A model based on the gradient boosting regressor to predict trends in the ratio of residence in relation to the age of homeless people in Colombia

Un modelo basado gradient boosting regressor para predecir tendencias de razón de residencia en relación a la edad de los habitantes de la calle en Colombia

Hugo Armando Ordóñez Erazo
Universidad del Cauca, Colombia

Víctor Andres Bucheli Guerrero
Universidad del valle, Colombia

Cristian Camilo Ordoñez Quintero
Fundación Universitaria de Popayan , Colombia

Recibido: 02/02/2023
Aceptado: 20/04/2023
Publicado: 18/05/2023

Correspondencia de autores:
hugoordonez@unicauca.edu.co

Abstract

Objective: To generate a machine learning model that is capable of predicting trends in the reasons that lead to being a street dweller during adolescent and adult ages. **Methodology:** The model is trained with information of the Census of Street Dwellers - CHC- 2021 of the National Administrative Department of Statistics (DANE), which contains 19.375 records and 22 variables. The population of street dwellers is identified to include children, young people, adults, the elderly, and even families, who, regardless of their age, gender, race, marital status, social or mental condition, or occupation, live there depending on a reason that somehow forces them to stay permanently or for periods of time. Then, different models were applied until the results described below were obtained. **Results:** The paper presents an ensemble model of machine learning algorithms based on Gradient Boosting Regressor to predict trends in the reason for street dwelling in relation to the age of street dwellers in Colombia. The results obtained in the model evaluation are promising, providing validity for the model to serve as a basis for government institutions for the formulation, management, and evaluation of policies, plans, and programs of the municipal administration in relation to street dwellers. **Conclusions:** It can be concluded that the proposed model can serve as a basis to support decision-making by government institutions for public policies, and programs of the municipal and local administration regarding the comprehensive care, rehabilitation, and social inclusion of street dwellers in Colombia..

Keywords: Street dwellers, Ensemble methods, Machine learning, Dataset, Evaluation metrics.

Resumen

Objetivo: Generar un modelo de inteligencia machine learning que sea capaz de predecir tendencias de las razones que conllevan a estar en estado de habitante de la calle en edades adolescentes y adultas. **Metodología:** Se identifica la población de los habitantes de la calle se encuentran niños, jóvenes, adultos, ancianos e incluso familias, que, sin importar su edad, sexo, raza, estado civil, condición social, mental u oficio, viven allí dependiendo de una razón que de alguna forma los obliga a estar de forma permanente o por periodos de tiempo, después se aplicaron diferentes modelos hasta llegar a obtener los resultados descritos a continuación. **Resultados:** Los resultados obtenidos en la evaluación del modelo son prometedores, dando validez para que el modelo sirva de base a instituciones gubernamentales para la formulación, gestión y evaluación de las políticas, planes y programas de la administración municipal en relación a los habitantes de calle. **Conclusiones:** Se puede concluir que el modelo propuesto puede servir de base como soporte a la toma de decisiones a instituciones gubernamentales para la formulación, gestión y evaluación de las políticas, planes y programas de la administración municipal y de las localidades, respecto a la atención integral, rehabilitación e inclusión social de los habitantes de la calle en Colombia.

Palabras clave: Habitantes de la calle, Métodos de ensamble, machine learning, dataset, métricas de evaluación.

Como citar (IEEE): H. Ordóñez-Erazo., C. Ordoñez-Quintero., y V. Bucheli-Guerrero. “Un modelo basado gradient boosting regressor para predecir tendencias de razón de residencia en relación a la edad de los habitantes de la calle en Colombia". *Investigación e Innovación en Ingenierías,* vol. 11, n°1, 145-157, 2023. DOI: https://doi.org/10.17081/invinno.11.1.6572

Introduction

According to [1], the concept of the homeless or vagrants, comes from the ambition of European expansion and the colonization process, where the marginalized people were recruited and sent as the first wave of conquest, regardless of the risk to their lives. These people took advantage of the donations from charitable Christians. Nowadays, homelessness arises as a response to economic development programs, which leave a percentage of the population in marginality, exclusion, and absolute poverty [2].

Currently, the world's homeless population is composed of children, young people, adults, the elderly, and even families, who, regardless of their age, gender, race, marital status, social status, mental condition, or occupation, live there permanently or for extended periods, making life on the street [3]. In Colombia, the homeless population has increased due to political, economic, and cultural factors that cross social organization. These factors include displacement, armed conflict, violence within families, unemployment, and an increase in the consumption of psychoactive substances, which make this situation critical [4, 5].

