Talamas-Carvajal, J.A. (2023). Research plan on the effects of interventions on dropout predictions for Higher Education Institutions. In Proceedings of the 11th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2023). Braganca, Portugal

Research plan on the effects of interventions on dropout predictions for Higher Education Institutions

Juan Andrés Talamás-Carvajal[1[0000-0002-6140-088X]]

[1] Tecnologico de Monterrey, Monterrey, Nuevo Leon, Mexico
juan.talamas@tec.mx

Abstract. One of the main challenges that Higher Education Institutions face currently is dropout/ student retention. In most cases, identifying this group of students is no easy task, and doing so on time is even harder. This challenge requires both speed and accuracy, which makes it a prime candidate for the use of machine learning models and predictions. We are currently developing a series of models capable of early identification of students at risk of dropping out, with one key difference from classic approaches: we want to not only find out who these students are, but how we can best help them avoid that prediction. By developing methodologies capable of identifying and measuring the effects of a series of interventions (academic guidance courses, extra-curricular encouragement, diminished course load, etc.), we intend to develop a system capable of providing counterfactuals (what the student needs to change or do to reverse a prediction) based on the causal effects of the previously mentioned interventions. In this manner, we would not only identify groups of students at risk of dropping out, but would be doing so on time, and with a viable and specific strategy for each individual to improve.

Keywords: Dropout, Intervention effects, Higher Education, Educational Innovation.

1 Introduction and motivation

The COVID-19 pandemic brought with it one of the most important changes in education in history. For the first time, distance learning was brought front and center into the spotlight of education due to the overwhelming necessity of an alternative to in-person education. Curriculums had to be changed, rules updated, and evaluations transformed in order for education to continue in that new reality [1]. Even after the world returned to a relative "normality", the effects can still be felt. Saqr et al [2] summarized some of the research regarding the different topics of the pandemic (school closures, online education, challenges, etc.), and found that most articles dealt with problem understanding, challenges, and impact, and that the feeling of urgency may have made meaningful research difficult. The changes brought on by this period, together with the fact that we are teaching the most interconnected generation in history has made it of the utmost importance to quickly innovate and find appropriate methods of teaching, even at a distance.

Throughout the history of education there have been a multitude of approaches and techniques that have been used, with different degrees of success. We have seen content-based, problem-based, case-based, and even competence-based approaches, with each one focusing on a different objective. However, no matter the approach, figuring out how to help students at risk of dropping out has proven to be no easy task, even for Institutions with large amounts of data to use.

Using student data to better understand and improve teaching and learning is not a new idea. Going back as far as 2012, Learning Analytics and Student Analytics were terms that were being used when talking about data mining for academic purposes [3]. It has been shown that a data driven approach can help improve overall student performance when using it to help faculty provide meaningful feedback to students [4]. Both data mining techniques and resources have improved almost exponentially in recent years, it is reasonable to think we could improve and build on their results.

One of the biggest challenges for universities when trying to improve themselves are the several obstacles they face when attempting any type of change. Aside from the extremely long time that is needed before results can even be seen, there is also the problem of determining if the change was beneficial or not to the students. A large enough change would make direct comparison impossible.

Our systematic literature review found that, while there was a popularity boom for online courses and overall data gathering and data-based model development, there is a large area of opportunity for interventions to be used and their effects recorded, as very few explicit model replications were found, there is difficulty applying them in contexts different form their original one.

Due to these factors, we believe that there is value to be found in the development of prediction and classification models that take into account the effect of interventions, both during model development and afterwards, when evaluating the effects of such models. Specifically, we believe these models could be well suited to predict Student Dropout at early stages of their studies, as well as measure the effects of different intervention strategies like academic guidance courses or workshops to prevent dropout.

Machine learning in education usually tries to determine one of two common features, student success (usually simplified to grades) and student dropout. We believe that that the different probabilities and prediction can be enhanced by the addition of intervention based on those predictions, as well as a causal analysis to determine the specific effects of those interventions and the previously used model features.

