International Journal of Business and Economic Development
Vol. 2, No. 4; December, 2021. pp 18-31
ISSN 2692-5591 (print), 2692-5583 (online)

Published by International Center for Promoting Knowledge

Application of Data Mining in Student Retention Strategy Using Decision Tree Module

Dr. Yang Huo
Hospitality Management/Organizational Leadership Department
Woodbury School of Business
Utah Valley University
United States

Abstract

This paper addresses the issue of why students want to drop out from a course and suggest appropriate strategies to enhance student retention. We empirically examine with a sample of 261 hospitality management students, drawing on Tinto's theory and applying a behavior on student departure. We apply the data mining &decision tree using Classification and Regression Trees method as an analytic tool to identify a group, discover relationship between groups and predict future events and for segmentation. The results regarding the demographics indicate that the most effective attributes are job situation, earned credit hour. In need, our findings provide empirical evidence that financial situation and quality of teaching are the most effective attributes. Finally, we argue that students' financial situation and job availability during COVID-19 crisis can negatively impact on potential dropout decision and may be an important consideration of planned behavior.

Keywords: classification and regression trees (CRT), data mining, decision tree module, self-efficacy student retention, theory of planned behavior

Introduction

One of the biggest challenges that higher education faces is to improve student retention. Student retention has become a critical indicator for academic performance and enrollment management as well as the student success. The figures on student dropout in the United States provide an example of the magnitude of this problem, with the overall dropout rate for undergraduate college students is 40%, with approximately 30% of college freshman dropping out before their sophomore year (Hanson, 2021). Despite steady efforts for enrollment rates in higher institutions, failed student retention and high dropout rates remain persistent problems for institutions. The issue of student retention in higher education has recently received considerable attention in the academic literature. The importance of student retention relates to student dropout in higher education reflected in academic research has been carried out in recent years. Some scholars even claim that student retention and its strategy is highly correlated with institutional environments and student's demographics (Aljohani, 2016: Kerby, 2015: Mah, 2016: Lin, 2012: Tight, 2020). Olaya, et al., (2020) assert that student dropout is a genuine concern in private and public institution in higher education because of its negative impact on the well-being of students and the community in general. According to Seidman (2012), "students dropping out of college miss the opportunity afforded by universities to develop their critical thinking skills, will tend to earn less in their careers, often leave university with loans to be repaid; universities with retention problems suffer a significant loss of income; and for countries, higher education systems which can increase social mobility and provide the specialized intellectual and skills required in the 21st century are undermined by high levels of student attrition (p.2)." These claims have increased our attention to determine the reasons to dropout and retention strategy by observing from the student's perspective. The issue of student retention in higher education has recently received considerable attention in the academic literature. Student retention in higher education has been hypothesized to be a survival tool and asset because it is believed to: (1) lower student dropout rates and allow for successful strategy to respond to changing student learning environments and (2) lead to superior information sharing that improves learning approach and joint efforts between the instructor and students to minimize inefficiencies in learning cycle.

Dewberry & Jackson (2018) state that even though extensive literature has accumulated on the possible causes of student dropout, with the number of studies running into thousands (Berger, Ramírez, & Lyons, 2012), there has been little development in cognitive approach determining the relationship between the institutional and personal environments. Dewberry & Jackson (2018) also argue that the development that has taken place (Braxton et al., 2014;Guiffrida, 2006; Tinto, 1993) has generally focused on revisions to the dominant, sociological, student integration theory of retention developed by Tinto, 1975, Tinto, 1987and is based on systematizing and organizing research findings rather than on a set of underlying theoretical premises and little research has been able to measure a "causal effect." (e.g., Nora & Crisp, 2012a). However, does student retention really pay off in reducing student dropout, or does this retention strategy to student dropout bring a positive effect to the institutions as well as students achieving their academic goal and benefits? Although the literature on student retention is well developed, they collected data from the students who already had dropped out and had left the institution and obtained and observed the data from the institutional point of view. The empirical research on the students' perspectives who enroll currently is lacking. In other words, research techniques used by institutions are driven by institutional data and analytic methods rather than the surveys and theories (Delen, 2012). While possibly important for explaining reasons to dropout, the students' personal data are not available at every institution due to data protection and privacy act/law, information on student level of satisfaction for the course, individual fit of the institutional framework and diligence are not fully described (Berns et al., 2018).

Given that voluntarily dropping out of college involves an active choice and a range of cognitive and behavioral adjustments related to individual situation, learning environments and internal environments has considerable potential in explaining why some students complete their studies whereas others drop out (Dewberry & Jackson, 2018). To be successful to determine reasons to dropout and develop an appropriate retention strategy, it is critical and valuable to identify those students that are at risk of dropping out by constructing survey instruments and applying the qualitative, behavioral, and survey-based studies driven by theories. While most institutions do administer surveys to their students at various time during their academic career, these instruments meet other institutional goals and are usually not specifically designed to support theoretically based retention (Caison, 2007).

The purpose of this article is to focus on student's self-efficacy, internal situations/environments to determine how well a student can execute courses of action (i.e., dropout) required to deal with prospective external situations (institutional environments and learning environments) by applying a theory of planned behavior (TPB) (Ajzen, 1991). More specifically, we seek to answer the following questions and identify why students drop out from the course/program, why they transfer or leave college or university. We also explore empirically the extent to which external environmental factors create educational value for students and determine intervention options and appropriate strategies to enhance student retention for hospitality management major students and hospitality management program. In other words, our goal is to determine the reasons to fail/leave the course and program and to augment and develop the retention strategies in the ways theorized in the academic literature through a decision tree method by collecting a data from student who enroll currently. More specifically, we apply the data mining, which is the process of extracting valuable knowledge from a large amount data and a decision tree. The decision tree is one of classification techniques and analysis methods of data mining tools to identify a group (i.e., institutional environments, learning environments, self-efficacy of student, and student's demographics), discover relationships between groups and predict future events and for segmentation, stratification, data reduction and variable screening, interaction identification, and category merging. Furthermore, it provides a powerful classification algorithm to construct a tree for improving the prediction accuracy for the dropout and retention strategy.

