Journal of Scientometric Research, 2025; 14(1):309-318.
https://www.jscires.org

Research Article

Recognizing Patterns on Two Principal Causes of Death in the Mexican Population during the Period 2010-2021

Guillermo De Ita Luna[1,*], Erika Bonfil Barragán[1], Daniela Ximena De Ita Varela[2]

[1]Fac.Cs. de la Computación, Benemérita Universidad Autónoma de Puebla, 14 Sur Sur esq. Sn. Claudio, Edif. CCO1, C.P.72570, Puebla, MÉXICO.
[2]Fac. de Ingeniería, Universidad Popular Autónoma del Estado de Puebla, 21 Sur 1103, C.P.72410, Puebla, MÉXICO.

ABSTRACT

Two of the 11 primary causes of death among the Mexican population during the period 2010-2021 were analyzed in its correlation with Mexico's socio-economic situation and the identified patterns are presented. It has been confirmed that the application of learning algorithms, specifically the use of decision trees, is very useful in analyzing and correlating data from various sources, which is evidenced by the obtained results. The constructed decision trees illustrate and explain the patterns found between deaths from heart disease and suicide compared to certain socio-economic attributes from Mexico. This helps to characterize the population that experienced these deaths. Our results can be utilized to aid in the decision-making process for policies aimed at preventing health issues that affect the Mexican population, as well as to define other health public policies.

Keywords: Pattern Recognition, Learning Algorithms, Decision Trees, Heart Disease, Suicide in México.

Correspondence:
Guillermo De Ita Luna
Fac.Cs. de la Computación, Benemérita Universidad Autónoma de Puebla, 14 Sur Sur esq. Sn. Claudio, Edif. CCO1, C.P.72570, Puebla, MÉXICO.
Email: deitaluna63@gmail.com
ORCID: 0000-0001-7948-8253

Received: 25-07-2024;
Revised: 23-09-2024;
Accepted: 30-10-2024.

INTRODUCTION

Numerous variables can be utilized to measure the Life Expectancy (LE) such as demography, healthcare, economy, employment and security, among others. However, demography is renowned for its three principal components: mortality, birth rate and migration. Mortality, in particular, holds significant importance due to its strong correlation with economic and social variables.[1,2]

In Mexico, the National Institute of Statistics and Geography (INEGI) reports periodically death statistics and occasionally estimates some possible causes based on death public registries. Nevertheless, these studies do not provide insights in the relationships between death rates and social variables such as employment, poverty, internet use, among others. Thus, it is necessary to make research that truly allows us to identify and model these behaviors.

Several studies conducted by public institutions, such as INEGI and National Population Council (CONAPO), focus on presenting statistics and projections in demography, birth rate and mortality. There is some research in the field of mortality rates in Mexico and in the analysis of the poverty influence,[2] as well as the death rate analysis by causes.[3] Nonetheless, these studies adopt a purely statistical method and they often use more classical approaches like life tables[4] and model tables.[5]

Another relevant study is the one published by the Lancet Regional Health journal.[6] In this research the authors employed a Poisson regression model to predict specific causes of mortality during the COVID-19 pandemic period, based on the mortality data from 2015-2019. The authors showed that the largest increases in mortality were found in people with diabetes, respiratory infections, ischemic heart diseases and hypertensive diseases. On the other hand, there were some decreases in parasitic diseases, skin diseases, non-traffic related accidents and malignant neoplasm.

One common line of research focuses on dealing with just one specific illness or cause, as demonstrated in,[7] where the likelihood of mortality due to COVID-19 is measured by certain risk factors. In,[8] mortality prediction in breast cancer was conducted using Machine Learning techniques. Additionally, in[9,10] some statistical relations between the COVID-19 pandemic and the increased suicide rates in Mexico were analyzed.

Similarly, in Peru, a study concerning how the COVID-19 affected certain causes of death was conducted.[11] Some of the most important conclusions in the article highlighted that cases not related to COVID-19 were myocardial infarction, cardiac arrest and heart failure. The explanation was that most hospitals in Peru had partially disrupted services for circulatory emergencies

DOI: 10.5530/jscires.20251454

during the pandemic. However, the methods used in this research are not published.

In Ortiz Velásquez V.,[12] it is presented an analysis of the child death rate in Colombia by applying machine learning tools. However, in all previous studies, the analysis of the causes of death and its relation with economic variables applying machine learning techniques is not common.

With Life Expectancy (LE) stalling or declining in most parts of the world as well as the global economic downturn, it is critical to identify the most relevant factors affecting LE.[13] Identifying the most relevant factors from the less critical ones could provide a cost-effective way of boosting LE, since recent studies on LE have shown interesting prospects for the application of data science and machine learning to improve LE.[13] For example, in[14] a study on 132 counties revealed that a low mortality population has a strong correlation between life expectancy and the level of economic development, as well as between economic stagnation and LE stagnation.

We found almost no studies researching the effect of socioeconomic variables on death causes in the Mexican population and the studies that we found did not apply machine learning algorithms for the analysis of this data. Therefore, we provide a new contribution to the research of mortality rates in Mexico by using machine learning tools in order to analyze data related to the mortality rate in Mexico. Moreover, our purpose in this work is to correlate socio-economic attributes associated to the Mexican population with the main causes of death in Mexico.

In our study, we follow an orientation of data science. First, we start by recompiling from public repositories (INEGI, CONAPO and National Council for Evaluation of Social Development Policy-CONEVAL) special attributes about the socio-economic situations in Mexico during the period between 2010 and 2021. Throughout this compilation process, the missing annual reports of some socioeconomic attributes were requested via email from INEGI. In the case of the poverty attribute, whose measurement is carried out biannually, it was necessary to propose an interpolation polynomial to complete the information not reported by year.

With the objective of working under the same sample space, a normalization process was applied in which the percentage rate was obtained by dividing the affected population by attribute by the total Mexican population per one hundred thousand inhabitants. In equation (0) the normalization for the employment attribute is observed.

$$\frac{Working\ population\ 2010(100\ k)}{Total\ population\ 2010\ (100\ k)} * 100 = NormalizedEmployment2010 \qquad (0)$$

Subsequently, we did an initial pre-processing of the data: cleaning, adjusting and correlating the most important attributes versus the death causes. Afterwards, we apply machine learning tools on the recompiled data to get insights about mortality patterns in the Mexican population. We experiment with different automatic learning algorithms, i.e. applying neuronal networks, a support vector machine and decision trees. For this article, we have chosen the application of the algorithm CART and random-forest weka's utility to create decision trees in order to recognize patterns to model the relationships between causes of death and socio-economic variables in Mexico.

