CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, December 5
10:15 – 10:45 AM
Davidson Hall, Room 207

USING DATA MINING TO PREDICT SUCCESS IN A NURSING DEGREE PROGRAM

BENJAMIN DICKMAN

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: The Nursing department of a community college wanted to know if data mining could help them increase the student success rate in the Associate's Degree in Nursing program. Specifically, they wanted to know if data mining could predict students' grades in the "gatekeeper" first course, NURS1. This study reports the results of predicting the NURS1 numeric grade (range 0 to 100) using three modeling techniques that were compared with one another: Linear Regression, Neural Network and Random Forest. Of these three, the Linear Regression model had the lowest Mean Absolute Error. This error measure was lowest (just over 4 points, which is half of a letter grade step) for the letter grades C and C+, which also define the border between failure and success in the course. Thus, the identification of which students are borderline is a valuable outcome of the study, since it allows the faculty to focus intervention efforts on those students most likely to benefit.

Models were also generated to predict success or failure in the course as a binary target variable. The motivation was to enable the Nursing faculty to see the effects on the success rate of changing program admission criteria. The modeling techniques AdaBoost, Random Forest and Support Vector Machine gave overall error rates of under 25%, which were superior to the techniques Logistic Regression, Neural Net, Linear Model and Decision Tree. Missing data was filled in using multiple imputation, which gave slightly more accurate predictions than model segmentation. This study used the R programming language to generate graphics and to analyze the data.