Theoretical Framework & Literature Review

We draw on prior literature in defining retention and dropout as *retention is staying in school until completion of a degree and dropping out is leaving school prematurely. It seems simplistic that retention and dropout are just purely opposites* (Hagedorn, 2006). He also stresses that a dropout would then be defined in comparison to student outcome versus original intent and it is only when students leave college before achieving their goals that they should be labeled a dropout. This dropout would be explained as leaving course/program temporarily not permanently due to a surrounding environment of student and intention to come back when the environment situation is over or meet the expectation. Conceptually, institutions are not able to control student dropout; dropout is a sensitive and complex level phenomenon and has its foundation in individual students.

Research in student retention conducted for decades since student retention has become a critical indicator for academic performance and enrollment management of the university (Zhang et al., 2010). For a long time, student dropout and retention studies relied heavily on individual students' attributes. Bean's (1980) model of student retention identifies a causal relationship between student satisfaction in a course and institutional determinant that led to commitment to stay in a course or withdrawal. The previous research suggests that there are variables measured to predict to what degree a student will stay or withdraw a course (Wojciechowski & Palmer, 2005; Wray et al., 2012). Universities should build an intervention program that will target specific problems. Tinto cites five conditions that bet promote retention (Tinto, 2000): 1) having high expectations of students; 2) clearly explaining institutional requirements and providing good advice about academic choices since many students are not clear about their plans and need help in building a road map; 3) providing academic, social and personal support, particularly in and before the first year; 4) showing students that they are valued. Frequent contact with the staff is important, especially in the first year; 5) active involvement in learning - students who learn are students who stay and social learning, where students learn in groups, is particularly valuable, and can help foster friendship, which is another factor that encourages student persistence. The teachers who have positive attitudes towards the course or materials show greater tendency and enthusiasm to their jobs and their classes and students so that they can retain their students in the class and reduce the dropout rate eventually (Toraby&Modarresi, 2018).

Historically, academicians and institution administration have observed student retention as a critical tool to success in their program and institution as well as the students' academic success by observing sociological perspective in the context with academic institutional environments. Hermanowicz (2003) stated that the "structural sides" of universities (e.g., admission and transferred credit) and "cultural sides" (e.g., financial support and quality of learning) are both important elements since a higher rate of retention is often achieved when students find the institution environments in their university to be highly correlated with their interests which is major component of students' self-efficacy and their internal environments. Dyk&Weese(2019) recommended that the campus recreation program would assist undoubtedly assist in advancing the strategic imperative and make a significant difference in supporting higher education aspirations for indigenous students. Fong et al., (2017) suggested that cultivate psychological factors, such as "mind-set and achievement" are successful retention factors and student success. Sahin et al., (2016) insisted that absenteeism is one of the most basic indicators of to what extent the educational needs of students are met by schools regarding school dropouts and in analyzing retention. Sivakumar et.al. (2016), empties a decision tree algorithm in analyzing dropout factors. Ropeti (2015) focused on student perspectives as critical attributes for student retention and founded that student attitudes are critical factors for success as well as the teaching/learning environments. Mengo& Black (2016), determined a violence in the school and institutional environment as factors in dropping out. Kumar& Radhika (2014) examines multiple factors, including family factors, in determining success in school. Latif et al., (2015) found that economic effects such as financial problem is the most critical factor contributing to student dropouts. Peguero& Bracy (2014) determined that severe disciplinary practices create higher than necessary dropout rates and those alternative disciplinary measures can be more positive for retention and at the same time provide safety in the schools. Bean & Eaton (2001-2) proposed the psychological processes that lead to academic and social integration-based model that describes how successful retention program such as learning communities, freshman interest groups, tutoring, and orientation rely on psychological processes. These scholars' studies on student retention and its relationship with student dropout and their empirical studies' methods and findings are shown in Table 1.

Table 1 Evolution of Student Retention and Dropout Literature

Title	Authors/year	Methodology	Findings/recommendations
The Undeniable Role That Campus Recreation Programs Can Play in Increasing Indigenous Student Engagement and Retention	Chad Van Dyk and W. James Weese(2019)	Literature bases/content analysis	1. Create a campus recreation center & indigenous student advisory council. 2. Engage this council in the creation and evaluation of promotional materials and activities. 3. The group may also decide to design and deliver campus recreation program outreach activity days in local indigenous communities.
Psychosocial Factors and Community College Student Success: A Meta-Analytic Investigation	Carlton J. Fong and Young Wan Kim (2017)	Literature bases & meta-analysis	1.Motivation and self-perceptions were the most influential predictors for both achievement and persistence outcomes. 2. Regarding self-regulation as a predictor, there was a weak correlation associated with persistence but a stronger correlation with achievement, which was in line with previous meta-analytic results. 3.Regarding attributions, we found that for samples with greater percentages of minorities, internal attributions had a negative influence on community college achievement.
Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree	Subitha Sivakumar, Venkataraman and Rajalakshmi Selvaraj(2016)	Data mining/ decision tree algorithm	Experimental results proved that improved decision tree algorithm provides better prediction accuracy in educational data than that of traditional classification algorithms in the literature.
Causes of Student Absenteeism and School Dropouts	ŞeymaŞahin, Zeynep Arseven, Abdurrahman Kılıç(2016)	Semi-structured interview/content analysis	1. Establish a solution to prevent absences2. monitoring 3. Ministry of Family and Social Policies4. positive school atmosphere. 5. Elective courses should be varied depending on the interests. 6.Their anxieties of failure should be tried to be prevented.7. Cooperation between the school and the family. 8. Further larger-scale quantitative studies should be conducted on absenteeism and school dropouts.
Economic Effects of Student Dropouts: A Comparative Study	Latif, A.Choudhary, A., &Hammayun, A. (2015)	Literature bases/content analysis/comparative analysis	1. financial & economic positions of family 2. Lack of furniture, electricity, water and computer and science lab. 3. Establishment of teachers training programs4. Develop the interest of parent and students.
Student	SiamauaRopeti(	Open-ended	1. Student Positive attitude 2. Student

