

Factors Affecting Attrition among First Year Computer Science Students: the Case of University of Latvia

Juris Borzovs, Laila Niedrite, Darja Solodovnikova

University of Latvia, Faculty of Computing.

Address: Raina blvd 19, Riga, LV 1586, Latvia

Abstract. The purpose of our study was to identify reasons for high dropout of students enrolled in the first year of the computer science study program to make it possible to determine students, who are potentially in risk. Several factors that could affect attrition, as it was originally assumed, were studied: high school grades (admission score), compensative course in high school mathematics, intermediate grades for core courses, prior knowledge of programming. However, the results of our study indicate that none of the studied factors is determinant to identify those students, who are going to abandon their studies, with great precision. The majority of the studied students drop out in the 1st semester of the 1st year, and the dropout consists mostly of those, who do not really begin studies. Therefore, one of the main conclusions is such that the planned activities of informing about the contents of the program should be carried out, and the perspective students should be offered a possibility to evaluate their potential to study computer science before choosing a study program.

Keywords: Attrition rate, computer science education, data processing, data analysis.

I INTRODUCTION

In recent years, there has been a competition among applicants of the undergraduate programs of the Faculty of Computing at the University of Latvia. Besides, 90% of all entered students have identified the program as their 1st priority. This confirms that there has been a real selection, which identified relatively good and motivated students. However, the considerable number of students that annually abandon studies especially in the first study year (an average of 30% after the first semester, close to 50% over the 1st study year) and comparison with other countries encourages to investigate the causes of attrition.

One of the indicators used to analyze tertiary education, which also includes university level education, is "completion rate". It is defined as the ratio between the number of students in the study program, which obtained a degree in the reference year, and the number of students enrolled in the study program so many years ago, as the length of the full-time study program [1]. The "completion rate" of the university programs from 23 countries participating in OECD study varies, ranging from 93% in Japan and ending with indicator below 60% in several countries, including Sweden (49%, the worst indicator in the study).

According to OECD method of calculation in 2013, using data about the number of students, who obtained a degree in 2013 [2], and the number of students enrolled in 2009 [3], the "completion rate" of the computer science bachelor program of the University of Latvia was 47%. In comparison with the OECD study, the result is below the worst indicator, but it should be noted that OECD indicators characterize education at the national level, moreover, research shows that indicators of the engineering and information technology programs are worse than in other fields. However, this example and the situation in the first year computer science program mentioned at the beginning shows that there is room for improvement. The results of the other countries studies confirm that in different countries there are diverse factors affecting attrition and their combinations, therefore, local studies about causes for attrition in Latvia are justified.

In OECD studies a related indicator "non-completion rate" is also used. The average value of this indicator in all studied countries [1] was 30%. In other reports, the authors of the paper [4] use "dropout" concept, which is stronger, because in the calculation of the "non-completion rate" such students who fall behind the planned study schedule (for example, take a break in studies) are also taken into account, but the dropout examines only those leaving the studies without a degree. In many educational

studies the concept of drop-out is used, which is why exactly dropout will be analyzed in our study.

In terms of regularities, which the researchers worldwide are unanimous about, the dropout percentage varies depending on the academic year, besides the highest dropout is in the first study year and then it decreases [5]. For example, in the USA four-year colleges and universities, 38% of all dropouts are exactly in the first academic year, and 29% - in the second academic year [5]. The dropout rate of the second study year is closely related to the results of the first study year, so the first study year is the most important in terms of impact on further studies, including prevention of drop-outs. Therefore, the 1st study year has been selected for our study. Our study investigates students enrolled into the computer science bachelor and programmer professional programs in one year (2013) and analyzes what have these students been doing over time, including how many of them have dropped out of the University of Latvia and what is the ratio of these study leavers among enrolled students, i.e. „dropout percentage”.

II MATERIALS AND METHODS

The purpose of our research is to analyze a study model of the first year students enrolled in 2013 into the computer science bachelor program of the University of Latvia in order to evaluate, when the “dropout” takes place in the first year, what are the characteristics of “dropping out” students, what are the mutual influence of these characteristics, as well as to create a model to make it possible to identify students, who are potentially in risk.

As data sources, similar studies often use data from university management information systems, using data supplied by a student during enrolment, for example, age, gender, high school grades, etc., as well as study results [6]-[7]. The second data source is the survey [7]-[9], which is conducted beginning studies and leaving them, for example, to assess student's academic, psychosocial and cognitive characteristics. Studies tend to include both the data about all academic years during several years [6] and the data about the first year [8]-[9]. Besides, studies tend to analyze all university programs, or separate faculties, for example information technology programs [7], [10]-[11].

