA software package was used, that was developed at the University of Waikato in New Zealand. This package has been implemented in the software language Java and today stands out as probably the most competent and comprehensive package with algorithms of machinery learning in academic and nonprofit world (Machine Learning Group at University of Waikato, 2011).

In order to get a better insight into the importance of the input variables, it is customary to analyze the impact of input variables during students' prediction success, in which the impact of certain input variable of the model on the output variable has been analyzed. Tests were conducted using four tests for the assessment of input variables: Chi-square test, One R-test, Info Gain test and Gain Ratio test. The results of each test were monitored using the following metrics: Attribute (name of the attribute), Merit (measure of goodness), Merit dev (deviation, i.e. measure of goodness deviation), Rank (average position occupied by attribute), Rank and dev (deviation, deviation takes attribute's position). Different algorithms provide very different results, i.e. each of them accounts the relevance of attributes in a different way. The average value of all the algorithms is taken as the final result of attribute ranking, instead of selecting one algorithm and trusting it. The results obtained with these values are shown in Table 4.1.

Table 4.1. The results of all tests and their average rank

ATRIBUT	Chi-Squared	One R	Info Gain	Gain Ratio	AVG Rang
PO	1,3	1	1,3	1	1,15
URK	1,7	8	1,7	2	3,35
MAT	4,7	6	4,7	4,3	4,93
VRI	3,7	10,3	3,3	4	5,33
SS	7,7	5	7,7	6	6,6
VO	5,7	10,3	5,3	6	6,83
MPD	5,7	9,3	5,7	6,7	6,85
INT	7	7	7,3	6,7	7
VSS	8,7	4	9	9	7,68
S	9	5,7	9	9,3	8,25
UAS	11	5	11	11	9,5
BCD	12	6,3	12	12	10,58

In this aggregate table "Merit" columns are not applicable, because the algorithms use mutually incompatible metrics. The aim of this analysis is to determine the importance of each attribute individually. Table 4.1 shows that attribute PO (GPA) impacts output the most, and that it showed the best performances in all of the four tests. Then these attributes follow: URK (entrance exam), MAT (study material), VRI (average weekly hours devoted to studying). The following attributes had the smallest output impact: BCD (number of household members), UAS (distance of residence from the faculty) and S (sex).

We have carried out some experiments in order to evaluate the performance and usefulness of different classification algorithms for predicting students' success. The results of the experiments are summarized in Table 4.2, 4.3, 4.4 and 4.5.

Table 4.2. Predictive performance of the classifiers

EVALUATION CRITERIA	CLASSIFIERS		
	NB	MLP	J48
Timing to build model (in Sec)	0	4,13	0
Correctly classified instances	197	183	190
Incorrectly classified instances	60	74	67
Prediction accuracy	76,65	71,20	73,93

The performances of the three models were evaluated based on the three criteria: the prediction accuracy, learning time and error rate, which are illustrated in Figures 4.1, 4.2, and 4.3.

Table 4.3. Comparison of estimates

EVALUATION CRITERIA	CLASSIFIERS		
	NB	MLP	J48
Kappa statistic	0,3552	0,1958	0,1949
Mean absolute error (MAE)	0,2637	0,2856	0,3255
Root mean squared error (RMSE)	0,4204	0,4969	0,4431
Relative absolute error (RAE)	71,73 %	77,68 %	88.53 %
Root relative squared error (RRSE)	98,25 %	116,14 %	103.55 %

Table 4.4. Comparison of evaluation measures by class

CLASSIFIER	TP	FP	Precision	Recall	Class
NB	0,500	0,149	0,517	0,500	A
	0,851	0,500	0,843	0,851	B
MLP	0,371	0,179	0,397	0,371	A
	0,821	0,629	0,804	0,821	B
J48	0,290	0,118	0,439	0,290	A
	0,882	0,710	0,796	0,882	B

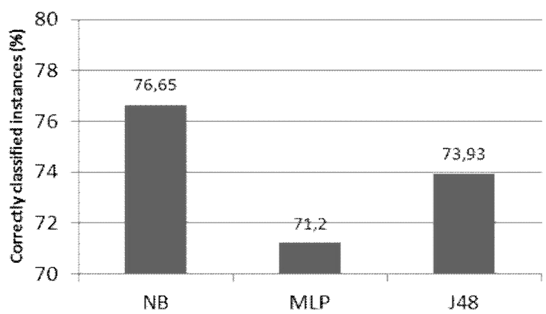

Figure 4.1. Prediction Accuracy

As shown in Figure 4.1. Naïve Bayes predicts better than other algorithms. Among the three classifiers used for experiment, the accuracy rate of Multilayer perceptron algorithm is the lowest.

