Determinants of Efficacy of Studying in the Republic Croatia - Comparing Neural Networks and Decision Trees: Research Framework Proposition

Alisa Bilal Zorić
Veleučilište Baltazar, Croatia

Abstract

Rapid technological development and progress lead to the need for better and more efficient education which should prepare the applicant for increasingly flexible labour market. The goal of this research is to create models for prediction of student's efficacy, compare them, find the key factors that contribute to more efficient studying in the Republic of Croatia, and finally determine how efficient studying is related to first employment. Models will be based on students' data and hypothesis will be tested using multivariate statistical methods (multiple regressions, Cronbach's alpha), decision trees and neural networks. Data will be collected by structured questionnaire and will consist of demographic and economic data, information about previous education, attitudes towards learning, and goals after completing studies and information about the first employment. Students' efficacy will be measured by grade point average in college. This research will try to increase our understanding of how different factors influence students' performance and how students' efficacy affects the speed and conditions of finding the first employment.

Keywords: efficacy of studying, education, neural network, decision trees
JEL classification: I230, C450

Introduction

With the accelerating technological development and progress, learning and education are taking on a more important role in the development of our society. Data mining has been used in industries for a long time and for a variety of applications in order to achieve profitability. Similarly, data mining techniques can be applied on universities. This field is called educational data mining. It concerns developing methods that discover knowledge from data that come from educational settings, and using those methods to better understand students (Romero et al., 2008). Educational Data Mining focuses on developing new tools and algorithms for discovering data patterns. EDM uses techniques from statistics, machine learning and data mining to analyse data from educational environment ((Romero et al., 2013).

The discovered knowledge can be used to improve teaching process, for better understanding of students' behaviour, for predicting enrolment of students into various courses, to assist instructors, for predicting student performance, to evaluate and improve e-learning systems, to improve curriculums, for dropout students and many other benefits (Ali, 2013). There are four main stakeholders involved in educational data mining process: Learners who are dealing with student needs and

methods to improve the learner's experience and performance, Educators who attempt to understand learning process and try to find methods they can use to improve their teaching methods, Researchers who focus on the development and the evaluation of data mining techniques for effectiveness and Administrators who are responsible for allocating resources for implementation in institutions (Romero et al., 2013).

In this paper, we will focus on predicting student performance. Among the popular techniques to predict student performance are: classification/regression methods such as Bayesian networks (Minaei-Bidgol et al., 2003, Bekele et al., 2005, Thai-Nghe et al., 2009), logistic regression (Cen et al.,2006), linear regression (Feng et al., 2009), decision trees (Thai-Nghe et al., 2007, Baradwy et al., 2012), Romero et al., 2008), neural networks (Romero et al., 2008) and support vector machines (Thai-Nghe et al., 2009). We want to make a predicting model from student data which will predict students' performance based on demographic and economic data, information about previous education, attitudes towards learning and goals after graduation and also, we want to see if there is a relationship between academic performance and first employment.

Data will be collected by a structured questionnaire and will include the data from 300 students from different Croatian faculties of economics and business. The main hypothesis is that there are different factors which affect students' performance and that more successful students (with higher grades) find better employment (higher first salary, shorter job search).

Research model

We want to isolate factors that influence students' success measured by the average grade. There are many researches dealing with different factors that influence student's success, but there are only a few researches that are dealing with the combination of factors, especially in Croatia (Zekić-Sušac, 2009).

There are many studies that examine the impact of parents' education on the educational achievement of their children (Ermischet al., (2001); Johneset al., (2004); Black et al. (2005); Ishitani (2006), Ortiz et al., (2008) and Stratton et al., (2008)). There are also many studies dealing with the influence of age on students' success, but the results are inconsistent. Some studies show a negative relationship (Clark et al., 1990), while others suggest that more mature students with a clear idea of the future profession achieve better results (McInnis et al., 1995).

Based on this, the first research question is:

1. What is the relationship between demographic characteristics and students' success?

Following hypotheses are developed to answer this research question:

H1a: Age has a positive, statistically significant impact on students' success measured by the average grade.

