World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:7, 2019

Main Cause of Children's Deaths in Indigenous Wayuu Community from Department of La Guajira: A Research Developed through Data Mining Use

Isaura Esther Solano Núñez, David Suarez

***Abstract*—**The main purpose of this research is to discover what causes death in children of the Wayuu community, and deeply analyze those results in order to take corrective measures to properly control infant mortality. We consider important to determine the reasons that are producing early death in this specific type of population, since they are the most vulnerable to high risk environmental conditions. In this way, the government, through competent authorities, may develop prevention policies and the right measures to avoid an increase of this tragic fact. The methodology used to develop this investigation is data mining, which consists in gaining and examining large amounts of data to produce new and valuable information. Through this technique it has been possible to determine that the child population is dying mostly from malnutrition. In short, this technique has been very useful to develop this study; it has allowed us to transform large amounts of information into a conclusive and important statement, which has made it easier to take appropriate steps to resolve a particular situation.

***Keywords*—**Malnutrition, datamining, analytical, descriptive, population, wayuu, indigenous.

I. Introduction

THE Wayuu community, an existing community in Colombia, is being afflicted by a serious condition of malnutrition and deaths in children. The cause of these problems is a combination of several factors. Some of them are: the absence of the State in that area, the scarcity of drinking water, the guarantees of a health service, the customs they preserve as an indigenous community, the difficult access to that territory, the climatic conditions and extreme poverty, among others [1]. Additionally, based on information from the National Institute of Health on the deaths of minors [17], it is found that the cause of death – weather immediate or associated – is malnutrition; in particular, there is a very high sub record of the information reported to the *Sivigila* (National PublicHealth Surveillance System). This system has been created to provide systematic and timely provision of information on the dynamics of events that affect or may affect the health of the Colombian population). It turns out, then, that the 48.5[1] due to acute diarrheal disease (ADD) at the national level corresponds to the indigenous population. This demonstrates the serious situation of the rights of this population, according to [2].

Through data mining, we intend to obtain a clearer analysis such as: common factors, age range that most afflict this problem, the most affected gender, areas with the highest malnutrition and mortality rate, presence of government entities, etc. Some factors to consider in this analysis are gender and age, mainly because they are repeated in all the reports of the entities on which we have found information about this exploitation of data and study. And they are very important elements because through them we will be able to show clear studies of the weak points and how to offer alternatives, predictions, and decisions against the problem of the deaths in children under age in the Wayuu community.

II. Wayuu Community Features

A. Geographical Location

The Wayuu community is an indigenous people that inhabit the arid peninsula of La Guajira in northern Colombia and northwestern Venezuela, on the Caribbean Sea. It is a region with a warm, dry and inhospitable climate, which was previously bathed by the Rancheria and El Limón rivers (Venezuela). The DANE 2005 Census reports 270.413 people who recognize themselves as belonging to this community [6]. The Wayuu are not evenly distributed in their traditional territory. The population density in and around Nazareth and other areas of Alta Guajira are located in the surroundings of *Uribía,* the *Serranía de Jalaala* and in the savannahs of *Wopumuin*, in the municipalities of *Maicao* and Manaure, according to [3], [15].

B. Socio-Political Organization

Its socio-political organization is marked by clans, which are defined by the maternal line. Members of a clan share the same social status and a common ancestor. Currently there are 22 clans and the largest percentage of population is found in the *Epieyu* clans and the kinship of this ethnic group is linked to succession and inheritance in the clan of the eldest maternal nephew [4].

C. Economy

Its economy is conditioned by the climate where the scarcity of rainfall prevails. Underground wells are ecological determinants that condition their subsistence and reproduction. They are mainly dedicated to the breeding of goat cattle, which links the largest number of people to grazing. They also carry out fishing activities throughout the coastal area, and

Isaura Solano Nuñezis with the San Mateo Educación Superior, Colombia (e-mail: isolano@sanmateo.edu.co).

[1]The figure quoted by the Sivigila is not set in percentage but as an absolute value in their article**)**

they exploit salt in the *Manaure* mines. Women are dedicated to handicrafts and embroidery of clothing and fabric elaboration [5].

D. Language and History

The Wayuu language belongs to the Arawak linguistic family, which has some dialectal differences depending on the room area (high, medium or low Guajira), but in fact they are minimal. Nowadays, due to the multiple cultures they prefer Spanish, especially the youth community who also keeps their mother tongue, which represents an important factor of ethnic and cultural identity for them [6]. From this period, the *"guajiros"* established commercial and cultural ties with the majority society, as they have done throughout their history. In the 20th century they were affected by the oil exploitation of Lake Maracaibo and the opening of the coal mine *El Cerrejón* and its port in Alta Guajira, in the mid-1980s [7].

