

Associated Rule Mining Algorithm For Market Basket Analysis: An Iterative Approach for Generating Rules For Data Science Applications

Dr. Sunil Khilari Dr. Chandrani Singh*

Abstract: *In today's competitive e-business era it is essential to observe frequently occurring patterns and associations or correlations from data-sets found in various kinds of large transactional databases and relational databases. Traditionally this can be achieved with the help of market basket analysis to some extent. Researchers in this paper explore an appropriate solution that can be deduced with the help of Association Rule Mining (ARM) algorithm. Further the researcher has identified some use case scenarios of Associated Rule Mining Algorithm for Market Basket Analysis.*

Keywords: Associated Rule Mining (ARM), Market Basket Analysis (MBA)

1. Associated Rule Mining (ARM) Algorithm

Association rules are "if-then" statements that show the probability of relationships between data items, within large data sets in various types of databases⁹. Association rule mining has a huge number of applications and is extensively used to help discover sales correlations in transactional data or in data sets. It is used to find associations between many variables. This algorithm is also used in data mining through which organizations have to convert raw data into useful information. By using software tools to look for patterns in large batches of data, appropriate mining methodologies show which items appear together in a transaction. Businesses can learn more about their customers to develop more effective marketing strategies, increase sales, improve services and decrease costs. ARM now day coming as recommendation engines are used to know what customer buy next. It is essential that an unsupervised non-linear algorithm uncovers how the items are associated with each other.

2. Market Basket Analysis (MBA)

Mainly business are accumulating large quantities of data from their day-to-days operations. For example, customers buying and selling transactions are commonly known as market basket transactions. Further in Market Basket Analysis, a modelling technique that is based upon the theory states that if a certain group of items are bought, then more or less likely another group of items can be bought. This analysis traditionally assists the businesses for developing market strategies but up to limited extent. Such data can be further analyzed for understanding the micro-patterns of transactions made by huge number of customers with the help of Associated Rule Mining Algorithm⁹.

The purpose of market basket analysis is to determine what products customers purchase together; it takes its name from the idea of customers throwing all their purchases into a shopping cart.

3. Types of Association and Rule Evaluation Metrics

There are three common ways to measure association and evaluate the rules

Support - It says how popular an item is, as measured in the proportion of transactions in which an item set appears.

$$\text{Support } (A \Rightarrow B) = P(A \cup B)$$

Confidence - It says how likely item B is purchased when item A is purchased, expressed as $\{A \Rightarrow B\}$. Thus it is measured by the proportion of transaction with item A in which item B also appears. Confidence might misrepresent the importance of association.

$$\text{Confidence } (A \Rightarrow B) = \frac{P(A \cup B)}{P(A)}$$

Lift - It says how likely item B is purchased when item A is purchased while controlling for how popular item B is.

$$\begin{aligned} \text{Lift } (A \Rightarrow B) &= \frac{\text{Confidence } (A \Rightarrow B)}{P(B)} \\ &= \frac{P(A \cup B)}{P(A) P(B)} \end{aligned}$$

4. Use Cases for Association Rule

In data science, association rules are used to find correlations and co-occurrences between data sets. They are ideally used to explain patterns in data from seemingly independent information repositories, such as relational databases and transactional databases². The act of using association rules is sometimes referred to as "association rule mining" or "mining associations."

