Anuario de Psicología Jurídica (2026) 36, e260485, 1-11

Colegio Oficial de la Psicología de Madrid

Anuario de Psicología Jurídica 2026

https://journals.copmadrid.org/apj

Factors Associated with Family Members in Prison: An Imbalanced Machine Learning Approach

Mariana Politti[1,2], Luis A. Calvo Pascual[3,7], Gisela Delfino[3,2,4,5], and Solange Rodríguez Espínola[1,6]

[1]Pontifical Catholic University of Argentina; [2]National Scientific and Technical Research Council, Argentina; [3]Comillas Pontifical University, Spain; [4]Andrés Bello University, Chile; [5]University of Buenos Aires, Argentina; [6]Argentine Social Debt Observatory, Argentina; [7]Institute for Research in Technology (IIT), Spain

ARTICLE INFO

Article history:
Received 24 March 2025
Accepted 13 November 2025

Keywords:
Imbalanced machine learning classification
Socioeconomic factors
Incarceration
SHAP analysis
Argentina

A B S T R A C T

This study explores the family environment of individuals deprived of liberty in Argentina through an imbalanced machine learning approach. Based on a national representative dataset of 17,139 individuals from various urban regions, machine learning techniques, including mutual information, balanced random forest, and SHAP analysis, were employed to identify structural socioeconomic conditions associated with having incarcerated family members. The main factors identified include lower socio-economic status, overcrowded living conditions, lack of homeownership, and limited education. The Balanced Random Forest model achieved an F1 score of .6357 and a PR AUC of .7063, demonstrating its effectiveness in handling the imbalanced dataset. SHAP analysis provided interpretability to the results by illustrating how the previous factors are related to having an incarcerated family member. This study reveals the need for targeted policy interventions that address structural inequalities and promote social mobility. The findings underscore the importance of shifting from punitive approaches to preventive strategies that focus on improving education, housing conditions, and economic opportunities for vulnerable populations.

Los factores asociados con tener familiares en prisión: un enfoque de aprendizaje automático con datos desbalanceados

Palabras clave:
Clasificación de aprendizaje automático con datos desbalanceados
Factores socioeconómicos
Encarcelamiento
Análisis SHAP
Argentina

R E S U M E N

Este estudio explora el entorno familiar de sujetos privados de libertad en Argentina mediante un enfoque de aprendizaje automático desbalanceado. Utilizando un conjunto de datos representativo a nivel nacional de 17,139 individuos provenientes de diversas regiones urbanas, se aplicaron técnicas de aprendizaje automático, como información mutua, bosques aleatorios balanceados y análisis SHAP con el objetivo de conocer las condiciones estructurales de índole socioeconómica vinculadas con tener familiares privados de libertad. Los principales factores encontrados fueron tener un estatus socioeconómico bajo, vivienda compartida por muchas personas, falta de propiedad de la vivienda y educación limitada. El modelo de bosques aleatorios balanceados tuvo unas métricas F1 de .6357 y un área bajo la curva de precisión (PR AUC) de .7063, demostrando su efectividad al tratar con una base de datos desbalanceada. El análisis SHAP permitió interpretar los resultados al describir cómo las anteriores variables se relacionan con tener un familiar encarcelado. Este estudio revela la necesidad de intervenciones públicas dirigidas que aborden las desigualdades estructurales y promuevan la movilidad social. Los resultados subrayan la importancia de pasar de enfoques punitivos a estrategias preventivas que se centren en mejorar la educación, las condiciones habitacionales y las oportunidades económicas para las poblaciones vulnerables.

The complexity of periodically collecting information related to Argentina's penitentiary system is due, more than anything else, to the diverse legal competencies governing prisons. While federal penitentiary institutions exist nationwide, others operate under provincial penitentiary services. According to the National System of Statistics on the Execution of Sentences (SNEEP, 2021), in 2021, there were 114,074 people deprived of liberty, with the Province of Buenos Aires housing 46% of the total. The prison population is predominantly characterized by young, male, Argentine residents with low levels of education. The main offenses leading to incarceration include robbery (and attempted robbery), sexual abuse (rape), intentional homicide, and violations of Law 23.737 (1989) on narcotics.

Cite this article as: Politti, M., Calvo Pascual, L. A., Delfino, G., & Rodríguez Espínola, S. (2026). Factors associated with family members in prison: An imbalanced machine learning approach. *Anuario de Psicología Jurídica, 36,* Article e260485, 1-11. https://doi.org/10.5093/apj2026a18

Correspondence: gidelfino@comillas.edu (G. Delfino).

According to the information provided by the detention units, in 2021, there was an average overpopulation rate of 15.7%, showing significant challenges in the Argentine penitentiary system. Prison overcrowding leads to a lack of privacy, poor hygiene, and significant obstacles to accessing work, education, and health services. These violate their rights and undermine the objective of resocialization and social reintegration (Procuración Penitenciaria de la Nación, [PPN] 2022).

Latin American Context

In contrast to global trends, the prison population in Latin America has been increasing at a rate surpassing that of the general population for several years (PPN, 2022). According to the United Nations Office on Drugs and Crime (UNODC, 2021), the incarceration rate per 100,000 inhabitants in Latin America has risen by nearly 70% over the past two decades. This increase is closely linked to persistent inequality in the region, which has been shown to contribute to higher crime rates. Recent research analyzing new data on inequality and crime in Latin America and the Caribbean confirms that disparities in income and social mobility play a crucial role in shaping criminal behavior and incarceration trends (Schargrodsky & Freira, 2023).

Sociodemographic Factors and Incarceration

Delinquency in the Latin American region cannot be considered without analyzing its psychological and social contexts. Educational inequalities are closely linked to the socioeconomic reality of families. Although it has been shown that education is not a determining factor for social improvement, educational systems in Latin America are unequal to the point of reproducing and maintaining structural differences. The disparity in educational opportunities for those with less purchasing power presents a significant obstacle to both the perception and the real possibility of social mobility for young people in the region (Mendoza & Ewing, 2020; Travis et al., 2014).

Several studies focus on the sociodemographic variables associated with a higher probability of incarceration. Factors like race (Bagaric et al., 2022), personality traits such as anger (Cornell et al., 1999), mental health, substance abuse, and having been detained at a young age (Pedrosa, 2020) have been linked to a greater propensity for deprivation of liberty. Beyond individual traits, structural economic factors play a significant role in incarceration rates in Latin America. Studies have shown that crime and violence impose significant economic and social burdens on the region, with the costs of crime estimated to absorb nearly 3.5% of Latin America's economic output, affecting public spending on education, employment opportunities, and social safety nets (Jaitman, 2015).

The criminalization of poverty has long been a focus of study, with the understanding that it is not poverty, but inequality, that generates an increase in the crime rate. Consumer societies set goals that not all members of society can achieve, generating frustration and aggression that can lead to criminal behavior (Cid Ferreira et al., 2017). This perspective aligns with broader findings in Latin America, where research has shown that socioeconomic inequity, rather than absolute poverty, plays a crucial role in driving crime rates. Structural disparities limit access to opportunities, fueling social tensions and increasing the likelihood of criminal activity, particularly in marginalized communities (Niño, 2010).