The author in Aguirre Agustín R.[15] made a study based on multiple learning algorithms that were applied to correlate causes of death with socio-economic variables in Mexico. This study mentioned the use of convolutional Neural Networks (ANN) and a Support Vector Machine (SVM), but it was found that the best results were achieved when decision trees were utilized. After reviewing previous literature on the topic and experimenting with different learning techniques such as SVM, ANN and decision trees, as well as analyzing the resulting graphics from previous methods, it was determined that decision trees would be the most appropriate machine-learning tool for this task.

The purpose of this work is to analyze the main causes of death in the Mexican population between 2010 and 2021 and to identify potential correlations with socio-economic variables in Mexico. The goal is to uncover significant patterns related to the individuals who died from heart disease and suicide during this period.

Our paper is structured as follows: Section 1 provides a general introduction. Section 2 introduces the methodology employed in data science and applied to this study. In Section 3, we present the main technique used for our analysis of the recompiled data, in this case, the application of decision tree algorithms. Section 4 presents the results of the decision tree algorithms used in this study. The final section 5 offers conclusions drawn from this work.

METHODOLOGY

The emergence of new tools in machine learning and data science has facilitated the generation of valuable insights, particularly from data, enabling a deeper understanding of life expectancy factors. In this study, the focus was on finding correlations between causes of death and socio-economic variables. The objective of this article is to show that learning algorithms are useful tools to explain correlations between the causes of death and the socioeconomic conditions of a population, which is why several machine learning algorithms were experimented with, including convolutional neural networks and vector support machine. However, the analysis revealed that decision trees were the most effective method to identify and illustrate the sought-after correlations.

The outcomes of implementing the J48 and random forest algorithms using the WEKA system are presented here, as well as the results from the CART Python module. We applied these systems to the data gathered from two primary causes of death in

Mexico and we connected these causes with variables that encode Mexico's socio-economic status from 2010 to 2021.

While our study focused on analyzing the 11 primary causes of death among the Mexican population in the period from 2010 to 2021, due to the length of this article and the inherent relevance for understanding some of the causes of death, we decided to present the results obtained on mortality due to only heart disease and suicide in Mexico by applying decision tree algorithms which identified patterns that characterize the population affected by these causes. This study highlights the usefulness of this method in understanding and characterizing such deaths in Mexico.

Data science is the application of scientific processes, methods and algorithms to extract insights and knowledge from data to inform decisions. Meanwhile, the learning model's primary objective is to uncover the underlying patterns within the data previously collected.[16]

This work utilized a common methodology employed in data science, which involves:

Collecting a dataset. This is an important step in machine learning as it provides the necessary information for the model to learn and make predictions.

Once the dataset has been collected, the next step is to algorithmically build a model based on the data. This involves training the model using various techniques and methods.

Selecting the right algorithm for the problem at hand. This is crucial step in order to achieve accurate and reliable results. It requires a good understanding of the different algorithms and their strengths and weaknesses.

Finally, analyzing and giving meaning to the results obtained, which is essential in interpreting the predictions made by the model. This involves examining the accuracy and validity of the results and making any necessary adjustments.

According to,[17] there are different metrics that are used when we do a regression instead of a classification, the most common are Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Mean Squared Error (MSE) computes the squared difference between the predicted and true values of single outputs is averaged. In equation (1), we can see its definition.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (1)$$

Where: y means the target value, $\hat{y}$ means the predicted value, i is the index for current sample and n means the number of predictions. This metric is used in the CART algorithm, which is one of the algorithms to be applied in our study.

Other relevant metric used for the regression process, is the Mean Absolute Error (MAE). MAE is denoted as the average of the absolute difference between the predicted and the true values of single outputs. In equation (2) we see its definition.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

Although these previous metrics are commonly used in regression process, they are also utilized in algorithms for building classification trees.

Data Repositories

There were three main data repositories used in the search and compilation of information on the socioeconomic situation of Mexico and the main causes of death in Mexico. We worked with data extracted from INEGI (National Institute of Statistics and Geography), CONEVAL (National Council for Evaluation of Social Development Policy) and CONAPO (National Population Council). The difficulty was that for Mexico very few socioeconomic variables are recorded and they are only found on very short periods. Despite this, a data set with 15,233,229 useful records and 103,367 incomplete records was compiled.

Based on the data obtained, the following variables have been selected for this study:

Employment: The number of people with formal employment, which was obtained from INEGI (National Survey of Occupation and Employment).

Safety: The incidence of crime, which was obtained from INEGI (National Survey of Victimization and Perception of Public Security).

Poverty: The number of people living in poverty according to CONEVAL.

Internet: The number of individuals with access to the internet has been measured and projected through the National Survey on the Availability and Use of Information Technologies conducted by INEGI.

PIB: Gross Domestic Product, which has been steadily increasing. Retrieved from INEGI (Goods and Services Accounts).

Access-Health: The number of people who are affiliated to a health care service. This number has had a significant rise in the last couple of years and is retrieved from INEGI (System of National Accounts of Mexico, SALUD).

Population: Registration of the country's population over the period from 2010 to 2021. This is retrieved from CONAPO (Population and Projections).

In some cases, such as the case of the 'poverty' data, an interpolation polynomial was applied due to the lack of annual information. The reason for this is that institutions do not typically publish data in an annual format and for certain data sets like poverty, the information is recorded biannually. In order to conduct this study, we had to acquire the general mortality records that were published by INEGI for the period from 2010 to 2021.

From these records, the data were classified according to the Mexican list of diseases in order to obtain the main causes of deaths within the same period. Based on our analysis, the main causes of death among the Mexican population during this period are: Heart Disease, Diabetes, Tumors, Influenza, Liver Disease, Cerebrovascular Disease, Assault, Accidents, Lung Disease, Kidney Failure, Perinatal deaths and Suicide. The graphs below depict the trends in the number of cases for each cause of death in the Mexican population from 2010 to 2021.

