International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 37, No. 4, August, 2023, pp. 817-824
Journal homepage: http://iieta.org/journals/ria

Determination of Steel Area in Reinforced Concrete Beams Using Data Mining Techniques

Jose Manuel Palomino Ojeda[1], Nancy Pérez Herrera[1], Lenin Quiñones Huatangari[1*], Billy Alexis Cayatopa Calderón[2]

[1] Instituto de Investigación de Ciencia de Datos, Universidad Nacional de Jaen, Jaén 06800, Peru
[2] Instituto de Investigación en Sismológica y Construcción, Universidad Nacional de Jaen, Jaén 06800, Peru

Corresponding Author Email: lenin.quinones@unj.edu.pe

https://doi.org/10.18280/ria.370401

Received: 16 April 2023
Revised: 16 May 2023
Accepted: 25 May 2023
Available online: 31 August 2023

Keywords:
artificial intelligence, computational models, data mining, rectangular beams, reinforced concrete

ABSTRACT

This study aimed to determine the area of reinforcing steel in rectangular reinforced concrete beams, a critical concern given that a significant proportion of residential structures in Peru do not conform to technical design regulations. Data were collected through a structured form encompassing various design variables, yielding a comprehensive data matrix. The methodology adopted involved Knowledge Discovery in Databases (KDD), executed in several stages: (1) selection, where the InfoGainAttributeEval algorithm was utilized to identify variables influencing the reinforcing steel area; (2) preprocessing, during which anomalous and duplicate data were purged using the Python libraries, Pandas and Numpy; (3) reduction, and (4) data mining. For the latter, Decision Stump, Hoeffding Tree, J48, Logistic Model Trees (LMT), and Random Tree classification algorithms were employed, facilitated by Weka 3.9.4. Accuracy rates of these algorithms were found to be 25, 51.70, 77.97, 78.39, and 88.98% respectively. The Random Tree algorithm, in conjunction with the GP_04 model, provided estimations of the steel area in the beams with a reliability exceeding 88%. The application of this model could enable the optimization of beam design, facilitate cost savings on materials, and enhance structural safety. This research thus presents a significant contribution to the field of structural engineering, particularly in regions where compliance with technical design standards is suboptimal.

1. INTRODUCTION

Most Latin American countries have a "housing problem" [1]. In Peru, 70% are informal and have no structural design, causing material losses. These problems become evident when seismic events occur, causing the collapse of structural elements because they have not been designed by a civil or structural engineer, do not comply with technical standards, and the personnel is not qualified for construction [2]. Generating demolition and reconstruction of structures, with rehabilitation and strengthening being a more cost-effective alternative [3]. This affects most of the self-built houses in Peru, making them uninhabitable and vulnerable to structural collapse [4]. In addition, structural deformation must be controlled to ensure the safety of concrete structures [5]. These are designed primarily for strength, using standards such as the American Concrete Institute (ACI) and Technical Standard E.060 (Reinforced Concrete), where the structure can reach a limited state of damage based on the loss of compressive strength capacity of the concrete [6].

Reinforced concrete structures are exposed to various factors that can cause functional and structural damage over time [7]. Understanding the cracking behavior and failure load of reinforced beams is crucial for designing robust and strong structures [8]. Cracking may occur in the early stages, followed by a possible decrease in yield strength and ultimate load capacity [9]. This is because they have not been designed to withstand loads according to their use [10]. Reinforced concrete beams are subjected to loads and their design must ensure ultimate load capacity, stiffness, ductility, and energy absorption capacity [11], being destined to fail typically by steel creep followed by concrete failure [12]. In addition, increasing the reinforcement ratio can reduce deflection and deformation while maintaining the same stiffness. However, increasing the reinforcement ratio will also cause the tensile strength to be underutilized, resulting in economic waste and brittle failure [13]. Mohammed [14] recommends reinforcing beams with a composite section by reusing the beam web and reinforcing it with epoxy, but in Peru, they are not compatible with design standards, have construction errors, or are exposed to unforeseen loads [15]. The most common beam failures are critical cracks, diagonal cracks in the web [16], out-of-limit deflections, and multiaxial stress states [17].

Proper design of beams is of great importance because it ensures the safety, stability, durability, and economy of structures by distributing loads along their length, avoiding excessive concentration of stresses at certain points on other elements, controlling structural deformations, is resistant and the materials used for its manufacture are low cost compared to steel beams [18].

The method used to calculate the steel in the beams is based on the ACI and Technical Standard E-0.60 (Reinforced Concrete), which uses several iterative processes. The beam section is pre-dimensioned according to the free span and the dead load is determined. Then, the nominal moment is calculated and checked if it is greater than or equal to the bending moment divided by the load factor (Ø). This process is repeated until a sufficient steel area is obtained, which makes it difficult to match the section's resisting moment with the ultimate applied moment due to the beam's weight. Therefore, the design process is slow and has no economic analysis [19]. Other methods used in the design of reinforced concrete beams are serviceability limit state design and fire resistance design using bar diameter, camber, and loading regime. Here, increasing the beam diameter or decreasing the camber thickness is an effective method to reduce the maximum crack width [20]. These methods are complex and require the use of specialized seismic and structural analysis software, which requires time and effort to learn how to use them correctly. Faced with this problem, data mining algorithms have been used to discover patterns, trends, and relationships in large data sets due to their applications in structural data management, helping engineers to organize, store, and analyze large amounts of information about existing structures. This facilitates the design of structural elements, monitoring of structural health, and informed maintenance and repair decisions.

