"Reducing dropout rates: the challenge of Learning Analytics in Higher Education Institutions"

PhD Candidate: Marta Cannistrà

Advisor: Prof.ssa Anna Maria Paganoni

Co-Advisor: Prof. Tommaso Agasisti

The PhD thesis primarily focuses on exploring the dynamic interplay between Learning Analytics and university dropout. Learning Analytics forms the methodological cornerstone for my research and the overall development of the thesis. In parallel, university dropout emerges as a crucial policy concern, serving as the central theme of this study. Essentially, the objective of my PhD thesis is to confront the formidable challenge of mitigating dropout rates in Higher Education by employing Learning Analytics tools to enhance the decision-making process.

Firstly, the thesis is dedicated to contextualizing my research within the existing scholarly landscape. It introduces Learning Analytics as the culmination of an extensive discourse aimed at integrating research into the administrative fabric of universities. This integration is crafted to streamline decision-making and tap into the institutional expertise inherent within educational settings. Following this, the emphasis is on elucidating critical facets of university dropout rates. This involves delving into the academic dialogue surrounding dropout rates, tracing its evolution, exploring diverse definitions of dropout, and examining the reasons compelling institutions to address this challenge.

Finally, the thesis synthesizes the concepts of Learning Analytics and university dropout rates. It draws upon pertinent examples from the literature and situates them within the context of Politecnico di Milano in Italy. A framework is presented to conceptualize a workflow supporting decision-making processes within Higher Education Institutions. The four papers constituting this thesis are positioned within this framework, and their principal characteristics are expounded upon.

The first paper aimed at dealing university dropout by adopting Learning Analytics (LA) tools: utilizing advanced techniques from statistical and Machine Learning (ML) domains, offer promise in predicting at-risk students early on. Effective algorithms could identify students needing intervention for better retention. However, the debate on the best predictive models is ongoing, and practical implementation of LA solutions is significant.

This study contributes to this area by developing innovative methods to predict at-risk students early in their academic journey, using administrative data from Politecnico di Milano. The research investigates the performance and interpretation of various algorithms, specifically Machine Learning and Generalized Linear Models, in predicting dropout. This exploration informs potential interventions to support students at risk of dropping out and bridges the gap between data-driven approaches and theoretical foundations. The study also introduces novel methods that account for the nested structure of data, such as considering different degree courses, enhancing the accuracy and robustness of dropout predictions.

The second paper in the thesis examines dropout from a temporal perspective, addressing a notable gap in existing academic literature where the time component of this phenomenon is often overlooked. Understanding not only the determinants of withdrawal but also when it is likely to occur provides a crucial advantage for policymakers aiming to implement timely interventions to retain at-risk students. This paper focuses on analyzing student trajectories at Politecnico di Milano with a dual purpose: investigating factors influencing dropout over time and pinpointing the earliest accurate moment for dropout predictions. The goal is to enable universities to implement early interventions through targeted preventive actions, emphasizing not only identifying at-risk students but also determining when they are most vulnerable. The paper employs survival analysis methods, studying each unit of analysis in a follow-up period to detect potential dropout within a specified time-window. A methodological innovation involves considering frailties to model potential dependencies across degree courses. Results indicate that students enrolled in the same courses exhibit similar trajectories over time. Additionally, the paper constructs an Early Warning System that identifies the optimal point for predicting the time-to-dropout for students at Politecnico di Milano, suggesting that policymakers should balance early intervention with a reliable estimation of risk.

This paper assesses the impact of a nudging intervention designed to reach out to students at risk of dropout and inform them about the university's free tutoring services. The rationale behind directing students to tutoring stems from research indicating that dropout is often linked to academic unpreparedness, a issue that tutoring can address. Additionally, previous literature establishes a causal relationship between traditional remedial interventions, like tutoring, and academic improvement. The study, conducted at Politecnico di Milano, unfolds in two phases: a quasi-experiment in the 2020/21 academic year and a Randomized Controlled Trial (RCT). In the quasi-experiment, high-risk students receive a communication encouraging them to enroll in tutoring, while the RCT involves reaching all at-risk students with diverse communications sharing the same objective as the quasi-experiment but with different content. The evaluation focuses on measuring the differential effect of receiving one message over another. Through random assignment, this experiment establishes a solid basis for exploring causality between treatment (varied nudging communication) and the desired outcome (tutoring enrollment). Results indicate that nudging communication leveraging peer behavior significantly increases the likelihood of students enrolling in tutoring activities. Table 4 provides an overview of the main characteristics of this paper.

While numerous studies explore factors influencing individual student dropout risk, cross-country evidence remains limited. Variations in data collection methods, definitions, and usage across European institutions hinder comparative research. Yet, many HE institutions routinely gather student-level data for dropout prediction and intervention, opening doors for cross-institutional comparisons. Despite data-driven dropout prediction, it's uncertain whether these models are country-specific or applicable to other institutions and contexts.

This study conducts a comparative analysis involving two European universities, one in the Netherlands and one in Italy, to better grasp dropout prediction. The research investigates if commonly used dropout predictors hold consistent value across different countries and institutions. Models trained on one country's university data are assessed for their predictive capability in a different country. Shared student background and academic performance variables are adjusted for both settings. First, the study evaluates if these shared variables effectively explain dropout within their original contexts. Then, it examines the external validity of predictive models across different institutions and countries at various points in the academic year. By comparing dropout prediction models across countries, policymakers aim to identify universal factors that can enhance student success.