



# Association Rule Mining for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm

Jesús Silva<sup>1</sup>(✉), Mercedes Gaitan Angulo<sup>2</sup>, Danelys Cabrera<sup>3</sup>,  
Sadhana J. Kamatkar<sup>4</sup>, Hugo Martínez Caraballo<sup>5</sup>,  
Jairo Martínez Ventura<sup>6</sup>, John Anderson Virviescas Peña<sup>7</sup>,  
and Juan de la Hoz – Hernandez<sup>6</sup>

<sup>1</sup> Universidad Peruana de Ciencias Aplicadas, Lima, Peru  
jesussilvaUPC@gmail.com

<sup>2</sup> Corporación Universitaria Empresarial de Salamanca (CUES),  
Barranquilla, Colombia  
M\_gaitan689@cuesw.edu.co

<sup>3</sup> Universidad de la Costa, St. 58 #66, Barranquilla, Atlántico, Colombia  
dcabrera4@cuc.edu.co

<sup>4</sup> University of Mumbai, Mumbai, India  
sjkamatkar@mu.ac.in

<sup>5</sup> Universidad Simón Bolívar, Barranquilla, Colombia  
hugo.martinez@unisimonbolivar.edu.co

<sup>6</sup> Corporación Universitaria Latinoamericana, Barranquilla, Colombia  
academico@ul.edu.co, juandelahoz03@gmail.com

<sup>7</sup> Corporación Universitaria Minuto de Dios - UNIMINUTO,  
Bello, Antioquia, Colombia  
john.virviescas@uniminuto.edu

**Abstract.** Customer's segmentation is used as a marketing differentiation tool which allows organizations to understand their customers and build differentiated strategies. This research focuses on a database from the SMEs sector in Colombia, the CRISP-DM methodology was applied for the Data Mining process. The analysis was made based on the PFM model (Presence, Frequency, Monetary Value), and the following grouping algorithms were applied on this model: k-means, k-medoids, and Self-Organizing Maps (SOM). For validating the result of the grouping algorithms and selecting the one that provides the best quality groups, the cascade evaluation technique has been used applying a classification algorithm. Finally, the Apriori algorithm was used to find associations between products for each group of customers, so determining association according to loyalty.

**Keywords:** Data mining · Apriori algorithm · Dates product · Association rules · Hidden patterns extraction · Consumer's loyalty

## 1 Introduction

Marketing focuses on the establishment, development and maintenance of continuous relationships between client and seller as a source of mutual benefits for the parties [1]. In this sense, for marketing policies to be effective in a context of highly competitive marketing, the literature proposes to consider relational benefits and customer segmentation [2]. Through the definition of consumer segments that value the benefits of the relationship to varying degrees, a company can design marketing strategies according to the characteristics of each type of customer [3]. Based on what has been described, the purpose of this research is to perform the customer's segmentation according to their level of loyalty on a sample of companies belonging to the SMEs sector in Colombia through the application of Data Mining techniques.

## 2 Theoretical Review

### 2.1 RFM Analysis

The RFM (Recency, Frequency, Monetary) analysis is a marketing technique used for the analysis of the customer's behavior [4], which is achieved by examining what the customer has purchased, using three factors: (R) purchase Recency, (F) Frequency of purchase, and (M) amount of purchase in Monetary terms. According to theories and researchers, customers who spend more money or buy more frequently in their company, are those customers who end up being more sensitive to the information and messages that the company is transmitting.

### 2.2 Data Mining Methodologies

#### 2.2.1 SEMMA

The SAS Institute, developer of this methodology, defines SEMMA as the process of selection, exploration and modeling of large amounts of data to discover unknown business patterns. The name of this terminology is the acronym corresponding to five basic phases of the process: Sample (Sampling), Explore (Exploration), Modify (Modification), Model (Modeling), Assess (Evaluation) [5].