2. LITERATURE REVIEW

Authors have addressed the problem as Coello Coello et al. [19] present an optimization model for the design of rectangular beams, using genetic algorithms to minimize the cost of the beam in the design procedures, taking into account the cost of concrete, steel, and formwork. They present their methodology for adjusting the parameters: population size, crossover rates, mutation, and the maximum number of generations, obtaining a cost of 47.80%, 41.50%, and 10.70% of the total cost in steel, concrete, and formwork, respectively [19]. Chakrabarty [21] proposes a nonlinear programming model for the optimal bending design of a reinforced concrete beam, where the input values are the bending moment and other parameters, and the output variables are the beam cross-sections. Other authors use data mining techniques in other areas as a solution to various problems; Zhang et al. [22] uses data mining algorithms, regression tree, and gradient boosting, to predict mechanical properties of concrete in the context of dam construction, evaluating models using mean absolute error, mean square error, correlation coefficient, and improved synthesis index, using K-fold cross-validation to verify the robustness and sensitivity of prediction models. Fernández-Ceniceros et al. [23] designed an algorithm based on data mining techniques to optimize costs in concrete structures; combining the different components and materials used in construction to minimize the final price and energy consumption to obtain optimal typologies, allowing the exploration and analysis of different combinations of components and materials to find more efficient solutions from an economic and energy point of view, the limitations of the study were the data dependency because these models require representative data. Kang and Choi [24] propose BIM (Building Information Modeling) based data mining methods for decision-making and management of energy consumption in buildings. Extracting insights and hidden patterns in building energy consumption data to make more informed decisions in managing energy consumption and identifying opportunities for improvement. In order to identify energy inefficiencies, optimize control systems, and implement energy efficiency strategies, the study was limited by the quality and availability of the data included in the model. If the BIM data is not accurate or complete, the results obtained may not be reliable or representative. Mataei et al. [25] is based on a data mining method to determine the optimal maintenance and rehabilitation (M&R) policy for Iran's road network by modeling historical data, identifying deterioration patterns to propose efficient and effective M&R strategies, allocating resources optimally, the limitations of the research were the availability and quality of data and consideration of other variables such as available budget, logistical constraints, and region-specific needs. Sun et al. [26] employs decision trees and other Methods for bridge damage detection based on dynamic fingerprints, which are measurements obtained from the vibrational responses of a bridge, which seek to identify patterns or anomalies that may indicate the presence of damage to the bridge structure to identify possible structural damage, the limitations of the research were the availability and quality of data are limited or unrepresentative, the models may be less accurate or generalizable. Mansouri et al. [27] use decision trees and other methods for predicting the behavior of concrete embedded in the fiber-reinforced polymer, which is a composite material used in structural engineering to improve the strength and load-bearing capacity of concrete. Allowing the analysis and processing of data related to the mechanical properties and performance of concrete, the limitations of the models were the availability and quality of data used in the analysis. Kang et al. [28] uses decision tree algorithms to perform failure analysis in rigid pavements. These algorithms allow the analysis of data related to pavement conditions and the types of failures observed to identify patterns and relationships that help in making decisions in pavement management, the limitations of the study were the quality and availability of data used in the failure analysis. When data are limited or unrepresentative, the results obtained may not be accurate or applicable in all cases.

Data mining is the process of discovering patterns and generating knowledge from large amounts of data [29], data sources can be databases, information repositories, or data dynamically fed into the system [30]. To understand what is extracted from models describing classes of important data [31], they are also called classifiers, one of them is decision trees, which have a flowchart-like structure in which each internal node indicates a test on an attribute, each branch manifests a test result and each leaf node contains a class label [32]. The methods used for the design of reinforced concrete beams have been little studied as an alternative to traditional methods using data mining algorithms.

The objective of the research was to apply data mining algorithms to determine the steel area in rectangular reinforced concrete beams. The need for an efficient and accurate methodology to determine the steel area in reinforced concrete beams motivates this research; since it is crucial to ensure the safety and adequate load capacity of the structures, resulting in a complex and laborious process, especially when multiple variables and conditions are considered. The contributions of the study were: the development of a predictive model, based on data mining algorithms, to estimate the steel area in reinforced concrete beams and the optimization of the design process, reducing the time and resources required while ensuring the safety and adequate load capacity of the structures. Data mining algorithms make it possible to analyze large data sets and find complex patterns that may be missed by conventional methods, save time and resources by automating and streamlining the structural design process, and improve informed decision-making in the design and evaluation of reinforced concrete beams.

3. MATERIALS AND METHODS

3.1 Data matrix

The information was collected from technical files hosted on the SEACE (Electronic State Contracting System) website, using a total of 593 reinforced concrete beam designs obtained from 2019 to 2022; forming a data matrix of 593 instances and 8 variables. The variables collected were: Ultimate moment, length, width, height, effective camber, concrete strength, steel creep, and reinforcing steel (see Figure 1).

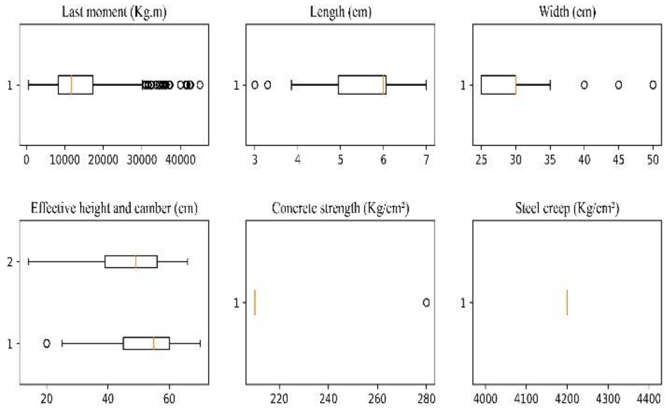


Figure 1. Descriptive statistics of the variables

The reinforcing steel present in the data matrix includes 9.5 mm, 12.7 mm, 15.9 mm, 19.1 mm, and 25.4 mm bars corresponding to bars #3, #4, #5, #6, and #8 according to the "Standard Specification for Deformed and Plain Carbon-Steel Bars for Concrete Reinforcement" (ASTM A-615), placed in one and two layers. The design with the highest percentage was 2 Ø #5 +3 Ø #6 and the lowest was 3 Ø #4, with a total of 38 different combinations (see Figure 2). The description, unit, and type of variables are shown in Table 1.