In relation to the increase in homelessness, a series of works have been carried out to help understand and identify the social problems that this phenomenon entails. Thus, [4] hypothesizes through a census that the entry into street living in Bogotá occurs at any age, and when this happens, the stage of despair with which these people's deadly cycle begins is anticipated. On the other hand, [1] analyzes the homeless phenomenon in Medellín, where references to concept, history, and intervention strategies have been implemented by different municipal administrations. They propose projects to attack the phenomenon through the components of the Adult Street Resident Care System. Likewise, [6] conducted a study based on surveys to capture perceptions regarding health and education aspects of 90 homeless people in the city of Santiago de Cali. The study allowed analyzing and demonstrating the advances made by the mayor's office of Santiago de Cali in terms of a social public policy for homeless people, which has the support of the Ministry of Social Protection, who carries out a comprehensive action for these people to provide them with a dignified life and guarantee their needs.

Furthermore, all the works carried out focus on analyzing information through censuses, or surveys, and although the contributions are significant. However, according with the literature review works related to machine learning were not found. In this paper we present a model of machine learning, which is a branch of artificial intelligence [7]. Similar models are developed in other contexts to support decision-making, such as theft trends in Colombia [8], analysis of kidnappings [9], predicting terrorist attacks [10], among others. In this sense, this article defines an ensemble model of machine learning algorithms based on Gradient Boosting Regressor to predict trends in the reason for street residency in relation to the age of homeless people in Colombia. To define the model, data from the Street Residents Census - CHC- 2021 of the National Administrative Department of Statistics (DANE), which contains 19.375 records and 22 variables. The results allow identifying the reasons for street residency in relation to the age of each homeless person, and this output can serve as support for municipal administrations to define social policies for homeless people in Colombia.

The article is organized as follows: section two presents the motivation scenario, section three presents the proposed model, section four presents the evolution and results obtained, and finally, section five presents the conclusions and future work.

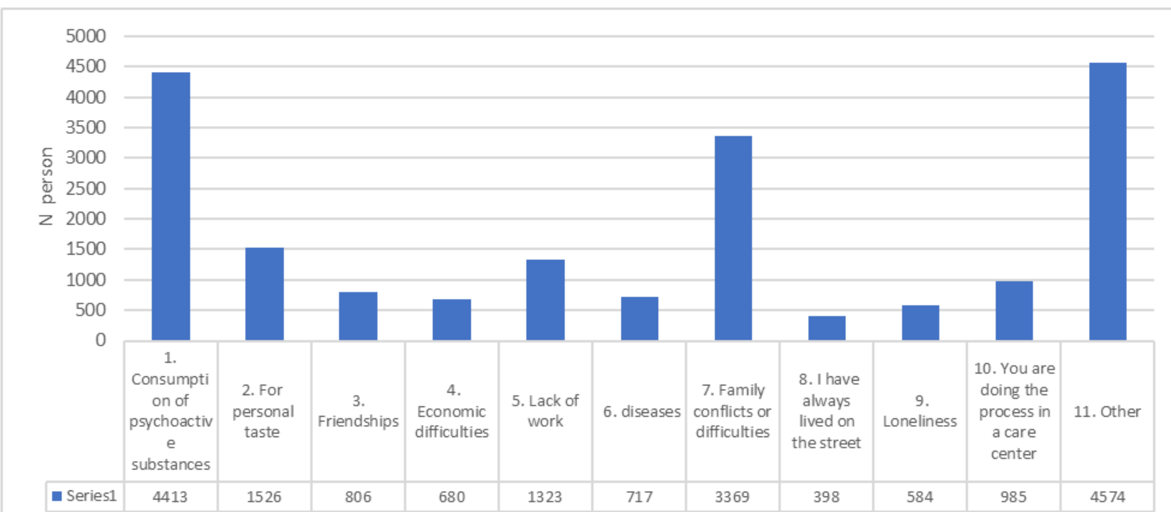
Metodology

Motivation scenario

In Colombia, street life has critical impacts, resulting in factors such as substance abuse, constant sexual danger, and delinquent activities [11]. The condition of being a street dweller and homelessness is linked to a process of community and family disconnection, meaning a distancing from these traditional and formally established aspects of society. According to [3], "For many of these people, the street is the place where they can stay, finding satisfying relationships and, in some way, finding the opportunity for an affective community, but also facing the attacks of chance and marginalization, poverty and exclusion; mistreatment and pain; displacement, loneliness and orphanhood, crime and unproductive leisure, psychoactive drugs, glue, begging and scavenging. From this perspective, the street is significant as a space for survival."

