

# Information systems lab assignment 1

Erik Bijl (s2581582)  
Emilio Oldenziel (s2509679)  
Group 5

November 2018

## Tasks

We will describe how we solved the tasks that were given in the assignment.

## 1 Histogram

We have plotted a histogram of the amount of groceries in the data. The histogram is shown in Figure 1. Observing this Figure shows us that the amount of yogurt is most occurring in the data. Also the amount of packaged fruit/vegetables, root vegetables and soft cheese is relatively high. Therefore, the support for these groceries will be high.

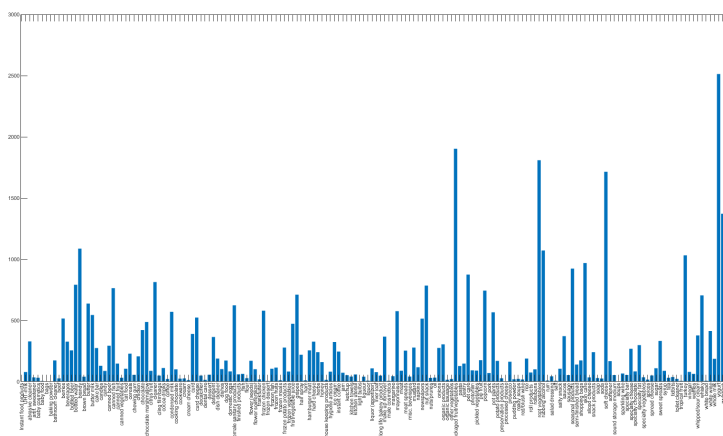


Figure 1: The histogram of the amount of groceries

## 2 Improvement to code

A possible improvement to the existing Matlab code is to add the anti-monotone property. In general, this property does not hold for rules with different item sets. However this anti-monotone property holds for rules with the same item set. The anti-monotone property can be explained as follows:

$$L = A, B, C, D$$

$$confidence(ABC- > D) \geq confidence(AB- > CD) \geq confidence(A- > BCD)$$

From this we can conclude that when an consequence D does not meet a given confidence, each other consequence containing D will also not meet this confidence (given that the item set stays the same). As a result the program does not has to calculate the confidence of the consequences containing D and can calculate the next.

Using the existing Matlab code, this can be implemented by first iterating over the consequences. If we encounter a consequence that did not meet the minimum confidence we store in a ‘skip’ list. Using this list we can skip the calculation of rules where the consequence is a superset of an item in the skip list.

## 3 Timing of improvement

We have timed our code with improvement to the original code given in the assignment. The timing of these can be seen below in Table 1. One can observe that we did not obtain a lower execution time. The reason for this could be that our implementation introduces too much overhead when checking the supersets. This could be consequence of implementing too much statements/loops in order to satisfy the anti-monotonic property for the same item sets.

Algorithm	Execution time (seconds)
Original	9.65
With skipping	12.5

Table 1: The execution of the two programs.

## 4 Rules satisfying parameters

We want to observe the amount of rules that satisfy a particular set of parameters. The rules must follow the following parameters: support = 0.001 and confidence = 0.8. Our program gives 413 rules that satisfy these parameters.

## 5 Top 30 results

The top 30 results of the mining in the previous question with their [support, confidence] sorted on confidence are shown below.

```
1 { 'flour , root vegetables , whipped/sour cream -> whole milk  
  [0.00,1.00] '}  
2 { 'oil , other vegetables , root vegetables , yogurt -> whole  
  milk [0.00,1.00] '}  
3 { 'rice , sugar -> whole milk [0.00,1.00] '}  
4 { 'butter , domestic eggs , other vegetables , whipped/sour  
  cream -> whole milk [0.00,1.00] '}  
5 { 'citrus fruit , root vegetables , tropical fruit , whipped/  
  sour cream -> other vegetables [0.00,1.00] '}  
6 { 'canned fish , hygiene articles -> whole milk [0.00,1.00] '  
  }  
7 { 'brown bread , pip fruit , whipped/sour cream -> other  
  vegetables [0.00,1.00] '}  
8 { 'cream cheese , domestic eggs , napkins -> whole milk  
  [0.00,1.00] '}  
9 { 'cream cheese , domestic eggs , sugar -> whole milk  
  [0.00,1.00] '}  
10 { 'bottled water , other vegetables , pip fruit , root  
  vegetables -> whole milk [0.00,1.00] '}  
11 { 'ham , pip fruit , tropical fruit , whole milk -> other  
  vegetables [0.00,1.00] '}  
12 { 'oil , root vegetables , tropical fruit , yogurt -> whole milk  
  [0.00,1.00] '}  
13 { 'butter , domestic eggs , soft cheese -> whole milk  
  [0.00,1.00] '}  
14 { 'butter , hygiene articles , pip fruit -> whole milk  
  [0.00,1.00] '}  
15 { 'butter , rice , root vegetables -> whole milk [0.00,1.00] '  
  }  
16 { 'citrus fruit , root vegetables , soft cheese -> other  
  vegetables [0.00,1.00] '}  
17 { 'curd , domestic eggs , sugar -> whole milk [0.00,1.00] '}  
18 { 'hygiene articles , pip fruit , root vegetables -> whole  
  milk [0.00,1.00] '}  
19 { 'hygiene articles , root vegetables , whipped/sour cream ->  
  whole milk [0.00,1.00] '}  
20 { 'butter , fruit/vegetable juice , tropical fruit , whipped/  
  sour cream -> other vegetables [0.00,1.00] '}  
21 { 'butter , other vegetables , pork , whipped/sour cream ->  
  whole milk [0.00,1.00] '}  
22 { 'butter , other vegetables , root vegetables , white bread ->
```

```

    whole milk [0.00,1.00] '}'
23 { 'citrus fruit ,pastry ,rolls/buns ,whipped/sour cream ->
    whole milk [0.00,1.00] '}'
24 { 'grapes ,tropical fruit ,whole milk ,yogurt -> other
    vegetables [0.00,1.00] '}'
25 { 'ham ,pip fruit ,tropical fruit ,yogurt -> other vegetables
    [0.00,1.00] '}'
26 { 'newspapers ,rolls/buns ,soda ,whole milk -> other
    vegetables [0.00,1.00] '}'
27 { 'rolls/buns ,root vegetables ,sausage ,tropical fruit ->
    whole milk [0.00,1.00] '}'
28 { 'oil ,other vegetables ,root vegetables ,tropical fruit ,
    yogurt -> whole milk [0.00,1.00] '}'
29 { 'cream cheese ,other vegetables ,sugar -> whole milk
    [0.00,0.94] '}'
30 { 'root vegetables ,sausage ,tropical fruit ,yogurt -> whole
    milk [0.00,0.94] '}'

```