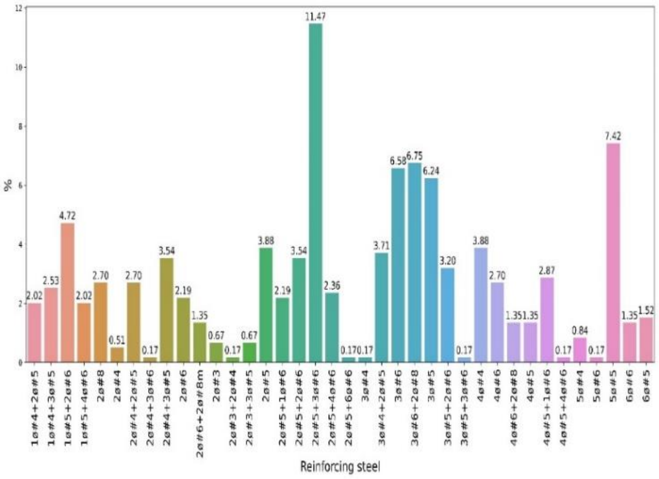


Figure 2. Types of reinforcing steel contained in the database

Table 1. Description, type, and range of variables collected

Variable	Description	Unit	Type
Last moment	It is the bending moment, generated by the application of loads on the beam.	Kgf.m	Discreet
Length	It is the free distance of the beam, measured from the axis of the supports.	m	Discreet
Width	It is the base of the beam.	cm	Discreet
Height	It is the height of the beam.	cm	Discreet
Effective camber	It is the height from the point of the fiber at the most extreme compression, to the centroid of the layer of rods.	cm	Discreet
Concrete strength	It is the capacity to support a load per unit area of concrete.	Kgf/cm^2	Discreet
Steel creep	It is the stress that corrugated steel withstands.	Kgf/cm^2	Discreet
Reinforcing steel	Is the amount and diameter of corrugated steel.	-	Nominal

3.2 Methodology for employing data mining

The Knowledge Discovery Databases (KDD) approach was used, which is characterized by automatically obtaining a deep knowledge of the rules implicit in the information [33]. This process includes data preparation, statistical analysis, the algorithm used for data mining, and the evaluation and interpretation of the data, resulting in knowledge, which is a non-trivial process of identifying valid, novel, potentially useful, and understandable patterns from the data [34].

3.2.1 Selection

The need to determine the reinforcing steel in rectangular reinforced concrete beams was identified through data mining. A data set was created on which the discovery process was performed. For this stage, the Weka software was used, using the "InfoGainAttributeEval" algorithm for the selection of variables, due to its ability to evaluate and classify the characteristics of the data set through information gain, for the assignment of weights to each variable by calculating the entropy, which measures the degree of "impurity". The closer it is to 0, the fewer impurities there are in the data set, therefore a good attribute contains the most information, reducing the entropy to the maximum [35].

After classifying the attributes, variables with entropy greater than 0.5 were selected to filter out those variables that do not have a significant influence on the determination of the steel area in reinforced concrete beams, reducing the complexity of the model in terms of relevant variables to improve efficiency. We worked with five input variables (see Table 2), which are those that provide the most information to the output variable (reinforcing steel).

Table 2. Selected variables that provide the most information

Variables	Symbol	Weights
Last moment	M	2.27
Effective camber	d	1.20
Height	h	1.20
Length	L	1.13
Width	b	0.52

3.2.2 Pre-processing

It ensures the quality and suitability of the data for classification. It includes data cleaning, integration, normalization, and transformation to improve the quality and relevance of training the classification algorithm [36]. The cleaning tasks are aimed at solving two common problems: the absence of values in some datasets and the existence of duplicate data [37]. Data cleaning was applied to improve data quality by eliminating anomalous data. This was done by replacing missing values using the average, median, and regression algorithms, as well as eliminating duplicate data using Python's Set (Data) command and anomalous data using the Panda library's boxplot, which allows visualization of the distribution and detection of outliers by identifying points outside the boundaries of the graph.

3.2.3 Transformation

Characteristics were sought to represent the data as a function of the steel to be used. Dimensionality reduction methods include principal component analysis, which transforms a set of correlated variables into uncorrelated variables called principal variables. It allows dimensionality reduction by projecting the data into a lower dimensional subspace that preserves most of the variability of the data [38], Linear Discriminant Analysis maximizes the separation between classes in a labeled data set, taking into account class information where samples are separated, is useful when the goal is to reduce dimension while maintaining discrimination between classes [39], Singular value decomposition divides a matrix into three matrices, allowing important features to be extracted and the data to be represented in a lower dimensional subspace [40], Feature selection is based on obtaining a subset of relevant features and discarding irrelevant or redundant ones, using statistical techniques such as significance or correlation tests, as well as methods based on machine learning algorithms such as genetic algorithms [41]. In this sense, the feature selection method was used to reduce variables according to the statistical analysis to maximize the separation between classes and the data set, taking into account the labels and the effective number of variables under consideration to find representations in the data, see Table 3.

Table 3. Variables that make up each group

Group	Variables
Group I (GP_1)	Last moment Width Height
Group II (GP_2)	Last moment Height Length
Group III (GP_3)	Last moment Effective camber Height
Group IV (GP_4)	Last moment Width Effective camber
Group V (GP_5)	Last moment Length Width

3.2.4 Data mining

WEKA 3.9.4 software, written in Java and developed by the University of Waikato for machine learning and data mining, was used. Through the graphical interface, the "Explorer" function and the "Classify" tool were used, using the k-folds cross-validation k=15 test recommended by Sunyani et al. [42], because it provides a better estimation in the model accuracy, reducing the variance by dividing the dataset into more folds to improve random partitions of the data in the final results.

The decision tree algorithms used were Decision Stump, which constructs a single-level binary decision tree applied to either nominal or numeric data sets. Hoeffding Tree is an incremental decision tree induction algorithm that can learn from massive data streams at any point in time, assuming that the examples used to generate the distribution do not change over time. Hoeffding trees take advantage of the fact that a small sample may be sufficient to choose an optimal partitioning attribute. J48 is an algorithm that provides the ability to stop before reaching the leaves in each subtree; this results in less refined trees and avoids overfitting. LMT is a classifier algorithm for constructing logistic model trees with logistic regression functions on the leaves. The algorithm can handle binary and multiclass target variables, numeric attributes, and nominal and missing values. Random Tree, an algorithm that constructs a tree that considers K randomly selected attributes at each node, does not perform pruning. In addition, it has the ability to perform class probability estimation based on a holdout set (back-fitting) [43]. These algorithms were used for their simplicity and ability to build decision trees that consider nominal and numeric data, allowing efficient and adaptive learning in situations where data arrive in stream form, combining logistic regression features in the leaves of the tree, being able to handle binary and multi-class target variables, as well as numeric and nominal attributes.

