

Essay Big Data

Jasper Robeer
3802337

Utrecht University
February 17, 2017

1 Terminology

The discovery of **frequent itemsets** is one of the major families of techniques for characterizing data. To explain how frequent itemsets work, we will take a look at the *market-basket* model first. The market-basket model and frequent itemsets problem originated in the analysis of true market baskets. Retailers wanted to learn what items are frequently bought together. In the *market-basket* model we look at the many-to-many relationship between the *items* and the *baskets*. The *frequent itemsets* problem follows from this model. We are concerned with finding the sets of items that appear in many of the same baskets. To give a more formal definition: we have a number s , which is known as the *support threshold*. The *support* for a set of items I is the number of baskets for which I is a subset. We say that a set of items I is frequent if its support is s or more.

Association Rules