



---

*United States Military Academy*

---

# VALUE ADDED OUTCOMES IN MILITARY SERVICE EDUCATIONAL PROGRAMS:

## Comparing New and Old Technologies and Methods

LTC Rod Sturdivant and MAJ Krista Watts

United States Military Academy, West Point, New York

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

# Outline

- Introduction
- Goals
- Methods
- Example
- Conclusions
- Future Work

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

# Introduction

- Focus includes predictors known prior to admission to USMA and corresponding performance and choices at USMA
- USMA has a long-standing interest in candidate qualifications and the subsequent performance of those selected for admission
- Candidates submit a host of information in application for admission, including: GPA, class rank, test scores and measures of leadership potential and also answer “subjective” survey questions related to education and leadership development prior to USMA and personality traits
- Once selected, attention focuses on the academic, physical, and military performance, and retention while at the military academy – and as an officer
- Traditional evaluation techniques used to examine how admission variables are related to performance/retention include chi square, correlation, regression
- Present study compares traditional evaluation techniques with a broader approach utilizing several data mining approaches

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

# Goals

- Explore traditional statistical techniques and new multi-disciplinary methods
  - Create a transportable methodology to apply to similar problems
- Identify traits that indicate potential for success in the Army

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

# Method

- Data were drawn from the USMA Class of 1989 (N = 1425 at entry in 1985 and 1059 at graduation in 1989)
- Predictors include a variety of measures taken prior and upon entry to USMA (including demographic and aptitude data) and cadet performance indexes, finalized at the time of graduation (including academic, physical and military performance scores)
- Criterion measures include retention, military rank, and selection for Army schools
- Statistical analyses were conducted independently utilizing either traditional or newer approaches examining the predictor-criterion relationships

---

*Duty, Honor, Country*

---



# Same Problem, Different Techniques

- Traditional
  - Software: SAS
  - Algorithms: Multivariable linear regression, multivariable logistic regression, Survival Analysis
- Contemporary
  - Software: Clementine modeling software with SPSS statistical package
  - Algorithms:
    - Decision Trees
    - Neural Networks
    - Clustering
    - Association Rules
- Combinations
  - Principle Components and Factor Analysis
  - Combining Models



---

*United States Military Academy*

---

# Example

- Measure of Success – Commissioned Officer remained in service past Active Duty Service Obligation (ADSO)
  - Exceeded ADSO defined as 61 months
  - Removed Aviation, Special Forces, and Medical Corps
- Predictors
  - Admissions, USMA performance, and other demographic data
  - Survey data

---

*Duty, Honor, Country*

---



# “A” Logistic Regression Model

Predictor	p-value
Final Military Development average	0.015
Basic branch army (CA,CS, CSS, other)	0.039
Q3 – any honors or AP course in HS	0.044
Q22 – student council president	0.025
Q35 – Eagle scout	0.008
Q46 – Strength exercises done	0.045
Q84 – USMAPS tactical officer	0.005
Q87 – 1 <sup>st</sup> priority for college	0.005
Q94 – Decision at last moment	0.027
Q5 – 1 <sup>st</sup> or 2 <sup>nd</sup> in HS class	0.082
Q100 – Had few options	0.070

- Overall model  $p < 0.0001$
- Misclassification rate (using test set of 185): 36.82%



# Logistic Interpretation

- Higher Military Development grade increases odds of remaining past ADSO (1.5 times as likely for increase of 1 grade)
- CSS only 0.5 times as likely to remain as CA
- No honors/AP 1.4 times as likely to stay
- Student body presidents and Eagle scouts nearly 2 times as likely to remain past ADSO
- No or negative impact of USMAPS tactical officer decreases odds of staying past ADSO
- Priority for college “Army career opportunity” more than 2 times as likely to stay than “Quality of Academic program”
- Those waiting until the last possible moment to decide on USMA MORE likely to stay past ADSO
- Top of HS class (1<sup>st</sup> or 2<sup>nd</sup>) MORE likely to stay past ADSO
- Those with no options for college more likely to stay past ADSO
- Hard to interpret...those doing pushups and situps remain past ADSO more than those doing just pushups



