VIRGINIA COMMONWEALTH UNIVERSITY

Predicting Failure/Success Probabilities of Students Enrolling in a Bottleneck Course

SAIR 2016

Khoi D. To, PhD
Enterprise Analytics and Advanced Research
VCU Office of Planning and Decision Support

Overview

- Introduction
- Problem Identification
- Approach
- Data, Methods, and Software
- Findings
- Implications
- Further Study

Introduction

❖ **Virginia Commonwealth University (VCU)**

- A major public research university located in Richmond, the state capital of Virginia.
- Classified as a Research University-Very High Research Activity, the highest ranking by the Carnegie Foundation.
- Total enrollment of 32,000; 222 degree and certificate programs, 67 of which are unique in the state of Virginia.
- One of the largest academic health centers in the nation. The VCU Medical Center was named the No. 1 hospital in the state in 2013 by U.S. News & World Report.

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Problem Identification

- Bottleneck courses: courses with high enrollment and high percentage of students receiving **D/F/W** final grades.
- They have been a common problem nation-wide and can have consequences for both institutions and students:
  - ✓ Slower progress towards graduation and lower graduation rates → adversely impacting overall performance of institutions and students
  - ✓ Increased cost of attendance for students and operational costs for institutions.

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Problem Identification (Cont.)

CHEM 101	FINAL GRADE DISTRIBUTION				TOTAL
	SUCCESS (A/B/C)		FAILURE (D/F/W)		
	#	%	#	%	#
Spring 2012	546	57.7	401	42.3	947
Fall 2012	831	68.1	390	31.9	1,221
Spring 2013	614	64.4	339	35.6	953
Fall 2013	902	71.4	361	28.6	1,263
Spring 2014	496	58.4	354	41.6	850
Fall 2014	807	60.7	523	39.3	1,330
Spring 2015	512	55.8	405	44.2	917
Fall 2015	899	58.4	640	41.6	1,539
Spring 2016	444	50.1	443	49.9	887

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Problem Identification (Cont.)

Is there a way VCU can identify "at-risk" students before the final exam takes place???

- → To provide timely feedback to students
- → To identify weak students so that they can be provided with further academic assistance
- → To establish a framework for an adaptive tutoring/student support system in the future

(Bydžovská, 2015)

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Approach

- Using logistic regression as a data mining technique to predict the probability of a student failing(*) a bottleneck course based on a set of available predictors, **progressively—before mid-term exam and after mid-term exam.**
- (Most studies on this topic used all performance indicators available right before the final exam (quizzes, home assignments, mid-term exam, class participation…).

  → It may be too late to intervene for better final grades.)

(*) *Failing means getting D/F/W as final grade.*

Limitations

- Mid-term grades of several students were missing. Those missing values were coded as "U" (Unknown). If all mid-term grades had been available, it could have improved the overall predictability of the model.
- Grades of quizzes, home assignments, class participations… were not available. Those data could have been useful in predicting a student's failure or success, especially at an early stage (before the mid-term).

Data

- Data collected from Banner ODS modules (Admissions, Enrollment, Courses, Financial Aid)
- Students taking CHEM 101 in two most recent semesters (Fall 2015 and Spring 2016; 2,523 observations)
- Final grades serve as the target variable. A/B/C final grades are grouped as "SUCCESS" and D/F/W final grades as "FAILURE".
- Four groups of predictors: demographics, financial aid, pre-college, and in-college

Data (Cont.)

Demographics	Financial Aid
Residency	Dependent/Independent
Gender	Applied for FASFA or not
Race/Ethnicity	Amount of Pell grant received
First generation	Percent of need met
	Median income of zip code

Data (Cont.)

Pre-College	
High school GPA	SAT Combined/Math/Verbal
IB/AP credits recognized	
In-College (before current semester)	
First-time (N) or transfer (T)	Cumulative credits with grades D/F/W
Number of times repeating the course	Latest academic standing

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Data (Cont.)

College (current semester)	
STEM major	Athlete/Honors
Full-time/Part-time	Term credits attempted
Student class (FR/SO/JR/SR)	Instructors for the course
College/School enrolled	On-campus/Off-campus

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Methods and Software

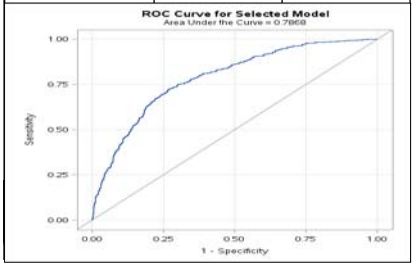
- Software: SAS Base
- Data mining technique: logistic regression
- Missing values are imputed.
- Two-step analyses:
  - → Failure/Success prediction without mid-term grades
  - → Failure/Success prediction with mid-term grades
- Original data set is split into two: 60% for training and 40% for validation.