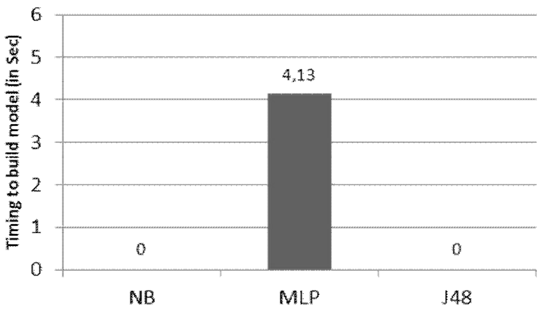

Figure 4.2. Learning time of tree classifiers

Figure 4.2 illustrates the learning time of the three schemes under consideration. Multilayer perceptron, the neural network classifier consumes more time to build the model. The Naïve Bayes and decision tree classifier learn more rapidly in the time to build a model for the given dataset. Figure 4.3 show the correctly classified instances vs. incorrectly classified instances.

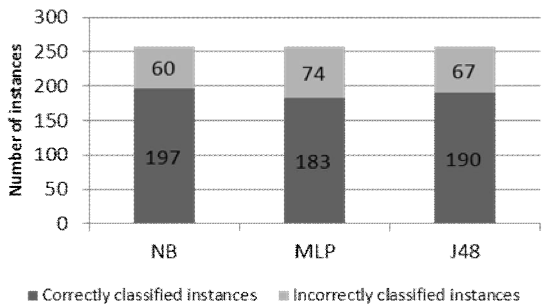

Figure 4.3. Error rate

The performance of the learning techniques is highly dependent on the nature of the training data. Confusion matrices are very useful for evaluating classifiers. The columns represent the predictions, and the rows represent the actual class. To evaluate the robustness of classifier, the usual methodology is to perform cross validation on the classifier.

Table 4.5. Confusion matrix

CLASSIFIERS	A	B	
NB	31	31	A
	29	166	B
MLP	23	39	A
	35	160	B
J48	18	44	A
	23	172	B

In general, cross validation has been proved to be statistically good enough in evaluating the performance of the classifier. Good results correspond to large numbers down the main diagonal and small, ideally zero, off-diagonal elements. From the confusion matrix given in Table 8, it is observed that MLP, NB and J48 produce relatively good results. The results strongly suggest that data mining can aid in the predict success in a course (either passed or failed). It is hoped that more interesting results will follow on further exploration of data.

On the other hand, in an educational problem it is also very important for the classification model obtained to be user friendly, so that teachers can make decisions to improve student learning. Nonetheless, some models are more interpretable than others (Romero, Ventura, Espejo, and Hervás, 2008):

- Decision trees are considered easily understood models because a reasoning process can be given for each conclusion. Knowledge models under this paradigm can be directly transformed into a set of IF-THEN rules that are one of the most popular forms of knowledge representation, due to their simplicity and comprehensibility which professor can easy understand and interpret (Figure 4.4).

- Statistical methods and neural networks are deemed to be less suitable for data mining purposes. Knowledge models obtained under these paradigms are usually considered to be black-box mechanisms, able to attain very good accuracy rates but very difficult for people to understand.

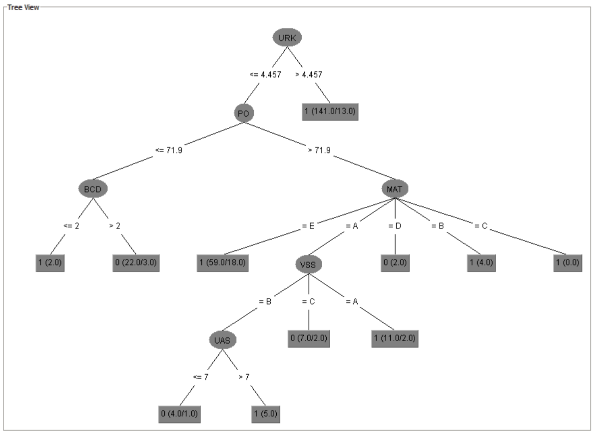

Figure 4.4. Obtained decision tree model

The model (see Figure 4.4.) is easy to be read and understood. This model can give professor interesting information about student and provides guidance to teacher to choose a suitable track, by analyzing experiences of students with similar academic achievements.