Perspectives Regarding School Failure at the American Samoa Community College	2015)	interview/typ ological analysis	Experience on learning 3. Student participation in class 4. Teacher teaching/learning environment
Violence Victimization on a College Campus: Impact on GPA and School Dropout	Cecilia Mengo and Beverly M. Black (2016)	Case files/paired sample t-test	sexual and physical/verbal victimization are associated with decreases in GPA. Students' GPA influences their retention.
A Survey on Predicting Student Performance	Dinesh Kumar,Dr.V.Ra dhika(2014)	educational data mining techniques	student's grade in senior secondary exam, living location, medium of teaching, mother's qualification, father's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.
The psychology underlying successful retention practices	John Bean, Shevawn Bogdan Eaton (2001)	Literature bases/content analysis	Develop program helps students1. develop positive attitudes toward the college and her/himself 2. learn strategies to approach new academic and social situations 3. feel academically or socially successful4. feel in control of their lives as students at their institutions

In this study, using a data collected from classes offered during a semester along with a decision tree data mining technique, we constructed an analytical algorithm to determine drop out reasons. In order to identify the important predictors, we adopt the TPB (i.e., self-efficacy) as an internal environment by utilizing a high impact engaged demand (HIED) survey. HIED is a survey of a student perceptions of teaching (SPOT). SPOT is a survey about to "increase the student retention and well-being in a course and to augment the course materials to meet the needs and demands of students: self efficacy"(Gallup: Virginia Tech, 2018). As implemented in the course, SPOT collects the perceptions of all elements foundational to HIEDs: high performance expectations, significant investment of time, interaction with peers, institutional environment, frequent and timely feedback, structured opportunities to reflect on learning, real-world relevance, and public demonstration of competence (Kuhet al., 2017). However, currently, many institutions have not conducted studies that measure the factors thought to contribute to improve learning and retention from a result of SPOT. HIED survey fosters and encourages instructors to understand the student needs and wants from the student's point of view called as a "pull" factor rather the course designed exclusively by the instructor called as a "push" factor in classroom setting and environments. In other words, the student and instructor engage in the learning process, and in the academic realm. As such, the practice and implementation of results and finding from HIED survey has much potential to improve retention and completion. In addition, it is beneficial for non-traditional, first-generation, and minority students and learn more about the benefits of diversity and how to work effectively in a diverse academic environment.

Methodology

Sample & Data Collection

The data for this study was collected from a single institution, a teaching-oriented state university located in the mid-west region of the United States with an average enrollment of 40,000 students, of which 85% are the residents of the same state and 81% are working while taking classes. A survey administered to students in most failed courses. This included six sections in spring 2020 and three sections in Summer First Block 2020, and three sections in Fall 2020 for a total of approximately 260 students.

Operational Measures

We chose a high impact engaged demand (HIED) survey to operationalize internal environments interaction of a variety of characteristics such as personal variables/self-efficacy, institutional variables, and learning variables to increase students' engagement in course content and learning outcomes. Students in a course complete a high impact engaged demand (HIED) survey to determine the student perceptions of teaching (SPOT). As part of this survey, students demonstrate how students and professors drive student engagement, reflect on their learning and teaching, evaluate the course modules and engagement skills, and set goals for academic improvement. The students complete the HIED survey twice throughout the semester; the first one was performed after the first exam and the second one was performed after the second exam. The course instructors review the HIED survey results and findings to determine how to augment and develop methods for increasing retention in a course. Furthermore, the instructors review the second results and findings to determine any progress in the course regarding the retention. The engaged demand survey and its outcome provide the opportunity to measure what degree a student completes a course. This research has the potential to specifically impact learning for identifying potential issues on why students want to drop out, why they transfer or leave the course not only in the department also across campus.

Self-Efficacy: demographic variables

(1) Personal: age, gender, residence, marital status, occupation.
(2) Previous education: previous academic experience, transferred credit, age of transfer.
(3) Study: course of study (major), type of enrollment (i.e., full or part time), numbers of credit per semester, GPA, workload, mid-term grade.

Each scale item was measured on a seven-point Likert scale (1 = not at all; 5 = to a very great extent).
Institutional environments

Consistent with previous studies, we operationalized institutional environments using multiple scale items designed to measure the extent to which the institutional environments provide influential factors to the students' reasons to dropout from the class as well as institutions. Institutional environments were operationalized as the sum of the following sub measures and attributes that are reflections of a single unidimensional construct:

(1) The extent to which the institution provides adequate financial support.
(2) The extent to which the institution offers appropriate major selection.
(3) The extent to which the institution provides appropriate facilities, classroom, lab, campus programs & activities

Each scale item was measured on a seven-point Likert scale (1 = not at all; 5 = to a very great extent).
Learning Environments

(1) The extent to which the program provides adequate learning/course content, module.
(2) The extent to which the faculty/instructor offers quality teaching.
(3) The extent to which the faculty/instructor encounter with students in proper manner.

Each scale item was measured on a seven-point Likert scale (1 = not at all; 5 = to a very great extent).

