Pre-course Prediction of At-Risk Calculus Students

James Cunningham[1], Raktim Mukhopadhyay[1], Rishabh Ranjit Kumar Jain[1],

Jeffrey Matayoshi[2], Eric Cosyn[2], Hasan Uzun[2]

[1] Arizona State University, Tempe, AZ, USA
[2] McGraw Hill ALEKS, Irvine, CA, USA

Abstract. Identifying students who are at-risk of failing a mathematics course at the earliest possible moment allows for support and scaffolding to be applied when it can have greatest impact. However, because risk of non-success can arise from a complex interaction of factors, early detection of struggling students is difficult. Machine learning is particularly suited to modeling this challenging interplay of variables. In this study, we measure how well machine learning models can identify at-risk students before an entry-level university calculus course begins. Five classification algorithms were applied to data combined from the student information system, an adaptive placement test, and a student survey. We were able to produce predictions before class start that were competitive with other studies using course activity data after coursework began. In addition, important features of the model provided insights into possible causes of student non-success.

Keywords: Predictive modeling, Machine learning, Placement tests, Calculus.

1 Introduction

Using machine learning (ML) for early alert systems of students at-risk in higher education has been an important application of learning analytics [1–5]. While there are many levels of early alert systems, our focus has been on course-level predictions especially in mathematics. The importance of completing calculus in the student's first attempt is critical in university majors related to science, technology, engineering, and math (STEM) [6–9]. However, calculus represents a substantial barrier to completing these majors especially for female students and students from underrepresented populations [10–11]. Early detection of students who may be struggling in these calculus courses is critical for intervention, scaffolding and support.

Often by the time gradebook data, used by many instructors for assessing risk, has made it clear that a student is failing calculus, over half the course is complete, making changing the outcome of that student difficult. While machine learning models are a viable alternative to instructor gradebooks for assessing risk of failure, most studies featuring predictive models depend heavily on course activity data from a learning management system (LMS) or a mathematics learning platform to classify students, making early, accurate predictions difficult when interventions are critical but course activity data is sparse [12–13]. For this reason, there is a lack of studies using machine learning predictive models before higher education courses begin [2–4].

This research contributes to this field of study by using placement assessment data as an alternative to course activity data for very early predictions in a higher education mathematics course. While a placement exam is a predictive model in itself, we hypothesized that combining data from the student information system (SIS) with placement data could yield classifications related to risk of failure with similar predictive power to classifications made in other studies with course activity data. Because these predictions would be available before the course begins, early intervention and allocation of limited resources for support and scaffolding could be strategically targeted when they could have the greatest impact.

One other source of pre-course data that was available to us for this study was a survey regarding math background that students filled out before the placement exam. Although this data was self-report, we were interested to see if this could also make a significant contribution to our model.

To this end, we conducted this study with these three research questions in mind:

RQ1: How would a machine learning predictive model, limited to data only available before the course starts, compare to predictive models using course activity data in accurately identifying at-risk calculus students?

RQ2: Which features of the model would be most important in predicting student risk?

RQ3: How much lift would be contributed to the predictions of the model from data derived from the SIS, placement test, and the self-report student survey?

2 Method

The data used in this study came from historical data of 6,380 undergraduate students enrolled in the course, Calculus for Engineers I. The label used for the supervised learning classification was "At-Risk" for students who achieved a final letter grade of 'C' or below, and "Not At-Risk" for students who achieved a final letter grade 'C+' or above with the reasoning that students who barely passed the course with a 'C' grade might be underprepared and have more in common with at-risk students than not at-risk students. Of our total population, 2,739 (43%) were labeled "At-Risk," and 3,641 (57%) were labeled "Not At-Risk." So, a baseline model for this data based on the majority class should be considered 57%.

