CodeClause

Allocated Project

Task 2 -Market Basket Analysis using Apriori Algorithm

Inference

Introduction

Some of us go to the grocery with a standard list; while some of us have a hard time sticking to our grocery shopping list, no matter how determined we are. No matter which type of person you are, retailers will always be experts at making various temptations to inflate your budget.

Remember the time when you had the "*Ohh, I might need this as well.*" moment? Retailers boost their sales by relying on this one simple intuition.

People that buy this will most likely want to buy that as well.

People who buy bread will have a higher chance of buying butter together, therefore an experienced assortment manager will definitely know that having a discount on bread pushes the sales on butter as well.

Data-driven strategies

Huge retailers pivot on a detailed market basket analysis to uncover associations between items. Using this valuable information, they are able to carry out various strategies to improve their revenue:

- Associated products are placed close to each other, so that buyers of one item would be prompted to buy the other.
- Discounts can be applied to only one of the associated products.

Association Rule Mining

But how exactly is a Market Basket Analysis carried out?

Data scientists are able to carry out Market Basket Analysis by implementing Association Rule Mining. Association Rule Mining is a **rule-based machine learning method that helps to uncover meaningful correlations between different products** according to their co-occurrence in a data set.

However, one of the major pitfalls is that it consists of various formulas and parameters that may make it difficult for people without expertise in data mining. Therefore, before sharing your results with stakeholders, make sure that the underlying definitions are well-understood.

Apriori Algorithm

It is an algorithm that uses frequent itemset to generate association rules. It is based on the concept that a subset of a frequent itemset must also be a frequent itemset.

Support : It is the frequency of item a or combination of item A and B.

Confidence: It tells us how often the items a and b occur given that a is bought.

Lift: It tells us the strength of the rule.

Support = freq(A,B)/N

A and B Products

N is total Transactions

Confidence = freq(A,B) / freq(A)

Lift = Support / support(A) * Support(B)

Gathering information about products

Totally 541909 items were sold in 305 days throughout 24 months.