Fall 9-11-2020

Mining High School Data to Predict and Increase Student Success in College

Timothy Barczak
Ohio State University, barczak.7@buckeyemail.osu.edu

Hemant Jain
University of Tennessee at Chattanooga, hemant-jain@utc.edu

Keith White

Follow this and additional works at: https://aisel.aisnet.org/sais2020

Mining High School Data to Predict and Increase Student Success in College

Timothy Barczak
Ohio State University
timbarczak11@gmail.com

Hemant Jain
University of Tennessee at Chattanooga
hemant-jain@utc.edu

Keith White
Public Education Foundation
KWhite@pefchattanooga.org

ABSTRACT

In this research we work with local educators to develop and assess predictive data dashboards that can be used to help increase student success in high school and college. Starting with longitudinal data from previous academic years, we develop predictive models that can reliably determine student archetypes. With the help of local teachers, in-school counselors, and administrators we built dashboards that leveraged current school student data to point educators towards interventions that are most pertinent to specific students.

KEYWORDS

Education, Data Mining, SAS, Community Partnership

EXTENDED ABSTRACT

In United States alarmingly low percentage of students entering postsecondary institutions persist and complete a degree. These rates for low-income and minority background students are significantly lower. Recent studies coming from Seattle and Chicago have elucidated this issue and have demonstrated the need for understanding how to best serve these populations (Grubb & Hernandez, 2018) (Nagoaka & Seeskin, 2018). Grigorenko et al., (2009) validated the use of measures such as GPA and self-efficacy measures in predicting academic outcomes; however, the study was limited to a subset of students already deemed high achievers in a broader academic context and only went so far as to estimate future GPA in a secondary schooling environment. Additional studies have also validated the use of test-scores and GPA as a predictor of success in a postsecondary environment. However, there is a long history of caution surrounding the use of test scores and GPA as the sole predictors of student ability and many colleges have begun making standardized test scores optional in the admissions process (Linn, 1990) (Hoffman & Lowitzki, 2005) (Zwick & Green, 2007). Thus, one is able to find a wealth of inquiry into student success factors. However, much of this research focuses on single factors within a student's academic career instead of a holistic assessment of the entire spectrum of interactions within the education system.

We use data mining approach on large quantity of longitudinal data to develop models for predicting the probability of a student completing a degree at postsecondary institution based on his/her complete high school record. We use these models to build a dashboard that can be used by teachers and principals to serve as an early warning system and allows them to take corrective action.

Working with a local school district and a public education foundation we focus on four schools with predominantly low-income and minority student population. We collected longitudinal data from students who graduated from these high schools between 2011 and 2017. We pair that data to corresponding National Student Clearinghouse information regarding matriculation, persistence, and graduation from colleges. Using SAS Enterprise Guide and SAS Enterprise Miner, we conducted multiple data mining processes including decision trees, regression analysis, neural networks, and a combination thereof to identify common patterns of student success. We then developed dashboards based off of these models using current student information to give teachers and administrators a quick and easy way to assess student progress and take corrective actions to ensure successful outcome. The dashboard has gone through multiple iterations based on feedback received from users. We evaluate the usefulness and effectiveness of dashboard by interviewing users.

With an understanding of the ethical implications of student tracking, we worked with schools throughout the implementation process to ensure the dashboards were used in an ethical manner. The intent of these dashboards were not to section students into different factions based on their current performance, instead, the aim was to give practitioners the resources to quickly identify the areas in which students need the most support for success.

REFERENCES

1. Grubb, B., & Hernandez, J. (2018). To and Through: Community and Technical College Pathways in South Seattle and South King County. Seattle, WA: Community Center for Education Results.

2. Grigorenko, E. L., Jarvin, L., Sternberg, R. J., Diffley III, R., Goodyear, J., & Shanahan, E. J. (2009). Are SSATs and GPA Enough? A theory-Based Approach to Predicting Academic Success in Secondary School. Journal of Educational Psychology, 101(4), 964-981.

3. Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting College Success with High School Grades and Test Scores: Limitations for Minority Students. The Review of Higher Education, 28(4), 455-474. doi:https://doi.org/10.1353/rhe.2005.0042

4. Linn, R. L. (1990). Admissions Testing: Recommende Uses, Validity, Differential Prediction, and Coaching. Applied Measurement in Education, 3(4), 297-318.

5. Nagoaka, J., & Seeskin, A. (2018). The educational attainment of Chicago Public Schools students: 2017. Chicago, IL: University of Chicago Consortium on School Research.

6. Zwick, R., & Green, J. G. (2007). New Perspectives on the Correlation of SAT Scores, High School Grades, and Socioeconomic Factors. Journal of Educational Measurement, 44(1), 23-45.