In this regard, a recent study in Bangladesh (Haque & Muniruzzaman, 2020) found that impoverished living conditions increase the possibility of individuals engaging in criminal activities. Results showed that limited access to basic household facilities, such as electricity, sanitation, and transportation, can contribute to heightened frustration and increase the chances of criminal behavior. For instance, findings indicate that individuals living in conditions of extreme deprivation, where access to food, electricity, and essential goods is severely restricted, may experience greater socioeconomic stressors that can push them toward illicit activities.

Additionally, different forms of economic deprivation can shape the crimes individuals are more likely to commit. The patterns and levels of criminal behavior also vary based on factors such as gender, marital status, and a person's level of education and professional background. Similarly, Jaitman (2015) highlights that individuals from lower socioeconomic backgrounds face greater exposure to financial instability, housing insecurity, and unemployment, all of which contribute to a higher likelihood of incarceration.

Research on incarceration has shown that both micro- and macro factors play a role in shaping incarceration rates. A study in the United States found that education, especially parental education, along with gender and age, can influence a person's possibility of ending up in the criminal justice system (Barnert et al., 2021). At the same time, research on U.S. counties found that when mental and behavioral health services are more accessible and affordable, incarceration rates tend to be lower (Fahmy & Mitchell, 2022). This suggests that while individual factors matter, broader issues like healthcare access and economic stability also shape incarceration outcomes. Communities with better mental health and substance use treatment resources may be able to support individuals in ways that reduce reliance on punitive measures.

Furthermore, individuals with co-occurring mental health and substance use disorders are at a substantially elevated risk of reincarceration compared to those with no disorders or mental health needs alone, highlighting the necessity for integrated correctional, health, and social services (Butler et al., 2024). However, there is a notable lack of comprehensive studies on predictive variables in the Latin American context. This is particularly relevant given that socioeconomic status is one of the most frequently examined factors related to incarceration in this region. Despite this focus, no studies have specifically analyzed socioeconomic level as a predictor variable for incarceration, leaving a critical gap in understanding how structural inequalities contribute to imprisonment in Latin America.

Impact on Families

Recent high-quality evidence from a large-scale systematic review and meta-analysis confirms that exposure to family incarceration is robustly linked to elevated risks for a wide range of psychiatric symptoms in affected families (Liu et al., 2025). These include heightened risks of externalizing problems, substance use disorders, and psychological distress, with powerful effects for children and populations in middle-income countries. The systematic review highlights that these consequences persist across various countries and demographic groups, and points to the need for comprehensive psycho-legal interventions targeting the entire family.

From a forensic psychology perspective, having a family member incarcerated is not only a marker of social vulnerability but also directly activates the requirement for psychosocial and forensic assessment within judicial proceedings. This includes the evaluation of psychological harm, parental fitness, child custody decisions, and recommendations for supervised visitation. Such assessments are crucial for protecting the rights and well-being of minors and other vulnerable family members. They are routinely requested by courts to inform sentencing, custody, and rehabilitation measures. Importantly, the responsibility to implement these resources and interventions lies with the State and its relevant institutions, which must ensure that support and evaluation extend beyond the person deprived of liberty to address their family's comprehensive needs and management (Liu et al., 2025; Ríos, 2023; Thew & Terry, 2025; Wildeman & Wakefield, 2014).

Recent qualitative research with Argentine families (Politti & Delfino, 2023) shows that the deprivation of liberty deeply affects the entire family system, requiring significant emotional and organizational adaptation. Beyond the psychological impact on the incarcerated individual, family members themselves experience profound changes in their daily lives, relationships, and sense of belonging. These changes include economic strain, emotional distress, stigma, and disruption of family roles. Analyzing the family as a dynamic system, this perspective highlights that incarceration triggers a process of systemic adjustment, often with lasting consequences for all members.

These families often face social and occupational discrimination, treated as if they were "contaminated" by the offender's actions. The disruptions in family life carry economic and psychological burdens, often borne by families already experiencing social marginalization (Halsey & Deegan, 2015; Jardine, 2017; de Mendonça & Barsaglini, 2020; Paredes Blandón, 2019). Besides the research conducted by Campo Aguzzi et al. (2019), no studies have delved into the social reintegration of people deprived of liberty in the Argentine context. Nevertheless, the rates of recidivism and reiteration increase year by year. It peaked in 2020, with 20% of the prison population being part of the first group and 10% of the second (SNEEP, 2020).

Within the framework of the studies conducted by the Observatory of the Argentine Social Debt, it was observed that one of the most significant impacts suffered by the families of incarcerated people is the reduction of economic resources in the household. This generates financial instability and material difficulties that often cause insecurity in fulfilling basic needs and substantially affect their quality of life (Cadoni et al., 2019). Could these characteristics, presented as a description, have some predictive value? Children and adolescents with an incarcerated family member predominantly live in low-income environments, with 21.7% residing in slums and settlements and 42.2% in medium or medium-low urban areas. Additionally, 39.8% live in single-parent households, and nearly half (48.8%) belong to families with a low educational level.

Research Objective

This article aims to explore the various sociodemographic variables associated with the imposition of incarceration in Argentina. It offers valuable insights into incarceration's broader social and economic impacts on families and communities. It is important to emphasize that this study does not analyze individual characteristics of adjudicated people, nor the factors directly associated with criminal behavior. Instead, it focuses on identifying the socioeconomic, housing, and educational structural correlates of households where at least one member is or has been deprived of liberty. In this way, the work provides evidence on the family and community conditions that increase the likelihood of incarceration, without examining the individual determinants of criminal conduct.

Understanding the relationship between these variables and imprisonment is crucial for identifying key factors that can guide the development of more targeted and effective crime prevention strategies and social interventions. Additionally, exploring the social and economic factors behind incarceration can reveal hidden biases and unfairness in the criminal justice system, paving the way for much-needed reforms. Developing predictive models based on these variables can support early intervention strategies and improve risk assessment within the criminal justice system.

Method

This cross-sectional, correlational study employs advanced statistical and machine learning techniques to identify key socioeconomic variables of households with an incarcerated member in Argentina.

Data Collection

The dataset uses three waves of data collection of the Argentina National Debt Survey (EDSA, in Spanish). The study comprises a comprehensive set of sociodemographic variables related to individuals and families affected by incarceration derived from a survey of 17,139 individuals from various regions of Argentina collected within the Argentine Social Debt Observatory of the Pontifical Catholic University of Argentina (UCA) from 2017 to 2019. The sample consisted of individuals aged 18 to 98 and was designed to represent urban clusters in Argentina with at least 80,000 inhabitants, based on data from the 2010 census. A multistage sampling approach was employed, beginning with a clustering stage followed by stratification. Sample clusters within each urban group and stratum were selected randomly, with weighting based on the number of households. Within each sample point, blocks and households were chosen through systematic random sampling, whereas individuals within households were selected using a quota system that considered sex and age. About one-third of the households from the initial probabilistic sample participated in multiple survey rounds, enabling the creation of panel data. To be included in this panel subgroup, individuals had to continue living in the household, willingly take part each year, and complete the entire survey in every wave. Socio-economic stratification was conducted by classifying and ranking census clusters according to the average educational level of the household head. The overall margin of error is less than ± 3% at a 95% confidence level, assuming maximum dispersion (p and q = .50) and accounting for a design effect of 23 (Rodríguez Espínola et al., 2020).