#### 2.2.2 CRISP-DM

The CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is a method that has proven to guide the Data Mining works. It was created by the group of companies SPSS, NCR and Daimler Chrysler in the year 2000, and is currently the most used reference guide in the development of Data Mining projects [6]. This method structures the process in 6 phases: Understanding the business, Understanding the data, Preparing the data, Modeling, Evaluation, and Implementation. The succession of phases is not necessarily rigid, and each phase is broken down into several general tasks in a second level.

### 3 Materials and Methods

The data was provided by the Chamber of Commerce of Barranquilla, Colombia, and corresponds to customer records and sales taken from 2015 to 2018 for a group of companies belonging to the SMEs sector in Colombia, Caribbean region [7].

The data collected were categorized as follows:

- Clients: covers the personal data of clients, provides geographical and demographic descriptors such as ID, RUC, address, age, gender, marital status, telephone, e-mail, workplace, profession, etc.
- Sales: this category has the daily billing records for sale, which provide the description of each purchase made by customers during the period 2015–2018.

The databases that contain the data of interest for the analysis are the following [8]:

- Clients: contains personal information of the company's clients. It has a total of 44,800 customer records.
- Client Type: contains three records representing end customers, distributors, and franchisees.
- Institution: defines if the client belongs to a public institution/company, to a private company, or a natural person.
- Invoice: all the billing information recorded by the company during the study time period. It has a total of 136,278 invoice records.
- Invoice Details: the products that have been purchased in each invoice. It has a total of 403,159 invoice detail records.
- Products: contains the records of all the products that the company sells. It has a total of 11,127 product records.
- Product Groups: the groups or categories to which the products belong. It has a total of 58 product categories.
- Brands: the brands of the products marketed by the company. It has a total of 396 trademark registrations.

The method used to carry out the process of Data Mining was the CRISP-DM ([9] and [10]) which consists of five phases: Sample (Sampling), Explore (Exploration), Modify (Modification), Model (Modeling), Assess (Valuation), each of them covering a set of activities that must be followed to carry out a mining process with high quality results.

For the segmentation of Master PC customers based on their purchasing behavior, these normalized variables were taken into account: Receipt, Frequency, and Amount. Considering that there is a wide range of clustering algorithms, an analysis was performed on some of them, corresponding to the most used in this type of case. This step allowed selecting the segmentation algorithms that were applied in the present research, which are the following: self-organized maps (SOM) of Kohonen, K-means, and CLARA algorithm (Cluster for Large Applications) which is an extension of the k-medoids algorithm [11].

For the selection of the number of groups, the techniques of internal evaluation applied sum of error squared and the silhouette index. After applying the segmentation

algorithms, the results determine which of them provide the best results based on the described cascade evaluation method in [12].

### 4 Results

Based on the results, it was determined that the most appropriate method for customer segmentation in the study sample on the RFM attributes is the CLARA algorithm which belongs to the group of k-medoids methods. As a result, the following loyalty levels were discovered: Group 1 High, Group 2 Low, Group 3 Medium, and Group 4 Very Low, see Table 1.

**Table 1.** Loyalty groups profile

Group	RFM punctuation			Characteristic
	R	F	M	
High	4	3	4	Customers belonging to this group have a high level of loyalty, with a high level of recency, meaning that their last purchase was made a short time ago, in an average of 1 year ago. It also presents a level of frequency between medium and high, that is, they have bought several times, in an average of once a year, and a high amount that indicates that they have invested a lot of money in their purchases with an average of 982 dollars
Medium	3	1	5	Customers belonging to this group have a Medium level of loyalty, made their last purchase some time ago, in an average of two years and two months. The number of purchases made on average is one time, but have a very high average purchase amount of 830 dollars. This indicates that they have invested a lot of money in their purchases, considering that they have a low purchase frequency
Low	4	1	2	Customers belonging to this group have a Low loyalty level, have a high Recency level, that is, they have made their last purchase a short time ago, in an average of 1 year ago, but the average number of times they have purchased is 1, and the average amount spent is 22.3 dollars, which indicates that they have invested little money in their purchases. Customers of this group could also be considered new customers
Very low	1	1	2	Customers belonging to this group have a Very Low loyalty level, have a very low Recency, which indicates that they have made their last purchase a long time ago, with an average of 3.6 years ago. They also have an average frequency of a single purchase, and a Low amount indicating that they have invested little money in their purchases, with an average of 38.6 dollars. Customers of this group could be considered almost lost customers

### 4.1 Generating Association Rules (Product-Product) by the Apriori Algorithm

An association rule is a rule of the form  $X \Rightarrow Y$ , where  $X$  and  $Y$  are sets of elements. The meaning of this rule is that the presence of  $X$  in a transaction implies the presence of  $Y$  in the same transaction.  $X$  and  $Y$  are respectively called the antecedent and the consequent of the rule [1, 4, 7].