# Classification Tree

- Key Variables:
  - Basic Branch in Army
    - AG and FI more likely to stay in
    - AD, CM, MI, MP, MS, OD, QM, TC more likely to get out
  - Minority Code
  - How often did you do strengthening exercises
  - Having to listen and not talk back
  - Camp Counselor
  - USMA Liaison Officer
  - Rank in High School
  - I made up my own mind
  - I waited until the last possible moment to decide
- Misclassification Rate – 42.33%



---

*United States Military Academy*

---

# Neural Network

- Key Variables
  - First Priority
  - I waited until the last possible moment to decide
  - Being neatly and properly dressed at all times
  - I had few options
  - Basic Branch
- Misclassification Rate – 44.97%

---

*Duty, Honor, Country*

---



# Comparison of Techniques

- Naïve Model
  - Misclassification Rate – 48.68%
- Logistic Regression
  - Best Misclassification Rate – 36.82%
- Common Key Variables:
  - Branch
  - Decided at last moment
  - Few Options
- Classification Tree
  - Best Misclassification Rate – 42.33%
- Neural Network
  - Best Misclassification Rate – 44.97%



# Logistic Regression Issues

- Missing Data
  - Excluded ACT and 2 questions in step-wise
  - SAT missing several values
- Small “cell sizes”
  - Categorical variables with small numbers in some categories
  - “Collapsing” possible but takes time (collapsed branches)
  - Creates equivalent linear combinations of predictors
- Several “good” competing models
- Correlations among predictors – confounding makes model fitting difficult
- Number of potential predictors (even with reasonable data set) – model supports at most
- Model fit very time consuming – over 80 models fit after initial stepwise (for model using all predictors)



# Classification Tree Issues

- Vastly different results with different training/test sets
  - Test set possibly not representative
- Easy to interpret but hard to understand results
  - If your basic branch at USMA was Engineers and you were not a camp counselor and you did strengthening exercises 3-5 times per week in high school and your high school rank was above 40 and if you never had contact with a USMA liaison officer or your impression of him was very negative – then predict that you will stay in past your ADSO
- Not always logical splits



# Neural Network Issues

- Easily Influenced by irrelevant predictors
- Many different options/algorithms can be overwhelming
- Optimal choices depend largely on data set (not easily adaptable)
- Could not replicate the results without Clementine (Black Box)



# Strengths and Weaknesses

- Traditional Techniques

- Strengths

- Easily interpretable
- Replicable Results
- More widely understood

- Weaknesses

- Can be unwieldy with large # of variables
- Difficult to deal with data issues – tempting to give up

- Non-traditional Techniques

- Strengths

- Data easily manipulated
- Can handle large data sets and large # of variables relatively quickly
- “What-ifying” fairly easy

- Weaknesses

- Different algorithm/seed can give vastly different results
- Often not easily interpretable
- Modeling problems may not be obvious
- Black Box



# Future Efforts (Traditional)

- More work on categorical variables
  - Collapsing categories
  - Check for sparse cells
  - Consider whether meaningful
- Factor analysis (identify correlated predictors) as predictors
- Survival analysis (time in Army not just past ADSO)
- Other outcomes...



---

*United States Military Academy*

---

# Future Efforts (Non-traditional)

- Other Algorithms
- Combining Techniques
- Clustering and Association Rules
- Principle Components as predictors
- Other outcomes...

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

# References

Butler, R. P. (1974). Correlates of Officer Performance. Report No. 0B1.02-74-021. Office of the Director of Institutional Research, United States Military Academy, West Point, New York.

Houston, J. (1979). Attrition and Admission Scores, Classes of 1979-80-81. Report No. 80-001. Office of the Director of Institutional Research, United States Military Academy, West Point, New York.

Burke, W. (1987). Indicators of Non-selection to Captain for USMA Class of 1984. Report No. 87-005. Office of the Director of Institutional Research, United States Military Academy, West Point, New York.

Introduction to Clementine (Rev. ed.). (2005). Chicago, IL: SPSS Inc.

Johnson, Mike LTC, Office of Plans and Analysis, USMA

Matthews, Mike Dr., Department of Behavioral Sciences and Leadership, USMA

Kelly, Dennis Dr., Office of Plans and Analysis, USMA

Moore, Kerry MAJ, Department of Mathematical Sciences, USMA

---

*Duty, Honor, Country*

---



---

*United States Military Academy*

---

? ? ? ? ?  
**QUESTIONS**  
? ? ? ? ?

---

*Duty, Honor, Country*

---