- a) Amazon knows what customer wants to buy when they order something on their site- using Artificial Intelligence, Machine Learning
- b) what song customer wants to listen to next – using Neural Network, Deep Learning
- c) Grocery stores software systems
- d) e-commerce websites
- e) Online Social Media Apps

5. Working of Association Rule Mining (ARM) Algorithm

Association rule mining involves the use of machine learning models to analyze data for patterns, or co-occurrences, in a database. It identifies frequent if-then associations, which themselves are the association rules. An association rule has two parts: an antecedent (if) and a consequent (then). An antecedent is an item found within the data. A consequent is an item found in combination with the antecedent. Association rules are formed by searching data for frequent if-then patterns and using the criteria support and confidence to identify the most important relationships⁹. Support is an indication of how frequently the items appear in the data. Confidence indicates the number of times the if-then statements are found true. A third metric, called lift, can be used to compare confidence

with expected confidence, or how many times an if-then statement is expected to be found true. Association rules are calculated from item sets, which are made up of two or more items¹¹. If rules are built from analyzing all the possible item sets, there could be so many rules that the rules hold little meaning. With that, association rules are classically created from rules well-represented in data.

6. Use case for R-code and R-Package

Market Basket dataset consists of 1000 observations with item's Purchase Date, Items purchase qty, Rates The date variable ranges from 31/07/2020 to 31/12/2020. Item is a categorical variable or Item is a categorical variable that links with a purchase and rate. Quantity purchased is a quantitative variable that helps in differentiation of quantity and rate. R packages are collections of functions and data sets developed by the community. To perform Association Rule Mining in R, we use the `arules` and the `arulesViz` packages in R. If one does not have these packages installed in system, following commands are needed to install them.

```
install.packages("arules")
```

```
install.packages("arulesViz")
```

```
#setting working data source
```

```
setwd("C://Users/sunil/Desktop/Data Science-MCA-II Sem-III/data")
```

```
# installing Packages and Loading packages
```

```
install.packages("arules")
```

```
install.packages("arulesViz")
```

```
#Data import in Data Frame
```

```
df=read.csv("C://Users/sunil/Desktop/Data Science-MCA-II Sem-III/data/purchase1.csv")
```

```
print(df)
```

