VCU
VIRGINIA COMMONWEALTH UNIVERSITY

Predicting Six-Year Graduation Probabilities of First-Time Full-Time Freshmen

CSRDE NSSR 2016

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

Overview

- Introduction
- Problem Identification
- Conceptual Framework
- Data, Methods, and Software
- Findings
- Further Study

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

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.

Key Metrics on Student Success

	First-year retention (F2014 cohort)	Six-year graduation (F2009 cohort)	High-school GPA (F2015 cohort)	Student/ Faculty Ratio (F2015)
Virginia Commonwealth University	86% **(4)**	62% **(4)**	3.64 **(4)**	16:1 **(1)**
University of Alabama, Birmingham	79%	55%	3.66	18:1
University of Cincinnati	88% **(1)**	65%	3.48	18:1
University of Illinois, Chicago	82%	60%		
University of Louisville	79%	53%	3.60	16:1 **(1)**
University of South Carolina	88% **(1)**	72% **(1)**	4.07 **(1)**	18:1
University of South Florida, Tampa	88% **(1)**	68%	3.94	24:1

(All numbers were obtained from 2015-16 Common Data Set.)

Problem Identification

- VCU six-year graduation rate is 62% → there is still a lot of room for improvement.
- Enhancing six-year graduation rate lines up with VCU's commitment to student success.
- Higher graduation rates:
  - → Higher institutional reputation and ranking
  - → Less costs for students and their families
  - → More meaningful achievement for both institution and students

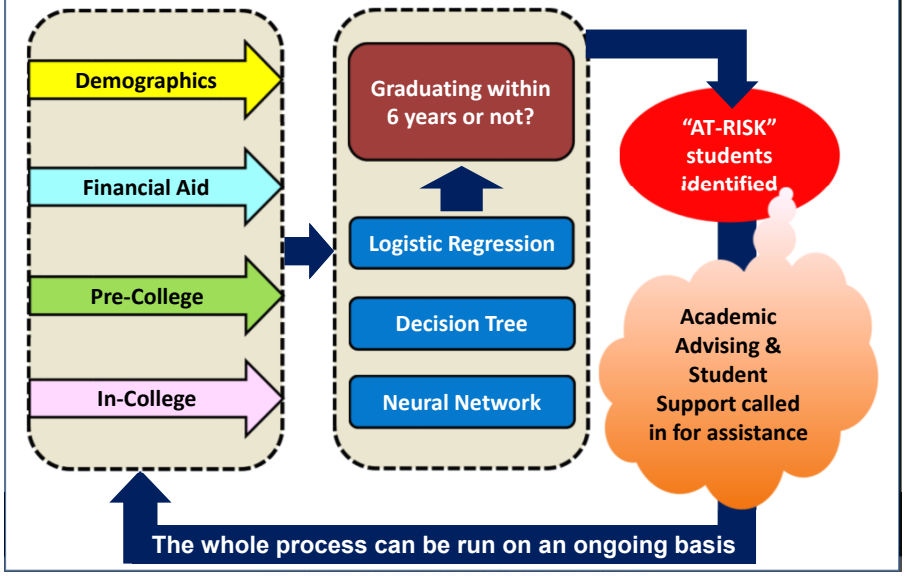
VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Problem Identification (Cont.)

Being able to predict "at-risk" students at an early stage (after the first semester) and provide them with necessary assistance to graduate in time is crucial for all parties involved.

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Conceptual Framework



Data

- Data collected from Banner ODS modules (Admissions, Enrollment, Financial Aid) after the first semester (fall)
- Fall 2009 full-time first-time freshman cohort (3,644 students)
- 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

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

Data (Cont.)

Pre-College	
High school GPA	SAT Combined/Math/Verbal
IB/AP credits recognized	
In-College (at the end of the first semester)	
Transfer hours recognized	On-campus/Off-campus
STEM major	College/School enrolled
Student class (FR/SO/JR/SR)	Athlete/Honors

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

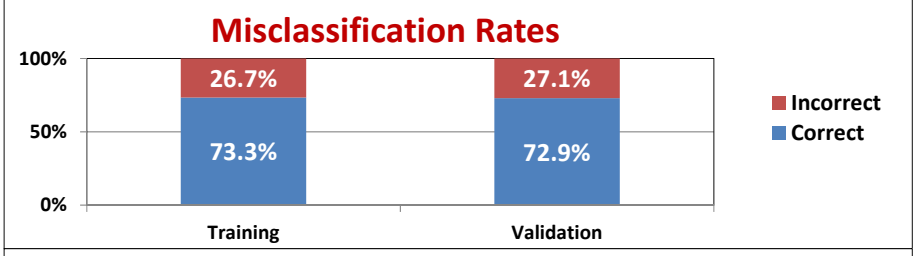
Data (Cont.)

