

Predicting politicians' misconduct: Evidence from Colombia

Jorge Gallego^{1,*} , Mounu Prem² and Juan F. Vargas¹

¹School of Economics, Universidad del Rosario, Bogota, Colombia

²Einaudi Institute for Economics and Finance, Rome, Italy

*Corresponding author. E-mail: jorge.gallego@urosario.edu.co

Received: 07 January 2022; **Revised:** 09 August 2022; **Accepted:** 14 October 2022

Key words: Colombia; corruption; machine learning; prediction

Abstract

Corruption has pervasive effects on economic development and the well-being of the population. Despite being crucial and necessary, fighting corruption is not an easy task because it is a difficult phenomenon to measure and detect. However, recent advances in the field of artificial intelligence may help in this quest. In this article, we propose the use of machine-learning models to predict municipality-level corruption in a developing country. Using data from disciplinary prosecutions conducted by an anti-corruption agency in Colombia, we trained four canonical models (Random Forests, Gradient Boosting Machine, Lasso, and Neural Networks), and ensemble their predictions, to predict whether or not a mayor will commit acts of corruption. Our models achieve acceptable levels of performance, based on metrics such as the precision and the area under the receiver-operating characteristic curve, demonstrating that these tools are useful in predicting where misbehavior is most likely to occur. Moreover, our feature-importance analysis shows us which groups of variables are most important in predicting corruption.

Policy Significance Statement

Resources to combat corruption are often scarce. While tools such as audits have proven effective in some contexts, they are often costly and difficult to scale and sustain. Therefore, anticipating, with a certain level of precision, where politicians are most likely to misbehave is key in the fight against this problem. Using data from Colombian municipalities, our approach shows how governments can use artificial intelligence tools to predict where irregularities are most likely to occur. This setup also allows us to identify the features with the greatest predictive power to anticipate where misconduct will occur, which is useful for discussions on institutional reform to curb corruption.

1. Introduction

Corruption, the misuse of public office for private gain (Svensson, 2005), has pervasive consequences on the development of countries and the well-being of the population. Corruption affects governments' abilities to collect taxes (Olken and Pande, 2012), provide public goods and services (Olken, 2007), and correct negative externalities (Bertrand et al., 2007). The private sector is also affected by malfeasance, as

 This research article was awarded an Open Data badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

firms' production choices and investment decisions are less efficient as a consequence of the higher levels of uncertainty generated by corruption (Sequeira and Djankov, 2014). Moreover, corruption acts as a barrier to entry for new firms, thus limiting market competition (Colonnelli and Prem, 2022). Not surprisingly, corrupt countries tend to have lower levels of GDP per capita, worse indicators of human capital, and lower levels of trade openness and political freedom (Svensson, 2005).

Despite its relevance, fighting corruption is tough, in part because it is difficult to measure and detect. Traditionally, most corruption measures have been based on perception surveys, in which key actors, such as businessmen or journalists, are asked their opinion regarding how corrupt a country is. These measures, some of which are still in force, were largely developed to make comparisons between countries.¹ However, comparisons within countries, for example between cities, are also important as they can determine how resources need to be allocated in the fight against corruption. In recent years, there has been notable progress toward the use of more objective and granular indicators of malfeasance. Some of these measures use, as a source of information, either the data contained in the public procurement platforms of each country, or the results of the audits carried out by the anti-corruption agencies.

In fact, the evidence suggests that top-down accountability, whose main tool are audits conducted by high-level agencies, is an effective tool in the fight against corruption (Olken, 2007; Ferraz and Finan, 2008, 2011) and can have positive effects on economic performance (Colonnelli et al., 2022; Colonnelli and Prem, 2022). But audits are a costly and scarce resource, in terms of time, money, and human capital. They cannot be carried out all the time everywhere, therefore they need to be assigned efficiently if what is sought is to maximize their effectiveness. In this context, recent advances in the field of artificial intelligence can be useful, since they allow us to anticipate where corruption is most likely to occur. Decarolis and Giorgiantonio (2020) and Gallego et al. (2021b) apply these tools to predict malfeasance in public procurement at the contract level in Colombia and Italy, respectively; Colonnelli et al. (2022) and de Blasio et al. (2020) use them to develop corruption indicators at the municipal level in Brazil and Italy, respectively; while Salles and Delles (2020) follow a cross-country approach.