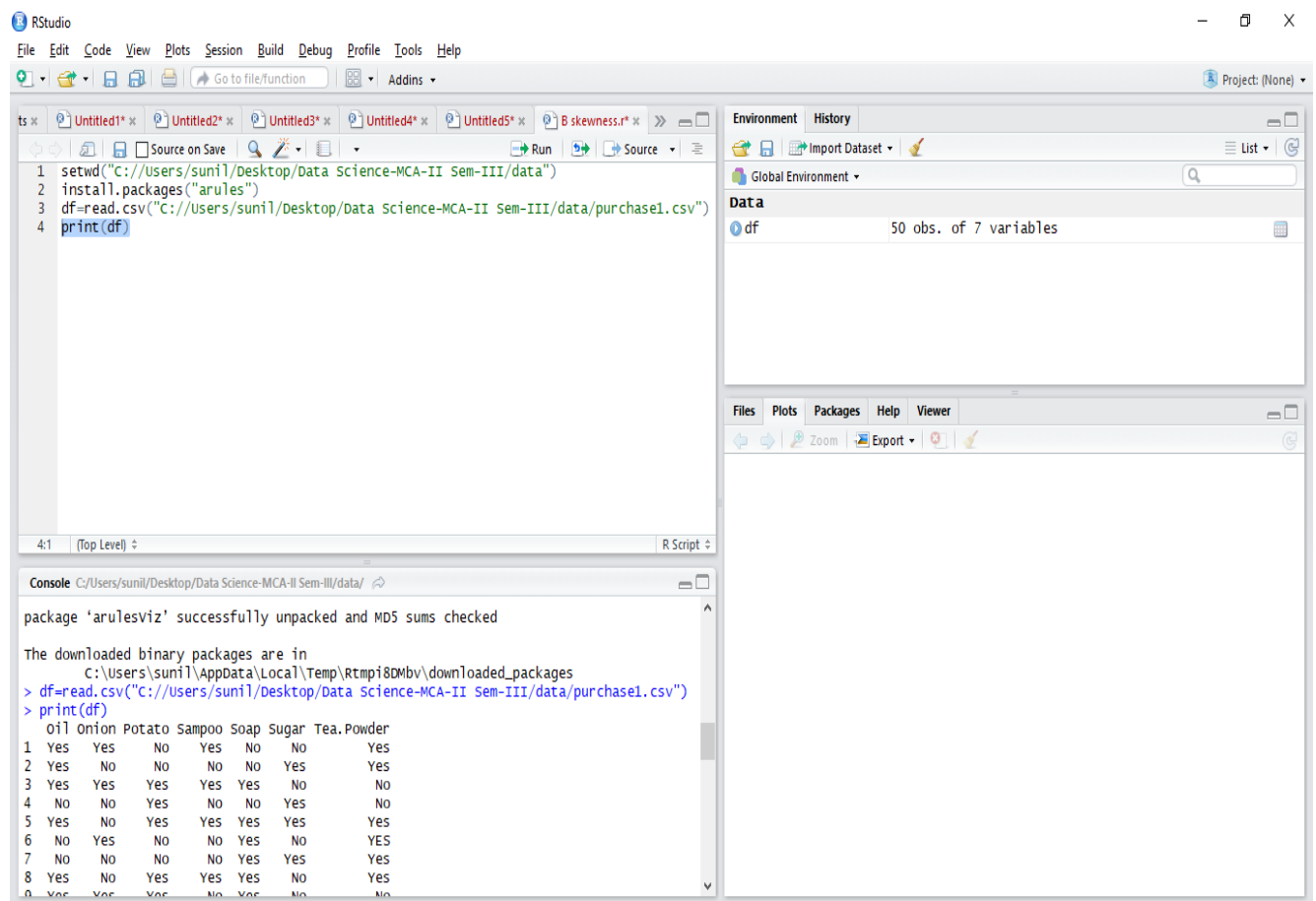


Fig.1 Importing bulk data in R-Data Frame

View(df)

library(arules)

rules<-apriori(df)