H1b: Parent education has a positive, statistically significant impact on students' success measured by the average grade.

Magdalena (2013) found a strong positive relationship between the previous and future education, and confirmed that the positive or negative attitude towards learning appears in very early schooling. Sulaiman and Mohezar (2006) showed that a student's undergraduate grades are the best indicator of the student's success.

Based on these researches, the second research question was developed:

2. What is the relationship between students' previous education and students' success?

Following hypotheses are developed to answer this research question:

H2a: Previously adopted learning habits have a positive, statistically significant impact on students' success measured by the average grade.

H2b: The success of the final test has a positive, statistically significant impact on students' success measured by the average grade.

Stratton et al. (2008) found that socio-economic background and economic factors such as unemployment rates have a major impact on full-time students. Lecompte, Kaufman and Rousseeuw (1983) found that financial problems are a common reason for dropping out. Robbins et al. (2004) emphasize the importance of financial aid during studies. Financial aid has a direct effect on the motivation of a student if high performance is required to keep receiving scholarships. Students who receive a scholarship are persistent and rarely drop out (St John et al., 2005). Although many studies (Cabrera et al., 1992.; Gross et al., 2007; Singell, 2004) confirmed the positive effect of receiving scholarships to student success, there are studies that have found a negative relationship (St. John and Starkey, 2005; Dowd and Coury, 2006). Studies that are dealing with the impact of living in a dorm on students' success are also inconsistent. Based on those researches, the third research question was developed:

3. What is the relationship between financial conditions during studies and students' success?

Following hypotheses were developed to answer this research question:

H3a: Higher household income has a positive, statistically significant impact on students' success measured by the average grade.

H3b: Living in a dorm has a positive, statistically significant impact on students' success measured by the average grade, while living in the parents' house has a negative, statistically significant impact on students' success measured by the average grade

H3c: Scholarships have a positive, statistically significant impact on students' success measured by the average grade

H3d: Working during studies has a negative, statistically significant impact on students' success measured by the average grade

Himelstein (1992) found that students with a clear vision of their future career goals achieve better results and are less likely to drop out than students without a clear vision. Students with clear objectives about their future careers are more confident and their success is more likely than of their colleagues without a clear objective after studies (Garcez , 2007). In that sense, the fourth research question is:

4. What is the relationship between students' career goals and students' success?

The following hypothesis was developed to answer this research question:

H4: Clear career goals have a positive statistically significant impact on students' success measured by the average grade.

Based on a research conducted by Wise (1975), in which he demonstrated a positive relationship between the salary and average grades in college, but in which the speed of finding the first job was not explored, we developed the fifth research question and hypothesis:

