STUDENTS WHO HAVE UNSUCCESSFULLY STUDIED IN THE PAST – ANALYSIS OF CAUSES

Petr Mazouch[1✉]**, Veronika Ptáčková**[2]**, Jakub Fischer**[2]**, Vladimír Hulík**[3]

✉[1]University of Economics, Prague, Czech Republic, mazouchp@vse.cz

[2]University of Economics, Prague, Czech Republic

[3]Ministry of Education, Youth and Sport, Prague, Czech Republic

Highlights

- *Students who definitely do not work at their preferred college are 6.3 times more likely to have unsuccessfully completed tertiary education in the past than students who definitely study in their preferred university*
- *Men are 1.426 times more likely to have an unsuccessful past tertiary education than women*
- *Students with a health disadvantage are 1.3 times more likely to have an unsuccessful past tertiary education than students who do not suffer from health complications*

Abstract

With the increase in the number of university students, the number of those who do not finish successfully the tertiary education is also increasing. The article uses a specific data source and analyses only a part of the group of unsuccessful students who re-enroll. This is a specific group of students - they did not finish the tertiary study in the past, but after some time they returned to education. The aim of the paper is to find significant factors that influence the decision whether the student changes the studied school or field of study. Factors will be searched using decision trees and binary logistic regression. Both methods were significant for gender and the fact that a student is studying his preferred university. Logistic regression adds to the student's health disadvantage. The data were obtained from the EUROSTUDENT survey, which was held in the Czech Republic in 2016 under the auspices of the Ministry of Education, Youth and Sports. The results can be used to identify a risky candidate or student at the beginning of tertiary education.

Keywords

Drop-out, EUROSTUDENT, failed studies, human capital, tertiary education, universities in the Czech Republic, unsuccessful students

Introduction

The phenomenon of recent years is the growth of the university-educated population in the Czech Republic. Nevertheless, together with the interest in tertiary education, there is also a growing number of those who fail to complete university studies. This topic is not only about the Czech Republic but also about other EU countries.

In this article, we focus on a specific group of students who have re-enrolled their studies again when the previous studies were unsuccessful. The data from the international survey EUROSTUDENT VI organized by the Ministry of Education, Youth and Sports were used for the analysis. In order to find statistically significant factors of unsuccessful study in the past, we use following statistical methods: binary logistic regression and decision trees - specifically, the CART (Classification And Regression Trees) method. The results of both methods are compared and confronted with conclusions from foreign and Czech studies. This identifies factors which can help characterising a risky candidate or a student at the beginning of a course. The results may help to reduce the proportion of unsuccessful students, which could be interesting for a policy of tertiary education, as well as for study advisers of individual universities and faculties.

Literature Review

The general term for unsuccessful study is usually "drop-out". It does not distinguish whether it captures a course, a degree program or an educational level. There is no uniform definition of this term in the Czech Republic or the world. Most often,

the drop-out is translated into Czech as "early departure from education" or "unsuccessful termination of education".

The drop-out calculation is often complicated. Problems can occur both due to the lack of a clear definition of the concept and the structure of the analysed data. International organizations do not analyse individual programs but levels of education. For example, the Organization for Economic Co-operation and Development (OECD) includes all students who have completed a given level of education without qualifying (Hraba, Hulík, Hulíková Tesárková, 2016). For Eurydice, which deals with the situation of higher education institutions across countries, the following definitions have been used for the Czech Republic: "Unsuccessful termination of tertiary education means a situation when the student fails to appear again as a tertiary student after another unsuccessful graduation for the next three years" (Hraba, Hulík, Hulíková Tesárková, 2016).

The Czech Ministry of Education, Youth and Sport recommends calculating the cohort rate of failure. This rate is associated with the registration year of study. We can calculate it as the ratio between the number of unsuccessfully completed studies in each year of study and the total number of studies commenced in that year of enrolment. The problem is that is focused on the study, not on the student. (Ministry of Education, Youth and Sports, 2017)

At the national level in the Czech Republic, we can use the cohort rate of failure which is connected with the student. It is monitored all years in the tertiary education. This rate is calculated as the ratio between the number of unsuccessfully

completed studies in the concrete cohort (and concrete level of study) and the total number of students who come to the concrete level of study for the first time in the concrete year (Vlk et al., 2017). For more details about drop-out definition, see Vlk et al. (2017).

