Using artificial intelligence in the request of pretrial measures: approaching the uncertainty in recidivism risk

Andrés Felipe Camacho
Daniel Mejía
Lina Navas

This version: Nov - 2023

Abstract

This paper presents an innovative approach to enhance decision-making processes related to pretrial detentions in the Colombian judicial system. By incorporating a machine learning model for predicting recidivism risk, the study aims to provide an objective, impartial, and uniform methodology to assess the benefits of using decision-aid tools. The prosecutors in Colombia face significant challenges in deciding whether a defendant should go to prison based on their recidivism risk. Current practices are scrutinized for their reliance on incomplete and non-uniform information, leading to two critical types of decision-making errors: Type I, involving the unwarranted detention of low-risk individuals, and Type II, the failure to detain high-risk individuals with potentially detrimental effects on public safety. The proposed analytical tool is designed to minimize these errors and biases, ensuring judicial decisions align with principles of equality and impartiality. The anticipated outcome is a more cautious use of detention measures, effectively balancing the rights of the accused with community safety and reducing the number of crimes by recidivists without proportionally increasing detention rates. There is evidence of welfare gains by using this approach when comparing with actual decisions through different levels of leniency in prosecutors.

1. Introduction

Between 2017 and 2018, the National Prosecutor's Office of Colombia charged 230,034 individuals with the alleged commission of a crime before the control of pretrial judges. Of these individuals, restrictive liberty measures were requested for 116,297 (just over 50%) by the case prosecutor, and of these, 102,071 (87.8%) were granted by the control of guarantee judge; in 88.9% of the cases, the measure was restrictive of liberty (76% incarceration and 24% in house arrest). Hundreds of decisions like these, where the Prosecutor's Office requests a restrictive liberty preventive measure for an individual charged with committing a crime and a control of guarantee judge grants or denies it, are made daily in Colombian courts. In making these decisions, prosecutors and judges must evaluate the risk of recidivism of the accused and its consequences on citizen security vis-à-vis the costs (for the accused and society) of sending to preventive detention or house arrest an individual charged with the alleged commission of a crime.

In addition to some material evidentiary elements that indicate the possible relationship of an individual with the commission of a crime, Law 906 of 2004 establishes three grounds that prosecutors must substantiate before control of guarantee judges to request the restrictive measure of the liberty of a person who has been captured for the alleged commission of a crime: (i) ensuring the appearance of the accused and the fulfillment of the sentence, (ii) preventing the obstruction of justice, or (iii) protecting the safety of the community and victims by avoiding the possible risk of recidivism of the individual being processed.

In practice, for most crimes related to citizen security, such as homicides, personal injuries, aggravated theft, etc., prosecutors rely on the third cause (protection of community and victims' safety) to request the intramural preventive measure of an accused individual, which means that defendants have to wait for their trial in jail. Both the request for the pretrial measure made by the prosecutor and the decision to grant or deny it by a judge, are based essentially on a prediction of the possible risk of criminal recidivism of the defendant; this prediction is often made with partial and incomplete information available at the time of the pretrial hearing where the security measure is requested, such as the crime attributed to the accused, the narrative of the facts, the circumstances of the arrest, and in some exceptional cases, information about the criminal history of the individual being charged. With this information, both the prosecutor and the judge must attempt to predict partially and objectively the risks to the safety of the community and the victims associated with a possible risk of criminal recidivism of the individual in question and, based on this inference, decide to request (the prosecutor) and granting (the judge) the custodial measure in a penitentiary center of the country. In essence, judges and prosecutors must carry out an objective assessment exercise that weighs at least two fundamental rights: the right to freedom of an individual accused of committing a crime to be tried at liberty and the right to safety and integrity of the victims and the community against a possible risk of criminal recidivism (i.e., affecting citizen security) by the individual being accused.

On the one hand, if an individual is subjected to intramural detention, the ramifications on their employment prospects, social reputation, and family dynamics can be profoundly detrimental in the short and long term. Conversely, should an individual with a high likelihood of reoffending be released during the trial period, the risk of them perpetrating a new crime poses a serious threat to public safety. This scenario could escalate to the extent of endangering the lives and well-being of others. In determining whether to apply (and approve) the preventive detention of an individual accused of a crime, prosecutors (and judges) face a critical prediction challenge: appraising the probability of criminal recidivism for the defendant. At present, these crucial decisions hinge on predictions made in an ad hoc manner, relying on prosecutors and judges to mentally synthesize partial, inconsistent, and often disparate information specific to each case. Nonetheless, the accuracy of such predictions or inferences is greatly enhanced by the availability of comprehensive, uniform, and dependable information. This approach markedly reduces the likelihood of judicial errors. We will specifically discuss two types of errors, defined as follows:

- Type I Error: imposing a preventive measure on a defendant with a low risk of recidivism.
- Type II Error: failing to impose a preventive measure on a defendant with a high risk of recidivism, affecting the safety of the community and the victims.

Type I Error affects the right to freedom of individuals who, if objectively and impartially evaluated, have a low probability of recidivism and, therefore, should not be deprived of liberty since they pose a danger to society[1]. Type II Error, on the other hand, can affect the community's security by granting freedom to a person with a high risk of criminal recidivism.

In this document, we present the development of a tool that uses machine learning models to help predict, in a more objective, impartial, and uniform manner, the probability of criminal recidivism of individuals who are accused of allegedly committing a crime and over whom prosecutors and judges must decide whether to impose an intramural preventive measure. The goal is to evaluate the performance of these prediction models and compare them to the observed decisions of prosecutors and judges in various dimensions, such as the number of Type I and II errors, the rate of criminal recidivism, and the total number of people who are sent to prison facilities as indictees. As we will show in this document, the use of data analytics tools (machine learning) like the one we propose in this work allows for more objective, uniform, and accurate predictions about the probability of recidivism of individuals who are captured and prosecuted, thereby reducing the errors described above. The available evidence that we analyze in this work shows that there are significant biases and errors in the decision-making process regarding preventive measures, leading to a substantial number of people who are accused but have a low probability of criminal recidivism, ending up being covered by intramural preventive measures (Type I Error).

Additionally, we demonstrate that at the right end of the recidivism risk distribution, a considerable number of Type II errors occur; that is, charged individuals with a very high recidivism risk are not covered by intramural detention measures. This document also shows that the use of data analytics tools, such as the one we propose, could help the criminal justice system minimize these errors and biases while ensuring that the information used by prosecutors and judges when making this decision is consistent across all hearings, which in turn guarantees adherence to the principles of equality and impartiality, the guiding principles of the Code of Criminal Procedure. Finally, this recidivism risk prediction tool can significantly reduce the number of crimes committed by repeat offenders without increasing the number of people subject to intramural detention measures. In other words, using mechanisms like the one proposed in this work would lead to a more rational, proportional, and efficient use of intramural detention measures and scarce prison capacities.

The risk prediction models we propose in this document use all available data to generate an individual-level prediction of the risk of criminal recidivism for individuals charged with the alleged commission of a crime. To construct these criminal recidivism risk prediction models, we use individual-level information from the National Police (total arrests), the information from the case management and information system - SPOA of the Prosecutor's Office (cases, charges, security measures, convictions, etc.), and the information from the National Penitentiary and

[1] Unless, of course, the request of the measure is made arguing the causes (i) and (ii) described above: to ensure the appearance of the accused and the compliance with the sentence, and/or to prevent the obstruction of the administration of justice.

Prison Institute - INPEC (number of entries to prison centers, days in prison, participation in resocialization programs, etc.). With the available information on individuals charged between 2012 and 2017, we built algorithms to predict the risk of recidivism (specifically, we used extreme gradient boosting decision tree models based on Friedman (2001) and Kleinberg et al. (2018)) that allow us to evaluate the performance of this recidivism risk prediction tool and compare it with the decisions made by judges and prosecutors in those years.

Using risk prediction models for recidivism that we propose to support the request (and granting) of intramural assurance measures can generate several benefits, both in terms of justice and efficiency in using scarce prison resources and in reducing criminal recidivism. First, from the perspective of justice administration and ensuring the proportionality of the measures adopted, the use of these types of tools can significantly reduce the number of people who, having a low objective risk of criminal recidivism, are covered by intramural assurance measures (Type I error) and the number of people who, having a high risk of recidivism, are not covered with intramural measures (Type II error). Specifically, the estimates indicate that, of the 36,823 identified individuals who were charged in 2018 and had a low risk of recidivism, 54.3% (19,982 individuals) were requested by the case prosecutor to be subjected to intramural assurance measures. Of this group, 45.7% (16,841 individuals) were granted these measures by the judge. On the other end of the recidivism risk distribution, of the 36,794 charged in 2018 with a relatively high risk of recidivism, the case prosecutor did not request intramural assurance measures for 14,463 (39.3%). Of the 22,331 individuals charged in 2018 with a high recidivism risk for whom prosecutors requested such measures, the judge did not grant them to 8,584 (38.4%). These numbers reflect the magnitude of Type II and Type I errors, respectively, that are made when requesting and granting intramural detention measures in the Colombian criminal justice system.

The results are even more surprising when we examine the two extremes of the recidivism risk distribution. In the first decile of the recidivism risk distribution (i.e., the least risky 10% of charged individuals in terms of criminal recidivism risk), 43.7% were requested to be subjected to assurance measures, and of these, 50.3% were actually granted such measures by a judge. In the two years following the charge, the observed recidivism rate for individuals in this first decile is only 3.57%. On the other end, in the riskiest decile of the distribution, prosecutors requested intramural measures for 62.2% of the charged individuals; of these, 56.5% were granted such measures by the judge. When we look at the recidivism rate of individuals belonging to the top 10% of defendants with the highest risk of reoffending, 65.2% of these re-offend within two years of being charged. These results indicate that the decisions currently made by prosecutors and judges are not based on an objective prediction of the risk of recidivism. As we will show later, the so-called current offense bias explains a significant part of these decisions (see Sunstein, 2018). This bias manifests in both prosecutors and judges processing high-risk individuals as if they were low-risk when the current crime for which they are being charged is a relatively minor offense, and vice versa. The use of risk prediction models partially corrects these biases. It utilizes the criminal history of the accused individuals to predict their objective risk of recidivism, giving a more appropriate weighting to the severity of the current offense.

Secondly, from an efficiency standpoint, by helping to address the composition issue of individuals subjected to intramural preventive measures, the proposed tools can significantly reduce the number of crimes committed by repeat offenders. Specifically, the results we will present in this paper indicate that, while maintaining the same number of people annually subjected to intramural preventive measures (approximately 24,332 in 2018), this tool can reduce crimes committed by repeat offenders by up to 25%. In other words, using these tools could reduce criminal recidivism by one-quarter without increasing the number of prison spaces for defendants (merely changing who is sent to pre-trial detention and who is not). Alternatively, while maintaining the same level of criminal recidivism, the number of people sent to intramural measures could be reduced by 36%, approximately 9,000 prison spaces per year. Again, this is without increasing the level of crimes committed by repeat offenders. Third, the use of this tool would ensure that in hearings requesting preventive measures, the same information about criminal records and annotations of the accused individuals is used, thus avoiding disparate treatment between similar cases and guaranteeing the principles of equality and impartiality (guiding principles of the Colombian Criminal Procedure Code, numbers 4 and 5).

The rest of the document is organized as follows: Section 2 describes the context of the Colombian Penal System and the main descriptive statistics of the official data. Specifically, this section provides an overview of requests and grants of intramural preventive measures and descriptive statistics of the main patterns of criminal recidivism in Colombia. Section 3 describes the methodological aspects of the four risk prediction models for recidivism in: (i) any crime, (ii) crimes against life and integrity, (iii) property crimes, and (iv) other crimes such as drug trafficking and manufacture, carrying and trafficking of weapons, and conspiracy to commit crime. The fourth section describes the main results, emphasizing the benefits of using this tool in potentially reducing errors and biases in preventive measure decisions and its potential effect on crime reduction. The fifth section presents the main conclusions.

2. Literature Review

The academic literature on statistical prediction tools in criminal justice systems is extensive and dates back to the 1990s. Most studies have focused on using these tools in the United States, concentrating on the potential gains from adopting them in decisions such as pre-trial detention, parole, bail use, and sentence duration determination. One of the pioneering studies (Berk et al., 2009) developed a criminal recidivism risk prediction tool for the Philadelphia Adult Probation and Parole Department (APPD). This tool aimed to classify captured offenders into three categories according to the risk of criminal recidivism they represented (by type of crime) and, based on this classification, helped determine the terms and conditions of parole to be granted to convicted offenders. For example, for those individuals considered high risk for recidivism in violent crimes, the requirement to report more frequently to an APPD supervision officer was imposed, compared to those considered low-risk. An experimental design impact assessment of this tool established that the system reduced the burden on parole without significantly increasing criminal recidivism rates. In a more recent study, Berk (2017) evaluated the effect of introducing machine learning tools that predict the probability of future re-arrest on the decisions to grant parole to individuals accused of committing violent and non-violent crimes, also in the state of Pennsylvania.

