Luc Nikiema

Design: Association Rule Mining

Initially, the data is preprocessed by converting the all the non-zero values to one (to indicate the presence of the products), transforming it into a pandas dataframe with the headers as the product names from the products file. After this if the item is present then its name is added to what will be a list of 1-item frozen sets. Then the dataset is converted to set form as well.

The dataset is then scanned for items that meet the minimum support threshold which are then appended to a list of frequent items. At the same time support values were kept in a dictionary. With that, potential item sets are generated using the value of k and the list of frequent items while the number of items is above zero. The dataset is scanned once again with the new potential item sets and the support data along with list of frequent items are updated. Each item in the list is a list of frozen sets with k items.

The rules are then generated from the frequent items, using the support data and confidence threshold. For sets with at least 2 items, a list of potential right candidates is generated from the set, which leads to new rules being generated. After this the confidence and lift values are calculated and if a rule meets the appropriate requirements (min confidence and lift >= 1) it is appended to a list and returned.