The dataset consists of 22 variables: "liberty_deprivation" (the target variable) and 21 additional socioeconomic and demographic indicators. The target variable does not refer to whether the surveyed individual is or has been in prison, but rather to whether any member of their household has experienced deprivation of liberty. Thus, the analysis focuses on the family environment's characteristics and not on adjudicated persons' individual particularities. A detailed description of each variable, including its definition and category coding, is provided in Table A1 in the Appendix. A comparative analysis based on "liberty_deprivation" (0 or 1) of bar charts for each categorical variable and descriptive statistics for the numerical variables is available in the Supplementary materials.

Feature Selection

The target variable, "liberty_deprivation", represents having a family member in prison, which is highly imbalanced (roughly 5% prevalence, not in prison = 16,410, in prison = 729), making it unsuitable for traditional statistical methods. The process begins with feature selection, which ranks and identifies the most relevant variables. To enhance the robustness of our feature selection, we employ several distinct methods: Mutual Information, Random Forest, XGBoost, and a logistic model approach. Mutual Information captures non-linear relationships between variables and the target, making it invaluable for detecting complex dependencies that simpler linear methods might overlook. On the other hand, tree-based methods, such as Random Forest and XGBoost, naturally model non-linear interactions and often provide feature importance measures based on splits across ensembles of decision trees, including a logistic method provides a more traditional, linear perspective on variable relevance. By comparing these different techniques, we gain a more comprehensive understanding of each feature's contribution, regardless of its linearity or complexity.

For each method, we created a nested cross-validation workflow. The dataset is partitioned into features (X) and target (Y), and a five-fold outer cross-validation defines train/test splits for final evaluation. A custom preprocessing pipeline is built inside each outer fold to detect and appropriately transform binary, categorical, and numerical columns. By doing this strictly within the training portion of each fold, we minimize data leakage and preserve the integrity of our performance estimates. We derive feature importance in the same outer fold (e.g., via a model-based metric or any other technique) by applying a bootstrap procedure. Multiple resamples of the training data are created and the chosen importance metric is recalculated for each resample. Aggregating these results across all bootstraps provides mean importance scores and standard deviations, quantifying the variability in those estimates. After completing each fold, the script stores and merges all results to produce a global ranking of features, including their average importance scores.

After computing feature importance via each method, we normalize the resulting scores so they can be directly compared, despite originating from different feature selection algorithms.

Imbalanced Machine-learning Models

To address the challenge of our imbalanced dataset, we employed a strategy focused on algorithm selection rather than data manipulation techniques. We explored various classification algorithms, each with distinct capabilities in handling imbalanced data: Random Forest, XGBoost, MLP Classifier, Logistic Regression, CatBoost, Balanced Random Forest, and AdaBoost. This selection encompasses machine learning paradigms, including tree ensembles, gradient boosting, neural networks, and linear methods. Additionally, we included a baseline model using a Dummy Classifier that always predicts the positive class ("liberty_deprivation" = 1), which aligns with our primary interest in identifying these cases. This baseline serves as a reference point to contextualize the performance of the trained models.

By comparing models with distinct strengths (e.g., robustness to outliers, automatic feature interactions, or inherent handling of imbalanced data), we aimed to identify the most effective algorithm for our specific task. For highly imbalanced classification problems, traditional accuracy metrics can be misleading. Therefore, we focused on the F1 score, the harmonic mean of precision and recall, as it balances false positives and negatives. Additionally, we employed Precision-Recall Area Under the Curve (PR AUC) to further evaluate performance in terms of recall (the correct identification of positive instances) relative to precision (the minimization of false positives). PR AUC is especially informative in scenarios with skewed class distributions, as it highlights differences in performance even when true negatives dominate the dataset. This approach has been employed in other articles, such as Castro Corredor and Calvo Pascual (2023).

In addition to these primary metrics, Table 1 includes each model's full confusion matrix, along with sensitivity (Recall) and specificity (True Negative Rate, TNR) values. These additional indicators provide a more detailed view of classification performance, helping to reinforce the validity of our results.

Model Training and Evaluation Procedure

Once the most influential features were determined, we adopted the following procedure for each model. First, we reserved 80% of the data for training and 20% for testing, ensuring class proportions were maintained via stratification. Within each cross-validation fold, a RandomUnderSampler was integrated into a pipeline to produce a more balanced subset, refining our estimates of model performance. We employed Bayesian Optimization through the bayes_opt library to optimize hyperparameters, using a five-fold stratified cross-validation to compute the average F1 and PR AUC scores at each iteration. After convergence, the best hyperparameters were used to refit the pipeline on the entire training set, and the final model was evaluated on the untouched test set. To gain a clearer view of predictive capacity across classes, we also created a balanced test set by randomly undersampling the majority class, thereby offering more profound insights into model performance for both minority and majority classes.

SHAP [SHapley Additive exPlanations] Analysis

Like most ensembled methods, Balanced Random Forest is a "black box" whose internal decision-making is not easily interpretable. The lack of transparency in such models has been widely discussed in literature, particularly in high-stakes decision-making contexts, where understanding the reasoning behind predictions is crucial (Rudin, 2019). We employed SHAP values to quantify how much each feature contributes to a prediction to gain insight into the decision-making process. By leveraging the SHAP Python library (shap.TreeExplainer), we visualized global and local explanations of the Balanced Random Forest outputs. Specifically,

Table 1. Model Performance Metrics: F1 Score, PR AUC, Recall, TNR, and Confusion Matrix

Model	F1	PRAUC	Recall	TNR	Confusion matrix
Random Forest	.6286	.6963	.6849	.6027	$\begin{pmatrix} 100 & 46 \\ 58 & 88 \end{pmatrix}$
XGBoost	.6273	.6991	.7260	.5822	$\begin{pmatrix} 106 & 40 \\ 61 & 85 \end{pmatrix}$
MLP Classifier	.4744	.6799	.8767	.3493	$\begin{pmatrix} 128 & 18 \\ 95 & 51 \end{pmatrix}$
Logistic regression	.6097	.6840	.7192	.5616	$\begin{pmatrix} 105 & 41 \\ 64 & 82 \end{pmatrix}$
Catboost	.6022	.6411	.6644	.5753	$\begin{pmatrix} 97 & 9 \\ 62 & 84 \end{pmatrix}$
Balanced Random Forest	.6357	.7063	.6918	.6096	$\begin{pmatrix} 101 & 45 \\ 57 & 89 \end{pmatrix}$
Ada Boost	.6182	.6908	.6986	.5822	$\begin{pmatrix} 102 & 44 \\ 61 & 85 \end{pmatrix}$
Dummy Classifier	.6667	0.5	0	1	$\begin{pmatrix} 0 & 146 \\ 0 & 146 \end{pmatrix}$

