## Week 6 Weka Assignment (75 points)

## **Assignment Due Date and Time**

* Sunday at 11:59 p.m. ET.

## **Purpose**

To gain an understanding of how to apply the apriori algorithm.

## **Description**

Consider the **usnews** data set. This dataset contains college data taken from the U.S. News & World Report’s Guide to America’s Best Colleges.

You are to construct association rules over the whole data set using the **apriori** algorithm available in the associate tab of Weka.

The data set represents raw problem domain data which you will need to first translate into the ARFF format and transform in order to build association rules. Once you have the data in the appropriate format you might want to consider the following data preparation questions:

1. Should instances with missing values be deleted?
2. Should attributes with missing values be deleted?
3. Should missing values be coded in a special way and then used in the data mining task?
4. Are some of the attributes categorical even though their levels are expressed as numbers?

The apriori algorithm can only handle categorical data, therefore you will need to discretize the attributes. For this you should use the “discretize” filter available in Weka.

Your report should discuss the following questions:

1. What are your top five rules?
2. What is their support/confidence?
3. Are they intuitive from the domain perspective or completely surprising?
4. How do the association rules change if you increase/decrease the number of bins (levels) in the attributes?

## **Your Assignment Report**

## Write a description of your experiment and your findings and submit this together with the discovered association rules and answers to the questions above.

## **Submission Instructions**

## Upload your written report to the Week 6 Weka Assignment submission area.