Figure 1. shows the reasons why street dwellers decide to reside there. One of the most important reasons is the consumption of psychoactive substances. Thus, the Colombian government formed the Drug Observatory of Colombia (ODC), It integrates information from all sectors responsible for carrying out actions related to the drug problem, to facilitate the formulation and adjustment of policies, plans, and intervention strategies [12]. Another important reason is family problems or conflicts, which can be due to abuse (domestic violence), separation or divorce of parents, arrival of a step-parent, death of a close relative, among others.

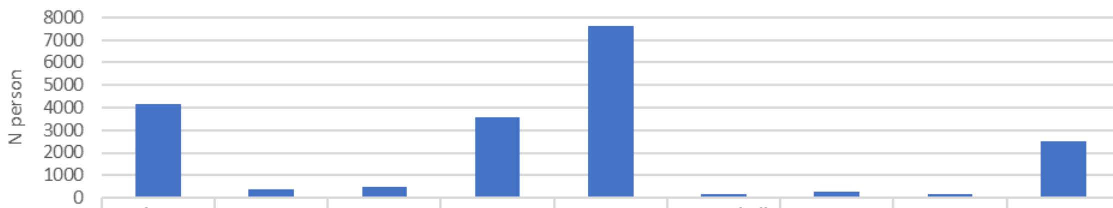
Figure 1. Reasons for residing on the street



Source: own elaboration

Another important aspect of street dwellers is the source of their incomes, as shown in Figure 2. One of the main sources of income is recycling, which is very notable in large cities. Similarly, activities such as cleaning windshields and watching over cars become sources of income and employment for these individuals. Begging is also a strong and widely practiced source of income, although not prohibited, it is considered a social problem directly related to inequality and poverty. In fact, not all street dwellers use theft or robbery to survive, only a small percentage of them have this activity as income.

Figure 2. Sources of incomes

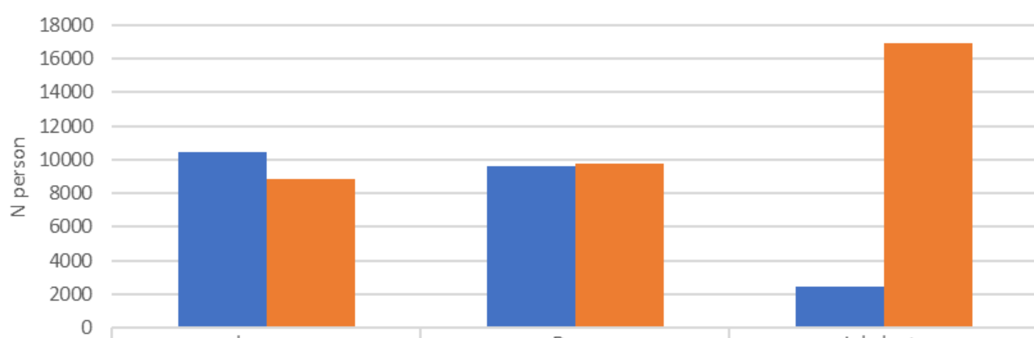


	1. Cleaning windows, taking care of cars, touching tires, selling on the...	2. Singing, juggling, counting crafts or other similar	3. Carpentry, electricity, constructio n or other similar	4. Asking, begging, begging	5. Collecting recyclable material	6. As a bell ringer, ticket seller, selling or transporting psychoactiv e...	7. Stealing or mugging	8. Practicing prostitution	9. Other
Series1	4186	357	501	3582	7612	164	263	183	2522

Source: own elaboration

As shown in Figure 1, the consumption of psychoactive substances is the main reason for street dwelling. In this sense, Figure 3 shows that over 50% of street dwellers consume bazuko. According to [13], most street dwellers are polydrug users, but the most problematic substance for patients is bazuko, as those who consume it become increasingly addicted, making it difficult to leave the streets. Similarly, marijuana is consumed, and it is prone to being penalized for the crime of trafficking, manufacturing, or carrying, as the consumption of a personal dose. It is decriminalized and carrying it is not considered a typical conduct within the criminal framework. Finally, inhalants are consumed by a low percentage, where people between the ages of 8 and 20 are found.

Figure 3. Type of drugs consumed



	bazuco	Dope	inhalants
yes	10487	9605	2472
No	8883	9765	16898

Source: own elaboration

Based on the Figures presented in the lines above, it is the data used to train machine learning models. The machine learning algorithms contribute to the development of prediction methods or trend detection of behaviors. It was applied to the street dwellers in Colombia, which can serve as a basis for the design of state programs and policies to strengthen. In the same way that was proposed in [14], increase the capacities and potential of street dwellers to achieve well-being within the framework of the life project.

