BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 1

Sofia • 2013

Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2013-0006

# Predicting Student Performance by Using Data Mining Methods for Classification

*Dorina Kabakchieva*

*Sofia University "St. Kl. Ohridski", Sofia 1000*
*Email: dorina@fmi.uni-sofia.bg*

***Abstract:*** *Data mining methods are often implemented at advanced universities today for analyzing available data and extracting information and knowledge to support decision-making. This paper presents the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management.*

***Keywords****: Educational data mining, predicting student performance, data mining classification.*

## 1. Introduction

Universities today are operating in a very complex and highly competitive environment. The main challenge for modern universities is to deeply analyze their performance, to identify their uniqueness and to build a strategy for further development and future actions. University management should focus more on the profile of admitted students, getting aware of the different types and specific students' characteristics based on the received data. They should also consider if they have all the data needed to analyze the students at the entry point of the university or they need other data to help the managers support their decisions as how to organize the marketing campaign and approach the promising potential students.

This paper is focused on the implementation of data mining techniques and methods for acquiring new knowledge from data collected by universities. The main goal of the research is to reveal the high potential of data mining applications for university management.

The specific objective of the proposed research work is to find out if there are any patterns in the available data that could be useful for predicting students' performance at the university based on their personal and pre-university characteristics. The university management would like to know which features in the currently available data are the strongest predictors of university performance. They would also be interested in the data – is the collected data sufficient for making reliable predictions, is it necessary to make any changes in the data collection process and how to improve it, what other data to collect in order to increase the usability of the analysis results.

The main aim of this paper is to describe the methodology for the implementation of the initiated data mining project at the University of National and World Economy (UNWE), and to present the results of a study aimed at analyzing the performance of different data mining classification algorithms on the provided dataset in order to evaluate their potential usefulness for the fulfillment of the project goal and objectives. To analyze the data, we use well known data mining algorithms, including two rule learners, a decision tree classifier, two popular Bayes classifiers and a Nearest Neighbour classifier. The WEKA software is used for the study implementation since it is freely available to the public and is widely used for research purposes in the data mining field.

The paper is organized in five sections. The rationale for the conducted research work is presented in the Introduction. A review of the related research work is provided in Section 2, the research methodology is described in Section 3, the obtained results and the comparative analysis are given in Section 4. The paper concludes with a summary of the achievements and discussion of further work.

A short summary of the results is already presented (in poster session) and published in the Conference Proceedings of the 4th International Conference on Educational Data Mining (EDM'2011) [19], conducted on 6-8 July 2011 in Eindhoven, the Netherlands.

## 2. Review of the related research

The implementation of data mining methods and tools for analyzing data available at educational institutions, defined as Educational Data Mining (EDM) [15] is a relatively new stream in the data mining research. Extensive literature reviews of the EDM research field are provided by R o m e r o and V e n t u r a [15], covering the research efforts in the area between 1995 and 2005, and by B a k e r and Y a c e f [2], for the period after 2005. The problems that are most often attracting the attention of researchers and becoming the reasons for applying data mining at higher education institutions are focused mainly on retention of students, improving institutional effectiveness, enrollment management, targeted marketing, and alumni management.

The data mining project that is currently implemented at UNWE is focused on finding information in the existing data to support the university management in better knowing their students and performing more effective university marketing policy. The literature review reveals that these problems have been of interest for

various researchers during the last few years. L u a n discusses in [9] the potential applications of data mining in higher education and explains how data mining saves resources while maximizing efficiency in academics. Understanding student types and targeted marketing based on data mining models are the research topics of several papers [1, 9, 10, 11]. The implementation of predictive modeling for maximizing student recruitment and retention is presented in the study of N o e l-L e v i t z [13]. These problems are also discussed by D e L o n g et al. [5]. The development of enrollment prediction models based on student admissions data by applying different data mining methods is the research focus of N a n d e s h w a r and C h a u d h a r i [12]. D e k k e r et al. [6] focus on predicting students drop out.