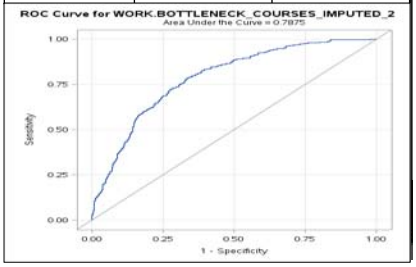
VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

A - Failure/Success prediction (without mid-term grades)

Training (1,412 obs)	Predicted	
	FAILURE	SUCCESS
	Row %	Row %
Actual FAILURE	67.52	32.48
SUCCESS	22.74	77.26

Validation (941 obs)	Predicted	
	FAILURE	SUCCESS
	Row %	Row %
Actual FAILURE	65.05	34.95
SUCCESS	23.38	76.62





A – Failure/Success prediction (Cont.)

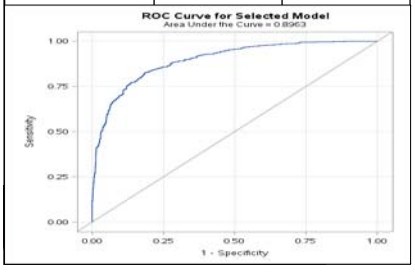
- Model assessment: Goodness-of-Fit Test (Hosmer and Lemeshow, Chi-square=9.752, p=0.283); Area under ROC curve: 0.787 (training) and 0.788 (validation)
- Significant predictors: high school GPA, SAT Math, cumulative GPA, and instructor teaching the course.
- **65%** of those who would eventually fail the course could be predicted correctly **even before the first class started**.

  → Early monitoring can be set up for those identified as "at-risk" students.

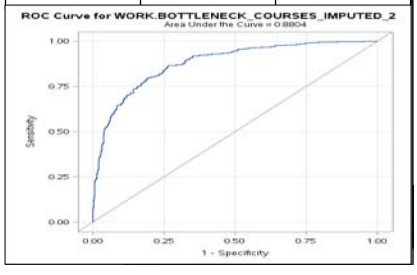
VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

B - Failure/Success prediction (with mid-term grades)

Training (1,412 obs)	Predicted	
	FAILURE	SUCCESS
	Row %	Row %
Actual FAILURE	74.40	25.60
SUCCESS	12.07	87.93



Validation (941 obs)	Predicted	
	FAILURE	SUCCESS
	Row %	Row %
Actual FAILURE	74.31	25.69
SUCCESS	15.52	84.48



B – Failure/Success prediction (Cont.)

- Model assessment: Goodness-of-Fit Test (Hosmer and Lemeshow, Chi-square=1.781, p=0.987); Area under ROC curve: 0.880 (training) and 0.880 (validation)
- Significant predictors: mid-term grade, high school GPA, SAT Combined, cumulative GPA, gender, and instructor teaching the course.
- **74%** of those who would eventually fail the course could be predicted correctly **after mid-term grades are available**.
  - → Warning can be sent to those at-risk.

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

B – Failure/Success prediction (Cont.)

- There are a few things at-risk students can do upon receiving warning:
  - → Seeking academic assistance from instructors/tutors in order to do better in coming quizzes, home assignments, and final exam
  - → Withdraw from the course to avoid an F and retake the course later. (Dropping a class may have several consequences, though--not enough cumulative credits for graduation, financial aid, health insurance…)

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Further Study

- If available, grades of quizzes and home assignments before mid-term can also be used as inputs for better prediction.
- Social behavior data can be added to improve the model: posts and comments in discussion forums, e-mails statistics, publication co-authoring, or files sharing. (Bydžovská, 2015)
- Optimal Equal Width Binning and Synthetic Minority Oversampling Technique (SMOTE) can also be employed for better accuracy. (Jishan et al., 2015).

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Thank You

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

References

- Bydžovská, H. 2015. A Comparative Analysis of Techniques for Predicting Student Performance. *Proceedings of the 9th International Conference on Educational Data Mining,* pp. 306-311.
- Jishan, S. T. et al. 2015. Improving Accuracy of Students' Final Grade Prediction Model Using Optimal Equal Width Binning and Synthetic Minority Over-sampling Technique. *Decision Analytics (2015) 2:1.*
- Majeed, E. A. and Junejo, K. N. 2016. Grade Prediction Using Supervised Machine Learning Techniques. *e-Proceeding of the 4th Global Summit on Education 2016.*
- Meier, Yannick et al. 2016. Predicting grades. *arXiv:1508.03865v2 [cs.LG] 18 Mar 2016.*