Although the introduction of these prediction tools did not change the overall rate of parole usage, it seems to have affected the composition of the convicted individuals to whom this measure was granted, distinguishing between offenders whom the algorithm predicts will be re-arrested for violent crimes, and those it predicts will be re-arrested for non-violent crimes. Additionally, introducing these tools reduced the re-arrest rate for violent and non-violent crimes. These prediction tools have become very common in various judicial districts in the United States. Closely related to the tool we present in this article are the risk prediction tools for recidivism developed more recently in different U.S. judicial districts to determine which individuals arrested for the alleged commission of a crime should be held in pre-trial detention before trial. The Laura and John Arnold Foundation has promoted and funded the development and use of these tools, which combine nine factors to predict whether an individual arrested will re-offend or fail to attend court trial hearings if released.

The risk prediction tool for criminal recidivism we developed in this article for the Colombian case is methodologically based on the work of Kleinberg et al. (2018). In particular, the authors use machine learning models to predict the risk of defendants in the state of New York failing to appear at trial hearings, and with these predictions, compare different outcomes such as incarceration rates, criminal recidivism, racial biases, etc., between what the tool would recommend regarding whether to release or detain an individual accused of committing a crime or to leave them on parole. The authors demonstrate that using this prediction tool can yield significant gains in reducing incarceration rates without increasing crime and in reducing crime without increasing incarceration rates. Specifically, the authors demonstrate that using this tool can reduce crime by 24.7% without increasing incarceration rates. Alternatively, incarceration rates can be reduced by 41.9% without increasing crime. Similar to the recidivism risk prediction tool we developed in this article, what underlies these results is a change in the composition of individuals who are incarcerated rather than in the number of individuals who are incarcerated. In other words, the use of these tools, by correcting biases in judges' decisions and thereby reducing Type I and Type II errors (i.e., reducing the number of low-risk individuals sent to pre-trial detention and increasing the number of high-risk individuals covered by this restrictive measure of freedom), has the potential to generate significant welfare gains in terms of reducing crime and incarceration rates.

A recent study evaluating the real-life implementation of these particular statistical tools for predicting recidivism risk to support judges in their sentencing decisions shows more ambiguous results (Stevenson and Doleac, 2019). Specifically, the authors show that although judges' decisions are indeed affected by the recidivism risk predicted by these algorithms, leading to longer sentences for offenders with a higher risk of criminal recidivism and shorter for those with a lower risk, this did not result in a statistically significant reduction in recidivism. The authors explain that this is due to the discretion judges exhibit when using the results of these algorithms, being more lenient with younger individuals despite their high risk of recidivism. These findings draw attention to a fact that the construction of these algorithms can overlook and which can lead to an overestimation of the potential gains in terms of reducing crime or incarceration rates, namely, the real-life objectives that judges have when making decisions about pre-trial detention or the duration of sentences may go beyond the reduction of crime or incarceration rates. In particular, the authors argue that when making these tools available to judges and prosecutors to manage real judicial cases, it is essential to consider how humans interact with these algorithms when making decisions.

Finally, a recent issue addressed in the literature on statistical tools for predicting recidivism risk in judicial decisions concerns potential changes in levels of discrimination and racial disparities resulting from judicial decisions that use these predictive statistical tools. The main argument of Kleinberg et al. (2018) is that while proving patterns of discrimination in human decisions is usually very difficult, if not impossible, when algorithms are involved in the decision-making process, proving discrimination should be more feasible, and they can be designed in such a way. Reducing discrimination can be achieved by regulating the process through which these algorithms are designed. This approach would lead to greater transparency in judicial decisions while explicitly highlighting the policy trade-offs faced in decision-making.

3. Context and Descriptive Statistics

According to the current legal framework in Colombia, within the first 36 hours after a person has been arrested (either in flagrant or by judicial order) for the possible commission of a crime, the prosecutor handling the case must legalize the arrest, charge the individual, and decide whether to request an intramural preventive measure or another restrictive measure on the freedom of the accused. The case prosecutor makes this request to a guarantee control judge. It is essential to clarify that the purpose of the hearings for legalizing an arrest and requesting preventive measures is not to establish the guilt of the individual arrested for the alleged crime they were charged with. Instead, the objective is to ensure that their procedural rights are respected, such as verifying the legality of the arrest and determining (during the hearing for requesting preventive measures) whether the individual should remain in custody while the investigation and trial stages proceed.

In the hearing for requesting a preventive measure, if the prosecutor requests a restrictive measure on the accused's freedom, their argument must be based on three possible grounds established in Law 906 of 2004: (i) to ensure the accused's appearance and compliance with the sentence, (ii) to prevent obstruction of the administration of justice, and (iii) to protect community and victim safety, avoiding the potential risk of recidivism of the individual being processed. In practice, however, for most public security crimes, prosecutors rely on the third ground to request a restrictive freedom measure from the judge. For this reason, in the analyses that follow, we will focus on this ground to examine the requests for intramural preventive measures and the subsequent observed patterns of criminal recidivism. In the hearing for requesting a preventive measure, the prosecutor and judge have firsthand information about the crime the arrestee is accused of, the circumstances of the arrest, and, occasionally, partial and incomplete information about the criminal history of the processed individual. Based on this information, the prosecutor must infer the probability of the accused's criminal recidivism and, based on this, request from the guarantee control judge the appropriate preventive measure for the accused. This prediction varies according to the prosecutor and judge's information about the accused. Still, it can also be affected by secondary aspects such as how lenient the prosecutor and judge are processing the case or the appearance and behavior of the accused during the hearing.

It is worth noting at this point that although in Colombia, criminal law is based on the act, meaning that a person investigated for the alleged commission of a crime can only be judged for the act committed and not for their past criminal history, this only applies to the investigation and trial stages of the criminal process, and not to the decision on preventive deprivation of liberty. In other words, in the decision on preventive measures that restrict freedom, the criminal history of the accused individuals can be taken into account when assessing the dangerousness of the processed individual and the potential risk of criminal recidivism. Therefore, using predictive statistical tools that systematically utilize the accused's criminal history does not violate any fundamental rights.

3.1 Data

The data used to predict recidivism corresponds to the criminal information of each person at three different links in the chain of criminal policy. In the first stage, there is data from the National Police's Statistical, Criminal, Contraventional, and Operational Information System (SIEDCO, in Spanish), which records the arrest warrants and arrests in flagrante carried out by the National Police since 2004. This information is recorded in a database by the Directorate of Criminal Investigation of the National Police of Colombia following strict established protocols and endorsed by the National Department of Statistics (DANE) (Buitrago et al., 2015). The second source is the database on investigations and actions in the Oral Accusatory Penal System (SPOA) of the Attorney General's Office (FGN in Spanish), which records information on all criminal proceedings in Colombia and the actions taken by the Prosecutor's Office in each of the investigations since its implementation in 2005[2]. Although the SPOA was initially designed as a management tool within the Attorney General's Office, where prosecutors must record all actions carried out in the context of a criminal investigation, in recent years, this entity has made

[2] Although the SPOA [Oral Accusatory Penal System] began in 2005, it was not until 2008 that it was extended throughout the entire country.

significant efforts to consolidate the SPOA as a system that tracks criminal records at the national level.

In order to construct the recidivism risk prediction tool presented in this work, the SPOA database allows for extracting individual-level records of any criminal investigation a person has undergone as well as details about alleged crimes, previous preventive measures, accusations, and convictions, among others. The third and final source of information used corresponds to the Comprehensive Systematization of the Penitentiary and Prison System (SISIPEC) of the National Penitentiary and Prison Institute (INPEC), which contains information on individuals with preventive measures or convictions, whether they are in intramural or house arrest, covering the population as of 2008 (stock) and the flows of people from then onwards until 2019. The SISIPEC system is the primary source of information for the penitentiary, prison, and judicial authorities regarding the conditions of confinement of each of the individuals deprived of liberty who are under the custody of Colombia's Penitentiary and Prison System. Graph 1 shows these three stages of information gathering and institutional databases.

Graph 1. Sources of Information

While unifying the databases, information from the SPOA database was cleansed so that each observation represents a criminal event, defined as the unique combination of the person's identification number and the date of the incident. For each criminal event, the criminal history recorded in SPOA before the current event (both charged and uncharged), the type of crimes committed in the event, the person's age at the time of the crime, and the gender were calculated.

The database of events in SPOA was consolidated with information from the other two data sources. From SIEDCO, the number of arrests each person had before the SPOA event was obtained[3]. Lastly, from SISIPEC, information is available on the number of times and the accumulated days the accused has been in prison before the current crime event, whether they have engaged in any sentence reduction activities (study, work, or teaching) during those events, as well as the proportion of time the person spent in these activities relative to the total time they were incarcerated, and the proportion of SISIPEC events in which they received visits.

To ensure that we only take into account recidivism in events that actually occurred and in which the prosecutor considered sufficient evidence in the case, only the records from the SPOA database

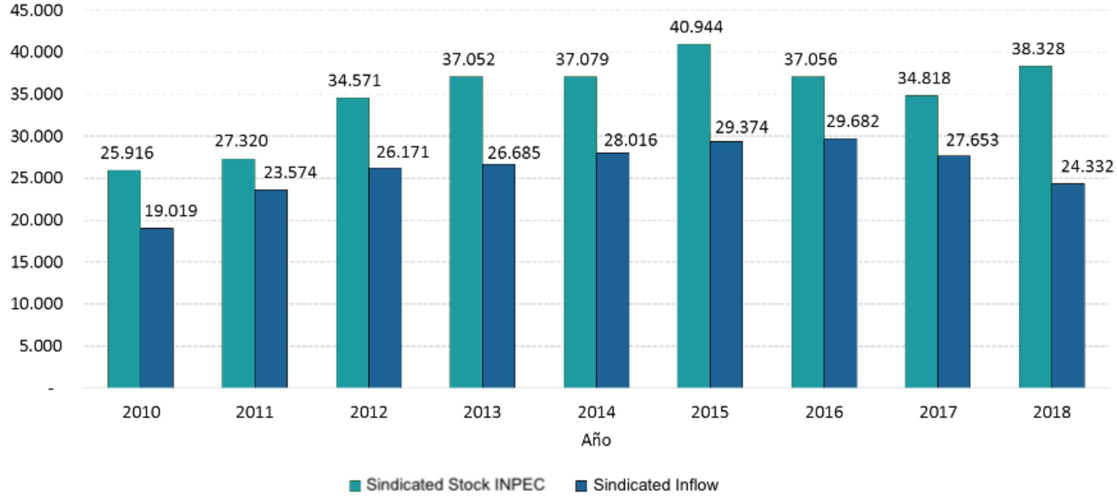
[3] Not all arrests result in an SPOA event

where the accused was charged are retained. As a result, we obtain a large database with 5,943,122 events from January 2005 to March 2019, of these events, 4,183,958 are identified with the individuals' citizenship identification numbers (70.4%). Of this total, 1,874,880 do not record a date of birth, resulting in a final database with 2,309,078 events, of which 994,141 (43%) are charged. The resulting 994,141 events correspond to 744,255 individuals, of which 67% have arrests recorded in SIEDCO between 2004 and 2019. Additionally, 13.32% of the individuals in the final database have records in the SISPEC system of INPEC.

3.2 Descriptive Statistics on Intramural Preventive Measures

In 2018, an average of 173 hearings for imputing charges and requesting preventive measures were held per business day in Colombia. Additionally, between 25,000 and 30,000 individuals were annually subjected to intramural preventive measures between 2010 and 2018[4], and the total number (stock) of defendants in national prisons in the country has fluctuated on average in recent years between 35,000 and 40,000 (see Graph 2).

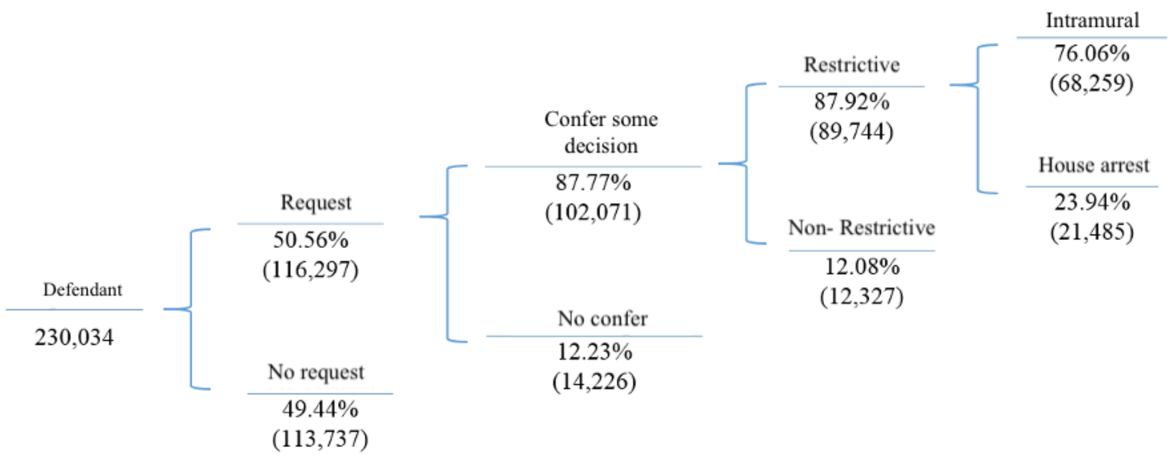
Graph 2. Population Charged with Intramural Measures in National Prisons



Within the approximately 230,000 individuals charged by the prosecutors between 2017 and 2018, Graph 3 shows that in 50.6% of the cases the prosecutor requested a preventive measure, and in 87.8% of these cases, the judge granted it. Of the measures granted, 87.9% were deprivation of liberty, and of these, in 76.1% of the cases, the measure granted was intramural, and in the remaining 23.9%, the restrictive measure of liberty was house arrest. In summary, of the nearly 230,000 individuals charged between 2017 and 2018, just over 68,000 ended up with an intramural measure (29.7% of the cases).

