DiSEA: Analysing Success and Dropout in Online-Degrees

Monique Janneck
Technische Hochschule Lübeck
monique.janneck@th-luebeck.de

Agathe Merceron, Petra Sauer
Beuth University of Applied Sciences Berlin
{merceron, sauer}@beuth-hochschule.de

ABSTRACT: Although several research works show that students at risk of dropping out of a course or a study program can be predicted with relatively high accuracy, this information has so far often not been accessible to course directors, teachers, or students. The DiSEA project aims to research this issue and close this gap in the context of online-degrees. Building on previous research, machine learning methods will be used to identify risk and success factors. The overall aim is to develop an integrated model to predict success in digital study programs and derive recommendations and interventions for course design, student counseling, and student self-reflection. A user-centered design involving all stakeholders will be followed.

Keywords: Online-degree, dropout, data analysis, dashboard, user-centered design

1 INTRODUCTION

High dropout rates are a problem for students as well as for universities. Young people experience a failure or even a serious cut in their career path. Economically, this represents a waste of educational resources. Relatively high rates of dropouts are leading universities to take innovative measures to actively address this problem. These include improved analytical studies of existing data.

For this purpose, models are developed, trained, and evaluated using machine learning methods (Aulck et al. 2019). The available data play an important role in the quality of the model (Schneider et al. 2019). These models can be used in early warning systems to identify students at risk earlier and provide more targeted advice.

One example of this is the early warning system *FragSte* (Berens et al. 2019, Schneider et al. 2019). *FragSte* only uses data that every university in Germany collects anyway. *FragSte* was evaluated at a state university and a private university of applied sciences. It was shown that the performance of this early warning system improves considerably as soon as data on academic performance can be accessed. Both FragSte and other (international) research projects (for example in Dekker et al.2009, Aulck et al. 2019) examine data from traditional study formats, i.e.

face-to-face study programs with predominantly "classic" first-year students who start their studies in early adulthood or after leaving school.

In addition to traditional study formats, which still make up the majority of study offers, digital study offers are becoming increasingly important and will probably come into even greater focus as a result of the Corona pandemic. Digital study programs take place entirely or predominantly as online or distance learning offerings. Presently, in terms of the characteristics of students, they differ significantly from students in traditional study programs: The target group of digital study formats are typically people who are already full-time in professional life and would like to gain new or further qualifications or who, due to their personal life situation - e.g. bringing up children, caring for relatives, illnesses - would find it difficult to realize a study program in face-to-face teaching and appreciate the flexibility of online offers. Accordingly, these people also show different study behavior: They often study part-time, which goes hand in hand with a - planned - significantly longer duration of studies and often also lower prioritization of studies. The demands on students' self-regulation and motivation skills are significantly higher: although online formats offer a lot of flexibility to adapt studies to personal life situations, they also require a high degree of discipline and organizational skills (Minks et al. 2011). This means that the reasons for dropping out of modules or studies in digital study programs might also differ from traditional study formats. For working people, for example, the semester is often more difficult to plan; unforeseen work-related demands can prevent a module from being successfully completed. In some cases, students underestimate the amount of work required at the beginning of their studies, or the compatibility of studies and work turns out to be more difficult than previously assumed. In some cases, the obtention of a formal degree can become less important in the course of studies - for example, if the acquisition of qualifications can also be proven by intermediate certificates or employers already reward the competencies acquired during studies through better pay or promotion, even if a formal degree is still not completed. The higher demands for independence and self-organization are also not successfully mastered by all students. All these aspects mean that dropout rates are generally higher in online formats (Diaz 2002, Beard & Harper 2002, Baker et al. 2015).

Up to now, research on academic success and dropout has focused primarily on traditional study formats. Digital study formats are only marginally addressed. For this reason, it is questionable whether and how these results can be generalized for digital study formats.

The paper addresses these challenges and presents as research in progress the DiSEA project. DiSEA will focus on analysing and identifying factors for success/failure and dropout, especially in digital study formats, investigating the transferability of previous research results to digital study formats.