Proposed model

The model assembles machine learning algorithms to make predictions about the reason for homelessness based on age. The algorithms used are Random Forest, Lasso regressor, and Ridge regressor, specifically oriented towards regression problems, which are assembled by Gradient Boosting Regressor and Bagging Regressor (see Figure 4).

Dataset: The dataset was obtained from the DANE open data portal, whose information is related to the volume and main socio-economic and demographic characteristics of homeless people in Colombia. The data was downloaded in CSV format since it is generated from a relational database. A relational database was also designed for data integration, since the data comes from several years and DANE portals. Finally, datamart was generated through SQL statements to train the model. Table 1. presents the variables and their description.

Tabla 1. Dataset and variables description

Variable	Description
age	Age at the time of the census
gender	gender of the street dweller
place_where_he sleeps	Place where you usually sleep or reside at night
hypertension	If you suffer from health problems related to hypertension
diabetes	If you suffer from health problems related to diabetes
cancer	If you suffer from health problems related to cancer
tuberculosis	If you suffer from health problems related to tuberculosis
HIV AIDS	If you suffer from health problems related to AIDS
reason_to live_on_street	What is the main reason for living on the street?
time_living_on_street	How long have you been living on the street?
reason_permanence_on_street	Reason for staying on the street, which is what forces you to live on the street
source_of_income	What is your source of income, how do you get the money to live.
cigarette	Consume or smoke cigarette
alcohol	Consume or are addicted to alcohol
dope	Consume or are addicted to marijuana
inhalants	Consume or are addicted to inhalants
cocaine	Use or are addicted to cocaine
basuco	Consume or are addicted to bazuco
heroin	Use or have a heroin addiction
Kernels	Consumes or is addicted to pepas
other_drugs	What other drugs do you usually take?
fear_for_his_life	He feels fear for his life living on the street
beating_victim	Has been the victim of beatings while on the street
shooting_victim	He has been the victim of shots during his stay on the street
knife_arm_victim	Has been the victim of a stab wound while on the street

Source: own elaboration

Data preparation: This process was carried out following the CRISP-DM methodology [15] and began with the phase of exploratory data analysis to obtain an understanding of the data. In this process, the variables that did not contribute to the solution were removed, duplicate data was eliminated, missing values were completed, and records still containing null values were deleted. Finally, in the case of regression, the original data was normalized with the Min-Max method, transforming the values into a range between (0-1). The distributions of the variables and the patterns were analyzed, and how the variables are related was identified. Regression models were developed in Python (programming language) using the scikit-learn, pandas, numpy, and matplotlib libraries. In the data preparation phase for analysis, the RobustScaler method was used to normalize the data and avoid outliers affecting the algorithm's results. Additionally, the data was split into training and testing sets, with 30% for testing and 70% for training.

The Machine Learning algorithms

The model combines machine learning algorithms, including Random Forest, Lasso regression, and Ridge regressor, to predict the reason for homelessness based on age. The assembled algorithms are focused on regression problems and are brought together by Gradient Boosting Regressor and Bagging Regressor. The hyperparameter tuning was performed using the scikit-learn library, using RandomizedSearchCV, as it is possible to achieve results as precise as those obtained with GridSearchCV, but with a significant reduction on time, due to the sampling of hyperparameters in the defined distribution. RepeatedKFold cross-validation was also used to improve the estimated performance of each model and avoid overfitting. The data was randomly divided into subsets and optimized with the mean squared error loss function.

- Random Forest Regressor, different values were experimented with in the number of trees, the number of features to consider at 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 method of sample selection to train each tree.
- Lasso Regressor, the sum of the absolute values of the weights for the penalty parameter n_samples was analyzed, which is the number of observations to analyze the performance of the regressor.
- Ridge Regressor, Alpha was evaluated with a value of 1 equivalent to ordinary least squares, and the tolerance for optimization with small values. The seed of the pseudorandom number generator was taken as 'random' to generate a random coefficient at each iteration. To increase the variance of the estimates, a positive value of 1 was used for the Alpha hyperparameter. Moreover, the maximum number of iterations for the conjugate gradient solver was defined as 15000. The solver parameter was set to automatically use the solver based on the data type.

