CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Thursday, April 19
12:00 – 1:30 PM
Maria Sanford, Room 203

COMPARING CLASSIFICATION ALGORITHMS IN DATA MINING

SAMPSON ADU-POKU

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: In this thesis, support vector machines, neural networks, logistic regression, naïve Bayes, classification and regression trees, the C4.5 algorithm, QUEST, CHAID, and discriminant analysis have been implemented on nine real-world data sets. The goal of the research is to evaluate the performance of the classification algorithms on both multiclass problems and binary classification problems using a variety of performance metrics, such as precision, recall (sensitivity), specificity, and lift charts / gains charts. We found that there is no universally best classification algorithms. Rather, the optimal algorithm depended on the type of problem being considered, the dataset characteristics, and the performance metrics used.