The paper is organized as follows. Section 2 reviews related literature. Section 3 describes the goals and research questions of the DiSEA project. In Section 4, we focus on the challenges with the involvement of different types of stakeholders and show how a Human-Centered-Design approach can meet them. Section 5 concludes the paper.

2 RELATED WORKS

Different kinds of data have been used to predict success and dropout in university programs. The first kind of data comes from the registrar's office. This data quite often includes demographic features as well as the academic performance of students, such as enrollments and marks. Different studies show that demographic data or data collected before enrolment are less meaningful for predicting academic success than data on academic performance. Such a finding is reported in (Aulck et al. 2019) that analyzed data from the academic administration of an American university of more than 66,000 students who began their studies between 1998 and 2010. Berens et al. (2019) analyzed data from two universities in Germany, one with 23,000 students and 90 different undergraduate programs and one with about 6,700 students and 26 undergraduate programs, and obtained the same result. This conclusion was also reached earlier by Dekker et al. (2009), although in this study only data from the first year of study of 648 students from a single degree program were considered.

Studies differ with about the machine learning models they use to predict dropout. For example, Dekker et al. (2009) achieved the best results with decision trees, Aulck et al. (2019) with logistic regression, and Berens et al. (2019) with the ensemble method AdaBoost.

Berens et al. (2019) and Aulck et al. (2019) use models that can be applied across universities. Therefore, academic performance is described with so-called "global features" that are not specific to a study program (such as the number of courses passed, average grade, or the number of courses taken). Manrique et al. (2019) compared the performance of models with "global features" and "local features" (i.e. program-specific performance such as grades in certain courses) and were able to show that better prediction results can be achieved with local attributes than with global attributes. The findings are inconsistent concerning the number of semesters considered: Berens et al. (2019) achieve better results the more semesters they consider for the prediction, while the study by Manrique et al. (2019) showed the opposite picture. One possible explanation is that the less good results of Manrique et al. (2019) are due to the small amount of data, as there are fewer dropout data in higher semesters.

In addition to studies that look at academic success globally, there are attempts to develop early warning systems for individual courses. In these attempts, the data is quite often the interactions stored by the learning management system, such as frequency of use, completion of assignments and their assessment, or by some specific learning software. “Course Signals", for example, aims to predict which students are at risk of failing a course (Arnold & Pistilli 2012) in a classical context, while Baneres et al. (2020) also examine specific risk factors in individual courses in an online-context and report high accuracy in predicting completion a course. Baker et al. (2015) and Kuzilek et al. (2015) were able to show for individual online courses that students who engage with the online materials early and regularly in the course and complete corresponding course assignments are more likely to pass the course successfully. Van Goidsenhoven et al. (2020) show that it is possible to achieve accurate predictive models of student success based on log data from an online course and that the models provide reasons for student success. Akçapınar et al. (2019) were also able to show for an e-book-based course that students who achieved better course outcomes had interacted more frequently and intensively with the online materials. These findings suggest that it is promising to look at data from learning management systems in digital study formats.

What form a subsequent intervention should take is not easy to decide. One early warning system described by Jayaprakash et al. (2014) is used in such a way that teachers contact students identified as being at risk. However, in addition to an improvement in student outcomes, this paper also reports a significant increase in course dropouts, which was not intended. Note, that this study took place in a classical context of a college education. By contrast, Baneres et al. (2020) report a slight reduction of dropout in the context of online-degrees: dashboards for students and teachers have been developed to warn students (and their teachers) who are at risk of failing a course. An experiment shows that dropout was slightly reduced for students who consented to the experiment, though it is not clear whether the warning-system itself is the reason for this reduction. Several dashboards have been designed and integrated into the learning management system or virtual learning environment to support students' reflection on their learning. Presently, there are not many studies measuring their usage and their impact. In a pilot study, de Quincy et al. (2019) reported that about 25% of the students view the dashboard weekly. More research to understand their usage, impact, and usefulness is needed.