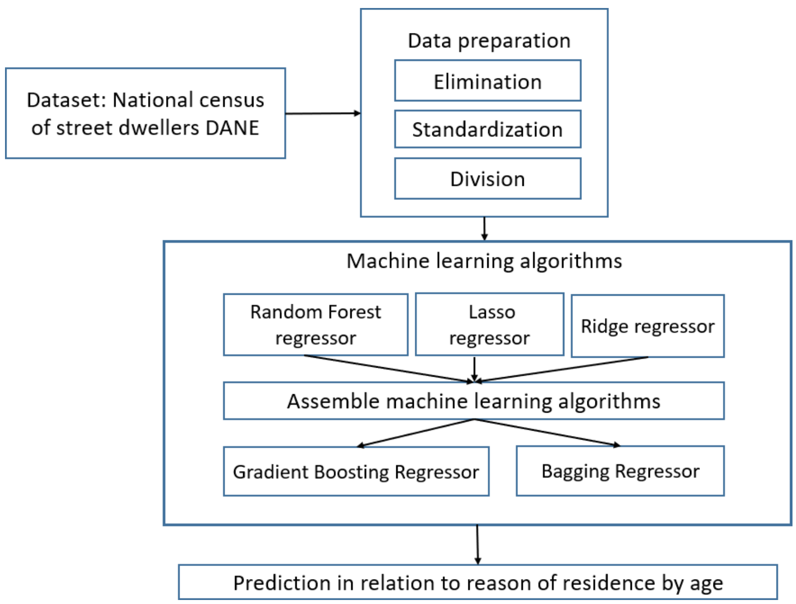
***Ensemble Machine Learning algorithms*.**

This work proposes an ensemble learning model that seeks better predictive performance by combining the predictions of multiple algorithms. Although there are several algorithm combinations that can be developed for predictive modeling, there are three methods that dominate the field of ensemble learning. In fact, each one is a field of study that has generated many more specialized methods. The three main classes of ensemble learning methods are bagging, stacking, and boosting. For the present model, boosting and bagging were implemented.

Boosting: it consists of 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 aggregating the predictions of all individual trees and it makes up the model. This method has the advantages of automatically selecting predictors, can be applied to regression and classification problems, handles numerical predictors as well as categorical predictors without having to create dummy variables, and is not highly influenced by outliers.

Bagging: it allows the creation of better predictive performance compared to a single model. The basic idea is to learn from a set of predictors (experts) and allow them to vote. Bagging decreases the variance of a single estimate, as it combines several estimates from different models. The result can be a model with higher stability. Bagging is a homogeneous model of weak listeners that learn independently in parallel and combine to determine the model's average.

Figure 4. Proposed ensemble learning model



Source: own elaboration

Results

The model evaluation process was carried out according to CRISDM. The evaluation of the predictor algorithms is performed by the identification of the best machine learning algorithm, which was used as the base for the ensemble. The root mean squared error (RMSE) (equation 1.) was used to measure the standard deviation of the prediction errors.

To evaluate the algorithms, a Pearson correlation analysis was performed, Thus, the variables were analyzed and identified the strength of the relationship between the dependent variables, length of homelessness and age. Table 3 shows a strong (increasing) relationship between the length of homelessness and age, as well as strong correlations between addiction to psychoactive substances. Then, substance addiction is a significant factor for homelessness. Furthermore, according to [16, 17], the majority of homeless individuals are exposed to physical and sexual abuse, and substance.

For the evaluation of machine learning, default hyperparameters were defined, except for ccp_alpha for random forest regressor and alpha for Lasso regressor and Ridge regressor, with values of (0.1, 0.3, and 0.5) due to their similarity and applicability in the algorithms. The ccp_alpha is a complexity parameter used for pruning that measures the cost of minimum complexity. This parameter chooses the subtree with the smallest ccp_alpha value.

The evaluation metrics are presented in Table 2. It shows that the random forest regressor achieves the best performance for all defined hyperparameter values. In sum, 8 and 9% lower RMSE values for the 0.1 hyperparameter value, and stable values for 0.3 and 0.5. Thus, it indicates that the decision trees implemented by the random forest regressor are much closer to the ideal or real values of the predicted variables, resulting in a lower mean squared error.

Table 2. Evaluation metrics

Algorithm	RSME	RSME	RSME
Random Forest Regresor	ccp_alpha = 0.5	ccp_alpha = 0.3	ccp_alpha = 0.1
	6.97	6.97	6,85
Lasso Regresor	Alpha =0.5	Alpha =0.3	Alpha =0.1
	7.74	7.74	7,76
Ridge Regresor	Alpha =0.5	Alpha =0.3	Alpha =0.1
	7.77	7,77	7,77

Source: own elaboration

Table 3. Correlation of variables

Reason for staying on the street	
Reason for staying on the street	1.000000
bazooko	0.305033
dope	0.174580
age	0.173700
cigarette	0.150682
inhalants	0.110148
heroin	0.094807
cocaine	0.076637
alcohol	0.068716
Kernels	0.062504
source of income	0.060399
other drugs	0.051753
beating victim	0.046844
reason for residing on the street	0.046722
time living on the street	0.039877
shooting victim	0.039086
knife victim	0.038293
gender	0.020854
place where he sleeps	0.006627
fear for his life	0.006530
Vih	-0.015049
Cencer	-0.015209
Diabetes	-0.033799
tuberculosis	-0.035920
hypertension	-0.076484

Source: own elaboration

Ensemble Model Evaluation

The ensemble with Boosting and Bagging were performed taking into account the Random Forest Regressor algorithm with better performance. In addition to RMSE, the metrics presented in Table 4 were used for this process.

