From Data to Intervention: Predicting Students At-Risk in a Higher Education Institution

Irene-Angelica Chounta, Kaire Uiboleht, Kersti Roosimäe, Margus Pedaste, Aune Valk
University of Tartu, Estonia
{chounta, kaire.uiboleht, kersti.roosimae, margus.pedaste, aune.valk}@ut.ee

ABSTRACT: In this paper, we present a computational approach to assess study success in a Higher Education academic institution. To that end, we employ data-mining and machine-learning methods to identify factors that may contribute to students' decision to drop out from their studies and to assess the risk of dropping out for each individual student. In order to communicate the results of the risk assessment, we employ an institutional dashboard – that is, a dashboard that presents the risk assessment per student and the reasons behind this assessment. The institutional dashboard aims to inform academic stakeholders, namely program directors and specialists in academic affairs about reasons that may contribute to dropouts in their programs and to help them identify students that may need further support.

Keywords: Learning Analytics, predictive modelling, dropout, students at risk, learning dashboards.

1 INTRODUCTION

In this paper, we present a research initiative at the University of Tartu in Estonia that aims to employ an evidence-based approach to identify students who may be at risk of dropping out from their studies. As dropouts, we define students' exmatriculations from the respective program for reasons that may reveal students' unwillingness to continue their studies, low academic achievement, lack of motivation or lack of interest. In order to achieve this, we propose a computational approach for assessing students' dropout risk using students' data as recorded by the study information system of the academic institution. The goal is to communicate the results of the risk-assessment through institutional dashboards to academic stakeholders (such as curriculum developers and program directors) so that they can identify bottlenecks in their programs and to provide appropriate feedback and support to students, if needed, in a timely manner.

Securing study success in Higher Education (that is, successful completion of studies leading to an academic degree) is among the goals leading the Europe 2020 strategic agenda[1]. Europe aims to scaffold innovation, productivity and also to support social justice by fostering high-level skills through Higher Education. To do that, one of the goals is to increase the rate of young, higher-education graduates by reducing the dropout rates in Higher Education. Estonia has established a number of policies to achieve this goal. However, according to an Annual Report from the Estonian Ministry of Education, the dropout rates for Bachelor students were approximately 51% in 2016[2] across all

[1] http://publications.europa.eu/resource/cellar/d9de3b17-0dcf-11e6-ba9a-01aa75ed71a1.0001.01/DOC_1

[2] https://www.hm.ee/sites/default/files/annual_analyses_2016_1.docx

disciplines. This finding is supported by related studies showing that dropout rates in Estonian Higher Education Institutions can come up to two thirds depending on the field of study (Kori & Mardob, 2017).

The University of Tartu (UT)[3] is Estonia's oldest university and leading centre of research and training. It consists of four faculties: the Faculty of Arts and Humanities, the Faculty of Medicine, the Faculty of Social Sciences and the Faculty of Science and Technology. In 2019 overall, 13400 students – out of which 1660 are international students – study in UT either in the bachelor, master, or PhD programs. In this context, UT launched an initiative in 2019 aiming to support students of mainly Bachelor and Master levels to successfully completing their studies but also to help other academic stakeholders (in this case, program directors and specialists in study affairs) to identify potential reasons that may contribute to dropouts in their programs or curricula and to provide appropriate feedback and support to students.

2 RELATED WORK

Detecting students at risk of dropping out of their studies is a prominent topic of research since dropout rates have a strong impact on the individual (student), the institutional (academic institution) and the national (country) level. Many frameworks have been proposed to evaluate academic success and to identify factors that influence it. For example, Tinto proposed a theoretical model of students' dropouts from college that built on work from social psychology and economics of education (Tinto, 1975, 2017). Tinto's model identifies two dimensions in the model as fundamental for academic success: student's characteristics (such as family background and individual attributes; e.g. goals to study in college) and student's experience with the academic system (such as performance and interactions with teachers and peers). Tinto specified that students need both academic and social integration to ensure retention in studies. According to the model, academic success is affected by the student's individual commitment to their goal along with the student's commitment to the academic system itself.

