Understanding Factors of Engineering Student Persistence Using Predictive Modeling

Dr. Daniel P. Kelly, Texas Tech University

Dr. Daniel P. Kelly is an Assistant Professor of STEM education at Texas Tech University in the Department of Curriculum and Instruction. He earned his doctorate in Technology Education from North Carolina State University where he also served on the faculty. Previously, he worked as a middle and high school science, technology, and engineering teacher in North Carolina. Dr. Kelly serves as the Associate Editor of the Engineering Design Graphics Journal and Editor-in-Chief and Founder of the Journal of Foster Care. Dr. Kelly studies how STEM education and engagement can improve the educational outcomes of students at risk of not completing high school due to academic, behavioral, or social needs. Of particular interest are children in foster care and other non-parental custody arrangements.

Dr. Jeremy V. Ernst, Embry-Riddle Aeronautical University

Dr. Jeremy Ernst is Professor of Technology and Associate Chancellor for Research within the Worldwide Campus at Embry-Riddle Aeronautical University. He has had prior academic and administrative appointments at Virginia Tech as well as North Carolina State University. His efforts center on curriculum research and development in STEM education to provide evidence-based models that promote engagement, development of cognitive competency sets, and performance-based application abilities of students at-risk.

Dr. Aaron C. Clark, North Carolina State University at Raleigh

Aaron C. Clark is a Professor of Technology, Design, and Engineering Education within the College of Education, as well as the Department Head for the Department of Science, Technology, Engineering and Mathematics Education. He has worked in both industry and education. Dr. Clark's teaching specialties are in visual theory, 3-D modeling, technical animation, and STEM-based pedagogy. Research areas include graphics education, game art and design, scientific/technical visualization and professional development for technology and engineering education. He is a Principle Investigator on a variety of grants related to visualization and education and has focused his research in areas related to STEM curricula integration.

Mr. Erik Schettig, North Carolina State University at Raleigh

Erik is a Ph.D. student in the Learning and Teaching in STEM program at NC State University with a focus on Engineering and Technology. He has over 10 years of K-12 teaching and education outreach experience. In addition, Erik has been a director for STEM Summer programs and lead instructor for NC State pre-college programs.

Understanding Factors of Engineering Student Persistence Using Predictive Modeling

Abstract

Student persistence in higher education is a topic of discussion in the academic literature and within our colleges and universities. This is especially relevant as university programs continue to focus on equity, inclusion, and support for student populations that are historically underrepresented in higher education and within specific disciplines. Engineering education has been attempting to address these issues for some time and with the graduation rates for engineering programs averaging up to 50%, understanding why students stay or leave these programs is crucial information. The reasons students persist or leave higher education programs are important data points for any university program. However, traditional statistical analysis methods may not be robust or accessible enough to understand and communicate these factors. To determine these factors, machine learning and predictive analysis software were employed to examine these factors of persistence for engineering education students. Dozens of variables including academic scores, non-cognitive and skill-based assessments, and demographic information for 300 students in an introductory engineering graphics course were used to develop a model capable of predicting whether a student will persist with nearly 94% accuracy. This research indicated that age, gender, three-dimensional modeling self-efficacy, and parental career were the most influential factors of persistence. Using this information, combined with the theoretical underpinnings of these constructs, may provide areas in which to focus and specifically target in order to improve persistence rates in engineering education.

Introduction

Compared to other degree programs, persistence rates of undergraduate engineering programs are low. Engineering programs have up to 50% of students who persist and graduate with an engineering degree whereas persistence rates of other majors such as education (81%), business (80%), and humanities (64%) are greater [1], [2], [3]. Programs develop curricula, establish course structures, provide resources, and implement support intended to improve student persistence [4], [2], [5]. When analyzing underrepresented minorities in engineering programs, persistence rates are even lower [5]. The purpose of this study was to identify factors of persistence in engineering programs and to identify how this information can be applied to develop instruction, support mechanisms to increase persistence. Traditionally, the first year of engineering programs is seen as a time to "weed out" students who may not meet program standards or who may possess misconceptions about engineering; however, it is also possible that students dropped out of the program due to insufficient support and resources [1].