The goal of this project is the development of machine learning models for the retention or dropout of students, with the added step that different interventions will be conducted towards the identified groups to help reverse their dropout prediction. As part of our research, we will be looking into which one of the various classification or prediction approaches (Neural Networks, Naïve Bayes, Logistic Regression, SVM, to name a few.) are best suited to our needs. However, we are also aiming to develop our models under an explainable AI principle: we need to be able to understand what the educational models are doing well and what they are doing wrong in order to be able to provide accurate information to stakeholders.

The objectives of this paper are as follows: Development of accurate machine learning models for dropout/retention of students. Measurement and integration of the effects of interventions for identified students in the developed models. Integration of Explainable AI methodologies and strategies into our models for optimal explainability and understandability.

Our research questions are these: 1) Do the specific electives that students take prior to choosing their specialization (classes as interventions before a decision) affect their performance in terms of academic performance (grades) and dropout? Should some of these electives be pre-requisites for others? 2) How do interventions affect students that were predicted to dropout/were at risk of dropping out? Which type of intervention generated the most positive outcome?

The rest of this paper will go into some background for the techniques used, background needed, research details, and results and future work.

2 Background

Data based decision-making has been proven to be a reliable tool for improving student performance in academic settings [4]. This is not limited to just data visualization or statistical insight, but to predictive models as well. However, the most important component in all of those is data itself. No valid insights, statistics, or predictions can be found or made without a large enough amount of quality data. The techniques used to make sense of that data are usually known as data mining, and in education, either Educational Data Mining or Learning Analytics.

2.1 Educational Data Mining and Learning Analytics

EDM was defined in the Journal of Educational Data Mining in 2008: The Journal of Educational Data Mining states that EDM is "concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students and the settings which they learn in" [5]. The Educational Data Mining Conference states that it entails Data Mining to "answer educational research questions, including exploring how people learn and how they teach. Educational data mining considers a wide variety of data types, including log files, student-produced artifacts, discourse, learning content and context, sensor data, and multi-resource and multimodal streams. The overarching goal of the Educational Data Mining research community is to support learners and teachers more effectively by developing data-driven understandings of the learning and teaching processes" [6].

Learning Analytics has been defined by Long et al. [7] as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" at the first-ever Learning Analytics and Knowledge conference in 2011. The

SOLAR Society (Society for Learning Analytics Research) mentions that there are 3 key methodologies when using LA: Descriptive Analytics, Diagnostic Analytics, and Prescriptive Analytics [8]. These correspond to what happened, why it happened, and advice on what might happen. Furthermore, the 2017 and 2022 versions of the Handbook of Learning Analytics state that LA is not easily condensed to a simple definition, and instead deals with the problems, opportunities, inquiries, communities, and refer to the "specific incarnation of the bigger shift to an algorithmically pervaded society". [9,10]

Learning Analytics projects usually focus on using insights and discoveries for educational purposes, such as individual and tutoring recommendations, while Educational Data Mining tends to go for discoveries that come from educational databases. An example of this can be seen in the article by Pascual and Cobos [11], in which they use an initial dropout prediction to provide interventions for at-risk students in the form of information regarding their course. Pardo et al. used pre-determined feedback to provide feedback regarding activity performance and found positive effects on both grades and feedback perception [12]. Iraj et al. performed a similar study in which they found a relationship between early engagement with feedback and course results, and previous engagement with future engagement [13]. Generally, some differences between EDM and LA could theoretically be used to distinguish them, but both aim to improve education by better understanding the reasons for something (EDM) or looking for the best responses (LA). As time has gone by, both fields have shown a much deeper overlap, with researchers publishing papers on both sides of the field.

EDM has been classically applied to analyze student performance and dropout (at-risk cases) and identify over and underperforming students. Several of these articles have historically reported excellent results, with correct results ranging between 90% to 99% of their predictions (also called accuracy). For example, Asif attempted to predict students' graduation performance after only half of their program was completed and reported accuracies of 83.65% [14]. Moreno-Marcos reported AUC values for identifying dropout in MOOCs of 99%, which can be considered excellent in many models [15]. Another example can be seen in a paper by G. Akçapınar, where the authors mention obtaining an accuracy of 84% [16].