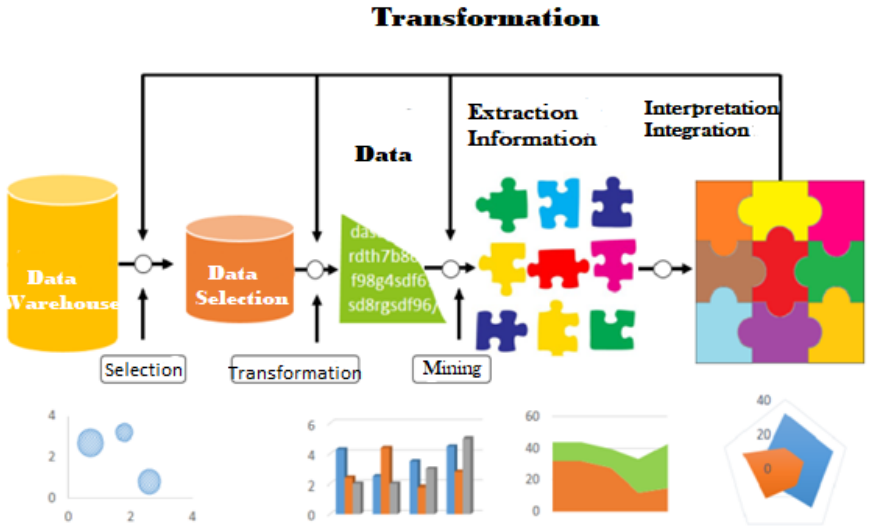

Fig. 1 Data mining (Design developed by the authors)

Graphic Explanation: Process of Data Mining (DM) in broad terms, the DM process can be divided into 6 steps:

1. Data set selection: Here we decide which will be the objective variables (those that we want to predict or infer), the independent variables and the selection of records (data) to be used.
2. We will look for atypical values or data information that we will represent by using histograms and/or scatter diagrams.
3. Transformation or pre-processing of the input data set: In this step, data are normalized to the same scale. It is also decided how to deal with missing, atypical or doubtful data.
4. Selection and application of data mining techniques: A model is built, which will be used on data to predict classes by classification or to discover similar groups by segmentation.
5. Knowledge extraction: Once the previous step has been applied, behavioral patterns are sought in the values of the problem variables or association relationships between these variables.
6. Interpretation and evaluation of data: The model must be validated by checking that the conclusions drawn are valid and satisfactory. If the final model does not exceed this evaluation, the process can be repeated from the beginning or from any of the previous steps.

III. Conceptual Framework

In data mining we will use operators such as read file and filter example. These operators allow us to read files of various formats in order to load them into a Business Intelligent tool, and then to filter the information contained in each one of them, allowing us to make a data mining process that contributes to the decision making [6].

New tools in data mining are based on dealing with data to find and obtain patterns that throw these analyses of huge volumes of Big Data. These methods are artificial intelligence, automatic learning, statistics, database systems, business intelligence, prediction, analytical, etc.

From mining we will explore databases provided for the planning, lifting, processing, analysis and dissemination of the official statistics of Colombia, by entities of the Colombian State that ensure the welfare of minors and by the institution that officially carries statistics of the deaths of Colombian citizens. The objective is to find patterns that can provide us with valuable information in making future decisions [8]. The data mining process usually has four main stages:

- Objectives determination
- Data processing
- Model determination
- Analysis of the results

Choosing the best algorithm for a specific analytical task is a great challenge and it depends on the problem to be solved. There are basically five different problems in data mining: classification, regression, segmentation, association and

sequence analysis [9]. To solve these problems there are many algorithms to use; the main ones are: the association, the clusters, the decision tree, the linear regression, the naive Bayesian classifier, the neural network, the sequence clusters and the serial temporary. Finally, the last step is the analysis of the results.