Such insufficiency or lacking of resources may impact the self-efficacy of students as well as the demographics of future engineering workplaces when considering the low rates of persistence among underrepresented populations [4], [5]. Adding further support for increasing overall persistence rates may fulfill gaps in engineering workplaces due to a possible increase in qualified individuals [4], [5]. In engineering fields, underrepresented populations consist of women as well as individuals who are African American, Hispanic American, and Native American [5], [6], [7]. Along with underrepresented populations, students at risk of non-matriculation also require support to promote persistence [8]. While it may vary between institutions, students at risk of non-matriculation can be those whose GPA falls below a certain point, such as 2 on a four-point scale [8]. Students pursuing engineering degrees come from various backgrounds and possess various factors that can impact their chance of persistence. Understanding how these factors impact persistence can enable the development of practices, curriculum, textbooks, supports, and environmental conditions that can positively influence self-efficacy as well as engagement to increase persistence in engineering programs [9], [10].

Method

A sample of 300 introductory engineering graphics course students completed surveys which revealed that 75% (225 students) of the sample remained enrolled in engineering majors three semesters after taking the course. From the survey, factors of persistence were identified through logistic regression which resulted in several significant variables, including student score on the pretest three-dimensional modeling self-efficacy (3DSE) assessment, gender, age, and whether or not the student had a parent with professional engineering backgrounds. The three-dimensional self-efficacy instrument consisted of nine questions, each being a 7-point Likert type item, designed to measure students' self-efficacy related to modeling three-dimensional objects [11]. Logistic regression could not identify for which subgroups of students the variables were most significant. For these reasons, machine learning analytics software was used to examine the predictors, and their interactions, that led to persistence in engineering degree programs. Machine learning has gained popularity over recent years due to its ability to store multiple algorithms and use that information to draw connections as well as make predictions across various platforms [12]. Essentially, the computer machine can learn algorithms to analyze large amounts of data.

Survey responses and institutional data were inserted into machine learning data mining and visualization software for analysis. This technique analyzed the impacts of variables as well as their association with persistence. Additionally, software analysis was utilized to classify groups for which distinct variables had the most significant impact using a classification and regression tree model. Figure 1 shows the classification tree as displayed by the software.

Figure 1.
Example of a classification tree derived by the analytics software.

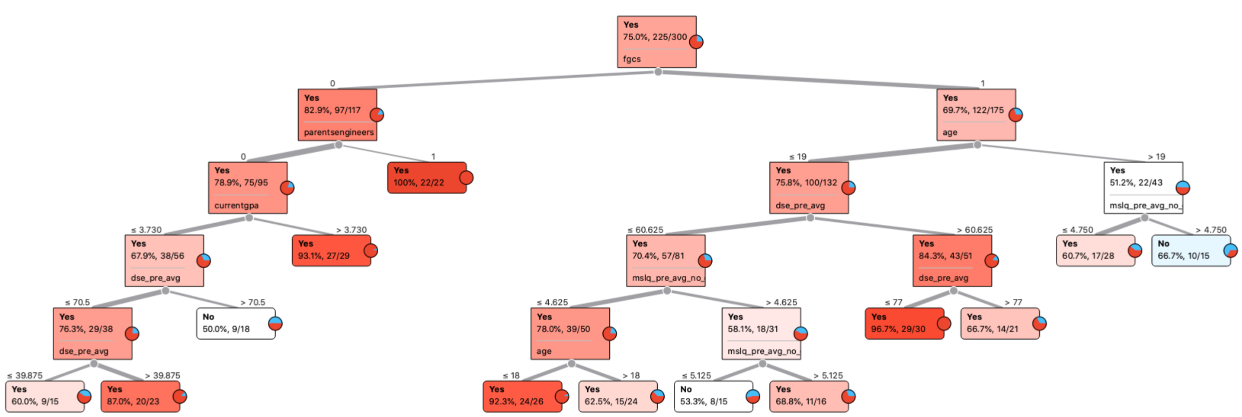

Figure 1 displays all initial variables with first-generation college students as the root node. The classification tree splits the observations into binary categories based on the variable values in each observation. In figure 1's case, the binary classification is persistence in engineering [Yes] and non-persistence [No]. For categorical variables, the split is which variable value exists within a particular observation. For continuous variables, the algorithm applies a regression analysis that determines the splitting point at a mathematically logical point. This model provides the advantage of adaptive ability and will self-adjust with new data. The darker color implies a great proportion of persistence. Next to the elements are icons that provide a visual of the proportion of persistence. As part of the analysis, random samples, 70% of the observations, were used as training data with the remaining 30% reserved to test the model. This process was repeated ten times to obtain a classification accuracy of the model of 70.3%.