In the last decade, data mining and machine learning studies on education have focused heavily on classification problems regarding student performance. Most studies attempt to predict if a student will fail, is at risk of failing, or will pass by classifying (separating) students into these respective groups; a few more even go as far as to try to predict if a particular student will do exceptionally well on a course [17, 18, 19, 20, 21, 22, 23]. Classification models such as these are excellent tools for improving the general performance of school cohorts, improving student retention, and the school's general reputation.

To achieve this result, classification models use several techniques, such as Decision trees, Naïve Bayes, Logistic regression, and Artificial Neural Networks. Such techniques have produced 70% to 85% accuracy in identifying failing students [21, 23].

2.2 Educational Models

While identifying students who are either at risk of failing or dropping out is of the utmost importance, academic performance does not correlate directly with adult life economic success. Even though success can be somewhat subjective, we know that, in general, the annual income of a person depends on the highest degree they managed to obtain. In Mexico, a person with the equivalent of an undergraduate degree can expect to earn twice as much as the national average, while someone with a postgraduate degree can expect 3 times that same amount or more. A similar relationship can be observed on data from the U.S. [24, 25]. This phenomenon has been widely documented and reported, to the point it has become common knowledge, at least regarding education up to an undergraduate level.

However, these relationships correspond to the completion of a degree, and not the overall academic success obtained. As a higher education institution, graduate employability and earning potential are key elements of what makes up the school's reputation, as well as the primary reason students and their parents look for specific schools and programs. As such, identifying features that could predict these two would help greatly with improving study plans, creating opportunities for students to obtain soft skills from extracurricular activities, or even changing the requirements for graduation so that they include these features.

One of the steps being taken by educational institutions towards improving employability is the development of curriculums and models that develop the necessary skills, knowledge, and attitudes required of graduating students to transition to the workplace in their chosen subjects. The combination of those 3 factors is called a competence. These competences are declared by the universities and industries and make up the graduate profile of the numerous majors available to students. It is around those graduate profiles that competence-based models are designed, also known as competence-based education, or CBE.

A competence-based model has as its objective the "mastery of knowledge, skills and abilities that demonstrate learning" [26]. While competence-based models are relatively new to higher education, they have been in use for close to 30 years in programs like secondary education [27] and in cases where skills and knowledge should not be without each other, as is in medical care [28].

The educational model proposed by Tecnologico de Monterrey (TEC21) is a challenge-based model with 3 distinct phases in its curriculum: discovery, focus, and specialization. New students have the option of enrolling in an undergraduate degree directly or spending 3 or 4 semesters on a more general path (engineering, social science, bio-sciences, etc.). The students that do not choose an undergrad directly have to take a series of elective courses for them to familiarize themselves with their different options and make an informed decision about their future.

Challenge based models have started to show promise in terms of competence development against traditional systems, as can be seen by the EPICS programs at Purdue University, for example [29]. Purdue University has a history of being at the vanguard of educational research, which gives some additional weight to their use of CBE. Before their EPICS programs, they were among the first institutions to use

EDM and LA as part of their decision-making process, as could be seen in Arnold and Pistilli's paper [30].

3 Research approach and methods

For our case study, be will we looking into the TEC21 model by Tecnologico de Monterrey, which is certified by the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC), the Mexican Federation of Private Institutions of Higher Education (FIMPES), and the Accreditation Board for Engineering and Technology, Inc. (ABET), among various others. The TEC21 model is a challenge-based educational model (also called challenge-based learning), and combines aspects of competence, problem, and case-based methodologies. The main objectives of this teaching model are for the student to acquire the fundamentals related to their chosen area, as well as developing the competencies and acquiring the skills they will need in their future. In this model, challenges are the central learning unit, with everything else focusing on delivering the needed tools for solving them.

3.1 Research Method

We will be implementing a CRISP-DM [31] methodology for our experiments, data manipulation, and model development. This methodology covers a series of steps that follow a set order but encourages iterations between steps as a way of improving the overall quality of the project. These steps are as follows: Business understanding: Dissection of the problem/project; Data understanding: Identifying the necessary data; Data preparation: Dataset cleaning and transformations; Modelling: Development of the model; Evaluation: The overall results are reviewed and considered in the context of the problem/project; Deployment: Model release and final collection of data.

