

# Fairness and Bias in Prediction of Students' Dropout in Higher Education

Haoze Zhu; Zhanyang Sun

January 2025

Student dropout has always been a concern worldwide, with higher education facing growing challenges in keeping students enrolled due to various reasons. AI-driven models are increasingly used to predict students' dropout in higher education and to identify associated at-risk students to implement early interventions. However, if these models are biased, they can reinforce existing educational inequalities among certain student groups.

Previous research on this matter has mainly focused on qualitative methods, with few implementing quantitative approaches. We find that even among those who are tackling this matter from a quantitative perspective, deep learning is not widely used [1]. Nevertheless, numerous researchers have demonstrated that deep learning models can often provide remarkable results compared to traditional machine learning models. In this project, we want to explore whether the inclusion of age, race, and other socioeconomic factors will affect the accuracy and fairness of student dropout prediction models.

In this project, our dataset, 'Predict Students' Dropout and Academic Success', is obtained from the UC Irvine Machine Learning Repository [4]. The dataset contains 36 features, including information known at the time of student enrollment (academic path, demographics, and social-economic factors) and students' academic performance at the end of the semester. We decide to experiment with deep learning models on the dataset and investigate its classification results by checking fairness and bias in the model. What's more, we want to look into the systematic biases of the model by inspecting metrics such as false positive and false negative rates across different student subgroups.

## References

- [1] Lovenoor Aulck, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. Predicting student dropout in higher education. In *ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*, 2016.
  - [2] Efthymoulos Drousiotis, Panagiotis Pentaliotis, Lei Shi, and Alexandra I. Cristea. Capturing fairness and uncertainty in student dropout prediction – a comparison study. In Ido Roll, Danielle McNamara, Sergey Sosnovsky, Rose Luckin, and Vania Dimitrova, editors, *Artificial Intelligence in Education*, pages 139–144, Cham, 2021. Springer International Publishing.
  - [3] Liga Paura and Irina Arhipova. Cause analysis of students’ dropout rate in higher education study program. *Procedia - Social and Behavioral Sciences*, 109:1282–1286, 2014. 2nd World Conference on Business, Economics and Management.
  - [4] Vieira Martins Mónica Machado Jorge Realinho, Valentim and Luís Baptista. Predict Students’ Dropout and Academic Success. UCI Machine Learning Repository, 2021. DOI: <https://doi.org/10.24432/C5MC89>.
- [1] [2] [3]