Results

The tree identified in figure 1 was simplified through pruning where misclassified or non-influential variables were removed. In this instance, 12 variables were reduced to four, and the model's accuracy of predicting persistence was increased to 93.8%. The remaining variables that had elevated levels of predicting persistence included gender, parent(s) is/are engineers, age, and three-dimensional modeling self-efficacy (pre-test). Factors associated with non-persistence are nearly limitless and not easily specified. Influences of these factors can include personal, social, academic, financial, and health elements [13], [14]. Additionally, the factors can be influenced by the uniqueness of various populations and individuals. For these reasons, only factors of persistence can be reliably assessed with the current data. When looking at predicting non-persistence, the model was only able to predict non-persistence with 22.2% accuracy (see table 1)

Table 1
Confusion Matrix for Pruned Classification Tree (Four Variables)

		Predicted	
		No	Yes
Actual	No	**22.2%**	77.8%
	Yes	6.2%	**93.8%**

Discussion

The model identifies factors that can predict persistence with no single variable predicting student persistence nor the ability to accurately predict non-persistence. The use of conventional analysis techniques or software was not able to effectively identify the distinctive factors related to persistence or non-persistence as nested groupings were not apparent, the comparison of all possible variable combinations was not practical or feasible, and the point within variables where differences in persistence rates were significant could not be determined. In addition, persistence rates based on categorical variables are inaccurate due to nested groupings and a lack of sensitivity for non-binary variables. Information on identified factors can apply to the improvements of course structures, environments, resources, and support to possibly mimic effects of factors that are unable to be directly addressed. An understanding of how and why these factors impact persistence can be developed through study and further implemented in potential or already existing delivery of information as well as the structure of programs such as mentoring or outreach.

Increased study related to the elements and complexity of variable associations is required to further identify connections of variables that lead to persistence or non-persistence. Further studies should focus on the variables that lead to the successful persistence of engineering programs instead of variables that lead to non-persistence. While factors, such as whether a parent has an engineering background, provide more rationale for an impact on persistence, other variables were not as evident.

Conclusion and Limitations

The objectives of this study were to identify factors of persistence in engineering programs that can establish a foundation for further research as well as a framework for improving persistence rates in engineering programs. A unique model was developed to explore the variables that most impact persistence in engineering programs three semesters after an introductory engineering graphics course. Variables identified in the study as having the greatest impact were three-dimensional modeling self-efficacy, gender, age, and having a parent with an engineering background. The developed model is able to predict persistence, among the studied population, with a high degree of accuracy; however, there were limitations in the ability to predict non-persistence. Using machine learning in conjunction with classification modeling, an effective method to visually represent and identify subpopulations presented value in predicting persistence. The lack of literature using such methods in engineering degree programs provides an opportunity for future research using these as well as similar methods. Future research is needed to further analyze the interconnectedness of the variables and to identify additional variables that increase model accuracy. Additionally, using a larger sample size and a more diverse population, further evaluation can broaden the application of these results. The present results should be considered exploratory and interpreted within the context of study limitations. A manuscript is in development with more detailed information related to the theoretical underpinnings of the variables, suggestions for the specific use of the information, and further detail into the methods used. Details are limited in this format and this paper is meant to introduce a larger project to this audience.

This material is based upon work supported by the National Science Foundation under Grant No. 1900348.

References

[1] B. Christe & C. Feldhaus., "Exploring Engineering Technology Persistence and Institutional Interventions: A Review of the Literature," in *Journal of Engineering Technology*, vol. 30. pp. 44-51, Oct. 2013.

[2] B. King., "Changing College Majors: Does it Happen More in STEM and Do Grades Matter?," in *Journal of College Science Teaching*, vol. 44, pp. 44–51, Jan. 2015. [Online]. Available: https://doi.org/10.2505/4/jcst15_044_03_44 [Accessed Dec. 15, 2020].