3.2 Data acquisition

We will be using student data provided by Tecnologico de Monterrey, specifically information regarding their Tec21 model that is currently running. All datasets provided are obtained and managed by the IFE Living Lab and Data Hub. They make sure that privacy issues, and all details concerning the collection, curation, and publication of student data were validated with Tecnológico de Monterrey's Data Owners and the Data Security and Information Management Departments. As of the writing of this article, two datasets from calls for papers have been used, one for dropout research and one for competency research. The dropout dataset contains 121,576 unique students with 49 distinct features and an identification number, while the competency dataset contains 33,319 unique students with 47 features and randomized identification numbers. Future datasets will be either similarly from calls for papers, or custom datasets with similar features to the 2 mentioned before. The Interventions are planned to be either specific guidance courses or student activities.

3.3 SHAP as an explainability tool

SHapley Additive exPlanations (SHAP) is an explainability tool for machine learning algorithms that is used to identify relevant features in complex machine learning models, to better understand their effect on final predictions. SHAP is based on the use of shapley values, a concept in game theory where the players (variables) interact with one another to determine the individual contribution to the final result. The resulting shapley values make up the difference between the expected value (the average) and the individual result. SHAP applies shapley values into machine learning to see through the “black box” models. Some mathematical properties of SHAP makes them valuable resources for model explainability: they are locally accurate, meaning they approach the true model as values are removed; missing values or values with no effect don’t impact the model; and if a variable’s contribution increases or stays the same, this is reflected on the resulting shapley value. [32]

4 Results to date and future work

Up to the writing of this document, a state-of the art systematic review has been performed regarding data mining in education. Specifically, we looked into its more prominent objectives (dropout, grade prediction), most commonly used methods (classification, numerical prediction, clustering), most relevant input variables (socio-demographics, grades, log-data), and prevalent model deployment (small percentage of models were reported to be deployed outside the original context).

Stemming from those results, we developed an early dropout detection model using a low number of variables available in any Higher Education institution. This model is accompanied by a method for model viability based on an expected value framework, with a break-point for minimum required retention for it to be self-sustainable. An article was written around this model and has since been published [33].

Continuing from those results, we then started to investigate possible methods for intervention identification and effect measurement and have since started the development of methodologies to measure the effects of interventions. As an initial step towards this goal, we are currently working with a SHAP approach to dropout prediction with the intent to measure intervention effects, which was presented at the LAK23 conference as a poster presentation and awarded an Honorable Mention. [34]

At the time of writing this article, we have developed an early identification model that runs with just a few variables generally available at any higher education institution allowing for a relatively easy deployment. Along with the model, we describe a method for its viability using an expected value framework. This framework allows for a cost-benefit analysis and provides a breakpoint for the models. We have also done research into what features (educational or socio-demographic) are relevant when looking to develop competencies in higher education in the Tec21 model, and on the effects of the pandemic on competence development at different levels. Two articles are currently under review regarding these two topics.

Using SHAP values, we have been able to show initial differences in feature importance between student groups of different educational models, which we are attempting to validate in interviews with practitioners. If confirmed, these could serve as a basis for practitioners looking into student mentoring.

As future work, we intend to develop a method for measuring the effect of interventions in machine learning models that can be used in conjunction with model predictions in order to empower practitioners to make better data-based decisions. Along with this, we also intend to measure the causal effect of educational interventions in student's risk of dropping out.

5 Conclusion

Modern education has recently started to accelerate its rate of change due to the popularity boom of alternative sources of instruction like MOOCs and the forced innovation brought by the COVID-19 pandemic. As new methods and educational models arise, it's imperative that we make sure those changes truly benefit students and the teaching-learning process. It is in this aspect that data mining, and specifically, modern machine learning algorithms, can make a sizeable contribution.

