

CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

DATA MINING
THESIS PRESENTATION

Wednesday, May 4
12:00 – 1:00 pm in NC 155

**NETPIX: A Method of Feature
Selection Leading to Accurate
Sentiment-Based Classification
Models**

James Steck

(MS candidate in Data Mining)

Abstract: The online DVD rental company N E T F L I X® says that their service “is the best way to rent movies.” Though N E T F L I X® claims they enable customers to “find and discover movies they will enjoy”, consistently renting movies that meet personal tastes and standards still remains an elusive task. An intelligent data mining model that recommends movies according to each viewer’s personal preferences—their “net picks”, so to speak—would likely increase customer satisfaction. Researchers have proposed several techniques that accurately classify the underlying sentiment found in reviews. In several cases, these techniques rely on adjectives as likely indicators of subjectiveness, sentiment, or opinion. This thesis describes a method that extracts useful features from a collection of movie reviews and uses them to build data mining models capable of accurately classifying a new review as either “Good” or “Bad.” The experiments described in this thesis use attribute selection methods in WEKA to evaluate each feature’s relevance, with respect to the task of movie review classification. Subsets of the ranked features are then programmatically input to Bayesian-based classifiers in WEKA to generate classification results. These methods are proven to produce highly-accurate classification models with results often more competitive than those reported in current literature.

Refreshments will be served.

For further information:

gotchevi@ccsu.edu (860) 832-2839