The genders of the students in this study were 23% female and 77% male, and the age breakdown was 83% at or below 22 years old, 11% 23-30 years old, and 6% over the age of 30. Two proxies were used for socio-economic status. Students who were first generation college students made up 28% of our sample, and Pell eligible students made up 33%. The ethnicity breakdown was Asian 15%, Black 4%, Hawaii/Pac < 1%, Hispanic 21%, Native Am. 1%, No Report 3%, Two or more 5% and White 50%.

This data was merged with other academic and grade information from the SIS that would have been available before students began the calculus course, data from the placement test and the student self-report survey, and some engineered features from the SIS data that we thought might be predictive of risk.

The placement data used was from the ALEKS Placement, Preparation, and Learning (PPL), a specialized adaptive placement test developed to offer recommendations for placing students in post-secondary mathematics courses [14]. All this data was merged with three features from the self-report survey that was attached to the placement test. These three features were “last math level,” “last math class,” and “last math grade.”

Five ML methods were used for classification comparison: Logistic Regression, Support Vector Machines (SVM), K-Nearest Neighbors, Random Forest, and CatBoost. All of these methods except CatBoost were accessed through the Scikit-learn Python machine learning library [15]. CatBoost, a newer method, seeks to mitigate prediction shifts that are present in most implementations of gradient boosting by means of ordered target statistics associated with categorical variables [16–17]. The dataset was split into subsets, 80% for training and 20% for testing. Ten-fold cross validation was used to limit overfitting in our training set and increase generalization. The GridSearch CV library from Scikit-learn, which exhaustively considers all parameter combinations, was used for hyperparameter tuning.

A post-hoc algorithm was employed to extract feature importance from the black-box models and measure the lift of the different datasets. There are several new methods for model explainability; however, we chose Permutation Feature Importance (PMI) because it does not suffer from bias toward categorical variables as do some other methods [18–20].

3 Results

A comparison of the performance of the differing ML methods is presented in Fig. 1.

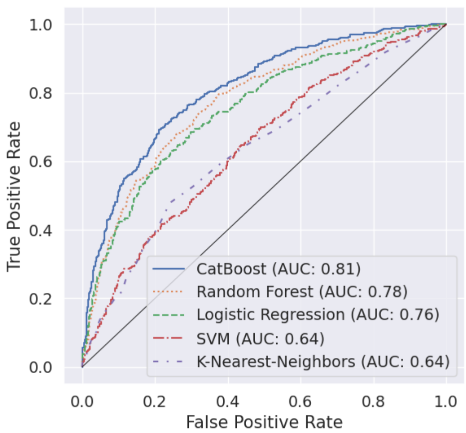

Fig. 1. Comparison of ROC curves for the five methods tested.

The receiver operating characteristic (ROC) is a performance measure of models at various threshold settings and is used to summarize the performance of models over a wide range of conditions. Of the five methods tested, CatBoost outperformed the other four machine learning methods with an area under the curve (AUC) of 0.81. The overall accuracy of our best model was 0.74, with a recall of 0.73, and an F_1 score of 0.71.

The Permutation Feature Importance algorithm scored each of the 46 features in terms of the contribution to the predictive power of our best model. The top five features were previous term GPA, last math class, part-time, placement test, and faculty difficulty with PFI scores of 0.049, 0.038, 0.035, 0.029 and 0.024 respectively.

By grouping the features from each dataset and using the PFI algorithm, we were able to score the impact of each dataset on the predictive power of the model. The scores for the three datasets were: SIS (0.148), ALEKS PPL (0.051), and survey (0.047).

4 Discussion and Conclusions

RQ1: Our first research question was aimed at measuring how well the models could predict who was at-risk of failing calculus before the course started without any course activity data. The ROC curves in Fig. 1 demonstrate that our predictions are comparable to other models using activity data in the first few weeks of a course [12–13]. **RQ2:** Using a post-hoc, model explainability algorithm, we were able to determine five features that were most important in early prediction of Calculus for Engineers I: previous term GPA, the last math class taken, official part-time status of the student, placement test data, and how hard an instructor typically grades their students. All three datasets used in this study had features represented in the top five. **RQ3:** Of the three datasets, features derived from the student information system were most predictive, with the placement test being second, and survey data coming in third.