To generate the association rules, the Apriori algorithm, the most commonly used algorithm for the generation of these rules was applied. The Apriori algorithm is a method to discover sets of frequent elements and generates association rules on a set of transaction data [8]. It first identifies the frequent individual elements through the transactions and then extends to the increasingly large element sets until the resulting element sets reach a specified frequency threshold (support) [13]. This algorithm is implemented within the rules package [9] of R.

In order to elaborate the transaction data set, the following data were used: Invoice, Detail\_Invoice, Product, Product Group, and Clients with their respective Loyalty groups. In this way, the data set was made up of an identifier of the transaction and the name of the category of product purchased in the transaction. Finally, a set of data was prepared for each customer loyalty group.

In an initial generation of rules, association rules were generated from 57 product categories, for which an amount that exceeded 20,000 rules was obtained. According to the inspection of the rules, it was possible to see that almost all of them were made up of the following categories: CASES and CHASSIS, HARD DRIVES, MEMORIES, PROCESSORS, MOTHERBOARDS, MONITORS, INTE-EXTE MEMORY READERS, DVD WRITERS AND DVD PLAYERS, SUPPRESSOR REGULATORS.

#### 4.1.1 Generation of Association Rules for Product Recommendation to High Loyalty Customers

For the generation of rules for High loyalty customers, an acceptable level of support was selected according to the distribution of products in the set of transactions and a fairly high confidence level. The values of these parameters are described below in the Table 2.

**Table 2.** Parameters to generate high loyalty customer association rules

Parameter	Value	Percentage
Support	0.01	1%
Confidence	0.8	80%
Maximum length	3 (items per rule)	
Minimum length	2 (items per rule)	

After applying the Apriori algorithm, a set of 84 rules was obtained but, to guarantee its quality, only those rules that present a lift value greater than 3 were selected, leaving a total of 70 rules. The association rules obtained served as the basis for making recommendations to any High loyalty customer. Figure 1 shows the first 10 with the highest level of confidence, and Table 3 presents the interpretation of some of them.

```
> inspect(head(sort(subrules, by="confidence"),10));
  lhs                                rhs      support confidence   lift
1 {FLASH MEMORY,                     => {MOUSES Y MINIMOUSE}  0.02951638  0.9317618  9.172869
  PORTATAILES}
2 {MOCHILAS Y ESTUCHES,             => {PORTATAILES}        0.03106884  0.9294533  20.890940
  SOFTWARE}
3 {FLASH MEMORY,                     => {MOCHILAS Y ESTUCHES} 0.02930021  0.9249380  13.191513
  PORTATAILES}
4 {IMPRESORAS,                       => {MOUSES Y MINIMOUSE}  0.01243933  0.9227405  9.084058
  PORTATAILES}
5 {IMPRESORAS,                       => {MOUSES Y MINIMOUSE}  0.01240002  0.9225146  9.081834
  MOCHILAS Y ESTUCHES}
6 {MOCHILAS Y ESTUCHES,             => {MOUSES Y MINIMOUSE}  0.03242478  0.9171762  9.029280
  PORTATAILES}
7 {FLASH MEMORY,                     => {SOFTWARE}           0.02904475  0.9168734  10.649838
  PORTATAILES}
8 {MOCHILAS Y ESTUCHES,             => {MOUSES Y MINIMOUSE}  0.03063651  0.9165197  9.022816
  SOFTWARE}
9 {IMPRESORAS,                       => {PORTATAILES}        0.01228212  0.9137427  20.537820
  MOCHILAS Y ESTUCHES}
10 {IMPRESORAS,                      => {MOCHILAS Y ESTUCHES} 0.01228212  0.9110787  12.993852
  PORTATAILES}
```