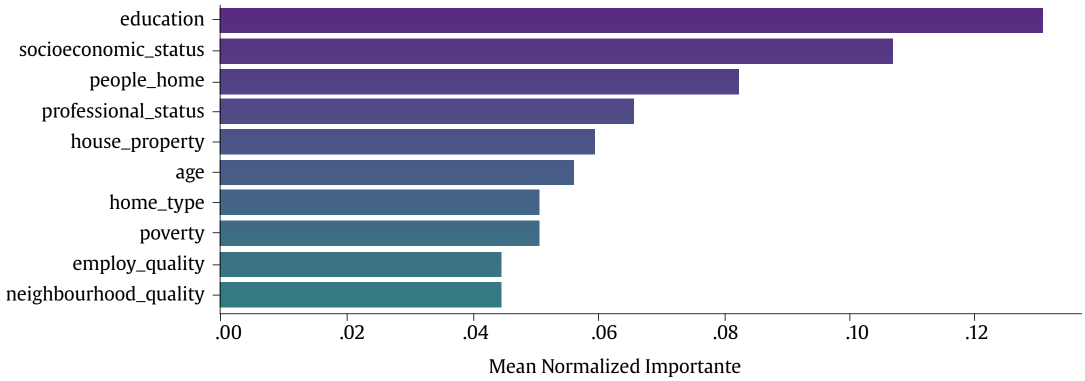

Figure 1. Top 10 Features according to Mean Normalized Importance.
Data source: EDSA-Agenda for Equity (2017-2025), Argentine Social Debt Observatory, UCA.

we generated SHAP summary plots to assess the average impact of each feature across the dataset. Additionally, we created a Waterfall Plot highlighting a high predicted probability case, allowing us to analyze how the main features interacted to produce an extreme outcome.

Results

Based on the results from mutual information and other feature selection methods, the most influential variables explaining liberty_deprivation were: education (a binary variable distinguishing incomplete high school or high school and superior), socioeconomic_status (very low, low, middle-low, middle-high), people_home (number of individuals residing in the home), professional_status (a categorical variable indicating if it is a subsistence worker, skilled worker, traditional middle-class or professional middle-class) and house_property (occupants, renters, homeowners, etc.). (See Figure 1).

The selected model was the Balanced Random Forest, as it achieved the highest F1 score (.6357), PR AUC (.7063), and specificity (TNR .6096) among all models evaluated, except for the Dummy Classifier in F1 and specificity. However, the Dummy Classifier has a poor PR AUC of .50 and zero Recall. In contrast, the Balanced Random Forest obtained a robust PR AUC of .7063 and a Recall that, while slightly lower than the highest models, remained comparable (see Table 1).

The precision-recall curve (Figure 2) shows a descending slope, with an area under the curve (AUC) of approximately .7064. This indicates that the model reasonably balances precision (minimizing false positives) and recall (correctly identifying positive cases).

Figure 3 presents the features with the strongest associations with liberty_deprivation. These are low socio-economic status, living in an overcrowded household, a lower level of educational attainment, not owning a home, and a lower professional status.

Each dot represents an individual instance, with the horizontal axis showing the SHAP value, which indicates the impact of the feature on the model output. Positive SHAP values correspond to an increased likelihood of liberty deprivation, while negative values decrease this likelihood. The color gradient reflects the feature value, ranging from low (blue) to high (red).

Individuals with a low or very low socioeconomic status, represented in blue, exhibit higher SHAP values, indicating a strong positive correlation with the likelihood of having a family member in prison. Similarly, lower levels of education (incomplete high school), lower levels of house_ property (occupants, renters, etc.), and lower levels of professional_status (subsistence or skilled workers), all represented in blue, contribute to an increased probability of incarceration among family members. In contrast, having more family members living together, people_home, represented in red, is also associated with a higher probability of having a relative in prison.

Finally, we present the results shown in Figure 4, which displays the SHAP Waterfall Plot for a single instance predicted with a high probability of liberty deprivation, $f(x) = .941$. The most influential feature in this case was people_home, contributing + .16; specifically, the respondent lived with seven people in the same household. The second most impactful factor was house_property, adding + .12, as the individual resided in borrowed housing. Additionally, socioeconomic_status contributed + .09. The individual also had an education level below high school, contributing + .05 to the final probability.

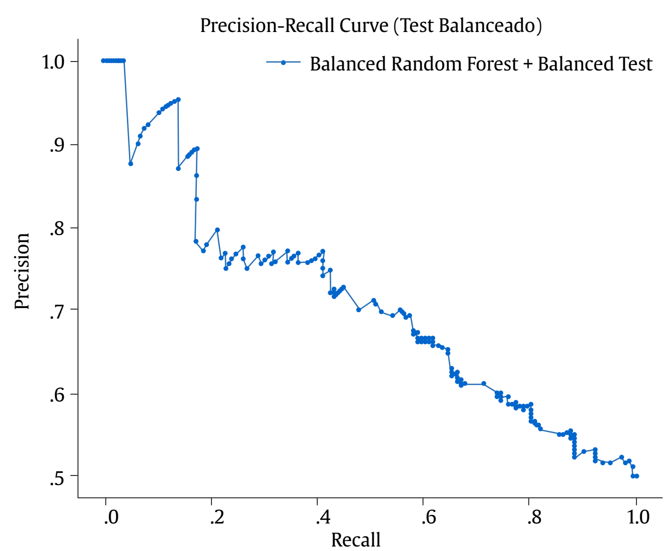

Figure 2. Precision-Recall Curve of the Balance Random Forest Model.

Additionally, to address whether gender and age might moderate the relationship between key predictors and having a family member in prison, we conducted exploratory analyses. We computed SHAP values for the main model predictors and compared their distributions by sex (using the Mann-Whitney test) and age group (using the Kruskal-Wallis ANOVA). For the variable education, we found statistically significant differences between men and women ($p < .05$), suggesting that the influence of lower educational attainment on incarceration risk may differ by gender, with men having a higher probability. In the case of age, although no overall significant interaction was detected, a closer examination of participants younger than 25 revealed that the variable people_home (number of people residing in the household) tended to contribute more substantially to the predicted risk in this subgroup. This finding may reflect that, among younger respondents, living in a large household, often corresponding to families with multiple children, is particularly associated with a higher predicted risk of having a family member in prison. These exploratory results are preliminary and should be interpreted with caution. However, they provide valuable insights for future research on the potential moderating role of demographic factors in the relationship between structural conditions and family incarceration.

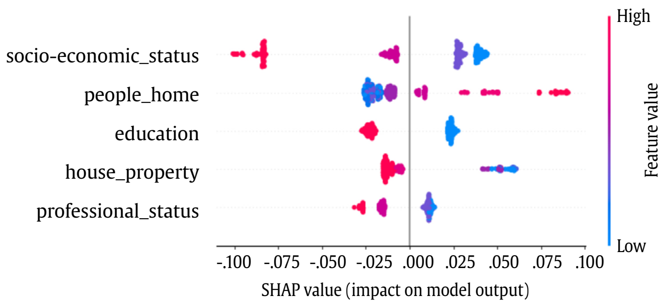

Figure 3. SHAP Values of the Balance Random Forest Model.

