

CCSU  
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, April 25  
12:30 – 1:00 PM  
Davidson Hall, Room 207

**PREDICTING CHANGE IN COUNTY-LEVEL  
PRESIDENTIAL ELECTION VOTER TURNOUT  
USING DATA MINING METHODS**

**RICHARD ACEVES**

(Data Mining MS Thesis Presentation)

**CENTRAL CONNECTICUT STATE UNIVERSITY**

**Abstract:** This thesis explores the use data mining techniques to model voter turnout change in the 2012 Presidential Election. Specifically, the percentage of difference in turnout from the 2008 Presidential Election is examined at the county level. A majority of counties experienced a decrease in turnout of one to six percent, however, the differences ranged from a decrease of 31 percent to an increase of 28 percent. In all, eight models were created. The first model was a regression model which used  $k$ -means clusters as inputs. The  $k$ -means clustering algorithm grouped counties according to their socio-demographic profiles. A second regression model used Principal Components, a dimension reduction technique which allowed for 80 percent of the information in the dataset to be represented by twelve indices. The next three models employed the variable selection techniques of Forward Selection, Backward Elimination, and the Stepwise Procedure. In addition to the five regression models, a CHAID decision tree model and a Neural Network model were created. Of the seven models, the Forward Selection and Stepwise regression models proved most accurate against the holdout test set, which was set aside for model scoring. These two models showed a 35 percent improvement over the baseline model. The eighth and final model performed the best, with a 42.5 percent improvement over the baseline. This model was a combination model, incorporating the predictions of the Stepwise, CHAID, and Neural Network models.

***For further information:***

[gotchevi@ccsu.edu](mailto:gotchevi@ccsu.edu) 860-832-2839

<http://www.math.ccsu.edu/gotchev/colloquium/>