Following the works of (Aulck et al. 2019) and (Berens et al. 2019), we intend to use primary data on academic performance to predict whether students are at risk of dropping out of their degree. We will investigate whether local or global features work the best in our context. However, it is not clear whether the features considered in the literature can be overtaken as is in our context where most of the students study part-time. Therefore, feature engineering will be investigated further also to provide understandable explanations of the prediction to students. Predicting "dropping out of the degree" will be complemented by "predicting dropping out of the course" making use of the data stored by the learning management system as done in (Baneres et al. 2020). It is an innovative aspect of this project to investigate how both predictions can be combined and conveyed constructively to students.

3 GOALS AND RESEARCH QUESTIONS

The DiSEA project aims to identify risk factors for dropout as well as success factors for online study programs. On the one hand, the transferability of previous research results to digital study formats will be investigated. On the other hand, analyses of data on learning behaviour will be combined with analyses of academic data.

For this purpose, the extensive experience and data from the university network "Virtual University of Applied Sciences", or VFH (Virtuelle Fachhochschule, https://www.vfh.de/), will be used. The VFH network was founded in 2001 as part of an extensive German research programme. Currently, 13 higher education institutions from several federal states and one from Switzerland belong to the network; students come from all over Germany, some from abroad. The VFH currently offers 12 joint accredited Bachelor's and Master's degree programmes as online degree programmes.

The VFH study programs rely on a common learning management system (Moodle), thus providing extensive user data.

Our project addresses the following main research questions:

1. *Generalizability of prior findings on dropout risk factors to digital study formats*: We will analyze if and to what extent current results and prediction models from traditional face-to-face study programs can be applied to online degrees. Furthermore, we will investigate which specific factors need to be taken into account to predict success and failure in digital study formats.

2. *Analysis of data on learner behavior*. As prior research shows, models predicting study success are relatively weak in the first semesters, when data on academic achievements is scarce. However, the introductory phase can be decisive for later success. In digital study formats, data from learning management systems (LMS) is available from the onset, providing insights into learning behavior (e.g. frequency and intervals of use). Therefore, we will analyze Moodle user data to enhance models predicting study success.

3. *Using dashboards to enhance learners' self-reflection.* In online study programs, students' self-regulatory capacities are crucial for success, as students need to structure and organize their learning activities themselves to a much higher extent than students in traditional face-to-face programs. In this regard, it is essential that students receive feedback on their learning activities to recognize and reflect problematic habits and the need for change. Learning management systems may use so-called dashboards to visualize user data, learning activities, deadlines, and assignments, etc. (e.g. Brandenburger et al. 2019, Constapel et al. 2019). In our project, we will develop a dashboard providing learning and study-related data to enhance students' self-regulatory competencies. These dashboards will be evaluated in selected courses to analyze their impact on learning behavior and success as well as user acceptance. Providing learning analytics dashboards for students is a rather new direction in learning analytics research, as teachers have long been the predominant target group (cf. Schwendimann et al. 2016).

The overall aim is to develop an integrated model to predict success in digital study programs. We expect that from these findings, we can derive additional recommendations e.g. for course design and student counseling.

4 CHALLENGES REGARDING THE INVOLVEMENT OF STAKEHOLDERS

To successfully conduct our research activities and enable practical changes various stakeholders need to be involved, first of all, students and lecturers, but also program managers, heads of department, student counselors, data security officers, etc. Especially students' active involvement in designing research activities is crucial. Prior studies show that students accept data analyses if they are well informed and convinced of potential benefits (cf. Ifenthaler & Schumacher 2016, Slade et al. 2019).

In our project, we will design our research and development activities in a *Human-Centered-Design* approach. Essential stakeholders will be included in all phases of the project:

- In the *requirements analysis* phase we will conduct workshops and interviews to include stakeholders' views. As pointed out in Martinez-Maldonado et al. (2016), a challenge of this first stage is the identification of "possible new and radical features that can be offered by the data to address stakeholder needs, but where the stakeholders may not realize this". For example, many students are not aware and even might not believe that dropout from a degree can be predicted with pretty high accuracy at the end of the first semester of study as shown in various works (Berens et al. 2019, Wagner et al. 2020).