In this article, we train a set of machine-learning models to predict corruption at the municipality level in Colombia.² For this purpose, we use the results of the prosecutions for disciplinary offenses carried out by the Office of the Inspector General against mayors of the country's municipalities, for the 2008–2011 and 2012–2015 mayoral periods. Using this indicator as the outcome variable, we train four canonical machine-learning models—Random Forests, Gradient Boosting Machine (GBM), Lasso, and Neural Networks—and ensemble their predictions using the Super Learner approach (Poley et al., 2011). These models are fed by 147 municipality-level predictors, which in turn, are grouped into 10 categories of interest that allow us to understand what are the most important predictors of corruption. Our results show, on the first place, that misconduct at the municipality level in Colombia can be predicted with tolerable levels of precision. The performance of our models is acceptable, reaching levels of precision of 84% and an area under the receiver-operating characteristic (ROC) curve (AUC) of 0.71, two metrics commonly used in the literature to evaluate algorithms. Secondly, our feature importance analysis allows us to understand which characteristics have the greatest predictive power when forecasting corruption. The results show that, in the Colombian case, variables associated with the development of the financial sector and human capital have the greatest predictive power. Variables associated with the public sector, local politics, or crime have a lower weight, despite being frequently associated with corruption. Finally, characteristics typically related to the Colombian context, such as armed conflict, illicit activities, or dependence on natural resources, have the least predictive weight.

We consider that the construction of this type of indicators, based on observable characteristics of the municipalities and using cutting-edge artificial intelligence models, represents an important tool in the fight against corruption. In fact, in a recent application in the context of the economic and health crisis caused by the coronavirus disease-2019 (COVID-19) pandemic in Colombia, Gallego et al. (2020b) finds

¹ See, for instance, Transparency International's *Annual Corruption Perception Index* or the World Bank's *Control of Corruption Index*.

² Municipalities are the third administrative layer in Colombia, equivalent to counties in the U.S.

that municipalities with a higher risk of corruption according to a predicted index based on a machine-learning model, show a greater increase in the use of discretionary (non-competitive) contracts to respond to the emergency. This result is consistent with the fact that several mayors have been investigated and convicted of misusing resources in the midst of the emergency, demonstrating that corruption indexes based on machine-learning models can indeed serve to anticipate in which places misconduct is more likely to occur.

From a public policy perspective, this work represents an important contribution as it illustrates how artificial intelligence can be used to allocate scarce resources. Anti-corruption audits are effective but expensive, so predicting where they will have the greatest impact may help increase government efficiency. Critical events, such as the COVID-19 pandemic, in which governments have to spend a lot in a short time, create opportunities for corruption (Gallego et al., 2021a). Therefore, these types of tools are crucial for governments to respond quickly to the transparency challenges created by these events. Additionally, compared to other studies that predict malfeasance at the contract level (e.g., Decarolis and Giorgiantonio, 2020; Gallego et al., 2021b), in this article, we focus on the municipal level, which is important given the way audits are allocated in some countries. Colonnelli et al. (2022) and de Blasio et al. (2020) also predict at the municipal level for Brazil and Italy, respectively. The advantage of our study is that by focusing on a country like Colombia, we can analyze the role played by key variables in this context, such as those related to the armed conflict, drug trafficking, or certain types of natural resources.

2. Context and Data

2.1. Context

Colombia is considered a country highly affected by corruption. It is ranked 96, out of 180 countries, in Transparency International's Annual Corruption Perception Index. Furthermore, a recent opinion poll reveals that 78% of the population considers that corruption is getting worse in the country, and 29% think that this is the worst problem in Colombia. A recent report on this phenomenon, based on national and regional press reports, found that 69% of corruption cases in the country are associated with the municipal order (Transparencia por Colombia, 2019).³ This is particularly serious, considering that since the decentralization process that began to consolidate in the 1990s, mayors began to decide on increasingly larger fractions of the public budget.⁴ In fact, mayors are in charge of providing basic public services in important areas such as education, health, drinking water, and sanitation.

The Office of the Inspector General (PGN for its Spanish acronym) is the autonomous agency in charge of monitoring the behavior of public officials so that their actions comply with the current disciplinary code. The Inspector General is elected for 4 years by the Senate from a list proposed by the President, the Supreme Court of Justice, and the Council of State. The PGN is empowered to initiate, carry out, and rule on investigations conducted against public servants due to disciplinary offenses (Cetina et al., 2020). These investigations may be induced by press news, its own audits, reports from other agencies, or tip-offs (Martinez, 2019). After starting a process, the officials assigned to the case verify and analyze the information received, assess the facts and the responsibility of the people involved, and if they find sufficient evidence, file charges. In this way, the PGN is independent of the three branches of government in Colombia. However, unlike in other countries, the assignment of anti-corruption audits is not random.

