Trends in Addiction to Psychoactive Substances Among Homeless People in Colombia Using Artificial Intelligence

Hugo Ordoñez[1]*, Ricardo Timarán-Pereira[2], Juan-Sebastián González-Sanabria[3]*

[1] Universidad del Cauca, Popayán (Colombia)
[2] Universidad de Nariño, Pasto (Colombia)
[3] Universidad Pedagógica y Tecnológica de Colombia, Tunja (Colombia)

* Corresponding author: juansebastian.gonzalez@uptc.edu.co (J. S. González-Sanabria), hugoordonez@unicauca.edu.co (H. Ordoñez).

Received 31 July 2023 | Accepted 22 November 2023 | Published 5 February 2024

Abstract

Introduction: Currently, homelessness should not be seen as just another problem, but as a reality of inequality and the absence of social justice. In this sense, homeless people are subjected to social disengagement, lack of job opportunities or the instability of these, insecurity circumstances, these aspects being one of the causes associated with the consumption or addiction to psychoactive substances. *Data*: To define the proposed approach, data from the Census of Street Inhabitants - CHC- 2021 of the National Administrative Department of Statistics (DANE), which contains 19,375 records and 25 columns, were used. *Methodology*: This article presents an artificial intelligence approach that implements a model based on machine learning algorithms for identifying addiction trends to psychoactive substances in street dwellers in Colombia. *Conclusions*: Based on the results obtained, it is evident that the approach can serve as a support for decision making by municipal administrations in the definition of social public policies for the street-dwelling population in Colombia.

Keywords

Addiction, Artificial Intelligence, Homelessness, Psychoactive Substances, Social Justice.

DOI: 10.9781/ijimai.2024.02.002

I. Introduction

In today's world, homelessness should not be seen as just another problem but as a reality of inequality and lack of social justice. In this sense, it is common to observe citizens in large cities who often transit or live permanently on the streets: children, young people, adults, older people, and even families, who, regardless of their age, sex, race, marital status, social condition, mental condition, or occupation, live there permanently or for prolonged periods, making life in this context a transitory or long-lasting option [1].

The homeless population has increased due to political, economic, and cultural factors that affect social organization. Among these, we can name displacement, armed conflict, domestic violence, and unemployment. As a particular case and focus of attention and social degradation, the consumption of psychoactive substances is identified as the most substantial and explosive factor that generates the phenomenon of homelessness [2]-[3].

Homeless people are subject to social disengagement, lack of employment opportunities or instability in this area, and circumstances of insecurity. Additionally, the experiences of loss, abandonment, and domestic violence are some causes associated with illicit psychoactive substance use or addiction [4]. The problem at the beginning of the addiction of these people is strongly related to the rupture of meaningful or affective bonds, family issues, and hostile environments. Similarly, emotional abandonment, neglect or permissiveness on the part of the caregiver or parents increases the risk of addiction [5].

Moreover, addiction arises due to the properties of illicit psychoactive substances, which, by acting more rapidly within the organism and being eliminated more swiftly, foster an increased compulsion for the individual to acquire them to sustain their consumption. The street is the most accessible place to acquire these substances [6].

Currently, as an alternative to face transcendental problems in society, technological tools have allowed to establish new state or governmental policies based on data evidence. For example, artificial intelligence tools, specifically machine learning (ML) algorithms, are being used to modernize services and assist governments in their decision-making regarding social policy issues [7]-[11].

Therefore, this paper presents an ML model for identifying trends of addictions to illicit psychoactive substances in homeless people

Please cite this article in press as:
H. Ordoñez, R. Timarán-Pereira, J. S. González-Sanabria. Trends in Addiction to Psychoactive Substances Among Homeless People in Colombia Using Artificial Intelligence, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 3, pp. 52-60, 2025, http://dx.doi.org/10.9781/ijimai.2024.02.002

in Colombia. The data of the Homeless People Census (CHC) from 2021 of the National Administrative Statistics Department (DANE), containing 19,375 records and 25 columns, was used to define the proposed model.

The model evaluation results allow identifying the trends of addictions to psychoactive substances in relation to the age of each homeless person. Moreover, the results can help in the decision-making of municipal administrations to define social public policies for the care of these people in the country.

This paper is structured into five sections. Section one describes the work related to the subject of the study. Section two presents the motivation context. Section three explains the proposed model. Section four presents the evaluation. Finally, section five develops the conclusions and future work.

II. Motivation and Related Works

Consumption, abuse, and addiction to psychoactive substances, licit or illicit, are a matter of public health. Thus, additional efforts must be made to evaluate this phenomenon in homeless people since the lifestyle of this group is related to higher drug dependence. Several studies have addressed this topic.

One study [12] addresses the control of anxiety in cocaine users undergoing outpatient addiction treatment. It applies six steps of the intervention mapping approach: needs evaluation, creation of performance objective matrices, method selection and practical strategies, program development, adoption and implementation, and evaluation, to develop the interpersonal nursing theory for anxiety in the intervention of persons with illicit psychoactive substance use disorders.

Similarly, other studies [13]-[14] address some differences between women who are mothers and those who do not have children in a sample of women living on the street in Madrid, Spain. The information was collected through a structured interview. The results evidence that the women living in the street were mothers who had experienced traumatic situations from an early age and had higher levels of illicit psychoactive substance abuse. Furthermore, they had issues with the judiciary system, which may have negatively impacted the lives of their children and their relationships.

Likewise, another research [15] aimed at identifying the structure of the social representations homeless people have of persons in the same situation who consume drugs. This was based on the social representations theory and addressed 158 homeless people in the historic center of Salvador, Bahía, Brazil. The data were collected through free word association using the inductor homeless people who consume drugs. The data were analyzed with two software programs. The results identified that the participants were mainly young Afro-descendant men who had finished primary school. It was concluded that these people have their lives at risk, are excluded, and need help.

Another study [16] focused on analyzing the knowledge and experiences with new psychoactive substances among users in the homeless population. The participants were selected from support charitable organizations in the United Kingdom through convenience sampling. Descriptive and logistic regression statistics were applied to analyze the obtained data. The results showed that the street dwellers consumed illicit psychoactive substances to escape reality and self-medicate, and they stopped consuming due to the adverse side effects. These effects were reported by most of the participants and caused over 20% of them to require medical treatment.

In the same sense, some studies [17]-[18] examined aspects such as sociodemographic characteristics, access to economic resources, social support, addiction chronicity and access to new technologies of the homeless people in León, Nicaragua. A questionnaire was used to collect the data. The results showed that homeless people have social difficulties and high chronicity levels. Despite the major cultural and developmental differences between Spain and Nicaragua, there are significant similarities among the homeless people in both countries.

