**FUNDAMENTALS OF DATA SCIENCE**

**FINAL\_PROJECT\_REPORT**

Name: Dasaradh Sravanam

UUID: dsrvanam

U-number: U00830073

****

**Abstract**

In this work, Market Basket Analysis has been done using the Apriori Algorithm. Market Basket Analysis is a very efficient and effective technique used very frequently by the business analysts now a day. This analysis extracts the association information among the products sold by the retailers. The association among the sold product is very much important information that helps the retailer to understand the demand of the buyers.

According the demand of the buyers’ retailers create their product stocks and promote the associated items when an item being sold. This technique helps the retailers to sell more products and consequently the profit of the retailers goes high. On the other hand, the buyers get knowledge about the associated products when they buy some items.

This Market Basket Analysis helps both the buyers and sellers at the same time. The Apriori Algorithm is one of the most efficient algorithms used by the business analysts for the Market Basket Analysis. This algorithm works on relational databases by learning with the association rule and this algorithm is very much useful for item mining. This algorithm works initially for a single item and then it makes a large set of items related to the item chosen initially. In this work, the Apriori algorithm has been applied on the retailer database for the Market Basket Analysis.

**Introduction**

Although, the name of the Market Basket Analysis looks like it is related only with the marketing using basket. However, the Market Basket Analysis has a lot more application. A very pleasant application of the Market Basket Analysis has been found in the web application of the Amazon. In the web application of Amazon, when a buyer select a product for buying the web application automatically put some suggestions for associated items. This is the most common and appropriate example of application of the Market Basket Analysis. This analysis is very much useful for business growth in all types of business sectors.

The Market Basket Analysis is working on the base of the ‘if this, then that’ rule. The main idea behind the rule is that if one item has been selected then the probability of the other items has been chosen becomes high. At first, the business analysts try to understand the purchase behaviors of the users. Then they apply the association rule to find the association rule among the products. The Apriori algorithm is based on statistical approach used for finding the interesting association rule among the purchased products. The Apriori algorithm is very much useful for the Market Basket Analysis.

**Related Work**

The market basket analysis is very much important method for analysing the business sectors. Data analysts find the analysis method very suitable for the growth of the business. Mainly the retailers use this analysis in large scale to grow their business. The Apriori algorithm is also used for the analysis very frequently. Other than the retail business, the market basket analysis is very much important for the other business sectors. Some of the related areas where the market basket analysis applied quite successfully are as follows. Telecommunication sector use the Market Basket Analysis to enforce direct marketing approach in the business. The application of the Market Basket Analysis helps the telecommunication sector to overcome the churn rate.

In the insurance field use the Market Basket Analysis for identifying the frauds in this sector. In the healthcare, the Market Basket Analysis applied very efficiently.

In medical sectors, the condition of comorbid patients has been analysed through the Market Basket Analysis. The related symptoms of the comorbid patients can be identified by the Market Basket Analysis.

There are other factors are also there such as banking sectors. In the banking sectors also use the Market Basket Analysis for the finding of actual credit card users to stop the fraud credit card users.

**Design**

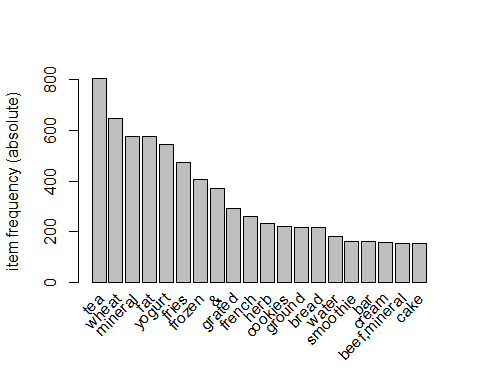
The proposed work is done by using a dataset called “Market\_Basket\_Optimisation.csv”. After successfully loading the dataset and called the required library packages, I created a frequency plot of a certain number if items here.