As we can distinguish in Figure 1, one of the most relevant causes of deaths in Mexico is heart disease, which has a consistent percentage of almost 20% in average. Meanwhile, the second and third most common cause of death is diabetes and tumors. Another important factor that we can recognize is the increase on cases related to influenza and a new peak point with this cause of death on 2020. On the other hand, other causes such as lung, assaults and renal decrease considerably on the year 2020. Lastly, we can see that the percentages of total deaths related to suicide is almost consistent with around 0.8% in all years.

Learning algorithms based on the construction of a decision tree

The domain of pattern recognition is focused on the automated identification of consistent patterns within data through the utilization of computer algorithms. Additionally, it pertains to the application of these identified patterns for the purpose of decision-making and the systematic categorization of data into distinct classes.[16] Decision trees are among the most popular machine learning algorithms given their intelligibility and simplicity.[18]

Decision or classification trees are a type of machine learning model that uses a database to create diagrams of logical construction, which can be used to explain the patterns that are hidden behind the data. Decision tree classification algorithms are widely used as they have proven to be robust methods for correlating variable-attributes with class values. These algorithms report a hierarchy of significance on attributes values to determine class values, which makes them very useful for different applications.

Decision trees are also known as hierarchical segmentation. It is a decomposition technique that uses a sequential, iterative and descending division process. It begins with a dependent variable and forms homogeneous groups specifically designed with the

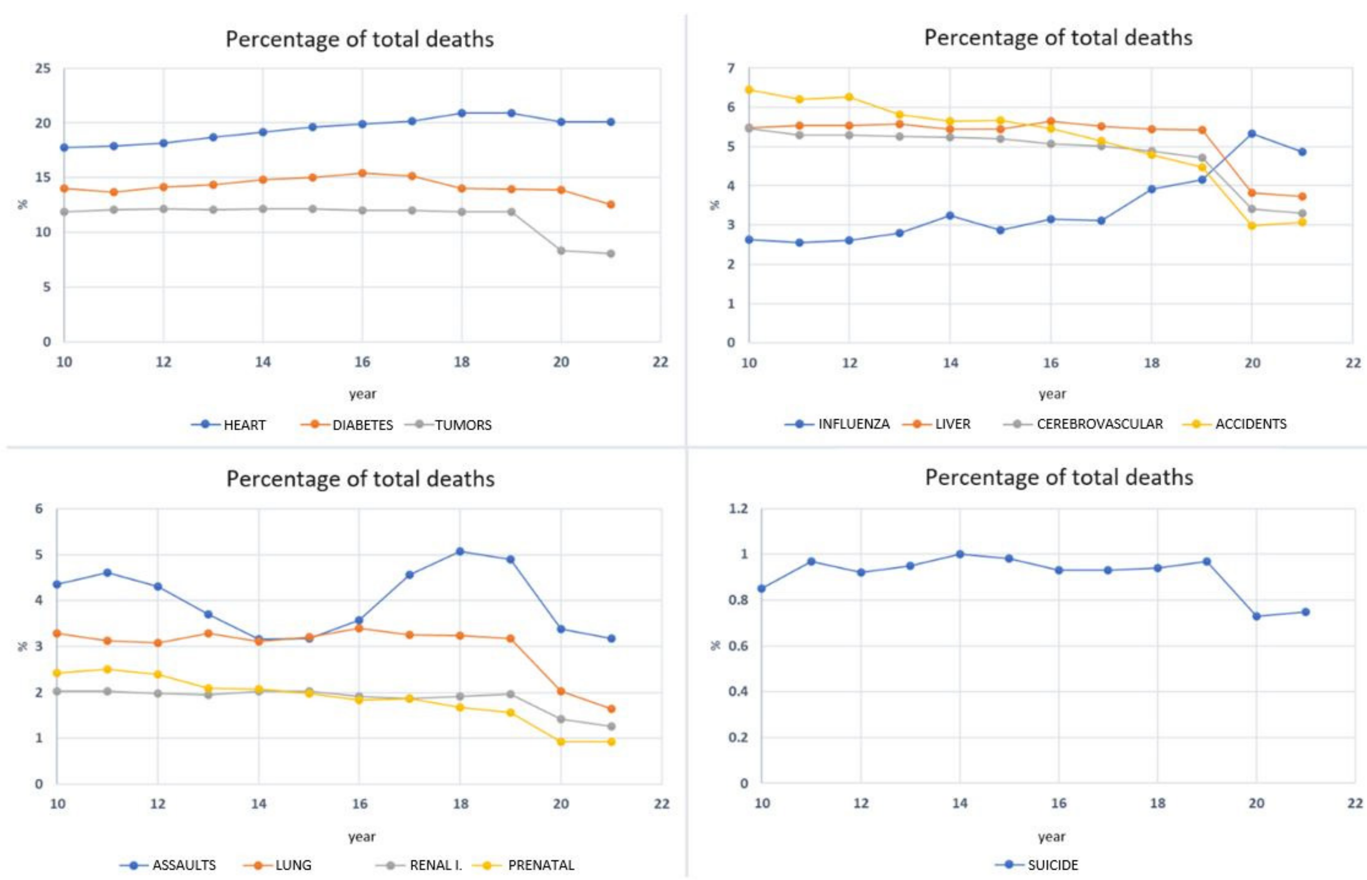

Figure 1: Percentage of total deaths for each cause of death during the period from 2010 to 2021.