Because accurate early detection is possible, scarce resources, scaffolding and support can be targeted to students who need it the most when the impact of those interventions can help students at-risk get off to a strong start. Moreover, because this kind of data is typically available for all entry level math courses at the university, it is possible to construct similar models for other critical math courses as well. Future work will focus on developing these models for other courses and combining these models with course activity data after classes start for even more accurate student modeling and weekly predictions that can guide interventions throughout the course for increased student success.

References

1. Arnold, K., Pistilli, M.: Course Signals at Purdue: Using learning analytics to increase student success. In: Proceedings International Learning Analytics & Knowledge Conference (LAK'12), ACM, (2012).

2. Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., Liao, S.: Predicting academic performance: a systematic literature review. In: Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE'18 Companion), pp. 175–199 (2018).
3. Namoun, A., Alshanqiti, A.: Predicting student performance using data mining and learning analytics techniques: a systematic literature review. Applied Sciences 11, pp. 1–28 (2021).
4. Rastrollo-Guerrero, J., Gómez-Pulido, J., Durán-Domínguez, A.: Analyzing and predicting students' performance by means of machine learning: a review. Applied Sciences 10, 1–16 (2020).
5. Romero, C., Ventura, S.: Educational data mining and learning analytics: an updated survey. WIREs Data Mining and Knowledge Discovery 10, 1–21 (2019).
6. Ayerdi, J.: Relative rates of success of students in Calculus I. Honors Theses. 2904. https://scholarworks.wmich.edu/honors_theses/2904
7. Bigotte de Almeida, M., Queiruga-Dios, A., Cáceres, M.: Differential and Integral calculus in first-year engineering students: a diagnosis to understanding the failure. Mathematics 9(61), 1–18 (2021).
8. Dibbs, R.: Forged in failure: Engagement patterns for successful students repeating calculus.
9. Garaschuk, K., Predicting failure in first-term calculus courses. UBC Faculty Research and Publications, https://open.library.ubc.ca/cIRcle/collections/facultyresearchandpublications/52383/items/1.0357414, last accessed 2021/02/06.
10. Nortvedt, G., Siqveland, A.: International Journal of Mathematical Education in Science and Technology 50(3), 325–343 (2019).
11. Sanabria, T., Penner, A.: Weeded out? Gendered responses to failing calculus. Social Sciences 6(47), 1–14 (2017).
12. MacFadyen, L., Dawson, S.: Mining LMS data to develop an "early warning system" for educators: a proof of concept. Computers in Education 54, pp. 588–599 (2010).
13. Hlosta, M., Zdrahal, Z., Zendulka, J., Ouroboros: early identification of at-risk students without models based on legacy data. In: Proceedings International Learning Analytics & Knowledge Conference (LAK'17), ACM, (2017).
14. McGraw Hill ALEKS PPL: Pave the path to graduation with placement, preparation, and learning, last accessed 2021/02/2011.
15. Pedregosa et al.: Scikit-learn: machine learning in Python. JMLR 12, 2825–2830 (2011).
16. CatBoost is a high-performance open-source library for gradient boosting on decision trees. catboost.ai, last accessed on 2021/02/11.
17. Prokhorenkova et al.: CatBoost: unbiased boosting with categorical features. arXiv preprint arXiv:1706.09516 (2017).
18. Permutation feature importance. https://scikit-learn.org/stable/modules/permutation_importance.html, last accessed 2021/02/11.
19. Strobl C., Boulesteix A., Kneib T., Augustin T., Zeileis, A.: Conditional variable importance for random forests BMC Bioinformatics 9(1), 1-11 (2008).
20. Strobl, C., Boulesteix A., Zeileis A. HothornT.: Bias in random forest variable importance measures: Illustrations, sources, and a solution. BMC Bioinformatics 8(1), 1-21 (2007).