Analysis of UA First-time Freshmen Success in English 111 by Course Delivery Type

Chantelle J. McGinness
University of Alaska Fairbanks
Spring 2013

Committee:
Dr. Ronald Barry (Advisor)
Dr. Julie McIntyre
Dr. Margaret Short

Author Note

Chantelle McGinness, Department of Mathematics and Statistics, University of Alaska Fairbanks.

Chantelle McGinness is also employed by UA Institutional Research and Analysis, University of Alaska.

This research was preformed to satisfy degree requirements for a Master of Science in Statistics at the University of Alaska Fairbanks.

Correspondence concerning this article should be addressed to Chantelle McGinness, UA Institutional Research and Analysis, University of Alaska, PO Box 755260, Fairbanks, AK 99775. E-mail: cjmcginness@alaska.edu

TABLE OF CONTENTS

Abstract

This study examines course success for

1. Introduction

Considering other emerging trends in e-Learning, Lee and Choi (2011) preformed a literature review of recent research th

2.2 Variables

The response variable was course success, which

Several academic variables were also used as predictors, and included the following:

1. Term: This variable identifies which term a student was enrolled in English 111: fall 2011, spring 2012, summer 2012, or fall 2012.
2. High School Grade Point Average: This variable is a continuous variable on a 0.0 to 5.0 scale.
3. Degree Level: Degree level is the primary degree program in which a student is enrolled. Categories include certificate, associate, or bachelor.
4. Full-time/part-time Status: Full-time/part-time status is determined each term by a student's credit hour load. Students who were enrolled in 12 or more credits were classified as full-time students. Students who were enrolled for fewer credits were classified as part-time students. Audited student credit hours were not included for computing full-time status.
5. Student Credit Hour Load: Student credit hour load is based on the number on non-audit hours taken by students in credit courses: 0-3, 4-6, 7-8, 9, 10-11, 12-14, 15 or more.
6. Discipline Area: Discipline area is the student's broad field of study, based on the first two digits of discipline (CIP) code for a student's primary major. Examples include education, engineering, health, natural resources, etc.

3. Logistic Regression

3.1 Method of Analysis

The logistic regressions for this study were generated using SAS software, Version 9 of the SAS System for Linux. Copyright © 2002-

Developed during the late 19th cen

SUCCESS BY COURSE DELIVERY TYPE

Main effects models for each data set were built using forward stepwise selection. This procedure first estimated a parameter for the intercept model, then evaluated the explanatory variables using the residual chi-

3.2 Results: Course Delivery Method

The initial logistic regression investigated whether course success was statistically related to course delivery method, traditional versus e-Learning. In total, 2,255 first-time freshmen were identified as having taken English 111 between fall 2011 and fall 2012. After 207 records were excluded from the analysis because of missing high school GPA, 2,048 records were analyzed. Table 1 reports the demographic characteristics by course delivery type.

Table 1. Demographic Characteristics by Course Delivery Type

	Traditional		eLearning	
	Headcount	Percent of Total	Headcount	Percent of Total
Gender				
Female	[illegible]	[illegible]	[illegible]	[illegible]
Male	[illegible]	[illegible]	[illegible]	[illegible]
UAA Race				
AK Native/American Indian	[illegible]	[illegible]	[illegible]	[illegible]
Hawaiian/Pacific Islander	[illegible]	[illegible]	[illegible]	[illegible]
Asian	[illegible]	[illegible]	102	5.46
Black	3	1.67	65	3.48
White	[illegible]	[illegible]	[illegible]	[illegible]
Other	[illegible]	[illegible]	[illegible]	[illegible]
Origin State at Entry				
Alaskan	173	96.11	1713	91.70
Out-of-State	7	3.89	17	0.91
Foreign	[illegible]	[illegible]	138	7.39
Age Group				
Under 20	94	52.22	1590	[illegible]
20-24	46	25.56	161	8.62
25-29	22	12.22	73	3.91
30-39	10	5.56	22	1.18
40 or Older	8	4.44	14	0.75
First Generation Student				
No	123	68.33	1278	68.42
Yes	57	31.67	590	31.58
Received Financial Aid				
No	112	62.22	1060	56.75
Yes	68	37.78	808	43.25
Total	180	100.00	1868	100.00

SUCCESS BY COURSE DELIVERY TYPE

3.3 Results: Courses Delivered via e-Learning

For students who took English 111 via e-Learning, 40 records were excluded from the analysis because of missing high school GPA. Therefore, 180 observations were used for the analysis.