**Fig. 1.** Main association rules for recommendation of product-customers with High loyalty Original screen in Spanish.

**Table 3.** Interpretation of the main association rules for recommendation of product-customers with high loyalty

Rule	Interpretation of rules for High loyalty customers
1.	If a high loyalty customer buys product from the [FLASH MEMORY] and [PORTABLE] categories, the likelihood that he/she will also buy [MOUSES AND MINIMOUSES] on the same visit is 93%
2.	If a High loyalty customer buys product from the [BACKPACKS AND CASES] and [SOFTWARE] categories, the likelihood that he/she will also buy [PORTABLE] on the same visit is 92%
3.	If a High loyalty customer buys product from the [FLASH MEMORY] and [PORTABLE] categories, the likelihood that he/she will also buy [BACKPACKS AND CASES] on the same visit is 92%
4.	If a High loyalty customer buys product from the [PRINTERS] and [PORTABLE] categories, the likelihood that he/she will also buy [MOUSES AND MINIMOUSES] on the same visit is 92%
5.	If a High loyalty customer buys product from the [PRINTERS] and [BACKPACKS AND CASES] categories, the likelihood that he/she will also buy [MOUSES AND MINIMOUSES] on the same visit is 92%
6.	If a client of High loyalty buys product from the categories [BACKPACKS AND CASES] and [PORTABLE], the probability that he/she will also buy [MOUSES AND MINIMOUSES] in the same visit is 91%
7.	If a High loyalty customer buys product from the [FLASH MEMORY] and [PORTABLE] categories, the likelihood that he/she will also buy [SOFTWARE] on the same visit is 91%
8.	If a High loyalty customer buys product from the [BACKPACKS AND CASES] and [SOFTWARE] categories, the likelihood that he/she will also buy [MOUSES AND MINIMOUSES] on the same visit is 91%
9.	If a High loyalty customer buys product from the [PRINTERS] and [BACKPACKS AND CASES] categories, the likelihood that he/she will also buy [PORTABLE] on the same visit is 91%
10.	If a High loyalty customer purchases product from the [PRINTERS] and [PORTABLE] categories, the likelihood that he will also buy [BACKPACKS AND CASES] on the same visit is 91%

### 4.1.2 Generation of Association Rules for Product Recommendation to Average Loyalty Customers

For the generation of rules for average loyalty customers, an acceptable level of support was selected according to the distribution of the product categories in the transaction set and a fairly high confidence level. The values of these parameters are described in Table 4. The initially generated size of the rules was 952. From this total of rules, those with higher quality were selected, i.e. those with the highest lift values.

The base of lift was set to 3, but the number of rules was still being high, which would hinder its interpretation within the marketing area of the company, so finally, those rules that have a lift value greater than or equal to 5 were selected, leaving a total of 125 rules. The association rules obtained will serve as a basis for making recommendations to any customer of Average loyalty. Figure 2 shows the first 10 rules according to the level of trust for the customers of Average loyalty and, in Table 5, the interpretation of some of them is presented.