Data Analysis

We applied the data mining & decision tree (DT)/regression tree using Classification and Regression Trees (CART) method as an analytic tool to identify a group, discover relationship between groups and predict future events and for segmentation, stratification, data reduction and variable screening, interaction identification, and category merging the paths from root to leaf represent classification rules. Decision Trees (DTs) are a supervised learning technique that predict values of *responses* by learning decision rules derived from *features*. They are utilized in both a regression and a classification context. For this reason, they are sometimes also referred to as Classification and Regression Trees (CART) (James et al., 2013). Roiger (2017) defines data mining as the process of finding interesting structure in data and the structure may take many forms, including a set of rules, a graph or network, a tree, one or several equations and it uses one or several algorithms for the purpose of identifying interesting trends and patterns within data and the ultimate goal is to apply what has been discovered to new situations. Quinlin (1986) states that decision trees are part of the induction class of data mining techniques. Quadril&Kalyankar (2010) argue that modern data mining models such as decision trees can more accurately predict risk than current models, educational institutions can predict the results more accurately, which in turn can result in quality education. The decision tree module can help spot students 'at risk' in the context of major selection, course/class management by evaluating the course or module suitability, and tailor the interventions to increase student success and retention and improve student's performance. Specifically, DTs are analytic tool to identify a group (i.e., institutional environments, learning environments, and self-efficacy of students), discover relationships between groups and predict future events and for segmentation, stratification, data reduction and variable screening, interaction identification, and category merging. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g., whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules. Tree based learning algorithms are considered one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map nonlinear relationships quite well. They are adaptable at solving any kind of problem at hand (classification or regression). Decision Tree algorithms are referred to as Classification and Regression Trees (CART, Breiman et al., 1984). The Decision Tree/Regression Tree method is an analytic tool to identify a group, discover relationships between groups and predict future events and for segmentation, stratification, data reduction and variable screening, interaction identification, and category merging. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g., whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules. Tree based learning algorithms are considered one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map nonlinear relationships quite well. They are adaptable at solving any kind of problem at hand (classification or regression). Educational data mining is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context (Sivakumar et al., 2016). First, we ran the cross tabulation/contingency table to describe a basic picture of the interrelation between the demographics and dropout for gender, academic level, job, etc.

Second, we apply the data mining & Decision Tree/Regression Tree using CART (Classification and Regression Trees) method as an analytic tool to identify a group, discover relationship between groups and predict future events and for segmentation, stratification, data reduction and variable screening, interaction identification, and category merging the paths from root to leaf represent classification rules. It also monitors a student's profile, analyzes student academic behavior, and provides a basis for efficient intervention strategies. Moreover, we explore how data mining and decision tree modules can help spot students 'at risk' in the context of major selection, course/class management by evaluating the course or module suitability, and tailor the interventions to increase student success and retention and improve student's performance through CART. IBM SPSS Decision Trees 26 was used to run the data mining for creating a tree-based classification model.

Results and Discussions

The simple descriptive statistic for the cross tabulation displays the frequency distribution of dropout for students' demographics show are shown in Table 1. The descriptive statistics indicate that the following individual attributes when they had a dropout experiences: female students (68%) vs male students (32%), age (21-25, 56%), senior (64%), no job (48%), married (56%), Caucasian (76%), Utah resident (84%), self-financial support (57.3%), hospitality management major (44%), GPA (3.1-3.5, 56%), taking 11-15 credit Hours/semester (68%), respectively. The descriptive analysis implies that institution administration should pay more attention to a female age in between 21-25 with no job and self-financial support situation and with 31-3.5 GPA and taking 11-15 credit hours since this cluster/group/segment had a dropout experience, and it represents the chance to dropout in the future will be much higher than others. While it is not possible in this study to determine the in-depth cross sectional contribution weight that accounts for these relative dropout factors, the fact that demographics of students was strongly correlated to high dropout ratio and risk suggests that a student retention strategy for understanding demographics improves low dropout ratio. To confirm and verify the results of the descriptive analysis run through cross-tabulation for the students' demographics, we examined the decision tree analysis. The dependent variable of this decision tree is dropout experience that has two classes, yes or no. The root of this tree contains all 161 observations in this dataset. The results indicate the most influential attribute to determine whether to make a decision to dropout or not is the occupation attribute. The 21.21% students in our sample that had no occupation or no job indicates higher dropout ratio (48%) compared to those who are working and taking a course at the same time The institution should pay attention to this group or cluster (see Figure 2). The decision tree results also predict to classify new observations. Another new observation is the number of earned credit hours that shows students who have earned 91-120 credit hours represents the highest dropout experience ratio (57.14%).

The results of the decision tree analysis for the prediction of dropout model regarding the self-efficacy, institutional environments, and learning environmentare shown in Figure 3. The dependent variable of our decision tree is "Dropout" which has two cases, Dropout before (yes) or No dropout (No). The root of this tree contains 161 observations in this dataset. The most important influential attribute(factor) for dropout is the financial situation (income, job, scholarship) from the demographic attribute while the "quality of faculty teaching is the most critical factor from the institutional environments. The financial situation is more severe in the future due to the pandemic circumstance. It implies that students would focus on the job rather than the academics. Finally, the students' confidence level for current courses at the time responding this survey shows that 93% students indicate that they have confidence that can perform well in the course while 7% students indicate that their confidence level is lower.

The findings from the descriptive and decision tree analysis show that one of the major reasons for dropout is the personnel-related factor (financial situation, job, family issue, etc.) rather than the institutional environment such as quality of teaching, facilities of the classroom (i.e., tutoring support, lab, etc.). However, faculty still needs to improve the quality of teaching to meet the needs of students in order to retain the students in the class, major program, or even in the college.

As this is the first time to determine the critical factors for student's dropout from the courses and the college at participated institution based on the empirical study from the students' perspective rather the academic record or data, this project contributes to find the factor and co-relationship with the demographics and academic environment (i.e., learning environments), and institutional environments affecting student success or failure.

Conclusions & Implications

The purpose of this study is to identify issues on why students want to drop out, why they transfer or leave the institution and to follow up with intervention options and appropriate strategies to enhance student retention. This study involves undergraduate students and identifies a high impact practice with extensive benefits and value for students, faculty, and universities since it determines the value driven needs and wants. In addition, our findings validate previous theoretical arguments that the augment the classroom setting and ultimately determine whether there is a pattern in everyone's (i.e., self-efficacy) reasons in the decision to withdraw, and how faculty and instructors should implement the intervention program to increase student retention and completion in a course.

