USING DATA MINING TECHNIQUES TO DISCOVER THE CORRELATION BETWEEN HIGH SCHOOLS FINISHED AND HIGHER EDUCATION PROGRAMMES CHOSEN BY STUDENTS

Dragana Knežević[1], MSc; Marija Blagojević[2], PhD; Ivana Marinković[1], MA
[1]Business and Technical College of Applied Sciences, Užice, SERBIA;dknezevic28@gmail.com
[2]University of Kragujevac, Faculty of Technical Sciences, Čačak, SERBIA; marija.blagojevic@ftn.kg.ac.rs
[1]Business and Technical College of Applied Sciences, Užice, SERBIA; ivana.marinkovic@vpts.edu.rs

Abstract: Although used as a basis for various analyses and research, large databases as such are insufficient. Accumulated data should be further analysed using appropriate techniques. Databases kept by student service departments are good examples of this. This paper explains the correlation between the high schools finished by students of the Business and Technical College of Applied Sciences (BTC) and the study programmes they have enrolled on. Real data have been used in this research, without revealing students' names and personal information in order to ensure the privacy of each individual. The aim is to discover whether students enrolled at higher education institutions (BTC in this case) choose the same fields of science as in high schools or study programmes in entirely different fields. Furthermore, a subsection provides information on students' high school academic achievement, as well as a comparative overview of study programmes regarding this issue. Another section provides a graphical overview of the results obtained using the decision tree technique. The results may be used for planning the promotion of specific study programmes to potential students. Any higher education institution can benefit from such analyses.

Key words: data mining, decision tree, students, high school, academic achievement.

1. INTRODUCTION

The existence of large databases as such is not sufficient even though they contain all the data relevant for specific research. The aim is to extract only 'interesting' information out of an enormous set of data. This 'interesting' information is in fact some non-trivial, non-concealed, but still unknown, and potentially useful information. This is why large databases are generally considered data-rich, but poor in information with real life applications unless the data are processed, most often using a data mining technique. Numerous examples from practice show how important data mining is in all spheres of life. Significant results in different fields have been obtained owing to data mining. An important characteristic of data mining is that it gives answers not only to the questions raised prior to the very beginning of research, but also to those that have not even been raised but have imposed themselves as laws governing the given situation, that we had not even been aware of. Unlike traditional statistical analyses based on testing previously suggested hypotheses, data mining relies on research using software as a tool. Software can be used to determine the connection and interdependence between numerous factors, thus providing answers that can help to solve numerous problems. In this way, for example, the degree to which a student's high school academic achievement and entrance exam success are interrelated can be determined [1]; also, proper teaching methods can be defined based on student behavior, especially when it comes to problematic students [2, 3]. Furthermore, with the help of proper techniques, and based on students' grades in certain high school subjects, it becomes possible to predict which study programme would be the right choice for them, and which learning techniques would be the most efficient for certain subjects [4, 5], etc. Using the decision tree technique and answers to previously raised questions, this paper graphically presents the results obtained using data processing software. The paper is aimed at discovering whether students enrolled in higher education institutions (BTC in this case) choose the same fields of science as in high schools or study programmes in entirely different fields.

2. RESERCH METHODOLOGY

The subject of this paper is discovering enrollment patterns in the Business and Technical College of Applied Sciences in Užice (BTC). The paper focuses on high schools finished by students enrolled on specific study programmes offered by the BTC. The results are explained in the following section.

The research sample involved 1029 undergraduate students enrolled in the first year at the BTC during the last three academic years, i.e. 2014/2015, 2015/2016 and 2016/2017. Please note that the data presented in the paper are real life data, extracted from the database kept by the Student Service Department of the BTC, but in order to protect the privacy of students, their real names and personal information are not revealed, and they are referred to as 'Student 1', 'Student 2', 'Student 3', etc. As all the records are kept within the database created for the needs of the Student Service Department itself, first of all the data had to be converted from the existing form into the format convenient for processing by data mining software. After the conversion, the data were reviewed; format and spelling inconsistencies removed, and checks for duplicate entries done. Generally speaking, before the very beginning of data processing using data mining techniques, it is good to remove data or entire data groups which are not necessary, and will not be used subsequently during the process. For example, from the data set used for the purpose of this research, the columns entitled 'Student's current residence address' and 'Student's phone number' have been removed as they contain irrelevant data subject to changes that students usually do not report to the Student Service Department, and as such, provide unreliable information about particular students. As soon as the data were prepared, their analysis using the appropriate software started. The data used in this paper were processed using the Cytoscape software [6], which provides graphical presentation of the given data.