5. What is the relationship between students' success and the first employment?

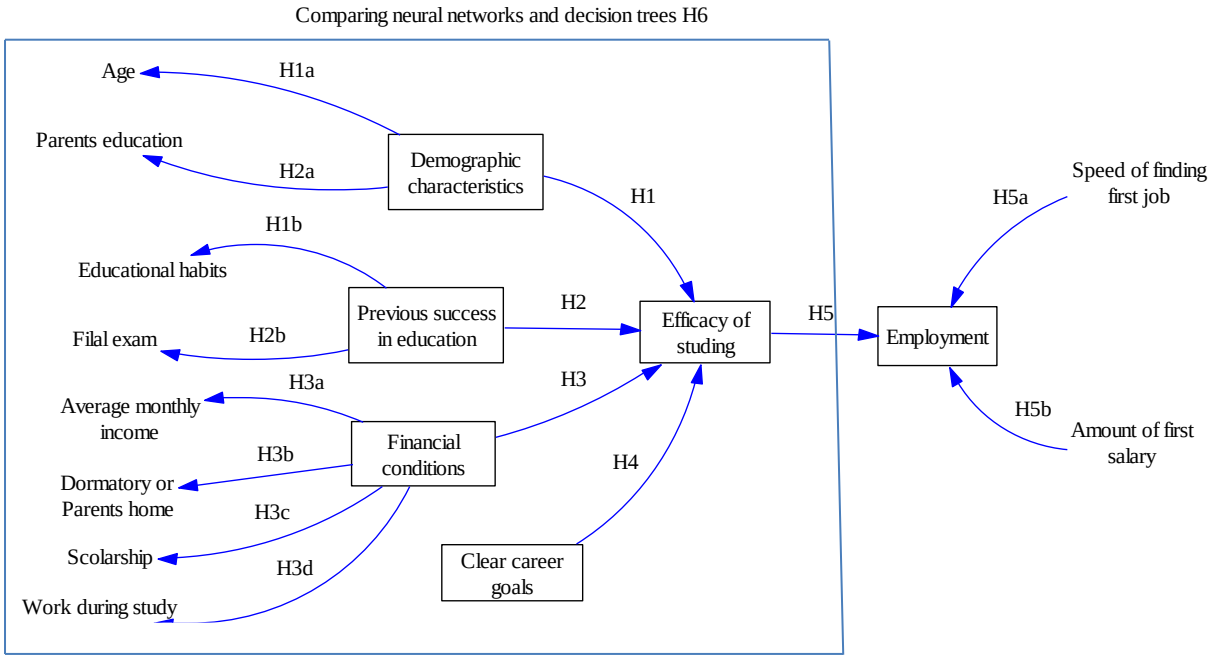
H5a: Better grade point average has a positive, statistically significant impact on the speed of finding the first employment.

H5b: Better grade point average has a positive, statistically significant impact on the amount of the first salary.

In the last hypothesis, we will compare two prediction models, one created with the help of a decision tree and another created with a neural network algorithm.

According to the mentioned research questions and hypotheses, the following research model is proposed (Figure 1).

Figure 1 Proposed Research Model



Source: Author`s work

Research Methodology

A structured questionnaire will be used to collect data from 300 students from different Croatian faculties of economics and business who graduated between 2012 and 2014. Communication with respondents will be carried out via e-mail. This research will be conducted in two phases.

The first stage will be a preliminary survey testing, which aims to test the understanding of the questionnaire. In the preliminary part of the study a small number of respondents will be contacted to check whether the survey questions are unambiguous and whether respondents understand them. Based on the results of the first phase of the research, the questionnaire will be adapted.

The second phase will be a descriptive study using a structured questionnaire. This part of the research will be conducted using an online survey to which respondents will be invited via e-mail. The survey questionnaire will consist of demographic and economic characteristics, information about previous education, attitudes towards learning, and goals after completing studies and information about the first employment. Students' efficacy will be measured by grade point average in college. An analysis of the data will be conducted using SPSS. Following statistical methods will be employed: exploratory and descriptive statistical methods, multivariate and simple linear regression, Cronbach's alpha test, decision trees and neural networks.

The first hypothesis (H1) presumes that students' performance is positively influenced by demographic characteristics. The first hypothesis will be tested using the supporting H1a sub-hypothesis (the impact of age) and H2a sub-hypothesis (the impact of parents' education. This hypothesis will be tested using the results of primary research and linear multiple regression model.

The second hypothesis (H2) presumes that students' performance is positively influenced by students' previous education. The second hypothesis will be tested using the supporting H2a sub-hypothesis (the impact of previously adopted learning habits) and H2b sub-hypothesis (the impact of the success on the final test). This hypothesis will be tested using the results of primary research and linear multiple regression model.

The third hypothesis (H3) presumes that students' performance is positively influenced by financial conditions during studies. The third hypothesis will be tested using the supporting H3a sub-hypothesis (the impact of higher household income), H3b sub-hypothesis (the impact of living in a dorm) and H3c sub-hypothesis (the impact of scholarships) and H3d sub-hypothesis (the impact of working during studies). This hypothesis will be tested using the results of primary research and linear multiple regression model.

The fourth hypothesis (H4) presumes that students' performance is positively influenced by clear career goals. This hypothesis will be tested using the results of primary research and linear multiple regression model.