This study applies data mining and decision tree modules to monitor student's profile, analyze student academic behavior, and provide a basis for efficient intervention strategies. In this paper we discuss how data mining and decision tree modules can help spot students 'at risk' in the context of major selection, course/class management by evaluating the course or module suitability, and tailor the interventions to increase student success and retention and improve student's performance. The project and its findings are compared with previous studies on the student success/failure studies and their research models/tools in order to determine the best way to determine the accuracy of the models and its implementation in the professional field in higher education.

However, we should caution that the validity of the implied casuals and prediction of decision tree analysis and its model are limited by the students' perspective. In addition, our findings may only be generalized to the hospitality and business management areas with similar characteristics (i.e., the hospitality industry is more service-oriented and service-specific asset, and the products cannot be stored for the next day or future sale and its daily quota).

We believe that students' personal environment and self-efficacy may be an important critical factor toward potential dropout, however, hospitality management faculty need to augment their teaching module and consider key factors contributing for students why they drop out from the class and even the college and determine the best strategies for retention. Specifically, they need to utilize this study's findings to augment/modify/develop pedagogy in their classes especially for having a high dropout ratio in the class more corrective and effective measures in collaboration with the college, institutional research center.

The study has an impact on teaching and learning in higher education in the context of student retention and completion in a course generally through dissemination of the findings. The study examines how students perceive, experiencing, applying and developing learning skills in conjunction with the value driven mind set such as benefits vs time, tuition, course fees, etc. faculty and student experiences with this tool are needed, particularly to determine learning value. As such, this study provides practice related to high impact engaged learning in higher education.

The HIED survey method along with SPOT is relatively new to determine and identify students' perception on drop out, therefore, potential future directions of research include comprehensive longitudinal research on 1)panel study that involves sampling a cross-section of individual, 2) cohort study that involves selecting a group based on a specific attributes, and 3) retrospective study that involves looking to the past by looking at historical information such as students' academic record could embrace on both the generalizability of our findings to other programs as well as the validity of the causal links for intention to dropout.

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

Aljohani, O. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. Higher Education Studies, 6(2). 1-18.

Bean, J. & Eaton, S.B. (2001-2002). The psychology underlying successful retention practices. Journal of College Student Retention, 3(1), 73-89.

Berger, J.B., Ramirez, G.B., & Lyons, S. (2012). Past to present: A historical look at retention. In A. Seidman (Ed.). College Student Retention. New York: American Council on Education.

Berns, J. Schneider, K., Gortz, S., Oster, S., &Burghoff, J. (2018). Working Paper: Early detection of students at risk - Predicting student dropout using administrative student data and machine learning methods. CESifo Working paper, No. 7259. http://hdl.handle.net/101419/185457.

Braxton, J. M., Doyle, W., Hartley, H.V., Hirschy, W.A., Jones, W.A., McLendon, J.K. (2014). Rethinking College Student Retention. San Francisco; Jossey-Bass. 177-179.

Breiman, L., Friedman, J., Olshen, R,,& Stone, C. (1984). Classification and Regression Trees. Monterey, CA.Wadworth International Group.

Caison, A. L. (2007). Analysis of institutionally specific retention research: A comparison between survey and institutional database methods. Research in Higher Education, 48(4), 435-451.

Delen, D. (2011-2). Predicting Student Attrition with Data Mining Methods. Journal of College Student Retention. 13(1), 17-35.

Dewberry, C. & Jackson, D. (2018). An application of the theory of planned behavior to student retention. Journal of Vocational Behavior, 107, 100-110.

Dyk, C.V. &Weese, J. (2019). The undeniable role that campus recreation programs can play in increasing indigenous student engagement and retention. Recreational Sports Journal, 43(2), 126-136. https://doi.org/10.1177/1558866119885191.

Fong, C.J., Davis, C.W., & Kim, Y. (2016). Psychosocial factors and community college student success: A meta-analytic investigation. Review of Educational Research, 87(2), 388-424. https://doi.org/10.3102/0034654316653479.

Gallup; Virginia Tech. (2018). Virginia Tech Student Survey: Understanding students' experience and perceptions of Virginia Tech. Retrieved from file:///C:/Users/10009444/Documents/VT_Student_Survey_Findings_070918_digital_FINAL.pdf

Guiffrida, D.A. (2006). Toward a cultural advancement of Tinto's theory. Review of Higher Education, 294, 451-471.

Hagedorn, (2012). How to define retention: A new look at an old problem. In Alan Seidman (2nd ed). College Student Retention: Formula for Student Success, 81. UK. Rowman & Littlefield Publishers.

Hanson, M. (2021). College dropout rates. Eductiondata.org. Retrieved fromhttps://educationdata.org/college-dropout-rates/

Hermanowicz, J. C. (2003). College attrition at American Research Universities: Comparative case studies, New York: Agathon Press.

Kerby, M.B. (2015). Toward a new predictive model of student retention in higher education: An application of classical sociological theory. Journal of College Student Retention: Research, Theory & Practice. 17(2), 138-161.

Kuh, G. D., O'Donnell, K., & Schneider, C. G. (2017). HIPs at ten. *Change: The Magazine of Higher Learning, 49*(5), 8-16. Retrieved from https://doi.org/10.1080/00091383.2017.1366805

Kumar, A.D. & Radhika, V. (2014). A survey on predicting student performance. International Journal of Computer Science and Information Technologies, 5(5), 6147-6149.

James, G., Witten, D., Hastie, T.,&Tibshirani, R. (2013). *An Introduction to Statistical Learning*, Springer.

Latif, A., Choudhary, A., &Hammayun, A. (2015). Economic effects of student dropouts: A comparative study. Journal of Global Economics, 3(2), 137.

Lin, S.H. (2012). Data mining for student retention management. Consortium for Computing Sciences in College. JCSC 27(4). CCSC; Southwestern Conference. 92-99.