3.2.5 Data evaluation

After training the five models with the database, three for validation, the groups of variables were selected that showed the highest percentage of prediction during training, namely GP_1, GP_2, GP_3, GP_4, and GP_5. The collected data and the proposed data mining models were used for validation.

After determining the steel area, the different models are evaluated through the confusion matrix according to Zambrano et al. [30], which allowed us to evaluate the algorithms during prediction, see Table 4 and Eqs. (1) to (4).

Table 4. Confusion matrix

Current	Detected	
	Positive	Negative
Positive	Positive True (PV)	Negative False (NF)
Negative	Positive False (PF)	Negative True (NF)

Correctly classified instances (CCI): The proportion of correctly classified instances divided by the total number of instances.

$$\mathrm{CCI} = \frac{PV+NV}{PV+NF+NV+PF} \tag{1}$$

Incorrectly classified instances (ICI): The proportion of incorrectly classified instances divided by the total number of instances.

$$\mathrm{ICI} = \frac{PF+NF}{PV+NF+NV+PF} \tag{2}$$

PV Rate (TPV): The rate of positive cases that are correctly identified or the proportion of cases that test positive and are positive.

$$\text{TPV} = \frac{PV}{PV+NF} \tag{3}$$

Accuracy (AC): The ratio of the total number of correct positive predictions over the total number of instances classified as that class.

$$\text{AC} = \frac{PV}{PV+PF} \tag{4}$$

4. RESULTS

The data matrix consisted of 593 rectangular beam designs (reinforced concrete) and 8 variables collected from the technical files of public and private works in Peru. After the preprocessing stage, outliers were eliminated using data analysis libraries implemented in Python through exploratory analysis using descriptive statistics and box plots (see Figure 3). This made it possible to identify outliers outside the limits and eliminate them using the Pandas package, which is based on the standard deviation above the upper and lower limits of the threshold, filtering the rows of the DataFrame to keep values within the range. Thirteen designs were eliminated after the cleanup analysis.

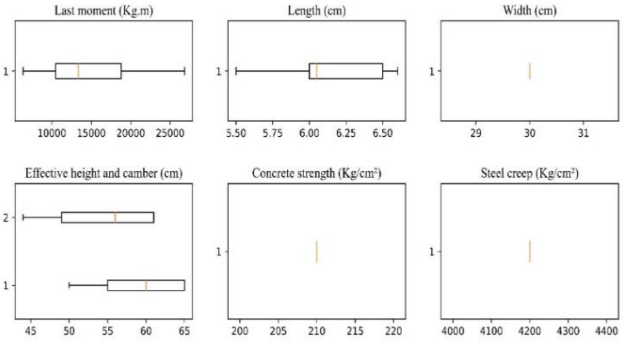


Figure 3. Database without outliers

Weka's InfoGainAttributeEval is a measure based on information theory to evaluate the importance of attributes about the target variable (reinforcing steel); it calculates the information gain provided by entropy reduction in classification. This method was chosen because it allows us to rank the attributes according to their importance, which helps to select an optimal subset of attributes for the ranking model.

The ranking of the attributes allows us to prioritize those that are considered more relevant and discard those that may have less influence on the ranking, allowing us to evaluate the value of each variable by measuring the information gained regarding the output variable (reinforcing steel). In addition, the weka.attributeSelection.The Ranker search method allowed us to rank the variables by their scores, see Table 5. The value of k-folds for the subset of training and test data was selected according to Table 6.

The results of the rebar prediction with the different trained and validated data mining algorithms are presented in Figure 4 and Table 7, where groups GP_02, GP_03, and GP_04 present the lowest error in the rebar estimation. Where the RandonTree algorithm with GP_04 presents an estimation of 88.98% because the last moment, width, and effective camber of the variable provide more information about the output variable, see Table 2. In addition, the algorithm used randomness in the selection of attributes, which reduces the correlation between the constructed trees and improves the diversity in the set of trees, for the other algorithms. For future work, it is recommended to evaluate other variables that influence beam design, such as the addition of carbon fiber, glass, and geogrids to increase beam strength at large spans. The decision tree generated by the Random Tree algorithm is shown in Figure 5.

Table 5. Ranking of variables according to the InfoGainAttributeEval

Variables	Symbol	Weights
Last moment	M	2.27
Effective camber	d	1.20
Height	h	1.20
Length	L	1.13
Width	b	0.52
Concrete strength	F'c	0.09
Steel creep	Fy	0.07

Table 6. Choice of k-folds value for training and testing

Group	Instances correctly classified (%)			
	k=10	k=15	k=20	k=30
GP_1	86.86	87.71	86.02	86.86
GP_2	87.29	88.56	87.29	87.71
GP_3	87.29	88.14	87.29	87.71
GP_4	87.29	88.98	87.71	87.71
GP_5	87.29	87.71	86.86	87.29