Discussion

The findings of this study provide valuable insights into the socioeconomic factors associated with incarceration in Argentina, utilizing imbalanced and explainable machine learning techniques to address the challenges of imbalanced data. The results of this study should be interpreted within the context of a household-level analysis, not of adjudicated individuals. The research did not investigate the individual causes of criminal behavior or the profiles of those deprived of liberty, but rather the structural conditions that characterize families affected by incarceration. This perspective helps guide public policies focused on prevention and reducing social vulnerability, rather than on interventions targeting the convicted individual.

Education, Socioeconomic Status, and Professional Status

An influential predictor identified by the model was education, with lower educational attainment associated with an increased likelihood of incarceration. Socioeconomic status emerged as another critical factor, consistent with previous research on the relationship between inequity and incarceration (Cid Ferreira et al., 2017). As Mendoza and Ewing (2020) and Travis et al. (2014) have stated, inequality of educational opportunities is closely linked to socioeconomic background and professional prospects, perpetuating structural differences. The findings underscore the importance of ensuring that individuals remain in the education system, improving the quality of education provided, and fostering pathways to meaningful employment. Strengthening educational access and outcomes, alongside promoting professional development, could serve as a key strategy for reducing incarceration rates and addressing broader social inequalities that could ultimately improve social mobility.

Moreover, the analysis revealed that individuals in lower professional categories, particularly subsistence workers and skilled workers, face a higher risk of incarceration. This aligns with prior research emphasizing the intersection between labor market stratification and criminal justice outcomes (Haque & Muniruzzaman, 2020; Jaitman, 2015). Previous studies have found that precarious employment and economic instability contribute to heightened stressors that can lead to engagement in illicit activities or increased policing of marginalized groups (Desmond & Gershenson, 2016; Ríos, 2023). Similarly, the structural barriers lower-status workers face, including job insecurity and informal labor conditions, reinforce cycles of socioeconomic disadvantage and criminal justice involvement (Herbert et al., 2015).

Rather than poverty alone, it is the unequal distribution of resources, opportunities, and access to stable employment that exacerbates vulnerabilities and increases the probability of involvement in the criminal justice system. This perspective is consistent with broader findings in Latin America, where research has shown that inequality, and not absolute poverty, is a key driver of crime. Structural disparities limit access to education, employment, housing, and professional advancement, reinforcing cycles of marginalization and increasing the likelihood of criminal behavior, particularly among disadvantaged populations (Niño, 2010; Schargrodsky & Freira, 2023).

Housing and Family Characteristics

The model also identified living in overcrowded households and a lack of homeownership as significant predictors of incarceration. These findings align with previous research suggesting that economic deprivation and inadequate living conditions create stressors that may increase the chance of engaging in illicit activities (Haque & Muniruzzaman, 2020). However, while prior studies have established a link between impoverished living conditions

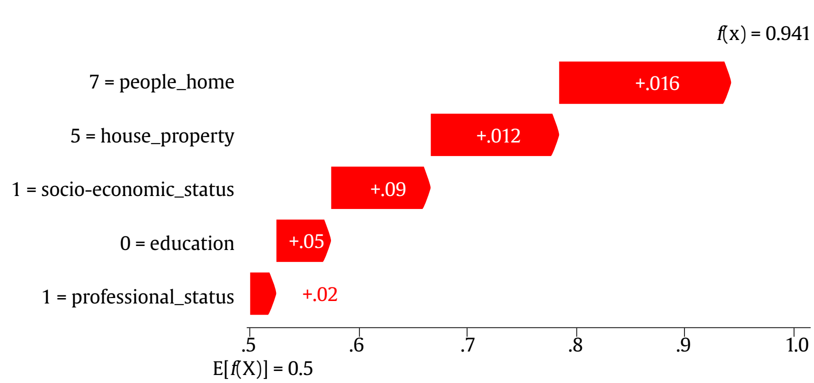

Figure 4. SHAP Waterfall Plot for the Case with the Highest Predicted Risk (Balanced Random Forest Model).

and crime, this study goes further by demonstrating that specific aspects of socioeconomic disadvantage, such as precarious housing situations, directly predict incarceration. Overcrowding and lack of homeownership indicate financial hardship and reflect more profound structural inequalities that limit opportunities for social mobility, access to stable employment, and community support systems (Desmond & Gershenson, 2016).

Previous research has shown that in Argentina, children and adolescents with an incarcerated family member are more likely to come from low-income backgrounds and face educational disadvantages (Cadoni et al., 2019). This study builds on that evidence by demonstrating that, beyond general socio-economic hardship, specific housing-related factors play a crucial role in incarceration risk, underscoring the impact of housing instability on the likelihood of imprisonment (Herbert et al., 2015). For example, Figure 4 presents a case where the individual lived with seven members, resided in borrowed housing, and had an education level below high school. This case illustrates how multiple socioeconomic disadvantages can interact to exacerbate vulnerabilities.

This study reinforces the idea that individual and structural factors shape incarceration risk. While prior research has established that education, gender, and age influence incarceration outcomes (Barnert et al., 2021) and that access to mental and behavioral healthcare services can lower incarceration rates (Fahmy et al., 2022), this study highlights the predominant role of socioeconomic status and housing conditions in Argentina. Despite the frequent examination of socioeconomic factors in Latin America (Jaitman, 2015; Schargrodsky & Freira, 2023), there has been little focus on their association with incarceration.

At the same time, subjective factors, such as happiness or personal projects, appear less relevant concerning incarceration. However, this may be influenced by how these variables were coded in the analysis, given that they were dichotomized. A more refined approach that considers individual perceptions and lived experiences could provide deeper insights into how personal and psychological factors interact with socioeconomic conditions in shaping incarceration risk. Future research should explore these elements more comprehensively to capture the complexity of incarceration dynamics.

The results support previous research and offer valuable insights for developing targeted interventions and social support programs. In particular, they emphasize the importance of improving access to durable goods and housing conditions to reduce socioeconomic inequities. Addressing these structural disadvantages could help mitigate risk factors associated with incarceration and promote greater social mobility.

It is essential to clarify that this research does not aim to predict criminal behavior or assume a deterministic relationship between the structural conditions identified and individual criminal acts. Instead, this analysis focuses on the structural and household correlates of family incarceration that are often present in the backgrounds of those who end up imprisoned. These factors are not direct causes of crime, but rather indicators of social vulnerability, and recognizing their role is crucial for designing preventive, rights-based policies that address inequality at its root.

Implications for Criminal Justice Policy

Addressing socioeconomic disparities and improving access to education and housing could play a crucial role in reducing incarceration rates. While punitive approaches remain the dominant response to crime, this study reinforces the need for policies that tackle the root causes of incarceration (Wildeman & Wakefield, 2014). In Argentina, prison overcrowding is a persistent issue, with an average overpopulation rate of 15.7% in 2021, which exacerbates challenges in accessing education, employment, and healthcare, undermining resocialization efforts (PPN, 2022).

Ensuring access to stable housing, quality education, and economic opportunities could serve as effective strategies to prevent criminal justice involvement. Prior studies have emphasized that inequality in educational opportunities is closely linked to socioeconomic background, reproducing structural disparities that limit social mobility and increase incarceration risks (Mendoza & Ewing, 2020; Travis et al., 2014). Research has also shown that incarceration disproportionately affects individuals from low-income backgrounds, who are more likely to experience financial instability, housing insecurity, and unemployment before imprisonment (Jaitman, 2015).