Mah, D-K. (2016). Learning analytics and digital badges: Potential impact on student retention in higher education. Technology, Knowledge and Learning 21, 285-305.

Mengo, C. & Black, B. M. (2016). Violence victimization on a college campus: Impact on GPA and school dropout. Journal of College Student Retention: Research, Theory & Practice. 18(2), 234-248.

Nora, A. & Crisp, G. (2012). Future research on Hispanic students: What have we yet to learn? And what new and diverse perspective are needed to examine Latino success in higher education? Vtechworks.lib.vt.edu.

Peguero, A. & Bracy, N. (2014). School order, justice, and education: Climate, discipline, practices, and dropping out. Society for Research on Adolescence, 1-15.

Olaya, D., Vasquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift modeling for preventing student dropout in higher education. Decision Support Systems, 134, 113420.

Quadril, M. &Kalyankar, N. (2010). Drop out feature of student data for academic performance using decision tree techniques. Global Journal of Computer Science and Technology, 2(10), 2-5.

Quinlin, J.R. (1986). Induction of Machine Learning. Machine Language 1. 81-106. https://doi.org/10.1023/A:1022643204877

Roiger, R. J. (2017). Data Mining: A tutorial-based primer. 2nd Ed. Boca Raton, FL. CRC Press.

Ropeti, S. (2015). Student perspectives regarding school failure at the American Samoacommunity college. Walden University ScholarWorks. Walden Dissertation and Doctoral Studies Collection.

Sahin, S., Arseven, Z., &Kilic, A. (2016, January). Causes of student absenteeism and school dropouts. International Journal of Instruction. 9(1), 195-210.

Seidman, Alan. (2012). Formula for student success. College Student Retention 2nd Edition. American Council on Education. Maryland. Rowman & Littlefield Publishers, Inc.

Sivakumar, S., Venkataraman S., & Selvaraj, R. (2016, January). Predictive modeling of student dropout indicators in educational data mining using improved decision tree. Indian Journal of Science and Technology, 9(4), 1-5.

Tight, M. (2020). Student retention and engagement in higher education. Journal of Further and Higher Education,

44(5), 689-704.
Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (Second ed.), Chicago University Press, Chicago.
Tinto, V. (2000). Taking student retention seriously: Rethinking the first year of college. NACADA Journal 19(2) 5-10.
Tinto, V. (2006). Research and practice of student retention: What next? Journal of College Student Retention: Research, Theory & Practice, 8 (1), 1-19.
Toraby, E. &Modarresi, G. (2018, April). EFL Teachers' emotions and learners' views of teachers' pedagogical success. International Journal of Instruction, 11(2), 513-526.
Virginia Tech. (2019). Office of Academic Decision Support. Retrieved fromhttps://ads.vt.edu/spot.html
Wojciechowski, & Palmer, A. (2005). Individual student characteristics: Can any be predictors of success in online classes? Online Journal of Distance Learning Administration, 8(2), 13.
Wray, J., Barrett, D., Aspland, J., & Gardner, E. (2012). Staying the course: Factors influencing pre-registration nursing student progression into Year 2 – A retrospective cohort study. International Journal of Nursing.
Zhang, Y., Oussena, S., Clark, T., & Kim, H. (2010). *Use data mining to improve student retention in higher education – A case study. London. Middlesex University Research Repository.* Retrieved fromhttps://pdfs.semanticscholar.org/50f7/863aeb043c7fdf851bdcd57dbf6617ca83ea.pdf?_ga=2.179977188.1559167417.1572386974-1892701416.1572386974

Figure 1

Theory of Reasoned Action/Theory of Planned Behavior/Model for Prediction of Dropout

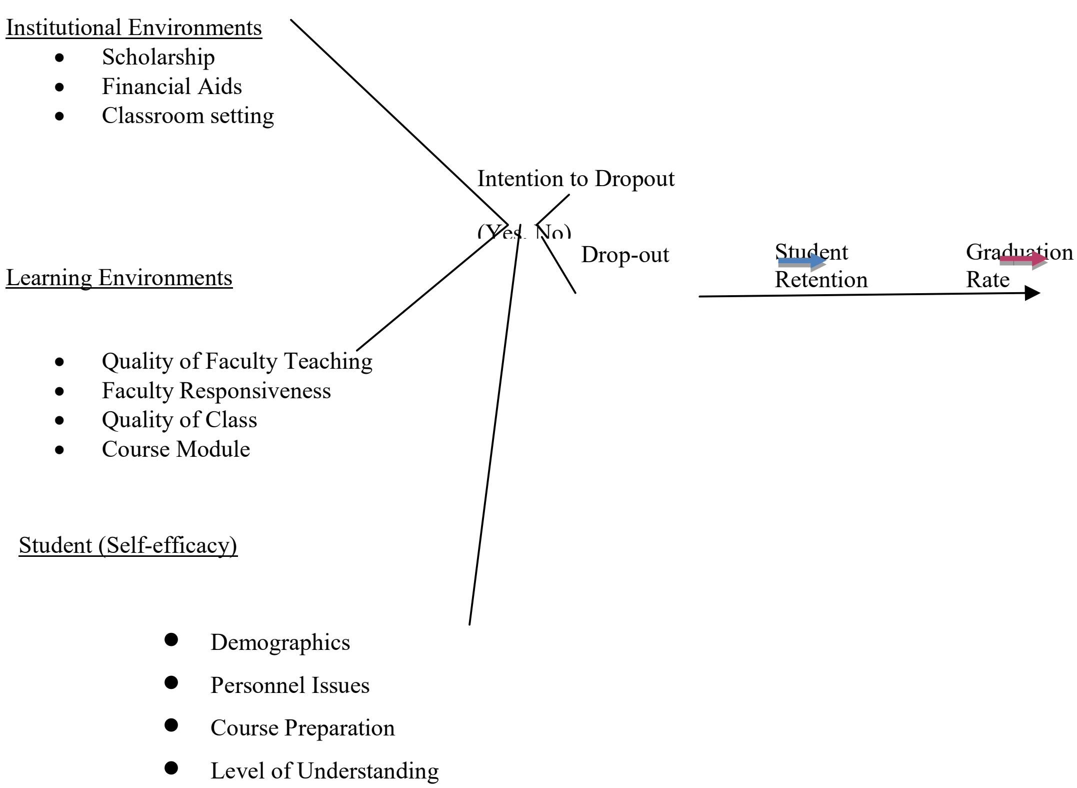


Table 2 Cross-Tabulation: Highest Ratio of Dropout (Demographics)