Arnold and Pistilli (Arnold & Pistilli, 2012) proposed a student success system – Course Signals – in order to support faculty members of a Higher Education Institution (Purdue University) in providing meaningful feedback to students. The system used machine learning algorithms and data mining to predict students who may be at risk of dropping out their studies. The system used data about students' earned credits, student's effort in terms of interaction with the learning environment, students' performance in earlier studies - for example, high school Grade Point Average (GPA) or performance in standardized tests – and other information, such as demographics. Also, Barber and Sharkey (Barber & Sharkey, 2012) proposed the use of predictive models identifying students at risk in the University of Phoenix. In this case, the model combined data from the learning management system, the financial aid system, and the study information system to assess the risk of any given student failing at the course level. Earlier research in Estonian HEIs is based mainly on self-report surveys among dropouts (Kori et al., 2016; Must et al., 2015). The studies show that student dropout is often related to the combination of reasons that include individual and curriculum-level factors: for example, dissatisfaction with the quality or organization of studies, inefficient academic and social

[3] https://www.ut.ee/en/university

environment, wrong choice of studies, inefficient study skills and low motivation, working during the studies and financial reasons. In this work, we use the findings of related research in context, to predict students' dropout and to inform our research with lessons learned and successful practices from similar studies.

3 METHODOLOGY

The overarching goal of this research is to provide a holistic assessment of students' performance (Chounta et al., 2019) in three ways:

- by using various kinds of data, for example information on the course level from the learning management system the university uses and also from the students' feedback questionnaires;
- by applying multilevel analytical approaches – for example social network analysis and pattern mining – to analyze various data sources and to complement insights; and
- by supporting stakeholders through learning analytics dashboards that will present multimodal feedback, for example textual feedback along with visualizations.

Currently, in order to assess risk of dropping out, we use students' data as recorded in the study information system of the academic institution (University of Tartu). Our dataset includes information about students' demographics, their prior academic background and their progress while studying at the institution. After consulting with the university's academic commission about potential issues regarding privacy and ethics, we decided to exclude demographical information – such as gender or citizenship – or potentially private or sensitive information – such as postal address – when assessing whether a student is likely to drop out or not. To take into account differences between student populations that can be attributed to the curricula or the faculties, we modelled these factors as random effects. For the purpose of this work, we employed a computational model that predicts risk on three dimensions:

a) *academic background*. That is, information that may relate to student's previous academic experience, such as: admission grade, number of degrees that the student has acquired and how many times a student has been enrolled in the university's study programs;
b) *effort in terms of participation*. To assess effort, we used the following features: the amount of registered courses and credits, the amount of credits the student cancelled, the amount of credits registered for extra-curricular courses, the time a student spent on academic leave, the time a student spent studying abroad and the student's workload (full or part time);
c) *performance in terms of academic achievement*. To assess performance, we used the following features: the number of successfully completed courses, the number of failed courses, the number of no-showups in exams, the amount of earned credits and the differentiated scores (for example, amount of A's, number of B's, and so on).

For each of these three dimensions, the computational model – in this case a logistic regression classifier – provides a binary assessment, that is whether the student is likely to dropout or not. We decide on the "severity" of the risk assessment based on the following rule:

- Students who are predicted to drop out on three dimensions are classified as "high-risk";
- Students who are predicted to drop out on at least one dimension are classified as "medium-risk";
- Students who are not predicted to drop out on any dimension, are classified as "low-risk".

The results of this assessment will be presented to program directors through an institutional dashboard. This process is described in Figure 1. With the term "institutional dashboard", we mean an online interactive and dynamic interface that will be accessible through the study information system of the institution to program directors. The rationale is to inform them so that they can assess potential risks for their respective program, to help them redesign their program if necessary and to support them in identifying specific cases where an intervention might be needed. Neither teachers nor students will have access to the information presented in the institutional dashboard. The reason is that we do not want to create or support any bias either on the student or the teacher level and to potentially affect student's motivation negatively.

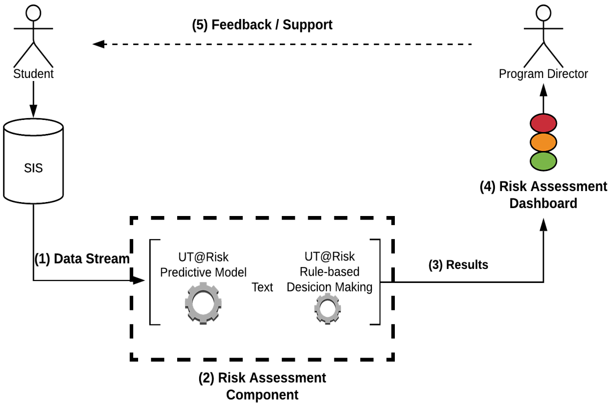


Figure 1. Risk Assessment process for detecting students at-risk of dropping out.