There have also been studies in Colombia regarding homeless people. One of them [19] described their health situation, concluding that one in five people has a health issue, such as dental problems, respiratory problems, abdominal pain or injuries caused by third parties. The most common chronic diseases are hypertension and diabetes.

Similarly, another research [20] described that homelessness affects health and makes these people vulnerable. The study was based on a program called Mobile Care Center for Drug-Dependent People (CAMAD). Its objective was to interpret the experiences of a group of homeless people in the Rafael Uribe area in Bogotá, Colombia, focusing on health.

In conclusion, several studies and research on homeless people worldwide address issues related to their addiction to illicit psychoactive substances and their health. Although these studies have had outstanding results, none have taken advantage of the potential of current technologies, specifically artificial intelligence, to analyze the data collected. Therefore, the ML model presented in this paper analyzes socioeconomic, health, economic activity, gender, and types of addiction variables to predict how these addictions may be related to the age of street dwellers in Colombia.

In Colombia and most of the world, the homeless person is subjected to social disengagement, job instability, and precarious conditions. The pressure of other homeless people or friends, the emotional suffering, anxiety, depression, and the environmental stress experienced on a daily basis when living on the street can be factors that increase the risk of drug use [21].

In Colombia, there are 22,790 homeless people; 12% are women, and 88% are men [22]. Fig. 1 shows the number of homeless people per department. The departments of Valle del Cauca and Antioquia are noteworthy because their capitals, Cali and Medellín, respectively, have a multicultural population. Additionally, they receive many migrants from the center of the country due to their increased industrialization and development.

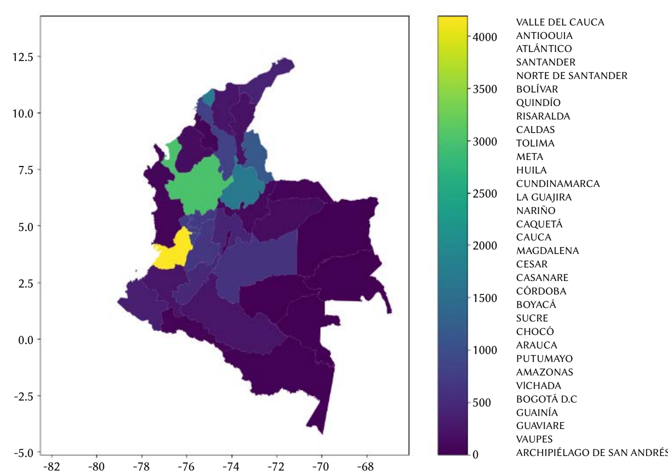

Fig. 1. Population of homeless people per department in Colombia.

Illicit psychoactive substances contain natural or synthetic compounds that act on the nervous system generating alterations in the functions that regulate thoughts, emotions, and behavior. These substances are not freely distributed and are punishable by law [23].

Some of them are:

- Marijuana or cannabis: It directly affects brain function, particularly the brain parts responsible for memory, learning, attention, decision-making, coordination, emotions, reaction time, relaxation and euphoria, anxiety, fear, distrust, and panic.
- Basuco: It is a toxic substance whose main risks when consumed are related to neurological and physical deterioration. It destroys brain tissue and causes irreversible memory loss; the effects are immediate: the skin turns yellow, the lips get dried, the tongue gets numb, the pupils dilate, and the body trembles. Its dissolution in the bloodstream is swift, which makes it very addictive.
- Cocaine: It affects the nervous system and the rest of the body immediately. These affectations include vasoconstriction, mydriasis, hyperthermia, tachycardia, and hypertension. Additionally, the effects derived from the euphoria, mainly during the first 30 minutes, are hyperstimulation, the sensation of less tiredness, and a state of greater mental alertness.
- Heroin: It is highly addictive, and its effects are very pleasant, which causes a continuous and repetitive consumption behavior. Heroin gets to the brain rapidly and adheres to the opioid receptors of the cells, especially those associated with pain and pleasure and those controlling the heart rate, sleep, and breathing. It produces dryness in the mouth, reddening and heating of the skin, heaviness in arms and legs, nausea and vomiting, intense itching and clouding of the mental faculties.

Psychoactive substance consumption is a problem affecting society in general, regardless of age, culture or social status. The consumption pattern of these substances in homeless people depends mainly on the age and type of substance. Fig. 2 presents the number of homeless people using each drug. It can be noted that the age range of basuco consumption is between 16 and 64 years, with a concentration in people between 25 and 50 years. Basuco is the most consumed substance among homeless people. Its characteristics increase addiction because its effects appear and vanish rapidly, making the person feel a greater need for consumption [24].

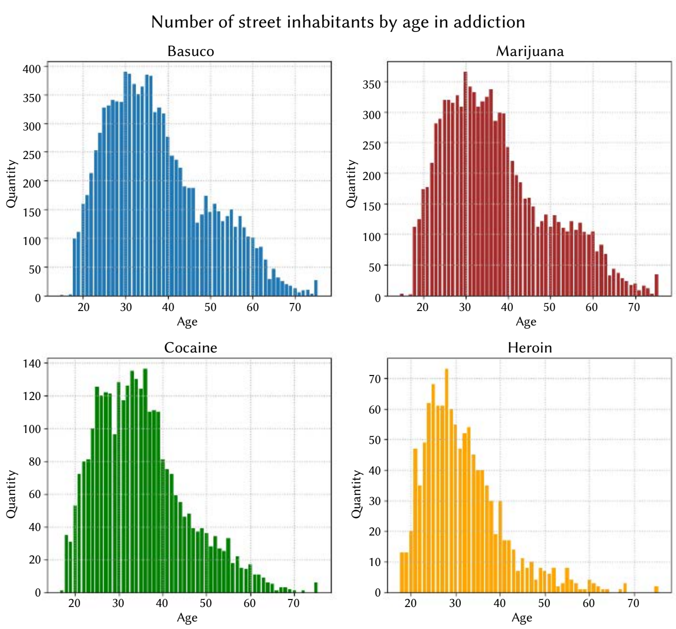

Fig. 2. Number of consumers per age according to the type of psychoactive substance.

Marijuana is the second most consumed substance by homeless people due to its effects: relaxation, drowsiness, a sensation of slowness in the passage of time, disinhibition, excessive joy, and eye reddening. The consumption age range for this drug goes from 19 to 60 years. Basuco and marijuana are the cheapest and easiest substances to acquire on the street [25].