Some explanations of reasons for unsuccessful studies at universities are based on the theoretical model of student residence in an academic environment designed by Tinto (1975). Tinto's sociological-anthropological model states that a student successfully completes university studies, not only when properly fulfilling study duties but also actively integrating into natural social structures in the academic environment. Tinto points out that if the student more communicates with classmates and faculty, his/her chances of successfully completing the studies are increasing. It emphasizes the responsibility of the school to support the student's academic and social integration (Tinto, 1997).

Tinto (1997) identifies two types of study leaving. The first is the termination of studies because of insufficient learning outcomes (involuntary leaving) and the second voluntary leaving from studies, which can be affected by a number of factors. Tinto (1997) points out that the school should define its duties and obligations towards the student (as well as the student to school). In addition, he identified six basic conditions that support the success of the study: the duty of the school to enhance student success, student expectations, student support, feedback on student performance, student-to-student relationships, and student learning. Tinto (1999) says that the critical period of study is the first year.

Jensen (2011) divides factors into three levels: individual (academic performance, student attitude and satisfaction with study), institutional (conditions created by the school: pro-social climate in school, support services, awareness of student needs, opportunity to participate out-of-school activities) and external social standards (social support: support from parents, friends, schoolmates).

The German Center for Higher Education and Scientific Research[1] (Heublein, 2014) has drawn up a model that highlights the fact that unsuccessful completion of studies cannot be described as an individual failure or problem of the education system but as a complex problem that can be divided into three phases. The preliminary stage is affected by the social status and family background, the content of the study program, the study itself and socialization in the educational process. The second phase reflects the relationship between the internal (motivation, performance, psychological and physical possibilities of the student) and external (study, accommodation) factors. The final decision is the third phase. (Heublein, 2014).

When analysing the effects of terminating studies at German public schools, the factors were divided into three groups: predisposing (social and demographic factors, personality traits, the initial level of knowledge and motivation), important life events (work and family responsibilities), and institutional factors (methods of studies, teachers, administrative support). Fully-employed students, migrants, and women - who have higher expectations than men in the study program and the environment - are included in the risk group. On the contrary, older students with higher motivation for professional and personal growth, and students with a child have a higher odd to graduate (Stoessel et al., 2015).

Wolter, Diem and Messer (2014) found a higher drop-out rate for men and older students. It also depends on the education and employment of the student's parents, the results of admissions,

integration, and motivation. The study highlights the influence of the Bologna process when a lot of master's programme was divided into bachelor and follow-up master's programme. Due to that change, the rate of unsuccessful women decreased.

Kingston (2008) emphasizes emotional intelligence and satisfaction with the learning environment. Vnoučková et al. (2017) point out the importance to have student's feedback not only about ongoing subjects but also at the end of the subject, but also about all their tertiary studies – it can help to increase the quality of the university and student's satisfaction. Kearney and Levine (2016) address the problem of income differentiation and early school leaving. They point out that boys are more responsive to family and economic disadvantages. They suggest that governments should invest more in human capital (low-income students) at secondary schools. Early intervention can thus positively influence decisions on further study.

Pikálková, Vojtěch and Kleňha (2014) confirmed that the number of unsuccessful students has risen in recent years. They assume that half of the students - who attended college in 2012 - have had not finished it. Higher risk of abandonment is attributed to secondary school postgraduate graduates and secondary vocational schools with graduation. According to the authors, the rate of departures varies with the field of study. Students of technical disciplines are more likely not to have over-pressure in admissions. Mathematics, physics or agriculture students also leave more often.

Fučík and Slepičková (2014) emphasize that students who went to study as a so-called deferred choice are more likely to leave (the students went to college for which they were admitted and then left). Again, this is a conflict between expectation and reality. Also, family and professional opportunities have influence. Charvát et al. (2014) stress the importance of interest and satisfaction with the study. Rubešová (2009) shows the connection between the success of the university studies with the result of the admission procedure and the secondary school achievement. Konečný, Basl and Myslivečel (2010) confirmed these results. They say that students from grammar schools are less risky because they have better preparation for entrance examinations and study.

Hloušková (2014) points to internal factors of incomplete university studies, low socio-economic and cultural status, unfavourable family environment, fostering and educational aspirations of parents. External factors are the difficulty of study, university environment, teaching teacher skills and the rules of the educational institution. In addition, she mentions the influential events of pregnancy, injury, illness or poor school choice.

Menclová, Pacnerová and Vacek (2008) came up with the term "amotivation", which indicates little or no motivation to study at students who do not know what jobs they want to do in the future. They begin to study the field for which they successfully passed entrance examinations. They also work with the concepts of "leaving behind something" and "leaving as an escape". "Leaving for something" captures a situation when a student stops studying for work or family reasons. "Leaving as an escape" capture the termination of studies that arose from stress, crisis situations, conflict, inability to combine the field of study with personal interests, abilities, and talents.