Data used in our study come from multiple data sources. Admissions data are obtained from the management information systems of the University of Latvia. These include personal data and high school exam grades. To be admitted into the Faculty of Computing at the University of Latvia, grades of the centralized exams in math (75% of the admission score) and Latvian language (25% of the admission score) are used in the competition. The maximum total score is 1000 points. The outstanding applicants, who have additional points for participation in the national

math or informatics Olympiads, are considered separately. Since the admissions data are also stored in the data warehouse of the University of Latvia, several previously existing reports were used for the analysis of the distribution of applicants based on the admission score (without integration with other data).

To analyze the students' study process, data about intermediate grades (tests, homework, practical tasks and other grades) for several courses were used. These were the two compulsory courses "Algebra" and "Programming". The first one represents mathematics, the other is a specialty course. Data were available from Excel tables provided by the teachers of the courses.

The study also includes data about the "Compensative Course in High School Mathematics", which is compulsory to all students of the Faculty of Computing, whose admission score is below a certain number of points (700 out of 1000 maximum possible score). These data were obtained from the e-learning environment Moodle grade book data, which stores intermediate grades. In this case, these are data on whether a student has passed 2 tests and how many times he or she has taken the tests.

Data from two surveys were also used in the study. The first survey was carried out in 2013. 235 newly admitted students took part in it beginning their studies. The survey was anonymous, this is why only aggregated results were available. In our study we used aggregated answers to questions about students' prior knowledge in programming. The second of the surveys was carried out in January 2015 during the session. All the students who continued to study (out of those enrolled in 2013, i.e. "survivors") were interviewed. The same question of the first survey about the prior knowledge in programming was asked, but this time respondents were identified. The question contained several possible answers, such as "I studied programming in a school programming course", "I studied programming myself," etc., as well as the opportunity to answer that "I did not study programming before".

In studies of other authors, statistical methods are used the most often [6]-[7], [10]-[12]. Sometimes, data mining [13] and data warehouse [10] are used in studies. The latter is mainly used for the data preparation, integration and supply of data for systematic analysis.

In our study, the infrastructure of the data warehouse of the University of Latvia was used as an environment for many data source integration, since the part of data was already available in it. The data warehouse is implemented using Oracle database and Microstrategy tool for reporting. MS Excel tool was used for data analysis and results visualization for the purpose of this study.

III RESULTS AND DISCUSSION

In our study we examine the first study year. We pay more attention to the first semester, when the dropout percentage is the biggest throughout the whole study process (see Table 1 and Fig. 1).

TABLE 1.
DYNAMICS OF THE NUMBER OF STUDENTS DROPPING OUT

Changes of study status in time	Number
Students enrolled in both programs in total	254
Dropout after the 1 st semester	85
Dropout after the 2 nd semester	39
Total dropout after the 1 st study year	124

The fact of dropping out is detectable only at the end of the semester, rather than in relation to the time dimension (months) as in [11], because in our case, the withdrawal often occurs because the student does not register for the next study semester. Very few students write applications for the withdrawal of their own volition. This is why the possibility to carry out a survey and to determine the true causes of dropout is also limited. In addition to the examination of all students enrolled in 2013, we found out that for the analysis of dropout, it is necessary to distinguish and examine the subset of students, who actually do not start studies (however, they begin studies formally, since they sign the contract), because there are no intermediate results about fulfilled study exercises.

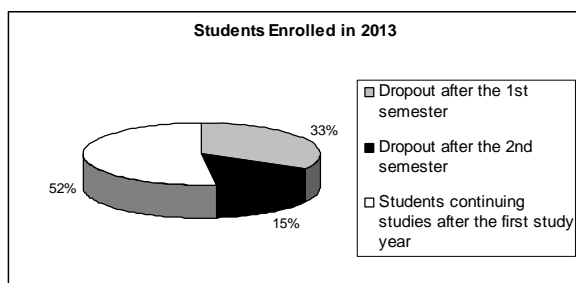


Fig.1. Distribution of enrolled students after the 1st study year

We studied a number of influencing factors, which were originally assumed to have a potential impact on attrition:

- 1) High school grades (admission score)
- 2) Compensative course in high school mathematics
- 3) Intermediate grades for core courses
- 4) Prior knowledge in programming

School Grades (Admission Score) and Dropout

The first factor that was analyzed, was the „high school grades”, which were analyzed by the total score (taking into account both the mathematic and the Latvian language exams), because this score is used in

the entry competition, as well as the mathematics exam was analyzed separately, which could have different results assuming that part of applicants are Latvian non-native speakers. The hypothesis of the study is that because of the weak school knowledge, it might be difficult to study. To analyze the total admission score, it was divided into intervals of 100 points, starting with 500, because the smallest score of 2013 applicants, who beat the competition, was 573.5 points.