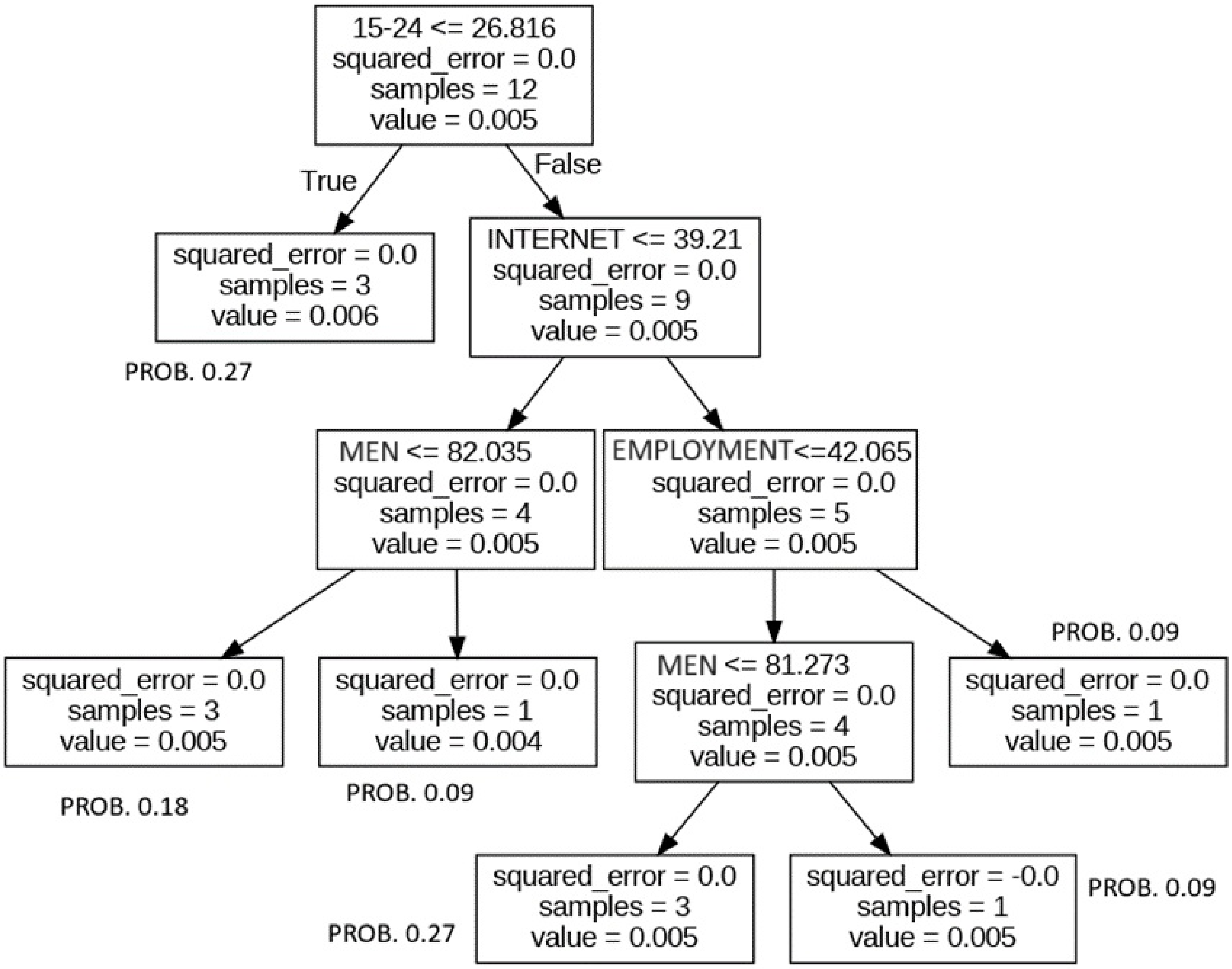

Figure 2: Results from CART algorithm for suicides in Mexico (2010-2021)

combinations of independent variables, ensuring that all the cases collected in the sample are included.

Our research revealed that decision trees, which are derived from machine learning, were the most suitable algorithms for identifying and prioritizing the relevance of attributes that characterize the population affected by each of the primary causes of death in Mexico.

One of the most influential algorithms in decision tree construction is the ID3 algorithm (Iterative Dichotomiser 3) proposed by Ross Quinlan.[19] The ID3 algorithm follows a top-down approach to construct decision trees and starts by addressing the question of "which attribute should be tested at the root of the tree?" To determine the answer, the algorithm evaluates each attribute using a statistical test that measures its information's entropy. This means that each attribute is evaluated using a statistical test that calculates the information gain concerning class values for each feature. This helps to determine how well the attribute ranks the data samples in the training set.

The attribute with the highest information gain is selected as the new node in the tree and then, a descendant of the current node is created. The branches of the tree are then created based on each possible value of the chosen attribute, while the training samples are sorted concerning the appropriate descending node (under the branch corresponding to the test value of this attribute). The entire process is repeated using the test samples associated with each descending node in order to select the best attribute at each level of the tree.

The ID3 algorithm is an avid form of search, which is an efficient way to build a decision tree in which the algorithm never backtracks to consider better options. The search strategy of the ID3 algorithm can be divided into two main parts.

It has a preference for shorter trees.

It selects trees with the highest information gain and those that are closer to the root.

On the other hand, the CART algorithm is the acronym for Classification and Regression trees designed by Breiman *et* *al.* in 1984.[20] The algorithm is capable of generating binary or multi-decision trees, meaning that the nodes in a CART tree can have more than two children. This feature allows the modeling of complex relationships within the data. This model can handle nominal, ordinal and continuous input and output variables, which makes it suitable for solving both classification and regression problems.

CART is indeed quite similar to the ID3 algorithm, but with one key difference-it can handle attributes with numerical and textual values, which is great for regression analysis. CART algorithm works by constructing binary trees, where each node is determined by the feature and threshold that provides the greatest gain of information. To measure the impurity of each node, CART uses the Gini index, which is calculated using the following pair of equations.

$$Gini\ Impurity(D) = \sum_{i=1}^{k} \left(\frac{n_i}{n}\right) Gini(D_i) \quad (3)$$

Where:

n_i is the number of cases for each value of class and

$$\boldsymbol{Gini(D_j) = 1 - \sum_{j=1}^{c} (p_j)^2} \quad \textbf{(4)}$$

Where:

D_j are the subsets of the class D.

p_j is the probability of samples belonging to class j at a given node.

c is the number of classes.

We have utilized Python's Sci-Kit Learn library and Graphviz library[17] to implement the CART algorithm, which also enabled us to render the graph-tree illustrated in Figure 2.

Results of the decision tree algorithms

Among the different decision trees algorithms applied to our data, we have selected the CART algorithm (provided by Phyton utilities) and the Randomforest algorithm included in the Weka system, since the obtained results were the most significant for our study. In the case of Randomforest, it is a machine learning algorithm that uses an ensemble learning technique. It creates as many trees on the subset of the data and combines the output of all the trees. In this way, it reduces the variance and the overfitting problem in decision trees; therefore, it improves the accuracy of the classification.