Cocaine and heroin are consumed by people in similar age ranges (18 and 60 years old), but fewer people are addicted to these substances due to their price, the difficulty of acquiring them, and the symptoms they generate, such as restlessness, irritability, and anxiety. They can also cause tremors, dizziness, muscle spasms or paranoia, and serious medical complications if consumed excessively [26].

Addiction to illicit psychoactive substances becomes a disease that disturbs the brain and behavior of the person who consumes them and results in an inability to control consumption. Then, despite the damage they cause, it is customary to continue consuming them. Addiction to multiple substances is common among homeless people. Furthermore, as consumption increases, it is increasingly difficult to live without them [27].

Fig. 3 shows addictions to multiple illicit psychoactive substances in homeless people according to their age. Most notably, the substances of greatest consumption are marijuana and basuco, which means that the person who is addicted to marijuana is also addicted to basuco, and this preference oscillates among people between the ages of 18 and 60 years old.

Likewise, and more concerning, a smaller number of people are addicted to marijuana and cocaine. The age range of this group is 19 to 55 years old, and it is possibly related to the higher cost of cocaine compared to that of marijuana and basuco.

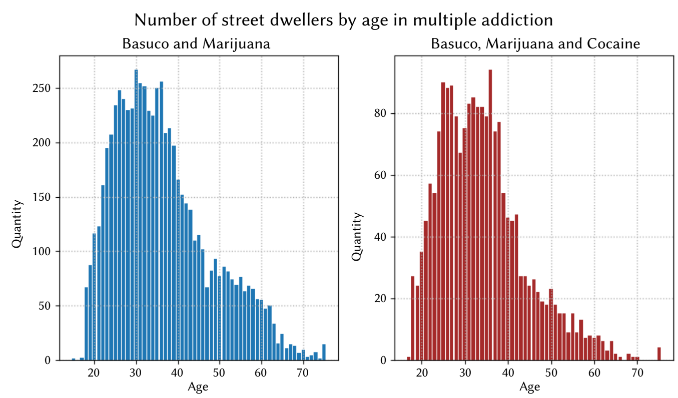

Fig. 3. Number of homeless people with multiple addictions.

Another relevant characteristic of homeless people is their source of income. As seen in Fig. 4, one of the primary sources of income is recycling, which is very common in major cities. Similarly, cleaning car windows and car guarding are ways of acquiring money and being self-employed for these people. Begging is also a widespread activity. Although not forbidden, it is considered a social problem directly related to inequality and poverty. It has to be noted that not all homeless people resort to theft to make ends meet, i.e., not all of them are criminals, only a small percentage resort to this activity.

People living on the street are exposed to malnutrition, have health problems, and are at risk of getting sick, dying, or experiencing diverse and continuous violent aggressions, which affect their health, due to the environment where they live [28]. Fig. 5 presents the most common aggressions suffered by homeless people. As it can be seen, the greatest fear is losing their life violently, followed by beatings (to which they are exposed daily), and assaults with cold weapons, which are very common in this context.

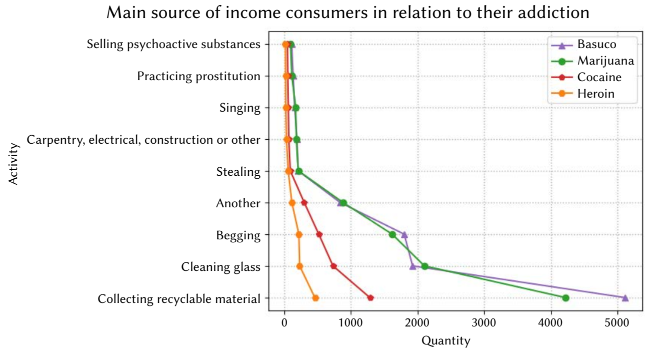

Fig. 4. Source of income of homeless people according to their addiction.

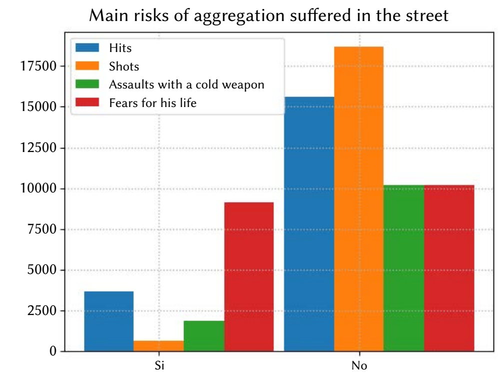

Fig. 5. Risks and fears of homeless people.

A high percentage of homeless people have addictions or are direct consumers of illicit psychoactive substances, and they are exposed to high health-related risks. Many of them are frequent victims of several crimes, and some of them commit crimes on a regular basis. The homeless person who consumes drugs lacks a source of employment, is concerned about having a permanent source of drug supply, uses drugs to "remedy" negative feelings, withdraws from friends and family, may trade friends for people who are regular users, and experiences increased tolerance and ability to process the drug. Thus, using technology, such as artificial intelligence, is relevant to support governmental entities in decision-making processes. Particularly when discussing programs that implement social strategies, such as the Comprehensive Policy for the Prevention and Care of Psychoactive Substance Use [29] and the Social Public Policy for Street People 2021-2032 [30].

III. Proposed Model

The predictive model for addiction to illicit psychoactive substances of the homeless people in Colombia is represented in Fig. 6. It applies a series of machine learning algorithms that integrate specific characteristics of a training dataset to make predictions and find trends. It is divided into three parts: data and pre-processing, machine learning models, and evaluation.

A. Data and Pre-processing

The data of the CHC from 2021 of the DANE containing 19,375 records and 25 columns were used to define the proposed model. The dataset related the information with the volume and the main socioeconomic and demographic characteristics of the homeless people in Colombia. It was built with comma-separated values (CSV) files containing information of the CHC from several years. A relational database was built based on these files, assimilating all the information. Subsequently, a minable view was generated through SQL statements to run the model. Table I presents the variables and their description.

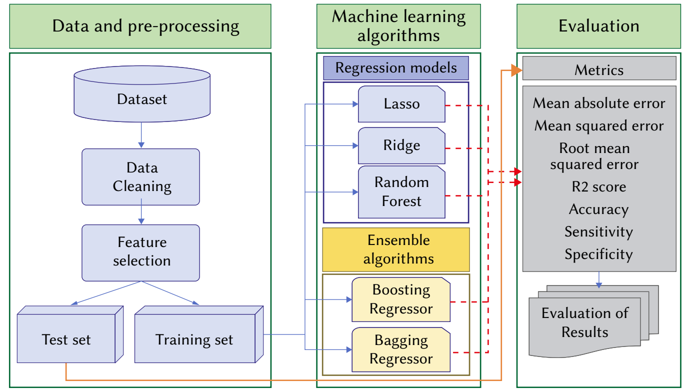

Fig. 6. Prediction model of addiction to illicit psychoactive substances based on artificial intelligence.