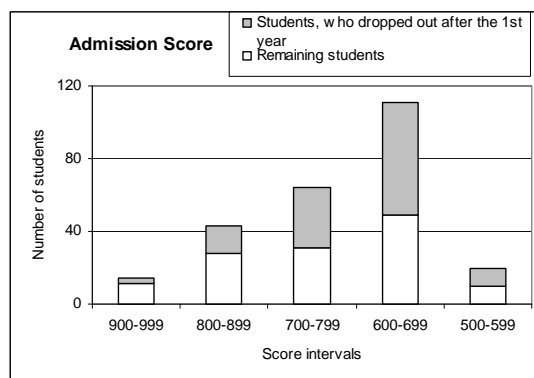


Fig.2. Distribution of remaining students and students, who dropped out, by admission score

The results of the study (see Table 2 and Fig. 2) show that the score at intervals of up to 800 points, which covers 78% of enrolment, the ratio between dropout and "survivors" is close to 50%. Slightly better results are among very good students (22% of the total number of students, score interval 800-1000), but the dropout is also present. Since it is not confirmed that majority of students with worse admission score drop out, but students with better admission score do not, then there is a basis for further research.

TABLE 2.
DYNAMICS OF THE NUMBER OF STUDENTS, WHO DROPPED OUT

Admission Score	Total Remaining Students	Dropout	Total Enrolled	Dropout (%)
900-999	11	3	14	21,43
800-899	28	15	43	34,88
700-799	31	33	64	51,56
600-699	49	62	113	54,87
500-599	10	10	20	50
Total	130	124	254	48,82

It is possible to distinguish and analyze students admitted outside of the competition, who were granted extra points for winning national Olympiads in mathematics or informatics (additional 1000 points for each Olympiad). In the following table (Table 3), it can be seen that none from the 6 outstanding students dropped out.

TABLE 3.
NUMBER AND STATUS OF ENROLLED WINNERS OF OLYMPIADS

Admission Score	Remaining Students	Dropout
2900-2999	1	0
2800-2899	1	0
1900-1999	1	0
1800-1899	3	0

Another hypothesis that was studied was the grades for high school mathematics exam as the dropout indicator.

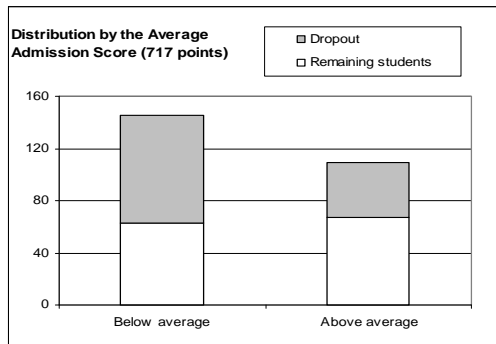


Fig.3. Remaining students and students, who dropped out, by the average admission score

A dropout analysis was carried out, taking into account only math admission examination scores and taking into account the total number of points. In this study, the number of students, who dropped out and the number of remaining students above and below the average number of points were compared: average admission score was 717 (Fig. 3), but taking into account only the math exam, the average score was 566 (Fig. 4). As shown in the charts, the significant changes in the distribution are not observed, therefore, the total score was used for further research, which officially served as the grounds for admission to the program.

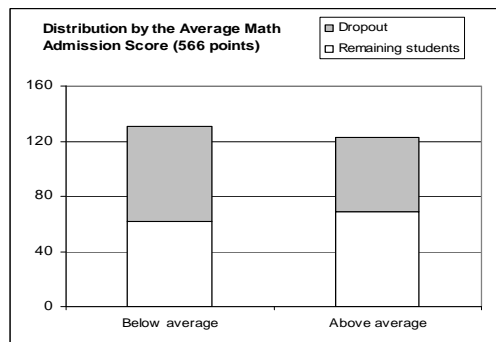


Fig.4. Remaining students and students, who dropped out, by the average math admission score

Compensative Course in High School Mathematics

The 2nd factor that also allows to analyze the impact of the students' school mathematics knowledge on the dropout is the „Compensative Course in High School Mathematics” organized by the Faculty of Computing. Students must participate in the course, learn topics of high school mathematics selected according to the needs of the program, and pass 2 tests. It is allowed to try to pass the tests several times.