3. RESEARCH RESULTS AND DISCUSSION

Only two out of many questions regarding the enrolled students have been chosen as the subject of this paper. The first one is: *After which high schools do students enroll on particular study programmes offered by the BTC?* The second one is: *What was students' academic achievement at the end of high school?* The figures in the following subsections show diagrams illustrating the answers to the questions. Each blue field ('house') represents one student.

3.1. After which high schools do students enroll on particular study programmes offered by the BTC?

As the BTC provides students with the education in two different fields – social science and humanities, and technical science and technology, one of the key points is the high school BTC students finished. In addition to whether they have finished a three-year or four-year high school, it is important to indicate the type of school in order to determine whether they have continued education in the same field of science or chosen a study programme in a completely different field. Such information is shown in the following diagrams (Figures 1-7).
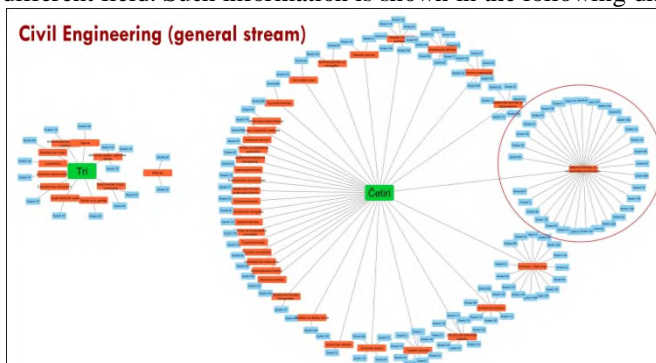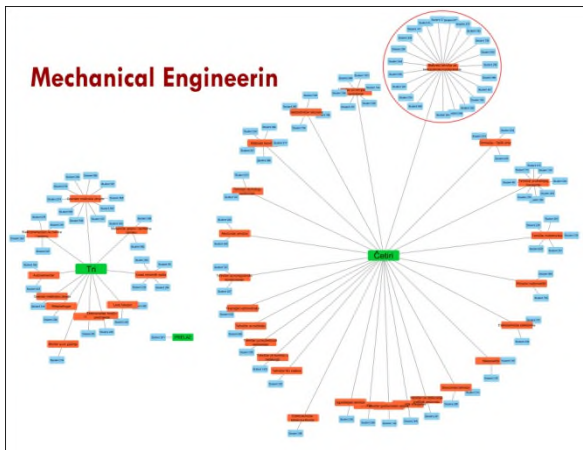

Figure 1: High schools finished by students enrolled on the Civil Engineering study programme
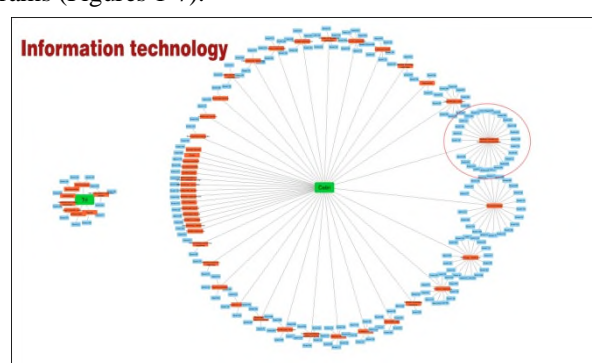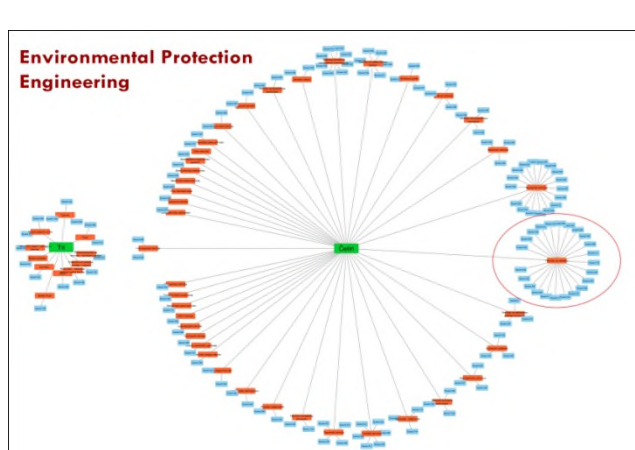
Figure 2: High schools finished by students enrolledon the Information Technology study programme