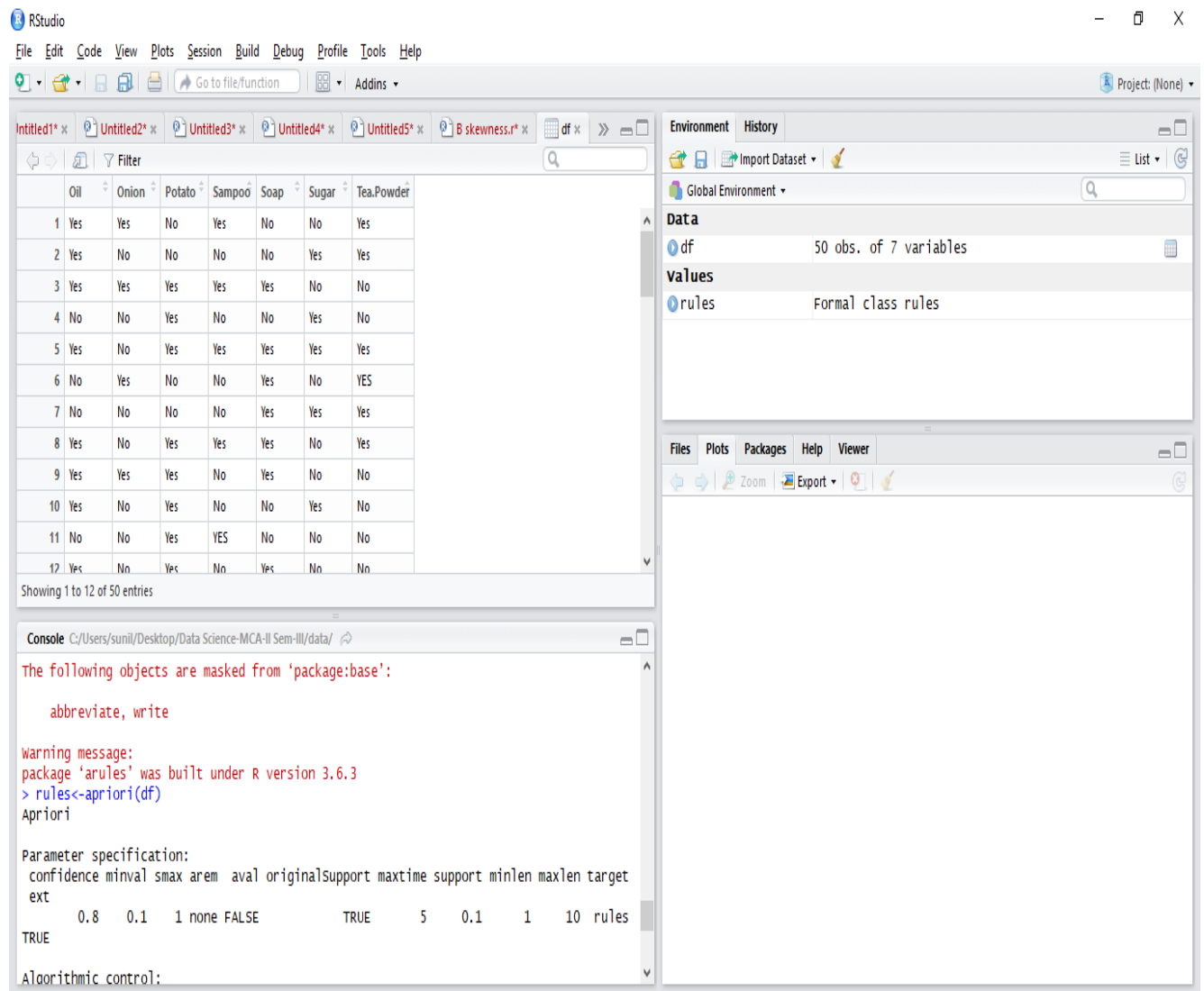


Fig.2 Association Rule generation

```
inspect(rules)
```

```
library(arulesViz)
```