TABLE I. Variables of the Dataset

Variable	Description
Age	Age at the moment of the census
Gender	Gender of the homeless person
place_where_they_sleep	The place where the person usually sleeps or stays at night
hypertension	If the person has health problems related to hypertension
diabetes	If the person has health problems related to diabetes
Cancer	If the person has health problems related to cancer
tuberculosis	If the person has health problems related to tuberculosis
hiv_aids	If the person has health problems related to AIDS
reason_for_living_on_the_streets	What is the main reason for living on the streets
time_living_on_the_streets	How long the person has been living on the streets
reason_for_staying_on_the_streets	Reason for staying on the streets, what forces the person to stay on the streets
source_of_income	What is the person's source of income, how do they access money to live
Tobacco	Uses or smokes tobacco
Alcohol	Uses or is addicted to alcohol
marijuana	Uses or is addicted to marijuana
inhalants	Uses or is addicted to inhalants
Cocaine	Uses or is addicted to cocaine
Basuco	Uses or is addicted to basuco
Heroin	Uses or is addicted to heroin
Pills	Uses or is addicted to pills
other_drugs	Other drugs that the person uses frequently
fear_for_their_life	Fears for their life while living on the streets
beating_victim	The person has been a victim of beatings while living on the streets
gunshot_victim	The person has been a victim of gunshots while living on the streets
cold_weapon_victim	The person has been a victim of cold weapon assaults while living on the streets

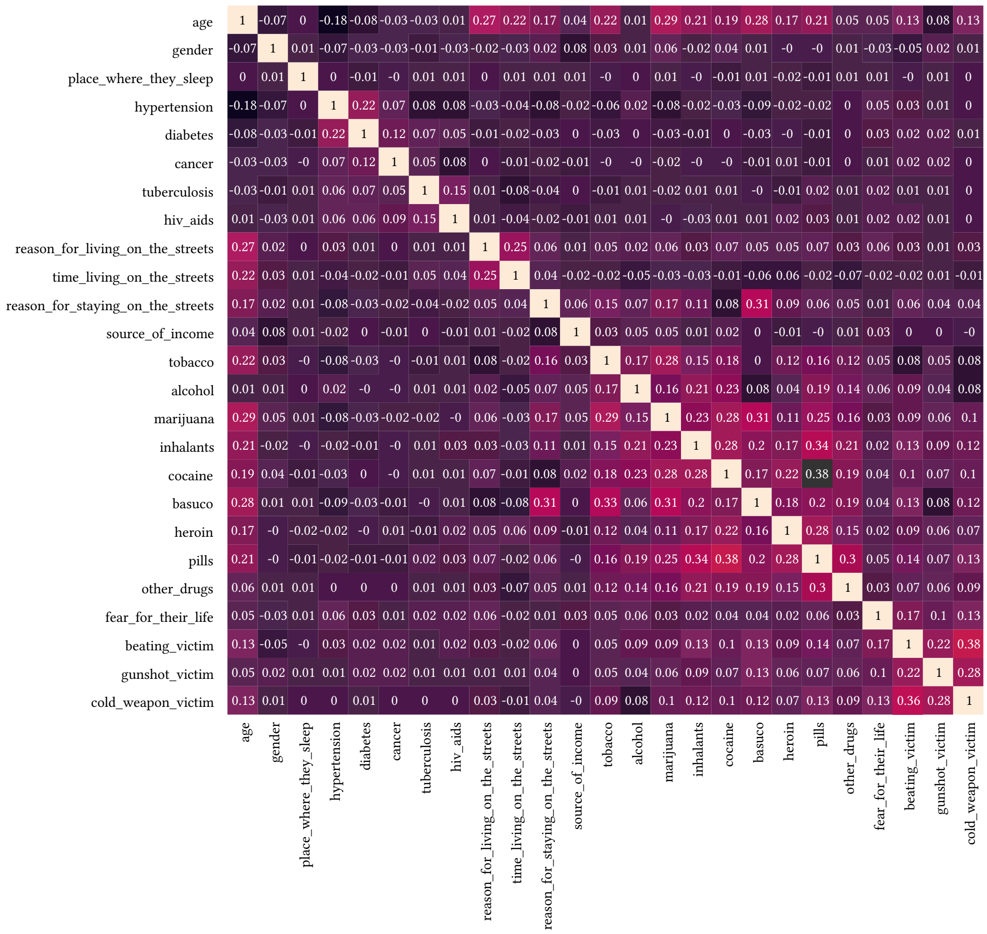

Fig. 7. Correlation (corr) among the variables of the study.

1. Data Cleaning

An exploratory analysis was carried out to understand the data. Columns or variables that did not contribute to the solution were removed, duplicate data were eliminated, and missing values were completed by data ingestion with the average between the previous and the next value in each column. Then, the records that still contained null values were eliminated. Finally, in the case of regression, the original data were normalized with the min-max method, transforming the values into a range between zero and one. The distributions of the variables and the patterns they presented were analyzed, and the relationships among the variables were identified.

2. Feature Selection

This process was carried out to achieve greater interoperability, reducing the computational cost during training and prediction, and avoid overtraining. Thus, some variables that did not contribute to the study were deleted, such as mobility, decision autonomy, personal interactions, among others. The Pearson correlation was used to quantify the linear dependence among the variables, given that the dataset includes continuous data.

Fig. 7 shows the ten variables that have the highest correlation with the study variables, showing that homeless people who use basuco, marijuana, cocaine, and heroin have a close relationship with the use of other substances, which is directly associated with their age and the reason or motive for living on the street. Regarding the test set and the training test, these were divided into training and tests, setting 70% for the first and 30% for the second.

B. Machine Learning Models

For all machine learning algorithms, hyperparameter optimization was performed with the scikit-learn library, using RandomizedSearchCV, since it is possible to obtain results as accurate as those obtained with GridSearchCV, although with a significant reduction in time, due to the sampling of the hyperparameters in the defined distribution. Cross-validation (RepeatedKFold) was also used to improve the estimated performance of each model and avoid overtraining. The data were randomly divided into subsets and optimized with the loss function: mean squared error.

