**3.4 Part 4: Association Finding**

* The datasets have similar no of attributes but vary in the number of observations. The supermarket1-small.arff dataset has binary values f (for false) and t (for true) present for each of the attributes. Whereas, the supermarket2-small.arff has t (for true) values and missing values (?) for each of the present attributes. This means that the missing values (?) could be an f (for false) values where attributes might have binary values just like the first dataset.
* As the dataset was huge in size, the Weka couldn’t process the dataset. Hence, 7 attributes (Bread and Cake, Frozen foods, Milk, Biscuits, Fruit, Party snack foods and vegetables) were selected for finding association.

Apriori algorithm was used with the minimum support: 0.1, minimum metric <confidence>: 0.9.

Three association rules were found. Rule 1 is:

* *biscuits=t frozen foods=t party snack foods=t milk-cream=t fruit=t vegetables=t 512 ==> bread and cake=t 466 <conf:(0.91)> lift:(1.26) lev:(0.02) [97] conv:(3.05)*

As we can see the rule is presented in antecedent ==> consequent format. No rules have coverage less than 0.91 i.e. 95% (466/512). The support is 466/978 or 47% for the first rule. The no besides antecedent is the coverage of the dataset instances whereas the no besides the consequent is the matching instances of both sides. All the rules have a consequent of ‘bread and cake’ and the biscuits, frozen foods, fruit and vegetables appear in all the 3 rules.

* Lift metric type was used with minimum support of 0.35 and minimum metric of 1.1. The cut off was 1.15 for lift and the fruit as the consequent appear most as compared to other parameters. The bread and cake, fruit and vegetables appear most in the antecedent of the rules.
* The supermarket2 dataset had 1281 instances with 7 attributes (same as supermarket1). The minimum support was 0.1 (128 instances) with confidence metric type. The confidence interval was between 91-93% on different rules. The ‘break and cake’ was the consequent in all the 10 rules. Also, biscuits, frozen foods and fruits were the most which appear in the rules.
* Lift metric type was used with minimum support of 0.4 (512 instances) and minimum metric of 1.1. The cut off was 1.11 ranging from 1.11 to 1.15. Bred and cake, fruit and biscuits appeared most in both antecedent and consequent of the rules.
* The FPGrowth algorithm was used (on supermarket2 dataset) with confidence interval between 91-93% for different rules. The ‘bread and cake’ was presented in all the 10 rules as a consequent. Moreover, fruits, biscuits and frozen foods were mostly visible in the antecedent of all the rules. The output of both the Apriori and FPGrowth shows similar readings. There is no significant difference between the findings of both algorithms on supermarket2 dataset.
* Based on frequent item set buying, such as in this case, we found biscuits, frozen foods, fruit and vegetables are set of items people buy on which we can convince people to buy bread and cake along with the other item set so that the transactional amount is maximized.
* Apriori algorithm was used by selecting some nominal attributes from hypothroid dataset. There was no association or any significant findings as the result were the obvious and already known i.e. if a person is no sick then surgery is not required.