Abstract for an Invited Paper
for the APR20 Meeting of
the American Physical Society

Using Machine Learning and Big Data to Understand the Retention of STEM Students[1]

JOHN STEWART, West Virginia University

Retention of STEM students is a critical national problem. Introductory physics classes play a key role in the retention of these students. Machine learning algorithms including decision trees and random forests are applied to understand the variables important in predicting retention through the first year of college. This analysis identifies being a successful student in high school and arriving on campus calculus-ready as critical predictors of success. The students progression through the network of introductory science and mathematics courses is then explored. Machine learning algorithms are applied to understand a students risk factors as they matriculate from Calculus 1 and Chemistry 1 through Physics 1 and Physics 2. This will show students who matriculate through the network along different paths have different risk factors and chances of success.