The project specific objective is to classify university students according to the university performance results based on their personal and pre-university characteristics. Modeling student performance at various levels and comparing different data mining algorithms are discussed in many recently published research papers. K o v a č i ć i n [8] uses data mining techniques (feature selection and classification trees) to explore the socio-demographic variables (age, gender, ethnicity, education, work status, and disability) and study environment (course programme and course block) that may influence persistence or dropout of students, identifying the most important factors for student success and developing a profile of the typical successful and unsuccessful students. R a m a s w a m i and B h a s k a r a n [14] focus on developing predictive data mining model to identify the slow learners and study the influence of the dominant factors on their academic performance, using the popular CHAID decision tree algorithm. Y u et al. [18] explore student retention by using classification trees, Multivariate Adaptive Regression Splines (MARS), and neural networks. C o r t e z and S i l v a [4] attempt to predict student failure by applying and comparing four data mining algorithms – Decision Tree, Random Forest, Neural Network and Support Vector Machine. V a n d a m m e et al. [16] use decision trees, neural networks and linear discriminant analysis for the early identification of three categories of students: low, medium and high risk students. K o t s i a n t i s et al. [7] apply five classification algorithms (Decision Tree, Perceptron-based Learning, Bayesian Net, Instance-Based Learning and Rule-learning) to predict the performance of computer science students from distance learning.

## 3. The research methodology

The initiated data mining project at UNWE is implemented following the CRISP-DM (Cross-Industry Standard Process for Data Mining) model [3]. The CRISP-DM is chosen as a research approach because it is non-propriety, freely available, and application-neutral standard for data mining projects, and it is widely used by researchers in the field during the last ten years. It is a cyclic approach, including six main phases – Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. There are a number of internal feedback loops between the phases, resulting from the very complex non-linear nature of the data mining process and ensuring the achievement of consistent and reliable results.

The software tool that is used for the project implementation is the open source software WEKA, offering a wide range of classification methods for data mining [17].

During the ***Business Understanding Phase*** an extensive literature review is performed in order to study the existing problems at higher education institutions that have been solved by the application of data mining techniques and methods in previous research projects. Formal interviews with representatives of the University management at university, faculty and departmental levels, are also conducted, for finding out the specific problems at the University which have not yet been solved but are considered very important for the improvement of the University performance and for more effective and efficient management. Some insights are gathered from informal talks with lecturers, students and representatives of the administrative staff (IT experts and managers). Based on the outcomes of the performed research, the project goal and objectives, and the main research questions are formulated.

The main project goal is to reveal the high potential of data mining applications for university management, referring to the optimal usage of data mining methods and techniques to deeply analyze the collected historical data. The project specific objective, to classify university students according to the university performance results based on their pre-university characteristics, in data mining terms is considered a classification problem to be solved by using the available student data. This is a task for supervised learning because the classification models are constructed from data where the target (or response) variable is known.

During the ***Data Understanding Phase*** the application process for student enrollment at the University is studied, including the formal procedures and application documents, in order to identify the types of data collected from the university applicants and stored in the university databases in electronic format. The rules and procedures for collecting and storing data about the academic performance of the university students are also reviewed. Discussions with representatives of the administrative staff responsible for the university data collection, storage and maintenance are also carried out. University data is basically stored in two databases. All the data related to the university admission campaigns is stored in the University Admission database, including personal data of university applicants (names, addresses, secondary education scores, selected admission exams, etc.), data about the organization and performance of the admission exams, scores achieved by the applicants at the admission exams, data related to the final classification of applicants and student admission, etc. All the data concerning student performance at the university is stored in the University Students Performance database, including student personal and administrative data, the grades achieved at the exams on the different subjects, etc.