- We will use a *rapid prototyping* approach to discuss and test conceptual ideas, especially with students and lecturers. That way they will be able to test design ideas and give concrete feedback and suggestions for improvements.
- Our *evaluation* concept includes qualitative as well as quantitative methods, e.g. interviews, usability tests and questionnaires. That way, we will be able to combine in-depth feedback and more subjective views with a large-scale quantitative evaluation.
- We will develop an integrated *Learning Analytics concept*, including best practice collections and recommendations for lecturers, student counselors, and program and course designers. This is aimed at providing hands-on advice on how to incorporate our research findings into everyday practice.

An overall challenge of this *Human-Centered-Design* approach is to motivate various stakeholders to be part of the adventure. As far as students are concerned, de Quincey et al. (2019) describe an interesting approach involving four student-ambassadors who reach out to teams. A pilot study described in Brun et al. (2020) reports the involvement of more than 300 students in the design of dashboards without describing in detail how they reached out to students. Similarly, Rodriguez-Triana et al. (2018) describe the involvement of teachers and students in the design of dashboards for teachers. We aim to make participation as easy and rewarding as possible, for example by offering incentives for participation in online surveys, but also including interviews and focus groups as part of course achievements. The last point is particularly important in the context of online-degrees where students are more mature, work full-time, have family, and, therefore, a tight schedule. Furthermore, a number of the degree programs offered are connected to Information Technology (IT) and Computer Science. They include courses like "Human-Computer Interaction", "Data Base", "IT-Law", "Algorithms", "Artificial Intelligence", "User Experience", or "User-Centered Design". Thus, group work and discussion involving topics of the project like "user interfaces", "data", "data privacy", "trust" or "explanations of models", to name a few, allow for a fruitful combination of teaching and research. Such a procedure would also ease the participation of different teachers in the project.

However, reliance on volunteers might also present a methodical problem, as volunteers might be more motivated than students in general, making it harder to generalize the results. Therefore, it will be crucial to motivate critical students who do not wish to release their user data to participate in other forms of data collection, e.g. interviews and questionnaires.

As a first step, data collection and analysis will be strictly voluntary. Students will be informed in detail about what and how data is used for analysis. There will be an easy opt-out possibility, only data from students who consented to data analysis will be included in our research. This approach will allow us to build trust and confidence with all stakeholders of the project. Depending on the outcomes of this approach, data collection and analysis might be broadened.

5 CONCLUSION

This article presents the DiSEA project. A key objective of the project is to provide personalized advice via early warning systems to those students who are at risk of dropping out. The focus of the project is on digital study formats. The project will use the extensive experience and data of the university network "Virtual University of Applied Sciences". The analysis of academic data from the participating universities, combined with data on learning activities from the Moodle learning management system available at the VFH, is intended to provide early indications of the risk of dropping out. A major challenge in the project will be how to communicate the analysis results via suitable dashboards to the students concerned and the other people involved, such as student counseling, program management, lecturers, and how an early warning system can be used in practical applications. Challenges in this context are to communicate results in an explainable and comprehensible way. The trust of students, in particular, must be awakened through early participation in a user-centered or even participatory design.

ACKNOWLEDGEMENT: The DiSEA project is funded by the German Ministry of Education and Research (Bundesministerium fuer Bildung und Forschung –BMBF) under the project funding number 01PX21001A.

REFERENCES

Akçapınar, G., Hasnine, M.H., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments 6*:4, https://doi.org/10.1186/s40561-019-0083-4

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*, April 29-May 02, 2012, Vancouver, British Columbia, Canada.

Aulck, L. S., Nambi, D., Velagapudi, N., Blumenstock, J. E., & West, J. (2019). Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In *Proceedings of the 12th International Conference on Educational Data Mining* (Montreal, Canada, July 2-4). EDM 2019, 9-18. http://educationaldatamining.org/edm2019/proceedings/

Baker, R.S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education E-Learning Courses. I*nternational Conference on Educational Data Mining (EDM),* Madrid, Spain.