From Figure 1, it can be seen that most of students enrolled on the Civil Engineering study programme have finished four-year high schools, in particular those educating mechanical engineering technicians for computer-aided design; then grammar schools, and schools educating civil engineering technicians for building construction, architectural technicians, road traffic technicians, etc.

From Figure 2, it can be seen that most of students enrolled on the Information Technology study programme have also finished four-year high schools, in particular those educating mechanical engineering technicians for computer-aided design; then high schools of economics, grammar schools, schools educating computer technicians, etc.
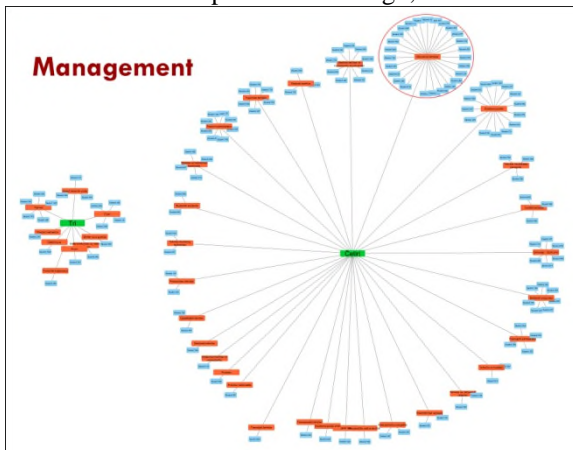
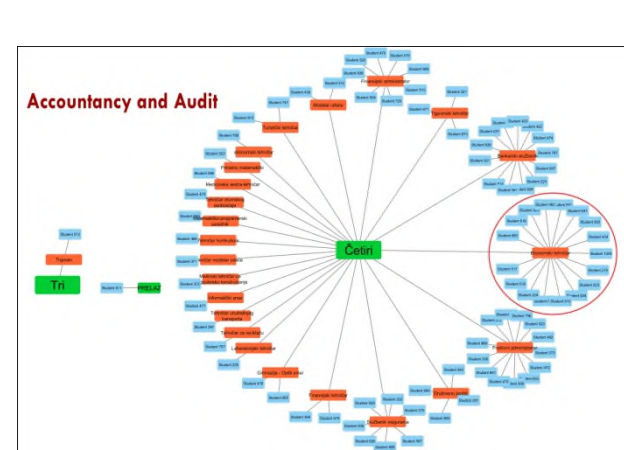**Figure 3:** High schools finished by students enrolled on the Mechanical Engineering study programme

Figure 4: High school finished by students enrolled on the Environmental Protection Engineering programme

As to the Mechanical Engineering study programme (Figure 3), most of students have also finished four-year high schools, such as those educating mechanical engineering technicians for computer-aided design, domestic transport technicians, mechatronics technicians, etc. However, unlike other study programmes, this one enrolls much more students who have acquired three-year secondary education, mostly machine operators.