**Table 4.** Parameters to generate association rules for average loyalty customers

Parameter	Value	Percentage
Support	0.01	1%
Confidence	0.8	80%
Maximum length	2 (items per rule)	-
Minimum length	2 (items per rule)	-

```

> inspect(head(sort(subrulesMedio, by="confidence"),10));
  lhs                                     rhs          support confidence   lift
1  {GAMERS-VOLANTES-VIDEOJUEGOS,      => {PARLANTES}      0.04988789  0.9569892  7.855531
   TECLADOS}
2  {CAMARAS, WEB , FILMADORAS,        => {MESAS Y SILLAS}  0.06240658  0.9515670  9.029763
   GAMERS-VOLANTES-VIDEOJUEGOS}
3  {MESAS Y SILLAS,                   => {TECLADOS}       0.05493274  0.9514563  8.044541
   MOUSES Y MINIMOUSE}
4  {GAMERS-VOLANTES-VIDEOJUEGOS,      => {CAMARAS, WEB , FILMADORAS} 0.06240658  0.9488636  7.152561
   MESAS Y SILLAS}
5  {GAMERS-VOLANTES-VIDEOJUEGOS,      => {MESAS Y SILLAS}  0.05997758  0.9441176  8.959074
   PARLANTES}
6  {GAMERS-VOLANTES-VIDEOJUEGOS,      => {CAMARAS, WEB , FILMADORAS} 0.04914051  0.9426523  7.105740
   TECLADOS}
7  {MESAS Y SILLAS,                   => {PARLANTES}      0.05437220  0.9417476  7.730419
   MOUSES Y MINIMOUSE}
8  {CABLES Y ADAPTADORES,             => {PARLANTES}      0.05287743  0.9401993  7.717710
   GAMERS-VOLANTES-VIDEOJUEGOS}
9  {AUDIFONOS, MICROFONOS,           => {CAMARAS, WEB , FILMADORAS} 0.05156951  0.9355932  7.052528
   GAMERS-VOLANTES-VIDEOJUEGOS}
10 {GAMERS-VOLANTES-VIDEOJUEGOS,      => {CAMARAS, WEB , FILMADORAS} 0.05941704  0.9352941  7.050273
    PARLANTES}

```

**Fig. 2.** Main rules of association for recommendation of product-customers with Average loyalty. Original screen in Spanish

**4.1.3 Generation of Association Rules for Product Recommendation to Low Loyalty Customers**

For the generation of rules for Low loyalty customers, the values of 0.01 for support and 0.8 for trust were selected as initial parameters, but under these conditions no association rule was found. The reason is that although there are several transactions for this group of customers, they rarely buy several products together, and those that are bought together do not satisfy high confidence levels.

Therefore, other parameters were established maintaining the support level, but decreasing the confidence level by half, which is also an acceptable value, although not as good as in the previous experiments.

The association rules obtained will serve as the basis for making recommendations to any Low loyalty customer. Table 6 shows the final values of the parameters for generating rules.

Despite having decreased the value of the acceptance parameters, only 3 rules were generated, one of which was discarded because the lift value was too low, leaving only 2 association rules. The association rules obtained will serve as a basis for making recommendations to any client with very low loyalty. The generated rules are described in Fig. 3 and the interpretation in Table 7.

**Table 5.** Interpretation of the main rules of association for recommendation of product - customers with Average loyalty.

Rule	Interpretation of rules for Average loyalty customers
1	If an Average loyalty customer buys product from the categories [GAMERS-FLYERS-VIDEO GAMES] and [KEYBOARDS], the probability that he/she will also buy [SPEAKERS] on the same visit is 95%
2	If an Average loyalty customer purchases product from the [WEBCAMS, VIDEOCAMERA] and [GAMERS-VOANTES-VIDEO GAMES] categories, the probability that he/she will also buy [TABLES AND CHAIRS] in the same visit is 95%
3	If an Average loyalty customer buys product from the [TABLES AND CHAIRS] categories and [MOUSES AND MINIMOUSE], the probability that he/she will also buy [KEYBOARDS] on the same visit is 95%
4	If an Average loyalty customer buys product from the categories [GAMERS-FLYERS-VIDEO GAMES] and [TABLES AND CHAIRS], the probability that he/she will also buy [WEBCAMS, VIDEOCAMERA] on the same visit is 94%
5	If an Average loyalty customer buys product from the categories [GAMERS-FLYERS-VIDEOGAMES] and [SPEAKERS], the probability that he/she will also buy [TABLES AND CHAIRS] on the same visit is 94%
6	If an Average loyalty customer buys product from the categories [GAMERS-FLYERS-VIDEO GAMES] and [KEYBOARDS], the probability that he/she will also buy [WEBCAMS, VIDEOCAMERA] on the same visit is 94%
7	If an Average loyalty customer buys product from the [TABLES AND CHAIRS] categories and [MOUSES AND MINIMOUSE], the probability that he/she will also buy [SPEAKERS] on the same visit is 94%

