**Computer Science 361 - Machine Learning**

Assignment 3 – Worth 5%

Due: 1st of May, 2020.

**Title: Association Rule mining**

**Data**

Supermarket.arff

**Question1). Description**

The data consists of three categories which are department, market items and total. Pre-processing done to help the understanding of data was to observe both .arff file and .csv file to identify what is really in the data set I will be rule mining. I did not want to make any changes to the data set as there wasn’t enough information on which I can delete and which I can keep (Just because some category don’t seem to fit in the process of rule mining, it doesn’t determine the value in the dataset).

**Question2). Selection of tool**

Implementation tool selected was the WEKA program.

**Question3). Mining**

For the mining of supermarket data, I have made few different set of results through adjustment of parameters. By setting different value of minimum support and min metric, slightly different result of rules were obtained. The different minimum values used for support and metrics gave similar rules but lined out some rules which was likely to occur due to frequency.

**Question4). Describing results**

As a summary of results, most fitting rule was selection of rules related to bread and cake. The interestingness measures were confidence and lift value observed from the outcome data of associator output using Apriori Algorithm. When the minimum metrics were dropped at a quite low level of 0.1, confidence only dropped to 0.7 which seems to be the lowest which this indicates that the data contains a lot of frequent items which shows repeated consistent rule. However, even though a lot of items related to bread and cake had similar confidence value the lift value was little higher for biscuits and milk-cream compared to vegetables and frozen food.