As to the Environmental Protection Engineering (Figure 4), most of students have finished four-year high schools educating recycling technicians, economics technicians, environmental protection technicians, mechanical engineering technicians for computer-aided design, etc.
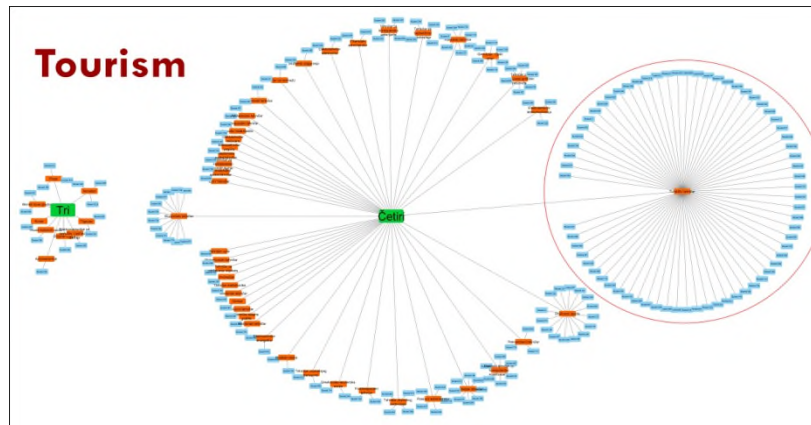

Figure 5: High schools finished by students enrolled on the Management study programme

Figure 6: High schools finished by students enrolled on the Accountancy and Audit study programme

As to the high school degree, most of students enrolled on the Management study programme (Figure 5) have finished four-year high schools, mostly high schools of economics, and grammar schools (the social studies and languages stream), etc.

The Accountancy and Audit study programme (Figure 6) has been enrolling students for the last three academic years, during which period they have enrolled only one student with three-year secondary education, and one student transferred from other college/faculty. Students enrolled on this study programme have mostly graduated from high schools of economics (economics technicians, business administrators, bank officers, financial administrators, etc).

Figure 7: High schools finished by students enrolled on the Tourism study programme

The Tourism study programme (Figure 7) mostly enrolls students with finished four-year high schools. A significant number of students finished high schools offering similar vocational education, i.e. technical high schools educating tourism technicians. There are also students who finished grammar schools – the social sciences and languages stream; mechanical engineering technicians for computer-aided design, etc.

The above given diagrams show that students mostly pursue further studies in similar vocational programmes. Namely, those who have finished technical high schools mostly enroll on study programmes in the field of technical sciences and technology. In a similar way, students who have finished high schools of economics and grammar schools mostly enroll on study programmes in the field of social sciences and humanities.

