CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, December 5
11:00 – 11:30 AM
Davidson Hall, Room 207

CLASSIFICATION MODELING
OF FREDDIE MAC HOME LOAN DELINQUENCY

ANDREW HENDRICKSON

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: Models developed to classify home loan delinquency of Freddie Mac data appear to offer substantial improvement, in terms of bank dollars gained, over baseline measures. Pas t economic research indicated the FICO score as a key variable in home loan delinquency modeling. The FICO score was found to be a key variable for classification models built in this study as well. Other research studies indicated that the riskiness of the borrower pool at the regional level is a factor in keeping a mortgage current (ie non-delinquent). For this study, separate data mining analysis was performed for mortgages from Minnesota (MN) and Florida (FL). The regional borrower risk was examined using the FICO score field in relation to the loan delinquency status for both MN and FL sets. For a FICO score range of 500-800, MN had a delinquency rate from 80% to 40% but FL had a delinquency rate from 80% to 20%. The data suggests that an 800 FICO score in FL is less of a loan risk (20% delinquency) than an 800 FICO score in MN (37% delinquency). Due to the coincidence of the mean loan amount requested being approximately equal to the mean loan interest, data driven misclassification costs were not able to be used for this analysis. A simple baseline measure was calculated for holdout test sets both MN and FL with the Bank of MN loans having a $309,000 loss and the Bank of FL loans having an $8,048,000 loss. Applying this measure for the classification model delinquency predictions, the Bank of MN loans would have a gain of $215,957,000 and the Bank of FL a gain of $410,015,000.