MOL2NET, International Conference Series on Multidisciplinary Sciences
http://sciforum.net/conference/mol2net-03

Creating a Model to Predict Student Success Using WeBWorK Data

Jose Muguira (E-mail: jose.muguira001@mymdc.net) [a],
Reinaldo Sanchez-Arias (E-mail: rsanchez-arias@stu.edu) [b]

[a] *Miami Dade College, Wolfson Campus, Miami FL, USA*
[b] *School of Science, St. Thomas University, Miami Gardens, FL, USA.*

Graphical Abstract

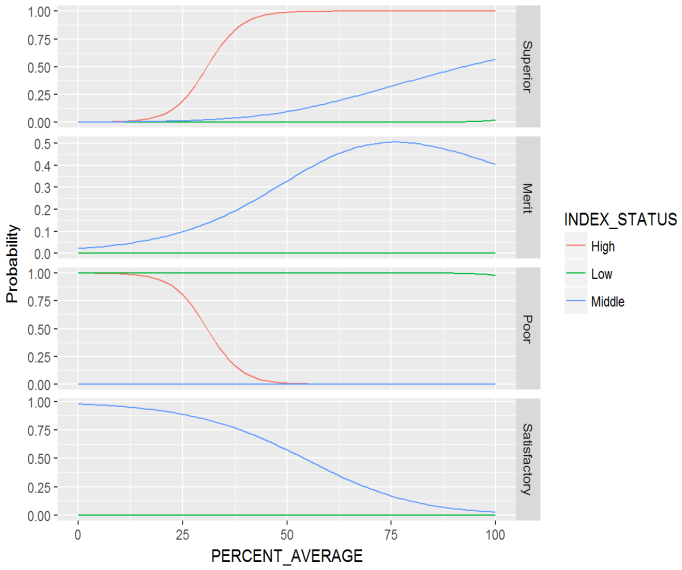

Fig. 3: Multinomial Logistic Regression model for a Calculus I course.

Abstract.

Student success is a major focus in the educational system, where a variety of predictors are used to estimate and measure how well students do in their different classes at the end of the academic year. Our research project aims towards proposing a model capable of demonstrating how student success can be predicted based on a series of indicators gathered from work submitted by the student throughout the semester. We studied the student's performance in an online homework assignment system for a mathematics course, taking into account the final score in a given assignment, but also the number of times every problem was tried by the student before obtaining a correct answer.

Keywords: *WeBWorK, data mining, logistic regression, predictive analytics.*

Introduction

For some of the mathematics courses at St. Thomas University, instructors use the open-source online homework system WeBWorK [1]. WeBWorK is supported by the Mathematical Association of America (MAA) and the National Science Foundation (NSF), and it provides students with a system that lets them work at their own pace and helps reinforce the different topics covered in a given course.

Results and Discussion

WeBWorK stores information of the scores for each assignment, as well as a variable called "*success index*", an indicator of the number of attempts made by the student to complete the assignment. This information was used to create our data sets, run a clustering analysis and differentiate between students that are very efficient in the assignments from those who need more trials before completing a given homework set. Data from one Pre-Calculus and two Calculus I courses at St Thomas University was anonymized and the open source statistical programming language R [2] was used for the data analysis. The main indicators used were labeled the "Homework Percent", which represents the score obtained by a student in a given assignment, and the "Homework Index", recorded by WeBWorK to measure the success indicator for the corresponding problem set. We used all the records from the Calculus course

and created a graph to help us visualize the different groups in the students. As observed in Fig. 2 we identified four groups, using a *K*-means algorithm in R: Students with high index and percent, students with average scores, students with low index but a high percent and students with low index and low percent.

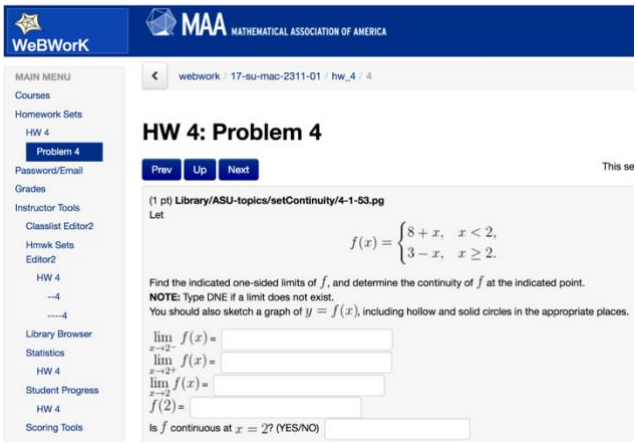

Fig. 1: Sample calculus assignment using WeBWorK

After all the data was successfully processed and having identified patterns, we created a mathematical model to interpret this analysis, grouping the grades into 4 categories: *Superior*, *Merit*, *Satisfactory*, and *Poor*. We divided the average of the index for every student into the categories of *High*, *Middle*, and *Low*. Using this information, we created a multinomial logistic regression model, capable of calculating the probability of a student to succeed on the course based on the average homework percent and the average homework index. The four different categories created can be then translated into an estimated final grade in the course. The method studied in this project showed high accuracy in the predictions and allowed for an easy interpretation of the model outputs as shown in Fig. 3.

Conclusions

In this research project we designed a model capable of predicting a student's probability of success in a given course, based on student's work submitted throughout the semester in the form of online homework assignments. The model takes into account the student's scores in all assignments during a semester, as well as the student's "success index" per assignment, a fairly good indicator of how well the student is grasping the concepts evaluated in every assignment.

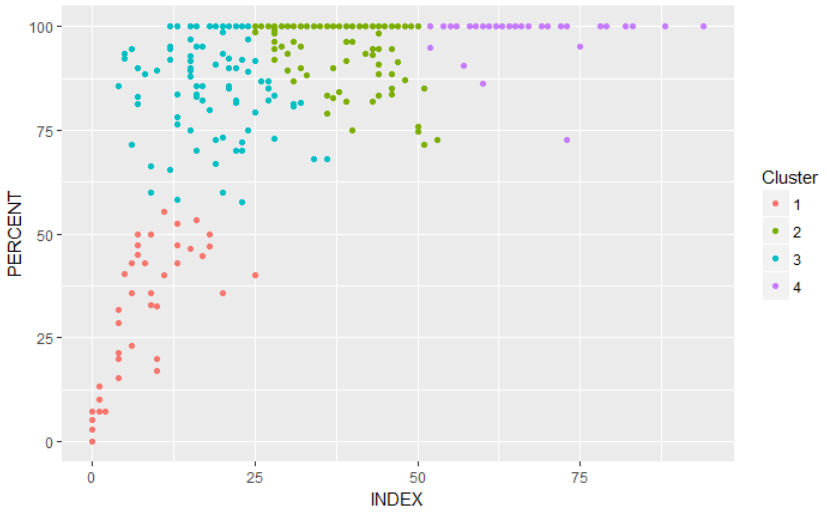

Fig. 2: K-means clustering of the different work submitted by students in a Calculus I course.

The model can be extended to include other predictors, and additional observations could be easily added to our framework for an even more robust model and prediction accuracy.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

Authors want to thank St. Thomas University facilities for completing this work during the SRI 2017. This project was supported, in part, by U.S. Department of Education grant award P03C1160161 (STEM SPACE), P031c160143 (STEM EngInE), P120A160036 (STEM ISLE), 1161177 (STEP Up), P120A140012 (SPARC).

References

1. WeBWorK. Mathematical Association of America (http://webwork.maa.org/)
2. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (http://www.R-project.org/)