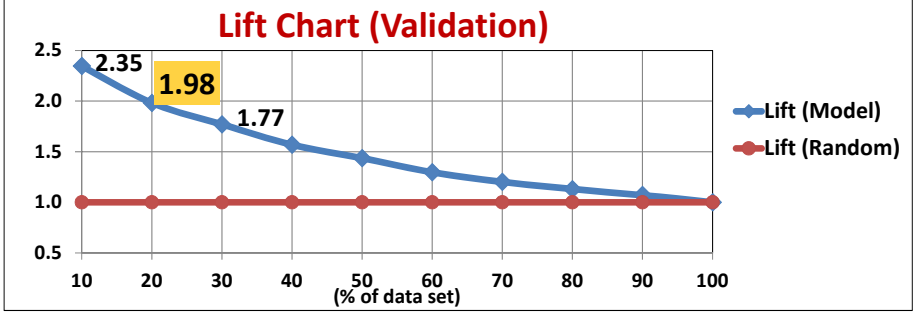
In-College (at the end of the first semester)	
Term hours attempted/earned	Number of Math/Physics/Chemistry courses taken
Term quality points/GPA hours	Academic standing
Number of D/F/W grades	Applied for transcript or not

Methods and Software

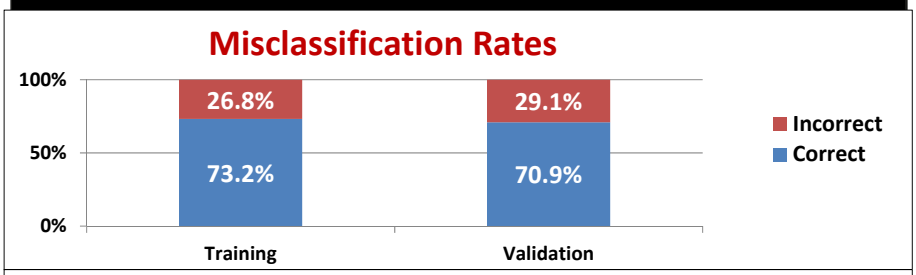
- Data are extracted and prepared with SAS Base.
- Imputation for missing values and modeling tasks are done with SAS Enterprise Miner.
- Three techniques: logistic regression, decision tree, and neural network models. The best model is selected based on misclassification rates.
- Original data set is split into two: 60% for training and 40% for validation.