Investigations can end in sanctions and removals from office of members of the executive, including the mayors, whenever misconduct is proved. According to *Transparencia por Colombia* (2019), the four most committed crimes in recent years are embezzlement (18%), wrongful awarding and signing of contracts (13%), falsification of a public document (12%), and conspiracy to commit a crime (11%). Just

³ In contrast to 25% at the departmental level and 6% at the national level.

⁴ In Colombia, there are 1,100 municipalities and the mayors are their main authority. Currently, they are elected to four-year terms, with no possibility of immediate reelection.

to give an example, the case of Samuel Moreno, elected mayor of Bogota for the 2008–2012 period, is quite illustrative. Moreno was suspended and removed from office by the Office of the Inspector General on 2012, accused of corruption in public procurement, in particular for the construction of a public transport trunk line. Moreno is currently in prison, serving an 30-year sentence for these crimes.⁵

2.2. *Measuring corruption*

Measuring corruption is challenging (Olken, 2007; Olken and Pande, 2012). More so measuring corruption for all the districts of a particular country.⁶ Rather than relying on perceptions of citizens or key actors—a common approach to measure corruption championed by international organizations—we follow Colonnelli et al. (2022) and Gallego et al. (2021b) and use a machine-learning approach to predict corruption based on factual corruption detections and observable characteristics of municipalities.

Using information from the Office of the Inspector General of Colombia and originally collected by Martinez (2019), we construct a dummy variable indicating if the mayor of each municipality was prosecuted by this anticorruption agency in the 2008–2011 or 2012–2015 mayoral periods. We define this measure as our outcome variable and combine it with a large set of municipal features to train four machine-learning models: Random Forests, GBM, Lasso, and Neural Networks.

Note that our outcome variable corresponds to PGN prosecutions for violations of the disciplinary code. It is important to highlight several aspects of this measure. First, according to Martinez (2019), in 95% of the cases in his database where the outcome of the prosecution is observed, the investigation leads to a disciplinary sanction. Therefore, this measure is really capturing instances of misbehavior on the part of public servants, mostly mayors (70%). Second, unfortunately, with the information available, we cannot identify which of the investigations correspond to major offenses or minor ones. However, at the municipal level and for the period of interest of this study, the correlation between having an official found guilty and having one removed from office is positive, highly significant, and large in magnitude (Pearson $\rho = 0.73$). Therefore, as this is a purely predictive exercise, this high correlation implies that detecting places where an official is likely to be found guilty means it is also highly likely that an official committed a serious offense to the code. Nonetheless, an important caveat of our analysis is that although the dependent variable certainly captures the misconduct of public servants, it does not always measure severe acts of corruption.

Finally, this measure can also suffer from what the machine-learning literature has called the *selective labels problem*. As we described before, the decision of when and who to investigate is up to the PGN, with no randomness in the process. Therefore, this decision is not free from bias since investigators could go after cases in which more resources are involved, they could be intimidated or bribed by illegal actors, or they could simply make errors of judgment. Therefore, the observed outcomes are not necessarily a random sample of the population, which makes the predictive exercise more complicated.⁷ However, it is reassuring that the results presented below, both in terms of the performance of the models and the feature importance, are not substantially different from those of Colonnelli et al. (2022), who use data from proven cases of corruption detected after random audits in Brazil. An interesting avenue for future research—which is outside the scope of this study—is to use the *contraction* method proposed by Lakkaraju et al. (2017), which “facilitates effective evaluation of predictive models even in the presence of unmeasured confounders (unobservables) which influence both human decisions and the resulting outcomes.” In any case, we acknowledge that these predictive

⁵ See <https://www.eltiempo.com/justicia/delitos/cuantos-anos-de-carcel-terminara-pagando-samuel-moreno-420648> (last accessed August 24, 2021).

⁶ Ferraz and Finan (2008) and subsequent papers by the same authors measure corruption using the results of public audits in Brazil, that are only available for a subset of municipalities. In the case of Colombia, *Transparencia por Colombia* (2017) produced an index that focused on 28 department capitals.

⁷ It could even be concluded that more than measuring the probability of corruption, our outcome variable measures the probability of being investigated.

exercises will be more accurate when audits are randomly assigned, which in itself is an important policy recommendation.

2.3. Covariates

We use a total of 147 municipality-level predictors, grouped into 10 categories and measured based on the electoral period before the time at which the outcome was measured. The categories are as follows: financial sector, conflict, crime, human capital, local politics, public sector, local demographics, economic activity, illegal activity, and natural resources. The financial sector category includes per capita measures of financial sector employees, bank deposits, bank credits, housing credits, bank offices, among others. Conflict and crime variables include guerrilla and paramilitary presence and attack indicators, demobilized combatants, kidnappings, homicides, robberies, and so forth. The human capital dimension includes several educational features.

Local politics refers to electoral variables such as the number of candidates, margin of victory, voter turnout, among others. The public sector dimension includes judiciary indicators, expenditures, transfers, and revenues, and state capacity indices. Local demographics refer to population, density, rurality, inequality, child development, access to public services, and so forth. In the case of economic activity, we include GDP measures for different sectors and nighttime lights. The illegal activity dimension includes variables related to coca production and illegal mining. Finally, natural resources refer to features describing the oil, gold, and palm sectors plus some measures of deforestation. Supplementary Table A1 includes the complete list of covariates used to train the models.

3. Machine-Learning Models

In this section, we describe the machine-learning models used to predict corruption as well as the training procedure and the different measures we use to assess the performance of the different models.

3.1. Models

In order to predict municipality-level corruption, we train a set of popular machine-learning models, which include Random Forests, Gradient Boosting, Neural Networks, and Lasso. Each of these models has its own weaknesses and strengths, and therefore we also rely on an ensemble model that combines the predictive power of all individual models to optimize the overall performance (Friedman et al., 2001). We ultimately allow the data to inform which of the models is best suited for this application based on their out-of-sample performance.

3.1.1. Lasso

The Lasso regression, first developed by Tibshirani (1996), is similar to a logistic regression model, but adds a penalization term based on the sum of the absolute values of the coefficients and a penalization term based on the sum of the square of the parameters. By adding these penalization terms, the parameters of the model are shrunk toward zero, leading to a more parsimonious model than the logistic regression. In this way, we end up with a simple and less prone to over-fitting model. The tuning parameter in the cross-validation are the weight of the penalization terms in the objective function (λ) and the relative weight of the absolute sum of coefficients (α) as the penalization term.

3.1.2. Random Forests

Random Forests are ensembles of many decision trees, where each one of them is a sequence of rules that divides the sample into leaves, that is, sub-groups, based on certain variable cutoffs. The prediction for each leaf, in the case of a classification task, is the most common outcome for the trained observations on that leaf. The trees are fit with the aim maximizing the information gain of the resulting partitions of the data. In the case of Random Forests, each tree is constructed by sampling a random subset of the training data and a

random subset of the predictors. Each of these trees end up generating a prediction, and the overall prediction of the Random Forest is the average (or the majority) of the predictions among all trees. In this application, we keep fixed the number of fitted trees (500) and use cross-validation to determine the optimal number of features available in every node.⁸

3.1.3. *Gradient Boosting Machine*

GBMs are ensembles of weak learners, in this case, decision trees. Under boosting, classification algorithms are sequentially applied to a reweighted version of the training data (Friedman et al., 2000). GBM is a variant of Random Forests, in which trees are not fitted randomly nor independently. Instead, each tree is fitted sequentially to the full dataset, in such a way that the weaknesses of trees are identified by using gradients in the loss function, allowing subsequent predictors to learn from the mistakes of the previous ones. In other words, a gradient descent procedure is used to minimize the loss when adding new trees. As opposed to Random Forests, in this case observations are not selected via bootstrapping, but as a function of past errors. By doing this each new tree offers a slight improvement of the model (Freund et al., 1999). In our models, we keep fixed the learning rate (shrinkage parameter) and the minimum number of observations in the terminal nodes to avoid overfitting, and use our cross-validation procedure to determine the optimal number of trees and the interaction depth.

3.1.4. *Neural Networks*

Neural networks capture the relationship between input and output signals through models that mimic the way biological brains work. These models are composed of three basic elements: an activation function, that for each neuron, transforms the weighted average of input signals (predictors) into an output signal; a network topology, which is composed by the number of neurons, layers, and connections used by the model; and a training algorithm, which determines the way in which connection weights are set with the task of activating or not neurons as a function of the input signals. This process determines the final prediction of the model. The optimization problem seeks to find the optimal weights of the input signals for each node. In our analysis, we keep fixed a logistic activation function and use cross-validation to determine the optimal number of units in the hidden layer (size) and the regularization parameter (decay).⁹

3.1.5. *Super learner ensemble*

Ensembles are collections of models which are grouped to each other, to give a final prediction. It is usually the case that ensembles—as they result from the combination of different models—perform better than their individual components. For our analysis, we use the Super Learner ensemble method developed by Polley et al. (2011). This model aims to find an optimal combination of individual models by minimizing the cross-validated out-of-bag risk of these predictions. Van der Laan et al. (2007) show that this ensemble model performs asymptotically as well as the best possible weighted combination of its constituent algorithms. Finally, we also use the Super Learner model not only to stack the individual predictions, but also to test for the relative importance of different groups of variables to predict politician's misconduct.

3.2. *Training and testing*

We use an indicator variable for politician's misconduct in mayoral term t as our variable of interest and all the predictors are measured as averages within the politician term. In this way, we end up with a

⁸ See Marsland (2011, chapter 13) for more details.

⁹ See Marsland (2011, chapter 3) for more details.

cross-sectional dataset with all the municipalities for the mayoral periods 2008–2011 and 2012–2015. To train our models, we conduct the following steps:

1. We divide our dataset into a training set that uses 70% of the data and a testing set that uses 30%.
2. In our training set we perform a fivefold cross-validation procedure to train our models and choose the optimal combination of parameters. This method divides the training set into five different equal-size samples at random. Then, a model is fit in four subsamples and then test it in the remaining one. We then repeat this procedure for each of the five subsamples, so each one of them end up being a validation set, and for each of the values of the tuning parameter grid of each model. Finally, the best-performing parameters are chosen.
3. We repeat the previous step 10 times with different random partitions. In this way, we are able to obtain 10 “optimal parameters.” Then we use as our optimal parameter the average of them. In the case of integer parameters, we round it to the closest integer.
4. Based on these optimal parameters we assess the performance of our models in the test set that has never been used for training purposes.

We standardize the data by the mean and standard deviation. [Table 1](#) shows the optimal parameters from our training procedure for each of our models.

3.3. Assessing models' performance

Once we have calibrated our model following the procedure explained above, we proceed to compare the performance of the different models using the test set. Our first performance measure of interest is the area under the ROC curve (AUC). This measure captures the trade-off between the true positive rate and false positive rate, as we vary the discrimination threshold. It can also be interpreted as the probability that, if we randomly select two observations, they will be correctly ordered in their predicted risk of corruption, that is, the probability that the municipality at a greater risk for corruption is assigned a higher probability of corruption. We complement this measure with each model's level of *accuracy*, which is defined as the proportion of municipalities correctly predicted as corrupt; model's *sensitivity* which is the proportion of actual positives identified correctly (true positives over true positives plus false negatives); model's *precision* which is the proportion of correctly identified positives (true positives over true positives plus

Table 1. Model's parameters

Model	Optimal parameters
Lasso	λ : 1
Random Forest	α : 0.011 Trees: 500 Mtry: 4
Gradient Boosting	Trees: 200 Depth: 9 Shrinkage: 0.01 Min obs: 10
Neural Networks	Size: 1 Decay: 0.9
Ensemble weights	Lasso: 0.53 Random Forest: 0.47 Gradient Boosting: 0 Neural Networks: 0

Notes. This table presents the optimal parameters for each of the prediction models we implement after the training procedure described in Section 3.2.

false positives), and models' *specificity* which is the proportion of actual negatives identified correctly (true negatives over true negatives plus false positives).

3.4. Identifying best predictors

We begin by assessing the individual municipality characteristics that best predict corruption. In the case of tree-based models, importance is measured as the information gain, achieved when splitting on each variable. In this case, importance is measured on a scale from 0 to 100, where 100 is value for the most important predictor and the rest of the variable's information gain is expressed relative to the variable with the highest one. For the Lasso model the importance is measured by the estimated coefficients, where larger coefficients (in absolute value) correspond to higher importance. Finally, for Neural Networks, importance is determined by the weights that connect neurons within the network.

We then move to the analysis of the predictive power of subgroups of related municipality characteristics to understand which categories matter the most. It could be the case that some groups do not have one particular variable that highly predicts corruption, but that the group as whole has a high predictive power. To do this, we estimate models including one category at the time (i.e., excluding all covariates that are not part of it) and compute the resulting AUC for the group. Then, we are able to rank them according to their AUC, and compare the computed AUC with a 50% level, which corresponds to the AUC of a random prediction "model." In this case, the group that increases the AUC by itself the most is the model with the highest level of predictive power. Finally, we assess the statistical difference in the predictive power between groups by computing confidence intervals at a 95% confidence level. We do this using a bootstrap procedure over the test set and computing the AUC for each sample.

4. Findings

In this section, we present the main results of our analysis. First, we focus on the overall performance of the predictive models. Then, we identify the best individual and group predictors and their link to the corruption literature.

4.1. Models' performance and the predictability of mayors' misconduct

Figure 1 plots the ROC curves for each of the four models and for the ensemble. There are several aspects to highlight: first, all the curves are far enough away from the 45° curve, which would correspond to a

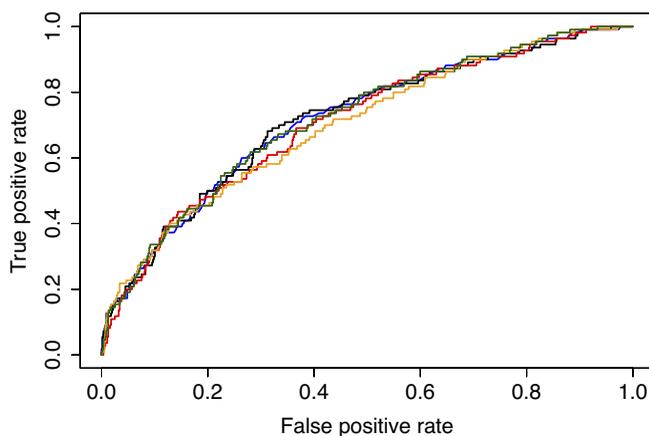


Figure 1. ROC curve. This figure presents the ROC curves for all our models. In blue, we present the ROC curve for the Random Forest model, in black, for the Gradient Boosting Machine, in red, for Lasso, in orange, for Neural Networks, and in green, for the Super Learner.

Table 2. Model performance

Model	Lasso	Random Forest	Gradient Boosting	Neural Networks	Ensemble
AUC	0.71	0.72	0.72	0.70	0.72
Accuracy	0.84	0.84	0.84	0.81	0.84
Sensitivity	0.50	0.63	0.58	0.39	0.63
Specificity	0.85	0.85	0.85	0.87	0.85
Precision	0.09	0.14	0.13	0.29	0.11

Notes. This table presents the model performance for all our prediction models. *AUC*, *accuracy*, *sensitivity*, *precision*, and *specificity* are defined in Section 3.3.

naive classifier that generates a false positive for each true positive. Second, the Neural Network achieves the worst performance of all according to this metric, which explains why it is not used in the ensemble, as shown in Table 1. Finally, the performance of the remaining models, in regards to the area under the ROC curve (AUC), is similar and acceptable, without being outstanding.

Table 2 corroborates the previous assertion. In terms of the AUC, Random Forest, GBM, and the ensemble achieve the highest performance (0.72), compared to the Neural Network, which achieves the lowest (0.70). Accuracy is similar in all five models, with a hit rate of 84% for Random Forests, GBM, Lasso, and the ensemble, and 81% for the Neural Network. Table 2 reports three additional metrics, sensitivity, specificity, and precision. The metrics suggest that the models tend to predict more false positives, as suggested by the low levels of precision.¹⁰ In sum, although the models do not reach performance levels as high as in other studies,¹¹ an accuracy of 84% and an AUC of 0.72 are still acceptable. Consequently, these models could be used by the authorities to decide where to conduct anti-corruption audits.

4.2. What are the best predictors of mayors' misconduct?

The analysis of the features that have the greatest predictive power to anticipate where mayors are most likely to misbehave, is divided into two parts: First, we group the 147 municipality-level characteristics into 10 dimensions of interest, to determine how much predictive power each one has. For this, we define the following dimensions: public sector, human capital, economic activity, local demographics, financial development, local politics, natural resources' exposure, illicit activity, crime, and conflict. We then disaggregate the analysis, by studying which individual characteristics, in each model, have the greatest predictive power.

