Data Mining – Homework 4

Yassine Berrehouma

1. Find out the pattern of all plant species (i.e., ids) in the forests with support threshold = 50%

* Stored the text file into a 2d array
* Counted the frequency of each plant type in the array
* Eliminated any plant type that has a frequency value less than the minimum support
* Enter a while loop that will keep doing the following steps until the value of frequency of a candidate itemsets is less than the minimum support which returns an empty result. That’s when we stop and return the previous result. The steps are as follow:
  1. Generate length (k+1) candidate itemsets from length k frequent itemsets
  2. Count the frequency of each itermset
  3. Eliminate any itemset with a frequency less than the minimum support
* Print all plant species in the forests with following the minimum support condition

2. Find out the max-frequent-plant species with the same support threshold.

Follow the same steps as the first question with an exception of printing only the max-frequent-plant species that applies the minimum support condition

3. with a confidence threshold of 70% (and support threshold 40%), find out all strong association rules.

Starting to solve this question by applying the same steps as the first question that follow a support threshold of 40% rather than 50%.

Once done with that and got our final output we:

1. Store each pattern applying the minimum support and calculate its association rules, following the metarule of selecting 2 inputs for testing association with 1 output.
2. If confidence of the latter patter is greater or equal to the confidence threshold then we consider it to be a strong association rule