Abstract

Predicting and characterizing students that are at risk of failing is important because it would allow At-Risk students to be identified and recommended to various approaches that could improve their grades and help them pass. Not only is it important to predict At-Risk students, but also important to understand the features describing At-Risk students. By characterizing their features, the educator together with the student could identify which features could be changed in order to improve the probability of passing the course. In addition, At-Risk students could be recommended to various interventions such as supplementary tutorials, online videos or food programs and support groups. However, although these interventions are seen as beneficial, there is a great need to measure how effective these interventions are. Without measuring their effectiveness, i.e the treatment effects, there is no way to know if they are helping the student or not. Machine learning and statistical methods were used to address these objectives.

Preliminary work required a thorough preparation of the dataset to be used for training. A major component of the data preparation was the imputation of missing data using the random forest imputation method *missForest*. Further important preliminary work included detailed and thorough model selection procedures to identify the best performing models to be used for the main objectives. Lasso, random forest, adaboost, gradient boosting and neural networks were used in this study. Gradient boosting and neural network models performed the best in both classification and regression applications and were subsequently used in addressing the objectives. It is believed that these preliminary steps were vital to achieving better performance in addressing the objectives.

To better characterize At-Risk students, global and local model-agnostic machine learning methods were introduced. These carried out model explanations, or interpretations, describing which features are most important. Model-agnostic methods can be applied to any black-box model, for which model interpretation is difficult. Global (referring to the entire cohort of students) methods included permutation importances and partial dependence plots. In addition to these global agnostic methods, lasso β coefficients and tree-based feature importances were used for characterization. The results suggest that

using these global methods in combination provided confidence in describing overall features of At-Risk students as well as providing a way, via the partial dependence plots, to identify sub-regions in continuous features that could more accurately identify At-Risk students.

Although the global explanations provided insight into the overall characteristics, the results suggest that the local, or individual, methods may provide the most help to the student. Local methods used were the Local Interpretable Model-Agnostic Explanation (LIME) and Shapley Additive Explanations (SHAP) methods, developed by Ribeiro et al. (2016) and Lundberg et al. (2017), respectively. This provides explanations for individuals as to why they are At-Risk. The educator together with the student is then able to understand which features could be changed to change At-Risk status. To do this, counterfactual explanations were further generated. Counterfactual explanations, introduced for the first time to this field, perturbs a feature over a specific range and measures how the probability of being At-Risk changes as the feature is perturbed. This provided potential solutions to the student to see which feature can be changed to change the At-Risk status.

A machine learning method first introduced (as far as is known) by Beemer et al. (2017) and Beemer et al. (2018), was analyzed and extended. This method used machine learning to predict counterfactuals and treatment effects in observational studies. Traditional ways of measuring treatment effects in observational studies are to use propensity score matching (PSM) and this was the first known study validating this technique by comparing it with the PSM method. The results were found to be comparable. This method was further extended to be applied to not only compute global treatment effects but individual treatment effects. It was found that although global treatment effects were small, the individual treatment effects varied greatly, implying individual treatment effects provided more insight. A method was further developed to (1) predict an individual student's or sub-group of students' final grades and (2) predict their treatment effect. This would not only help predict the At-Risk status of the student, but also by how much an intervention would help them. Continuing with measuring treatment effects, the statistical method difference-in-differences (DID) was further used to measure treatment effects of supplementary online videos on performance. It was found that although global treatment effects were not-significant, there were significant local treatment effects.

Perhaps the main implication of this study was that using machine learning and statistical methods to analyze global predictions, characterizations and treatment effects are valuable, but focusing on smaller groups or individuals showed more potential for improving student success. Students are different and require different care. Machine learning is a powerful tool that can be used in conjunction with experienced caring educators to provide the best education to students.