3.2. What was students' academic achievement at the end of high school?

In order to classify BTC students according to their high school academic achievement, it is necessary to create diagrams showing the connection between these categories. Figure 8 shows the total percentage of students with excellent, very good, good and sufficient achievement, and at the end (the smallest diagram) the percentage of students who have transferred college or faculty (marked as 'transfer').
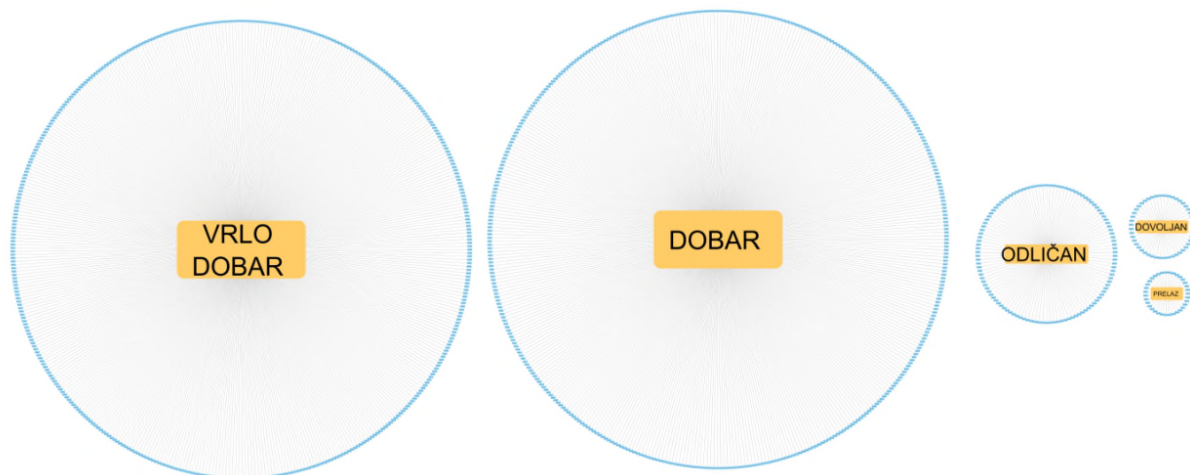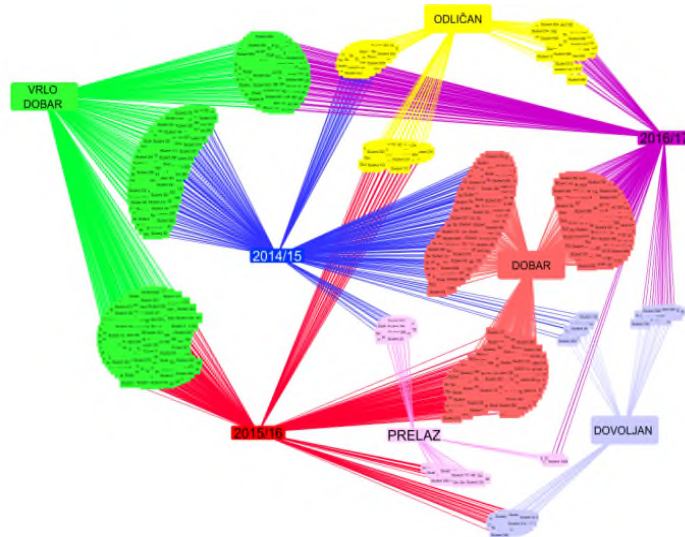

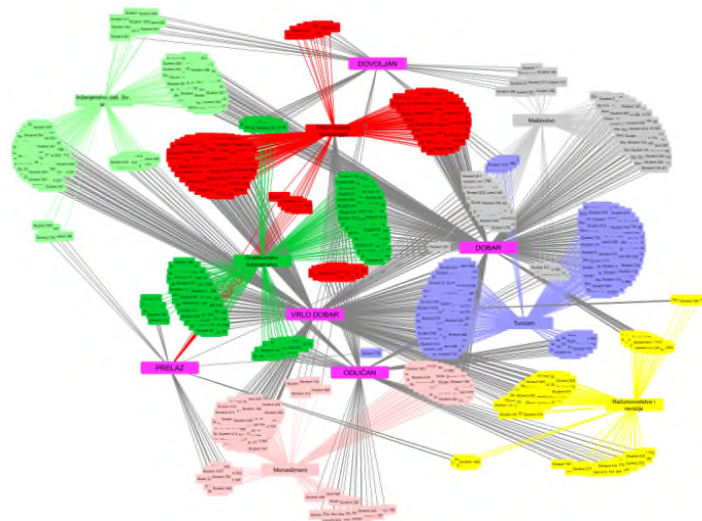**Figure 8:** Students' high school academic achievement

According to the results reported in Figure 8, 41.3% of students have attained very good, 39.5% good, 11% excellent, 4.7 sufficient academic achievement, and 3.4% of students have been transferred from another college or faculty.

If, in addition to students and their academic achievement, the enrollment year is included into the diagram as another criterion, a somewhat more complicated review is obtained, as shown in the following figures. In Figure 9, different colours are used to present each enrolled student, with his/her academic achievement and enrollment year. From the picture, it can be seen that the number of students with good high school academic achievement was more or less the same in the three analysed enrollment years, whereas the number of those with excellent high school academic achievement was the greatest in the 2016/2017 enrollment year.