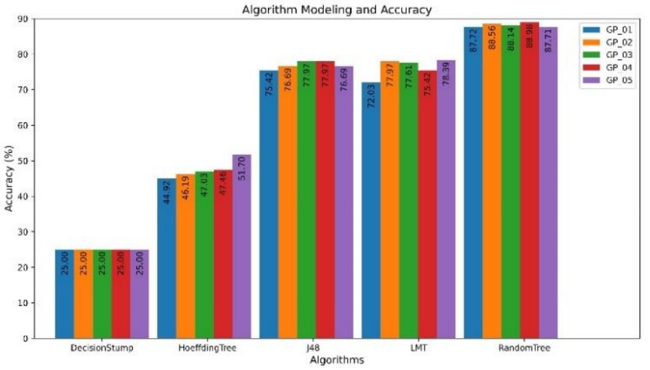
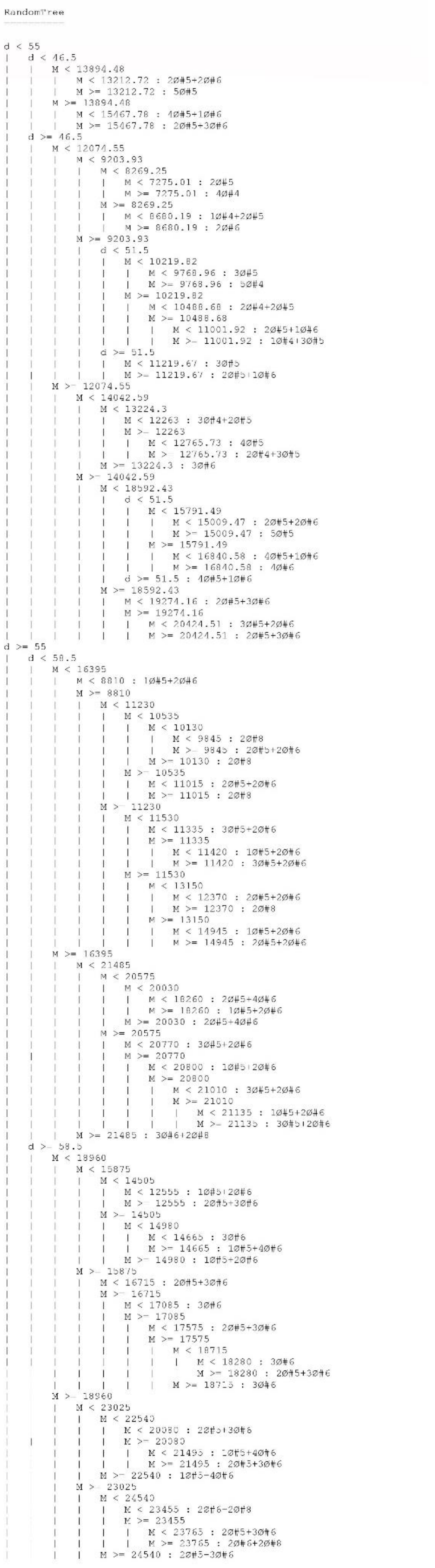


Figure 4. Accuracy of algorithms in the 5 groups of variables

Table 7. Accuracy and size of the tree according to the Random Tree algorithm

Group	Tree size	Kappa statistic	Accuracy (%)
GP_1	135	0.866	87.72
GP_2	135	0.876	88.56
GP_3	135	0.870	88.14
GP_4	135	0.880	88.98
GP_5	135	0.866	87.71



```
RandomTree
==========

d < 55
|   d < 46.5
|   |   M < 13894.48
|   |   |   M < 13212.72 : 2Ø#5+2Ø#6
|   |   |   M >= 13212.72 : 5Ø#5
|   |   M >= 13894.48
|   |   |   M < 15467.78 : 4Ø#5+1Ø#6
|   |   |   M >= 15467.78 : 2Ø#5+3Ø#6
|   d >= 46.5
|   |   M < 12074.55
|   |   |   M < 9203.93
|   |   |   |   M < 8269.25
|   |   |   |   |   M < 7275.01 : 2Ø#5
|   |   |   |   |   M >= 7275.01 : 4Ø#4
|   |   |   |   M >= 8269.25
|   |   |   |   |   M < 8680.19 : 1Ø#4+2Ø#5
|   |   |   |   |   M >= 8680.19 : 2Ø#6
|   |   |   M >= 9203.93
|   |   |   |   d < 51.5
|   |   |   |   |   M < 10219.82
|   |   |   |   |   |   M < 9768.96 : 3Ø#5
|   |   |   |   |   |   M >= 9768.96 : 5Ø#4
|   |   |   |   |   M >= 10219.82
|   |   |   |   |   |   M < 10488.68 : 2Ø#4+2Ø#5
|   |   |   |   |   |   M >= 10488.68
|   |   |   |   |   |   |   M < 11001.92 : 2Ø#5+1Ø#6
|   |   |   |   |   |   |   M >= 11001.92 : 1Ø#4+3Ø#5
|   |   |   |   d >= 51.5
|   |   |   |   |   M < 11219.67 : 3Ø#5
|   |   |   |   |   M >= 11219.67 : 2Ø#5+1Ø#6
|   |   M >= 12074.55
|   |   |   M < 14042.59
|   |   |   |   M < 13224.3
|   |   |   |   |   M < 12263 : 3Ø#4+2Ø#5
|   |   |   |   |   M >= 12263
|   |   |   |   |   |   M < 12765.73 : 4Ø#5
|   |   |   |   |   |   M >= 12765.73 : 2Ø#4+3Ø#5
|   |   |   |   M >= 13224.3 : 3Ø#6
|   |   |   M >= 14042.59
|   |   |   |   M < 18592.43
|   |   |   |   |   d < 51.5
|   |   |   |   |   |   M < 15791.49
|   |   |   |   |   |   |   M < 15009.47 : 2Ø#5+2Ø#6
|   |   |   |   |   |   |   M >= 15009.47 : 5Ø#5
|   |   |   |   |   |   M >= 15791.49
|   |   |   |   |   |   |   M < 16840.58 : 4Ø#5+1Ø#6
|   |   |   |   |   |   |   M >= 16840.58 : 4Ø#6
|   |   |   |   |   d >= 51.5 : 4Ø#5+1Ø#6
|   |   |   |   M >= 18592.43
|   |   |   |   |   M < 19274.16 : 2Ø#5+3Ø#6
|   |   |   |   |   M >= 19274.16
|   |   |   |   |   |   M < 20424.51 : 3Ø#5+2Ø#6
|   |   |   |   |   |   M >= 20424.51 : 2Ø#5+3Ø#6
d >= 55
|   d < 58.5
|   |   M < 16395
|   |   |   M < 8810 : 1Ø#5+2Ø#6
|   |   |   M >= 8810
|   |   |   |   M < 11230
|   |   |   |   |   M < 10535
|   |   |   |   |   |   M < 10130
|   |   |   |   |   |   |   M < 9845 : 2Ø#8
|   |   |   |   |   |   |   M >= 9845 : 2Ø#5+2Ø#6
|   |   |   |   |   |   M >= 10130 : 2Ø#8
|   |   |   |   |   M >= 10535
|   |   |   |   |   |   M < 11015 : 2Ø#5+2Ø#6
|   |   |   |   |   |   M >= 11015 : 2Ø#8
|   |   |   |   M >= 11230
|   |   |   |   |   M < 11530
|   |   |   |   |   |   M < 11335 : 3Ø#5+2Ø#6
|   |   |   |   |   |   M >= 11335
|   |   |   |   |   |   |   M < 11420 : 1Ø#5+2Ø#6
|   |   |   |   |   |   |   M >= 11420 : 3Ø#5+2Ø#6
|   |   |   |   |   M >= 11530
|   |   |   |   |   |   M < 13150
|   |   |   |   |   |   |   M < 12370 : 2Ø#5+2Ø#6
|   |   |   |   |   |   |   M >= 12370 : 2Ø#8
|   |   |   |   |   |   M >= 13150
|   |   |   |   |   |   |   M < 14945 : 1Ø#5+2Ø#6
|   |   |   |   |   |   |   M >= 14945 : 2Ø#5+2Ø#6
|   |   M >= 16395
|   |   |   M < 21485
|   |   |   |   M < 20575
|   |   |   |   |   M < 20030
|   |   |   |   |   |   M < 18260 : 2Ø#5+4Ø#6
|   |   |   |   |   |   M >= 18260 : 1Ø#5+2Ø#6
|   |   |   |   |   M >= 20030 : 2Ø#5+4Ø#6
|   |   |   |   M >= 20575
|   |   |   |   |   M < 20770 : 3Ø#5+2Ø#6
|   |   |   |   |   M >= 20770
|   |   |   |   |   |   M < 20800 : 1Ø#5+2Ø#6
|   |   |   |   |   |   M >= 20800
|   |   |   |   |   |   |   M < 21010 : 3Ø#5+2Ø#6
|   |   |   |   |   |   |   M >= 21010
|   |   |   |   |   |   |   |   M < 21135 : 1Ø#5+2Ø#6
|   |   |   |   |   |   |   |   M >= 21135 : 3Ø#5+2Ø#6
|   |   |   M >= 21485 : 3Ø#6+2Ø#8
|   d >= 58.5
|   |   M < 18960
|   |   |   M < 15875
|   |   |   |   M < 14505
|   |   |   |   |   M < 12555 : 1Ø#5+2Ø#6
|   |   |   |   |   M >= 12555 : 2Ø#5+3Ø#6
|   |   |   |   M >= 14505
|   |   |   |   |   M < 14980
|   |   |   |   |   |   M < 14665 : 3Ø#6
|   |   |   |   |   |   M >= 14665 : 1Ø#5+4Ø#6
|   |   |   |   |   M >= 14980 : 1Ø#5+2Ø#6
|   |   |   M >= 15875
|   |   |   |   M < 16715 : 2Ø#5+3Ø#6
|   |   |   |   M >= 16715
|   |   |   |   |   M < 17085 : 3Ø#6
|   |   |   |   |   M >= 17085
|   |   |   |   |   |   M < 17575 : 2Ø#5+3Ø#6
|   |   |   |   |   |   M >= 17575
|   |   |   |   |   |   |   M < 18715
|   |   |   |   |   |   |   |   M < 18280 : 3Ø#6
|   |   |   |   |   |   |   |   M >= 18280 : 2Ø#5+3Ø#6
|   |   |   |   |   |   |   M >= 18715 : 3Ø#6
|   |   M >= 18960
|   |   |   M < 23025
|   |   |   |   M < 22540
|   |   |   |   |   M < 20080 : 2Ø#5+3Ø#6
|   |   |   |   |   M >= 20080
|   |   |   |   |   |   M < 21495 : 1Ø#5+4Ø#6
|   |   |   |   |   |   M >= 21495 : 2Ø#5+3Ø#6
|   |   |   |   M >= 22540 : 1Ø#5-4Ø#6
|   |   |   M >= 23025
|   |   |   |   M < 24540
|   |   |   |   |   M < 23455 : 2Ø#6-2Ø#8
|   |   |   |   |   M >= 23455
|   |   |   |   |   |   M < 23765 : 2Ø#5+3Ø#6
|   |   |   |   |   |   M >= 23765 : 2Ø#6+2Ø#8
|   |   |   |   M >= 24540 : 2Ø#5-3Ø#6
```