The forward

where X_{1i} =1 if a student did not receive financial aid and 0 if a student did receive financial aid, and X_{2i} =high school GPA.

The Hosmer and Lemeshow test supported goodness of fit p=0.7837, as shown in Table 7, suggesting that there is not a significantly better model based on a nonlinear function of high school GPA and financial aid.

For a students who took the English 111 via e-Learning, regardless of financial aid status, an one unit increase in high school GPA had a multiplicative effect of 3.3 on the odds that a student succeed in the course.   Holding high school GPA constant, students who did not receive model including high school GPA and financial aid

student aged 20 and younger.

4. Classification and Regression Tree (CART) Analysis

4.1 Method of Analysis

Prediction trees are nonparametric models that are useful for easily predicting a response variable from a set of observed explanatory variables. Buja and Lee (2001) summarize tree construction as beginning with "a greedy growing phase driven by a binary splitting criterion, followed by a pruning phase based on cost-complexity measures and/or estimates of generalization error." Response where R(T) is a measure of lack of fit, which was deviance in this study, and >0 penalizes for tree size (Ripley, 1996).

The CART analyses for this study were generated using R software, Version 2.15.2, on Windows 7. R software is free software under the terms of the GNU General Public License as published by the Free Software Foundation.

The ‘tree’ package was used to generate the classification trees. The splitting criterion used in this study was deviance, which is defined to be

$$D(T) = 2[\ _{t}\, n_t \log n_t - \ _{c}\, n_{tc} \log n_{tc}],$$

where t is the leaf index and c is the class index, n_t is the number of observations that will reach leaf t, and n_{tc} is the number of each class at the node (Ripley, 1996). Using deviance as the splitting criterion results in an impurity index of

$$I(T) = D(T)/2n,$$

where n is the totals number of observations (Ripley, 1996). For each data set, an initial tree was developed using the tree() function with a stopping criterion of a minimum within-node deviance of $0.003*D_r$, where D_r is the deviance of the root node. Then 10-fold cross-validation was performed to determine the optimal size of the tree, using the function cv.tree(). Finally, the best tree having the number of terminal nodes identified by the cross validation was generated.

4.2 Results: Courses Delivered via e-Learning

Using all explanatory variables, an initial classification tree was generated for students who took English 111 via e-Learning, which is depicted in Figure 1 cross-validated deviance increasing for splits starting as early as the root node in Figure 2. Therefore, a final best tree was not generated for this data set.

In predicting course success, interpretation of a classification tree begins at the root node. In Figure 2, the root node resulted

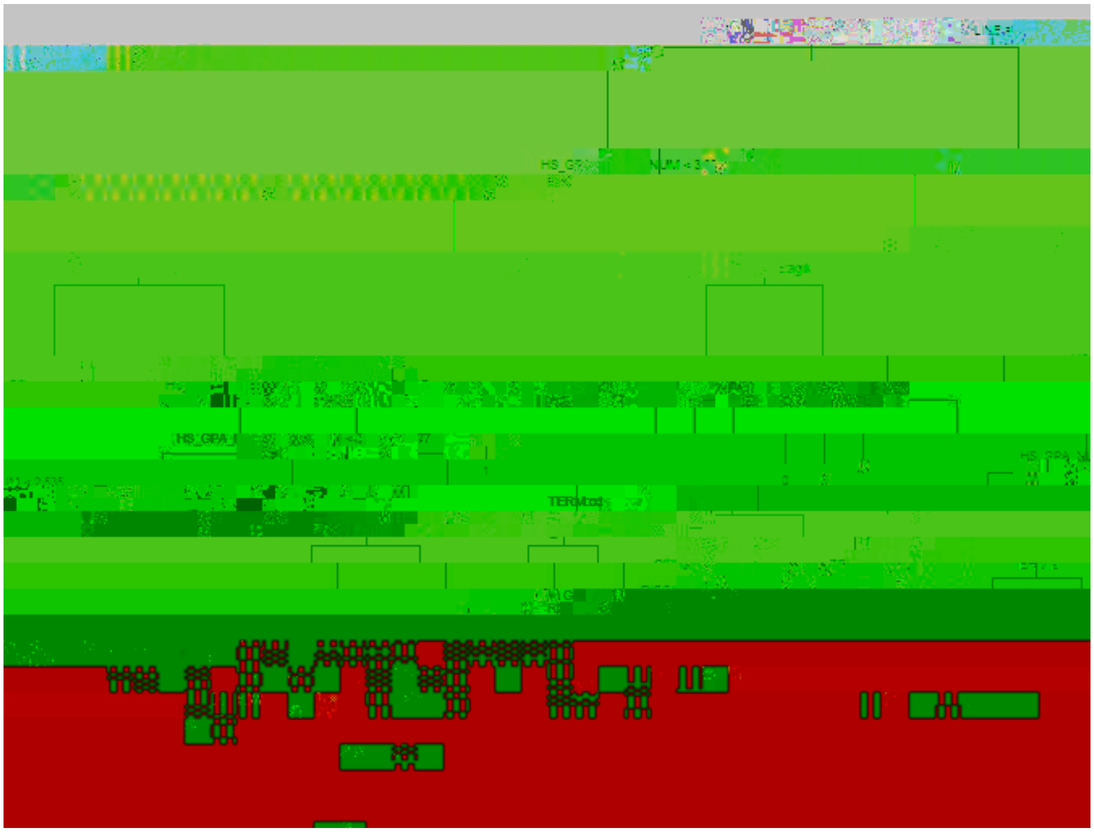

Figure 1: Initial classification tree for student who took English 111 via e-Learning (Diagram details can be found in Appendix A).

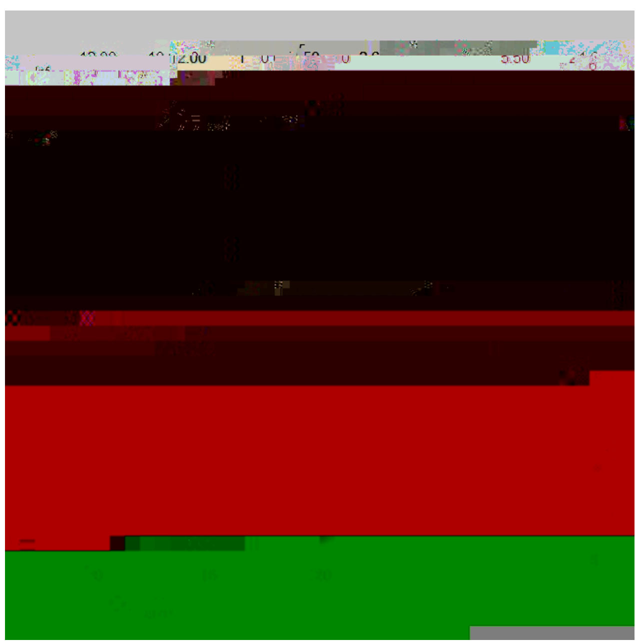

Figure 2: Tree size (horizontal axis) versus cross-validated deviance (vertical axis) for successive prunings of the initial classification tree for student who took English 111 via e-Learning. (The upper scale on the horizontal axis refers to the "cost/complexity" penalty.)

4.3 Results: Courses Delivered via Traditional Face-to-face Meetings

For students who took English 111 via traditional face-to-face meetings, all explanatory variables

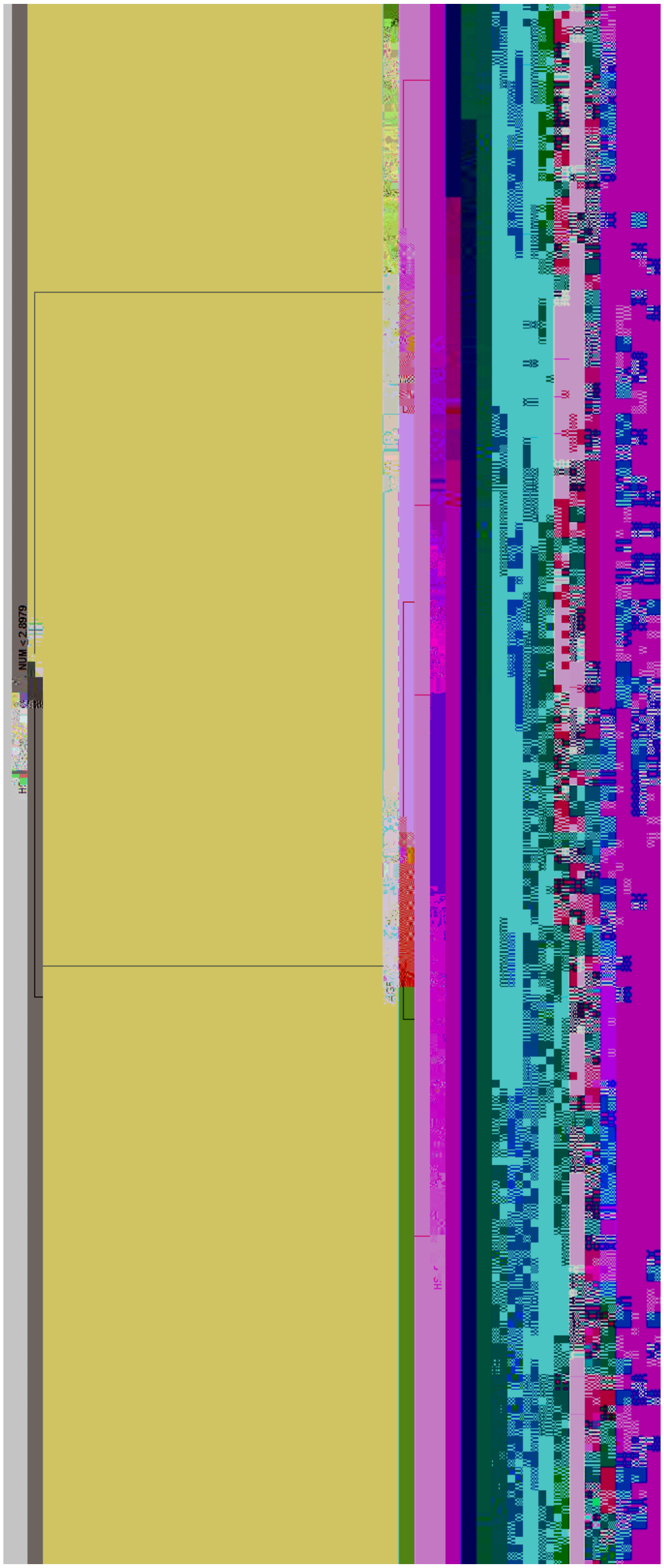

Figure 3: Initial classification tree for student who took English 111 via traditional face-to-face meetings (Diagram details can be found in Appendix B).

Figure 4: Tree size (horizontal axis) versus cross-validated deviance (vertical axis) for successive prunings of the initial classification tree for students who took English 111 via traditional face-to-face meetings. (The upper scale on the horizontal axis refers to the "cost/complexity" penalty.)

Figure 5: Final classification tree for student who took English 111 via traditional face-to-face meetings (Diagram details can be found in Appendix C).

5. Discussion

This study utilized two statistical methods, logistic regression and classification and regression tree (CART) analysis, in an effort to develop predictive models for course success in English 111 based on whether the course was delivered via eLearning or traditional face-to-face meetings. The CART analysis was less successful than logistic regression in producing viable predictive models for the data.

Regarding course success for traditionally delivered English 111, logistic regression produced vastly different results than the CART analysis. The final model that resulted from the logistic first-

SUCCESS BY COURSE DELIVERY TYPE

reports the minimum test scores that each MAU requires students to meet in order to be placed in English 111.

[redacted]

Based on the placement criteria, students at one MAU may be better prepared for the course and thus, more likely to succeed. Moreover, instructional technologies and methodologies may differ among the MAUs.

References

Blasius, J. & Greenacre, M. (1998). *Visualization of Categorical Data*, San Dieago, California: Academic Press.

Buja, A. & Lee, Y. (2001). Data mining criteria for tree-based regression and classification. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discover and Data Mining*, 27-36. doi: 10.1145/502512.502522

Davidson, P. (2009). University of Alaska unit cost analysis and other selected issues: Part 3, Distance Education. (Report Digest #45-30033c-09). Alaska State Legislature, Division of Legislative Audit.

Dupin-Bryant, P. (2004) Pre-entry variables related to retention in online distance education. *American Journal of Distance Education*, 18(4), 199-206.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied Linear Statistical Model*. New York, New York: McGraw-Hill/Irwin.

Lee, Y. & Choi, J. (2011). A review of online course dropout research: implications for practice and future research. *Educational Technology and Research Development*, 59, 593-618. doi: 10.1007/s11423-010-9177-y

Morris, L. V., Wu, S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *American Journal of Distance Education*. 11(2), 39-54.

National Center for Education Statistics. (2010). Web tables: Profile of Undergraduate Students: 2007-08. (NCES 2010-205). Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010205

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Radford, A. (2011). Learning at a distance: Undergraduate enrollment in distance education courses and degree programs. (NCES 2012-154). National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education. Washington, DC.

Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge, Great Britain: Cambridge University Press.

Ripley, B. (2012). tree: Classification and regression trees. R package version 1.0-33. http://CRAN.R-project.org/package=treeSAS Institute Inc. (1999).

SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

SAS OnlineDoc®, Version 8, Cary, NC. Retrieved from http://v8doc.sas.com/sashtml/stat/chap3/sect11.htm

SAS software, Version 9 of the SAS System for Linux. Copyright © 2001-2010 SAS Institute Inc.

University of Alaska. (2011). Distance education enhancements banner scheduling manual: How to schedule all your classes with the new distance education parameters. Retrieved from http://www.alaska.edu/files/studentservices/UA%20DE%20scheduler%20manual.pdf

University of Alaska. (2012). Report on University of Alaska e-Learning. (Reference 2). Board of Regents Meeting, Academic and Student Affairs Committee. Retrieved from http://www.alaska.edu/files/bor/121206Ref02_e-Learning_Rpt.pdf

Xi, D. & Jaggars, S. (2013). Adaptability to online learning: Differences across types of students and academic subject areas. Community College Research Center. New York, NY.

Appendix A

Below are the details for the classification tree depicted in Figure 1, which pertains to course success in English 111 via e-Learning:

node), split, n, deviance, yval, (yprob)

* denotes terminal node

1) root 180 221.600 1 (0.3056 0.6944)
2) DISCIPLINE: 01 Business and Public Administration,02 Computer and Information Science,03 Education,04 Engineering,06 Health,07 Letters, Comm., and Philosophy,10 Social Sciences,11 Visual and Performing Arts,12 Vocational Education 16 16 1

165) RACE: 05 White 17 15.840 1 (0.1765 0.8235)
330) DEG_LEVEL: 3_Associate 7 0.000 1 (0.0000 1.0000) *
331) DEG_LEVEL: 5_Bachelor 10 12.220 1 (0.3000 0.7000) *
83) TERM: 201201 5 0.000 1 (0.0000 1.0000) *
21) HS_GPA_NUM > 3.77 7 0.000 1 (0.0000 1.0000) *
11) DISCIPLINE: 02 Computer and Information Science,03 Education,04 Engineering,06 Health 21 0.000 1 (0.0000 1.0000) *
3) DISCIPLINE: 05 Foreign Languages,08 Math, Physical and Life Sciences,14 Interdisciplinary Studies 15 0.000 1 (0.0000 1.0000) *

Appendix C

Below are the details for the classification tree depicted in Figure 5, which pertains to course success in English 111 via traditional face-to-face meetings:

node), split, n, deviance, yval, (yprob)

* denotes terminal node

```
1) root 1868 1646.0 1 ( 0.16060 0.83940 )
  2) HS_GPA_NUM < 2.8979 596  738.4 1 ( 0.31040 0.68960 )
    4) AGE_GROUP: 01 Under 20 415  543.0 1 ( 0.36145 0.63855 )
      8) HS_GPA_NUM < 2.475 157  217.5 1 ( 0.48408 0.51592 ) *
      9) HS_GPA_NUM > 2.475 258  309.2 1 ( 0.28682 0.71318 ) *.316 _GROt 0 0 Tm /Tc1.4 (*)Td:
```