5. CONCLUSION

In this paper, three supervised data mining algorithms were applied on the preoperative assessment data to predict success in a course (either passed or failed) and the performance of the learning methods were evaluated based on their predictive accuracy, ease of learning and user friendly characteristics.

The results indicate that the Naïve Bayes classifier outperforms in prediction decision tree and neural network methods. It has also been indicated that a good classifier model has to be both accurate and comprehensible for professors. This study was based on traditional classroom environments, since the data mining techniques were applied after the data was collected.

However, it can be concluded that this methodology can be used to help students and teachers to improve student's performance; reduce failing ratio by taking appropriate steps at right time to improve the quality of learning. As learning is an active process, interactivity is a basic elements in this process that affects students' satisfaction and performance. It is important to answers these questions:

- How to obtain that predicting models are user friendly for professors or non-expert users?
- How to integrate data collection system of university and data mining tool?

For future work, the experiment can be extended with more distinctive attributes to get more accurate results, useful to improve the students learning outcomes. Also, experiments could be done using other data mining algorithms to get a broader approach, and more valuable and accurate outputs. Some different software may be utilized while at the same time various factors will be used.

REFERENCES

Books

1. Cios, K.J., Pedrycz W., Swiniarski, R.W. & Kurgan, L.A. (2007), Data Mining: A Knowledge Discovery Approach, Springer, New York.

2. Klosgen, W. & Zytkow, J. (2002), Handbook of data mining and knowledge discovery, Oxford University Press, New York.

3. Quinlan, J.R. (1993), C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco.

4. Witten, I.H. & Frank E. (2000), Data Mining – Practical Machine Learning Tools and Techniques, Second edition, Morgan Kaufmann, San Francisco.

5. Wu, X. & Kumar, V. (2009), The Top Ten Algoritms in Data Mining, Chapman and Hall, Boca Raton.

Articles

1. Kotsiantis, S.B. & Pintelas, P.E. (2005), Predicting Students Marks in Hellenic Open University, ICALT 2005. Fifth IEEE International Conference on Advanced Learning Technologies, Kaohsiung, pp. 664 – 668.

2. Kumar S. A. & Vijayalakshmi M. N. (2011), Efficiency of Decision Trees in Predicting Student's Academic Performance, First International Conference on Computer Science, Engineering and Applications, CS and IT 02, Dubai, pp. 335-343.

3. Minaei-Bidgoli, B., Punch, W.F. (2003), Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System, GECCO 2003 Genetic and Evolutionary Computation Conference, Chicago, pp. 2252-2263.

4. Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer G. & Punch WF (2003), Predicting student performance: an application of data mining metods with the educational web-based systems LON-CAPA, 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, pp.13-18.

5. Romero, C. & Ventura, S. (2007), Educational Data Mining: a Survey from 1995 to 2005, Expert Systems with Applications, Elsevier, pp. 135-146.

6. Romero C., Ventura S., Espejo, P.G. & Hervás C. (2008), Data mining algorithms to classify students, I International Conference on Educational Data Mining (EDM). Montreal, pp. 8-17.

7. Superby J.F., Vandamme J.P. & Meskens N. (2006), Determination of Factors Influencing the Achievement of the First-year University Students using Data Mining Methods, Proceedings of the 8th international conference on intelligent tutoring systems, Educational Data Mining Workshop, (ITS`06), Jhongali, pp. 37-44.

8. Zekić-Sušac M., Frajman-Jakšić A. & Drvenkar N. (2009), Neuron Networks and Trees of Decision-making for Prediction of Eficiency in Studies, Ekonomski vjesnik, Vol.No.2, pp. 314-327.

Website

1. Al-Radaideh, Q., Al-Shawakfa, E. & Al-Najjar, M. (2006), Mining Student Data Using Decision Trees, International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Available on:

   http://titania.addu.edu.ph/researches/DECISION%20SUPPORT/Mining%20Student%20Data%20Using%20Decision%20Trees.pdf [pristup 10.januar 2012.]

2. El-Halees, A. (2008), Mining students data to analyze learning behavior: a case study, Available on:

   http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf [accessed January 10 2012]

3. Machine Learning Group at University of Waikato, Available on: http://www.cs.waikato.ac.nz/ml/weka/ [accessed December 15, 2011]