4 FIRST INSIGHTS

In order to test our approach, we collected data from bachelor students and students in Bachelor's and Masters integrated programme (in Medical Faculty) who enrolled in the university from 2010 to 2014. The rationale was to use data of students whose nominal time of studies (3 years in case of Bachelor programme and 6 years in case of integrated studies) is as a rule over and most of whom should have had the opportunity to graduate. Overall, the dataset contained 3695 students from all four faculties. The distribution and dropout rates of students among the four faculties of the university are presented in Table 1.

Table 1. Number of students and dropouts per faculty

Faculty	Number of Students	Number of Dropouts
Arts and Humanities	599	296 (49%)
Medicine	786	162 (21%)
Social Sciences	1443	582 (40%)
Science and Technology	853	429 (50%)

To train and test the model, we split the dataset into two parts: the training and the test sets. For the training set, we used data of students who were matriculated from 2010 to 2013. For the test set, we used data of students who were matriculated in 2014. This resulted in a 70/30 split: 70% of the original dataset was used for training the model and the remaining 30% was used for testing the model. This decision was made in order to test whether using old data to predict dropouts for recent cases could provide accurate predictions. However, we acknowledge that this can have a negative impact on the accuracy of predictions, especially if the dropout rates have changed significantly over the years. We plan to explore the effect of changes in dropout rates on predictive accuracy in future work.

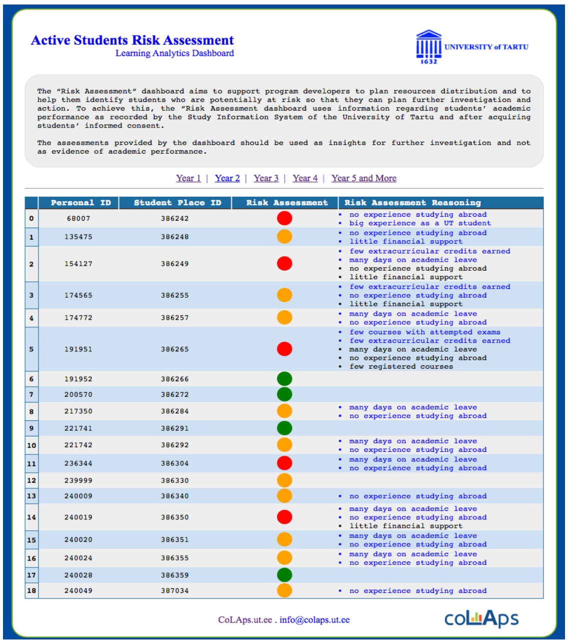
Table 2. Classification metrics for predictions based on the three independent classifiers and on their combination.

	Performance Classifier	Effort Classifier	Academic Background Classifier	Perf + Eff Classifier	Combined Prediction (RAC)
Recall	0.95	0.90	0.40	0.95	0.97
Precision	0.93	0.94	0.55	0.95	0.95
Accuracy	0.95	0.93	0.60	0.96	0.96
F-measure	0.94	0.92	0.46	0.95	0.96

We tested the performance of the Risk Assessment Component (RAC) on the test set. Overall, the test set consisted of 1248 students out of whom 544 students had dropped out from their studies. The RAC assessed that 514 students were on a high risk of dropping out their studies, 217 students were evaluated as medium-risk of dropping out while 517 students were assessed as low-risk. Out of the 514 students that were predicted as high-risk, 488 students indeed dropped out (96%). Similarly, out of the 217 students who were predicted as medium-risk, 39 students dropped out eventually (20%). Finally, out of the 517 students who were assessed as low-risk, only 17 of them eventually dropped out (3.3%). Table 2 shows the results per independent classifier and for their combinations. The results of the Performance and the Effort Classifier are highly correlated (ρ=0.92, p<0.001) while the correlations between these classifiers and the Academic Background Classifier are low (ρ<0.2, p<0.001). Nonetheless, including the Academic Background dimension in the classification process appears to provide the best results in terms of precision, recall and accuracy.

5 DISCUSSION

To communicate the risk assessments to the stakeholders, we designed an institutional dashboard following the traffic lights metaphor (Figure 1). That is, students who were assessed as high-risk, were followed by a red traffic light, students who were assessed as medium-risk were followed by a yellow traffic light, and students who were assessed as low-risk, were followed by a green traffic light. This design was presented as a mock-up to approximately 30 program directors from all faculties during one of their regular meetings and it was well-received. Additionally, the program directors indicated that they would like to receive information about the reasoning behind the model's predictions. That is, why the model predicted that a student belongs to a specific risk group. They also commented that it is important for them to receive this assessment in a timely manner – that is, in the beginning of the new semester, so that they have enough time to intervene, if needed.