- Random forest regressor: We experimented with different values for the number of trees, the number of features to consider in each split, the maximum number of levels in the tree, the minimum number of samples required to split a node, the minimum number of samples required at each leaf node, and the sample selection method to train each tree [31].
- Lasso regressor: The sum of the absolute values of the penalty weights was analyzed for the parameter n_samples, which is the number of observations to analyze the performance of the regressor [32]. Ridge regressor: The alpha was evaluated with a value of one, equivalent to an ordinary least square, and the tolerance for optimization with small values. The seed of the pseudorandom number generator was random to generate a random coefficient at each iteration. A positive value of one was used for the alpha hyperparameter to increase the variance of the estimations.

In addition, the maximum number of 15,000 iterations was defined for the conjugate gradient solver. Regarding the solver parameter, the solver was used automatically depending on the type of data [33].

TABLE II. Evaluation Metrics

Metric	Equation	Description
MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)^2$ (1)	It measures how close the points of the predictions made are to the regression line, according to the risk corresponding to the expected value of the squared error loss. The closer to zero, the better the performance of the model.
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-x_i)^2}$ (2)	It measures the difference between the actual values and those estimated by the models. The closer to zero, the better the performance of the model.
R2	$1-\frac{\Sigma(y_i-x_i)^2}{\Sigma(x_i-\underline{x_i})^2}$ (3)	It establishes how well the actual data approximate the regression line. The closer to one, the better the performance of the model.
MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-y_i\rvert$ (4)	It calculates the errors between the actual values and those predicted by the model. The closer to zero, the better the performance of the model.
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	It measures the fraction of predictions that the model made correctly. The closer to one, the better the performance of the model.
Sensitivity	$\frac{TP}{TP+FN}$	It measures the ability of a model to detect positive cases, i.e., the predictions actually made. A high value means the model correctly identifies most of the positive results.
Specificity	$\frac{TN}{TN+FP}$	It measures the proportion of true negatives identified correctly by the model. A high value means that the model correctly identifies most of the negative results.

1. Ensemble of Machine Learning Algorithms

In this study, an ML model ensemble strategy that significantly improves the predictions made by each one separately was implemented to optimize the efficiency of the model since it combines the predictions of multiple algorithms.

The ML ensembles can be classified into three types: 1) Average ensemble, which averages the results of several models to obtain a more accurate one. 2) Voting ensemble, in which the most accurate result is obtained by counting the votes of the individual models. 3) Blending ensemble, which combines the results of the individual models to obtain a more accurate one.

In turn, the three main classes of dataset learning methods are boosting, bagging, and stacking [34]. The first two were implemented in this model.

- Boosting regressor: It is formed by a set of individual decision trees trained sequentially where each new tree tries to improve the errors of the previous trees. The prediction of a new observation is obtained by adding the predictions of all the individual trees of the model. It has the advantages of automatically selecting predictors, can be applied to regression and classification problems, handles both numerical and categorical predictors without creating dummy variables, and is not significantly influenced by outliers [35].
- Bagging regressor: It allows for better predictive performance compared to one model. The goal is to learn from a set of predictors (experts) and allow them to vote. It decreases the variance of one estimation since it combines several estimations from different models. Thus, the result can be a more stable model. Bagging is a homogeneous model of weak listeners that learn from each other independently in parallel and are combined to determine the average [36].

C. Evaluation

To evaluate the performance of the proposed model, the metrics used were mean squared error (MSE), mean absolute error (MAE), root mean square error (RMSE), coefficient of determination (R2), accuracy, sensitivity, and specificity. Table II presents the equations, the description, and the performance criterion of each evaluation metric.

TP is a result where the model correctly predicts the positive class, and TN is where the model correctly predicts the negative class. FP is a result where the model incorrectly predicts the positive class, and FN is where the model incorrectly predicts the negative class. These data were extracted from the confusion matrix.

IV. Evaluation and Results

TP is a result where the model correctly predicts the positive class, and TN is where the model correctly predicts the negative class. FP is a result where the model incorrectly predicts the positive class, and FN is where the model incorrectly.

A. Evaluation of the Regression Models

After analyzing the relationships among the variables to be predicted, the RMSE metric was used to evaluate which of the regression algorithms (random forest regression, Lasso regression, and ridge regression) presented the best results in order to select the ensemble method.

Table III presents the results of each regression algorithm. It can be seen that random forest regression provided the best results for each of the analyzed variables. The RMSE values achieved by this model are the lowest, i.e., they tend more to zero, which means that the difference between the actual and the predicted values is low. Therefore, the predictions are closely related to the actual data.

TABLE III. Regression Algorithms Evaluation RMSE Metric

Algorithm	Basuco	Marijuana	Cocaine	Heroin
Random forest regressor	**0.3898**	**0.4255**	**0.3780**	**0.2566**
Lasso regressor	0.4228	0.4582	0.4184	0.2686
Ridge regressor	0.4239	0.4584	0.4184	0.2697

Once the base algorithm for the ensemble was selected, all the regressors were evaluated. Thus, the boosting and bagging regressors were added to the evaluation. The results are presented in Table IV. It can be seen that the boosting regressor had the best results, achieving an improvement of the variables in comparison to the random forest regressor: 0.0260 basuco, 0.0154 marijuana, 0.0168 cocaine, and 0.0150 heroin.

Moreover, the results of the variables in comparison to the bagging regressor were: 0.0113 basuco, 0.01234 marijuana, 0.01073 cocaine, and 0.01325 heroin. This means that the sequential ensemble made by the boosting regressor, in which each new tree tries to improve the errors of the previous ones, allows to have optimal prediction results since, according to the metric, they tend more and more to zero, with low difference between the actual values and the predicted ones, unlike the predictions made by the random forest regressor that only considers one solution for the regression. Similarly, the boosting regressor sequential ensemble achieved better results than the bagging regressor parallel ensemble.

TABLE IV. Regressor Evaluation

Algorithm	Basuco	Marijuana	Cocaine	Heroin
Random forest regressor	0.3898	0.4255	0.3780	0.2566
Lasso regressor	0.4228	0.4582	0.4184	0.2686
Ridge regressor	0.4239	0.4584	0.4184	0.2697
Boosting regressor	**0.3638**	**0.4100**	**0.3612**	**0.2417**
Bagging regressor	0.3751	0.4224	0.3719	0.2549

B. Evaluation of the Ensemble Model

Once it was identified that the boosting regressor achieved the best results, the ensemble was evaluated with the remaining set of metrics for each of the correlated variables. Table V presents the results obtained by the proposed ensemble model. Thus, concerning the MAE metric, the best results were obtained for the heroin variable, which means that the range of error in the prediction is low since the value is very close to zero. Likewise, for the rest of the variables it could be noted that the values obtained tended to zero, demonstrating that the prediction errors of the proposed model are low.

