

Market Basket Analysis

Master in Data Science

Alberto Bertoncini 983833

Massimo Cavagna 9838??

[Github repository](#)

May 31, 2021

Contents

1	Introduction	3
2	Dataset	4
3	Data structure	4
4	Preprocessing	4
5	Algorithms applied	4
6	Scaling of proposed solutions	4
7	Experiments	4
8	Results	4

1 Introduction

The purpose of this paper is to present some of the techniques used in order to perform the so called "market/basket" analysis for a huge amount of data. At first these techniques were exploited for the analysis of purchases in markets, trying to find some relationships among the goods bought by customers. The idea behind these algorithms is to find associations between goods so that can be claimed that if a customer buy item A he is also likely to buy item B and vice versa. This concept could be extended to association between sets of goods (not only single item pairs) and to generic items instead of just goods, so that, in the end, the aim of the algorithms that will be presented is to find frequent sets of items in all the baskets available. In particular, three main algorithms will be implemented:

- A-priori:
 - base
 - PCY
- SON
- Toivonen

Once the frequent sets are found, it is also important to check if all the items in one of these sets are actually associated one another: indeed, considering the environment these techniques come from, we could find that some goods, such as "milk" or "bread" are always frequent, but it cannot be claimed that there is an actual relationship with all the other items in the same frequent set, since it will be bought independently from the other.

- 2 Dataset
- 3 Data structure
- 4 Preprocessing
- 5 Algorithms applied
- 6 Scaling of proposed solutions
- 7 Experiments
- 8 Results

I/We declare that this material, which I/We now submit for assessment, is entirely my/our own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my/our work. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study.