Active Students Risk Assessment
Learning Analytics Dashboard

The "Risk Assessment" dashboard aims to support program developers to plan resources distribution and to help them identify students who are potentially at risk so that they can plan further investigation and action. To achieve this, the "Risk Assessment dashboard uses information regarding students' academic performance as recorded by the Study Information System of the University of Tartu and after acquiring students' informed consent.

The assessments provided by the dashboard should be used as insights for further investigation and not as evidence of academic performance.

Year 1 | Year 2 | Year 3 | Year 4 | Year 5 and More

	Personal ID	Student Place ID	Risk Assessment	Risk Assessment Reasoning
0	68007	386242	🔴	• no experience studying abroad • big experience as a UT student
1	135475	386248	🟠	• no experience studying abroad • little financial support
2	154127	386249	🔴	• few extracurricular credits earned • many days on academic leave • no experience studying abroad • little financial support
3	174565	386255	🟠	• few extracurricular credits earned • no experience studying abroad • little financial support
4	174772	386257	🟠	• many days on academic leave • no experience studying abroad
5	191951	386265	🔴	• few courses with attempted exams • few extracurricular credits earned • many days on academic leave • no experience studying abroad • few registered courses
6	191952	386266	🟢	
7	200570	386272	🟢	
8	217350	386284	🟠	• many days on academic leave • no experience studying abroad
9	221741	386291	🟢	
10	221742	386292	🟠	• many days on academic leave • no experience studying abroad
11	236344	386304	🔴	• many days on academic leave • no experience studying abroad
12	239999	386330	🟠	
13	240009	386340	🟠	• no experience studying abroad
14	240019	386350	🔴	• many days on academic leave • no experience studying abroad • little financial support
15	240020	386351	🟠	• many days on academic leave • no experience studying abroad
16	240024	386355	🟠	• many days on academic leave • no experience studying abroad
17	240028	386359	🟢	
18	240049	387034	🟠	• no experience studying abroad

CoLAps.ut.ee . info@colaps.ut.ee

Figure 2. Institutional Dashboard Mockup

Currently, we are re-designing the institutional dashboard and the predictive approach taking into consideration stakeholders' feedback and adding functionality, such as historical data about dropouts in the respective curriculum and current trends. Additionally, we plan to carry out extensive design workshops and pilots with the participation of program directors and curriculum developers as well as stakeholders from the university's government and administration. In this way we want to ensure that the institutional dashboard reflects existing needs and standards of the academic community and that the dashboard's interface is usable and useful for the target user population.

This paper presents work in progress and we acknowledge that significant improvements will be required until we reach the state of launching a viable solution. At the same time, we are aware of existing limitations. As aforementioned, program directors requested timely assessments – the earlier, the better. This is a challenging task especially for first-year students since we rely mostly on metrics of academic effort and performance and we do not take into account students' demographics. One potential solution would be to use student-entered data about their goals and expectations regarding the institution and their motivation for pursuing an academic degree. Another limitation is the way risk assessments should be used. At this point, we do not plan to use this information in any other way rather than for reflecting on our practices and policies. For example, to reflect on what measures could the university take to support students in pursuing their degree. Even though some stakeholders voiced their willingness to intervene with their own means in critical cases, this bears the questions: what would be an appropriate intervention taking into account that students in Higher Education are adults and what would be the cost of it?

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270.

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 259–262.

Chounta, I.-A., Pedaste, M., & Saks, K. (2019). Behind the Scenes: Designing a Learning Analytics Platform for Higher Education. *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19)*. International Conference on Learning Analytics & Knowledge.

Kori, K., & Mardob, K. (2017). *First-year higher education ICT studies and dropout rates–the Estonian case*.

Kori, K., Pedaste, M., Altin, H., Tõnisson, E., & Palts, T. (2016). Factors That Influence Students' Motivation to Start and to Continue Studying Information Technology in Estonia. *IEEE Transactions on Education*, *59*(4), 255–262.

Must, O., Must, A., & Täht, K. (2015). Programmi TULE uuringu „Haridustee vali kud ning õpingute katkestamise asjaolud Eesti kõrghariduses "aruanne. *Psühholoogia Instituut. Tartu Ülikool*.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89–125.

Tinto, V. (2017). Through the eyes of students. *Journal of College Student Retention: Research, Theory & Practice*, *19*(3), 254–269.