Question

1. Objectives: What is the domain and what are the potential benefits to be derived from association rule mining. This is high level - not find patterns, but what would improve because of the use of the patterns.

The dataset we were given has records about sales for a bakery shop.

Benefits that can be derived from association rule mining are:

* Easy to analyze and predict customer behavior
* To clearly understand future demand
* Improve sales
* Reduce maintenance cost
* Decrease intake of stock

2. Data set description: What is in the data, and what preprocessing was done to make it amenable for association rule mining. Where choices were made (e.g., parameter settings for discretization, or decisions to ignore an attribute), describe your reasoning behind the choices.

1. Find missing data (to ensure it will not lead to misleading result)
2. To check if the empty value in the table is na if not change it to na.
3. Set null as NA
4. Change column name to a suitable name (easier for doing data mining)
5. Replace the numeric data to alphabetical data (easier to visualize)
6. Create brand new table (lookup table) consisting of food id and food name.
7. Replace all the food id instances into food names by referring to the lookup table.
8. Receipt number inside item table is changed from integer to numeric.
9. Create a new table by splitting item table into half to make a transaction table.

3. Rule mining process: Parameter settings, choice of algorithm, and the time required.

The parameter we set were support is 0.02 and confidence is 0.8. The algorithm we chose is Apriori algorithm. Furthermore, the average time it took us to run is 0.56 seconds to 0.78 seconds.

4. Resulting rules: Summary (number of rules, general description), and a selection of those you would show to a client.

After removing the redundancies, we had 18 rules left out of the 54 rules before pruning.

|  |  |
| --- | --- |
| Description | Value |
| Minimum support | 0.02 |
| Maximum support | 0.04 |
| Minimum confidence | 0.8 |
| Maximum confidence | 1 |
| Min lift | 10 |
| Max lift | 15 |

We will show the client a selection of these which are rules with high support, confidence and lift value.

5. Recommendations: What should the client do because of the rules discovered?

From the rules we have selected, we observed that the customers’ purchasing pattern is sorted by the flavor they like. For instance, if the customer buys apple croissant they will most like buy apple tart. I would recommend the client to place the products that have same flavor in other areas, so that we can assume that customers will have to walk around the mall to get the things they wanted to buy. It may help them to purchase other items while they are looking for the items they wanted to buy. In the other words, it means it will help in increasing the sales to the client.