IV. Sources of Information

The achievement of the information is of great importance to guarantee the veracity of the information to which it can be concluded and, above all, due to the importance and sensitivity that deserves the topic of study. By selecting different primary sources, which have only been treated in their process of lifting, collecting and storage, we have:

- **NADS:** The National Administrative Department of Statistics is the entity responsible for planning, lifting, processing, analyzing and disseminating the official statistics of Colombia. It belongs to the executive branch of the Colombian State, with more than 50 years of experience.
- **National Institute of Legal Medicine and Forensic Sciences** is a public establishment of scientific technical reference that directs and controls the system of Legal Medicine and Forensic Sciences in Colombia. It provides forensic services to the community and the administration of justice based on scientific research and the suitability of human talent in a framework of quality, impartiality, competitiveness and respect for human dignity.

"The Colombian Institute of Family Welfare - CIFW is the entity of the Colombian State that works for the prevention and integral protection of early childhood, childhood, adolescence and the welfare of families in Colombia", quoted from [16].

Special attention must be paid to those in a position of threat or violation of their rights. They were chosen because at the governmental level they are the agencies in charge of processing, analyzing and providing the Colombian state with the results of studies in the different areas.

V. Data Processing

1) Extract

Based on the files of the NADS, CIFW and Legal Medicine, we will work with different factors of study such as age, cause of death, zone, etc.

2) Transform

We will convert these files, unify the formats of columns and rows for the databases, leaving the same number of records for all the sources.

3) Debug

We unify information and leave a large database to manage our analysis.

VI. Objectives of the Mining Technique of Data

Natural groups of people are selected according to their age, sex, area of occurrence of death, cause of death to find similarities in the cause of death and the year in which it occurred.

The aim is to be able to predict, analyze, make decisions about the problem that afflicts this population located in the north of Colombia, apart from wanting to avoid an extermination of one of the many indigenous communities, such as the *Wayuu* [6].

VII. Knowledge to Get

1) Due to the difficult access and climatic conditions of the region in Alta Guajira, it is necessary to know: In which socioeconomic conditions are the families where children die in malnutrition conditions? In order to find patterns that allow a possible contribution to the improvement of the activities performed by the head(s) of the family nucleus.

The operators used were: Excel, to upload data base. We filter the "average household income" field, so a design is generated.

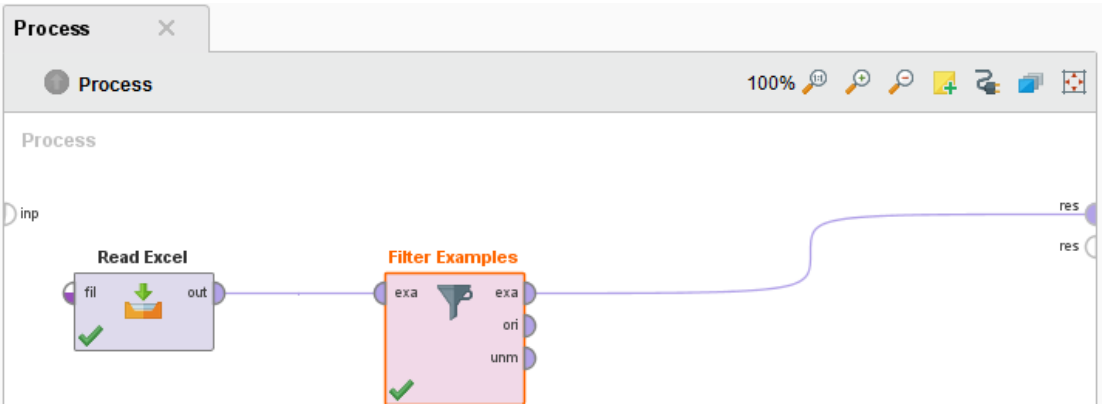

Fig. 2 Load of data

For the objective to be analyzed, a filter has been made for all those communities that earn less than $250,000 per month in order to know their purchasing power. It could be concluded that no one in that community managed to match or exceed the value to be analyzed.