Baneres, D., Rodriguez, M. E., Guerrero-Roldán, A.E. & Karadeniz, A. (2020). An Early Warning System to Detect At-risk Students in Online Higher Education. *Applied Sciences 2020*, 10(13), 4427; doi:10.3390/app10134427 https://doi.org/10.3390/app10134427

Beard, L.A., & Harper, C. (2002). Student Perceptions of Online versus on Campus Instruction. *Education, 122* (4).

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk –

predicting student dropouts using administrative student data from German universities and machine learning methods. *Journal of Educational Data Mining, 11*(3), 1–41. https://doi.org/10.5281/zenodo.3594771.

Brandenburger, J., & Janneck, M. (i.V.). A teamwork tool to support group work in online-based higher education: Exploring user experience and the use of support mechanisms by students. *Submitted*.

Brandenburger, J., Constapel, M., Hellbrück, H., & Janneck, M. (2019). Analysis of Types, Positioning and Appearance of Visualizations in Online Teaching Environments to Improve Learning Experiences. *Proceedings of AHFE 2019*, Springer.

Brun, A., Gras, B., Merceron, A.: Building Confidence in Learning Analytics Solutions: Two Complementary Pilot Studies In: Adoption of Data Analytics in *Higher Education Learning and Teaching*, edited by D. Ifenthaler & D. Gibson, Springer, ISBN: 978-3-030-47392-1, 2020. pp. 285 -303.

Constapel, M., Doberstein, D., Hoppe, H. U., & Hellbrück, H. (2019). IKARion: Enhancing a learning platform with intelligent feedback to improve team collaboration and interaction in small groups. *Proceedings of the 18th International Conference on Information Technology Based Higher Education and Training*.

de Quincey, Ed., Briggs, C., Kyriacou, T. & Waller, R. (2019). Student Centred Design of a Learning Analytics System. Proceedings of Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3303772.3303793.

Diaz, D.P. (2002) Online Drop Rates Revisited. *The Technology Source*, May/June 2002.

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Education Technology Research and Development, 64,* 923–938. https://doi.org/10.1007/s11423-016-9477-y

Jayaprakash, S.M., Moody, E.W., Lauría, E.JM., Regan, J.R., and Baron, J.D. 2014. Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics 1, 1 (2014), 6–47.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2016). LATUX: an Iterative Workflow for Designing, Validating and Deploying Learning Analytics Visualisations. *Journal of Learning Analytics*, *2*(3), 9-39. https://doi.org/10.18608/jla.2015.23.3

Minks, K.-H., Netz, N., & Völk, D. (2011). *Berufsbegleitende und duale Studienangebote in Deutschland: Status quo und Perspektiven*. HIS: Forum Hochschule 11/2011.

Schneider, K., Berens, J., & Burghoff, J. (2019). Drohende Studienabbrüche durch Frühwarnsysteme erkennen: Welche Informationen sind relevant? *Zeitschrift für Erziehungswissenschaft, 22,* 1121–1146. https://doi.org/10.1007/s11618-019-00912-1.

Schwendimann, B., Rodriguez-Triana, M., Vozniuk, A., Prieto, L., Boroujeni, M., Holzer, A., Gillet, D., Dillenbourg. P. (2016). Perceiving learning at a glance: A systematic literature reviewof learning dashboard research. IEEE Transactions on Learning Technologies.

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge LAK19*, March 2019, 235–244. https://doi.org/10.1145/3303772.3303796

Rodriguez-Triana, M.J., Prieto, L.P., Martínez-Monés, A., Asensio-Pérez, J.I. & Dimitriadis, Y.. (2018). The Teacher in the Loop: Customizing Multimodal Learning Analytics for Blended Learning. *Proceedings of Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, New York, NY, USA, pp.417-426, https://doi.org/10.1145/3170358.3170364

Wagner, K., Merceron, A. & Sauer, P., (2020). Accuracy of a Cross-Program Model for Dropout Prediction in Higher Education. In Companion Proceedings of the 10th Learning Analytics and Knowledge Conference (LAK'20), p. 744-749. Workshop on Addressing Dropout Rates in Higher Education, Frankfurt am Main, Germany, 2020. https://www.solaresearch.org/wp-content/uploads/2020/06/LAK20_Companion_Proceedings.pdf