With the use of precise prediction models, we can identify students at risk of dropping out in an early enough timeframe so that we can intervene on time and prevent that "prophecy" from fulfilling itself. We can even go one step further and figure out what kind of intervention would be best suited in every case: is it an academic guidance course, extracurricular activities, a smaller academic load? In more technical terms, we could create counterfactuals specifically tailored to individual students to reverse the original predictions, instead of "only" identifying these students on time.

Acknowledgements

This research is currently being carried out within Tecnologico de Monterrey as part of its Doctoral Program in Computer Science, based in the Monterrey Campus at Nuevo Leon, Mexico.

The author would like to acknowledge the Living Lab & Data Hub of the Institute for the Future of Education, Tecnológico de Monterrey, Mexico, for the data published through the Call "Bringing New Solutions to the Challenges of Predicting and Countering Student Dropout in Higher Education" used in the production of this work.

The author would like to thank the financial support from Tecnológico de Monterrey through the "Challenge-Based Research Funding Program 2022". Project ID # I004 - IFE001 - C2-T3 – T.

References

1. García-Peñalvo, F.J.; Corell, A.; Abella-García, V.; Grande, M. Online assessment in higher education in the time of COVID-19. Education in the Knowledge Society (2020). doi:10.14201/eks.23013.
2. Saqr, M., Raspopovic Milic, M., Pancheva, K. et al. A multimethod synthesis of Covid-19 education research: the tightrope between covidization and meaningfulness. Univ Access Inf Soc (2023). https://doi.org/10.1007/s10209-023-00989-w
3. Ferguson, R.: Learning analytics: drivers, developments and challenges. International journal of technology enhanced learning 4, (5-6), 304–317 (2012).
4. Arnold, K., Pistilli, M.: Course signals at Purdue. In: LAK '12, ACM, pp. 267–270. (2012)
5. EDM Homepage, https://educationaldatamining.org, last accessed 2023/05/30.
6. EDM 22 Homepage, https://educationaldatamining.org/edm2022/, last accessed 2023/05/30.
7. Long, P. D., Siemens, G., Conole, G., & Gašević, D. In: LAK'11: Proceedings of the 1st international conference on learning analytics and knowledge. (2011)
8. What is Learning Analytics? - Society for Learning Analytics Research (SoLAR), https://www.solaresearch.org/about/what-is-learning-analytics/, last accessed 2023/05/30.
9. Lang, C., Siemens, G., Wise, A., Gasevic, D.: Handbook of Learning Analytics. Society for Learning Analytics Research (SoLAR), 2017. doi: 10.18608/hla17.
10. Lang, C., Wise, A., Merceron, A., Gasevic, D., Siemens, G.: Handbook of Learning Analytics – Second Edition. Society for Learning Analytics Research (SoLAR), 2022. https://www.solaresearch.org/publications/hla-22/
11. Pascual, I., & Cobos, R.: A proposal for predicting and intervening on MOOC learners' performance in real time. CEUR Conference Proceedings vol. 3238, pp. 26–38. LASI-Spain 2022 (A. Vázquez-Ingelmo, Y. Dimitriadis, A. Martínez-Monés, and F. J. García-Peñalvo, eds.), Salamanca, Spain, June 20-21, 2022.
12. Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., Mirriahi, N.: Using learning analytics to scale the provision of personalised feedback. Br. J. Educ. Technol., vol. 50, no. 1, pp. 128–138. (Jan. 2019), doi: 10.1111/bjet.12592.
13. Iraj, H., Fudge, A., Faulkner, M., Pardo, A., Kovanović, V.: Understanding Students' Engagement with Personalised Feedback Messages. Proc. LAK 2020, pp. 438–447. doi: 10.1145/3375462.3375527.
14. Asif, R., Merceron, A., & Pathan, M. K.: Predicting Student Academic Performance at Degree Level: A Case Study. International Journal of Intelligent Systems and Applications, 7(1), 49–61 (2014). https://doi.org/10.5815/ijisa.2015.01.05
15. Moreno-Marcos, P. M., Muñoz-Merino, P. J., Maldonado-Mahauad, J., Pérez-Sanagustín, M., Alario-Hoyos, C., & Delgado Kloos, C.: Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. Computers and Education, 145, (2020), 103728. https://doi.org/10.1016/j.compedu.2019.103728
16. Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H.: Developing an early-warning system for spotting at-risk students by using eBook interaction logs. Smart Learning Environments, 6(1), (2019). https://doi.org/10.1186/s40561-019-0083-4
17. Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M.: Predicting achievement and providing support before STEM majors begin to fail. Computers and Education, 158, 103999 (2020). https://doi.org/10.1016/j.compedu.2020.103999
18. Bravo-Agapito, J., Romero, S. J., & Pamplona, S.: Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study.