[4] This does not include the number of indicted individuals who are in the district jails of major cities.

Graph 3. Overview of Requests and Grants of Preventive Measures



Source: Author's calculations based on SPOA information for 2017 and 2018

Tables 1 and 2 compare individuals with different measures requested from prosecutors and granted from judges, respectively, between 2013 and 2016. Table 1 shows a higher proportion of individuals for whom prosecutors requested preventive measures committed more serious crimes. While 9.3% of the individuals, for whom a preventive measure was requested, were being processed for homicide, only 2% of those for whom no measure was requested, were being processed for the same crime. The same pattern is observed for those charged with sexual crimes (4.7% vs. 2.4%) and for the crime of conspiracy to commit crime (12.2% vs. 2.3%).

Regarding the severity of the crime for which those individuals with a measure requested, the average penalty is 180 months, compared to a lower average penalty (118 months) for the current crime of those individuals for whom the prosecutor did not request a measure. The average severity of previous crimes committed by individuals for whom prosecutors requested preventive measures is lower than those charged for whom the prosecutor did not request the measure. The severity is measured by the average penalty of the crime established in the penal code. On average, the severity for those for whom a measure was requested was 169 months, while for the others the average severity was 186 months.

This is an early indication of what is known in the literature as the 'current offense bias' in the request for preventive measures. As the aim of the pretrial hearing is to estimate the risk of recidivism and not the severity or guilt of the accused, it would be expected in principle that the severity of the criminal history of individuals for whom prosecutors request preventive measures be higher, regardless of the particular crime they are being investigated for at that moment. However, the evidence in Table 1 suggests the opposite: individuals for whom a preventive measure is requested have an average severity of the crimes they have committed in the past that is lower than that of those individuals for whom no measure is requested.

Table 2 contains similar information but compares the accused individuals to whom judges granted intramural preventive measures with those who were not granted such measures, either because the judge did not grant the measure, or did grant it, but the restrictive measure of freedom established did not involve intramural detention. As in Table 1, in the case of requests by prosecutors, judges grant more intramural measures to people being judged for more serious crimes such as homicides or sexual offenses. However, contrary to requests by prosecutors, judges decide to grant preventive measures to individuals who have a higher historical average severity of crimes committed in the past compared to those who are not granted the measures, again, measuring the historical severity of previous crimes by the average penalty (186 months for those given the measures vs. 141.5 for those not).

Table 1. Request of preventive measures (2017-2018)

	No requested	**Requested**
Defendants	68,897	89,172
Theft*	30.32%	30.02%
Homicide*	2.06%	10.59%
Injuries	2.32%	1.53%
Sexual Offenses *	2.06%	4.80%
Narcotics	17.66%	24.20%
Domestic Violence *	13.49%	7.80%
Conspiracy *	2.28%	10.17%
Current crime sentences *	123.41	200.46
Average sum of historical crime sentences *	163.55	193.33

Table 2. Granting of Preventive Measures, given the request (2017-2018)

	No granted	**Granted**
Defendants	37,187	51,985
Theft	30.05%	29.99%
Homicide*	5.45%	14.26%
Injuries	1.49%	1.57%
Sexual Offenses *	1.99%	6.81%
Narcotics *	23.67%	24.58%
Domestic Violence *	11.14%	5.41%
Conspiracy *	6.21%	13.01%
Current crime sentences *	165.12	225.74
Average sum of historical crime sentences *	152.19	222.77

To document the current offense bias more formally, Table 3 estimates a simple regression where the dependent variable is the request (column 1) and the granting (column 2) of preventive measures based on, first, the penalty for the current crime for which each individual is being investigated, and second, the average sum of the penalties for historical crimes previously committed. The results indicate that both independent variables are significant, but the severity of the current crime holds more weight than that of historical crimes in determining the penalty. In the case of the request, a 10% increase in the penalty for the current crime increases the likelihood of a request by the prosecutor by 2.5 percentage points (pp), while a 10% increase in the average penalty for historical crimes only increases the likelihood of the prosecutor requesting intramural preventive measures by 0.01 pp. In the case of the granting of intramural measures by judges, the severity of the current crime also matters more (a 10% increase in the penalty for the current crime increases the probability of granting by about 1 pp) than the severity of the accused's criminal history (a 10% increase in the average penalty for historical crimes increases the probability of granting by 0.2 pp). This suggests that judges, unlike prosecutors, may be more influenced by the historical severity of an accused's criminal record when deciding to grant preventive measures. It indicates a divergence in the criteria used by prosecutors and judges in assessing the risk posed by individuals, with judges potentially giving more weight to past criminal behavior than the current offense being prosecuted.

Table 3. Logistic Regression of Request and Granting (2017-2018)

	(1) **Requested**	**(2)** **Granted intramural**
Current Crime Sentences	0.254***	0.144***
	(0.002)	(0.003)
Average sum of historical crime sentences	0.012***	0.027***
	(0.000)	(0.001)
Observations	158,069	89,172

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

3.3 Descriptive Statistics on Criminal Recidivism in Colombia

Of the 84,536 individuals charged for allegedly committing a crime in 2018, 51% had previous annotations in the SPOA. Of the more than 40,000 people charged in 2018 who had prior annotations in the SPOA, 3% were for homicide, 1% for sexual offenses, 10% for injuries, 8% for domestic violence, 24% for theft, and 20% for drug offenses (Table 4). When disaggregating the information by the crime charged in 2018, in virtually all cases (with the sole exception of sexual offenses), more than 45% of the charged individuals had previous annotations in the SPOA. For those charged with homicide, 9% had prior annotations for the same crime, 13% for personal injuries, 16% for theft, and 10% for drug trafficking and manufacturing. Table 4 shows that individuals charged with theft are those who most proportionally have previous events in the SPOA (62% of the 26,542 charged for this crime), and a significant majority of these (47%) had prior annotations for the same crime.

Table 4. Proportion of crimes committed by repeat offenders accused

	Charged events	%with previous events	Homicide	Sexual	Injuries	Domestic Violence	Theft	Narcotics
Total	84,536	51%	3%	1%	10%	8%	24%	11%
Homicide	5,311	50%	9%	1%	13%	7%	16%	10%
Sexual Offenses	2,849	33%	1%	8%	7%	7%	6%	3%
Injuries	1,662	49%	3%	1%	15%	12%	19%	6%
Domestic Violence.	9,551	48%	1%	2%	13%	21%	10%	4%
Theft	26,542	62%	3%	1%	12%	7%	47%	10%
Narcotics	16,438	47%	4%	1%	8%	5%	14%	24%

The data from Table 4 show that a very high percentage of the accused individuals are repeat offenders, and these data can be taken as an approximation of the percentage of total crime that repeat criminals commit.

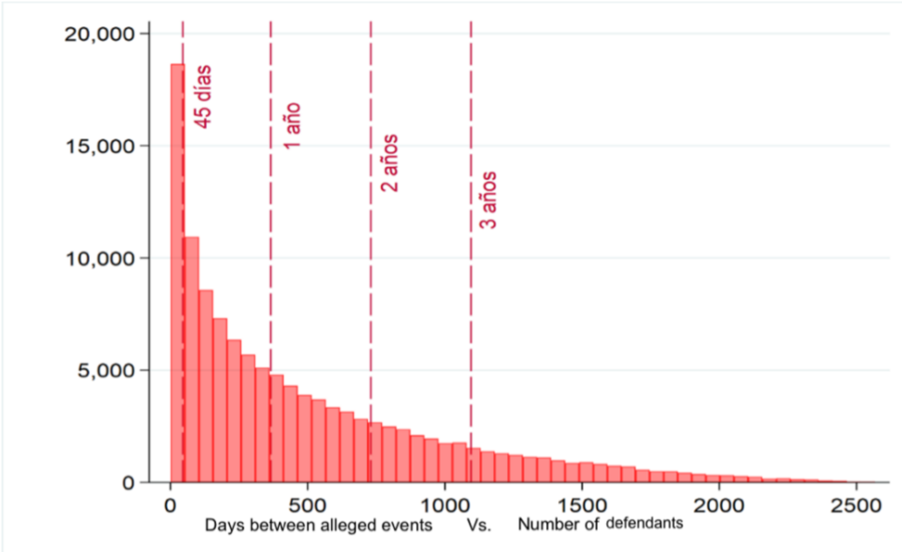
Table 5 presents similar information except using the arrest data from the National Police, instead of the charge information from the Attorney General's Office. Of the more than 215,000 individuals arrested by the National Police in 2018, 47% had previous arrests, of which 2% were arrested for homicide, 1% for sexual offenses, 5% for personal injuries, 3% for domestic violence, 19% for theft, and 22% for drug trafficking or manufacturing. Similar to the information presented for the charges by the Attorney General's Office in Table 4, in the case of arrests, the highest rates of recidivism in arrests are observed for the crimes of theft, drug trafficking, and manufacturing.

Table 5. Proportion of crimes committed by repeat offenders arrested

	Arrests	%with previous arrests	Homicide	Sexual	Injuries	Domestic Violence	Theft	Narcotics
Total	215.590	47%	2%	1%	5%	3%	19%	22%
Homicide	5.508	49%	14%	1%	7%	2%	14%	16%
Sexual Offenses	5.412	25%	0%	10%	3%	2%	4%	5%
Injuries	16.907	33%	1%	1%	10%	4%	11%	11%
Domestic Violence	10.734	35%	1%	1%	7%	11%	9%	11%
Theft	43.799	56%	1%	0%	6%	3%	42%	19%
Narcotics	58.734	56%	1%	0%	4%	2%	15%	45%

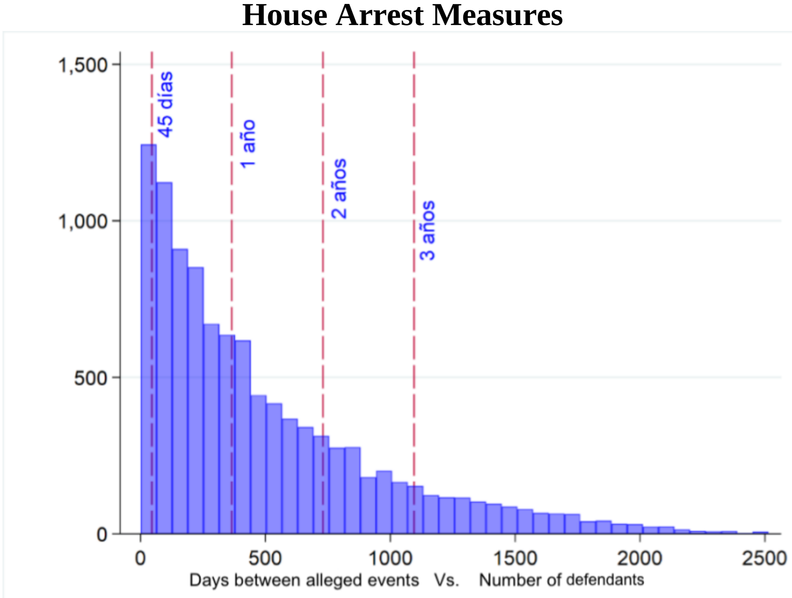
Graph 4 shows the distribution of the number of days elapsed between charged events for the 120,716 repeat offenders who had more than one charge event between 2012 and 2017. Of these recidivism events measured by charged crimes, 52% occur within the first year after a charge, and 74% within the first two years. 13.9% of the individuals who re-offend do so within the first 45 days after being charged.

Graph 4. Distribution of the Number of Days Between Charged Events



Of the 51,745 individuals who between 2012 and 2017 were subjected to house arrest measures, 10,349 reoffended (20%). Although the information systems do not allow to precisely establish whether these recidivism events occur within the period in which individuals are under house arrest, it is noteworthy that while the average duration of a house arrest measure is 953 days, 51% of recidivism events occur within the first year after the judge decrees this measure. This suggests that in many cases, criminal recidivism can occur during the duration of the house arrest measure (Graph 5).