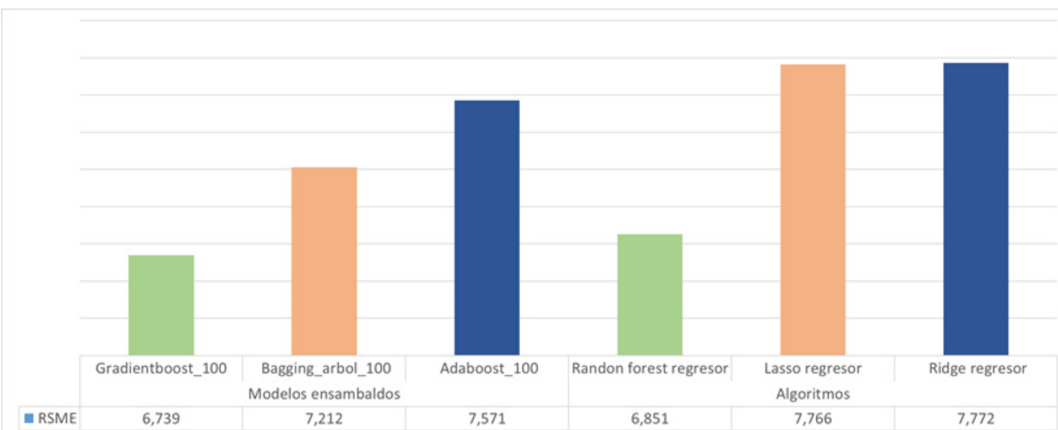
Table 4. Metrics for evaluating ensemble method

Metric	Equation	Description	Performance Criteria
MBE	$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$	Provides information on the long-term performance of the models.[18]	The closer it is to zero, the better the prediction result will be.
R2	$1 - \frac{\Sigma(y_i - x_i)^2}{\Sigma(x_i - \underline{x_i})^2}$	Establece qué tan bien se aproximan los datos reales a la línea de regresión. [18]	Oscila entre 0 y 1, entre más se aproxime a 1, mejor será el rendimiento del modelo.

Source: own elaboration

Regarding the RMSE of the ensemble models, the parameter n_estimators=100 was used for Boosting and Bagging. Figure 6 shows that the Boosting model, assembled with decision trees, obtains the best values, meaning it approximates zero more closely. This is because this model uses decision trees, like the Random Forest Regressor. The improvement in performance compared to the Random Forest Regressor and the rest of the models is due to the fact that in Gradientboost_100, individual decision tree ensembles are trained sequentially. Therefore, each tree improves the errors of the preceding trees, making the prediction calculated automatically based on the predictions of all individual trees that make up the model.

Figure 6. RMSE values of the ensemble models



Source: self made

The ensemble models were evaluated and it determined that Gradientboost_100 would be used to make predictions. According to the reasons for residing on the street presented in Table 5. In this regard, Figure 7 shows the results of the proposed model prediction, which indicates that for inhabitants under 10 years old, the main reasons for living on the street are influenced by friendships and economic difficulties, as people at this age are easily manipulated, abused, or do not have sufficient economic support at home.

For inhabitants aged between 15 and 55 years old, the main reasons for street life are, first, the consumption of psychoactive substances, followed by personal preference and friendships. Identifying the primary cause of residents living on the streets within this age group is complex, as they face multiple situations such as the use of psychoactive substances, unemployment, family problems, depression, or strong difficulties. These situations lead them to leave their homes and wander the streets, ultimately becoming homeless. In contrast, the primary reasons why residents over 55 years old end up living on the streets are typically related to economic difficulties, unemployment, health issues, conflicts, or family problems. These are common challenges faced by individuals in this age group who do not see any other viable solutions or opportunities, leading them to become homeless.

