

How can a Data-driven and AI-driven Model be Developed to Predict Early Student Dropout in Higher Education?

Joanna Alvarado Uribe, Paola Mejía, Ubaldo Martínez

joanna.alvarado@tec.mx, gabriela.almada@tec.mx, ubaldo.martinez@tec.mx

Living Lab & Data Hub, Institute for the Future of Education, Tecnológico de Monterrey

October 22 | 9:00-13:00 | Room 1

Registration Link

<https://forms.office.com/r/1h3rscgQex>

Abstract

If a student loses learning during a specific year, the loss is forever for the student, the institution, and society. The high dropout rate costs students, educational institutions, and society as a whole [1, 2]. The importance of having an accurate model to predict student dropout is that its results could be used to improve and develop retention strategies in Universities [3]. This tutorial proposes to build theoretically and practically an approach to collect, analyze, and predict early dropout of university students, using statistical techniques and Machine Learning algorithms, through the CRISP-DM methodology. For the case study, a dataset from Tecnológico de Monterrey [4] will be used to analyze the factors that influence early dropout and develop a model to predict it.

References:

1. Raisman, N. (2013). The Cost of College Attrition at Four-Year Colleges & Universities. Policy Perspectives. Educational Policy Institute.
2. Latif, A., Choudhary, A. I., & Hammayun, A. A. (2015). Economic effects of student dropouts: A comparative study. Journal of Global Economics.
3. Cardona, T. A. & Cudney, E. A. (2019). Predicting student retention using support vector machines. Procedia Manufacturing, 39, 1827-1833.
4. Alvarado-Uribe, J., Mejía-Almada, P., Masetto Herrera, A.L, Molontay, R., Hilliger, I., Hegde, V., Montemayor Gallegos, J.E., Ramírez Díaz, R.A., & Ceballos, H.G. (2022). Student Dataset from Tecnológico de Monterrey in Mexico to Predict Dropout in Higher Education. Data, 7, 119. <https://doi.org/10.3390/data7090119>

Tutorial 5



Target Audience

This tutorial is open to all researchers, faculty, and graduate students.

Pre-requisites

Knowledge of: Statistics and Machine Learning algorithms and Python programming language.
Furthermore, each participant must sign a “Terms and Conditions of Data Use” document that IFE Data Hub will provide before the tutorial in order to access the student dropout dataset.

Technical requirements

Laptop, Institutional email account, Jupyter notebook, and Internet connection.