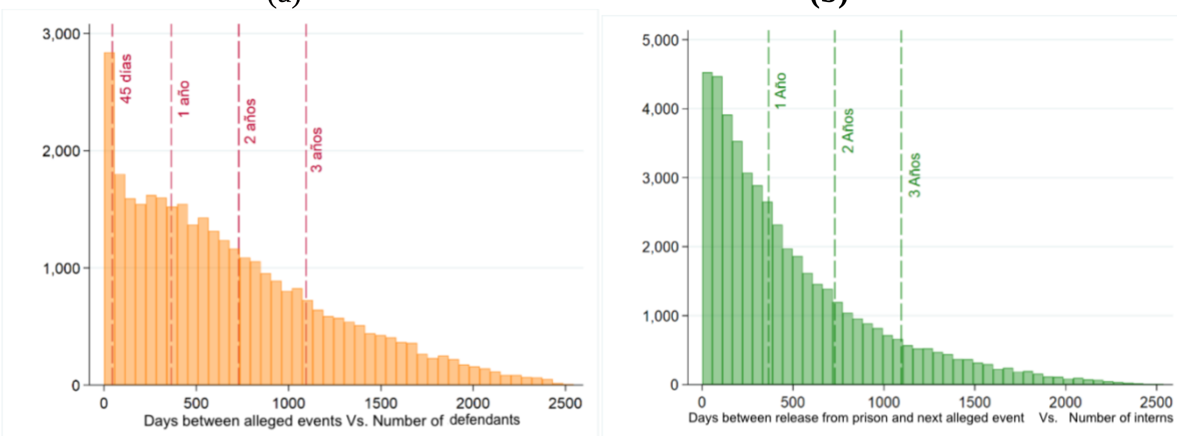
Graph 5. Distribution of the Number of Days Between Charged Events for Events Granted House Arrest Measures



When analyzing the criminal recidivism rates of individuals who have been subjected to intramural preventive measures or convicted in prison facilities in the country, the patterns are similar to those described earlier. First, of the 187,350 individuals who went through pre-trial detention between 2012 and 2017, 33,723 reoffended upon release (18%). Of these, 45% did so within a year of leaving pre-trial detention, and 61% within two years (Graph 6a). Second, for individuals who were released from a penal institution after having been charged or convicted (221,000 individuals between 2012 and 2017), 21.5% (45,517 individuals) reoffended (recidivism measured by charged events); 50% within the first year of leaving prison and 75% within the first two years (Graph 6b).

Graphs 6a and 6b. Distribution of Days Between Recidivism Events

(a) **(b)**



4. Methodology

We use machine learning tools to predict the likelihood of criminal recidivism of an individual charged with the alleged commission of a crime. In artificial intelligence, algorithms are capable of analyzing large volumes of information. These algorithms undergo training to unravel behavior patterns and understand which variables, in our case related to a person's criminal history such as previous arrests, previous charges, convictions, etc., allow for a more truthful and precise prediction of future criminal recidivism. The problem of predicting the risk of recidivism can be addressed using supervised classification algorithms, where, through a set of a priori classified data, patterns and characteristics are recognized that predict the class to which a new observation belongs. In other words, the algorithm fits a function $m(Xi)$ that relates the set of independent variables or inputs (all available variables related to an individual's criminal history) to predict the probability that the binary response variable y_i, in our case criminal recidivism, takes the value of 1. For each accused individual i, given their characteristics, their criminal history, and the crime they are being accused of, the goal is to predict the likelihood of their recidivism in the following two years:

$$\Pr(y_i = 1|X_i) = m(X_i) \qquad (1)$$

Specifically, we use Extreme *Gradient Boosting Decision Trees* prediction models. In machine learning models based on decision trees, the space formed by independent variables is segmented through binary partition rules, which can be summarized in a tree diagram. Each branch symbolizes a space partition, and the goal is to find partitions that minimize a loss function $L(y_i, m(X_i))$. The predicted value usually corresponds to the mean of the response variable y_i in each resulting region or leaf of the decision tree.

An extension of tree-based methods is the Gradient Boosting algorithms introduced by Friedman (1999), which consist of the sequential construction of multiple decision trees. In each iteration of the sequential process, more importance is given to observations predicted with less precision in previous iterations, using a negative gradient in the loss function. The final result corresponds to a weighted average of the predicted value in each iteration. As explained in a previous section, these algorithms have been previously used in the literature in similar contexts, for example, to estimate the risk of a defendant failing to appear in court in New York City (Kleinberg et al., 2018). In this work, we use the method of Extreme Gradient Boosting, developed by Chen et al. (2019), to implement the algorithms introduced by Friedman (1999), which have greater computational efficiency.

5. Definition and Evaluation of the Predictive Model

The analysis is divided into two stages: training the algorithm and evaluation. In the algorithm construction process, the sample is randomly split into 70% for the training stage and 30% for evaluation. This division prevents the problem of overfitting, which occurs when the model fits very well to the data with which it was trained and predicts with high precision the risk for these observations; however, its performance is poor when trying to predict the risk for a different sample (James et al., 2015).

The algorithm's complexity in the training stage depends on parameters such as the depth of the trees, the number of trees, and the weighting scheme. Following the strategy described by Kleinberg et al. (2018), these parameters are selected through cross-validation over five groups. Once selected, they are used to estimate the model with the entire training sample. This process is carried out using the free software R, through the “Caret” package, which specializes in implementing machine learning algorithms.

To predict the risk of recidivism from different dimensions, four models are proposed using the same set of independent variables (X_i) to predict different forms of recidivism (y_i) over a two-year time horizon. In the first model, recidivism is predicted from a general perspective, where the response variable (y_i) takes the value of 1 if the accused reoffends in any crime. In the second model, y_i takes the value of 1 if the next crime committed by the accused is economically motivated (theft, fraud, etc.). In the third model, y_i takes the value of 1 if the next crime committed by an individual is violent (homicide, personal injury, sexual violence, etc.). Finally, in the fourth model, y_i takes the value of 1 if the next crime is trafficking, manufacturing and carrying weapons or drugs, or conspiracy to commit a crime. Annex I provides a more detailed classification of property and violent crimes used to build the models.

For the evaluation stage, commonly used metrics in machine learning algorithms are considered: accuracy, sensitivity, specificity, and area under the ROC (receiver operating curve) or AUC. Given a threshold or cut-off in the probability predicted by the model, accuracy measures the percentage of observations correctly classified (2), sensitivity the percentage of hits in the positive class (3), and specificity the percentage of hits in the negative class (4). The area under the ROC curve (AUC) summarizes the previous metrics into a single measure of the model's overall performance, being the relationship between sensitivity and specificity for all possible threshold values. These measures are on a scale from zero to one, and the model with the value closest to one in each of these metrics is sought.

$$Accuracy = \frac{TP + TN}{P + N} \tag{2}$$

$$Sensitivity = \frac{TP}{P} \tag{3}$$

$$Specificty = \frac{TN}{N} \tag{4}$$

Where P is the number of observations that belong to the positive class $(y_i = 1)$ and N is the number of observations that belong to the negative class $(y_i = 0)$. Given a threshold t, TP (true positive) is the number of observations from the positive class that are correctly classified by the model $(\Pr(y_i = 1|X_i) \geq t|y_i = 1)$ and TN (true negative) is the number of observations from the negative class correctly classified $(\Pr(y_i = 1|X_i) < t\ |y_i = 0)$

Finally, from metrics (3) and (4), it is possible to calculate the probability of classifying as a recidivist an individual who will not re-offend, defined as Type I error (5), and the probability of classifying as non-recidivist an individual who will re-offend, defined as Type II error (6).

$$Type\ I\ error = 1 - Specificity \tag{5}$$

$$Type\ II\ error = 1 - Sensitivity \tag{6}$$

These metrics provide a comprehensive view of the model's performance, including its accuracy, error rates, and its ability to distinguish between different classes of outcomes. This evaluation framework is critical for assessing the effectiveness of a predictive model, especially in high-stakes contexts like criminal justice, where the consequences of misclassification can be significant.

6. Results

Table 6 presents a summary of the performance metrics for the four models of recidivism risk prediction. The economic crime model achieves the best fit in prediction, with the values of AUC, accuracy, sensitivity, and specificity closest to one and the lowest Type I and Type II errors. In terms of the area under the ROC curve, the violent crime model performs lower compared to the other models, with an AUC of 0.74; however, the fit in prediction is still good when compared to the result of the study conducted by Kleinberg et al. (2018) for the state of New York, where they sought to predict the risk of a defendant failing to appear at judicial hearings (Failure to Appear, or FTA) or being re-arrested, with an AUC of 0.707.

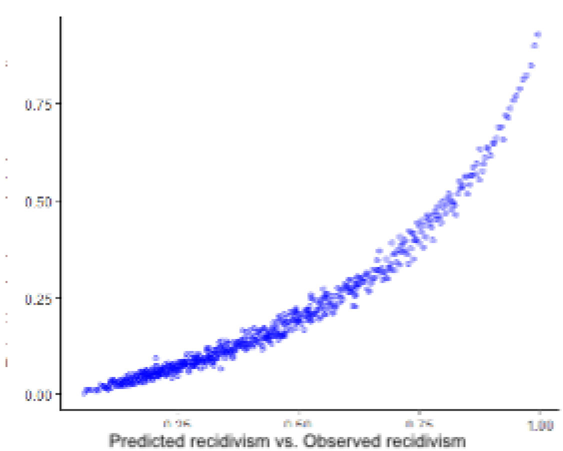
Table 6. Results of the recidivism risk prediction models by crime dimension.

	Recidivism prediction models			
	General crime	Economic crime	Violent crime	Other crimes
AUC	0.7735	0.8874	0.7406	0.7765
Accuracy	0.7168	0.7964	0.6725	0.6698
Sensitivity	0.6789	0.8212	0.6877	0.7397
Specificity	0.7271	0.7935	0.6722	0.6641
Type I Error	0.2729	0.2065	0.3278	0.3359
Type II Error	0.3211	0.1788	0.3123	0.2603

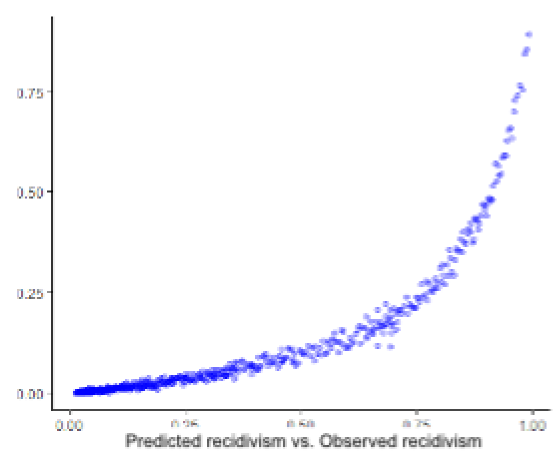
Additionally, Graph 7 shows that the algorithm's prediction aligns closely with the observed recidivism. In the general crime recidivism prediction model (see Graph 7.a), there is a trend close to the 45° line, where the model's fit would be perfect. This representation visually demonstrates the model's predictive accuracy across different crime categories. The closer the trend of the plotted points to the 45° line, the more accurate the model is in predicting recidivism.

Graph 7. Rate of recidivism predicted by the model versus observed recidivism rate for general crime (a), economic crime (b), violent crime (c), and other offenses (d). Each point represents one of 600 percentile groups.

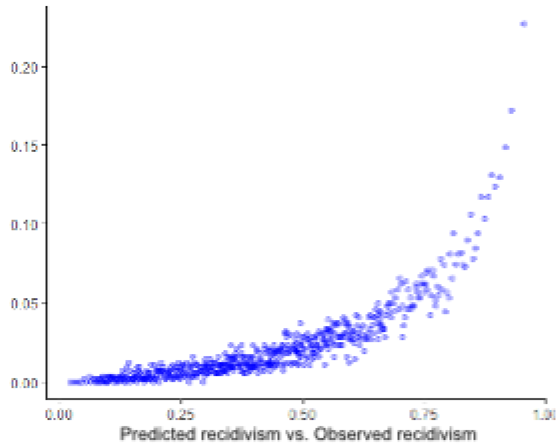
(a) General Crime



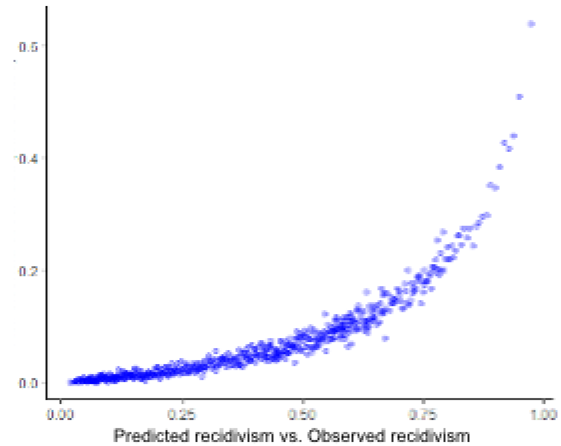
(b) Property Crime



(b) Violent Crime



(d) Other crimes



The importance assigned by the algorithm to different variables varies depending on the type of crime being predicted. The variables that most influence the likelihood of an accused individual reoffending in any crime are: the number of previous police arrests, the number of previous investigations for theft, and the age of the accused. In predicting recidivism in economically motivated crimes, the most important variables are whether the current crime is theft, the number of previous investigations for theft, and age. In contrast, in the violent crime model, the likelihood of recidivism is more significantly determined by the age of the accused, whether the current crime is domestic violence, and the average penalty for the current crimes. Meanwhile, the most important variables in the dimension of other offenses are age, the penalty for the current event, and previous police arrests. It is important to note that gender only has a relatively high importance in the violent crime model. Annex II shows each model's 15 most important variables and their relative importance. The following are detailed results for each model and the potential effect of using the tool on errors made in requesting and granting preventive measures, crime committed by recidivists, and the number of intramural preventive measures.