In this proposed work I have taken the item frequency plot for 20 grocery items. Once I have completed our initial observation over the dataset, to set the rules for Apriori algorithm approach here. I have created the rules by setting up the support to 0.001 and the confidence value to 0.8 respectively. By the selected rules of the Apriori approach have inspected as well as observed the top 5 rules for this particular set of work.

Finally, sorted the created rules by the confidence factor in a decreasing order and plotted the rules within a graph so that we can have a clear idea about the implemented Apriori algorithm approach here.

**Evaluation**

I have successfully implemented Apriori algorithm approach in this scenario by using the “Market\_Basket\_Optimisation.csv” dataset. Once loaded our dataset and the frequency plot for the top 20 items will look like the following.



Here observed that tea, wheat, mineral, yogurt etc. items are most transacted grocery items found here.

As the transactions are concerned, found that there are 5729 items are bought from the retail store where there are 7501 numbers of transactions are made. Again, we can observe that after sorting there are 354 numbers of recorded items are there after creating the rules of the Apriori algorithm.

Again after implementation of the rules for the Apriori algorithm here on that particular database, we got yogurt, fat, dark, wheat, cream etc. rules are the top 5 observed rule in this selected dataset. It has been covered the items like cookies, burgers, fries, escalope, mushroom etc. items respectively.

Now we have sorted the items by their value of confidence and plotted it within an interactive graph. This graph is defined as follows.



**Conclusion**

In this work, the Apriori algorithm has been applied quiet successfully for Market Basket Analysis on retailer database. Apriori algorithm is very much used algorithms for the Market Basket Analysis. Therefore, the Apriori algorithm has been applied in this work. By applying the Apriori algorithm, some interesting and useful association rules among the sold items has been found by analysing the retailer dataset. The association rule is very much applicable to find the associated items for buying. Appropriate values have been chosen for executing the Apriori algorithms. The Apriori algorithm has been used quiet successfully in this work for the Market Basket Analysis.

The Market Basket Analysis is very well accepted technique and it is used by almost all leading retailers for their business. The application of the Market Basket Analysis grows the profit of the retailers in large margin. Therefore, the Market Basket Analysis becomes very popular now a day. In this work, the Market Basket Analysis has been applied in one way that is very much appropriate and has been used efficiently. The Market Basket Analysis using the Apriori algorithm is very successful technique and, in this work, the same has been done successfully.

**Future Work**

The Market Basket Analysis is a very complex technique. As the number of items added by the retailers in the analysis, the complexity of the analysis raises. For small number of items, the analysis is comparatively easy. In this work, the Market Basket Analysis has been done on moderate dataset, where the complexity is also moderate. In future, the same work can be done on more complex datasets containing more items and the items with large hierarchy of set and sub sets.

Another challenge of the Market Basket Analysis is the need of huge amount of purchase history of the buyers in the retailer. The association rule has been learned on the purchase data from the retailers. The learning rate is completely dependent on the amount of purchase data supplied by the retailers. If the amount of data from the retailers is high then learning rate must be high. The Apriori algorithm has found some interesting association rule among the products sold by the retailers. More analysis on the retailer’s data using other algorithms used for market basket analysis can be done in future. After that comparison with the analysis reports found by the different algorithms will be done.

**Appendix**

---

output:

word\_document: default

html\_document: default

---

Load the libraries

```{r}

# Load the libraries

library(arules)

library(arulesViz)

```

Load the data

```{r}

Groceries <- read.transactions("Market\_Basket\_Optimisation.csv")

```

Create a frequency plot with 10 items

```{r}

itemFrequencyPlot(Groceries,topN=20,type="absolute")

```

Now we get the rules

here we set support to .001

and confidence to .8

```{r}

rules <- apriori(Groceries, parameter = list(supp = 0.001, conf = 0.8))

```

Here we see the etop 5 rules

```{r}

options(digits=2)

inspect(rules[1:5])

```

Sorting the rules by confidence

```{r}

rules<-sort(rules, by="confidence", decreasing=TRUE)

```

Now we see the rules in a graph

```{r}

library(arulesViz)

plot(rules,method="graph",interactive=TRUE,shading=NA)```