Data – EUROSTUDENT VI

The EUROSTUDENT - international project - seeks to obtain comparable data on the social dimension of European higher education. The survey should clarify issues related to the living conditions and attitudes of students in bachelor and master

---

1   Deutsches Zentrum für Hochschul-und Wissenschaftsforschung

programs taught in Czech in five key areas (German Centre for Higher Education Research and Science Studies, 2017):

- the permeability of studies,
- student relationship to school,
- living conditions of students,
- the foreign mobility of students and language skills,
- students with disabilities.

For the first time, EUROSTUDENT was organized by the European Higher Education Area (EHEA) in 1994. In recent years, the country has been striving to maximize EHEA, provide high-quality higher education, increase graduate employment, and improve student international mobility as a tool for improving learning outcomes. The financial and economic crisis has affected student living conditions (Hauschild et al., 2015). This is one of the reasons why today ministers seek public funding for higher education, reduce inequalities, and provide quality support to students during their studies, individual consultations and the diversity of the studied subject areas. They want to increase employment and student international mobility (Hauschild et al., 2015).

The sixth wave of this international survey was held in 2016. Respondents were: public, state and private universities in the Czech Republic which have accredited bachelor, master or postgraduate courses taught in the Czech language. Over 230,000 students were approached within the project. 22,207 students entered the questionnaire, but 16,602 students completed it. After a detailed analysis of the data, fifty-one questionnaires that were not filled completely but fulfilled minimum requirements were added to the calculation. Weights were assigned on the basis of data from the United Students Register Information System containing gender, age, type of study program, and college (Fischer et al. 2016).

The variables from EUROSTUDENT VI, which were selected on the basis of a literature review, came into the analysis. There were the social and demographic factors with which a student came to university and which could have influenced unsuccessful studies in the past:

- type of high school,
- gender,
- the social status of parents,
- mother's highest education,
- father's highest education,
- mother's job,
- father's job,
- the answer to the question: "Was your university preferred option?",
- health handicap.

A variable *unsuccessful study in the past* is a dependent variable that can acquire two values: "yes" and "no". Unsuccessful college studies are defined in the EUROSTUDENT VI survey as termination of study without a title (failure to meet study requirements, termination at their own request, etc.).

Methods

Two statistical methods were used to find significant factors: logistic regression and decision trees. The methods were chosen for the binary explanation of the variable and the character of the task solution. According to available sources, EUROSTUDENT data were processed for the first time in this way.

Decision trees

Structure of decision trees looks like a reversed tree that displays a hierarchical set of relationships between dependent and independent variables. The method can be used not only to classify individuals but also to classify a set where the starting population (e.g. respondents) is divided into smaller homogeneous groups (respondents who are characterised by some property). In addition, this method detects dependence between dependent and independent variables (Vild, 2012). Trees are formed by using different algorithms – they are different in optimal cleavage. In this case, the Classification and Regression Tree (CART) method was used, which is good for categorical and regression tasks. Trees arise from a recursive binary division. At the beginning of tree formation, all observations are brought to one node (root). The observations are divided gradually into two daughter nodes based on the value and the predictor *X*. The division to the other nodes is binary again (Breiman et al., 1984). Predictor *X* should divide the dependent variable so that the values of the dependent variable inside the node resemble as much as possible but different as much as possible between the nodes. The homogeneity of a node is determined by the Gini index, entropy, or classification error (Komprdová, 2012).

Classification forest will be created by a combination of classification trees. The value of the predictor vectors is determined by each tree in the given class. Voting is determined by the classification function. Regressive forests that contain regression trees are generated by a similar procedure, the resulting regression function is calculated as the average of regression functions of individual trees (Klaschka, Kotrč, 2004).

Logistic regression

Logistic regression is used to find the best - meaningful model. This model describes the relationship between the dependent variable and the group of independent variables. Binary logistic regression is used in this analysis because the dependent variable has only two values. An easy interpretation of the results is an advantage of this method (Řeháková, 2000). In addition, the output can be described as a mathematical model. A model displays the relationship of the dependent variable to the other independent variables. A model allows for the stepwise selection of the independent variables (Tufféry, 2011, Hosmer, 2000).