6.1 General Crime Prediction Model

Graph 8 displays the distribution of the probability of recidivism in any crime for individuals with events between 2012 and 2017. The mean predicted risk is 0.45, and 50% of the individuals have a general risk higher than 0.43. This highlights how the algorithm can distinguish various aspects influencing recidivism based on crime type, indicating a nuanced understanding of criminal behavior and its predictors. This differentiation is crucial for developing targeted interventions and policies in the criminal justice system.

Graph 8. Distribution of Risk Predicted by General Model

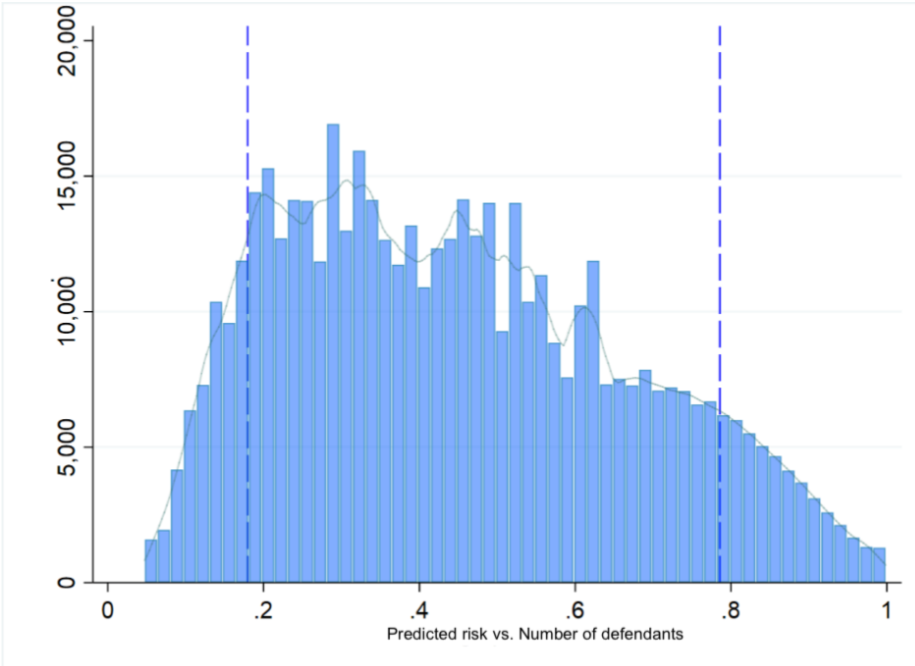


Table 7 compares the two extremes of the risk distribution in general crime. Individuals in the first decile of the recidivism risk distribution for any crime reoffend at a rate of 3.6%, and the average recidivism probability predicted by the model for this group is 0.13. Meanwhile, individuals in the highest-risk decile reoffend at a rate of 65.3%, with the model estimating an average recidivism risk of 0.86. Additionally, the highest-risk decile primarily consists of individuals charged with theft (78.3%) and drug trafficking, carrying, or manufacturing (10.1%), while the lowest-risk decile mainly comprises individuals charged with sexual offenses (18.27%) and domestic violence (8.99%).

Regarding penalties, although the penalty for current crimes is higher for less risky individuals, the average sum of historical crimes is 20 times higher in the most risky decile. Similarly, statistics on the number of previous charges and arrests, and the number of times in prison and duration of imprisonment, are significantly higher in the more risky group. Despite this, and the expectation of a large difference in the rate of requesting preventive measures between the two groups, the rate of requesting intramural preventive measures by prosecutors in the most risky decile (45.66%) is only nine percentage points higher than the request rate in the least risky decile (36.72%).

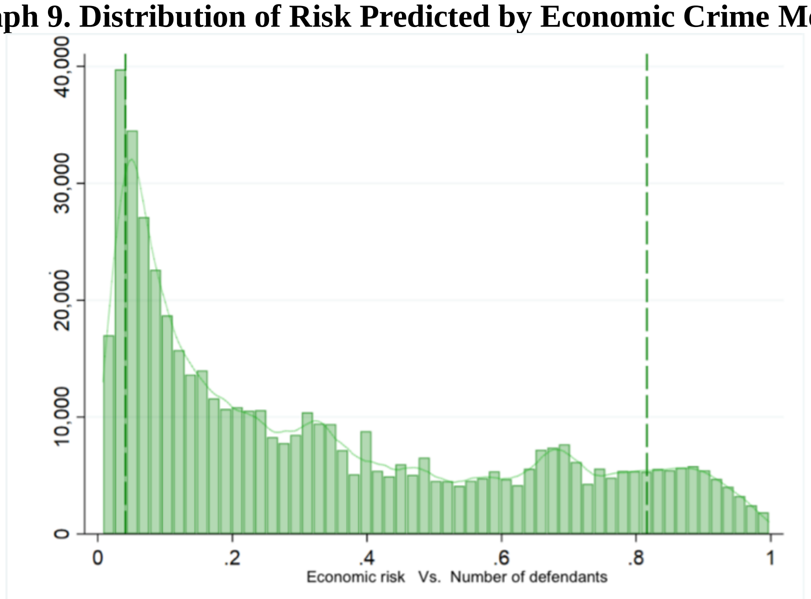
Table 7. Comparison of the lower and upper deciles by predicted risk in General Crime

	Least risky decile	Most risky decile
N	52,519	45,768
Estimated risk	0.13	0.86
Recidivism within 2 years | Does not grant	3.57%	65.23%
Recidivism economic crime	0.30%	50.26%
Recidivism violent crime	0.94%	2.43%
Recidivism other crimes*	0.77%	11.42%
Theft	0.22%	78.32%
Homicide	5.11%	1.14%
Injuries	5.00%	0.54%
Sexual Offenses	18.07%	0.43%
Narcotics	1.16%	10.07%
Domestic Violence	8.99%	1.53%
Conspiracy	3.18%	2.58%
Current crime sentences	133.75	102.89
Average sum of historical crime sentences	22.93	472.62
Previous arrests	0.43	6.43
Previous SPOA Charges	0.02	2.75
Times in Prison INPEC	0.01	1.10
Days in Prison	4.75	700.60
Requests Measure	36.72%	45.66%
Grants Prison Measure | Requests	56.95%	71.06%
Does Not Grant Prison Measure	43.05%	28.94%

6.2 Prediction Model for Recidivism in Economically Motivated Crimes

The distribution of the probability of recidivism in economically motivated crimes for individuals with events between 2012 and 2017 is shown in Graph 9. The average risk is 0.34, and 50% of the individuals have a risk higher than 0.25, indicating a leftward skew in the distribution.

Graph 9. Distribution of Risk Predicted by Economic Crime Model



This graph illustrates how the prediction model assesses the likelihood of individuals reoffending in economically motivated crimes. The leftward skew of the distribution suggests that a significant portion of the individuals have a lower probability of reoffending in these types of crimes, but there remains a substantial group with a moderate to high risk. Understanding this distribution is crucial for targeting interventions and resources effectively, especially for those who are at a higher risk of recidivism in economically motivated crimes.

Table 8 compares various aspects of the differences between the least and most risky deciles according to the risk of recidivism in economically motivated crimes. On average, individuals in the least risky decile reoffend within the first two years after an accusation at a rate of 6.2%, while the model predicts a reoffense rate of 3%; conversely, individuals in the most risky decile reoffend at a rate of 61.1%, with the model predicting an average probability of 89% for them. The least risky decile in terms of recidivism in property crimes is primarily composed of individuals charged with domestic violence (20.2%), drug trafficking and manufacturing (15.7%), and sexual offenses (15.5%), while the most risky decile is comprised of individuals charged with theft (95.8%).

Regarding sentences, the least risky decile in this case has an average sentence of the current charged crime of 123.7 months, while the most risky decile has an average sentence of 91.3 months for the current charged crime. This relates to the previous description, showing that while the least risky decile in terms of recidivism in property crimes mainly consists of people who have committed more serious violent crimes, the most risky decile is disproportionately composed of individuals who have committed theft crimes. However, analyzing the average of the total sentences for historical crimes committed by individuals in the upper decile, it is almost 8 times higher than that observed for individuals in the lower decile of the distribution (393 months vs. 50 months, respectively). In other words, from this analysis, it can be inferred that individuals in the lower risk decile of the property crime recidivism distribution commit, on average, more serious crimes (violent offenses like domestic violence and sexual crimes) whereas those in the upper decile commit less serious crimes (theft), but have done so much more frequently. This is reflected in the fact that individuals in the lower decile have, on average, 0.4 previous arrests and 0.15 previous charges, while those in the upper decile have an average of 6 previous arrests and 2.4 charges.

Similar to the general crime model described in the previous subsection, the difference between the upper and lower deciles in the rates of requests for intramural preventive measures by prosecutors and their granting by judges is only 11 percentage points (32% higher for the upper decile) and 9 percentage points (15.7% higher for the upper decile) respectively, when the observed and predicted recidivism levels by the model are 10 times and 30 times higher for the riskiest decile, respectively.

Table 8: Comparison of Lower and Upper Deciles Based on Predicted Risk in Economically Motivated Crimes

	Least risky decile	Most risky decile
N	53,504	46,602
Estimated risk	0.03	0.89
Recidivism within 2 years | Does not grant	6.26%	61.13%
Recidivism economic crime	0.22%	53.82%
Recidivism violent crime	1.83%	1.62%
Recidivism other crimes *	2.03%	5.18%
Theft	0.00%	95.77%
Homicide	3.24%	0.41%
Injuries	1.77%	0.39%
Sexual Offenses	15.50%	0.14%
Narcotics	15.69%	0.60%
Domestic Violence	20.24%	0.39%
Conspiracy	4.56%	1.67%
Current crime sentences	123.70	91.37
Average sum of historical crime sentences	49.96	393.44
Previous arrests	0.40	5.99
Previous SPOA Charges	0.15	2.40
Times in Prison INPEC	0.02	1.00
Days in Prison	5.80	595.52
Requests Measure	34.83%	45.97%
Grants Prison Measure | Requests	60.35%	69.81%
Does Not Grant Prison Measure	39.65%	30.19%

6.3 Predictive Model for Violent Crimes

In Graph 10, the distribution of the probability of recidivism in violent crimes is depicted for individuals with recorded events between 2012 and 2017. The average risk predicted by the model is 0.44, meaning that 50% of the individuals have a risk greater than 0.44 of committing a violent crime.

Graph 10. Distribution of Risk Predicted by the Violent Crime Model

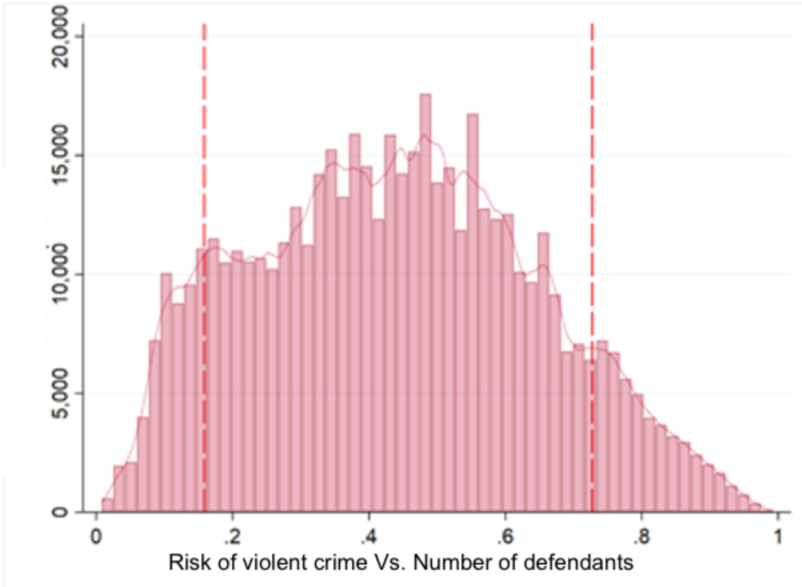


Table 9 shows that in the lower decile of the distribution, the estimated risk is 0.11, and individuals reoffended at a rate of 24.2% in violent crimes; whereas in the upper decile, the estimated risk is 0.81, and the observed recidivism rate was 9.15%. Unlike the previous two models, the most risky decile is mainly composed of individuals charged with domestic violence (45.76%) and homicide (20.09%), while those in the least risky decile are predominantly charged with drug-related offenses (34.93%) and theft (27.33%). Furthermore, it is observed that both the sentence for current crimes and the sum of sentences for historical crimes are higher in the more risky decile than in the less risky one. Conversely, the number of previous arrests and charges is lower in the decile of individuals with higher risk. Finally, there is a wider difference in the rate of requests for preventive measures between the two deciles, nearly 15 percentage points, being 58.30% in the upper decile and 43.71% in the lower decile. Regarding the granting of intramural measures by control judges, it is 50.4% in the lower decile and 71.5% in the upper decile.