[3] A. Patrick, M Borrego., & A.N. Prybutok., "Predicting Persistence in Engineering Through an Engineering Identity Scale," in *International Journal of Engineering Education*, vol. 34, pp. 351-363, Jan. 2018. [Online]. Available: https://par.nsf.gov/servlets/purl/10066205. [Accessed Dec. 15, 2020]

[4] M. Graham, J. Frederick., Byars-Winston, A., Hunter, A., & Handelsman, J., "Increasing Persistence of College Students in STEM," in *Science*, vol. 341, pp. 1455-1456, Sep. 2013. [Online]. Available: http://www.jstor.org/stable/42619390 [Accessed Dec. 15, 2020].

[5] A. Lisberg, & B. Woods., "Mentorship, Mindset and Learning Strategies: An Integrative Approach to Increasing Underrepresented Minority Student Retention in a STEM Undergraduate Program," in *Journal of STEM Education: Innovations and Research*, vol. 19, pp. 14-20, (2018). [Online]. Available: https://www.jstem.org/jstem/index.php/JSTEM/article/view/2280. [Accessed Dec. 16, 2020].

[6] D. MacPhee, S. Farro, and S.S. Canetto., "Academic Self-Efficacy and Performance of Underrepresented STEM Majors: Gender, Ethnic, and Social Class Patterns," in *Analyses of Social Issues and Public Policy*, vol. 13, pp. 347-369, Dec. 2013. [Online]. Available: https://doi.org/10.1111/asap.12033. [Accessed Dec. 16, 2020].

[7] C. Vallas & S. Donohue., "Identifying Factors Affecting Persistence Rates Among Undergraduate Engineering Students From Underrepresented Populations at the University of Virginia," in *American Society for Engineering Education-ASEE*. June 2007. [Online]. Available: https://peer.asee.org/2925. [Accessed Dec. 15, 2020].

[8] S. Gershenfeld, D. Ward Hood, & M. Zhan., "The Role of First-Semester GPA in Predicting Graduation Rates of Underrepresented Students," in *Journal of College Student Retention: Research, Theory & Practice*, vol. 17, pp. 469–488, April 2015. [Online]. Available: https://doi.org/10.1177/1521025115579251. [Accessed Dec. 18, 2020].

[9] N.A. Mamaril, E.L. Usher, C.R. Li, D.R. Economy, & M.S. Kennedy., "Measuring Undergraduate Students' Engineering Self-Efficacy: A Validation Study," in *Journal of Engineering Education*, vol. 105, pp. 366–395, April 2016. [Online]. Available: https://doi.org/10.1002/jee.20121. [Accessed Dec. 16, 2020].

[10] R.L. Navarro, L.Y. Flores, J.P. Legerski, J. Brionez, S.F. May, H.N. Suh, D.R. Slivensky, F. Tapio, H.S. Lee, P.O. Garriott, H.K. Hunt, C.D. Desjarlais, B.H. Lee, D. Diaz, J. Zhu, & A.K. Jung., "Social Cognitive Predictors of Engineering Students' Academic Persistence Intentions, Satisfaction, and Engagement," in *Journal of Counseling Psychology*, vol. 66,

pp. 170–183, Mar. 2019. [Online]. Available: https://doi.org/10.1037/cou0000319. [Accessed Dec. 15, 2020].

[11] C. Denson, D. Kelly, & A. Clark., “Developing a Scale to Investigate Student’s Self-Efficacy as it Relates to Three-Dimensional Modeling,” in *The Engineering Design Graphics Journal*, vol. 82, Fall 2018. [Online]. Available: http://www.edgj.org/index.php/EDGJ/article/view/705. [Accessed April 14, 2021]

[12] M.J. Petrilli., “Big Data Transforms Education Research,” in *Education Next*, vol. 18, Jan. 2018. [Online]. Available: https://www.educationnext.org/big-data-transforms-education-research-can-machine-learning-unlock-keys-to-great-teaching/. [Accessed Jan. 10, 2021].

[13] W. Pearson., “Persistence of Adult Students,” in *The Journal of Continuing Higher Education*, vol. 67, pp. 13-23, Jan .2019. [Online]. Available: https://doi.org/10.1080/07377363.2019.1627166. [Accessed Dec. 16, 2020].

[14] G. Vaquera & R. Maestas., “Pre-College Factors Impacting Persistence at a Diverse University,” in *Journal of College Student Retention: Research, Theory & Practice*, vol. 10, pp.425-445, Mar. 2009. [Online]. Available: https://journals.sagepub.com/doi/abs/10.2190/CS.10.4.b. [Accessed Dec. 15, 2020].