Model quality

Model quality is evaluated as a whole (not as a component). The ability to predict effectively the values of the dependent variable using independent variables based on observed data means a quality. Among the methods which that model evaluates belong: classification table, ROC[2] curve, statistics (Cox-Snell determinant, Nagelkerk determination factor, 2LL) (Hosmer, Lemeshow, 2000). The classification table records the number of correctly and incorrectly classified objects. On the main diagonal, we can find correct classified objects. As a consequence of the classification table, we can calculate sensitivity and specificity in our logistic regression model. Sensitivity is the probability that the object with the positive answer is classified correct. Specificity is the probability that unsuccessful object is classified as unsuccessful. The graph, which illustrates the relation between sensitivity and specificity, is called ROC curve. *X*-axis values are calculated as (1 – *specificity*), *Y*-axis values are *sensitivity* (Betinec, 2006). The theoretical ROC curve for a random predictor (i.e. for a zero-discriminatory test) leads from the lower left to the top right corner. ROC curve is drawn in a unit square. The closer ROC curve is to the top left corner

---

2   Receiver Operator Characteristic

of the unit square, the better is the discriminative quality of the test (Tufféry, 2011).

Random forest is another option for verifying the quality of a model. It consists of one thousand trees with the same dependent variable. The difference is that each time the data are randomly divided into the training and test set. The software used assesses the importance of the variables involved, whichever is closest to the root node. Subsequently, according to significance, for the explained variable (unsuccessful study), it is sorted downwards according to predictive and confidential significance.

There are two metrics calculated during calculation: Mean Decrease Accuracy and Mean Decrease Gini. Mean Decrease Accuracy says how the accuracy decreases on average when the given tree model variable in the given forest is dropped. Mean Decrease Gini is related to the Gini index for that independent variable. The figure says how much variability, resp. diversity, of the dependent variable can the independent variable explain. A variable with a higher value brings better results.

The calculations were performed using the statistical program R.

Results

Main results from the survey Eurostudent VI

In the Czech Republic, one-fourth of college students have experience with unsuccessful studies (24.8%). These students could identify a combination of factors in the questionnaire which played a role in deciding to leave tertiary education. The most frequent reasons were: dissatisfaction with the content of the study (45.3%), high study intensity (38.6%), dissatisfaction with the quality of teaching (19.6%), lack of social integration (17.2%) and the fact that completed study was only a "backup option" (15.9%). Men left the university because they have a job opportunity or lack of social integration. Women left for the health and family reasons, and because their study was a backup option for them (Fischer et al., 2016).

Decision trees

We used a fixed set of the statistical program R, which states that the trees cannot be more complex than the edge-end metric. The tree was formed by randomly dividing the data into a training and test set. The training set contains seventy percent of the analysed data. The decision tree was created based on this set. The data from the test section was subsequently used to rank in the correct class dependent variable *unsuccessful study.*

The biggest influence on the experience with failed studies in the past had the answers: "*rather not,*" "*certainly not*" to the question: *Was university (which you study nowadays) your preferred option?*[3]. Subsequently, the tree was divided by *gender*. More often, men leave and return to tertiary studies compared to women.

The quality of the model was evaluated by the classification table (Table 1) and the ROC curve. The decision tree (created by CART) has very good prediction capabilities.

		Predicted values	
		Yes	No
Real values	Yes	152	875
	No	119	3,828

Table 1: Classification table, Eurostudent VI (source: own calculation)

Out of 1,027 unsuccessful students, 152 were classified correctly and 875 were misclassified. It is 14.8%. Out of

3,947 successful students, 3,828 were classified correctly and 119 were misclassified. The prediction ability is 97.0%. The accuracy is the most important result. If we sum both the correct and incorrect classifications, we get 152+3,828=3,980 correct classified cases. The total sum of the objects is 4,974. We can calculate the accuracy as 3,980/4,974=0.8001. When we transform it into the percentages, the accuracy is 80.01%.

During calculation the metric Mean Decrease Accuracy (Figure 1), the biggest values were at independent variables: *Was university (which you study) your preferred option*, *father's highest education* and *mother's highest education*. When we remove the variable *Was the university (which you study) your preferred option* from the model, we can classify the wrong 151 students on the average. In the case we remove *father's highest education*, resp. *mother's highest education*, the misclassification can be 46, resp. 43, students on the average.
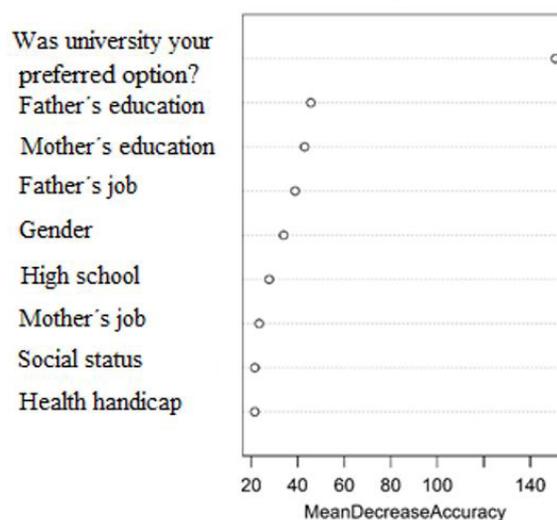