Table 9: Comparison of Lower and Upper Deciles Based on Predicted Risk in Violent Crime

	Least risky decile	**Most risky decile**
N	51,809	48,446
Estimated risk	0.11	0.81
Recidivism within 2 years | Does not grant	24.23%	23.57%
Recidivism economic crime	17.11%	7.22%
Recidivism violent crime	0.24%	9.15%
Recidivism other crimes	5.67%	6.83%
Theft	27.33%	11.35%
Homicide	0.54%	20.09%
Injuries	0.26%	2.92%
Sexual Offenses	0.10%	7.97%
Narcotics	34.93%	5.91%
Domestic Violence	0.00%	45.76%

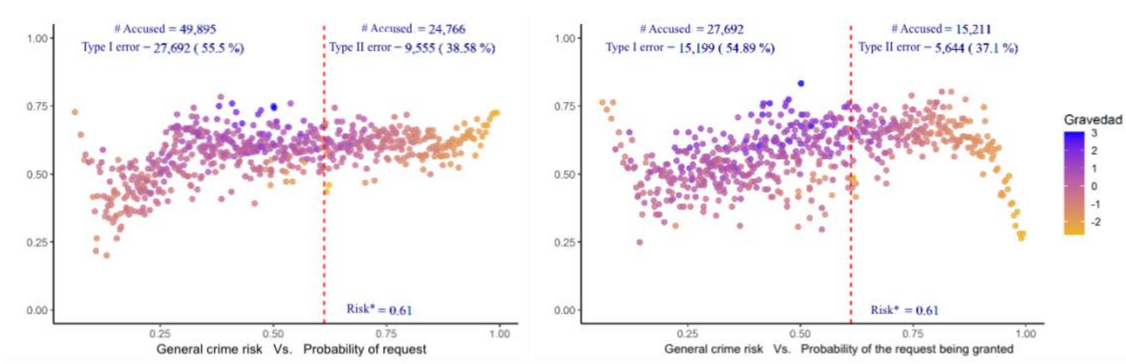
	Least risky decile	Most risky decile
Conspiracy	9.81%	8.49%
Current crime sentences	149.00	216.55
Average sum of historical crime sentences	151.93	310.41
Previous arrests	2.48	1.89
Previous SPOA Charges	1.07	0.78
Times in Prison INPEC	0.23	0.41
Days in Prison	107.05	279.22
Requests Measure	43.71%	58.30%
Grants Prison Measure | Requests	50.37%	71.46%
Does Not Grant Prison Measure	49.63%	28.54%

6.4 Potential Gains of Models Over Current Errors

To evaluate the potential effect of the models, we calculated the Type I and Type II errors currently committed given the predicted risk. Graph 11 shows the request rate (a) and the granting rate upon request (b), in relation to the risk predicted by the general crime model, for the events of 2018. If the 24,766 intramural measures granted in 2018 had been requested for the events involving the most risky individuals, they would have been requested from people with a risk of 0.61 or higher. However, prosecutors requested measures for 55.5% of the accused with a risk below this threshold (Type I error), and failed to request measures for 38.58% of those with a higher risk (Type II error). Regarding judges, they granted intramural measures to 54.89% of the accused with a risk below 0.61 (Type I error), and did not grant them to 37.1% of those with a higher risk (Type II error). Additionally, and concerningly, a drop in the granting rate for risks close to one (i.e., in the upper tail of the criminal recidivism risk distribution) is observed. When considering the severity of the current crime, it is noted that in cases with a higher risk of criminal recidivism, the sentence for the current crime for which these cases were being charged is low. Therefore, Type I and Type II errors may be partly due to what is known as the current crime bias.

Graph 11: Distribution of Request Rate (a) and Grant Rate (b) Relative to the Risk Predicted by the General Crime Model (2018)

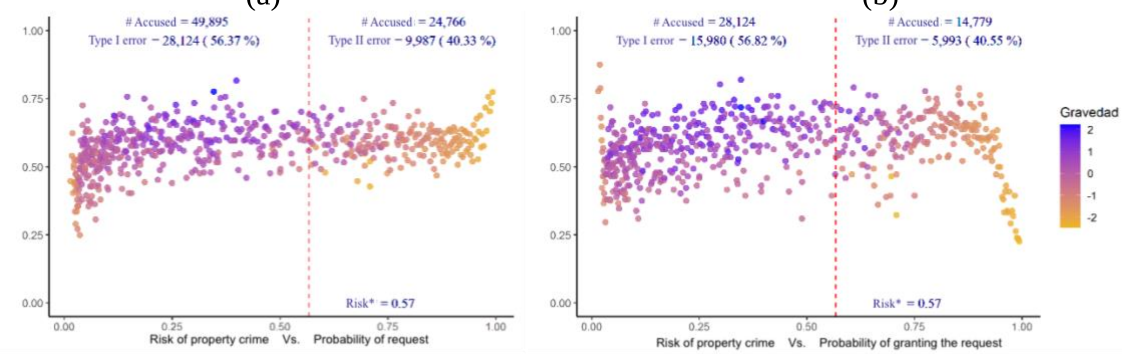
(a) (b)



In cases of economically motivated crimes, if the 24,766 intramural measures granted in 2018 had been requested for the events of the most risky individuals according to the risk prediction model for recidivism in such crimes, they would have been requested for individuals with a risk of 0.57 or higher. However, prosecutors requested measures for 56.37% of the accused with a risk lower than this threshold (Type I error), and failed to request measures for 40.33% of those with a higher risk (Type II error). In terms of judges, they granted intramural detention measures to 56.82% of the accused with an economic risk lower than 0.57 (Type I error), and did not grant them to 40.55% of those with a higher risk (Type II error). Although a positive trend in requests and grants in relation to risk would be expected, a very similar rate is observed across the entire risk distribution.

Graph 12: Distribution of Request Rate (a) and Grant Rate (b) Relative to the Risk Predicted by the Economic Motivation Crime Model (2018)

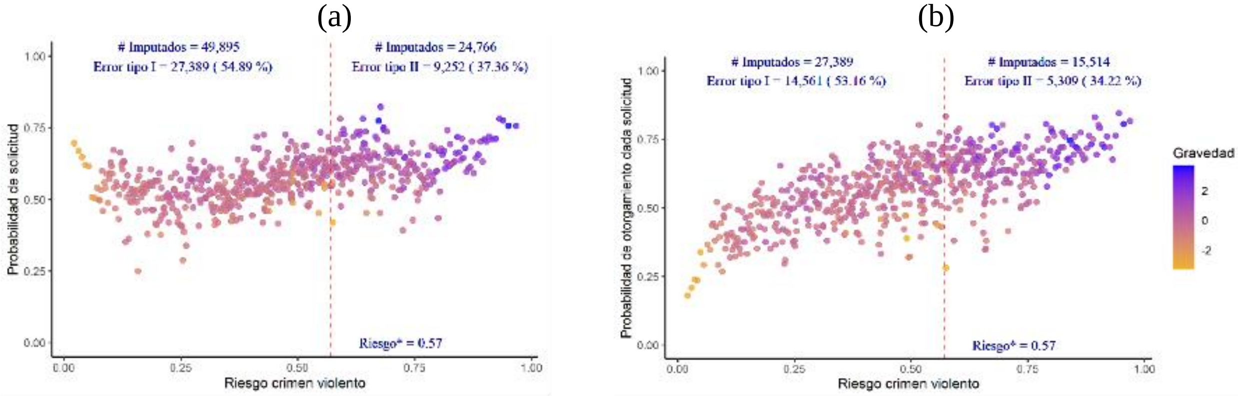
(a) (b)



When analyzing the predictive model for recidivism in violent crimes, if the 24,766 intramural measures granted in 2018 had been requested for those most at risk of reoffending in violent crimes, they would have been requested for individuals with a risk of 0.57 or higher (Graph 13). However, prosecutors requested measures for 54.89% of the accused with a risk lower than this threshold (Type I error), and failed to request measures for 37.36% of those with a higher risk (Type II error). In terms of judges, they granted intramural measures to 53.36% of the accused with an economic risk lower than 0.57 (Type I error), and did not grant them to 34.22% of those with a higher risk (Type II error). In this case, a slight positive trend is observed in the granting rate in relation to risk, and it is also noted that the severity of the current crime is greater in events with higher risk.

Graph 13: Distribution of Request Rate (a) and Grant Rate (b) Relative to the Risk Predicted by the Violent Crime Model (2018)

(a) (b)



7. The Problem of unobservables in crime prediction: are prosecutors and judges making mistakes?

Following the methodology proposed by Kleinberg et al. (2018), this section uses the severity of prosecutors who request preventive measures to address the "selective labels" issue, which will be detailed here. Using the unique identifiers of prosecutors who request these measures, incarceration rates by the prosecutor are calculated, and quintiles of severity are determined to more accurately estimate the reduction in crime achievable by using the criminal recidivism risk prediction tool in Colombia.

The addressed problem arises from the difficulty of knowing whether the algorithm's prediction can improve judicial decisions. With a binary variable Y and a set of variables X related to each individual's criminal history (previous arrests, charges, convictions, incarceration events, etc.), we define $P(Y = 1) = y$ as the probability that a person commits a crime in the following two years. Additionally, suppose that prosecutors observe a set of variables U that the algorithm does not, due to information available to prosecutors and judges at the time of these hearings that is not in the databases.

Following Kleinberg et al. (2018), we can define two unidimensional variables $x(X) \equiv E[Y|X]$ and $u(X, U) = E[Y|X, U] - E[Y|X]$. This allows us to identify a model in which the recidivism risk of charged individuals is characterized by observable variables x and unobservable variables u (the latter observed by prosecutors and judges but not by prediction models). There might be a w capturing additional aspects of the accused or even the mood of the prosecutor or judge that could affect the decision to request or grant intramural preventive measures, but this does not provide additional information about y. The risk is then defined as:

$$E[y|X, U] = E[y|x, u] = x + u$$

Following this conceptual framework, each prosecutor f makes a decision L not to request (L=1) or to request preventive custody (L=0). It is important to highlight that, unlike Kleinberg et al. (2018), who assume that judges are assigned cases randomly in terms of x,u,w, in Colombia, prosecutors are not randomly assigned cases; however, we tested this approach to give a more accurate estimate of the gains from using the prediction algorithm.

Given the above, we model the expected payoff π of each prosecutor f as a function depending on the propensity of the accused to commit crimes (y), the risk of non-appearance (p), the risk of affecting the evidence (e), and the decision to not request preventive custody, L:

$$\pi^f(y,p,e,L) = -a_f yL - c_f(p+e)L - b_f(1-L)$$

Both p and e can be assumed as procedural risks with a cost -$-c_f$ that each prosecutor idiosyncratically assigns. $-a_f$ represents the value that prosecutor f gives to the crime, and $-b_f$ is the cost of imprisonment immersed in the prosecutor's decision. Measures are requested if the prosecutor's prediction about the risk posed by an individual is above a certain threshold k_f, determined by how the prosecutor rates the crimes committed (a_f), the costs of imprisonment $\left(b_f\right)$ and the procedural costs (c_f). The expected payoff for the prosecutor, given a decision rule assumed by them, can be seen as $\Pi^f(\rho) = E[\pi^f(y,p,e,L)]$where the prosecutor chooses a decision rule ρ^f that maximizes their expected payoff.

The question to be resolved then is whether a prediction given by the algorithm $m(X)$ can improve the decisions of the prosecutors; that is, if there is a decision rule d that combines the prosecutors' prediction and that made by the algorithm about the risk of the accused continuing their criminal activity to improve the prosecutors' payoff. More precisely, a decision such that $\Pi^f(\rho^d) > \Pi^f(\rho^f)$ (ρ^f). The difference between these two payoffs is given by the rates of non-request of the preventive measure of the prosecutors' decision rules $\bar{L}^f$ multiplied by the probability of recidivism given that decision rule when the accused is not deprived of freedom, and by the weight that prosecutors give to each cost:

$$\Pi^f(\rho^d) - \Pi^f(\rho^f)$$
$$= -a_f(\bar{L}^d E[y|\rho^d = 1] - \bar{L}^f E[y|\rho^f = 1]) - c_f(p+e)(\bar{L}^d - \bar{L}^f) - b_f(\bar{L}^f - \bar{L}^d)$$

For simplicity, we consider the case in which a decision rule of the algorithm makes $\bar{L}^f = \bar{L}^d$, which implies keeping the rates of requesting measures constant and making the second and third terms of the right side of the equation become 0. In this case, the measurement problem is given by the first term, as although this cancels out when the tool and the prosecutor agree in their decision, the difference in payoffs is determined by the cases in which they do not agree.