For the MSE metric, the results obtained show that the predictions made are close to the regression line, i.e., they are closely related since all the values obtained with this metric are very close to zero.

Regarding the R2 metric, the values obtained tend to one, demonstrating that in the predictions made by the proposed model, the actual data approximate well to the regression line, with the highest point value being 0.8631 for the cocaine variable and the lowest value being 0.7674 for the basuco variable.

In the accuracy metric, the obtained values are very close to one, indicating that the fraction of predictions that the model made correctly is high, which demonstrates the excellent performance of the model. The sensitivity metric showed that the highest values were obtained for the basuco and marijuana variables, i.e., the model detected a greater proportion of positive cases for these two variables, as opposed to the cocaine and heroin variables, where the value of the metric was lower.

Finally, for the specificity metric, contrary to the previous ones, the highest values were presented in the cocaine and heroin variables since the model correctly identified a higher percentage of true negatives.

TABLE V. Results of the Evaluation of the Ensemble Model Proposed

Metric	Basuco	Marijuana	Cocaine	Heroin
MAE	0.0306	0.0247	0.0203	0.0123
MSE	0.0025	0.0013	0.0013	0.0006
R2	0.7674	0.8142	0.8631	0.7948
Accuracy	0.8193	0.7326	0.8349	0.9275
Sensitivity	0.9554	0.9194	0.3936	01966
Specificity	0.3785	0.3331	0.9664	0.99545
Matrix Confusion [[TP, FP] [FN TN]]	[[9999 467] [[2008 1223]]	[[8812 772] [[2743 1370]]	[[1237 1906] [[355 10199]]	[[229 936] [[57 12475]]

After analyzing the performance of the model, the results of the predicted consumption of illicit psychoactive substances, according to the age of the homeless people, were plotted. Regarding the basuco variable, Fig. 8 shows that the highest use of this substance begins, on average, at 18 years old, having a high concentration between 22 and 60 years old. According to the predictions, it can also be observed that the older the person is, the lower the consumption of this substance. A possible explanation is that an older homeless person may face greater difficulties in obtaining the money to sustain this addiction.

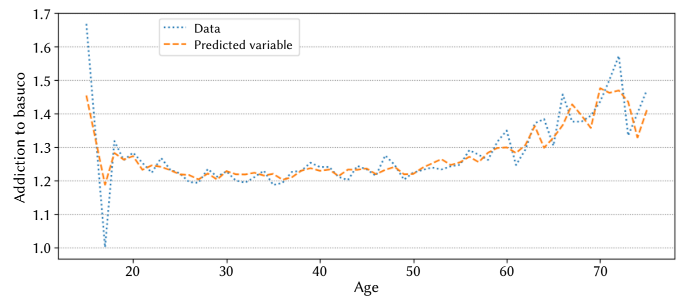

Fig. 8. Tendency to basuco addiction among homeless people according to their age.

Similarly, Fig. 9 presents the predictions made by the model regarding marijuana. It shows that most homeless people addicted to this substance are between 18 and 30 years of age, indicating that it has a greater preference among young people. Although marijuana use is not criminalized in Colombia, the number of homeless people over 40 who choose this drug is low. The direct relationship of the predictions of the proposed model with the reality of the data is remarkable.

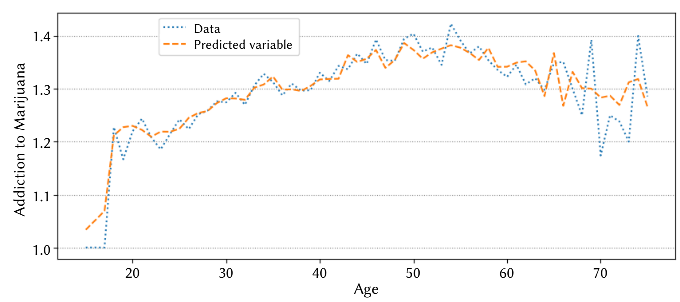

Fig. 9. Tendency to marijuana addiction among homeless people according to their age.

Fig. 10 presents the tendency to cocaine addiction of homeless people in Colombia. Most of the addicts are between 18 and 30 years of age. Its consumption decreases with the increasing age of homeless people due to its high cost, the difficulty of obtaining it on the street, and the fact that its trafficking is heavily penalized.

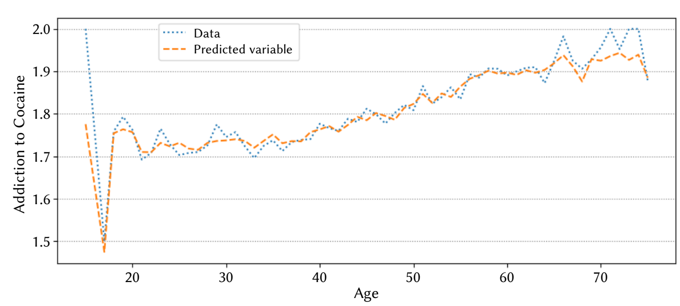

Fig. 10. Tendency to cocaine addiction among homeless people according to their age.

Regarding heroin, Figure 11 presents the prediction tendencies. This substance is similar to cocaine in that the highest concentration of addicts is found between the ages of 14 and 30 and decreases for older homeless people, for the same reasons as cocaine.

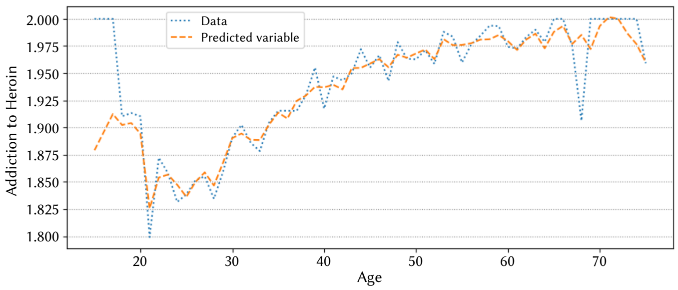

Fig. 11. Tendency to heroin addiction among homeless people according to their age.

The results of the evaluation of the applied metrics and the figures presented demonstrate that the predictions of the proposed model are close to the reality of the data.

V. Conclusions and Future Work

In this paper, an ML algorithm ensemble method was used. It was composed of three predictor algorithms and two ensemble methods. A dataset of the CHC from 2021 of the DANE containing 19,375 records and 25 columns was used to define the method. The Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology was applied to develop the research and process the data, which made it possible to understand the data set and conceptualize the data domain.

