Machine learning algorithms for predicting academic performance and identifying the contributing factors

Roland Molontay[1], Marcell Nagy[1]
[1]Budapest University of Technology and Economics, Budapest, Hungary

Data-driven approaches have received a lot of attention recently from higher education researchers and policy-makers as well. At Budapest University of Technology and Economics in fruitful cooperation with an industrial partner who developed the unified educational administrative system, an extract insights from big data stored in the administrative system have been done. Among many other questions, there has been a study curriculum prerequisite networks with a student flow approach, the effect of mathematical remediation, the impact of living on-campus on academic achievement, the connection between grade inflation and student evaluation of teaching and efficient visualization of student flows. However, one of the most burning problems in higher STEM education all over the world is dropping out. A predictive analytical approach for early detection of students at risk of academic failure is shown. Relatively high accuracy has being achieved, compared to the results of related works. A web application that is able to identify at-risk students mainly based on their high school results and matura results using machine learning algorithms (neural networks and gradient boosting trees) has been developed. The application can also be used to recommend tutoring sessions and remedial courses for at-risk students. The application is also capable of making suggestions for students which skills to improve in order to succeed in their university studies.

Keywords: machine learning, predictive analytics, predicting dropout, educational data mining, leaning analytics

REFERENCES

[1] Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). Predicting Student Dropout in Higher Education. *arXiv preprint arXiv:1606.06364*.
[2] von Hippel, P., & Hofflinger, A. (2017). The Data Revolution Comes to Higher Education: Identifying Students at Risk of Dropout in Chile. *Available at SSRN 3073912*.
[3] Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E., & Wollscheid, S. (2015). Dropout and Completion in Higher Education in Europe: Main report. European Union.
[4] Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems* (pp. 4765-4774).
[5] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.
[6] Nagy, M., & Molontay, R. (2018). Predicting Dropout in Higher Education Based on Secondary School Performance. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)* (pp. 389-394). IEEE.
[7] Nagy, M., Molontay, R., & Szabó, M. (2019). A Web Application for Predicting Academic Performance and Identifying the Contributing Factors. In 47th SEFI Conference.