CCSU

DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, December 4
10:15 – 10:45 AM
Davidson Hall, Room 204

DATA ANALYTIC APPROACHES TO PREDICTING SUCCESS IN BANK TELEMARKETING

CURTIS LENTZ

(Data Mining MS Thesis Presentation)

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: This paper uses a large data set from a Portuguese bank to use a data analytic approach to improving the results of telemarketing campaigns. The data was collected from 2008 to 2013, and was obtained from the University of California at Irvine's Machine Learning Repository. It contains 41,118 records and 20 variables, including the target variable of "yes" as a customer's response to the banks offer of a long-term deposit contract. The modeling is performed using IBM's Modeler software.

The first section of the paper establishes a foundation for Customer Relationship Management as a powerful and growing method of using data analysis techniques to better understand a company's customers and their relationship with the business. It also outlines the CRISP-DM approach to organizing a data mining project.

The data is analyzed in detail, including performing Exploratory Data Analysis (EDA) on all variables to get a detailed view of the data and the distributions of the variables, as well as an understanding of the relationship between the variables and the dependent variable. Further, Principal Component Analysis and K-means clustering is conducted on the data to identify potential correlations in variables and how groups of customers might respond similarly.

Logistical regression is used on four different models, the results are described and compared across a variety of criteria. Using the profit criteria, the model with the highest revenue per customer is selected as a model to be used for prediction of how to select the best potential customers for a future campaign. Four other modeling techniques: CART, C5.0, Support Vector Machines, and Neural Networks are tested to see if the predictive results can be improved. Logistic regression is recommended as the best approach due to its higher profit per customer, and its relative ease in interpreting its results over the other modeling techniques. The thesis ends with conclusions, recommendations and suggestions for future research.

For further information:

gotchevi@ccsu.edu 860-832-2839
http://www.math.ccsu.edu/gotchev/colloquium/