We start showing the results obtained considering suicide in Mexico as the class value for the classification tree. And, in a second phase, we show the decision tree obtained for deaths originated from heart disease. We have to consider that those data were recompiled for the Mexican population during the period from 2010 to 2021.

In Figure 3, we show the relevance of the main attributes found in the case of suicide. Analyzing the hierarchy of those attributes, we can identify that the attribute age in the range from 15 to 24 years represents the attribute with the maximum relevance for suicide and its impact is 76.73% during the interval from 2010 to 2021.

The second attribute with the greatest impact was the use of the Internet. This result highlights a new parameter which was not considered in previous statistical analyses of suicide in Mexico,[3,6,7,9,10] since it is an attribute that only recently came to be prominent. The weight of this attribute was 17.36 which show the importance of including this variable for future studies and on the prevention of this problem.

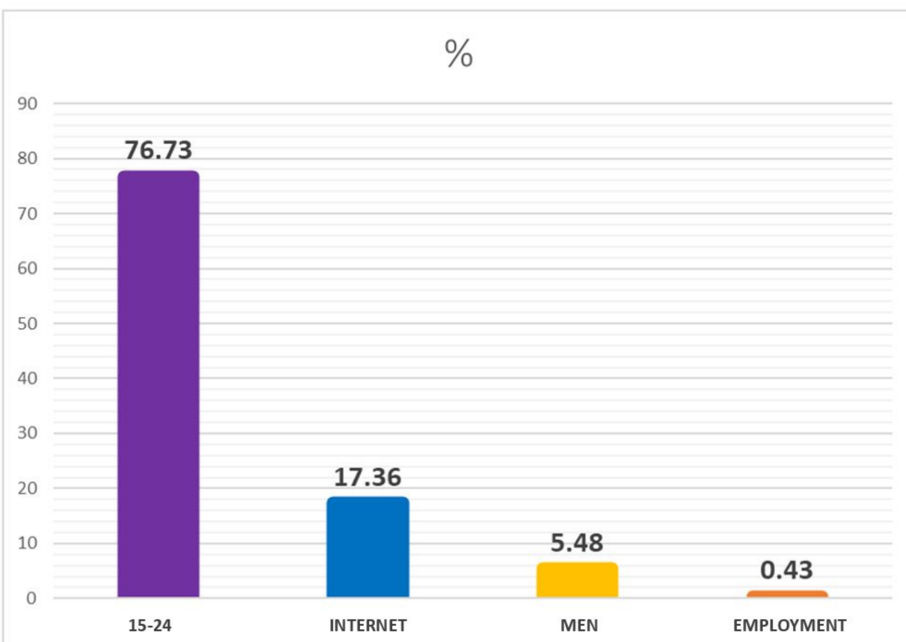

Figure 3: Relative weight among the principal attributes that characterize suicide in Mexico (2010-2021).

The third and fourth relevant attributes for this problem are gender and employment, respectively. Those values show that there is a difference between men and women who commit suicide, being that men are the ones who have a higher suicide rate. Meanwhile, the lack of employment is also a relevant characteristic among those who committed suicide. Notice that the mean squared error was used as an error function to obtain optimal classification nodes.

The Figure 4 displays the main attributes that define individuals who have committed suicide in Mexico from 2010 to 2021 and their relative weights.

Using the CART algorithm, we can derive a set of decision rules from the classification tree. These rules can then serve as a guide to build prospecting systems. Thus, the systems can guide us in categorizing new instances and help us make informed decisions based on the classification tree.

If AGE[15-24]<=26.816: VALUE=0.006 (PROBABILITY 0.27)

If AGE[15-24]>26.816 and INTERNET<=39.21 and MEN<=82.035:

VALUE=0.005 (PROBABILITY 0.18)

If AGE[15-24]>26.816 and INTERNET<=39.21 and MEN>82.035:

VALUE=0.004 (PROBABILITY 0.09)

If AGE[15-24]>26.816 and INTERNET>39.21 and EMPLOYMENT<=42.065 and MEN<=81.273:

VALUE=0.005 (PROBABILITY 0.27)

If AGE[15-24]>26.816 and INTERNET>39.21 and EMPLOYMENT<=42.065 and H>81.273:

VALUE=0.005 (PROBABILITY 0.09)

If AGE[15-24]>26.816 and INTERNET>39.21 and EMPLOYMENT>42.065:

VALUE=0.005 (PROBABILITY 0.09)

The discovered patterns of suicide are not only consistent but also augment other analyses regarding this health issue in the Mexican population.[6,7,9,10] This opens up an additional aspect for study and, particularly, for the prevention of this problem that significantly surged during the COVID-19 period, especially among the young population of Mexico. Due to these increases, an attribute has been added to the study in order to consider the number of deaths from COVID-19 in Mexico from the years 2020 and 2021. This analysis resulted in the following decision tree based on the Random forest algorithm included in the Weka system.

This new decision tree showed in Figure 5 has significant changes from the decision tree in Figure 2, since it considers now the COVID-19 attribute as the root node of the tree, which means that the pandemic of COVID-19 had a significant impact on the suicide population during 2020 and 2021. Moreover, the tree indicates that there is a correlation between suicide and the lack of access to healthcare, unemployment and insecurity. From different analysis,[9,10] it has been determined that over half of the individuals who committed suicide in Mexico during the years 2020 and 2021 were young adults between the ages of 15 and 33. In addition, the data suggests that there were significantly more cases of men than women in this population, as it can be seen in