Having added the enrolled study programme to the above given data as additional criterion, the review as shown in Figure 10 has been obtained. The picture shows each student, the study programme he/she enrolled on, his/her enrollment year and high school academic achievement.
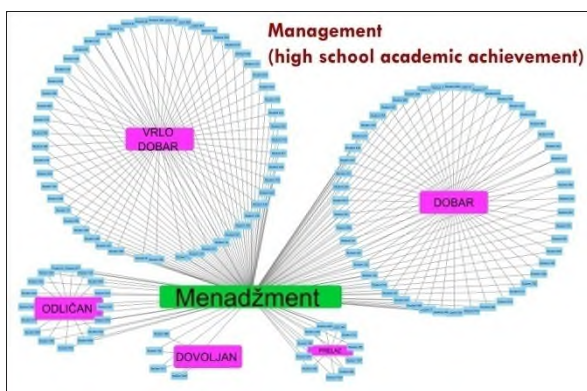
**Figure 9:** A summary review of students per enrollment year

Figure 10: A summary review of students' high school academic achievement and study programmes enrolled

The results of the analysis of students' high school academic achievement per study programme are shown in the following figures (Figures 11-17)
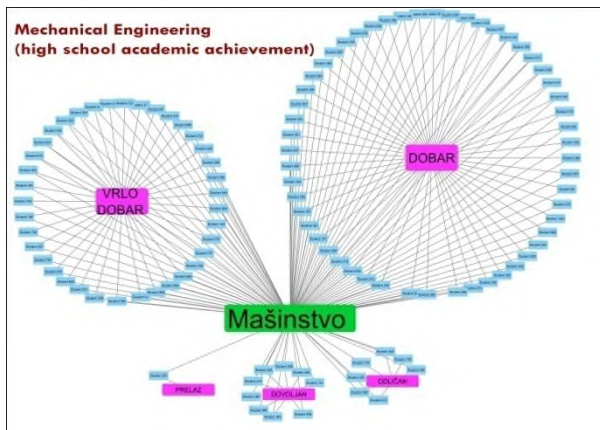

Figure 11: High school academic achievement of students enrolled on the Management study programme in the three analysed enrollment years
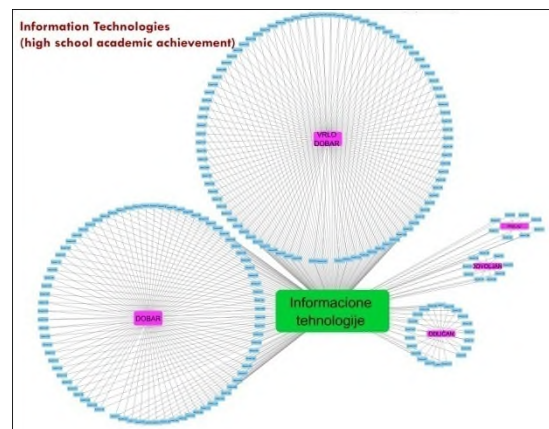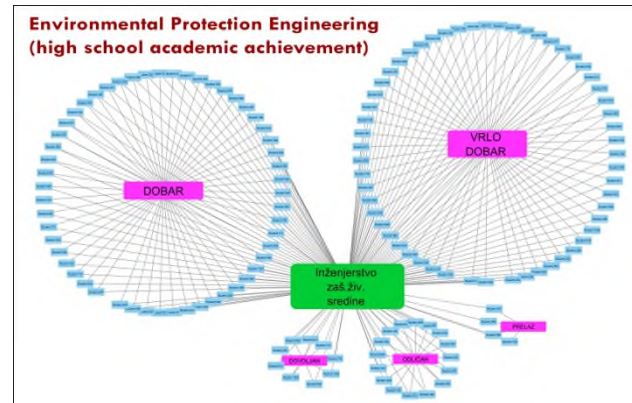

Figure 12: High school academic achievement of students enrolled on the Information Technology study programme in the three analysed enrollment years

Figure 11 shows students enrolled on the Management study programme and their high school academic achievement. The greatest number of students enrolled on this study programme have attained very good high school academic achievement, and are closely followed by those with good academic achievement, whereas the number of those with excellent and sufficient high school academic achievement, as well as the number of students transferred from other colleges or faculties is lower to some extent.