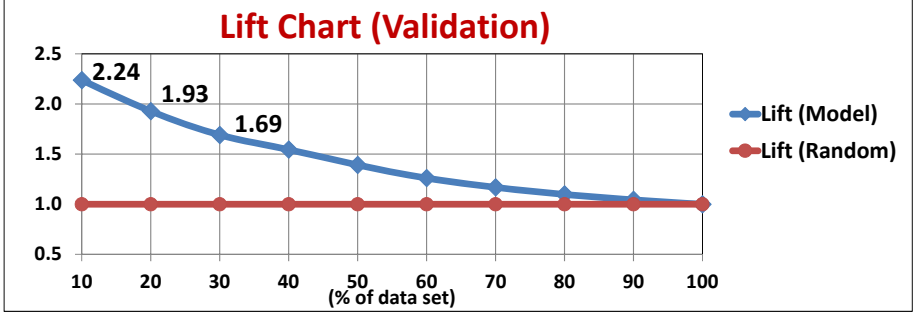
1 – Logistic Regression



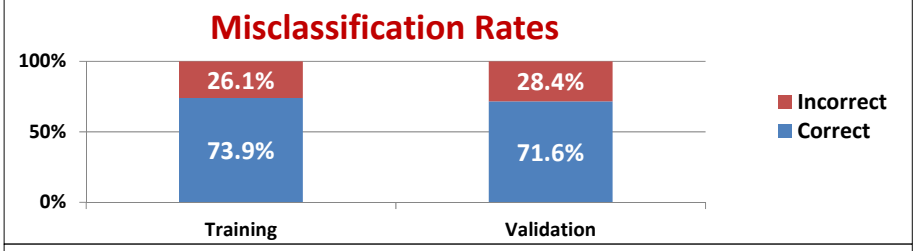


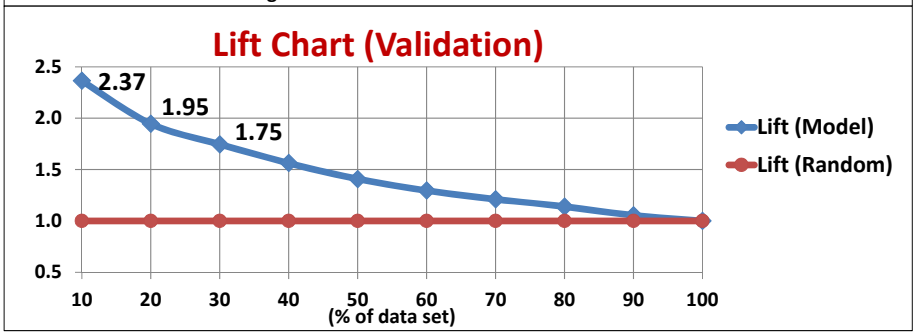
2 – Decision Tree





3 – Neural Network





Findings

- Logistic regression model is chosen based on misclassification rate, lift chart, and easiness of model interpretation.
- Significant predictors: first generation, academic standing, college, cumulative hours earned/GPA, SAT Verbal, high school GPA, percent of unmet need, and applied for transcript or not.

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Findings (Cont.)

- The lift of logistic regression at 20% is **1.98** → if the top 20% (sorted by highest to lowest probabilities) of total cohort were selected, the number of "at-risk" students captured by the model would be 1.98 times as many as when 20% of total cohort were selected at random.

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Findings (Cont.)

Total cohort (3,644 students): graduated in 6 years (2,269 students, or 62%), did not graduate in 6 years (1,375 students, or 38%)

	No modeling (random)	**Logistic regression**
Selecting 20% of total cohort	3,644*20% = 729 students	
→ **Number of "at-risk" students captured**	729*38% = 277	277***1.98** = 548
→ **If 70% of those "at-risk" students were helped to graduate in 6 years**	277*70% = 194	548*70% = 384
→ **Improved six-year graduation rate by targeting 20% of the total cohort**	(2,269+194)/3,644 = **68%**	(2,269+384)/3,644 = **73%**

Further Study

- Other variables can be introduced to the models to improve accuracy: average time spent in library, intent to complete a degree program (from SAT/ACT record)…
- Cluster analysis can be conducted on the predicted non-graduates to see if they shared any common characteristics.

Other Applications

- "High-risk" students can be monitored continuously on a semester basis and passed onto Academic Advising and Student Support for help so that they can graduate in time.
- A customized model can be developed for each school/college to help keep track of the progress of their own students.

Other Applications (Cont.)

- The same framework and methods can be developed to predict students' probabilities of **retention/attrition** at various levels (university, school/college, or department).

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

Thank You

VCU
VIRGINIA COMMONWEALTH UNIVERSITY
Make it real.

References

- Pyke, S.D. & Sheridan, P.M. Logistic regression analysis of graduate student retention. *The Canadian Journal of higher education, Vol. XXIII-2, 1993.*
- Bogard, M., Helbig, T., Huff, G., & James, C. A comparison of empirical models for predicting student retention. https://www.wku.edu/instres/documents/comparison_of_empirical_models.pdf

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.

References (Cont.)

- Karimi, A. Predictive modeling of student success. (Coffee Talk 11/15/13) https://www.fullerton.edu/analyticalstudies/_resources/pdfs/irct5.pdf
- Raju, D. Predicting student graduation in higher education using data mining models: A comparison. (Dissertation, 2012) http://acumen.lib.ua.edu/content/u0015/0000001/0000901/u0015_0000001_0000901.pdf

VCU VIRGINIA COMMONWEALTH UNIVERSITY Make it real.