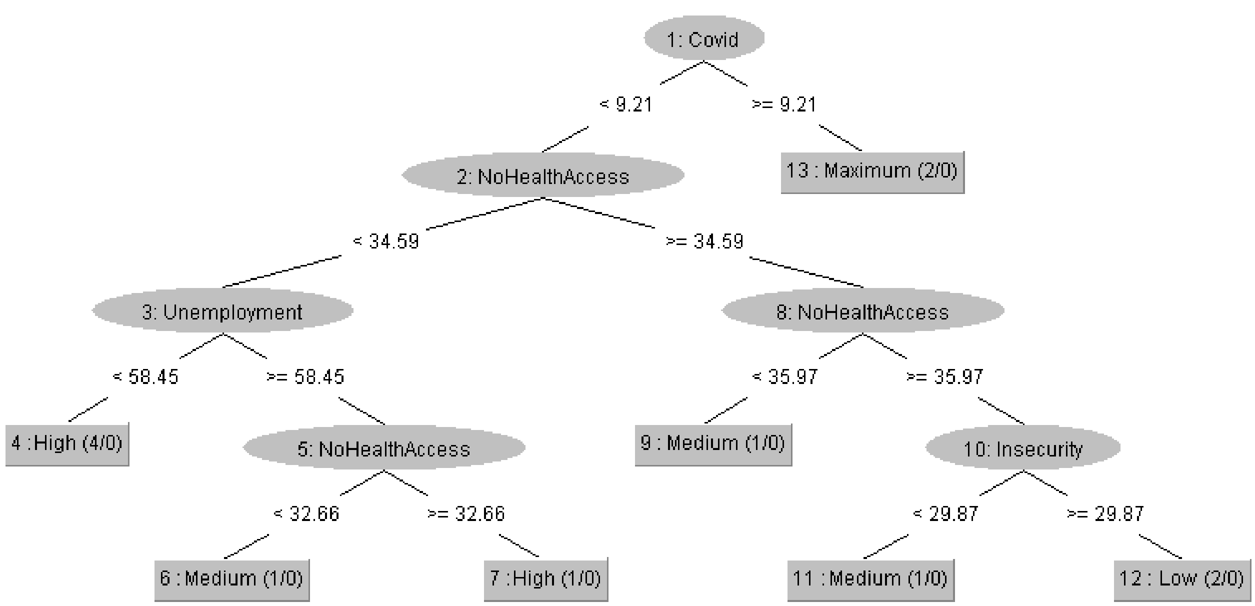

Figure 4: A decision tree on suicide data considering the number of cases of COVID-19 in Mexico.

Figure 6. The previous analysis confirms that suicide is a complex and multifactorial health problem, where the machine learning algorithms are extremely helpful to allow the establishment of new guidelines for prevention and health care strategies, especially of the young population.

As a result of the limited space available for this article, we will only share the outcomes of another decision tree. In this case, an analysis related to deaths caused by heart disease in Mexico is shown. It is worth noting that except from 2020 to 2021, heart disease has been the primary cause of death among the Mexican population throughout the decade spanning from 2000 to 2019.

The decision tree shown in Figure 5 was generated using the CART algorithm from the Python system. It appears that the population that died from heart problems consisted of a higher proportion of men than women. When we consider the relative weights of the main attributes that characterize the heart disease population, the percentage of men instead of women has a relative weight of 87.14%. On the other hand, poverty and access to health were found to have a high correlation with this population with 10.54% and 1.56% of weight for each attribute, respectively.

The acquired data from the tree diagram suggests that implementing preventive health programs and early detection measures for potential coronary issues could help prevent deaths from heart problems in Mexico. It is recommended to focus on the target population, specifically men aged 50 and above. In the following Figure, we show the weighting of the most relevant attributes for this population.

From the CART algorithm, we show now the set of decision rules that derive from the classification tree for the heart disease in Mexico in the period 2010-2021.

If MEN<=54.131 and POVERTY<=45.335 and HEALTH<=70.33:

VALUE=0.111 (PROBABILITY 0.25)

If MEN<=54.131 and POVERTY<=45.335 and HEALTH>70.33:

VALUE=0.122 (PROBABILITY 0.1666)

If MEN<=54.131 and POVERTY>45.335 and PIB<=13737.565:

VALUE=0.093 (PROBABILITY 0.25)

If MEN<=54.131 and POVERTY>45.335 and PIB>13737.565:

VALUE=0.100 (PROBABILITY 0.1666)

If MEN>54.131: VALUE=0.175 (PROBABILITY 0.1666)

These rules could serve as a helpful guide to develop a prospecting system that assists in categorizing new instances, or in the study of the expected trends related to this cause of death in Mexico. For example, according to our analysis on suicide in Mexico, it would be relevant to design social programs that support the mental health of the young population in Mexico, i.e. look on the internet