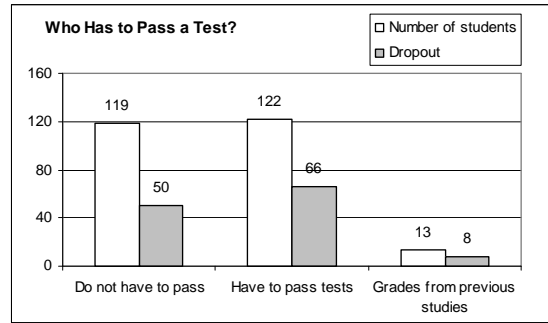


Fig.5. Student distribution in Compensative Course in High School Mathematics

Students, who have more than 700 points, do not pass tests and receive an automatic grade for both tests. So this group shows the same situation observed before, i.e. there is a big dropout (42%) among students with good knowledge of mathematics. The composition of all students and the dropout in comparison with the total number in each group, can be seen in the chart (Fig. 5).

For further research, those students are essential, for whom the effect of the compensative course can be examined in more detail; it is for those, who have to pass the tests (Fig. 6). The dropout rate is higher (54% of the total number), but the history of passing the course tests allows us to find out more information about the students passing the tests.

In this course (as it is also shown further in the paper in relevance to other subject analysis), the group can be distinguished that includes students, who do not do anything (there is no attempt to pass the tests) - 15 students, all of them were subsequently expelled.

Others can be divided into 2 groups. One group includes those students, who learn the course, sooner or later pass the tests (the chart shows only the results and does not reflect, how many times students try to pass the tests). A conclusion about this group can be made that these students meet the necessary requirements for knowledge of high school mathematics, and 30% of these students drop out.

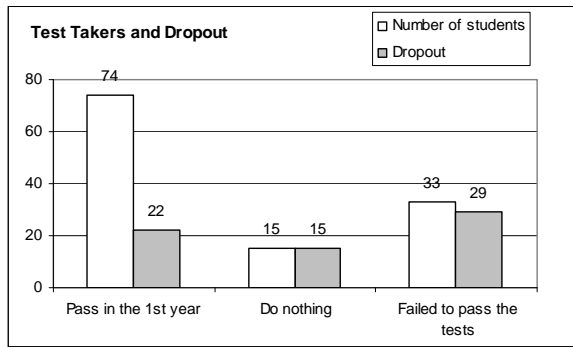


Fig.6. Outcomes of test takers and dropout

However, the group of students, who "failed to pass the tests" on the one hand confirms students' poor knowledge of high school mathematics, but the course also did not help (although, there are no data on whether the course was attended, or students just tried to pass the tests), and over time, students are expelled (at the time of data analysis, only 4 out of 33 students still had „student” status, so there was still a hope that they pass this debt). On the other hand, analyzing the data about repeated test passing by those students who "failed to pass the tests", it can be concluded that only 8 (out of 33) students tried to pass the tests repeatedly. Another ratio is observed among those, who passed the tests - 33 students are such, who passed the test on the first go, the other 41 students tried to pass the tests repeatedly, 6 of them did it more than 2 times. These data indicate that the outcome depends not so much on the school knowledge, but the motivation to work.

Intermediate grades for core courses

The 3rd factor, whose impact on dropout has been studied, was two core courses taught in the 1st semester. One of these two courses was "Algebra" that represented a mathematics course, and the second course was "Programming" that represented a specialty course. Intermediate grades for both courses were evaluated and an obvious group of "do-nothing" students was revealed. The first consecutive test was analysed for these students in both courses, whether there was or was not any grade for it (i.e. whether a student tried to write the test). Analysing other intermediate grades, there was also not more than one (typically, none) submitted assignment of other type (homework, practical work tasks, etc.). There are 45 students, who do nothing in both subjects (in the 1st semester). The number of "do-nothing" students separately for each course is slightly higher (Fig. 7). For example, there are 58 dropped out "do-nothing" students in the 1st semester Algebra course, but if we check whether any of the "do-nothing" group has "survived" (passed the debt or intend to do so later), then there are 2 such students. In Programming course, there are 49 dropped out "do-nothing" students. For comparison, it should be noted that the total dropout in the 1st fall semester is 85 students. It

can be seen that the majority of them are the "do-nothing" students of the Algebra course (68% of the fall semester dropout). A conclusion can be drawn that non-writing the 1st test in Algebra is a sign that the student will drop out, and here we do not mean unsatisfactory grades, but rather non-writing the test at all.