General Attributes	Frequency of Highest Percentage within each demographic's attributes has an experience on dropout from the class	
Gender	Female	68%
Age	21-25	56%
Academic Level	Senior (91-120 credits)	64%
Occupation	No Occupation (Full Time Students)	48%
Martial Status	Married	56%
Ethnicity	Caucasian	76%
Residency Status	Instate (Utah)	84%
Residing with	Married and Family	32%
Financial Support	Self	57.3%
Major	Business Management with Hospitality Management	44%
GPA	Between 3.1 - 3.5	56%
# of Credit Hours each semester	11-15	68%

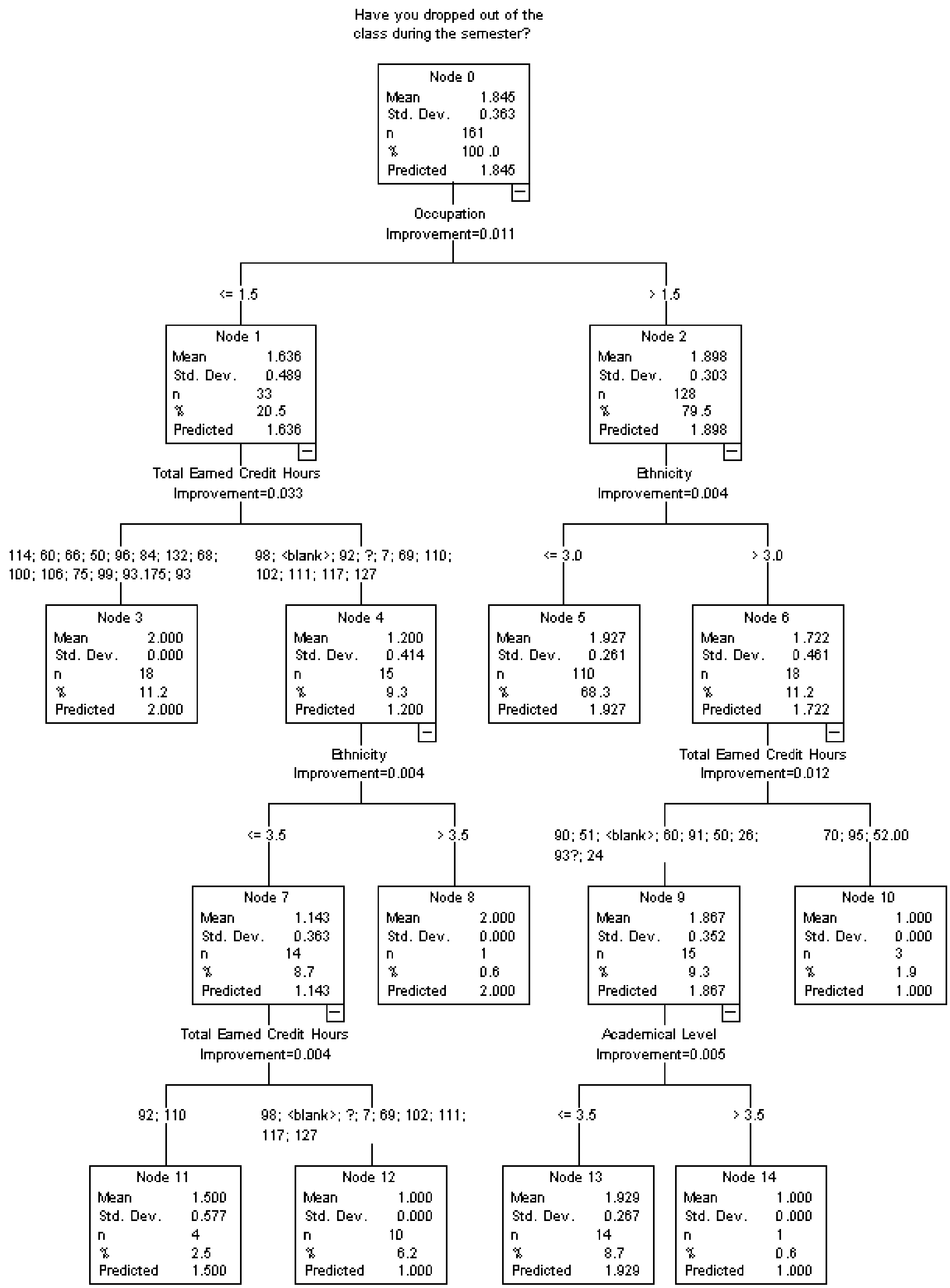
Table 3 Critical Factor for Dropped Out

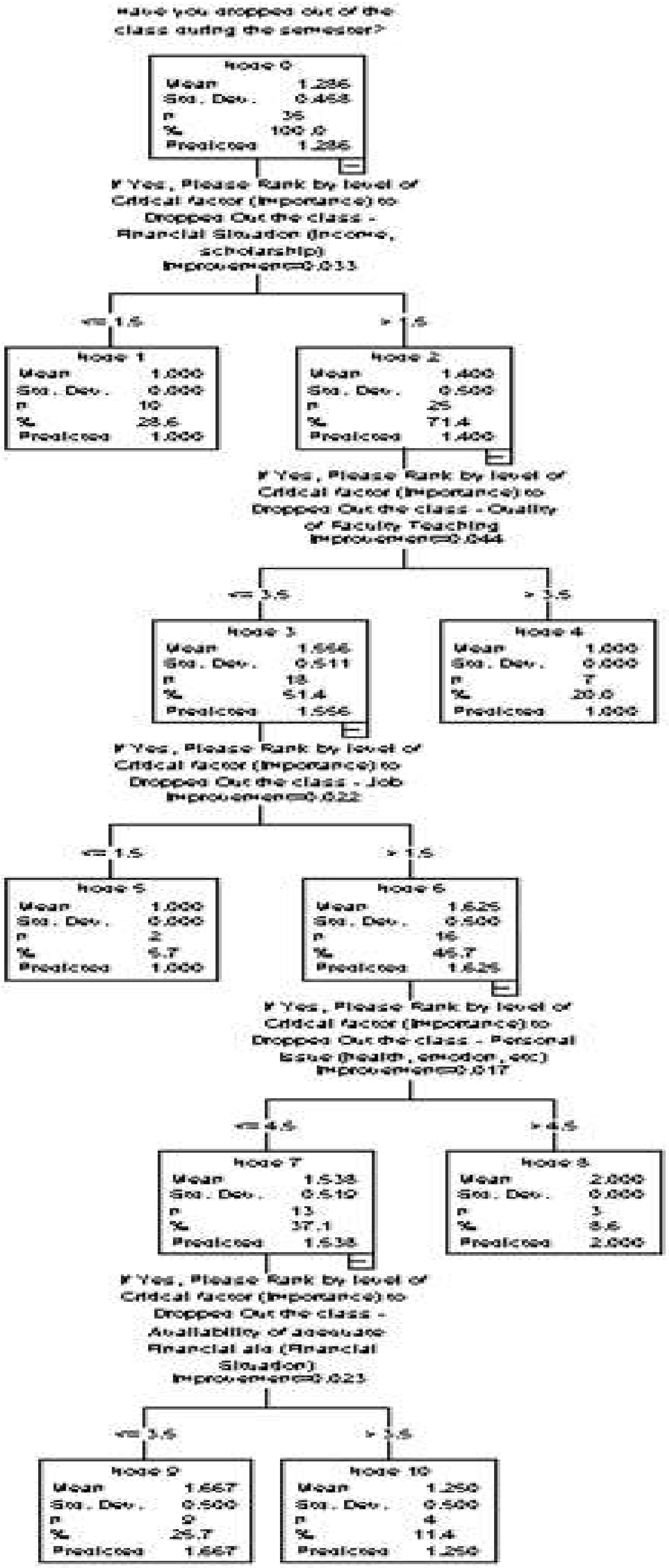
Critical Factor	Mean
Financial situation (Income)	3.58
Person issue	3.47
Quality of class (module)	3.03
Level of preparation for the course work	3.03
Job	3
Quality of teaching	2.97
Level of understanding the course contents	2.97
Family issue	2.84
Faculty responsiveness to student needs	2.84
Class workload	2.75
Classroom environment (peer collaboration)	2.62
Availability of adequate financial aid from school	2.59
Faculty feedback in a timely manner	2.48
Mid term grade	2.31
Other (specify)	2.31

Table 4 Critical Factor Impact on Possible Dropout

To be Critical Factor	(%) Proportion/Weight
Person issue	35.29
Financial situation (Income)	29.03
Class workload	18.75
Job Level of preparation for the course work	18.75
Quality of class (module)	15.63
Level of understanding the course contents	15.63
Quality of teaching	13.33
Availability of adequate financial aid from school	12.90
Faculty responsiveness to student needs	12.50
Family issue	10.34
Classroom environment (peer collaboration)	9.68
