**COSC 757 Data Mining Assignment 4**

Due date: 04/27/2016 11:59 pm

Instructions: This is an individual assignment. Use Blackboard to submit your answers on the due date (no hard copies please). Late submissions will receive a zero grade.

**Applied Frequent Itemset or Association Rule Mining:**  Choose a dataset that is well suited for frequent itemset or association rule mining. You can use any dataset that you would like to mine. A good number of datasets can be found in the UCI machine learning data repository (<https://archive.ics.uci.edu/ml/datasets.html>) but feel free to use any dataset that you want. You will want to stick with datasets that are **categorical** in nature. Categorical datasets can be found in the UCI Machine Learning Repository by selecting Categorical link in the Attribute Type box on the left site navigation menu. Numerical datasets will have to be discretized so that itemsets can be created.

Once you have selected a dataset, you can then use a tool such as the arules package in R, Weka, or RapidMiner to mine frequent itemsets or association rules in the dataset.

The deliverable for this project will be a report that details your experiment. The report should be in either **ACM or IEEE conference paper format** and should include an introductory section that details the dataset and the objectives of the analysis, a methodology section that explains the approach that you used to mine the dataset including the algorithms and parameters (e.g. confidence and support) as well as any steps that you had to take to preprocess the data, a results section that shows the results of your analysis and any interesting patterns that you found, and a conclusion section that summarizes your results and discusses the limitations of your approach and any difficulties that you had with your experiment.

Links to format templates:

<http://www.ieee.org/conferences_events/conferences/publishing/templates.html>

<http://www.acm.org/sigs/publications/proceedings-templates>