While efforts to reduce socioeconomic disparities and improve access to education and housing are essential to lowering incarceration rates, the data and literature show that a structural and intersectional analysis of penal systems must complement such approaches (Ríos, 2023; Wildeman & Wakefield, 2014). Our findings, in line with international research, demonstrate that incarceration in Argentina disproportionately affects individuals and families already marginalized by poverty, limited education, and precarious housing. Recent qualitative evidence reveals that these structural disadvantages intersect with gender, ethnicity, and migrant status to deepen vulnerability: women, especially migrants and those from racialized communities, not only shoulder greater financial and emotional burdens when a family member is incarcerated but also face specific forms of stigma, discrimination, and exclusion both within and beyond the justice system (Politti & Delfino, 2023).

Gender-specific analyses further highlight that women deprived of liberty themselves endure systemic neglect and abuse, including violent searches, punitive isolation, and inadequate healthcare, with reports of obstetric violence and other gender-based violations exacerbating their marginalization (Pérez Palomino, 2022; PPN, 2022). In provinces lacking facilities for women, these deficits are even more severe. These intersecting forms of exclusion and violence underscore that effective criminal justice policy must address not only the root socioeconomic drivers of incarceration but also the specific mechanisms, racialization, gendered bias, and xenophobia that reproduce and intensify vulnerability within both the penal system and affected families.

Additionally, expanding access to mental and behavioral health services, like what has been observed in prior research, may provide alternative support systems that reduce reliance on incarceration (Fahmy et al., 2022). Studies have also shown that individuals from lower-income backgrounds who experience unstable housing conditions and a lack of social support are more likely to face deprivation of liberty (Cadoni et al., 2019). Addressing these disparities through labor policies that promote employment stability and access to formal labor markets could serve as a preventive measure to mitigate incarceration risks among vulnerable populations. By shifting the focus from punishment to prevention, policymakers can work toward breaking cycles of incarceration and promoting long-term social reintegration.

Methodological and Ethical Considerations

The Balanced Random Forest model proved to be a useful tool for handling the imbalance in the dataset, allowing a good balance between precision and recall. Machine learning has already been used to explore incarceration trends, showing its potential to reveal patterns that traditional statistical methods might miss (Fahmy et al., 2022). As research in criminology and social sciences continues to evolve, these models could help improve predictions and guide policies to reduce incarceration rates.

SHAP analysis also played a key role in making the results more transparent and understandable. When dealing with sensitive issues like incarceration, knowing what the model predicts and why is crucial. By breaking down the impact of different factors, SHAP helps ensure that machine learning is not just a black box but a tool that can be used responsibly to inform real-world decisions (Rudin, 2019).

At the same time, the use of machine learning in criminal justice settings raises important ethical concerns. While the selected model is designed to identify risk factors to guide targeted interventions, ensuring it does not reinforce or amplify existing biases within the criminal justice system is essential. Transparency in how these models are developed and continuous efforts to assess fairness and equity are crucial to ensuring that predictive tools contribute to positive social change rather than perpetuating disparities.

In conclusion, this study highlights the potential of machine learning to identify key socioeconomic predictors of incarceration in Argentina. It provides valuable insights that can support evidence-based policies to reduce incarceration rates and promote social equity. However, it is important to approach these models with caution, recognizing that they should complement, rather than replace, human judgment in the criminal justice system (Barocas et al., 2019).

Limitations and Future Directions

Despite its predictive capabilities, this study has several limitations. One key constraint is that some variables were received in pre-categorized form, limiting the depth of analysis and restricting the ability to capture more nuanced relationships between socioeconomic factors and incarceration risk. Future research should aim to work with raw, continuous data to allow for a more detailed examination of these variables (Altman & Royston, 2006).

The reliance on survey data also introduces potential biases, and the cross-sectional design prevents conclusions about causality. Longitudinal studies could clarify how socioeconomic conditions influence incarceration risk over time (Farrington et al., 2016). Additionally, the model does not account for key factors such as mental health, substance use, or family history of incarceration, which are known to play a role in criminal justice outcomes (Pedrosa, 2020). Including these variables in future research could offer a more comprehensive view of the factors driving incarceration (Butler et al., 2024).

Finally, while advanced machine learning techniques like Balanced Random Forest were employed, the model's predictive performance remained moderate, with F1 score and PR AUC values indicating room for improvement. Sample size, feature selection, and the complexity of incarceration as a social issue may have influenced these results. Expanding datasets and refining feature selection could enhance predictive accuracy in future studies (Berk & Bleich, 2013).

Conclusions

The findings of this study highlight the significant role that socioeconomic and educational factors play in incarceration risk. Individuals from lower socioeconomic backgrounds, those living in overcrowded households, renters or non-homeowners, and those with lower educational attainment are more likely to have a family member in prison. These results reinforce the importance of addressing structural inequalities that contribute to cycles of incarceration, particularly in Argentina, where socioeconomic disparities remain a key challenge (Cid Ferreira et al., 2017). Our findings are in line with the latest meta-analytic research, which establishes clear links between family incarceration, psychiatric morbidity, and social risk at the population level (Liu et al., 2025). Therefore, international and national evidence converge to highlight the importance of policy and family-based interventions to mitigate these risks.

This study underscores the importance of shifting from punitive responses to policies that focus on prevention. Government policies and social programs should prioritize early interventions, such as improving educational access in marginalized communities, providing affordable housing options, and ensuring social safety nets for vulnerable populations. Investing in mental health and substance use treatment programs, as seen in other contexts, could also serve as a crucial alternative to incarceration (Butler et al., 2024).

While incarceration directly affects individuals, its consequences extend far beyond those deprived of liberty. Families of incarcerated individuals often experience economic hardship, social stigma, and emotional distress, reinforcing cycles of disadvantage that are difficult to escape (de Mendonça & Barsaglini, 2020; Halsey & Deegan, 2015; Jardine, 2017; Paredes Blandón, 2019). By identifying the key socioeconomic factors associated with incarceration, this study contributes to a broader understanding of how imprisonment affects not only those who are incarcerated but also their families and communities. Recognizing these impacts highlights the urgent need for policies that not only prevent incarceration but also support the reintegration of affected individuals and reduce the burdens placed on their families.

Finally, while machine learning models can offer valuable insights into incarceration risk, they must be implemented carefully to avoid reinforcing systemic biases (Barocas et al., 2019). Future research should continue exploring these relationships using longitudinal data and broader variables to develop more effective, evidence-based interventions. A holistic approach integrating education, social support, and equitable economic policies can help break the link between socioeconomic disadvantage and incarceration, ultimately fostering a more just and inclusive society.

Conflict of Interest

The authors of this article declare no conflict of interest.