Figure 1: Mean Decrease Accuracy in a random forest, Eurostudent 2016 (source: own calculation)

The most important variables, according to the metric Mean Decrease Gini (Figure 2), were: *Was university (which you study) your preferred option* (282.97), *type of high school* (147.03) and *father's highest education* (143.03). We observe that the satisfaction with the university is a key classifier for drop-out. The result was to be expected because many studies reported the fact that it is important for a student to be in a college he wished to study and was not just a backup option.

The type of secondary school studied was the second major factor. The studies have confirmed that students who come from grammar schools or continue to study in a field of study (which they have at a specialized high school) have a better chance of completing tertiary studies successfully. The following three variables (*father's education, social status, mother's education*) can be summarized into one - the student's social background. Parents with tertiary education lead the child to study at the university. For their child, this is a logical step for getting a job. In addition, parents with higher education have usually better financial background than parents with basic education. For poorer students, the financial situation can be the reason why they prefer to go to work than to the university. The university and government should discuss more intensively about the financial support of these students.

Predictive ability of the forest should be higher than the prediction ability of the decision tree. For this reason, the rate has been established. The prediction ability was approximately the same as for the decision tree: 79.92%.

Model quality can be verified graphically using the ROC curve (Figure 3). Due to a large number of random trees in the random

---

3   The respondent can choose answers: „certainly yes", „rather yes", „rather not" and „certainly not".

forest, the sensitivity and specificity for the most accurate tree will be determined (red point in the Fig. 3 – based on Euclidean distance), which is closest to the upper corner of the ROC curve. **This sensitivity is 0.684 and the specificity is equal to 0.719.**
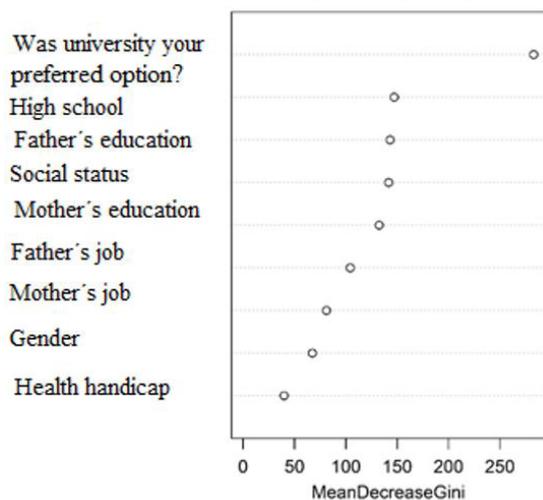

Figure 2: Mean Decrease Gini in a random forest, Eurostudent 2016 (source: own calculation)
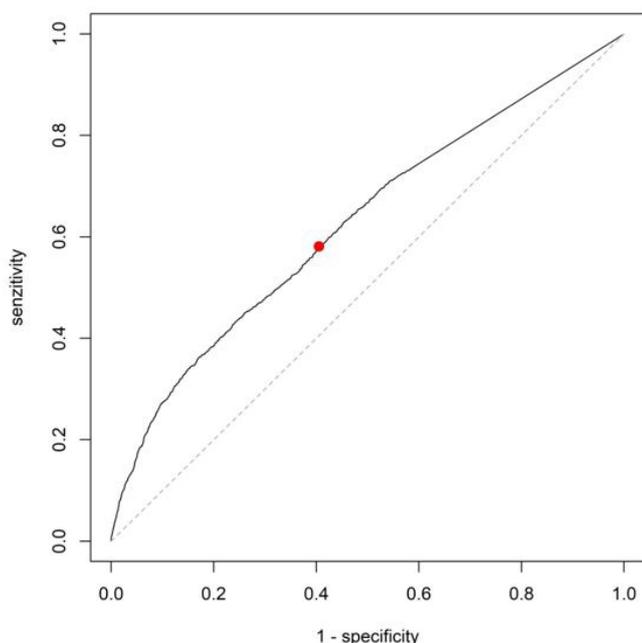

Figure 3: ROC curve in a random forest, Eurostudent 2016 (source: own calculation)

Binary logistic regression

Binary logistic regression was another way to find significant factors. Independent variables have been referenced to the reference category, which for each independent variable was the first category. The hypothesis was tested that there is no move between categories. In case of confirmation, the dependent variable in the model would be meaningless and could be removed. Conversely, the alternative hypothesis confirms its impact.