Computers in Human Behavior, 115, 106595 (2021). https://doi.org/10.1016/j.chb.2020.106595
19. Coussement, K., Phan, M., De Caigny, A., Benoit, D. F., Raes, A., Caigny, A. De, Benoit, D. F., & Raes, A.: Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model. Decision Support Systems, 135, 113325 (2020). https://doi.org/10.1016/j.dss.2020.113325
20. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q.: Predicting academic performance by considering student heterogeneity. Knowledge-Based Systems, 161, 134–146 (2018). https://doi.org/10.1016/j.knosys.2018.07.042
21. Riestra-González, M., del Puerto Paule-Ruíz, M., Ortin, F., Paule-Ruíz, M. del P., & Ortin, F.: Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers and Education, 163, 104108 (2021). https://doi.org/10.1016/j.compedu.2020.104108
22. Tomasevic, N., Gvozdenovic, N., & Vranes, S.: An overview and comparison of supervised data mining techniques for student exam performance prediction. Computers and Education, 143, 103676 (2020). https://doi.org/10.1016/j.compedu.2019.103676
23. Zeineddine, H., Braendle, U., & Farah, A.: Enhancing prediction of student success: Automated machine learning approach. Computers & Electrical Engineering, 89, 106903 (2021). https://doi.org/10.1016/j.compeleceng.2020.106903
24. Learn more, earn more: Education leads to higher wages, lower unemployment, https://www.bls.gov/careeroutlook/2020/data-on-display/education-pays.htm#:~:text=If%20you're%20wondering%20whether,rates%20decrease%20and%20earnings%20increase.&text=5.4-,Note%3A%20Data%20are%20for%20persons%20age%2025%20and%20over.,time%20wage%20and%20salary%20workers, last accessed 2023/05/30.
25. Xieling Chen, Haoran Xie, Di Zou, and Gwo-Jen Hwang. Application and theory gaps during the rise of Artificial Intelligence in Education. Computers and Education: Artificial Intelligence, 1:100002 (2020)
26. Simonds, J., Behrens, E., Holzbauer, J.: Competency-Based Education in a Traditional Higher Education Setting: A Case Study of an Introduction to Psychology Course (2017)
27. Sullivan, S., Downey, J.: Shifting Educational Paradigms: From Traditional to Competency-Based Education for Diverse Learners. American secondary education, 43(3) :4–19 (2015)
28. Harden R.: AMEE Guide No. 14: Outcome-based education: Part 1-An introduction to outcome-based education. Medical teacher, 21(1):7–14 (1999)
29. EPICS Homepage, https://engineering.purdue.edu/EPICS, last accessed 2023/05/30.
30. Arnold, K., AND Pistilli, M.: Course signals at Purdue. LAK ’12, ACM, pp. 267–270. (2012)
31. Wirth, R & Hipp, J.: CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (2000).
32. Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K. W., Newman, S. F., Kim, J., & Lee, S. I. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, *2*(10), 749–760 (2018). https://doi.org/10.1038/s41551-018-0304-0
33. Talamás-Carvajal, J.A., Ceballos, H.G.: A stacking ensemble machine learning method for early identification of students at risk of dropout. Educ Inf Technol (2023). https://doi.org/10.1007/s10639-023-11682-z

34. Talamas-Carvajal, J. A.: The Middle-Man Between Models and Mentors: SHAP Values to Explain Dropout Prediction Models in Higher Education [Poster presentation]. Learning Analytics and Knowledge Conference, Arlington, Texas, USA. (2023). https://www.solaresearch.org/wp-content/uploads/2023/03/LAK23_CompanionProceedings.pdf