During the ***Data Preprocessing Phase***, student data from the two databases is extracted and organized in a new flat file. The preliminary research sample is provided by the university technical staff responsible for the data collection and maintenance, and includes data about 10330 students, described by 20 parameters, including gender, birth year, birth place, living place and country, type of previous

education, profile and place of previous education, total score from previous education, university admittance year, admittance exam and achieved score, university specialty/direction, current semester, total university score, etc. The provided data is subjected to many transformations. Some of the parameters are removed, e.g., the birth place and the place of living fields containing data that is of no interest to the research, the country field containing only one value – Bulgaria, because the data concerns only Bulgarian students, the type of previous education field which has only one value as well, because it concerns only students with secondary education. Some of the variables containing important data for the research are text fields where free text is being entered at the data collection stage. Therefore, these variables are processed and turned into nominal variables with a limited number of distinct values. One such parameter is the profile of the secondary education which is turned into a nominal variable with 15 distinct values (e.g., language, maths, natural sciences, economics, technical, sports, arts, etc.). The place of secondary education field is also preprocessed and transformed to a nominal variable with 19 distinct values, by leaving unchanged the values equal to the capital city and the 12 biggest cities in Bulgaria, and replacing the other places with the corresponding 7 geographic regions – north-east, north-central, north-west, south-east, south-central, south-west, and the capital city region. One new numeric variable is added – the student age at enrollment, by subtracting the values contained in the admission year and birth year fields. Another important operation during the preprocessing phase is also the transformation of some variables from numeric to nominal (e.g., age, admission year, current semester) because they are much more informative when interpreted as nominal values. The data is also being studied for missing values, which are very few and could not affect the results, and for obvious mistakes, which are corrected.

Essentially, the challenge in the presented data mining project is to predict the student university performance based on the collection of attributes providing information about the student pre-university characteristics. The selected target variable in this case, or the concept to be learned by data mining algorithm, is the "student class". A categorical target variable is constructed based on the original numeric parameter university average score. It has five distinct values (categories) – "excellent", "very good", "good", "average" and "bad". The five categories (classes) of the target (class) variable are determined from the total university score achieved by the students. A six-level scale is used in the Bulgarian educational system for evaluation of student performance at schools and universities. "Excellent" students are considered those who have a total university score in the range between 5.50 and 6.00, "very good" – in the range between 4.50 and 5.49, "good" – in the range between 3.50 and 4.49, "average" – in the range between 3.00 and 3.49, and "bad" – in the range below 3.00.

The final dataset used for the project implementation contains 10330 instances (539 in the "excellent" category, 4336 in the "very good" category, 4543 in the "good" category, 347 in the "average" category, and 564 in the "bad" category), each described with 14 attributes (1 output and 13 input variables), nominal and numeric. The attributes related to the student personal data include gender and age.

The attributes referring to the students' pre-university characteristics include place and profile of the secondary school, the final secondary education score, the successful admission exam, the score achieved at that exam, and the total admission score. The attributes describing some university features include the admission year, the student specialty or direction, the current semester, and the average score achieved during the first year of university studies (the class variable). The study is limited to student data for three university admission campaigns (for the time period between 2007 and 2009). The sample contains data about equal percentage of male and female students, with different secondary education background, finishing secondary schools in different Bulgarian towns and villages. They have been admitted with 9 different exams and study at different university faculties.

During the ***Modeling Phase***, the methods for building a model that would classify the students into the five classes (categories), depending on their university performance and based on the student pre-university data, are considered and selected. Several different classification algorithms are applied during the performed research work, selected because they have potential to yield good results. Popular WEKA classifiers (with their default settings unless specified otherwise) are used in the experimental study, including a common decision tree algorithm C4.5 (J48), two Bayesian classifiers (NaiveBayes and BayesNet), a Nearest Neighbour algorithm (IBk) and two rule learners (OneR and JRip). The achieved research results are presented in the next paper section.

## 4. The achieved results

The study main objective is to find out if it is possible to predict the class (output) variable using the explanatory (input) variables which are retained in the model. Several different algorithms are applied for building the classification model, each of them using different classification techniques. The WEKA Explorer application is used at this stage. Each classifier is applied for two testing options – cross validation (using 10 folds and applying the algorithm 10 times – each time 9 of the folds are used for training and 1 fold is used for testing) and percentage split (2/3 of the dataset used for training and 1/3 – for testing).

### 4.1. Decision tree classifier

Decision trees are powerful and popular tools for classification. A decision tree is a tree-like structure, which starts from root attributes, and ends with leaf nodes. Generally, a decision tree has several branches consisting of different attributes, the leaf node on each branch representing a class or a kind of class distribution. Decision tree algorithms describe the relationship among attributes, and the relative importance of attributes. The advantages of decision trees are that they represent rules which could easily be understood and interpreted by users, do not require complex data preparation, and perform well for numerical and categorical variables.