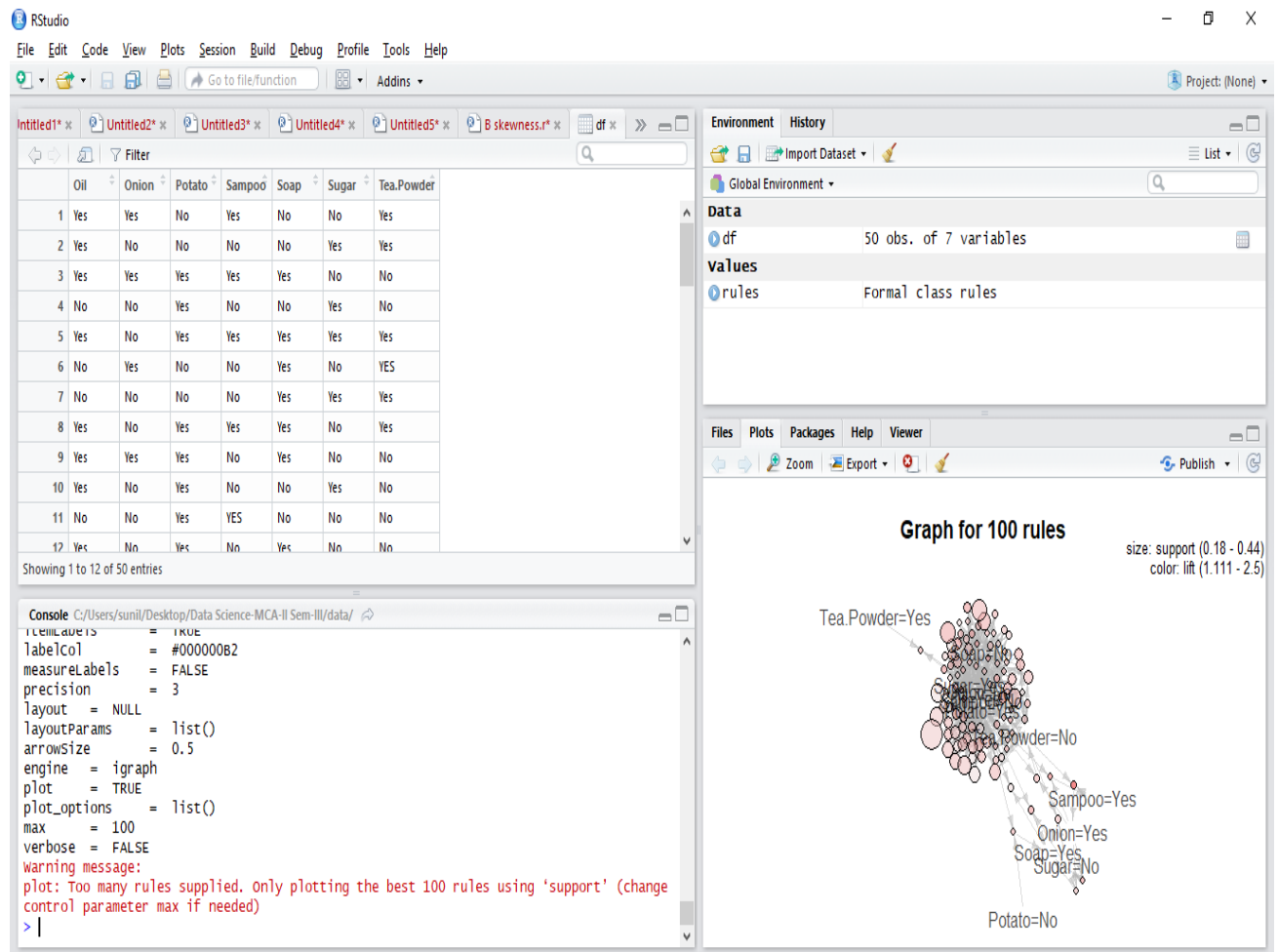
```
#setting seed
```

```
set.seed = 220
```

```
associa_rules = apriori(data = df ,parameter = list(support = 0.004,confidence = 0.2))
```

```
plot(rules,method="grouped")
```

```
plot(rules,method="graph",control=list(type="items"))
```