Figure 5. Algorithm that estimates the reinforcing steel in rectangular reinforced concrete beams

5. DISCUSSION

The RamdomTree algorithm used in GP_04 was the one that best estimated the steel area in reinforced concrete beams with an accuracy of 88.98%, for its application in the design of structural elements that guarantee the structural safety of buildings. This was the result of the evaluation of the 5 groups (GP_1, GP_2, PL_3, GP_4, GP_5) conformed by 3 different variables, and trained with 5 decision tree algorithms in Weka with which an accuracy of 87.72, 88.56, 88.14, and 87.71% was obtained.

The selected algorithms were DecisionStump because it captures linear relationships or simple rules of beam geometry and strength, HoeffdingTree because of its ability to adapt to massive data streams and its efficiency in memory usage and processing time, J48 because of its ability to handle numeric and nominal attributes, perform pruning, and avoid overfitting, LMT because of its ability to handle binary and multiclass target variables as well as numeric, nominal, and missing value attributes. RandomTree was used because it randomly builds decision trees to improve accuracy and avoid overfitting, as they best represent the patterns of a data set, according to Leon Atiquipa [44], who used DecisionStump, HoeffdingTree, J48, LMT, Random Forest, Random Tree, and REPtree algorithms in his research to select the best algorithm applied to the same data set.

The results obtained in the research were superior to the results of Mataei et al. [25] of 72.44, 81.29, and 80.66% of correctly classified instances, respectively. We contributed to the existing literature by evaluating and comparing classification algorithms in the field of civil engineering and construction in the accurate classification of reinforcing steel in beams to ensure the safety and efficiency of structures through proper material selection.

6. CONCLUSIONS

Data mining algorithms were used to determine the steel area in rectangular reinforced concrete beams. For this purpose, the technical files of buildings hosted in SEACE were used. These documents contain the structural calculation memory of buildings in Peru. With this information, a data matrix consisting of 593 records of rectangular beams (reinforced concrete) and 8 variables (seven discrete and one nominal) was created for training and validation of the models.