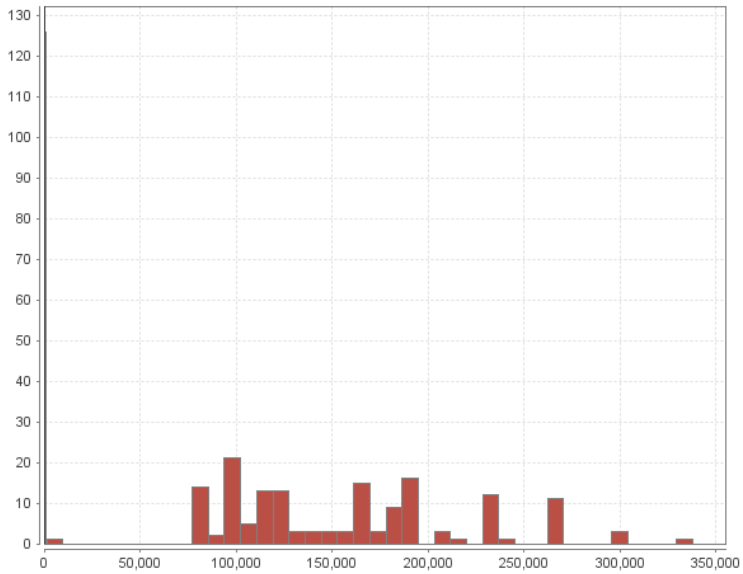

Fig. 3 Population income

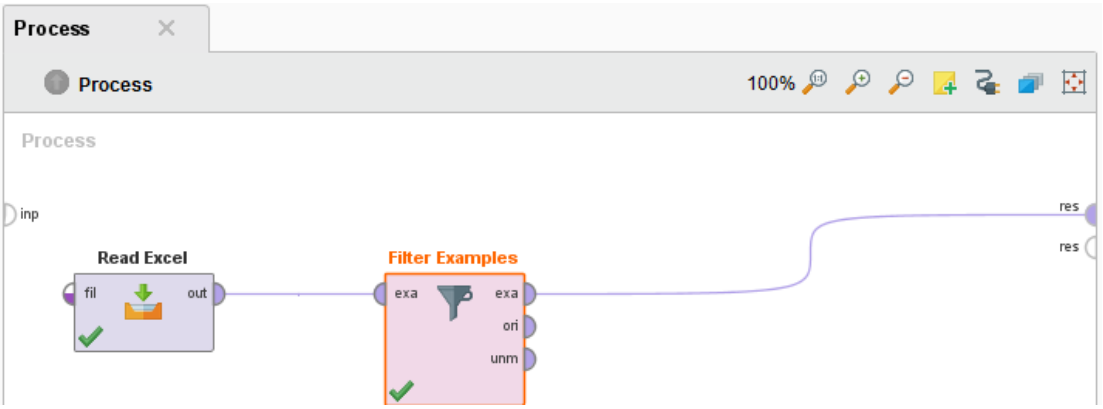

Fig. 4 Filter examples

2) It is essential to know the degree of vulnerability of children in the Wayuu tribe to determine the average age of death in children under malnutrition. Due to the difficult conditions in the area, where most of the findings point to scarce water resource, food and difficult access to the health system, it is urgent to find a way to provide information that can contribute to the reduction of deaths.

We filter the "age" field, so a design is generated.

For the objective to be analyzed, a filter has been made for all those ages by number of children.

3) The current available medical centers in La Guajira could be insufficient for the treatment of the community in general. Where more complex treatments are needed, the service would not be provided. It could possibly be transferred to a more complex health center, due to these reasons, it is necessary to know: what kind of medical treatment and social care (state entities) are given to children who die due to malnutrition?

Operators used: Read Excel: Upload information base, Filter Example: Filtered by the "Age" field; then the design is generated.

For the objective to be analyzed, it is determined that the attention provided by the State is insufficient to try to avoid deaths from malnutrition issues.

To achieve the stated knowledge, the data mining technique "Decision Tree" will be used [10]. Methodology, based on the separation of estimates, allows us to reach the consolidation of data necessary for decision making. Its objective is to use different ramifications analyzing the alternatives, the events, the probabilities and the results, to reach in a complete way all the possible solutions.

VIII. Methodology

Access to data is done directly to the primary sources, using search options and filtering by categories required by the study object.

The different sources of information allow to obtain data in time ranges, thematic, independent classifications that require a standardization treatment to match the format, dates, language, classifications and typologies used.

The loading of the information to the *RapiMinner* software is made by means of CSV files, flat text separated by commas,

using the import wizard and loading the data source [11]. Afterwards, data-operated applications are used, where different purpose and specialty algorithms are applied; additionally, the quantity and quality of the stored data must be taken into account [12].

IX. Design

The techniques used (linear regression, filters, classifiers, neural network) describe the way in which, through the RapidMiner tool and with the data, an output file is obtained with the result of the analysis of the same [13]. The images of *RapidMiner* components must be indicated as they were put, additionally, the results generated by data mining, interpretation and analysis of the resulting data must be explained [14].