At the 1% level of significance, significant variables were identified: *gender, health disadvantage*, *education* and *employment of the mother*, *type of high school* and the answer to the question: "*Was university (which you study) your preferred option?".*

Table 2 describes the results from binary logistic regression. In the first column (OR), we can see the ratio of probability which says the chance that student (with some concrete characteristic)

failed in the past in comparison to the reference category of the question. The answer to the question: "*Was university (which you study) your preferred option?"* had the biggest impact on the experience of an unsuccessful study. Students who definitely do not study their preferred college (their answer is "certainly not") have a 6.3 times higher chance of not completing tertiary education in the past than students who certainly study in preferred university (their answer is "certainly yes"). Students who do not attend the preferred college (their answer is "certainly not") do not complete the study successfully in the past 3 times more often than students who have placed their college at the same time in the first place (their answer is "certainly yes").

The variable *Gender* has also the influence. The man has 1.4 times bigger chance that he fails during the studies than women. Disabled students have 1.3 times higher chance to have unsuccessfully completed university studies in the past than student without health complication.

	OR	*p*-value
Mother's education = PhD (reference category = max. elementary school)	1.576	0.000
Gender = Man (reference category = woman)	1.426	0.000
High school = postgraduate graduates secondary vocational schools without graduation (reference category = Secondary vocational secondary school – excluding lyceum)	0.454	0.000
High school = Multi-year gymnasium (reference category = Secondary vocational secondary school – excluding lyceum)	0.830	0.001
Health handicap = Yes (reference category = No)	1.254	0.000
Was university (which you study) your preferred option? = Rather yes (reference category = certainly yes)	1.436	0.000
Was university (which you study) your preferred option? = Rather not (reference category = certainly yes)	3.286	0.000
Was university (which you study) your preferred option? = Certainly not (reference category = certainly yes)	6.307	0.000

Table 2*: *Binary logistic regression, Eurostudent VI (source: own calculation)

Discussion

Existing data sources as the database of Ministry of Education, Youth and Sport allow us to analyse the relationship between surveyed variables. Our results show more than we can find in results of both statistical methods - decision trees CART and binary logistic regression - a subjective response to the question: "*Was the university you are studying your preferred option?*" was a significant variable. Students who definitely do not study (their answer is "certainly not") at their preferred college are 6.3 times more likely to have unsuccessfully completed tertiary education in the past than students who study definitely in their preferred university (their answer is "certainly yes"). Those who are not currently studying their preferred institution (their answer is "rather not") are 3.3 times more likely to have experience with unsuccessful study than those who are definitely studying at their preferred college (their answer is "certainly yes"). It is clear, therefore, that students, after failing to complete their studies, choose the "backup" option and prefer to study afterwards, which they do not indicate as preferred. On the other hand, this result indicates that students after their unsuccessful studying can find another university (or study program) but with much less motivation to study it because this is not his or her preferred choice.

Decision trees, as well as logistic regression, have confirmed that men have a higher degree of failure than women. Men are 1.426 times more likely to have an unsuccessful past tertiary education than women. Wolter, Diem and Messer

(2014) published the same conclusion and as we combine this information with result from the same Wolter's work that men more often study mathematical and technical disciplines and that Pikálkova, Vojtěch and Kleňha (2014) published that in the same study programmes not to have over-pressure in admissions it could be reason why some students (especially males) would underestimate the difficulty of the university studies.

Also, health disadvantage can play a role in whether a student has unsuccessfully completed tertiary education in the past. Those who are at a disadvantage are 1.3 times more likely than students who do not suffer from health complications. It is, therefore, less possible for health-disadvantaged students to hide their strength against other students, but it can also indicate that schools cannot work with the disadvantaged in such a way as to provide them with the necessary conditions, and these students then go to study elsewhere.

