EduBoost: An Interpretable Grey-Box Model Approach to Identify and Prevent Student Failure and Dropout

Qin, Andy (School: Plano West Senior High School)

Student failure and dropout in universities are detrimental in many ways, costing $1700 on average in tuition fees, risking future financial aid, and prompting issues in mental health and motivation (College Post, 2021). Considering 220 million active users in MOOCs (Massive Open Online Courses), billions of dollars were lost and millions of student dropouts occurred. A lot of work has been done in industry (such as IBM SPSS) and relevant literature, targeting high performance but providing little meaningful feedback for at-risk students. This paper proposes EduBoost, a novel, intrinsically interpretable solution boasting comparable performance to Black-Box models. EduBoost contains a voting ensemble model formed using the best 3 Black-Box classifiers out of 4 tested (SVM, RF, MLP, KNN), a White-Box classifier CART, and two distinct Grey-Box model pipelines that use interpretable mimic learning. EduBoost is tested on two datasets: 1) student data from a mandatory university course and 2) a MOOC provided in the Open University Learning Analytics Dataset (OULAD), containing various features that gauge student performance and effort in class, i.e., grades, attempts, procrastination, etc. EduBoost showed increases of 3-11% in accuracy, precision, recall, and f1-score in comparison to the baseline White-Box model, peaking at 94.6% and 87.3% accuracy on the university and MOOC dataset respectively. More importantly, EduBoost's IF-THEN interpretability provides simple and helpful feedback to provide students with specific targets to improve. EduBoost has also been implemented as a cross-platform accessible application so students can view their progress and improve in real time.

Awards Won: