Predicting Academic Achievement in Fundamentals of Thermodynamics using Supervised Machine Learning Techniques

Paul Akangah*, Leotis Parrish*, Andrea Ofori-Boadu* and Francis Davis+

**North Carolina A&T State University/+Kwame Nkrumah University of Science & Technology*

Abstract

Supervised machine learning techniques were used to answer the central question: does excelling in reading quizzes, a good predictor of accurately predicting the passing rate in MEEN241 Fundamentals of Thermodynamics? Class assignments such as reading quizzes (RQ), quizzes (Q) and home-works (HW), Tests (T), and midterm (MT) were designed. The predictor variables analyzed are High GPA (>3.0), RQ, Q, and passing PHYS241 General Physics and these were used to develop two classifiers: CART with cross-validation and Random Forest. The CART and the Random Forest models identified Q, and Q and RQ, respectively as the best predictor, although quiz and reading quiz accounted respectively, for only 5% and 15% of the total weight. This suggests that students who devote the time and effort into doing the reading assignments and subsequently passing both the RQ and Q are likely to expend similar effort and time in other class assignments and preparation towards tests and examinations.

Keywords

Supervised machine learning, random forest, decision tree, thermodynamics, GPA.

Introduction

Dweck counseled that students with growth mindsets tend to outperform students with fixed mindsets [1]. Mindset is the characteristics mental attitude that determines how one would interpret and respond to situations [2]. The essence of Dweck's work is that students with a growth mindset believe that qualities such as intelligence, personality, and character can be developed through the process of hard-work, effort, and good strategies. Growth mindset students believe that these qualities are potentials that can be achieved by confronting challenges, profiting from mistakes, and persevering in the face of setbacks. Also, Dweck argued that students with a fixed mindset believe that these qualities are fixed and that they cannot be developed. This group of students finds challenges intimidating because they worry that their fixed characteristics may not be up to the task. As a result, these students are often afraid to leave their comfort zone and face challenges. They perceive mistakes and failures as demoralizing because such setbacks reflect severely on their level of fixed characteristics.

Engineering Thermodynamics is the basis of pure and engineering sciences [3], and an introductory course in Thermodynamics is taken by all engineering majors in the United States

[4]. However, students have found the subject very difficult to grasp. At the North Carolina A&T State University (NCAT), non-mechanical engineering majors are required to take a single three-credit introductory course in thermodynamics, although some departments are revising their curriculum and taking out the introductory course in thermodynamics. Mechanical engineering majors are required to take the introductory as well as the advanced courses.

Many researchers have used machine learning techniques to predict academic achievement (AA). Authors of [5] compare five classification algorithms, namely decision tree, Naïve Bayes tree, K-nearest neighbor and Bayesian network algorithms, for predicting engineering students' letter-grades. The complete dataset is used to build a classifier then the bootstrap method is used to improve the accuracy of the classifier. Their data set consists of 1000 instances and 18 attributes. Authors of [6] also used decision tree technique to access AA of undergraduate students using various qualitative attributes such as economic status, resource accessibility, living location etc. This study compared three decision tree algorithms—ID3, C4.5, and Classification and Regression Trees (CART). The decision tree learning algorithm (ID3) developed by Ross Quinlan [7], employs a top-down, greedy search through the given sets to test each attribute at every tree node, to select the attribute that has the highest information gain [5]. The primary advantage of CART algorithm is that it looks at all possible splits for all attributes. Once the best split is found, the algorithm repeats the search process for another node, continuing the recursive process until no further splitting is possible or a stopping criterion is reached. CART uses the Gini Index as an attribute selection measure. The Gini Index is a cost optimization method [5].

The objective of this work is to explore the relationship between a student passing Fundamentals of Thermodynamics (MEEN241) course, high GPA (GPA > 3.0), reading quiz (RQ), General Physics (PHYS241), and quiz (Q). The central question the authors seek to answer is: does passing RQ a good predictor to passing the MEEN241? To pass the RQ, students must obtain a minimum of 60% overall grade. Decision tree is a commonly used supervised classification technique [6] and is very popular technique because of its high accuracy [8]. Decision tree and random forest classifiers were used in this study.

Methods

Participants

The sample was composed of 39 undergraduate students enrolled in MEEN241 Fundamentals of Thermodynamics during the Fall 2016 semester. Class assignments and assessments were designed. Students are required to structure problem solution based on the problem-solving strategy outlined in the recommended textbook—An Engineering Approach, 8th Ed, by Yunus A. Cengel and Michael A. Boles [9]. In addition, students were assigned a total of seven reading assignments. After each reading assignment, a reading quiz was administered to access how thorough students understood the material assigned, which were selected sections from the textbook and these sections were part of the ABET accredited course syllabus. This approach is vital to help students become life-long learners. Questions on the reading quiz were conceptual. Also, students were encouraged to make notes while reading and could use their notes during the reading quiz. Four quizzes, based on materials discussed in class, were administered. In addition to having the course instructor as a resource for learning, students had access to office hours, where

they experienced personal student-faculty interaction, and teaching assistants, to assist them. Student-faculty interaction is important for student success [10].

A pass in PHYS241 General Physics with grade C or better prior to taking MEEN241 is optional. The GPA and pass/fail status in PHYS241 of each student was recorded at the beginning of the semester. All students registered for the course met the required prerequisite—MATH132 Calculus I and CHEM106 General Chemistry VI with a grade of C or better.

Data Preparation and Modeling

Measured Data

The measured data were taken from the AGGIE access database. This is a secure online portal, designed by Ellucian Company LP [11], that contains all information about students' academic progress. Data taken from this database included cumulative scores for RQ, Q, HW and pass/fail status in MEEN241 at the end of the Fall semester. GPA and pass/fail status in PHYS241 were taken at the start of the semester. The measured data was then stored in a comma-separated value (CSV) file format. Each student was randomly assigned a 2-digit number ranging from 10 to 90. The raw data was arranged in ascending order based on the randomly assigned numbers.

Analytic Data

The measured data is transformed into analytic data using Microsoft EXCEL. A pass in MEEN241 and PHYS241 is coded as a 1 while a fail is coded as 0. For the RQ and Q, a pass is 60% or above is coded as 1, otherwise, 0. A GPA of 3.0 and above is considered a high GPA and is coded 1, otherwise, 0. In addition to RQ and Q, students were also assigned HW, Tests (T), Midterm (MT) and Final Examination (FE). The weights assigned to these various assessments and the number of times an assessment was done, and students' success in the various predictor variables are shown in Table 1.

Table 1. Weights and Frequencies of Various Assignments and Students' Success

ASSESSMENTS			STUDENTS' SUCCESS	
Assessments	**# of Assessments**	**Weights**	**Predictor Variables**	**Success**
HWs	10	10%	GPA	48.7%
RQ	7	5%	PHYS241	74.4%
Q	4	15%	RQ	74.4%
T	2	20%	Q	59.0%
MT	1	20%		
FE	1	30%		

Although accounting for 70% of the total weight, FE, T, and MT were eliminated because of the low frequency of assessment. In addition, HW scores were also eliminated because independent student work could not be guaranteed. HW data showed that 89.7% (35 students) of students passed the HW assignments with a mean score of 79.7%. The highest score was 96.0%, minimum score, 20.9% and standard deviation of 14.5%. The pass-rate in MEEN241 is 64.1%.

Statistical Modeling

Supervised learning techniques, CART and random forest classifiers, were used. In each case, the response variable is MEEN241 and the predictor variables are GPA, PHYS241, RQ and Q.

RapidMiner [12] is an integrated extendable environment for machine learning, data mining, text mining and predictive analytics platform having excellent drag and drop graphics capability. It has powerful algorithms capable of solving many analytics problems. There are an open source and commercial versions of the software. The open source version is used in this study. This software was developed by RapidMiner Inc.

CART Decision Tree with Cross-Validation

In this model, the number of subsets was set to three. A linear sampling method was used in sub-setting the data. In cross-validation, two subsets were used in testing the data and one subset was used for testing. A minimum gini index gain of 0.1 was used. The resulting decision tree is shown in Figure 1. A confidence of 0.1 was used in making predictions based on the decision tree. This parameter specifies the confidence level used for the pessimistic error calculation of pruning [13]. Pessimistic pruning uses pessimistic statistical correlation test [14].

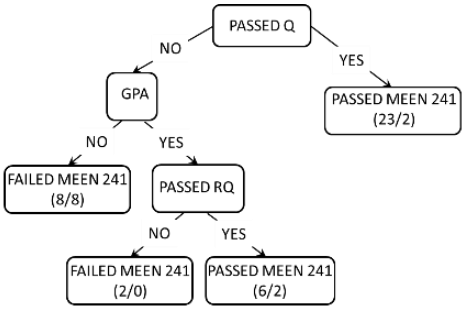

Figure 1. Decision tree for CART cross-validation

Random Forest

With the random forest method, there is no danger of overfitting. This method generates several decision-tree ensembles during training and outputting the class that is the mode of the classes in the case of classification problems or output the mean prediction of the individual trees, in the case of regression problems. Three subsets were created at random with replacement from the original data. This is the classic bootstrap method. The variables were also selected at random and the best split to the node is made. In this method, there is no pruning of the trees. A minimal gini index of 0.1 and a confidence of 0.1 were applied. Ten different decision trees were generated. Four of these trees are shown in Figure 2. A stratified sampling with three subsets was used in building these trees. This ensures that the class distribution in the subsets is the same as in the whole dataset.

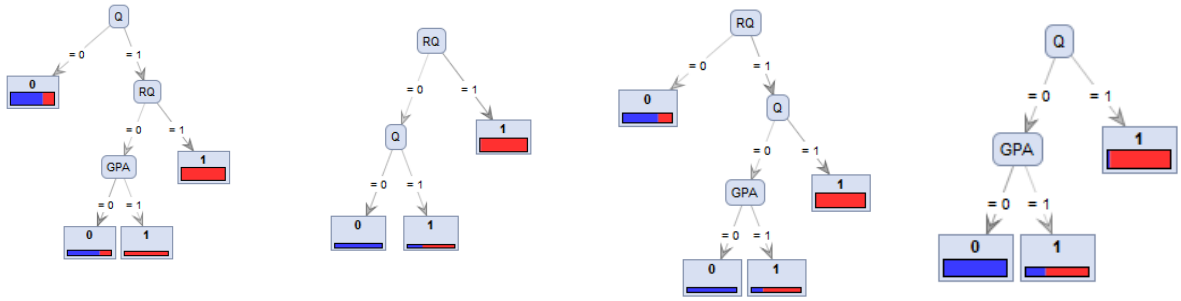

Figure 2. Decision tree for random forest

Results

The Confusion matrices are shown in Table 2.

Table 2. Confusion matrix for (a) CART Decision Tree with cross-validation, and (b) Random Forest model

TRUE	FAILED MEEN 241	PASSED MEEN 241	CLASS PRECISION
FAILED MEEN 241	9	1	90.00%
PASSED MEEN 241	5	24	82.76%
CLASS RECALL	64.29%	96.00%	

(a) CART Decision Tree with cross-validation

TRUE	FAILED MEEN 241	PASSED MEEN 241	CLASS PRECISION
FAILED MEEN 241	11	2	84.62%
PASSED MEEN 241	3	23	88.46%
CLASS RECALL	78.57%	92.00%	

(b) Random Forest model

The accuracy, classification error and kappa statistics for the CART model with cross-validation and Random Forest model are calculated with 95% confidence interval as: 84.62% +/- 6.28%, 15.38% +/- 6.28%, and 0.604 +/- 0.155; 87.18% +/-3.63%, 12.82% +/- 3.63%, and 0.719 +/- 0.062, respectively.

Discussion

This study explores how the predictor variables GPA, RQ, Q and PHYS241 affects the passing rate in MEEN241. The class assignments and assessments were designed along the problem-solving strategy outlined in the textbook and resources were also available to students to encourage learning. In addition, students were assigned reading assignments periodically. Supervised learning method was used and classification models were developed.

Attribute "PASSED_Q" is the root node in the CART model as it the best predictor among the predictor variables. This model rejects the central question that RQ is the primary predictor of passing MEEN241. However, there were instances where the Random Forest method generated models with "PASSED_RQ" as the root node. Perhaps, by combining Q and RQ, a more powerful central question could be obtained.

GPA is a very important parameter used in admissions, job recruitment decisions among others. However, various studies have shown that GPA can be easily influenced by reporting biases [15]. Some of these biases include tenures of faculty members [4], student-faculty interaction [10], motivation to achieve in college [16], whether students worked part-time during the semester, among others. Because of these reporting biases, it is not surprising that GPA is not overly important and therefore, not the best predictor of passing MEEN241. As Dweck aptly put it *"…an assessment of someone's abilities at one point in time has little value for predicting that person's potential ability in the future"* [17].

For the CART model with cross-validation, the kappa statistic ranged from 0.446 to 0.756, making the model to have a moderate to a substantial agreement with the data [18]. For the Random Forest model, the range is from 0.657 to 0.781, making this model to have a substantial agreement with the data.

Since PHYS241 is neither a root node nor a leaf node, statistically, this means that PHYS241 does not influence students' success in MEEN241. This finding is rather disturbing given that the syllabus of PHYS241 covers the fundamentals of heat and thermodynamics [19]. It is the opinion of the authors that taking PHYS241 should rather help students understand thermodynamics better. More data is needed to understand how PHYS241 impacts students' performance in MEEN241.

To pass both Q and RQ, students must have the discipline to put in the time and efforts needed to do the reading assignments and prepare towards reading quizzes. These qualities are also translated into solving problems for homework, preparing for tests and final examination. Students who are successful in doing these activities have the growth mindset and believe that the harder you work, the more your abilities will grow [1]. This study has demonstrated the importance of process—strategies that make investments in hard work, perseverance, and persistence in problem-solving translated in achievements by capitalizing on mistakes and confronting failures.

Conclusions

This paper explores the relationship between the pass-rate in MEEN241 (64.1%) and predictor variables such as GPA>3.0, "PASSED_PHYS241", "PASSED_Q", and "PASSED_RQ". Class assignments and assessments were designed. The problem-solving approach is based on the problem-solving strategy outlined in the textbook. Reading assignments were regularly given followed by RQ. Reading assignments are important because they encourage life-long learning skills in students. Developing life-long learning skills is one of the course outcomes for MEEN241.

The models have good accuracy, classification error and kappa values. However, the models are slightly better at predicting passing rate than they are at predicting failing rate. The kappa statistics suggested that the probability that chance alone will explain these predictions is low.

The models developed were unable to find any relationship between MEEN241 and PHYS241. The curriculum of PHYS241 covers some of the fundamental concepts in thermodynamics and should influence passing rate in MEEN241. This is an on-going research and data collection is in progress since the Spring 2017 semester. Since then, over 120 students have enrolled in the MEEN241. These data will clarify the role of PHYS241 as an introductory course to MEEN241, and to discover the process used by successful students in the course during the baseline study and after the implementation of experiential learning pedagogy.

The current study only has 39 cases. Notwithstanding the relatively low number of cases, the study has uncovered important relationships that needed to be researched further. The central question could be redefined to include both Q and RQ. Also, it is important to further study the effect of PHYS241 on the pass-rate of MEEN241. The result of this preliminary study led to a modest fellowship award by the NCAT STEM Center of Excellence for Active Learning to implement experiential learning pedagogy in this course.

Reproducible Research

The measured and analytic data are stored in the repository given below:

https://github.com/paulAkan/ThermoData

References

1 Dweck, C. S., Mindset: The new psychology of success, Random House Digital, Inc., 2008.

2 WordNet Search, Retrieved from http://wordnetweb.princeton.edu/perl/webwn?s=mindset, January 2017

3 Shell, M.S., Thermodynamics and Statistical Mechanics: An Integrated Approach. Cambridge University Press, 2014.

4 Karimi, A., and Manteufel, R. D., "Correlation of Prerequisite Course Grades with Student Performance" ASEE Annual Conference & Exposition, Atlanta, Georgia, ASEE, 2014.

5 Taruna, S., and Pandey, M., "An empirical analysis of classification techniques for predicting academic performance", IEEE International Advance Computing Conference (IACC), Gurgaon, 2014, pp. 523-528.

6 Lakshmi, T. M., Martin, A., Begum, R. M., and Venkatesan, V. P., "An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data", International Journal of Modern Education and Computer Science, 2013, pp. 18.

7 Quinlan, J. R., "Induction of Decision Tree", Journal of Machine learning, Morgan Kaufmann, 1986, pp. 81-106.

8 Nithyassik, B., and Nandhini, D. E. C., "Classification Techniques in Education Domain", International Journal on Computer Science and Engineering, 2010, pp. 1647-1684.

9 Cengel, Y. A., and Boles, M. A., Thermodynamics: An Engineering Approach. McGraw-Hill Education, 2015.

10 Lambert, A.D., Rocconi, L. M., Ribera, A. K., Miller, A.L., and Dong, Y., "Faculty lend a helping hand to Student Success: Measuring student-faculty interactions", Association for Institutional Research, New Orleans, Louisiana, 2012.

11 Ellucian Company LP, Retrieved from http://www.ellucian.com/, January 2017.

12 RapidMiner, Retrieved from www.rapidminer.com, February 2017.

13 RapidMiner User Manual, Retrieved from http://community.rapidminer.com/t5/tkb/communitypage, February 2017.

14 Quinlan, J. R., C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, Inc., 1993.

15 Felton, James and Koper, Peter T., "Nominal GPA and Real GPA: A Simple Adjustment that Compensates for Grade Inflation", Assessment & Evaluation in Higher Education, 2005, pp. 561-569.

16 Steinmayr, R., and Spinath, B., The importance of motivation as a predictor of school achievement. Learning and Individual Differences, 2009, 80-90.

18 Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

17 Dweck, C. S., The Business Source, Retrieved from http://ebsp.s3.amazonaws.com/pdf/mindsett.pdf, February 2017.

19 Mechanical Engineering Undergraduate Student Handbook 2016, Retrieved from http://www.ncat.edu/coe/departments/meen/ugprogs/MEEN_UG_HANDBOOK_2016_last.pdf, February 2017.

Paul Akangah

Paul Akangah is an Assistant Professor of Instruction at the North Carolina A&T State University. He earned his B.Sc. degree from Kwame Nkrumah University of Science & Technology, Kumasi, Ghana, M.Sc. (Energy Engineering, 2005) from The Royal Institute of Technology (KTH), Stockholm, Sweden, and Ph.D. (Mechanical Engineering, 2011) from the North Carolina A&T State University, Greensboro, NC. Dr. Akangah is currently teaching at the North Carolina A&T State University. His interests are in engineering pedagogy, metallic foam for thermal and structural applications, and advanced composites materials. Dr. Akangah may be reached at pmakanga@ncat.edu.

Leotis Parrish

Leotis Parrish is an Assistant Dean for Student Affairs in the College of Engineering at North Carolina Agricultural and Technical (NC A&T) State University. He earned a B.S. and M.S. degree in civil engineering from NC A&T State University and a Ph.D. in Higher Education Leadership from Northcentral University. In his role as an administrator and faculty member, Dr. Parrish develops and implements new programs for student recruitment/retention, student engagement, and leadership. In his role as a faculty member, he teaches the freshmen success seminar and an introductory engineering design course. His areas of interest are in student engagement and leadership development for students.

Andrea Ofori-Boadu

Andera Ofori-Boadu is an Assistant Professor of Construction Management at North Carolina A & T State University. She earned her B.S. degree (Building Technology, 1997) from the University of Science and Technology, Ghana, MS (Industrial Technology, 2004) from North Carolina A & T State University, and Ph.D. (Technology Management, 2012) from Indiana State University. Her research interests are in bio-modified cements, quality management practices, and STEM education. Dr. Andrea Ofori-Boadu may be reached at andreao@ncat.edu.

Francis Davis

Francis Davis is a Senior Lecturer of Mechanical Engineering at Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. He earned his B.S. degree from Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, MS (Mechanical Engineering, 2004) from Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, and Ph.D. (Mechanical Engineering, 2010) from the North Carolina Agricultural and Technical State University, Greensboro, North Carolina. Dr. Davis is currently teaching at the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. His interests are in design and manufacturing, maintenance and installation, design of experiment, and modeling. Dr. Davis may be reached at fdavis.coe@knust.edu.gh.