(continued)



**Table 5.** (continued)

Rule	Interpretation of rules for Average loyalty customers
8	If an Average loyalty customer buys product from the categories [CABLES AND ADAPTERS] and [GAMERS-FLYERS-VIDEO GAMES], the probability that he/she will also buy [SPEAKERS] on the same visit is 94%
9	If an Average loyalty customer buys product from the categories [HEADPHONES, MICROPHONES] and [GAMERS-FLYERS-VIDEO GAMES], the probability that he/she will also buy [WEBCAMS, VIDEOCAMERAS] on the same visit is 93%
10	If an Average loyalty customer buys product from the [HEADPHONES, MICROPHONES] and [SPEAKERS] categories, the probability that he/she will also buy [WEBCAMS VIDEOCAMERAS] on the same visit is 93%

**Table 6.** Parameters to generate Low loyalty customer association rules.

Parameter	Value	Percentage
Support	0.01	1%
Confidence	0.4	40%
Maximum length	2 (items per rule)	-
Minimum length	2 (items per rule)	-

```
> inspect(head(sort(subrulesBajo, by="confidence"),10));
lhs                rhs                support confidence lift
1 {SOFTWARE}        => {SERVICIOS INFORMATICOS} 0.02018943 0.7826087 2.854387
2 {REDES E INTERNET} => {CABLES Y ADAPTADORES} 0.01370887 0.4182510 4.547487
```

**Fig. 3.** Main association rules for product recommendation for Low loyalty customers.

**Table 7.** Interpretation of the main association rules for product recommendation for Low loyalty customers

Rule	Interpretation of rules for Low loyalty customers
1	If a Low loyalty customer buys a product from the [SOFTWARE] category, the probability that he/she will also buy [COMPUTER SERVICES] on the same visit is 78%
2	If a Low loyalty customer buys a product in the [NETWORKS AND INTERNET] category, the probability that he/she will also buy [CABLES AND ADAPTERS] in the same visit is 41%

#### 4.1.4 Generation of Association Rules for Product Recommendation to Very Low Loyalty Customers

For the generation of rules for Very Low loyalty customers, the parameters were initially set at 0.01 for support and 0.8 for trust but, as for the previous group, no rule was found under these conditions. The value of the confidence parameter was reduced to 40%, which is an acceptable value, see Table 8.

**Table 8.** Parameters to generate association rules for Very Low loyalty customers

Parameter	Value	Percentage
Support	0.01	1%
Confidence	0.4	40%
Maximum length	3 (items per rule)	–
Minimum length	2 (items per rule)	–

The final result was a total of 3 rules that meet a lift value greater than 2. In Fig. 4, these rules are presented, and their interpretation in Table 9.

```
> inspect(head(sort(rulesMuyBajo, by="confidence"),300));
  lhs                rhs                support confidence lift
1 {SOFTWARE}        => {SERVICIOS INFORMATICOS} 0.01143649 0.7857143 2.617234
2 {IMPRESORAS}      => {CABLES Y ADAPTADORES} 0.02755155 0.6556701 8.076568
3 {REDES E INTERNET} => {CABLES Y ADAPTADORES} 0.01031017 0.4630350 5.703682
```

**Fig. 4.** Main association rules generated for product recommendation for Very Low loyalty customers.

**Table 9.** Interpretation of the main association rules for product recommendation for Very Low loyalty customers.

Rule	Interpretation of rules for Very Low loyalty customers
1	If a Very Low loyalty customer buys a product from the [SOFTWARE] category, the probability that he/she will also buy [COMPUTER SERVICES] on the same visit is 78%
2	If a Very Low loyalty customer purchases a product from the [PRINTERS] category, the probability that he/she will also buy [CABLES AND ADAPTERS] on the same visit is 65%
3	If a Very Low loyalty customer buys a product from the [NETWORKS AND INTERNET] category, the probability that he/she will also buy [CABLES AND ADAPTERS] on the same visit is 46%