Mid-term grade	6.90
Faculty feedback in a timely manner	3.45
Confidence level	1.53
Other (specify)	0.70

Figure 2 Decision Tree of Demographics & All Attributes




Have you dropped out of the class during the semester?
Node 0
Mean 1.286
Std. Dev. 0.458
n 35
% 100.0
Predicted 1.286
If Yes, Please Rank by level of Critical factor (Importance) to Dropped Out the class - Financial Situation (Income, scholarship)
Improvement=0.033
<= 1.5
> 1.5
Node 1
Mean 1.000
Std. Dev. 0.000
n 10
% 28.6
Predicted 1.000
Node 2
Mean 1.400
Std. Dev. 0.500
n 25
% 71.4
Predicted 1.400
If Yes, Please Rank by level of Critical factor (Importance) to Dropped Out the class - Quality of Faculty Teaching
Improvement=0.044
<= 3.5
> 3.5
Node 3
Mean 1.556
Std. Dev. 0.511
n 18
% 51.4
Predicted 1.556
Node 4
Mean 1.000
Std. Dev. 0.000
n 7
% 20.0
Predicted 1.000
If Yes, Please Rank by level of Critical factor (Importance) to Dropped Out the class - Job
Improvement=0.022
<= 1.5
> 1.5
Node 5
Mean 1.000
Std. Dev. 0.000
n 2
% 5.7
Predicted 1.000
Node 6
Mean 1.625
Std. Dev. 0.500
n 16
% 45.7
Predicted 1.625
If Yes, Please Rank by level of Critical factor (Importance) to Dropped Out the class - Personal Issue (health, emotion, etc)
Improvement=0.017
<= 4.5
> 4.5
Node 7
Mean 1.538
Std. Dev. 0.519
n 13
% 37.1
Predicted 1.538
Node 8
Mean 2.000
Std. Dev. 0.000
n 3
% 8.6
Predicted 2.000
If Yes, Please Rank by level of Critical factor (Importance) to Dropped Out the class - Availability of adequate Financial aid (Financial Situation)
Improvement=0.023
<= 3.5
> 3.5
Node 9
Mean 1.667
Std. Dev. 0.500
n 9
% 25.7
Predicted 1.667
Node 10
Mean 1.250
Std. Dev. 0.500
n 4
% 11.4
Predicted 1.250