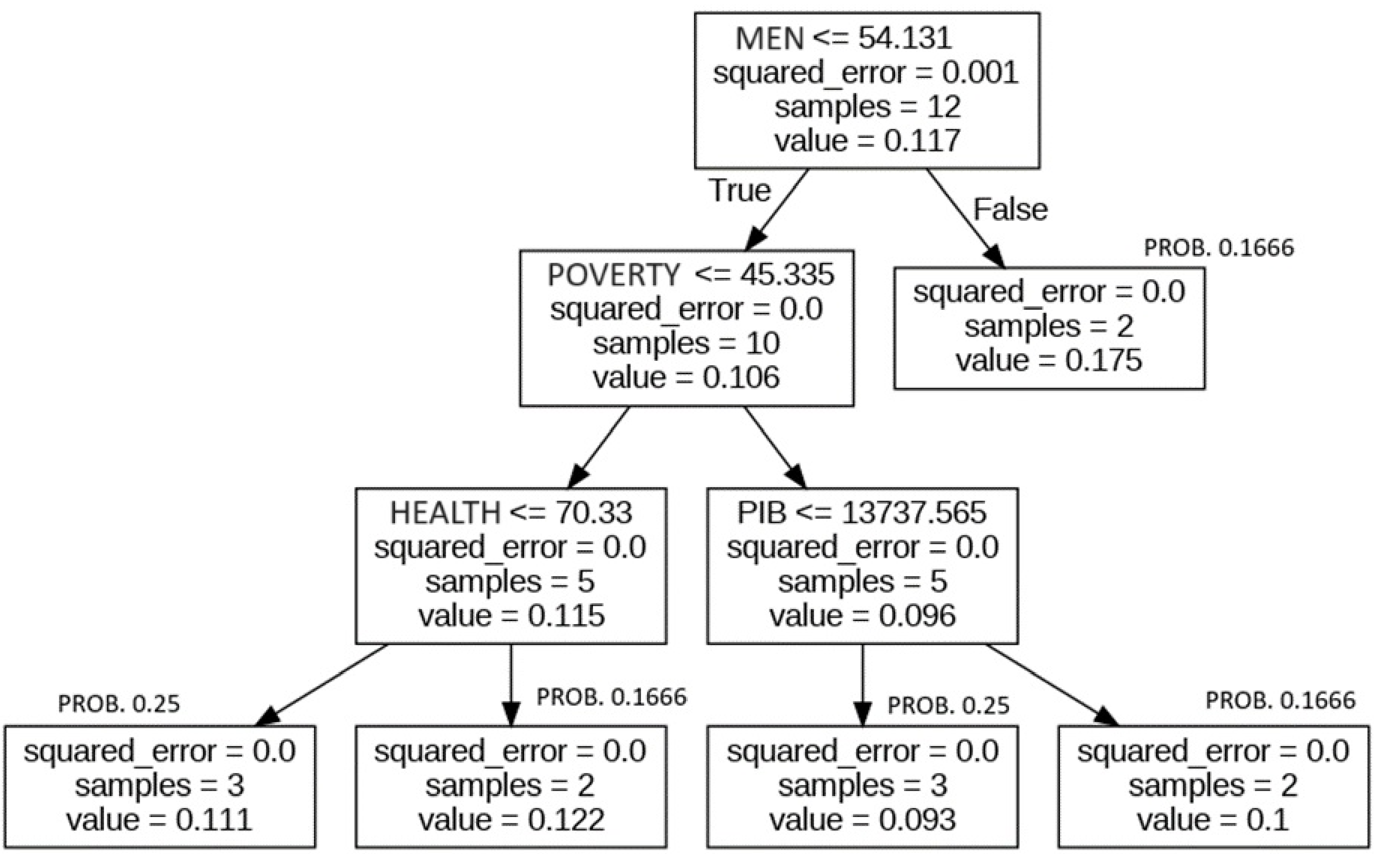

Figure 5: A decision tree on heart disease data for the Mexican population from 2010 to 2021.

for depression cases in the young population. Those tasks could be part of a social program to stop the increase of the suicide rate in Mexico. Furthermore, it is important to design preventive health programs and early detection measures for potential coronary issues that could help prevent deaths from heart illness in Mexico and could be, for example, mainly be focused on the men gender population.

CONCLUSION

In this article, we present the analysis of two out of the 11 leading causes of death in the Mexican population in the period from 2010 to 2021. Our results confirm the significant usefulness of learning algorithms, specifically the use of decision trees, in the recognition of patterns from the socioeconomic data and causes of death. Our analysis illustrates the strong correlation between the socio-economic variables and the two causes of death that were chosen to present in this article.

For example, for the population in Mexico that died from heart problems in the period from 2010 to 2021 and through its correlation with the socio-economic variables of the population, it was found that men had a higher incidence rate (54%) compared to women and identified poverty and access to healthcare as significant factors for this population.

Meanwhile, for those who died by suicide in Mexico during the period from 2010 to 2021, the attribute age in the range from 15 to 24 years represents the attribute with the maximum relevance for that cause of death. The second attribute with the greatest impact was the use of the Internet. Moreover, the men are the ones who have a higher suicide rate. Meanwhile, the lack of employment is also a relevant characteristic among those who committed suicide. However, our analysis also shows that the COVID-19 had the most significant impact on the suicide population during 2020 and 2021.

Our results can be utilized to aid in decision-making policies aimed at preventing health issues that affect the Mexican population and strengthen existing programs that improve the life expectancy of the Mexican population. For instance, providing mental health support to the youth and implementing suicide prevention programs could be effective ways to achieve this goal.

As future work, we aim to validate the patterns obtained using the data generated in the subsequent years and enrich this research by considering subgroup analysis, exploration of interactions and the evaluation of the impact of this study, since at the moment, we can only conjecture the future benefit, but not trace the effects of it.