X. Conclusions

Engineering in all its specialties, and since its beginnings has always involved experimentation, research, creation, innovation, and also the revolution in the most recognized specialties that are industrial, social, economic and political, taking into account that one of the most important is the knowledge revolution. And systems engineering, specifically data science, has been working for some years to improve the way data are treated and converted into information that is mixed and related to a context.

In the evaluation, an individual or collective analysis can be achieved, which can be in real or imaginary environments and with it to be able to determine future behaviors and trends of change in different aspects of human life.

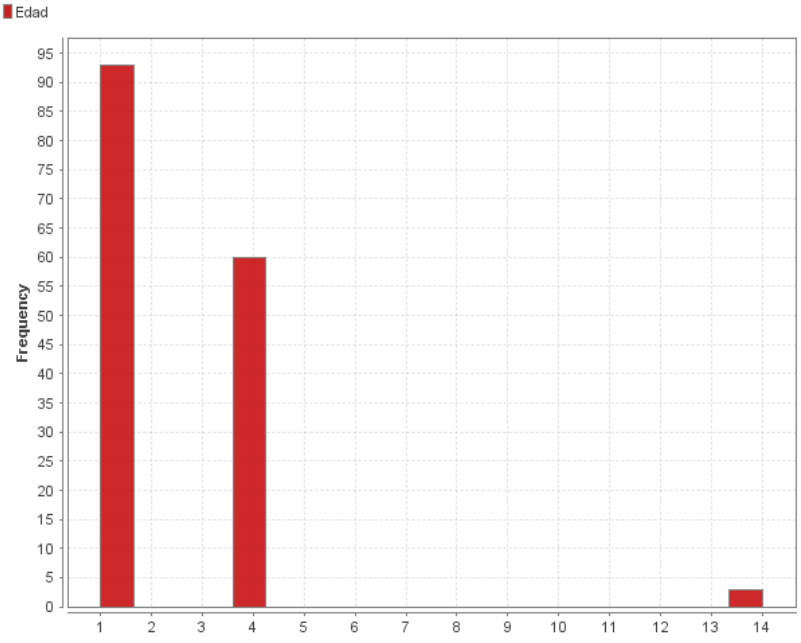

Fig. 5 Age of children

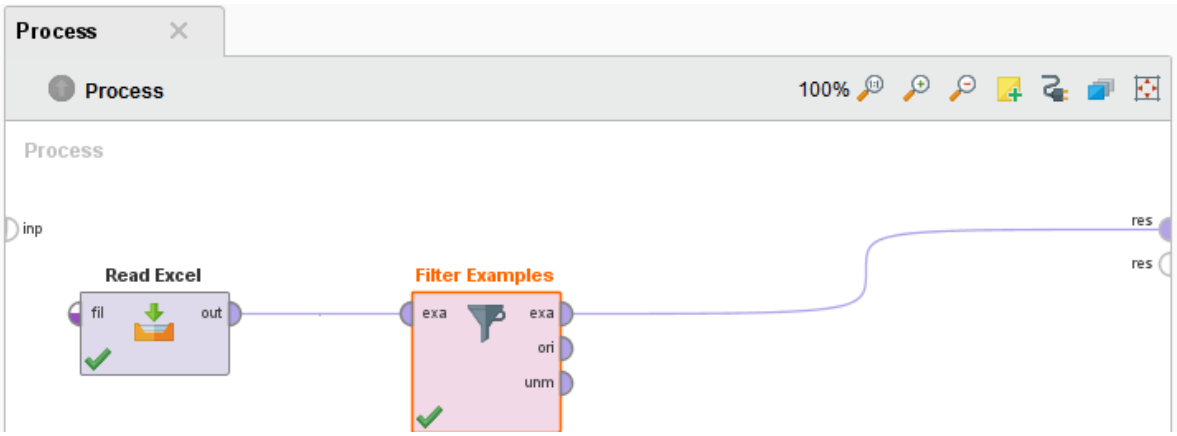

Fig. 6 Filter age of children

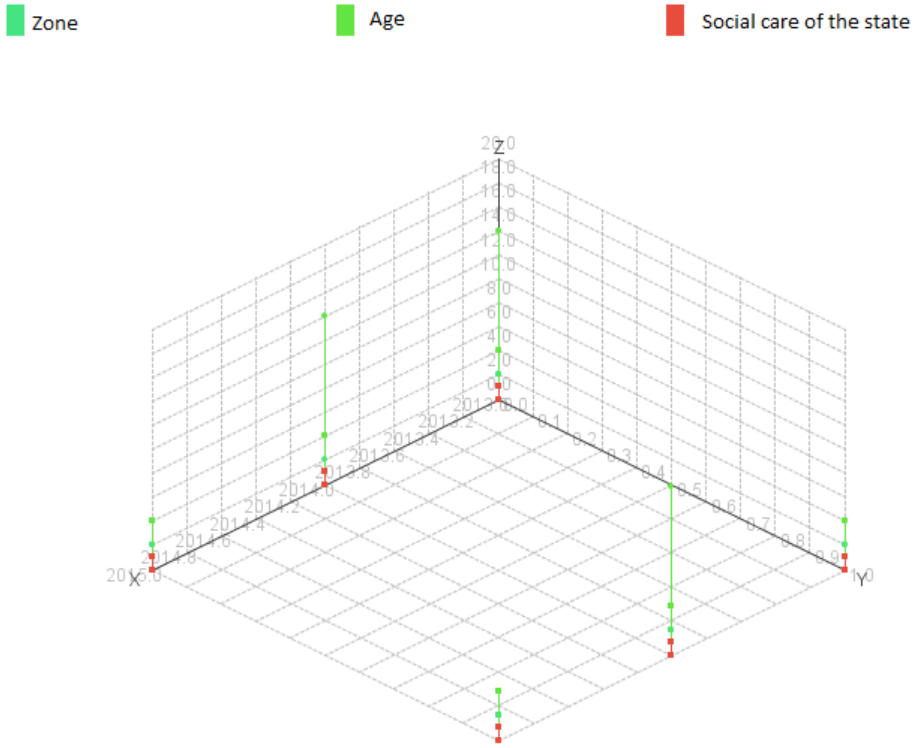

Fig. 7 Zone, Age and Social Attention

REFERENCES

[1] Alvarez, R. (1995). Caracterización general de la zona de sugerencia en la Guajira Colombia. Revista Academia Colombiana Ciencias , 679-694.
[2] Bogotá (provincia). Defensoría del Pueblo. (2014). Crisis Humanitaria en la guajira. Gómez, herrera, Y Cabrera: Autores.
[3] Bogotá (provincia). Departamento administrativo Nacional de estadística. (2005). Censo general de la Guajira. D.A.N.E.: Autor.
[4] Pedraja, R. (1981). La guajira en el Siglo Xix: Indigenas, Contrabando y carbón. Revista desarrollo y Sociedad, 329-358.
[5] Fajardo, G. (1979). Surgencias Costera en las Proximidades de la Peninsula Colombiana de la guajira. Centro colombiano de Datos Oceanográficos, 7-19.