Decision Stump, Hoeffding Tree, J48, Logistic Model Trees (LMT), and Random Tree classification algorithms were used using Weka 3.9.4 software, obtaining accuracies of 25, 51.70, 77.97, 78.39, and 88.98%. The amount of reinforcing steel in rectangular beams was determined by RandomTree and the GP_4 group, with an accuracy of 88.98%, this model is presented in Figure 5, where it is observed that the data pattern is based on the effective camber (d), which separates the predictions when d<55 obtaining more steel area and when d>55 the amount of steel decreases, because the higher the value of effective camber the beam strength increases requiring less steel. This algorithm allows engineers and structural designers to make informed and accurate decisions during the design and construction phases.

REFERENCES

[1] Castillo, R.F. (2021). Public housing policies in Peru

1946-2021 and contributions to a public housing policy 2021-2030. Paideia XXI, 11(2): 383-414. https://doi.org/10.31381/paideia.v11i2.4040

[2] Arevalo, A.S. (2020). Evaluación de la vulnerabilidad sísmica en viviendas autoconstruidas de acuerdo al Reglamento Nacional de Edificaciones en el A.H. San Jose, distrito de San Martin de Porres. Undergraduate Thesis, Universidad Peruana de Ciencias Aplicadas (UPC), Peru. https://doi.org/10.19083/tesis/648665

[3] Lin, X., Gravina, R.J. (2017). An effective numerical model for reinforced concrete beams strengthened with high-performance fiber-reinforced cementitious composites. Materials and Structures, 50(5): 1-13. https://doi.org/10.1617/s11527-017-1085-8

[4] Romero, J.Q., Ávila, T.A., Makedonski, P.M. (2005). El problema de la vivienda en el Perú, retos y perspectivas. Revista INVI, 20(53): 20-44. https://revistainvi.uchile.cl/index.php/INVI/article/view/62177.

[5] Hou, G., Li, Z., Wang, K., Hu, J. (2021). Structural deformation sensing based on distributed optical fiber monitoring technology and neural network. KSCE Journal of Civil Engineering, 25(11): 4304-4313. https://doi.org/10.1007/s12205-021-1805-z

[6] Noriega Barrueto, R. (2018). Estudio experimental de redistribución de momentos en vigas de concreto armado. Undergraduate Thesis, Pontifical Catholic University of Peru, Lima, Peru

[7] Hurukadli, P., Bharti, G., Shukla, B.K., Garg, P., Tripathi, A. (2023). Deflection of reinforced concrete beams: An analytical study on the impact of steel fibers. Materials Today Proceedings. https://doi.org/10.1016/j.matpr.2023.03.321

[8] Hanan, A.K., Muttashar, M.D., Abid, S.R., Yosri, A.M., Alsharari, F., Deifalla, A.F. (2023). Flexural performance of concrete beams internally reinforced with steel, geogrid, and GFRP meshes. Journal of Materials Research and Technology, 24: 9156-9170. https://doi.org/10.1016/j.jmrt.2023.05.146

[9] Hamoda, A.A., Eltaly, B.A., Ghalla, M., Liang, Q.Q. (2023). The behavior of reinforced concrete ring beams strengthened with sustainable materials. Engineering Structures, 290: 116374. https://doi.org/10.1016/j.engstruct.2023.116374

[10] Harmsen, T. (2005). Diseño de Estructuras de Concreto Armado. Fondo Editorial PUCP. https://drive.google.com/file/d/0B9nKI1tYMgmeR25qSUtqNzJUX0E/view?pli=1&resourcekey=0-HYWM08-rNBrVGl4VnRmQEQ.

[11] Balaji, S., Thirugnanam, G.S. (2018). Behaviour of reinforced concrete beams with SIFCON at various locations in the beam. KSCE Journal of Civil Engineering, 22(1): 161-166. https://doi.org/10.1007/s12205-017-0498-9

[12] Qeshta, I.M.I., Shafigh, P., Jumaat, M.Z. (2016). Research progress on the flexural behaviour of externally bonded RC beams. Archive of Civil and Mechanical Engineering, 16(4): 982-1003. https://doi.org/10.1016/j.acme.2016.07.002

[13] Costa, G., Cardoso, D.C.T. (2023). Nonlinear analysis of GFRP reinforced concrete beams using moment-rotation approach and conjugate beam method. Engineering Structures, 292: 116499. https://doi.org/10.1016/j.engstruct.2023.116499

[14] Mohammed, A.A., Ali, T.K.M. (2020). Flexural behavior of composite concrete–epoxy–reinforced concrete beams. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 44(2): 549-563. https://doi.org/10.1007/s40996-019-00255-1

[15] Singh, S.B. (2013). Shear response and design of RC beams strengthened using CFRP laminates. International Journal of Advanced Structural Engineering, 5(1): 16. http://dx.doi.org/10.1186/2008-6695-5-16

[16] Kagermanov, A. (2019). Análisis por elementos finitos de la rotura por cortante en vigas de hormigón armado y pretensado. Hormigón & Acero, 70(287). https://doi.org/10.1016/j.hya.2018.10.002

[17] Marí, A., Cladera, A., Bairán, J., Oller, E., Ribas, C. (2014). Shear-flexural strength mechanical model for the design and assessment of reinforced concrete beams subjected to point or distributed loads. Frontiers of Structural and Civil Engineering, 8(4): 337-353. https://doi.org/10.1007/s11709-014-0081-0

[18] George, G., Shreeram, P.K., Minalan, A.S., Lokesh, K., Mano, M., Prince, A. (2023). Numerical investigation on the flexural behavior of geopolymer concrete beam reinforced with different types of fiber-reinforced polymer bars. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2023.04.049

[19] Coello Coello, C.A., Christiansen, A.D., Hernandez, F.S. (1997). A simple genetic algorithm for the design of reinforced concrete beams. Engineering with Computers, 13(4): 185-196. https://doi.org/10.1007/BF01200046

[20] Chai, L.J., Guo, L.P., Chen, B., Sun, P.Y., Ding, C., Liu, Z.C., Wang, L.Y., Wang, Y.K. (2022). Design method of serviceability limit states of BFRP bar reinforced ecological high ductility concrete beam: Experimental and theoretical analysis. Structures, 40: 855-865. https://doi.org/10.1016/j.istruc.2022.04.065

[21] Chakrabarty, B.K. (1992). Model for optimal design of reinforced concrete beam. Journal of Structural Engineering, 118(11): 3238-3242. https://doi.org/10.1061/(ASCE)0733-9445(1992)118:11(3238)

[22] Zhang, M., Li, M., Shen, Y., Ren, Q., Zhang, J. (2019). Multiple mechanical properties prediction of hydraulic concrete in the form of combined damming by experimental data mining. Construction and Building Materials, 207: 661-671. https://doi.org/10.1016/j.conbuildmat.2019.02.169

[23] Fernández-Ceniceros, J., Martínez-de-Pisón, E., Martínez-de-Pisón, F.J., Lostado-Lorza, R., Celorrio-Barragué, L. (2010). Optimización de costes en estructuras de hormigón mediante técnicas de minería de datos. Aplicación a forjados unidireccionales.