It seems as there is the wider definition of the second factor defined by Jensen (2011) – not only university as the institution should form student but also secondary school has to prepare the student for next studies and it should be moderated at a high school in line with this fact. A definitely supportive solution is to raise the awareness of graduates about the conditions of study at universities, compulsory subjects and graduate profiles, which could also help to increase the intensification of the relationship between the students of the high and secondary schools themselves. The greatest degree of learning failure is concentrated in the first year of study. This is referred to as a "deferred choice" - students are poorly informed and when they start studying, they decide whether to stay or not (The Ministry of Education, Youth, and Sports, 2014). One of the reasons why students do not attend their preferred school is that they could study during their previous studies, but for some reason - financial, family, they did not have the study responsibilities - they left school. Questionnaire EUROSTUDENT VI does not answer this question. Secondary schools should better shape the student in his / her expectations due to his / her abilities.

Conclusion

The article has set the objective to analyse the defined segment of unsuccessful students who got into the studies again. The use of the EUROSTUDENT VI data source allowed a deeper but significantly more limited analysis of the reasons and factors of leaving the study in general, which is comprehensively published in the Czech Republic by Vlk et al. (2017).

Policymakers should be able to answer the question whether the fact that students tend more often to study an less solicited field after the unsuccessful study is ok, especially if this likelihood is higher for a group of people with health disabilities.

In addition to existing studies, these analysed data also show that the current system is not optimized and leads to a number of disbalances. It is not a realistic goal for all students to study a preferred field, although, as a theoretical goal this may. That fact should be much more integrated into the decision-making process at high school than at present. Current study programs should be better described with the correct keywords. Candidates should have better information about the study, the study requirements and the subsequent application. It is difficult to select a study program by name. Usually, the program (its name) can be found at more universities but they have different content each time. Still, it may be appropriate to ask why students study non-preferred disciplines, ask and then seek for the answers to how to improve this situation.

Acknowledgement

References

Betinec, M. (2006) *Použití ROC křivek pro hodnocení klasifikátorů,* Prague: JČMF.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees,* New York: Wadsworth Inc.

Fischer, J., Vltavská, K., Brázdilová, M., Lipovská, H., Mazouch, P., Ptáčková V. and Šimková, M. (2016) *EUROSTUDENT VI Základní výsledky šetření postojů a životních podmínek studentů vysokých škol v České republice*, [online], Available at: http://www.msmt.cz/uploads/odbor_30/TF/Analyticke_materialy/Eurostudent/E_VI_zaverecna_zprava.pdf [28 Feb 2018].

Fučík, P. and Slepičková, L. (2014) ‚Studenti, kteří odcházejí: Kvantitativní analýza nedokončených vysokoškolských studií', *AULA*, Vol. 22, No. 1, pp. 24–54.

German Centre for Higher Education Research and Science Studies (2018) *EUROSTUDENT social dymension of European higher education: eurostudent*, [Online], Available: http://www.eurostudent.eu/index_html [14 Feb 2017].

Hauschild, K., Gwosć,C., Netz, N. and Mishra, S. (2015) *Social and Economic Conditions of Student Life in Europe*. Bielefeld: W. Bertelsmann Verlag GmbH & Co. KG.

Heublein, U. (2014) ‚Student Drop-out from German Higher Education Institutions', *European Journal of Education*, Vol. 49, No. 2, pp. 497–513. http://dx.doi.org/10.1111/ejed.12097

Hloušková, L. (2014) ‚Mám základní vzdělání. Příčiny a důsledky předčasných odchodů ze studia a ze vzdělávání', *Studia paedagogica*, Vol. 19, No. 2, pp. 11-38. http://dx.doi.org/10.5817/SP2014-2-2

Hosmer, D. W. and Lemeshow, S. (2000) *Applied logistic regression, 2nd edition*, New York: John Wiley & Sons.

Hraba, J., Hulík, V. and Hulíková Tesárková, K. (2016) *Problematika neúspěšných ukončení vysokoškolského studia (drop-outs) v českém kontextu*, [online], Available at: http://kredo.reformy-msmt.cz/download/w-13-3/KREDO_prezentace_150313_2_Hulikova-Tesarkova-Hulik- Hraba.pdf [28 Feb 2018].

Charvát, M., Viktorová, L., Vobořil, L, Tošenová, M. and Opletalová, V. (2014). *Tvorba, administrace a analýza testů studijních předpokladů*. Olomouc: Palacký University in Olomouc.

Jensen, U. (2011) ‚Factors influencing student retention in higher education', *Kamehameha Schools Research & Evaluation*, Vol. 2, pp.1-4.