Supplementary Materials

Supplementary materials are available at https://doi.org/10.5093/apj2026a18

References

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *The BMJ, 332*(7549), Article 1080. https://doi.org/10.1136/bmj.332.7549.1080

Bagaric, M., Svilar, J., Bull, M., Hunter, D., & Stobbs, N. (2022). The solution to the pervasive bias and discrimination in the criminal justice system: Transparent and fair artificial intelligence. *American Criminal Law Review, 59*(1), 1-57. https://ssrn.com/abstract=3795911

Barnert, E. S., Perry, R., Shetgiri, R., Steers, N., Dudovitz, R., Heard-Garris, N. J., Nia, J., Zima, B., & Chung, P. J. (2021). Adolescent protective and risk factors for incarceration through early adulthood. *Journal of Child and Family Studies, 30*(6), 1428-1440. https://doi.org/10.1007/s10826-021-01954-y

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities.* https://fairmlbook.org/

Berk, R. A., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology & Public Policy, 12*(3), 513-544. https://doi.org/10.1111/1745-9133.12047

Butler, A., Nicholls, T. L., Samji, H., Fabian, S., & Lavergne, M. R. (2024). Mental health needs, substance use, and reincarceration: Population-level findings from a released prison cohort. *Criminal Justice and Behavior, 51*(7), 1054-1071. https://doi.org/10.21428/cb6ab371.bb39d19c

Cadoni, L., Rival, J. M., & Tuñón, I. (2019). *Infancias y encarcelamiento: condiciones de vida de niñas, niños y adolescentes cuyos padres o familiares están privados de la libertad en la Argentina* [Childhood and imprisonment: living conditions of children and adolescents whose parents or relatives are deprived of their liberty in Argentina]. Educa.

https://repositorio.uca.edu.ar/bitstream/123456789/8159/1/infancias-encarcelamiento-condiciones-vida.pdf

Campo Aguzzi, S., Cobbe Peregrin, J. M., Fanton, M., Ollinger, I. P., Zicarelli, J. P., & Cuadra, G. (2019). *Reinserción laboral de ex-reclusos: caso "espartanos"* [Reintegration of ex-prisoners into the labour market: the 'Spartan case'] [Bachelor's Thesis, Universidad Argentina de la Empresa]. Repositorio Institucional UADE. https://repositorio.uade.edu.ar/xmlui/bitstream/handle/123456789/8573/Campo%20Aguzzi.pdf?sequence=1&isAllowed=y

Castro Corredor, D., & Calvo Pascual, L. Á. (2023). Imbalanced machine learning classification models for removal biosimilar drugs and increased activity in patients with rheumatic diseases. *PLOS ONE, 18*(11), Article e0291891. https://doi.org/10.1371/journal.pone.0291891

Cid Ferreira, L., Lorenzo Pisarello, M., & Laks, R. (2017). *Observaciones sobre el delito en relación con los contextos económicos en la Argentina* contemporánea [Observations on crime in relation to economic contexts in contemporary Argentina]. Paper presented at the 1ª Jornada de Estudios Sociales sobre Delito, Violencia y Policía. La seguridad en cuestión. https://sedici.unlp.edu.ar/bitstream/handle/10915/113580/Documento_completo.pdf?sequence=1

Cornell, D. G., Peterson, C. S., & Richards, H. (1999). Anger as a predictor of aggression among incarcerated adolescents. *Journal of Consulting and Clinical Psychology, 67*(1), 108-115. https://doi.org/10.1037/0022-006X.67.1.108

de Mendonça, M. D. G. S., & Barsaglini, C. R. A. (2020). Being related to persons deprived of liberty: repercussions on the experiences of mothers and partners. *Revista Família, Ciclos de Vida e Saúde no Contexto Social, 8*(3), 337-348. https://www.redalyc.org/journal/4979/497963985001/497963985001_2.pdf

Desmond, M., & Gershenson, C. (2016). Housing and employment insecurity among the working poor. *Social Problems, 63*(1), 46-67. https://doi.org/10.1093/socpro/spv025

Fahmy, C., & Mitchell, M. M. (2022). Examining recidivism during reentry: Proposing a holistic model of health and wellbeing. *Journal of Criminal Justice, 83,* 1-13. https://doi.org/10.1016/j.jcrimjus.2022.101958

Farrington, D. P., Ttofi, M. M., & Piquero, A. R. (2016). Risk, promotive, and protective factors in youth offending: Results from the Cambridge study in delinquent development. *Journal of Criminal Justice, 45,* 63-70. https://doi.org/10.1016/j.jcrimjus.2016.02.014

Halsey, M., & Deegan, S. (2015). 'Picking up the pieces': Female significant others in the lives of young (ex) incarcerated males. *Criminology & Criminal Justice, 15*(2), 131-151. https://doi.org/10.1177/1748895814526725

Haque, I. E., & Muniruzzaman, M. (2020). Impoverished living conditions and crime in society: A study on prisoners at Jamalpur District Jail, Bangladesh. *Open Journal of Social Sciences, 8*(3), 33-51. https://www.scirp.org/journal/paperinformation?paperid=98736

Herbert, C. W., Morenoff, J. D., & Harding, D. J. (2015). Homelessness and housing insecurity among former prisoners. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 1*(2), 44-79. https://doi.org/10.7758/RSF.2015.1.2.04

Jaitman, L. (2015). *The welfare costs of crime and violence in Latin America and the Caribbean.* Inter-American Development Bank. https://doi.org/10.18235/0000170

Jardine, C. (2017). Constructing and maintaining a family in the context of imprisonment. *The British Journal of Criminology, 58*(1), 114-131. https://doi.org/10.1093/bjc/azx005

Law 23.7737 (1989). *Narcotráfico* [Drug trafficking]. Sanctioned on 10 October 1989. Published on 11 October 1989 National Gazette. Honourable Congress of the Argentinean Nation. https://www.argentina.gob.ar/normativa/nacional/ley-23737-138/actualizacion

Liu, H., Jingru Li, C., Lam Wong, E., Peng, Z., Wang, A., Chan, S. K. Y., & Hou, W. K. (2025). Family incarceration and mental health among 101,417 affected families: A systematic review and multilevel meta-analysis. *Trauma, Violence, & Abuse, 0*(0). https://doi.org/10.1177/15248380241306353

Mendoza, A. A., & Ewig, G. T. (2020). Youth violence in Latin America. *Oxford Research Encyclopedia of Criminology and Criminal Justice*, 1-31. https://doi.org/10.1093/acrefore/9780190264079.013.579

Niño, C. (2010). *Crime, poverty, and inequity in Latin America: Different sides of the same coin* [Conference paper]. Yale Law School. https://law.yale.edu/sites/default/files/documents/pdf/sela/Nino_Eng_CV_20100503.pdf

Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC, 2021). *Los datos importan: comprender los patrones de encarcelamiento en el mundo* [Data matters: Understanding patterns of imprisonment around the world]. https://www.unodc.org/documents/data-and-analysis/statistics/Data_Matters_1_prison_spanish.pdf

Paredes Blandón, D. (2019). *Dinámicas familiares en relación a la experiencia de privación de libertad* [Family dynamics in relation to the experience of deprivation of liberty]. [Bachelor's Thesis. Universidad de San Buenaventura]. Biblioteca Digital USB. https://bibliotecadigital.usb.edu.co/entities/publication/b972aac5-30c2-4528-b691-529149a603ec