The research process allows us to conclude that not all homeless people are criminals, as many believe. Similarly, it was identified that the main sources of income of these people are recycling, cleaning car windows, and begging, which are directly related to inequality and poverty.

The study also revealed that one of the main reasons these people live on the streets is a heavy dependence on illicit psychoactive substances, with basuco being the primary substance used, followed by marijuana and cocaine. Moreover, a high number of homeless people have multiple addictions, i.e., they are addicted to more than one illicit psychoactive substance at the same time; they use regularly basuco, marijuana, and cocaine. Due to these addictions, another significant characteristic of their daily life is suffering frequent assaults, such as beatings and cold weapons injuries.

Through the evaluation of the ensemble method, it was possible to identify that boosting regressor improves performance because the individual decision trees are trained sequentially. This reduces the errors of the predecessor trees, and the prediction is made automatically based on the forecasts made by all the individual trees that make up the model.

The results show that age is closely related to addiction to certain types of illicit psychoactive substances. For example, the use of basuco is intensely concentrated in the population of homeless people between 20 and 60 years of age, that is, between youth and maturity; marijuana is consumed between 18 and 30 years of age, which indicates that younger people strongly prefer it. Although cocaine and heroin addiction occur to some extent among homeless people, it is not as prevalent because of the high cost of these substances and the difficulty of obtaining them on the street.

Finally, it can be concluded that the proposed model can serve as a decision-making tool for governmental institutions in formulating, managing and evaluating policies, plans and programs of local and municipal administrations regarding the comprehensive care, rehabilitation, and social inclusion of homeless people in Colombia.

Future work is expected to test the method with other datasets, such as access to education or home ownership in Colombia and data found in the national open data system (ANDA). In addition, the proposed method could be complemented with other ensemble techniques, e.g., stacking, which presents a general procedure for assembling base models.

References

[1] P. C. Rosa, "Exclusiones del espacio público de los habitantes de la calle en la ciudad de Buenos Aires," *Territorios*, no. 39, p. 157, 2018, doi: 10.12804/revistas.urosario.edu.co/territorios/a.5632.

[2] P. Ruisoto and I. Contador, "The role of stress in drug addiction. An integrative review," *Physiology and Behavior*, vol. 202, pp. 62-68, 2019, doi: 10.1016/j.physbeh.2019.01.022.

[3] G. F. Koob and J. Schulkin, "Addiction and stress: An allostatic view," *Neuroscience and Biobehavioral Reviews*, vol. 106, pp. 245-262, 2019, doi: 10.1016/j.neubiorev.2018.09.008.

[4] E. Kpelly, S. Schauder, J. Masson, C. K. Kokou-Kpolou, and C. Moukouta, "Influence of attachment and psychotrauma in drug addiction," *Annales Médico-psychologiques*, vol. 180, no. 6, pp. S81-S87, 2022, doi: 10.1016/j.amp.2020.11.019.

[5] I. Sadło, E. Guz, A. Wójciuk, M. Brodowicz-Król, M. Kaczoruk, and P. Kaczor-Szkodny, "Addiction to psychoactive substances as a public health challenge," *Medycyna Ogólna i Nauki o Zdrowiu*, vol. 27, no. 1, pp. 70-76, 2021, doi: 10.26444/monz/133713.

[6] A. A. Moustafa et al., "The relationship between childhood trauma, early-life stress, and alcohol and drug use, abuse, and addiction: An integrative review," *Current Psychology*, vol. 40, pp. 579-784, 2021, doi: 10.1007/s12144-018-9973-9.

[7] C. Rudin and K. L. Wagstaff, "Machine learning for science and society," *Machine Learning*, vol. 95, pp. 1-9, 2014. doi: 10.1007/s10994-013-5425-9.

[8] C. Alexopoulos, V. Diamantopoulou, Z. Lachana, Y. Charalabidis, A. Androutsopoulou, and M. A. Loutsaris, "How machine learning is changing e-government," in *ACM International Conference Proceeding Series*, 2019, pp. 354–363, 2019, doi: 10.1145/3326365.3326412.

[9] P. Cadahia, A. Golpe, J. M. Martín-Álvarez, and E. Asensio, "Measuring anomalies in cigarette sales using official data from Spanish provinces: Are the anomalies detected by the Empty Pack Surveys (EPSs) used by Transnational Tobacco Companies (TTCs) the only anomalies?," *Tobacco Induced Diseases*, vol. 19, e98, doi: 10.18332/tid/143321.

[10] A. Andueza, M. Á. Del Arco-Osuna, B. Fornés, R. González-Crespo, and J. M. Martín-Álvarez, "Using the Statistical Machine Learning Models ARIMA and SARIMA to Measure the Impact of Covid-19 on Official Provincial Sales of Cigarettes in Spain," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 1, pp. 73-87, 2023, doi: 10.9781/ijimai.2023.02.010.

[11] A. Suruliandi, T. Idhaya, and S. P. Raja, "Drug Target Interaction Prediction Using Machine Learning Techniques – A Review," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 6, pp. 86-100, 2024, doi: 10.9781/ijimai.2022.11.002.

[12] C. F. Pereira, D. de Vargas, and L. S. Beeber, "An anxiety management intervention for people with substance use disorders (ITASUD): An intervention mapping approach based on Peplau's theory," *Frontiers in Public Health*, vol. 11, e1124295, 2023, doi: 10.3389/fpubh.2023.1124295.

[13] J. Ochieng, "Prevalence of Psychoactive Substance Use and Associated Behavioral Risks among Secondary School Students in Tanzania," *East African Journal of Education and Social Sciences*, vol. 3, no. 4, pp. 185–196, 2022, doi: 10.4314/eajess.v3i4.211.

[14] J. J. Vázquez, S. Panadero, and I. Pascual, "The Particularly Vulnerable Situation of Women Living Homeless in Madrid (Spain)," *The Spanish Journal of Psychology*, pp. 1-9, 2019, doi: 10.1017/sjp.2019.58.

[15] L. C. M. Campos, J. F. de Oliveira, C. Porcino, M. J. de O. U. Reale, M. V. S. Santos, and M. E. F. de Jesus, "Social Representations Held By Homeless Individuals Regarding Homeless Individuals Who Consume Drugs," *Revista Baiana de Enfermagem*, vol. 33, pp. 1–9, 2019, doi: 10.18471/rbe.

v33.26778.

[16] T. Coombs, T. Ginige, P. Van Calster, A. Abdelkader, O. Corazza, and S. Assi, "New Psychoactive Substances in the Homeless Population: A Cross-Sectional Study in the United Kingdom," *International Journal of Mental Health and Addiction*, e0123456789, 2023, doi: 10.1007/s11469-022-00988-7.