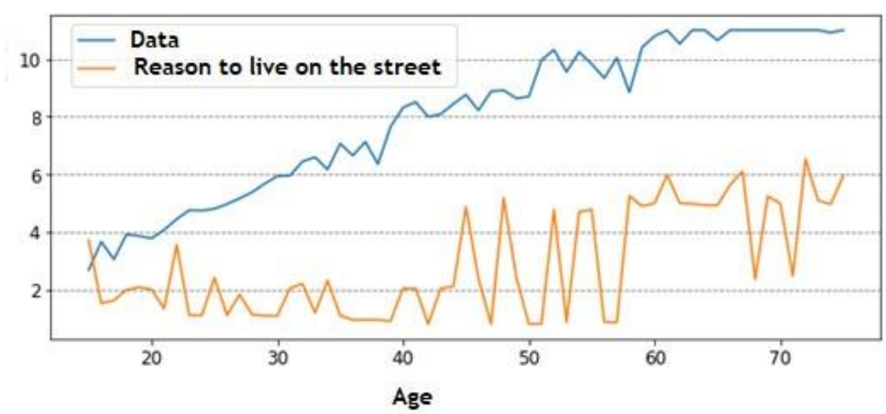
Similarly, it can be affirmed that the predictions made by the ensemble model are very close to the reality observed among street dwellers in Colombia. Several studies, including those conducted by the United Nations Office on Drugs and Crime, suggest that drug use is the primary reason for homelessness in Colombia.

Table 5. Reasons for residing on the street

Reasons for living on the street
1. Consumption of psychoactive substances
2. For personal taste
3. Friendships
4. Economic difficulties
5. Lack of work
6. Illness
7. Family conflicts or difficulties
8. He has always lived on the street
9. Loneliness
10. You are doing the process in a care center
11. Other

Source: own elaboration

Figure 7. Prediction of reasons for residing on the street by age



Source: own elaboration

Finally, the evaluation of the proposed model was executed with the metrics described in lines above (Table 6), in order to determine if the model will perform a good prediction. To evaluate de test data, and evaluate the performance of new predictions. Since new data may have unknown values. In this sense, Table 6. shows that regarding MBE, the achieved result approaches 1.742, meaning that the predictions made by the model are closely related to observation in the dataset-test. This can be corroborated with the data on the correlation of the producer variables, where psychoactive substances have a strong relationship, as mentioned previously. Similarly, for R2, the achieved value tends to 0.337, which confirms that each predictor variable does not affect the variation of the dependent variable due to the low percentage of variation of the independent variables. Therefore, the prediction made by the model is valid, as it is based on the relationship of the variables.

Table 6. Evaluated metrics

Metrics used	
MBE	1.742
R2	0.337

Source: own elaboration

Conclusions and future work

In this article, a method of ensemble machine learning algorithms was proposed, consisting of 3 predictor algorithms and 2 ensemble methods. To define the method, a dataset from the Census of Street Dwellers - CHC- 2021, provided by the National Administrative Department of Statistics (DANE), was used. This dataset contains 19,375 records and 22 variables. The CRISP-DM methodology was applied to research and treatment of the data, which allowed for understanding of the dataset and conceptualization within the domain.

The main sources of income for people living on the streets are recycling, window cleaning, car guarding, and begging. These activities are directly linked to inequality and poverty. Additionally, it has been found that most people living on the streets have addictions to psychoactive substances.

During the training process of the ensemble method, it was discovered that Gradientboost_100 improves performance due to the fact that individual decision trees are trained sequentially, reducing the errors of the preceding trees. As a result, predictions are made automatically based on the predictions of all the individual trees that comprise the model [19, 20].

The predictions generated by the method led to the conclusion that age is closely linked to the reason for living on the street. Children under the age of 10 are often vulnerable to influence, mistreatment, exploitation, or face economic and family problems. For people living on the street between the ages of 15 and 55, the reasons are diverse and include psychological problems, drug addiction, health problems, family problems, lack of family support, extreme poverty, armed conflict, or displacement.For people over the age of 55, reasons for living on the streets may include having no permanent family home, nowhere else to go, or facing challenges such as illnesses, disabilities, or a lack of employment opportunities after retirement.

In conclusion, the proposed model can serve as a tool for supporting decision-making by government institutions. It can be used to formulate, manage, and evaluate policies, plans, and programs that promote comprehensive care, rehabilitation, and social inclusion of people living on the streets in Colombia. As a future endeavor, the method will be tested with other datasets, such as accessibility to education or access to housing in Colombia, both of which can be found in ANDA, the National System of Open Data. Moreover, the proposed method will be enhanced with other ensemble techniques, such as the stacking technique, which uses a general procedure to assemble base models