Pedrosa, A. (2020). ¿A quién sancionamos? Un estudio exploratorio en prisiones del contexto español [Who do we sanction? An exploratory study in prisons in the Spanish context.] *Revista Internacional de Sociología, 78*(3), e163-e163. https://revintsociologia.revistas.csic.es/index.php/revintsociologia/article/view/1067/1471

Pérez Palomino, M. (2022). Violencia obstétrica en mujeres privadas de la libertad. Ixaya. *Revista Universitaria de Desarrollo Social, 12*(23), 137-143. https://revistaixaya.cucsh.udg.mx/index.php/ixa/article/view/7732

Politti, M. E., & Delfino, G. I. (2023). Family system and deprivation of liberty: Narratives about the past, present and future. *Interdisciplinaria, 40*(3), Article 12. https://ri.conicet.gov.ar/handle/11336/234738

Procuración Penitenciaria de la Nación. (2022). *Informe Anual 2021. La situación de los Derechos Humanos en las cárceles federales de la Argentina* [Annual Report 2021. The human rights situation in federal prisons in Argentina]. https://ppn.gov.ar/pdf/publicaciones/Informe-Anual-2021-final.pdf

Ríos, V. (2023). Prioridad macroeconómica del gasto público en la administración de justicia en Panamá. *Sapientia, 14*(4), 56-77. https://doi.org/10.54138/27107566.497

Rodríguez Espínola, S. S., Donza, E. R., Filgueira, P., & Paternó Manavella, M. A. (2020). Capacidad de Desarrollo Humano y derechos laborales en la población urbana al final de la década 2010-2019. El desafío de la equidad en la Argentina frente a la pandemia social y sanitaria. EDUCA. https://wadmin.uca.edu.ar/public/ckeditor/Observatorio%20Deuda%20Social/Documentos/2020/2020-OBSERVATORIO-DESARROLLO-HUMANO-TRABAJO.pdf

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Schargrodsky, E., & Freira, L. (2023). Inequality and crime in Latin America and the Caribbean: New data for an old question. *Economía, 22*(1), 175-202. https://www.jstor.org/stable/27302240

Sistema Nacional de Estadística sobre Ejecución de la Pena. (2020). *Informe ejecutivo SNEEP 2020* [SNEEP 2020 Executive Report]. Dirección Nacional de Política Criminal en materia de Justicia y Legislación Penal. https://www.argentina.gob.ar/sites/default/files/2021/10/informe_sneep_argentina_2020_0.pdf

Sistema Nacional de Estadística sobre Ejecución de la Pena. (2021). *Informe ejecutivo SNEEP* 2021 [SNEEP 2021 Executive Report]. Dirección Nacional de Política Criminal en materia de Justicia y Legislación Penal. https://www.argentina.gob.ar/sites/default/files/2022/10/informe_sneep_argentina_2021_3.pdf

Thew, K., & Terry, A. (2025). A review of the ripple effect of parental incarceration: Suggestions for system improvement from neonatal development through adolescence. *International Journal of Undergraduate Research and Creative Activities, 16*(1), Article 8. https://doi.org/10.61809/2168-0620.1350

Travis, J., Western, B., & Redburn, F. S. (2014). *The growth of incarceration in the United States: Exploring causes and consequences.* John Jay College of Criminal Justice. https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1026&context=jj_pubs

Wildeman, C., & Wakefield, S. (2014). The long arm of the law: The concentration of incarceration in families in the era of mass incarceration. *Journal of Gender, Race & Justice, 17*(2), 367-389. https://research.ebsco.com/linkprocessor/plink?id=fed8767e-c559-37af-881f-fd1b20e41820

Appendix

Table A1. Variables in the Model

1	liberty_deprivation Target variable. Refers to whether any household member has experienced deprivation of liberty at any time.	0 = no one in the household has ever been deprived of his or her liberty 1 = someone in the household is or has been deprived of liberty
2	urban_region It highlights socioeconomic disparities across different urban regions with four key urban concentrations:	1 = Autonomous City of Buenos Aires 2 = Greater Buenos Aires 3 = other metropolitan areas 4 = other urban areas in the country
3	professional_status The stratum of household membership through the status, type and occupational qualification, source of income and level of social protection achieved by the primary breadwinner of the household group.	1 = subsistence workers 2 = skilled worker 3 = traditional middle-class 4 = professional middle-class
4	socioeconomic_status Composite index that considers the educational capital of the head of household, the household's access to durable goods and the residential condition of the dwelling, which is recorded into socio-economic strata according to quartiles of the distribution.	1 = very low 2 = low 3 = middle-low 4 = middle-high
5	homelessness Indigent people are those households/individuals whose income does not allow them to acquire the value of the basic food basket. This includes a series of products required to cover a minimum threshold of food needs.	0 = homeless 1 = non-homeless
6	poverty Those households/individuals whose income does not exceed the threshold of the monetary income necessary to acquire the value of a basket of basic goods and services in the market.	0 = poor 1 = no poor
7	children_presence Presence of children in the household.	0 = without children 1 = with children
8	head_of_family The respondent is the head of the household.	0 = non-head of family 1 = head of family
9	sex Gender of the respondent.	1 = male 2 = female
10	education Respondent's education.	0 = up to incomplete high school 1 = complete high school or more
11	employ_quality Respondent's level of employment stability.	1 = stable 2 = precarious 3 = unstable 4 = unemployed
12	employ_status Refers to whether the respondent is employed.	1 = employed 2 = unemployed
13	happiness_deficit It measures the negative perception of mood that produces a feeling of dissatisfaction and sadness in the person.	0 = without deficit 1 = with deficit
14	personal_projects Measures perceived inadequacy in setting goals and objectives in terms of personal well-being.	0 = without deficit 1 = with deficit
15	satisfaction_deficit It measures the deficit in life satisfaction.	0 = without deficit 1 = with deficit
16	neighbourhood_quality Refers to the quality of the neighbourhood in which the person lives.	1 = neighbourhood with urban layout 2 = social housing neighborhood 3 = slums
17	house_type Measures the respondent's house type.	1 = house 2 = apartment 3 = bedroom in a share house 4 = precarious house 5 = bedroom in motel 6 = other
18	house_property Considering different tenure modalities, it refers to how the respondent accesses or owns their dwelling.	1 = homeowners 2 = renter 3 = homeowner but not land owner 4 = de facto occupants 5 = borrowed housing 6 = have housing at place of work 7 = other

Appendix (continued)

Table A1. Variables in the Model (continued)

19	home_type Classifies the household according to its composition and family structure.	1 = single-person household 2 = complete nuclear family household with children 3 = complete nuclear family household without children 4 = incomplete nuclear family household with children 5 = complete extended family household with children 6 = complete extended family household without children and with other relatives 7 = incomplete family household with or without extended children 8 = complete composite family household with children 9 = complete composite family household without children 10 = incomplete family household with or without composite children 11 = non-family multiperson household
20	family_life_cycle Stage of the life cycle in which the family is, considering the age of the children and the family structure.	1 = early-stage families 2 = families with young children 3 = families with school-age children 4 = families with teenage children 5 = families with older children 6 = empty nest families 7 = non-family home
21	age Age in numbers.	Numerical variable
22	people_home Number of people at home.	Numerical variable