The WEKA J48 classification filter is applied on the dataset during the experimental study. It is based on the C4.5 decision tree algorithm, building decision trees from a set of training data using the concept of information entropy.

The J48 classifier classifies correctly about 2/3 of the instances (65.94 % for the 10-fold cross-validation testing and 66.59 % for the percentage split testing), produces a classification tree with a size of 1173 nodes and 1080 leaves. The attribute Number of Failures appears at the first level of the tree, the Admission Score and Current Semester attributes appear at the second and third levels of the tree, the attributes University Specialty/Direction and Gender – at the third level of the tree, which means that these attributes influence most the classification of the instances into the five classes.

Table 1. Results for the decision tree algorithm (J48)

| Class | J48 – 10-fold Cross validation | | J48 – Percentage split | |
|---|---|---|---|---|
| | TP Rate | Precision | TP Rate | Precision |
| Bad | 0.83 | 0.851 | 0.84 | 0.892 |
| Average | 0.081 | 0.384 | 0.096 | 0.344 |
| Good | 0.729 | 0.665 | 0.742 | 0.667 |
| Very Good | 0.69 | 0.639 | 0.687 | 0.646 |
| Excellent | 0.015 | 0.211 | 0.032 | 0.429 |
| Weighted Average | 0.659 | 0.631 | 0.666 | 0.648 |

The results for the detailed accuracy by class, including the True Positive (TP) rate (the proportion of examples which were classified as class *x*, among all examples which truly have class *x*) and the Precision (the proportion of the examples which truly have class *x* among all those which were classified as class *x*), are presented in Table 1.

The results reveal that the True Positive Rate is high for three of the classes – Bad (83-84 %), Good (73-74 %) and Very Good (69 %), while it is very low for the other two classes – Average (8-10 %) and Excellent (2-3 %). The Precision is very high for the Bad class (85-89 %), high for the Good (67 %) and Very Good (64-65 %) classes, and low for the Average (34-38 %)) and Excellent (21-43 %) classes. The achieved results are slightly better for the percentage split testing option.

### 4.2. Bayesian classifiers

Bayesian classifiers are statistical classifiers that predict class membership by probabilities, such as the probability that a given sample belongs to a particular class. Several Bayes' algorithms have been developed, among which Bayesian networks and naive Bayes are the two fundamental methods. Naive Bayes algorithms assume that the effect that an attribute plays on a given class is independent of the values of other attributes. However, in practice, dependencies often exist among attributes; hence Bayesian networks are graphical models, which can describe joint conditional probability distributions. Bayesian classifiers are popular classification algorithms due to their simplicity, computational efficiency and very good performance for real-world problems. Another important advantage is also that the Bayesian models are fast to train and to evaluate, and have a high accuracy in many domains.

The two WEKA classification filters applied on the dataset are the NaiveBayes and the BayesNet. Both of them are tested for 10-fold cross validation and percentage split options. The achieved results are presented in Table 2.

Table 2. Results for the Bayesian Classifiers

| Class | NaiveBayes | | | | BayesNet | | | |
|---|---|---|---|---|---|---|---|---|
| | 10-fold Cross validation | | Percentage split | | 10-fold Cross validation | | Percentage split | |
| | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision |
| Bad | 0.821 | 0.791 | 0.835 | 0.804 | 0.817 | 0.813 | 0.835 | 0.819 |
| Average | 0.352 | 0.209 | 0.348 | 0.183 | 0.38 | 0.237 | 0.417 | 0.222 |
| Good | 0.521 | 0.644 | 0.545 | 0.649 | 0.598 | 0.626 | 0.597 | 0.633 |
| Very Good | 0.681 | 0.576 | 0.679 | 0.588 | 0.616 | 0.599 | 0.613 | 0.601 |
| Excellent | 0.184 | 0.277 | 0.14 | 0.268 | 0.237 | 0.312 | 0.199 | 0.264 |
| Weighted Average | 0.581 | 0.59 | 0.59 | 0.597 | 0.591 | 0.596 | 0.591 | 0.598 |