References

1. J. A. Jaramillo Serna, T. Fernández Cifuentes, and S. Bedoya Sepúlveda, "Habitantes de calle: entre el mito y la exclusión," *Poiésis*, vol. 1, no. 32, p. 179, 2017, DOI: 10.21501/16920945.2311.
2. 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.
3. M. E. Correa and J. Zapata, "La Otra Ciudad: Los Habitantes De La Calle," *Prospect. Rev. Trab. Soc. E Interv. Soc.*, vol. 12, pp. 181–204, 2007.
4. A. Arroyave, "El problema público de habitante de calle en Medellin 2012-2015. Un estudio desde el instrumento de regulación social," 2016.
5. C. Fern, "Una propuesta para la concepción y abordaje del habitante de la calle desde una perspectiva psicoanalítica," pp. 1–14, 2013.
6. A. Villegas, C. Zuluaga, R. Cardona, and C. Ruiz, "ANÁLISIS Y EJECUCIÓN DE LOS LINEAMIENTOS DE POLÍTICA PÚBLICA PARA HABITANTES DE LA CALLE EN MATERIA DE SALUD Y EDUCACIÓN EN CALI," *Proyección e Innovación Soc.*, vol. II, pp. 123–134, 2020.
7. J. Dessain, "Machine learning models predicting returns: Why most popular performance metrics are misleading and proposal for an efficient metric," *Expert Syst. Appl.*, vol. 199, no. March, p. 116970, 2022, DOI: 10.1016/j.eswa.2022.116970.
8. H. Ordóñez, C. Cobos, and V. Bucheli, "Machine learning model for predicting theft trends in Colombia," *RISTI - Rev. Iber. Sist. e Tecnol. Inf.*, vol. 2020, no. E29, pp. 494–506, 2020.
9. S. G. Alegría, "de víctimas de secuestro en Colombia Neural network model for predicting kidnapping victims trends in Colombia," vol. 8, pp. 38–49, 2020.
10. O. A. Olabanjo, B. S. Aribisala, M. Mazzara, and A. S. Wusu, "An ensemble machine learning model for the prediction of danger zones: Towards a global counter-terrorism," *Soft Comput. Lett.*, vol. 3, no. May 2020, p. 100020, 2021, DOI: 10.1016/j.socl.2021.100020.
11. J. A. Jaramillo Serna, T. Fernández Cifuentes, and S. Bedoya Sepúlveda, "Dynamic model for establishing policies of attention to the street dwellers," *Poiésis*, vol. 1, no. 32, p. 179, 2017, doi: 10.21501/16920945.2311.
12. C. Ministerio de Justicia, *estudio de mortalidad asociada al consumo de sustancias*, vol. 1. 2020.
13. C. Moreno Baptista, G. Espinosa Herrera, and L. Zapata Piedrahita, "Consumo de bazuco y la ruta hacia la situación de calle: crónicas y escenarios de la vivencia familiar," *Rev. Investig. Andin.*, vol. 21, no. 39, pp. 307–324, 2019, doi: 10.33132/01248146.1571.
14. G. Gesti, "Política Pública Social para Habitantes de la Calle 2021-2031," *Of. Promoción Soc. Colomb.*, vol. 1, 2021.
15. J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, "Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - Saber 11 Test," *Procedia Comput.* Sci., vol. 198, no. 2020, pp. 512–517, 2021, doi: 10.1016/j.procs.2021.12.278.

16. E. J. De la Hoz Domínguez, T. J. . Fontalvo Herrera, y A. A. Mendoza Mendoza, "Aprendizaje automático y PYMES: Oportunidades para el mejoramiento del proceso de toma de decisiones", *Investigación e Innovación en Ingenierías*, vol. 8, n.º 1, pp. 21–36, 2020. DOI: https://doi.org/10.17081/invinno.8.1.3506
17. C. A. Ruíz Ramírez, D. M. Montoya Quintero, y J. A. Jimenez Builes, "Un Ambiente visual integrado de desarrollo para el aprendizaje de programación en robótica", *Investigación e Innovación en Ingenierías*, vol. 9, n.º 1, pp. 7–21, 2021. DOI: https://doi.org/10.17081/invinno.9.1.3957
18. C. Tofallis, "Measuring Relative Accuracy: A Better Alternative to Mean Absolute Percentage Error," *SSRN Electron. J.*, no. January 2013, 2013, doi: 10.2139/ssrn.2350688.
19. P. Rückert, B. Papenberg, and K. Tracht, "Classification of assembly operations using machine learning algorithms based on visual sensor data," *Procedia CIRP*, vol. 97, pp. 110–116, 2020, DOI: 10.1016/j.procir.2020.05.211.
20. O. de las naciones unidas contra las drogas y el Delito, *Informe mundial sobre las drogas*. 2018. DOI: 10.18356/a9c60617-es.