We analysed the above-mentioned 45 students who did nothing in both core courses, and investigated what they did in compensative course in high school mathematics. We found out that 13 students failed to pass the tests (as research showed, these were mostly the students, who tried to pass the tests once and did not make an efforts to re-pass the test), 15 students did nothing in this course as well and only 1 student passed the test, while 16 students got an automatic grade (i.e. admission score was higher than 700). A conclusion can be drawn that there is a group of students that practically do not begin studies, i.e. 18% (45 out of 254 students, who began their studies). To conclude, a total of 33% of students were expelled in the 1st semester, i.e. a little more than a half of these students had not begun their studies at all. Speaking about the reasons for non-beginning studies, quite a large ratio of "do-nothing" students (16 students) among those with an admission score of more than 700 points rather indicate other reasons, which are not related to grades.

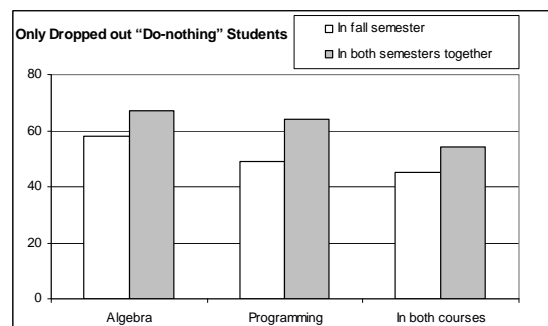


Fig.7. Dropped out "do-nothing" students – comparison by courses

Prior Knowledge in Programming

The 4th factor that was studied, was the impact of no prior knowledge in programming on dropout. Two surveys were carried out. The organization and data collection of them have already been described in the paper. The results of the surveys are given in the Table 4.

22% out of 235 students indicated that they did not have any prior knowledge in programming in the 1st initial survey before the start of the studies, but in the repeated survey of the "survivors" after the first year, 24% out of 114 respondents indicated the same answer. It can be concluded that the initial assumption that students with no prior knowledge in programming are at greater risk to drop out, is invalid, because the

initial ratio among students with and without prior knowledge remained the same.

TABLE 4.

RESULTS OF SURVEYS ABOUT PRIOR KNOWLEDGE IN PROGRAMMING

Questions	1 st survey (%)	2 nd survey (%)
I studied programming in a high school programming course	43	50
I took part in the informatics Olympiad	14	25
I studied at the training centre „Progmeistars”	4	5
I studied at the programming courses	7	8
I participated in the competition „Bebr@s”	4	2
I studied programming myself	49	48
I learned programming at workplace	13	12
I did not have any prior skills	22	24
Other	4	13

IV CONCLUSIONS

Looking at similar studies that analyze exactly engineering or information technology fields of study, since there is usually one of the highest dropout percentage, [10], [12], we have concluded that dropout is mostly affected by high school final grades, student's average grade during studying at a university, as well as selected program.

Engineering study programs were also studied in Latvia, at Latvian University of Agriculture [11] (34.4% drop out rate in the 1st study year), and results are similar that the main causes of dropout are poor knowledge of high school graduates and selected field of study.

The main conclusions of our research are such that there exists a large group of students, who in fact do not begin studies and who are later expelled. Failure to begin studies refers to the number of courses. In this study, among all analysed courses, dropped out students were best characterised by the fact that they did not write the 1st test in Algebra.

The majority of expelled students drop out in the 1st semester of the 1st year. These students include both students with weaker and with quite good high school grades. The exceptions are the applicants with excellent grades, where the situation differs with a smaller dropout. Since the 1st year dropout consists mostly of students, who have not really begun studies, so the high school grades do not have significant effect on dropout, but they may have an impact on the further study process.

The hypothesis that a programming background is an important factor influencing dropout was also rejected, since we found out that the ratio between

students with and without prior knowledge remained the same at the start and end of the first study year.

The research results show that universities should create an action plan to reduce dropout [10], but the University policy should not jeopardize the quality of studies (reduced study [11] or admission [8] requirements lower the level of graduates).