Kearney, M. S. and Levine, P. B. (2016) ´Income Inequality, Social Mobility, and the Decision To Drop Out of High School´, *NBER Working Paper*, (Spring 2016), pp. 1-73. http://dx.doi.org/10.3386/w20195

Kingston, E. (2008) ‚Emotional competence and drop-out rates in higher education', *Education + Training*, Vol. 50, No. 2, pp. 128-139. https://doi.org/10.1108/00400910810862119

Klaschka J., and Kotrč, E. (2004) *Klasifikační a regresní lesy*. Prague: JČMF.

Komprdová, K. (2012). *Rozhodovací stromy a lesy*. Brno: akademické nakladatelství CERM.

Konečný, T., Basl, J. and Mysliveček, J.. (2010) ,Přechod mezi střední a vysokou školou a role různých modelů přijímacích řízení', *Czech Sociological Review*, Vol. 46, No. 1, pp. 43–72.

Menclová, M., Pacnerová, H. and Vacek, J. (2008) *Příčiny předčasných odchodů ze středního školního vzdělávání u žáků ze sociokulturně znevýhodněného prostředí, průzkum pro projekt PROPOS*. Prague: Centrum adiktologie a IPPP.

Ministry of Education, Youth and Sports (2014) *Rámec rozvoje vysokého školství do roku 2020*, [online], Available at: http://www.vzdelavani2020.cz/images_obsah/dokumenty/ramec_vs.pdf [26 Feb 2018].

Ministry of Education, Youth and Sports (2017) *Věstník MŠMT*, [online], Available at: http://www.msmt.cz/dokumenty-3/vestnik-msmt-11-12-2017 [20 Feb 2018].

Pikálková, S., Vojtěch, J. and Kleňha, D. (2014) *Úspěšnost absolventů středních škol ve vysokoškolském studiu, předčasné odchody ze vzdělávání*, Prague: Národní ústav pro vzdělávání.

Ptáčková, V., Mazouch, P. and Fischer, J. (2017) *Plýtvá naše společnost lidským kapitálem*?. Prague: University of Economics, Prague.

Rubešová, J. (2009) ,Souvisí úspěšnost studia na vysoké škole se středoškolským prospěchem?', *Pedagogická orientace*, Vol. 19, No. 3, pp. 89–103.

Řeháková, B. (2000) ‹Nebojte se logistické regrese›, *Sociologický časopis/Czech Sociological Review*, Vol. 36, No. 4, pp. 475-492.

Stoessel, K., Ihme, T. A, Barbarino, M., Fisseler, B. and Stürmer, S. (2015) ,Sociodemographic Diversity and Distance Education: Who Drops Out from Academic Programs and Why?', *Research in Higher Education*, Vol. 56, No. 3, pp. 228-246. https://doi.org/10.1007/s11162-014-9343-x

Tinto, V. (1975) ,Dropout from Higher Education: A Theoretical Synthesis of Recent Research', *Review of Educational Research*, Vol. 45, No. 1, pp. 89-125. https://doi.org/10.3102/00346543045001089

Tinto, V. (1997) ,Classrooms as Communities', *The Journal of Higher Education*, Vol. 68, No. 6, pp. 599-623. https://doi.org/10.2307/2959965

Tinto, V. (1999) ,Taking Retention Seriously: Rethinking the First Year of College', *NACADA Journal*, Vol. 19, No. 2, pp. 5–9. https://doi.org/10.12930/0271-9517-19.2.5

Tufféry, S. (2011) *Data mining and statistics for decision making*. United Kingdom: Wiley and Sons Ltd.

Vild, J. (2012) *Analýza CHAID jako nástroj vyhodnocování marketingových kampaní*, Prague: University of Economics, Prague.

Vlk, A., Drbohlav, J., Fliegl, T., Hulík, V., Stiburek, Š. and Švec, V. (2017) *Studijní neúspěšnost na vysokých školách: Teoretická východiska, empirické poznatky a doporučení*, Prague: SLON.

Vnoučková, L., Urbancová, H. and Smolová, H. (2017) ,Factors Describing Students' Perception on Education Quality Standards', *Journal on Efficiency and Responsibility in Education and Science*, Vol. 10, No. 4, pp. 109-115. https://doi.org/10.7160/eriesj.2017.100403

Wolter, S. C., Diem, A. and Messer, D. (2014) ,Drop-outs from Swiss Universities: An empirical analysis of data on all students between 1975 and 2008', *European Journal of Education*, Vol. 49, No. 4, pp. 471–483. https://doi.org/10.1111/ejed.12096