[6] Gnanapriya, R. (2010). Data Mining Concepts and Techniques. Data Mining and Knowledge Engineering, 256-263.
[7] Bayona, G., Ochoa, F., Cardona, A., Jaramillo, C., Y Montes, C. (2007). Procesos Orogénicos del Paleoceno Para la Cuenca de Ranchería (Guajira, Colombia) y Áreas Adyacentes Definidos por Análisis de Procedencia. Revista Geología Colombiana, 21 – 46. 2007.
[8] Hegland, M. (2001). Data Mining Techniques. Acta Numerica, 313-355.
[9] Grouping Multidimensional Data. (2006). Berlin: Springer-Verlag Berlin Heidelberg. Kogan, J.: Author.
[10] Raval, K. M. (2012) Data Mining Techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 439-442.
[11] Kiani, R., Mahdavi, S., Y Keshavarzi, A. (2015). Analysis and Prediction of Crimes by Clustering and Classification. International Journal of Advanced Research in Artificial Intelligence, 11-17.
[12] Venkatesh Sharma, A. M. A. K. (2010). A Novel Approach in Extracting Medical Reports Using Mining Technique. Data mining and Knowledge Engineering, 221-226.
[13] Alghamdi, A. S. A. (2011). Efficient Implementation of FP Growth Algorithm-Data Mining. International Journal of Computer Science and Network Security, 7-16.
[14] Arpita, H. (2013). Comparative Analysis of Data Mining Tools and Techniques for Evaluating. International Journal of Computer Science and Applications, 232-237.
[15] Criales-Hernández, M. I. (2006). Flujos de Biomasa y Estructura de un Ecosistema de Surgencia Tropical en la Guajira, Caribe Colombiano. Revista de Biología Tropical, 1257 - 1282.
[16] Colombian Family Welfare Institute – ICBF Colombia. (2017). recovered from http://www.icbf.gov.co.