As to the Information Technology study programme, the results are almost the same (Figure 12). This study programme also enrolled most students with very good high school academic achievement; then those with good, excellent and sufficient academic achievement, and those transferred from other colleges or faculties.
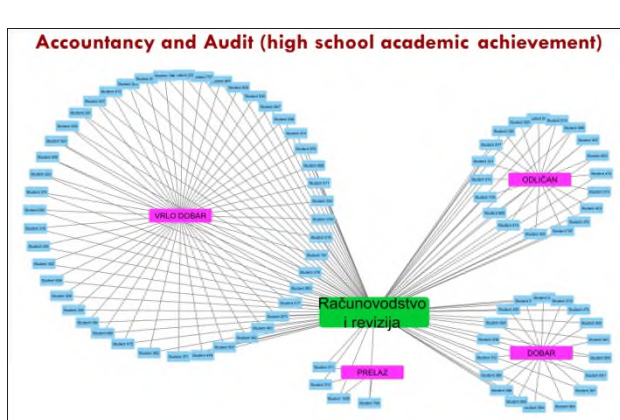

Figure 13: High school academic achievement of students enrolled on the Mechanical Engineering study programme in the three analysed enrollment years
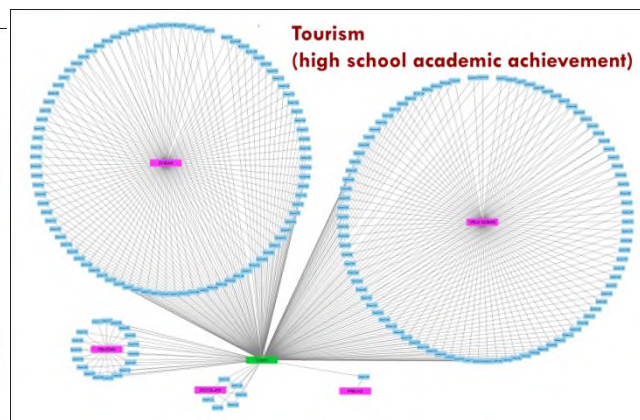

Figure 14: High school academic achievement of students enrolled on the Environmental Protection Engineering study programme in the three analysed nrollment years

Regarding the Mechanical Engineering study programme (Figure 13), the situation is somewhat different. The average high school academic achievement of most students was good; then come those with very good high school academic achievement, and then a small number of students with sufficient or excellent high school academic achievement.

Figure 15: High school academic achievement of students enrolled on the Accountancy and Audit study programme in the three analysed enrollment years

Figure 16: High school academic achievement of students enrolled on the Tourism study programme in the three analysed enrollment years
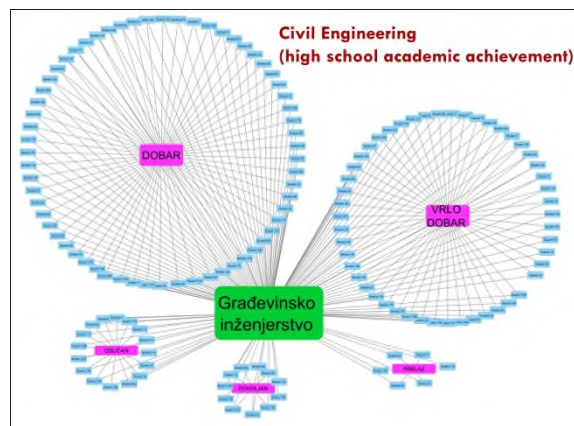
Figure 17: High school academic achievement of students enrolled
on the Civil Engineering study programmein the three analysed enrollment years

The average high school academic achievement of most students enrolled on the Environmental Protection Engineering, Accountancy and Audit, and Tourism study programmes was very good; whereas most Civil Engineering students had good high school academic achievement on average.

According to the above given figures, the conclusion can be made that most BTC students have finished high school with very good academic achievement, and generally speaking, the best students are enrolled on the Accountancy and Audit study programme, whereas the weakest ones are enrolled on the Mechanical Engineering study programme.