Student characteristics (profile) are different, but these differences are not taken into account in the university requirements (environment, process, etc.) ([10], for example, speak about the group of students above 25 years of age, who have more difficulties to adapt to the university requirements, which leads to dropout). It is necessary to carry out adapted activities for „early dropped out students” and „late dropped out students” [12]. For example, further data analysis is required for late dropped out students, to search for individual solutions promptly, but for early dropped out students, additional information should be provided to help adapt to the study environment.

Definitely, a sufficient information about the provided programs should be offered to potential students before they begin studies [6], [8], and qualified advice about the choice of courses should be given after the start, to reduce the dropout of students, for whom "the course was not such as they were expecting" [8].

The further research should focus on the analysis of grades for other courses in the 1st and later study years, to assess the impact of the outcomes of the 1st year on the further study process and dropout (academic debts, change of financing because of academic debts, etc.).

Since our study concluded that a majority of students really do not begin studies, the activities targeting both “early dropped out students” and candidates are considered. In relation to this issue, several ideas could be further studied, for example, the possibility for students to assess themselves their suitability for a chosen field of study, by offering a variety of self-assessment tests - personality tests, logic tests, mathematics tests. For example, the contents of the latter test could be based on the compensative course in high school mathematics, which is already conducted at the Faculty of Computing. It could be implemented as a computerized test for self-assessment that candidate students could take before they choose the study program.

Similarly, it is planned to help get more insight into "what is programming", because, although the study showed that a lack of prior knowledge in programming did not directly cause a dropout, there is still a large number of students, who do not really begin studies, which could be explained by the wrong choice of a study program. Therefore, it is necessary to offer a way to help get an idea about the profession by means of an interactive test. The groundwork in this direction has been started in the project “School

of Young Programmers” of the Faculty of Computing at the University of Latvia [14]. The purpose of this project is to promote the comprehension about programming in high schools.

A more detailed analysis of the planned activities, as well as a comparison with studies of authors from other countries and proposals to reduce the dropout rate is described in a separate paper and the activities are planned to implement in the near future.

V REFERENCES

- [1] OECD, “How many students drop out of tertiary education?”, in Highlights from Education at a Glance 2010, OECD Publishing, 2010.
- [2] Overview about higher education in Latvia, Available: <http://izm.izm.gov.lv/registri-statistika/statistika-augstaka/2013.html>, [Accessed: Mar.17., 2015].
- [3] Overview about higher education in Latvia, Available: http://izm.izm.gov.lv/upload_file/Registri_statistika/Pielikums_2009.xls, [Accessed: Mar.17., 2015]
- [4] OECD, “Education at a Glance 2004: OECD Indicators”, OECD Publishing, 2004.
- [5] V. Tinto., “Completing college: Rethinking institutional action”, University of Chicago Press, 2012.
- [6] F. Belloc, A. Maruotti, and L. Petrella, “University drop-out: an Italian experience”, Higher Education, vol. 60(2), pp. 127-138, 2010.
- [7] K. McKenzie and R. Schweitzer, “Who Succeeds at University? Factors predicting academic performance in first year Australian university students”, Higher Education Research & Development, vol. 20(1), pp. 21-33, 2001.
- [8] L. Grebennikov and M. Shah, “Investigating attrition trends in order to improve student retention”, Quality Assurance in Education, Vol. 20 (3), pp. 223 – 236, 2012.
- [9] A. L. Reynolds and M. J. Weigand, “The relationships among academic attitudes, psychological attitudes, and the first-semester academic achievement of first-year college students”, Journal of Student Affairs Research and Practice, vol.47 (2), pp. 175–195, 2010.
- [10] F. Araque, C. Roldán, and A. Salguero, “Factors influencing university drop out rates”, Computers and Education, vol. 53(3), pp. 563-574, 2009.
- [11] L. Paura and I. Arhipova, “Cause Analysis of Students’ Dropout Rate in Higher Education Study Program”, Procedia-Social and Behavioral Sciences, vol. 109, pp. 1282-1286, 2014.
- [12] N. Gury, “Dropping out of higher education in France: a micro-economic approach using survival analysis”, Education Economics, vol.19(1), pp. 51-64, 2011.
- [13] C. Marquez-Vera, C. Romero, and S. Ventura, “Predicting School Failure Using Data Mining”, In Educational Data Mining, pp. 271-276, 2011.
- [14] “School of Young Programmers” <http://www.df.lu.lv/nacstudet/jauno-datoriku-skola/>, [Accessed: Mar.17., 2015].