## 5 Conclusions

To assess the accuracy of the used algorithms, k-means, k-medoids, and Self Organizing Maps (SOM), classification rules were generated taking, as a decision attribute, the groups created by the algorithms mentioned in this research. Besides, based on the prediction level, the results suggest that the classification of the groups generated by the CLARA of k-medoids algorithm provide a higher accuracy. The groups of customers of the sample of companies in study, by means of data mining, revealed the levels of loyalty as: High, Medium, Low and Very Low. These results will allow the company to develop retention strategies to their customers. The application of the Apriori association algorithm on the set of transactions of each group of customers allowed to create

important association rules with quite high confidence levels, especially for the customers that belong to the highest loyalty groups because these customers are those who buy more products in the same transaction.

## References

1. Amelec, V.: Increased efficiency in a company of development of technological solutions in the areas commercial and of consultancy. *Adv. Sci. Lett.* **21**(5), 1406–1408 (2015)
2. Varela, I.N., Cabrera, H.R., Lopez, C.G., Viloría, A., Gaitán, A.M., Henry, M.A.: Methodology for the reduction and integration of data in the performance measurement of industries cement plants. In: Tan, Y., Shi, Y., Tang, Q. (eds.) *Data Mining and Big Data, DMBD 2018*. LNCS, vol. 10943, pp. 33–42. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-93803-5\\_4](https://doi.org/10.1007/978-3-319-93803-5_4)
3. Lis-Gutiérrez, J.P., Lis-Gutiérrez, M., Gaitán-Angulo, M., Balaguera, M.I., Viloría, A., Santander-Abril, J.E.: Use of the industrial property system for new creations in Colombia: a departmental analysis (2000–2016). In: Tan, Y., Shi, Y., Tang, Q. (eds.) *Data Mining and Big Data, DMBD 2018*. LNCS, vol. 10943, pp. 786–796. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-93803-5\\_74](https://doi.org/10.1007/978-3-319-93803-5_74)
4. Anuradha, K., Kumar, K.A.: An e-commerce application for presuming missing items. *Int. J. Comput. Trends Technol.* **4**, 2636–2640 (2013)
5. Larose, D.T., Larose, C.D.: *Discovering Knowledge in Data* (2014). <https://doi.org/10.1002/9781118874059>
6. Pickrahn, I., et al.: Contamination incidents in the pre-analytical phase of forensic DNA analysis in Austria—Statistics of 17 years. *Forensic Sci. Int. Genet.* **31**, 12–18 (2017). <https://doi.org/10.1016/j.fsigen.2017.07.012>
7. de Barrios-Hernández, K.C., Contreras-Salinas, J.A., Olivero-Vega, E.: La Gestión por Procesos en las Pymes de Barranquilla: Factor Diferenciador de la Competitividad Organizacional. *Información tecnológica* **30**(2), 103–114 (2019)
8. Prajapati, D.J., Garg, S., Chauhan, N.C.: Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment. *Future Comput. Inform. J.* **2**, 19–30 (2017). <https://doi.org/10.1016/j.fcij.2017.04.003>
9. Abdullah, M., Al-Hagery, H.: Classifiers' accuracy based on breast cancer medical data and data mining techniques. *Int. J. Adv. Biotechnol. Res.* **7**, 976–2612 (2016)
10. Khanali, H.: A survey on improved algorithms for mining association rules. *Int. J. Comput. Appl.* **165**, 8887 (2017)
11. Ban, T., Eto, M., Guo, S., Inoue, D., Nakao, K., Huang, R.: A study on association rule mining of darknet big data. In: *2015 International Joint Conference on Neural Networks*, pp. 1–7 (2015). <https://doi.org/10.1109/IJCNN.2015.7280818>
12. Vo, B., Le, B.: Fast algorithm for mining generalized association rules. *Int. J. Database Theory Appl.* **2**, 1–12 (2009)
13. Al-Hagery, M.A.: Knowledge discovery in the data sets of hepatitis disease for diagnosis and prediction to support and serve community. *Int. J. Comput. Electron. Res.* **4**, 118–125 (2015)