The fifth hypothesis (H5) presumes that first employment is positively influenced by students' performance. The fifth hypothesis will be tested using the supporting H5a sub-hypothesis (the impact of better grade point average on the speed of finding first employment) and H5b sub-hypothesis (the impact of better grade point average on the amount of first salary). This hypothesis will be tested using the results of primary research and linear multiple regression model.

In the sixth hypothesis, we will use neural networks and decision trees to create and compare two prediction models.

Conclusion

In this paper, we presented a case study in educational data mining. It showed how useful data mining can be in higher education, particularly in improving student performance. We applied data mining techniques to discover knowledge.

Findings and results will have future practical implications since they will allow the identification of those factors which may affect future students' studying performance. These results can be used in future discussions on the improvement of higher education system in Croatia.

The practical contribution of this thesis will consist in pointing out that successful studying has a beneficial effect on employment, that it affects the speed of finding the first job and that it has an effect on the amount of salary.

Major limitation to the proposed research is the fact that the data collected by the survey represents an insight into the trend of occurrences within a limited period of time. Hence, it does not provide an analysis of correlations and changes over an extended period of time.

References

1. Ali, D. M. M. (2013), „Role of data mining in education sector", International Journal of Computer Science and Mobile Computing, Vol. 2 No. 4, pp. 374-383.
2. Baradwaj, B. K., Pal, S. (2012), „Mining educational data to analyze students' performance", Computer Science, Vol. 2 No. 6, pp. 63-69.
3. Bekele, R., Menzel, W. (2005), „A Bayesian Approach to Predict Performance of a Student (BAPPS): A Case with Ethiopian Students", Artificial Intelligence and Applications, Vienna, Austria, pp. 189–194.

4. Cen, H., Koedinger, K., Junker, B. (2006), „Learning Factors Analysis A General Method for Cognitive Model Evaluation and Improvement", Intelligent Tutoring Systems, Springer Berlin Heidelberg, Vol. 4053, pp. 164–175.
5. Clark, E. E., & Ramsay, W. (1990), „The importance of family and network of other relationships in children's success in school", International Journal of Sociology of the Family, pp. 237-254.
6. Ermisch, J., Francesconi, M., (2001), „Family matters: Impacts of family background on educational attainments", Economica, pp.137-156.
7. Feng, M., Heffernan N., Koedinger, K. (2009), „Addressing the assessment challenge with an online system that tutors as it assesses", User Modelling and User Adapted Interaction, Vol. 3 No. 19, pp. 243–266.
8. Fu, T.C., Lui, C.L (2007), „Agent-oriented Network Intrusion Detection System Using Data Mining Approaches", International Journal on Agent-Oriented Software Engineering, Vol. 1 No. 2, pp. 158-174.
9. Garcez, C., (2007), „The Effects of Career Goals on Students", available at: http://aplawrence.com/Misc/cgcareergoals.html (4/10/2015)
10. Gross, J. P., Hossler, D., Ziskin, M. (2007), „ Institutional aid and student persistence: An analysis of the effects of institutional financial aid at public four-year institutions", Journal of Student Financial Aid, Vol. 1 No. 37, pp. 6.
11. Himelstein, H. C. (1992), „Early identification of high-risk students-using noncognitive indicators", Journal of College Student Development, Vol.1 No. 33, pp. 89-90.
12. Ishitani, T. (2006), „Studying attrition and degree completion behaviour among first-generation college students in the United States", Journal of Higher Education, Vol. 5 No. 77, pp. 861-885.
13. Johnes, G., McNabb, R. (2004), „Never give up on the good times: Student attrition in the UK", Oxford Bulletin of Economics & Statistics, Vol.1 No. 66, pp. 23- 47.
14. Lecompte, D., Kaufman, L., Rousseeuw, P., Tassin, A. (1983), „Search for the relationship between academic performance and some psychosocial factors", Acta Psychiatria Belgia, pp. 598-608.
15. Magdalena, S. M. (2013), „Research Directions on the Psychosocial Predictors of Education Success at the Beginning of School", Procedia-Social and Behavioural Sciences, Vol. 1 No. 76, pp. 785-789.
16. McInnis, C., James, R. H., McNaught, C. (1995), „First year on campus: Diversity in the initial experiences of Australian undergraduates", AGPS for Centre for the Study of Higher Education, University of Melbourne.
17. Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., Punch, W. (2003), „Predicting student performance: an application of data mining methods with an educational web-based system", Frontiers in education, Vol. 1 No. 33, pp. T2A-13.
18. Ortiz, E. A., Dehon, C. (2008), „What are the factors of success at university? A case study in Belgium", CESifo Economic Studies, Vol. 2 No. 54, pp.121-148.
19. Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., Carlstrom, A. (2004), „Do psychosocial and study skill factors predict college outcomes?", Psychological bulletin, Vol. 2 No. 130, pp. 261.
20. Romero, C., Ventura, S., (2013), „WIREs Data Mining KnowlDiscov", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol.1 No. 3, pp. 12–27.
21. Singell, L. D. (2004), „Come and stay a while: does financial aid effect retention conditioned on enrolment at a large public university?", Economics of Education Review, Vol.5 No. 23, pp.459-471.
22. St.John, E. P., Paulsen, M. B., Carter, D. F. (2005), „Diversity, college costs, and postsecondary opportunity: An examination of the financial nexus between college choice and persistence for African Americans and Whites", The Journal of Higher Education, Vol. 5 No. 76, pp. 545-569.
23. Stratton, L., O'Toole, D., Wetzel, J. (2008), „A multinomial logit model of college stopout and dropout behaviour", Economics of Education Review, Vol. 3 No. 27, pp. 319-331.