The overall accuracy of the Bayesian classifies is about (but below) 60 % which is not very high, and it is worse compared to the performance of the decision tree classifier (66-67 %). The detailed accuracy results for the Bayesian classifiers reveal that the True Positive Rate is very high for the Bad class (82-84 %), not so high for the Very Good (61-68 %) and Good (52-60 %) classes, low for the Average class (35-42 %), and very low for the Excellent class (14-24 %). The Precision is high for the Bad class (79-81 %), not so high for the Good (63-65 %) and Very Good (58-60 %) classes, and low for the Average (18-24 %) and Excellent (26-31 %) classes.

The Naïve Bayes algorithm classifies the instances taking into account the independent effect of each attribute to the classification, and the final accuracy is determined based on the results achieved for all the attributes. The BayesNet classifier produces a simple graph, including all input attributes at the first level.

### 4.3. The k-Nearest Neighbour Classifier

The k-Nearest Neighbor algorithm (k-NN) is a method for classifying objects based on closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its $k$ nearest neighbors ($k$ is a positive integer, typically small). The best choice of $k$ depends upon the data; generally, larger values of $k$ reduce the effect of noise on the classification, but make boundaries between classes less distinct. The accuracy of the k-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

The WEKA IBk classification filter is applied to the dataset, which is a k-NN classifier. The algorithm is executed for two values of the parameter $k$ (100 and 250), and for the two testing options – 10-fold cross validation and percentage split. The results are presented in Table 3.

Table 3. Results for the k-NN Classifier

| Class | k-NN Classifier | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *k*=100 | | | | *k*=250 | | | |
| | 10-fold Cross validation | | Percentage split | | 10-fold Cross validation | | Percentage split | |
| | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision |
| Bad | 0.358 | 0.944 | 0.335 | 0.972 | 0.154 | 1 | 0.078 | 1 |
| Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Good | 0.662 | 0.614 | 0.69 | 0.617 | 0.626 | 0.602 | 0.651 | 0.598 |
| Very Good | 0.712 | 0.592 | 0.705 | 0.6 | 0.733 | 0.576 | 0.727 | 0.586 |
| Excellent | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Weighted Average | 0.609 | 0.57 | 0.616 | 0.578 | 0.592 | 0.561 | 0.593 | 0.565 |

The k-NN classifier accuracy is about 60 % and varies in accordance with the selected value for *k*. The results are slightly better for *k* =100 if compared to *k* =250. This classifier works with higher accuracies for the Very Good (71-73 %) and Good (63-69 %) classes, with low accuracy for the Bad (8-36 %) class, and performs very badly for the Average (0 %) and Excellent (0 %) classes. The Precision is excellent for the Bad class (94-100 %), but not so high for the Very Good (58-60 %) and Good (60-62 %) classes.

## 4.4. Rule learners

Two algorithms for generating classification rules are considered. The OneR classifier generates a one-level decision tree expressed in the form of a set of rules that all test one particular attribute. It is a simple, cheap method that often produces good rules with high accuracy for characterizing the structure in data. This classifier is often used as a baseline for the comparison between the other classification models, and as an indicator of the predictive power of particular attributes. The JRip classifier implements the RIPPER (Repeated Incremental Pruning to Produce Error Reduction) algorithm. Classes are examined in increasing size and an initial set of rules for the class is generated using incremental reduced-error pruning. The results are presented in Table 4.

Table 4. Results for the rule learners

| Class | OneR | | | | JRip | | | |
|---|---|---|---|---|---|---|---|---|
| | 10-fold Cross validation | | Percentage split | | 10-fold Cross validation | | Percentage split | |
| | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision | TP Rate | Precision |
| Bad | 0 | 0 | 0 | 0 | 0.823 | 0.845 | 0.738 | 0.776 |
| Average | 0 | 0 | 0 | 0 | 0.043 | 0.313 | 0 | 0 |
| Good | 0.688 | 0.545 | 0.689 | 0.542 | 0.731 | 0.618 | 0.744 | 0.615 |
| Very Good | 0.584 | 0.555 | 0.572 | 0.553 | 0.625 | 0.634 | 0.614 | 0.636 |
| Excellent | 0.006 | 0.15 | 0.032 | 0.214 | 0.082 | 0.506 | 0.081 | 0.375 |
| Weighted Average | 0.548 | 0.481 | 0.543 | 0.479 | 0.634 | 0.621 | 0.63 | 0.601 |