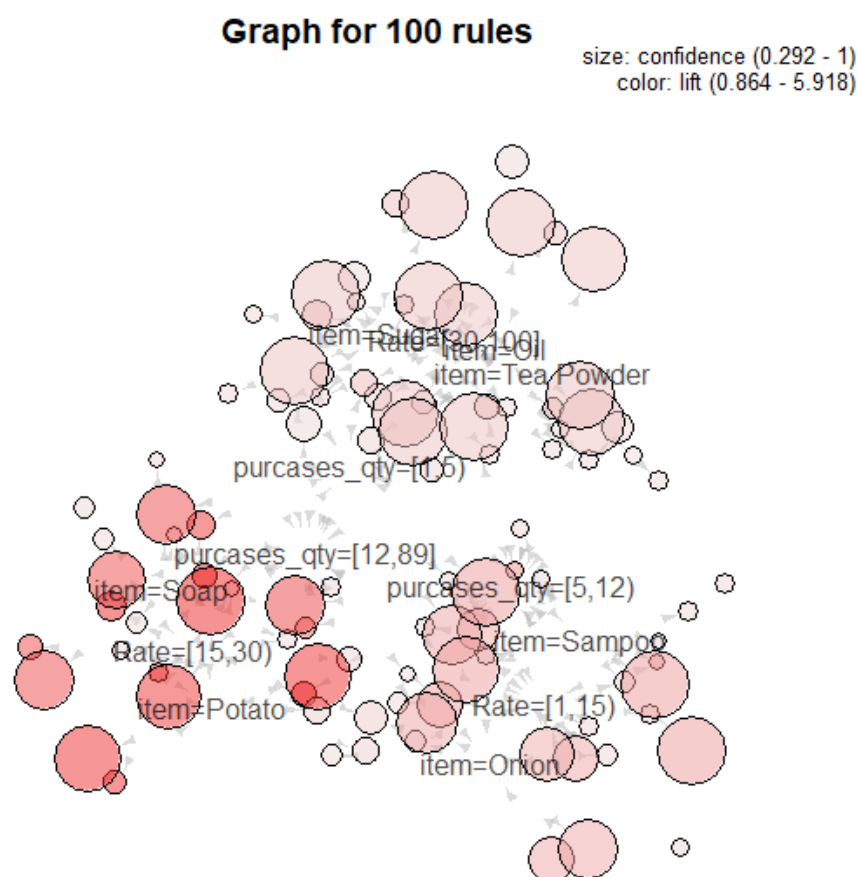


Fig.3. Visualization group of items and quantity purchased

7. Result and Inferences

In above business use case 1000 numbers of transactions (observations) made for 7 different items (Oil, Onion, Potato, Shampoo, Soap, Sugar, Tea Powder) and result found-

Support - Popular item identified –**Sugar**- 239 times i.e.23.9% (measured in the proportion of transactions in which an item set appears)

Confidence – likely purchased item along with Sugar, is identified as-**Tea Powder**-220 times i.e.22% (It says how likely item Tea Powder is purchased when item Sugar is purchased)

Lift – By controlling popularity of Sugar. The computed lift value is 5.69 which is greater than 1 implies **Oil** item is likely to be bought if **Tea Powder** is bought.

8. Conclusion

Researcher is trying to explore the strength of Association Rule Mining (ARM) Algorithm and its business uses case for extracting hidden sense from data for making business decision. This algorithm can achieve to perform market basket analysis using data mining technique. Also can be used to generate more number of healthy association rules for analyzing data for patterns, or co-

occurrences, in a large transactional and relational databases. ARM further identifies frequent if-then associations, which themselves are the association rules. This ARM is extension for linear regression techniques to identify the most important relationships between various dependent and independent variables used in business transaction processing systems.

ARM now a days have auto recommendations engines for business organizations to develop more effective marketing strategies, increase sales, decrease costs, grow and branch out the business interest.

9. References

9.1 Journal Article

- [1] Bernardini F. , Mittleman J., Rushmeier H., Silva C., 1999. *The ball-pivoting algorithm for surface reconstruction. IEEE Transactions on Visualization and Computer Graphics* 5, 4, 349–359.
- [2] Fei-Fei, L., Karpathy, A., Leung, T., Shetty, S., Sukthankar, R., & Toderici, G. (2014). *Large-Scale Video Classification with Convolutional Neural Networks*, by *IEEE Conference on Computer Vision and Pattern Recognition* (cited 865 times, HIC: 24 , CV: 239).
- [3] *Generative adversarial nets*, by Bengio, Y., Courville, A.C., Goodfellow, I.J., Mirza, M., Ozair, S., Pouget-Abadie, J., Warde-Farley, D., & Xu, B. (2014) *NIPS*. (cited 463 times, HIC: 55 , CV: 0).

9.2 Books

- [4] *“The Data Science Handbook: Advice and Insights from 25 Amazing Data Scientists”* by Carl Shan, William Chen, Henry Wang, and Max Song.
- [5] *“The Art of Data Science”* by Roger D. Peng and Elizabeth Matsui.
- [6] *“Learning deep features for scene recognition using places database”*, by Lapedr.
- [7] *“Doing Data Science: Straight Talk from the Frontline”* by Cathy O'Neil and Rachel Schutt.

9.3 Websites

- [8] <https://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining>.
- [9] <https://www.geeksforgeeks.org/regression-analysis-in-r-programming/>.
- [10] <https://towardsdatascience.com/association-rule-mining-in-r-ddf2d044ae50>.