The main problem arises from the impossibility of correctly estimating the change in crimes committed by those deprived of liberty preventively. That is, we can only measure with certainty the change in crime resulting from the accused to whom the prosecutor did not request measures but whom the algorithm did recommend depriving of liberty. However, due to the incapacitation effect produced by the deprivation of liberty in a prison, it is impossible to observe the criminal recidivism of those deprived of liberty but whom the algorithm suggested leaving free. To solve this problem, part of the literature suggests using the prediction based on observable variables (X) of the incarcerated people to solve the unobservables issue, an approach used in the first part of this document, estimating the propensity to commit crimes of incarcerated people based on the observed behavior of people with similar characteristics but to whom a preventive measure was not requested or granted. This procedure assumes that the expected crime rates of the people to whom a measure was requested $E[y|\rho^f = 0]$ are equal to those of those to whom it was not requested, having similar observable characteristics $E[y|\rho^f = 1, x]$. The problem lies in the unobservable variables that prosecutors and judges may consider when making these decisions. These unobservable variables may differ between people to whom a measure is requested or not, so a comparison based on unobservables can be problematic. This is precisely the "selective labels" problem posed in Kleinberg et al. (2018).

To solve the problem raised above, the contraction methodology proposed by Kleinberg et al. (2018) was used. In particular, we used the unique identifiers of the prosecutors from the SPOA system of the Prosecutor's Office, and the fact that different prosecutors have different rates of requesting measures: when prosecutors are divided into quintiles of severity, measured by the rate at which they request intramural preventive measures for similar cases, it is observed that the least severe quintile requests the intramural preventive measure in 12% of the cases, while the strictest quintile of prosecutors requests it in 80% of the cases.

Although cases are not assigned randomly to prosecutors, the question addressed with the contraction methodology in this section is: if we start from the cases of the least strict quintile of prosecutors and start requesting additional preventive measures according to the prediction of the tool, what results in terms of crime and rates of preventive measures would be reached?, and how do these results compare with the request rates of the other quintiles of stricter prosecutors?

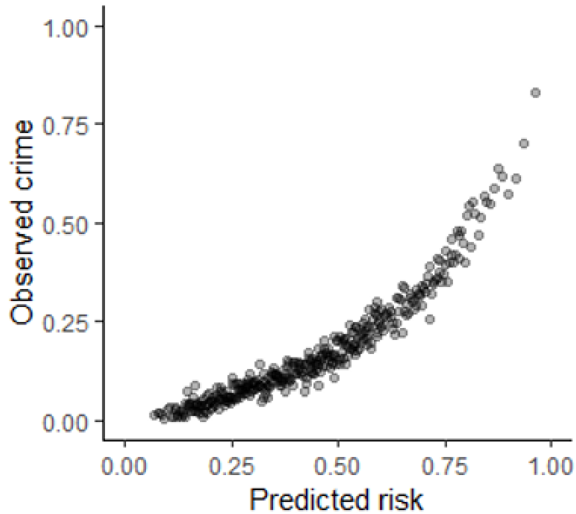
Conditioning on observable characteristics X, the assumption can be made that the quintiles of prosecutors on average have the same unobservable characteristics, then, when comparing between quintiles of severity, the measurement problem of unobservable characteristics can be solved since these cancel out between quintiles. In this way, the estimation of how much crime committed by reoffenders can be reduced can be more accurate. To implement this contraction procedure, 1,256 cells were created by the prosecutor's office section, year, and type of crime in the cases (violent, economic) characterized by having at least 5 prosecutors, each of them with at least 5 cases that apply for the request of intramural preventive measure, that is, whose current crime had a minimum sentence of more than 4 years. These 1,256 cells contain 50.2% of the original cases and 81.4% of the cases excluding the "test" sample on which the algorithm is tested.

7.1 Assumptions of the Severity Methodology

The methodology described above assumes that prosecutors from different quintiles equally value unobservable variables, that is, the unobservable characteristics of the people to whom an intramural preventive measure is not granted, are on average the same between quintiles (Kleinberg et al., 2018). To address this assumption, the general crime model is trained using the observations of the most lenient quintile and the fit of the model is evaluated on the observations of the other quintiles. If prosecutors in the most lenient quintile have a lower rate of request because they better discriminate risk based on unobservable characteristics, classifying fewer accused as high risk, it would be expected that they would predict a lower risk for the accused in the other quintiles. However, in Graph 14 it is observed that the fit of the model trained with observations from all quintiles is similar to the fit of a model trained with the most lenient quintile and evaluated in the other quintiles. Therefore, it is not evident that the predicted risk in relation to the observed crime is lower in the model trained with the most lenient quintile, so the risk assessment is not due to different unobservable variables between the quintiles.

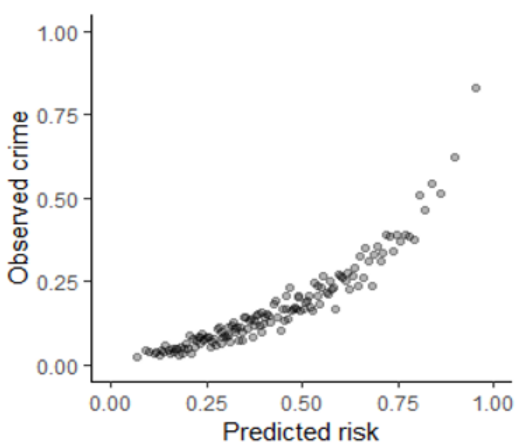
Graph 14: Evaluation of Unobservable Characteristics Among Quintiles

Predictions Using All Training Data

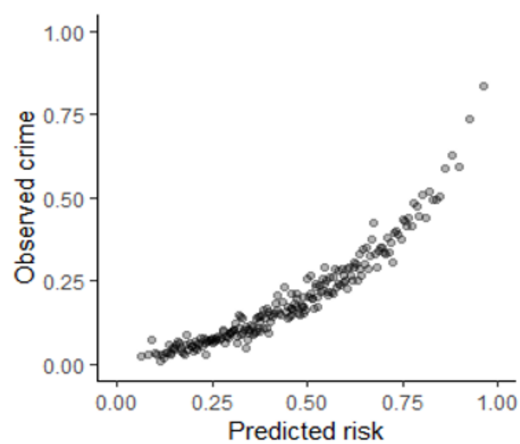


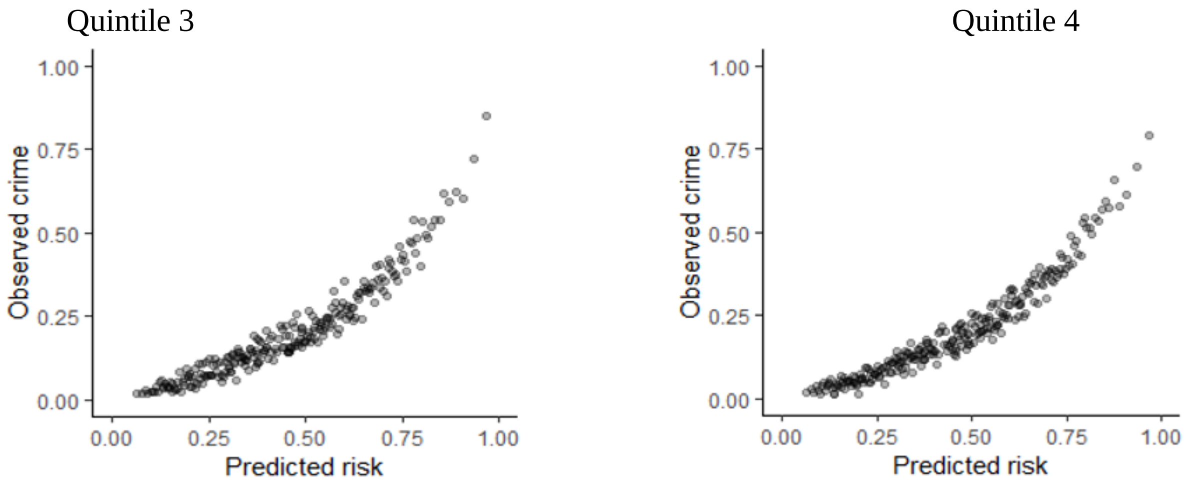
Predictions Using the Most Lenient Quintile

Quintile 1 (Least Lenient)

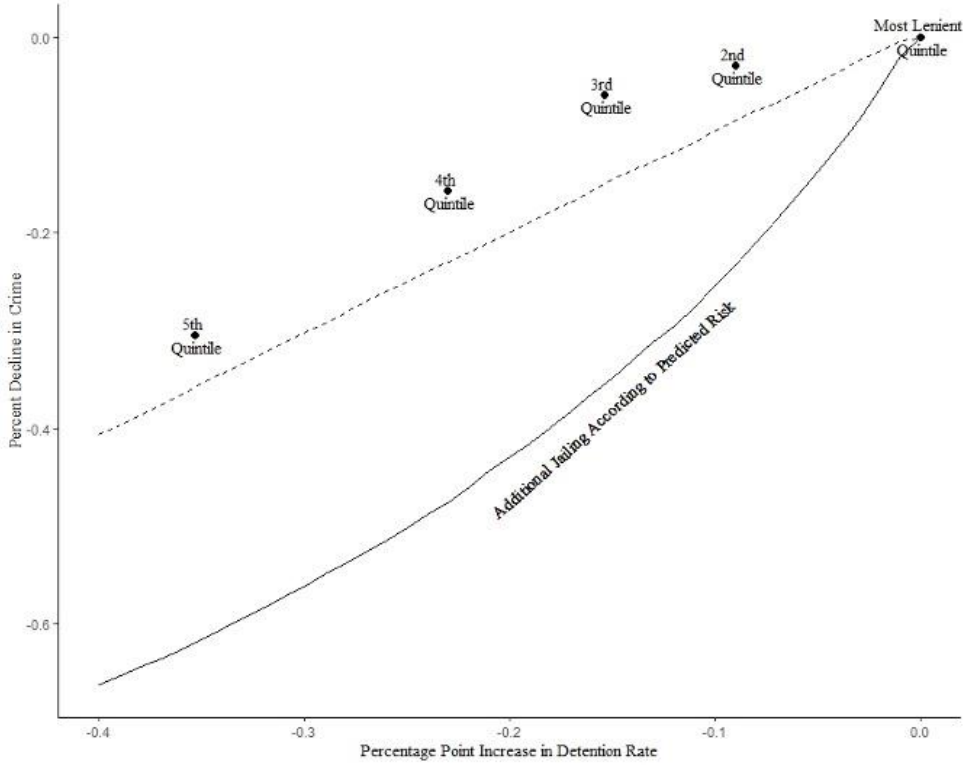


Quintile 2

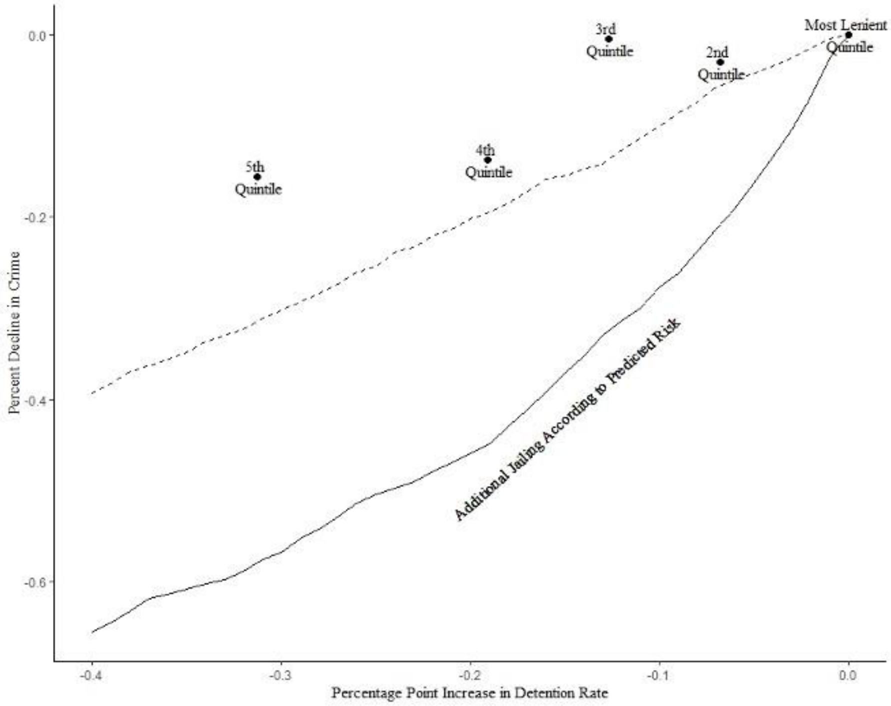




Graph 15. Changes in General Crime Achieved by Contracting the Population Released by the Most Lenient Quintile of Prosecutors



Graph 16. Changes in Violent Crime Achieved by Contracting the Population Released by the Most Lenient Quintile of Prosecutors



Graphs 15 and 16 illustrate the results of applying this methodology to reducing crime and its relationship with the incarceration rate. The solid line shows the decrease in crime that could have been obtained if the PRiSMA tool had been used to request additional measures based on the risk predicted by the algorithm. The points indicate the expected reduction when comparing the least strict quintile of prosecutors against each of the stricter quintiles. Since incarcerating individuals (regardless of their risk) can reduce the crime they commit due to the incapacitating effect of jail, the dotted line shows the results that can be obtained if measures are requested randomly. Prosecutors' decisions are no better than random measures. This may be due to the systematic allocation of different cases based on difficulty or recidivism, rather than random assignment.