The achieved results show that, as expected, the JRip rule learner performs better than the OneR rule learner. The overall accuracy of the JRip classifier is

about 63 %, and for the OneR classifier it is about 54-55 %. Both rule learners perform not so bad for the Good and Very Good classes, the JRip classifier showing slightly better results than the OneR classifier. Both are equally bad for the Excellent class. However, the two rule learners perform differently for the Bad and Average classes. The OneR classifier is absolutely unable to predict the Bad and Average classes, while the JRip classifier performs very well for the Bad class but badly for the Average class.

The OneR learner uses the minimum-error attribute for prediction and in this case this is the Admission Score. The JRip learner produces 25 classification rules, most of them containing the attributes Number of Failures, Admission Score, Admission Exam Score, Current Semester, University Specialty/Direction, and Secondary Education Score. These are the attributes that influence most the classification of the instances into the five classes.

### 4.5. Performance comparison between the applied classifiers

The results for the performance of the selected classification algorithms (TP rate, percentage split test option) are summarized and presented on Fig. 1.

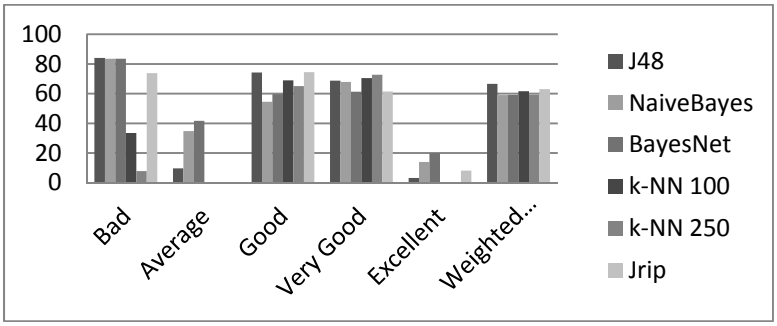


Fig. 1. Classification algorithms performance comparison

The achieved results reveal that the decision tree classifier (J48) performs best (with the highest overall accuracy), followed by the rule learner (JRip) and the k-NN classifier. The Bayes classifiers are less accurate than the others. However, all tested classifiers are performing with an overall accuracy below 70 % which means that the error rate is high and the predictions are not very reliable.

As far as the detailed accuracy for the different classes is concerned, it is visible that the predictions are worst for the Excellent class, and quite bad for the Average class, the k-NN classifier being absolutely unable to predict them. The highest accuracy is achieved for the Bad class, except for the k-NN classifier that is performing badly. The predictions for the Good and Very Good classes are more precise than for the other classes, and all classifiers perform with accuracies around 60-75 %. The decision tree classifier (J48) and the rule learner (JRip) are most reliable because they perform with the highest accuracy for all classes, except for the Excellent class. The k-NN classifier is not able to predict the classes which are less represented in the dataset. The Bayes classifiers are less accurate than the others.

## 5. Conclusions

The results achieved by applying selected data mining algorithms for classification on the university sample data reveal that the prediction rates are not remarkable (vary between 52-67 %). Moreover, the classifiers perform differently for the five classes. The data attributes related to the students' University Admission Score and Number of Failures at the first-year university exams are among the factors influencing most the classification process.

The results from the performed study are actually the initial steps in the realization of an applied data mining project at UNWE. The conclusions made from the conducted research will be used for defining the further steps and directions for the university data mining project implementation, including possible transformations of the dataset, tuning the classification algorithms' parameters, etc., in order to achieve more accurate results and to extract more important knowledge from the available data. Recommendations will also be provided to the university management, concerning the sufficiency and availability of university data, and related to the improvement of the data collection process.

