Homework Assignment 3

IT 531, Summer 2018

**Question 1:** Association Rule Analysis

Perform Association Rule Analysis (using Apriori algorithm) using R on following dataset:

Cars Dataset: <https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

Find more details for each dataset on the UCI repository page.

Deliverables:

1. Briefly describe the dataset.
2. Find top 25 rules for each dataset. Briefly describe top five rules.
3. Do some rule pruning. Make sure right hand side of the rule is the class attributes.
4. Submit a zip file containing the data and R script file and the document describing dataset and output results

**Question 2:** Classification Analysis

Perform classification using Decision Tree and Naïve Bayes using R on following dataset:

Spambase Dataset - <http://archive.ics.uci.edu/ml/machine-learning-databases/spambase/>

You can find more details for the dataset on the UCI repository page. For each algorithm and each dataset, do 10-fold cross validation.

Deliverables:

1. Briefly describe the dataset.
2. Generate confusion matrix on the test data. Calculate TP Rate, FP Rate, Precision and Recall with hand. Describe how you calculate these numbers.
3. Submit a zip file containing the data and R script file(s) and the document describing dataset, output results and TP, FP, precision and recall.