The results show significant gains in terms of crime reduction. For general crime, the second quintile of prosecutors reduces the general crime charged by 4% relative to the least strict quintile by increasing the request rate of measures by 9 percentage points. Using the algorithm (solid line), this same result could have been achieved by increasing measures by 1 percentage point. Equivalently, with the same number of measures granted by quintile 2, the tool would achieve a 25% decrease in crime by increasing the request for measures by 9 percentage points.

In the case of violent crime, moving quintiles from the least harsh to the harshest decreases crime in all cases except in quintile 3, where although there is a reduction in crime, it is much less than for quintile 2, which requests fewer measures. The results indicate that maintaining the same number of requested measures as quintile 2, using the algorithm would reduce violent crime by 22% simply by increasing the requested measures by 7 percentage points. Comparing all the points of the quintiles with the solid line, it can be concluded that the same levels of crimes can be achieved by reducing the number of measures requested, that is, moving to the right of the graph, maintaining the same decrease in crime. The results shown here suggest that improving efficiency and accuracy with more information in the decision-making process of prosecutors and judges could be promising for reducing incarceration rates in the country without increasing crime rates.

8. Conclusions

The use of statistical prediction tools has recently gained momentum in criminal justice systems. Decisions that judges and prosecutors make daily, such as granting bail to an individual under investigation for the alleged commission of a crime, the duration of sentences judges define, or the decision to detain a person while on trial due to their high risk of criminal recidivism, can be supported by prediction algorithms that help prosecutors and judges make better decisions. At no point is the promotion and use of these tools intended to replace the work of judges and prosecutors, but rather to complement them in making better decisions across various dimensions.

In this paper, we present the construction of the PRiSMA tool, Risk Profile for Recidivism for the request of pretrial measures, developed within the Attorney General's Office of Colombia to support the work of prosecutors who must make daily decisions on the request, before control judges, for the preventive detention of individuals who have been charged with the commission of a crime. Law 906 of 2004 establishes three reasons upon which prosecutors must base their arguments to judges when requesting pretrial detention for an individual under investigation for the presumed commission of a crime. For citizen security crimes, such as homicides, sexual offenses, injuries, and thefts, among others, the main argument prosecutors use to request pretrial detention is the possible impact on citizen security as a result of the individual's potential for criminal recidivism. However, when arguing this reason, prosecutors (and judges) have relatively little systematic information about the objective risk of criminal recidivism to make these decisions.

As a consequence, errors are made daily, such as detaining individuals who objectively do not have a high risk of criminal recidivism and releasing individuals who have a high risk of criminal recidivism and represent a danger to the community. Prediction algorithms like the one we present in this work can not only support prosecutors and judges to have more homogeneous, systematic, and transparent information in the hearings where these judicial decisions are made but can also help reduce the errors described above. Additionally, in this work, we show that by resolving these errors, significant reductions in crime could be achieved without increasing the number of people who are preventively deprived of liberty and/or reducing the number of people who are deprived of liberty without increasing criminal recidivism. In this work, we present two alternative ways to build these algorithms for predicting the risk of criminal recidivism.

In both models, the conclusions are similar: the appropriate use of these tools, with due precautions, monitoring, and protocols, can support prosecutors and judges in the decision-making process regarding the preventive detention of individuals whom the prosecutor has charged for the alleged commission of a crime. As with any tool of this kind, there are risks of manipulation and misuse, but with proper protocols and monitoring, such tools can support prosecutors and judges in making fairer and more efficient decisions, especially when the issue at hand is an individual's freedom. As has also been shown in some articles, the use of these tools can also contribute to reducing discrimination based on gender, race, or socioeconomic status.

With the right protocols, using these algorithms in criminal justice systems creates the potential for new forms of transparency and, therefore, opportunities to detect discrimination that otherwise would not have been detected. The specificity of these algorithms based on artificial intelligence also highlights the trade-offs that arise when making decisions like the preventive detention of an individual under investigation for the presumed commission of a crime. Although this specific issue was not addressed in this work, it is important to note that the individual-level databases available in Colombia can be utilized to continuously monitor the effects of using tools such as the one proposed in this study on discrimination levels in the criminal justice system and the different error types. This is essential to ensure that the use of such algorithms is transparent and any necessary adjustments can be made to promote fair, efficient, and transparent decision-making.

9. Bibliography

Berk, Richard (2017). "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism," *Journal of Experimental Criminology* 13 (2), 193–216.

Berk, Richard, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, and Lindsay Ahlman (2009). "Forecasting Murder within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning." *Journal of the Royal Statistical Society: Series A* 172 (1), 191–211.

Buitrago, J. R., Rodríguez, J. D. & Bernal, P. A. (2015). "Registros administrativos de policía para la consolidación de cifras de criminalidad en Colombia," *Revista Criminalidad* 57 (2): II-22.

Tianqi Chen and Tong He. xgboost : eXtreme Gradient Boosting. *R package versión 0.4-2*, pages 1–4, 2015. URL *https://cran.rproject.org/web/packages/xgboost/vignettes/xgboost.pdf.*

Friedman, Jerome H. (2001), "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* 29, 1189–1232.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018). "Human Decisions and Machine Predictions," *The Quarterly Journal of Economics* 133 (1), 237–293.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein (2019). "Discrimination in the Age of Algorithms," *Journal of Legal Analysis*, 04 (10), 113–174.

Stevenson, Megan and Jennifer Doleac (2019). "Algorithmic Risk Assessments in the Hands of Humans," IZA Discussion Paper No. 12853, December.

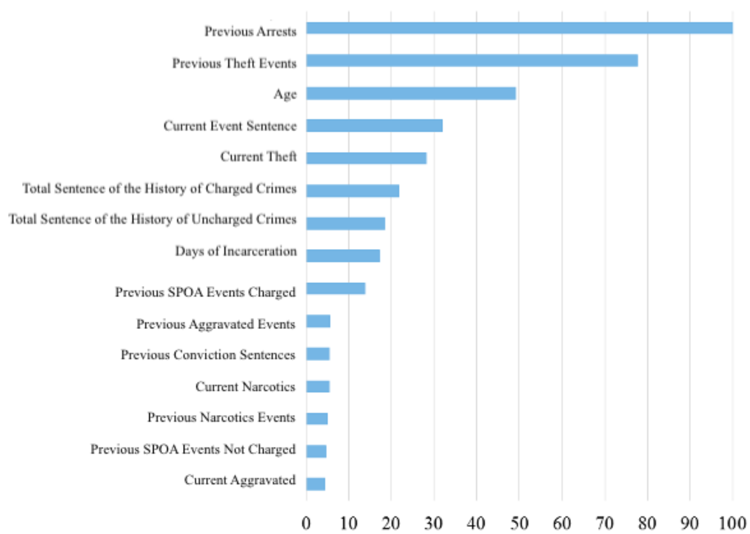
Annex 1

Table A1. Classification of property crimes, violent crimes, and other crimes

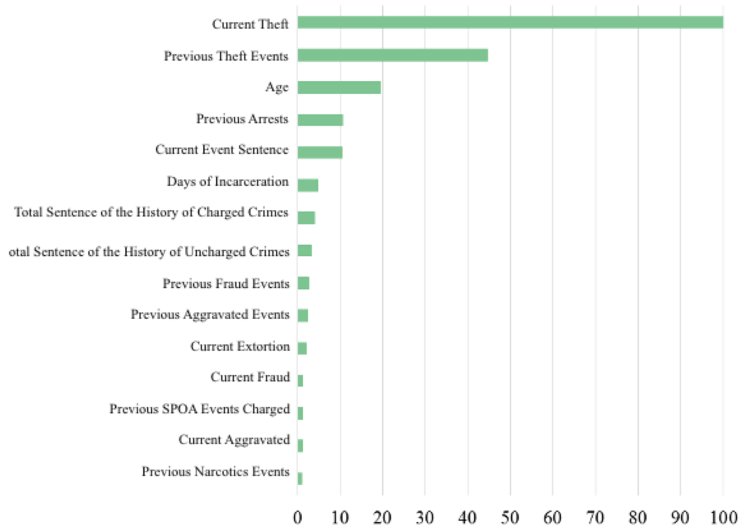
Category	Article of Criminal Procedure Code
Crime Against Property	Article 239. Theft
	Article 240. Qualified Theft
	Article 244. Extortion
	Article 246. Fraud
	Article 249. Breach of Trust
	Article 250. Qualified Breach of Trust
Violent crime	Article 103. Homicide
	Article 104A. Femicide
	Article 104B. Aggravating Circumstances of Femicide
	Article 105. Unintentional Homicide
	Article 106. Mercy Killing
	Article 108. Death of a Child Resulting from Violent or Abusive Carnal Access, or Non-consensual Artificial Insemination or Fertilized Ovum Transfer
	Article 111. Aggravated Injuries
	Article 119. Punitive Aggravation Circumstances
	Article 128. Abandonment of a Child Resulting from Violent or Abusive Carnal Access, or Non-consensual Artificial Insemination or Fertilized Ovum Transfer
	Article 135. Homicide of a Protected Person
	Article 138. Violent Carnal Access to a Protected Person
	Article 139. Violent Sexual Acts on a Protected Person
	Article 139A. Violent Sexual Acts on a Protected Person Under Fourteen Years of Age
	Article 141. Forced Prostitution of a Protected Person
	Article 141A. Sexual Slavery of a Protected Person
	Article 141B. Trafficking of Protected Persons for Sexual Exploitation
	Article 205. Violent Carnal Access
	Article 206. Violent Sexual Act
	Article 208. Abusive Carnal Access with a Minor Under Fourteen
	Article 209. Sexual Acts with a Minor Under Fourteen
	Article 210A. Sexual Harassment
	Article 213. Inducement to Prostitution
	Article 213A. Pimping with Minors
	Article 214. Coercion into Prostitution
	Article 217. Encouragement of Child Prostitution
	Article 218. Pornography with Persons Under 18
	Article 219A. Use or Facilitation of Communication Media to Offer Sexual Activities with Persons Under 18 Years
	Article 229. Domestic Violence

Category	Article of Criminal Procedure Code
	Article 230. Abuse by Restriction of Physical Freedom Article 230A. Arbitrary Exercise of Custody of a Minor Child Article 229A. Abuse by Neglect, Negligence, or Abandonment of a Person Over 60 Years Old
Other crimes	Article 340. Conspiracy to Commit Crime Article 340A. Advising Organized Criminal Groups and Organized Armed Groups Article 341. Training for Illicit Activities Article 343. Terrorism Article 345. Managing Resources Related to Terrorist Activities Article 346. Illegal Use of Uniforms and Insignia Article 347. Threats Article 348. Incitement to Commit Crime Article 349. Incitement to Commit Military Offenses Artículo 365. Fabricación, tráfico o tenencia de armas de fuego, accesorios, partes o municiones Artículo 366. Fabricación, tráfico y porte de armas, municiones de uso restringido, de uso privativo de las fuerzas armadas o explosivos Artículo 376. Tráfico, fabricación o porte de estupefacientes

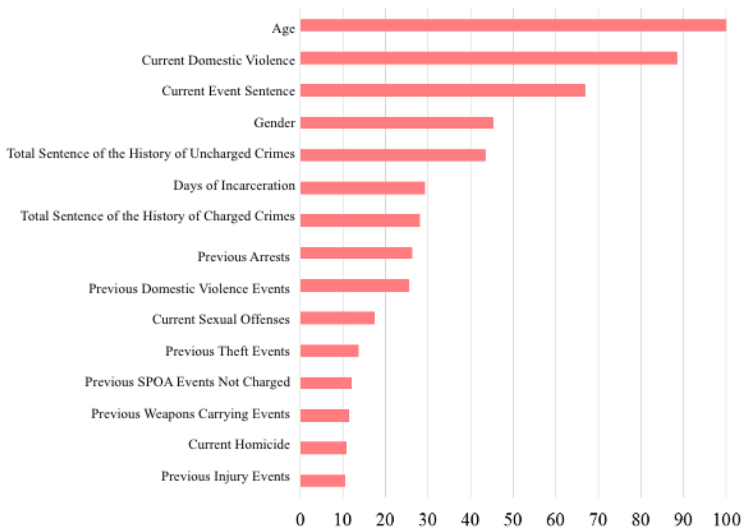
Annex 2
Graph A2.1. Importance of Variables in the General Crime Model



Graph A2.2. Importance of Variables in the Economic Crime Model



Graph A2.3. Importance of Variables in the Violent Crime Model



Graph A2.4. Importance of Variables in the Other Offenses Crime Model