[24] Kang, T.W., Choi, H.S. (2018). BIM-based data mining method considering data integration and function extension. KSCE Journal of Civil Engineering, 22(5): 1523-1534. https://doi.org/10.1007/s12205-017-0561-6

[25] Mataei, B., Nejad, F.M., Zakeri, H. (2021). Pavement maintenance and rehabilitation optimization based on cloud decision tree. International Journal of Pavement Research and Technology, 14(6): 740-750. https://doi.org/10.1007/s42947-020-0306-7

[26] Sun, S., Liang, L., Li, M., Li, X. (2018). Vibration-based damage detection in bridges via machine learning. KSCE Journal of Civil Engineering, 22(12): 5123-5132. https://doi.org/10.1007/s12205-018-0318-x

[27] Mansouri, I., Ozbakkaloglu, T., Kisi, O., Xie, T. (2016). Predicting behavior of FRP-confined concrete using neuro-fuzzy, neural network, multivariate adaptive regression splines, and M5 model tree techniques. Materials and Structures, 49(10): 4319-4334. https://doi.org/10.1617/s11527-015-0790-4

[28] Kang, M., Kim, M., Lee, J. H. (2010). Analysis of rigid pavement distresses on interstate highway using decision tree algorithms. KSCE Journal of Civil Engineering, 14(2): 123-130. https://doi.org/10.1007/s12205-010-0123-7

[29] Jian-hong, C., Hao-ren, R., De-ren, S., Wei, L. (2002). Data-mining massive real-time data in a power plant: Challenges, problems, and solutions. Journal of Zhejiang University - Science A, 3(5): 538-542. https://doi.org/10.1631/jzus.2002.0538

[30] Zambrano, S.J., Hidalgo Troya, A., Alvarado Pérez, J.C. (2016). Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional. Universidad Cooperativa de Colombia. https://doi.org/10.16925/9789587600490

[31] Sheikh Khozani, Z., Khosravi, K., Torabi, M., Mosavi, A., Rezaei, B., Rabczuk, T. (2020). Shear stress distribution prediction in symmetric compound channels using data mining and machine learning models. Frontiers of Structural and Civil Engineering, 14(5): 1097-1109. https://doi.org/10.1007/s11709-020-0634-3

[32] Becerra, R.M.N. (2007). Exploraciones sobre el soporte multi-agente BDI en el proceso de descubrimiento de conocimiento en bases de datos. https://www.uv.mx/personal/aguerra/files/2013/06/2007-Mondragon-Becerra.pdf.

[33] Zhang, S., Wang, L., Ou, J., Wang, G. (2007). Investigation of the structural form optimization methods of high-rise buildings. Frontiers of Architecture and Civil Engineering in China, 1(2): 182-187. https://doi.org/10.1007/s11709-007-0020-4

[34] Quiroz Gil, N.L., Valencia, C.A. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim. Revista Interamericana de Bibliotecología, 35(1): 97-108. https://doi.org/10.17533/udea.rib.13341

[35] Gerassis, S., Martín, J.E., García, J.T., Saavedra, A., Taboada, J. (2017). Bayesian decision tool for the analysis of occupational accidents in the construction of embankments. Journal of Construction Engineering and Management, 143(2): 04016093. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001225

[36] Jiao, C. (2020). Big data mining optimization algorithm based on machine learning model. Journal of Artificial Intelligence, 34(1): 51-57. https://doi.org/10.18280/ria.340107

[37] Cordero, M.P.O., Cedillo, P. (2020). Outlier detection with data mining techniques and statistical methods. Enfoque UTE, 11(1). https://doi.org/10.29019/enfoque.v11n1.584

[38] Jolliffe, I.T. (2002). Mathematical and statistical properties of sample principal components. En Principal Component Analysis, New York, NY, pp. 29-61. https://doi.org/10.1007/0-387-22440-8_3

[39] Xanthopoulos, P., Pardalos, P.M., Trafalis, T.B. (2013). Linear discriminant analysis. En Robust Data Mining, New York, NY, pp. 27-33. https://doi.org/10.1007/978-1-4419-9878-1_4

[40] Golub, G.H., Reinsch, C. (1982). Use of the singular value decomposition in regression analysis. The American Statistician, 36(1). https://doi.org/10.1080/00031305.1982.10482771

[41] Kumar, V. (2014). Feature selection: A literature review. Smart Computing Review, 4(3): 007. https://doi.org/10.6029/smartcr.2014.03.007

[42] Sunyani, Nti, I.K., Nyarko-Boateng, O., Aning, J. (2021). Performance of machine learning algorithms with different k values in k-fold crossvalidation. International Journal of Information Technology and Computer Science, 13(6), 61-71. https://doi.org/10.5815/ijitcs.2021.06.05

[43] Palomino, J. M., and Rosario, S. (2021). Estimación de la Vulnerabilidad Sísmica en Viviendas de Albañilería Confinada, Mediante Técnicas de Minería de Datos, en el Sector Pueblo Libre, Jaén – 2020. Universidad Nacional de Jaén. https://alicia.concytec.gob.pe/vufind/Record/UNJA_65443fe3339f70188102c7aab56848d6.

[44] Leon Atiquipa, H.E. (2018). Desarrollo de un modelo algorítmico basado en árboles de decisión para la predicción de la permanencia de un paciente en un proceso psicoterapéutico. Pontificia Universidad Católica del Perú. http://hdl.handle.net/20.500.12404/11868