24. Sulaiman, A., Mohezar, S. (2006), „Student success factors: Identifying key predictors", Journal of Education for Business, Vol. 6 No. 81, pp. 328-333.
25. Thai-Nghe, N., Busche A., Schmidt-Thieme L. (2009), „Improving Academic Performance Prediction by Dealing with Class Imbalance", Proceeding of 9th IEEE International Conference on Intelligent Systems Design and Applications (ISDA'09), Pisa, Italy, IEEE Computer Society,pp. 878–883.
26. Thai-Nghe, N., Janecek, P., Haddawy, P. (2007), „A Comparative Analysis of Techniques for Predicting Academic Performance", in: Proceeding of 37th IEEE Frontiers in Education Conference (FIE'07), Milwaukee, USA, IEEE Xplore, T2G7–T2G12,
27. Ventura, S., Espejo P. G., Hervs, C. (2008), „Data Mining Algorithms to Classify Students", in: 1st International Conference on Educational Data Mining (EDM'08), Montral, Canada, pp.8–17.
28. Wise, D. A. (1975), „Academic achievement and job performance", The American Economic Review, pp. 350-366.
29. Zekić-Sušac, M., Frajman-Jakšić, A., Drvenkar, N. (2009), „Neuronske mreže i stabla odlučivanja za predviđanje uspješnosti studiranja" ("Neural Nets and Decision Trees used for Prediction of Success of Studying"), Ekonomski Vjesnik/Econviews: Review of contemporary business, entrepreneurship and economic issues, Vol. 2 No.22, pp. 314-327.

About the author

Alisa Bilal Zorić is a doctoral student at the Faculty for Commercial and Business Sciences of Celje, Slovenia. She graduated from the Faculty of Science, Department of Mathematics, Zagreb and she has an MBA from the Faculty of Economics, Zagreb. Prior to returning to school to pursue her doctorate, Alisa spent ten years designing and developing business information systems for Siemens and Banksoft. She is currently working as an assistant at the University of Applied Sciences Baltazar Zaprešić, where she teaches math and computer courses. Her current research interests include educational data mining and knowledge management from a data mining perspective. The author can be contacted at **alisa.bilal.zoric@bak.hr.**