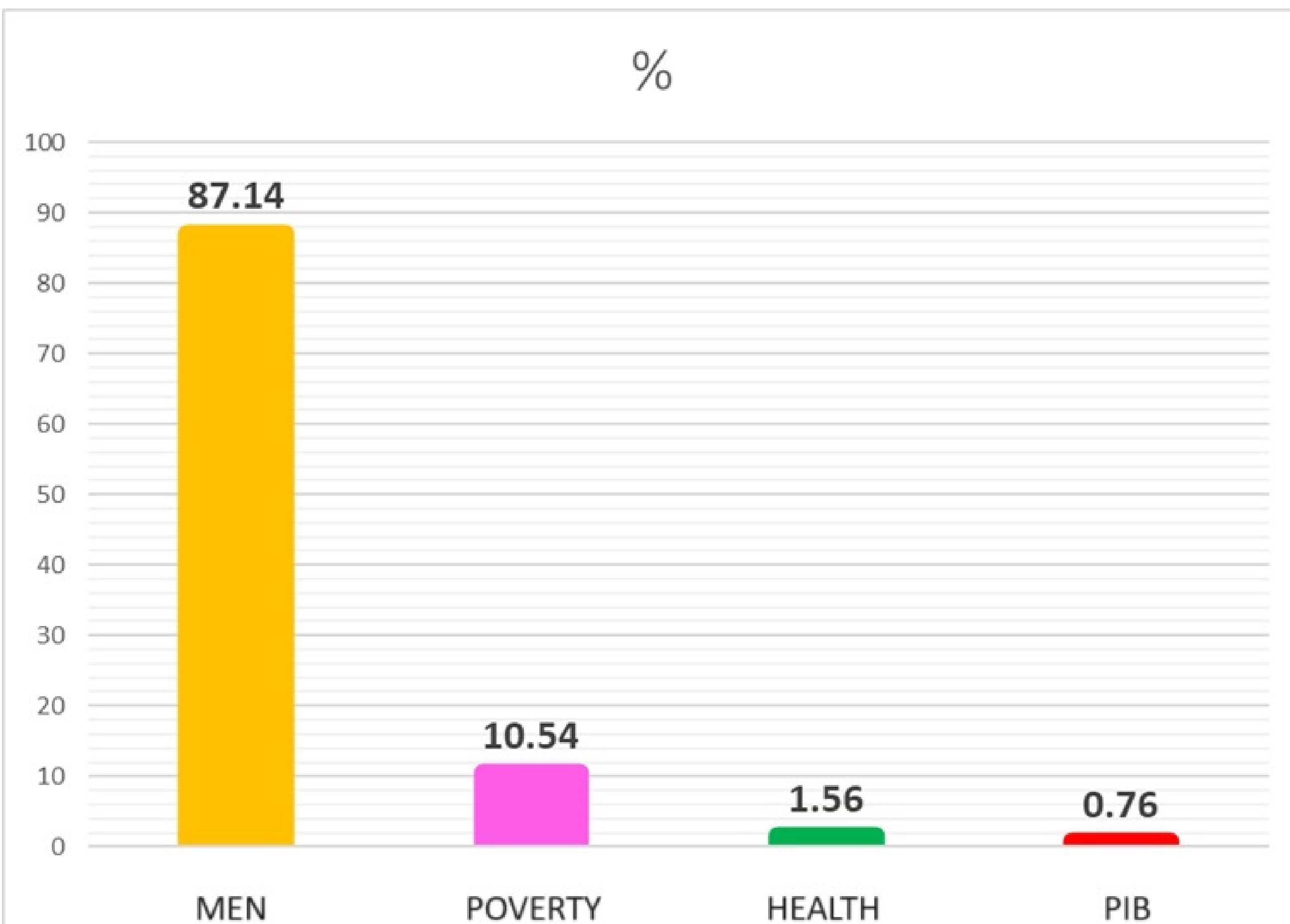

Figure 6: Relative weight among the principal attributes to characterize heart disease in Mexico.

ACKNOWLEDGEMENT

The first author thanks the SNII-México, while the third author thanks Conahcyt - México, for the support provided.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

LE: Life Expectancy; **INEGI:** National Institute of Statics and Geography of México; **CONAPO:** National Population Council of México; **CONEVAL:** National Council for Evaluation of Social Development of México; **ANN:** Neural Network; **SVM:** Support Vector Machine; **MAE:** Mean Absolute Error; **CART:** Classification and Regression Trees.

REFERENCES

1. Sen A. La vida y la muerte como indicadores económicos. Investig Cienc. 1993;1(202):6-13.
2. Zavala R, Isait D: Mortalidad y pobreza: México [Master's thesis]. Colegio de México; 2000.
3. Eligio S. La mortalidad por causas en México y las ganancias en las esperanzas de vida [Master's thesis]. Universidad Nacional Autónoma de México; 2002.
4. Cervantes CA. Análisis de la mortalidad por causas en México: tendencias y proyecciones al 2015 [Master's thesis]. El Colegio de México; 2008.
5. Cambría DJ. Mortalidad como indicador económico y social. Argentina y América Latina. Rev Salud Publ. 2012;16(2):57-66.
6. Palacio-Mejía LS, Hernández-Ávila JE, Hernández-Ávila M, Dyer-Leal D, Barranco A, Quezada-Sánchez AD et al. Leading causes of excess mortality in México during the COVID-19 pandemic 2020-2021: A death certificates study in a middle-income country. The lancet regional health – Americas. Vol. 13. Available from: https://www.sciencedirect.com/science/article/pii/S2667193X2200120X. doi: 10.0303(2022).
7. Becerra-Sánchez A, Rodarte-Rodríguez A, Escalante-García NI, Olvera-González JE, De la Rosa-Vargas JI, Zepeda-Valles G et al. Mortality analysis of patients with COVID-19 in México based on risk factors applying machine learning techniques. Diagnostics (Basel). 2022;12(6):1396. doi: 10.3390/diagnostics12061396, PMID 35741207.
8. COVIDSurg Collaborative, Machine learning risk prediction of mortality for patients undergoing surgery with perioperative sars-cov-2: the covidsurg mortality score. Br J Surg. 2021;108(11):1274-92. doi: 10.1093/bjs/znab183, PMID 34227657.
9. Baltazar-Gaytan E, Aguilar-Alonso P, Brambila E, Tendilla-Beltran H, Vázquez-Roque RA, Morales-Medina JC et al. Increased cell number with reduced nitric oxide level and augmented superoxide dismutase activity in the anterior-pituitary region of young suicide completers. J Chem Neuroanat. 2019;96:7-15. doi: 10.1016/j.jchemneu.2018.11.002, PMID 30423351.
10. García-Dolores F, Tendilla-Beltrán H, Flores F, Carbajal-Rimoldi LA, Mendoza-Morales RC, Gómez-Mendoza LE et al. Increased suicide rates in Mexico City during the COVID-19 pandemic outbreak: an analysis spanning from 2016 to 2021. Heliyon. 2023;9(6): e16420, Celpress. doi: 10.1016/j.heliyon.2023.e16420, PMID 37274652.
11. Liu Y, Zhao S, Yang L, Aliaga-Linares L, He D. All-cause mortality during the COVID-19 pandemic in Perú. IJID Reg. 2022;5:177-9. doi: 10.1016/j.ijregi.2022.10.005. PMID 36339931.
12. Ortiz Velásquez V. Aplicación de técnicas de aprendizaje automático para la segmentación y clasificación de características sociodemográficas asociadas a tasas de mortalidad infantil utilizando datos reportados por el DANE Colombia entre los años 2008 al 2017. Universidad de Bogotá; 2020. p. 1.
13. Ronmi AE, Prasad R, Raphael BA. How can artificial intelligence and data science algorithms predict life expectancy-An empirical investigation spanning 193 countries. Int J Inf Manag Data Insights. 2023;3(1):100168. doi: 10.1016/j.jjimei.2023.100168.
14. Caselli G, Drefahl S, Wegner-Siegmundt C, Luy M. Chapter 5. Future mortality in low mortality countries. In: World population & human capital in the twenty-first century: an overview; 2017. Available from: books.google.com.
15. Aguirre Agustín R. Reconocimiento de patrones de mortalidad en la población mexicana usando aprendizaje automático [Master's thesis in computer science]; Noviembre 2023, [Computer science Deparment, Benemérita Universidad Autónoma de Puebla].
16. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2007.
17. Kinsley H, Kukiela D. Neural networks from scratch in python: building neural networks in raw python. Kinsley: Harrison; 2020.
18. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H et al. Top 10 algorithms in data mining. Knowl Inf Syst. 2008-01-01;14(1): 1-37. doi: 10.1007/s10115-007-0114-2.
19. Quinlan JR. Induction of decision trees, machine learning. Vol. 1. Kluwer Academic Publishers; 1986. p. 81-106.
20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees, Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8; 1984.

Cite this article: De Ita GDL, Bonfil EB, De Ita DXDV. Recognizing Patterns on Two Principal Causes of Death in the Mexican Population during the Period 2010-2021. J Scientometric Res. 2025;14(1):309-18.