[17] J. J. Vázquez, A. E. Berríos, and A. C. Suarez, "Health, disability, and consumption of psychoactive substances among people in a homeless situation in León (Nicaragua)," *Social Work in Health Care*, vol. 59, no. 9-10, pp. 694-708, 2020, doi: 10.1080/00981389.2020.1835785.

[18] J. J. Vázquez, A. Suarez, A. Berríos, and S. Panadero, "Characteristics and needs of people living homeless in León (Nicaragua): Similarities and differences with other groups in severe social exclusion," *International Social Work*, vol. 65, no. 2, pp. 328–342, 2022, doi: 10.1177/0020872819896820.

[19] E. A. Salazar et al., "Inhabitants of the street in Colombia, some elements of your health," *Palarch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 8, pp. 3470–3476, 2021.

[20] S. Farigua, J. Pedraza, and R. Ruiz, "Experiencias de habitantes de calle que asisten al Programa de Salud Camad Rafael Uribe Uribe en Bogotá," *Revista Ciencias de la Salud*, vol. 16, no. 3, pp. 429–446, 2018.

[21] R. C. Fiorati, R. Y. D. Carretta, L. M. Kebbe, B. L. Cardoso, and J. J. D. S. Xavier, "Social ruptures and the everyday life of homeless people: an ethnographic study," Revista Gaucha de Enfermagem, vol. 37, e72861, 2017.

[22] J. C. Cubillos Álzate, M. Cárdenas, and S. Perea, *Boletines Poblacionales: Personas Adultas Mayores de 60 años Oficina de Promoción Social Ministerio de Salud y Protección Social*, Ministerio de Salud, Bogotá D. C., 2020. Available: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/boletines-poblacionales-envejecimiento.pdf [Accessed June 10, 2023]

[23] H. D. Whitehead et al., "Validated method for the analysis of 22 illicit drugs and their metabolites via liquid chromatography tandem mass spectrometry (LC-MS/MS) in illicit drug samples collected in Chicago, IL," *Forensic Chemistry*, vol. 33, e100475, 2023, doi: 10.1016/j.forc.2023.100475.

[24] D. Yajaira, B. Fernández, Á. Segura-Cardona, L. Montoya-Velez, and M. Hernández-Rendón, "Consumo de basuco en usuarios de drogas inyectables en Colombia," *Revista Cubana de Salud Pública*, vol. 42, no. 2, pp. 276-283, 2016.

[25] A. H. Sadaka et al., "Effects of inhaled cannabis high in Δ9-THC or CBD on the aging brain: A translational MRI and behavioral study," *Frontiers in Aging Neuroscience*, vol. 15, pp. 1-20, 2023, doi: 10.3389/fnagi.2023.1055433.

[26] S. E. Koch, J. A. Marckel, J. Rubinstein, and A. B. Norman, "A humanized anti-cocaine mAb antagonizes the cardiovascular effects of cocaine in rats," *Pharmacology Research and Perspectives*, vol. 11, no. 1, pp. 1–8, 2023, doi: 10.1002/prp2.1045.

[27] F. Gosetti, et al, "From the Streets to the Judicial Evidence: Determination of Traditional Illicit Substances in Drug Seizures by a Rapid and Sensitive UHPLC-MS/MS-Based Platform," *Molecules*, vol. 28, no. 1, e164, 2022, doi: 10.3390/molecules28010164.

[28] P. J. Cooper et al., "Understanding and controlling asthma in Latin America: A review of recent research informed by the SCAALA programme," *Clinical and Translational Allergy*, vol. 13, no. 3, e12232, 2023, doi: 10.1002/clt2.12232.

[29] Ministerio de Salud y Protección Social, *Política Integral para la Prevención y Atención del Consumo de Sustancias Psicoactivas*, p. 44, 2019. Available: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/politica-prevencion-atencion-spa.pdf [Accessed June 11, 2023]

[30] G. Gesti, *Política Pública Social para Habitantes de la Calle 2021-2031*, p. 231, 2021. https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PS/abece-habitantes-calle-2022-2031.pdf [Accessed June 11, 2023]

[31] S. Kwak et al., "Machine learning prediction of the mechanical properties of γ-TiAl alloys produced using random forest regression model," *Journal of Materials Research and Technology*, vol. 18, pp. 520-530, 2022, doi: 10.1016/j.jmrt.2022.02.108.

[32] P. Maranzano, P. Otto, and A. Fassò, "Adaptive LASSO estimation for functional hidden dynamic geostatistical model," *Stochastic Environmental Research and Risk Assessment*, vol. 37, pp. 3615-3637, 2022, doi: 10.1007/s00477-023-02466-5.

[33] S. Mohammadi, "A test of harmful multicollinearity: A generalized ridge regression approach," *Communications in Statistics - Theory and Methods*, vol. 51, no. 3, pp. 724-743, 2022, doi: 10.1080/03610926.2020.1754855.

[34] B. Das et al., "Comparison of bagging, boosting and stacking algorithms for surface soil moisture mapping using optical-thermal-microwave remote sensing synergies," *Catena*, vol. 217, e106485, 2022, doi: 10.1016/j.catena.2022.106485.

[35] M. Sipper and J. H. Moore, "AddGBoost: A gradient boosting-style algorithm based on strong learners," *Machine Learning with Applications*, vol. 7, e100243, 2022, doi: 10.1016/j.mlwa.2021.100243.

[36] P. W. Khan, S. J. Park, S. J. Lee, and Y. C. Byun, "Electric Kickboard Demand Prediction in Spatiotemporal Dimension Using Clustering-Aided Bagging Regressor," *Electric Vehicles: Planning and Operations*, vol. 2022, e8062932, 2022, doi: 10.1155/2022/8062932

Hugo Ordoñez

Full-time professor at the University of Cauca. PhD in Telematics Engineering Universidad del Cauca 2015, Master in Computing Universidad del Cauca 2011. Area of Interest: Business process discovery, information management.

Silvio Ricardo Timarán-Pereira

Full-time professor at the University of Nariño. Systems Engineer and Master Of Science in Engineering from Donetsk Polytechnic University.

Juan-Sebastián González-Sanabria

Systems and Computing Engineer, from UPTC Tunja, and has two specializations: one in Databases at UPTC and another from the National University of La Plata in Scientific and Technological Information Management. In addition, he has a Master's Degree from the International University of La Rioja in Software and Information Systems. In his teaching activity, he has worked in the subjects of Databases and Research. He has more than a dozen papers and scientific articles in different journals, mainly in the area of data analysis and development.