4. CONCLUSION

This paper presents the results of the research carried out using the decision tree technique and the true data about BTC students, obtained from the database kept by the Student Service Department. What makes such a method of research different from traditional statistical methods is that it does not necessarily follows previously determined rules, but allows researchers to use their creativity to discover some new, maybe even unexpected rules and characteristics during data analyses [7, 8]. Furthermore, traditional statistical methods imply certain previously determined research models and formulas and therefore certain expected results. On the other hand, without any previous assumptions regarding outcomes, data mining makes it possible to reach them, thus obtaining accurate results, because the less expected the obtained results, the lower the researches' influence on them. Unlike traditional statistical methods relying on testing previously suggested hypotheses, data mining focuses on research using software as a tool to define the connection and interdependence between a great number of factors, thus providing answers that can help to solve numerous problems [9, 10]. Therefore, the most important data mining feature is the fact that we can raise a question and obtain new objective and reliable answers that can be used to improve the existing conditions without any previous assumptions on the expected answers.

The main aim of this paper was to discover the important characteristics of high schools finished by BTC students by analysing the data about the first-year students, i.e. to discover whether they have chosen the same or similar field of study compared with their high school education, as well as to discover their high school academic achievement. The obtained results can be further used to predict the steps to take when planning the promotion of all or individual study programmes. For example, the discovered fact that an insignificant number of students who finished grammar schools enrol on the Mechnical Engineering study programme has led to the conclusion that no special efforts should be made to promote this study programme in such high schools. On the other hand, if there is a potential for such a thing, the promotion of certain study programmes should be enhanced in order to attract more students.

The results presented in this paper show that such research can help to define different parametres that can determine or influence the enrollment of students in the BTC Užice in this particular case, but the same procedure can be used by other higher education institutions as well.

REFERENCES

[1] Mladen Janjić, Vera Lazarević, 'Covariance and Corellation of High School Success and Entrance Exam Results', Technics and Informatics in Education, 5th International Conference, Faculty of Technical Sciences Čačak, 2014;

[2] Lakshmi Sreenivasa Reddy, Vijaya Bhaskar Velpula, B. Raveendrababu, 'Finding Peculiar Students From Student Database Using Outlier Analysis: Data Mining Approach', MOOC, Innovation and Technology in Education (MITE), 2014 IEEE International Conference, Patiala, India, 2014;

[3] Yue Zhang, 'Study on the Problems and Countermeasures of College Students' Psychological Education Based on Network Data Mining', AISTI, 2016;

[4] Ivan Budišćak, Vedran Mornar, Danijel Mlinarić, 'Analiza podataka u svrhu predlaganja studijskih programa u Nacionalnom informacijskom sustavu prijava na visoka učilišta', MIPRO 2014/miproBIS, Opatija, 2014;

[5] Zailani Abdullah, Tutut Herawan, Noraziah Ahmad, Mustafa Deris Mat, 'Extracting Highly Positive Association Rules from Students' Enrollment Data', In World Conference on Educational Technology Researches, 2011;

[6] http://cytoscape.org/manual/Cytoscape2_8Manual.html#Cytoscape User Manual, accessed in May 2017;

[7] Amjad Abu Saa, 'Educational Data Mining & Students' Performance Prediction', IJACSA, 2016;

[8] Velida Kijevčanin, ŠabanGračanin, PhD thesis: „Sve je u podacima, samo treba naći", Kragujevac, 2009;

[9] Miloš Radovanovic, Jure Ferlež, Dunja Mladenić, Marko Grobelnik, Mirjana Ivanović, 'Extending the IST-World Database with Serbian Research Publications', 9th International Multi-conference on Information Society, pp. 251-254, Ljubljana, Slovenia, 2006;

[10] Raheela Asif, Agathe Merceron, Syed Abbas Ali, Najmi Ghani Haider, 'Analyzing Undergraduate Students' Performance Using Educational Data Mining', Elsevier Ltd, 2017.