## References


1. Antons, C., E. Maltz. Expanding the Role of Institutional Research at Small Private Universities: A Case Study in Enrollment Management Using Data Mining. – New Directions for Institutional Research, Vol. **131**, 2006, 69-81.
2. Baker, R., K. Yacef. The State of Educational Data Mining in 2009: A Review and Future Visions. – Journal of Educational Data Mining, Vol. **1**, October 2009, Issue 1, 3-17.
3. Chapman, P., et al. CRISP-DM 1.0: Step-by-Step Data Mining Guide 2000. SPSS Inc. CRISPWP-0800, 2000.
   **http://www.spss.ch/upload/1107356429_CrispDM1.0.pdf**
4. Cortez, P., A. Silva. Using Data Mining to Predict Secondary School Student Performance. EUROSIS. A. Brito and J. Teixeira, Eds. 2008, 5-12.
5. DeLong, C., P. Radclie, L. Gorny. Recruiting for Retention: Using Data Mining and Machine Learning to Leverage the Admissions Process for Improved Freshman Retention. – In: Proc. of the Nat. Symposium on Student Retention, 2007.
6. Dekker, G., M. Pechenizkiy, J. Vleeshouwers. Predicting Students Drop Out: A Case Study. – In: Proceedings of 2nd International Conference on Educational Data Mining (EDM'09), 1-3 July 2009, Cordoba, Spain, 41-50.
7. Kotsiantis, S., C. Pierrakeas, P. Pintelas. Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques. – Applied Artificial Intelligence, Vol. **18**, 2004, No 5, 411-426.
8. Kovačić, Z. Early Prediction of Student Success: Mining Students Enrolment Data. – In: Proceedings of Informing Science & IT Education Conference (InSITE'2010), 2010, 647-665.
9. Luan, J. Data Mining and Its Applications in Higher Education. – New Directions for Institutional Research, Special Issue Titled Knowledge Management: Building a Competitive Advantage in Higher Education, Vol. **2002**, 2002, Issue 113, 17-36.
10. Luan, J. Data Mining Applications in Higher Education. SPSS Executive Report. SPSS Inc., 2004.
    **http://www.spss.ch/upload/1122641492_Data%20mining%20applications%20in%20higher%20education.pdf**

11. M a, Y., B. L i u, C. K. W o n g, P. S. Y u, S. M. L e e. Targeting the Right Students Using Data Mining. – In: Proceedings of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, 2000, 457-464.
12. N a n d e s h w a r, A., S. C h a u d h a r i. Enrollment Prediction Models Using Data Mining, 2009. **http://nandeshwar.info/wp-content/uploads/2008/11/DMWVU_Project.pdf**
13. N o e l-L e v i t z. White Paper. Qualifying Enrollment Success: Maximizing Student Recruitment and Retention Through Predictive Modeling. Noel-Levitz, Inc., 2008. **https://www.noellevitz.com/documents/shared/Papers_and_Research/2008/QualifyingEnrollmentSuccess08.pdf**
14. R a m a s w a m i, M., R. B h a s k a r a n. A CHAID Based Performance Prediction Model in Educational Data Mining. – IJCSI International Journal of Computer Science Issues, Vol. **7**, January 2010, Issue 1, No 1, 10-18.
15. R o m e r o, C., S. V e n t u r a. Educational Data Mining: A Survey from 1995 to 2005. – Expert Systems with Applications, Vol. **33**, 2007, 135-146.
16. V a n d a m m e, J., N. M e s k e n s, J. S u p e r b y. Predicting Academic Performance by Data Mining Methods. – Education Economics, Vol. **15**, 2007, No 4, 405-419.
17. W i t t e n, I., E. F r a n k. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers, Elsevier Inc., 2005.
18. Y u, C., S. D i G a n g i, A. J a n n a s c h-P e n n e l l, C. K a p r o l e t. A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year. – Journal of Data Science, Vol. **8**, 2010, 307-325.
19. K a b a k c h i e v a, D., K. S t e f a n o v a, V. K i s i m o v. Analyzing University Data for Determining Student Profiles and Predicting Performance. – In: Proceedings of 4th International Conference on Educational Data Mining (EDM'2011), 6-8 July 2011, Eindhoven, The Netherlands, 347-348.