Figure 2 shows the results of the first analysis. Surprisingly, variables associated with financial development rank first, followed by the measures of local demographics, local politics, and human capital. The result is surprising since dimensions that the literature usually associates with corruption, such as those related to the public sector, crime, or conflict, occupy intermediate positions (Rose-Ackerman and Palifka, 2016; Fisman and Golden, 2017). Shaxson (2007), for example, suggests that the resource curse is explained by the higher levels of corruption that exist in countries where these resources are abundant. Other studies suggest that the size and quality of the public sector is a determining factor in the level of corruption (Robinson and Verdier, 2013; Colonnelli et al., 2020b; Gallego et al.,

¹⁰ This result suggests that the models tend to over predict false positives. From a public policy perspective, the severity of this problem depends on the cost of conducting anti-corruption audits. If they are very expensive, models that generate a high number of false positives are not desirable because many resources would be wasted. On the other hand, if the audits are cheap, it is preferable that the false positives be numerous if that means that there are few false negatives. In any case, the researchers have some level of control over this result because the models can be tuned to optimize metrics such as accuracy.

¹¹ Colonnelli et al. (2020a), in the context of Brazilian municipalities, achieve performance metrics usually above 95%. The difference may be that in the Brazilian case, there is better information on the characteristics of the private sector.

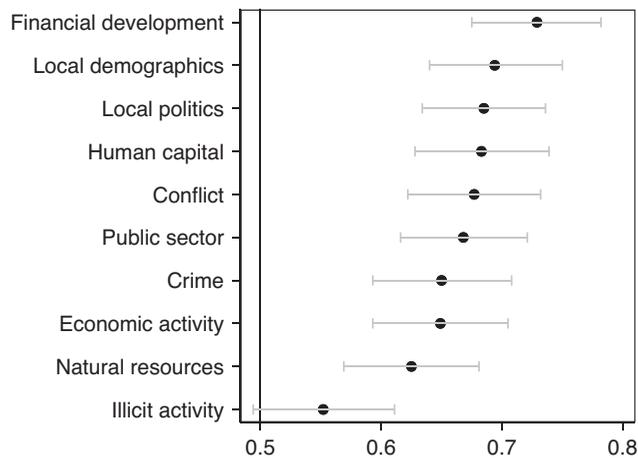


Figure 2. Group importance. This figure presents the relative importance of group of covariates as described in Section 3.4.

2020a). However, our results challenge all these explanations by suggesting that the main red flags of corruption are in the financial sector (Cooray and Schneider, 2018).

Moreover, variables usually associated with the Colombian context, such as those related to the dependence on natural resources and illicit activities, occupy the last places in this ranking, challenging the view that other manifestations of state weakness would predict where corruption is most likely to occur. In sum, these results corroborate what Colonnelli et al. (2022) find for Brazil, in the sense that variables associated with the private sector have a preponderant weight when explaining the level of corruption within a country, while the characteristics related to the public sector have less weight. These findings contrast with the great emphasis that is usually given to the public sector and public officials when thinking about anti-corruption reforms (Olken and Pande, 2012).

Finally, Figure 3 shows which individual variables have the greatest predictive power for each machine-learning model. It is interesting that in three out of four cases, a financial sector variable ranks first: The number of financial sector workers for the Random Forests and the GBM, and the size of the housing credit market in Lasso. Other financial variables appear consistently in the models, like the number of bank offices.¹² This result reaffirms what was found above, in the sense that the level of development of the financial sector is key to understanding why some places are more corrupt than others.¹³

Delving into why the development of the financial sector is an important predictor of corruption is outside the scope of this study, among other things, because this exercise is purely predictive and not causal. However, we propose a hypothesis: the degree of concentration and competition in the financial sector, reflected in the size of its main players, can determine the levels of rent-seeking, money in politics, and influence in government decisions. Anecdotal evidence from the Colombian case would give suggestive support to this hypothesis. The AVAL group, the leading financial conglomerate in the country, was related to the so-called Lava Jato scandal and the Brazilian multinational Odebrecht. According to judicial investigations, the Brazilian company and officials from the financial group bribed

¹² Although the set of variables varies from model to model, it is also noteworthy that the municipal development index appears consistently in all cases.

¹³ It is important to clarify that these graphs only tell us which variables matter most in each model, but not what is the direction of the relationship. To determine this direction, it is necessary to calculate the *partial dependency plots* of each variable. We omit such graphs to save space, but are available upon request.

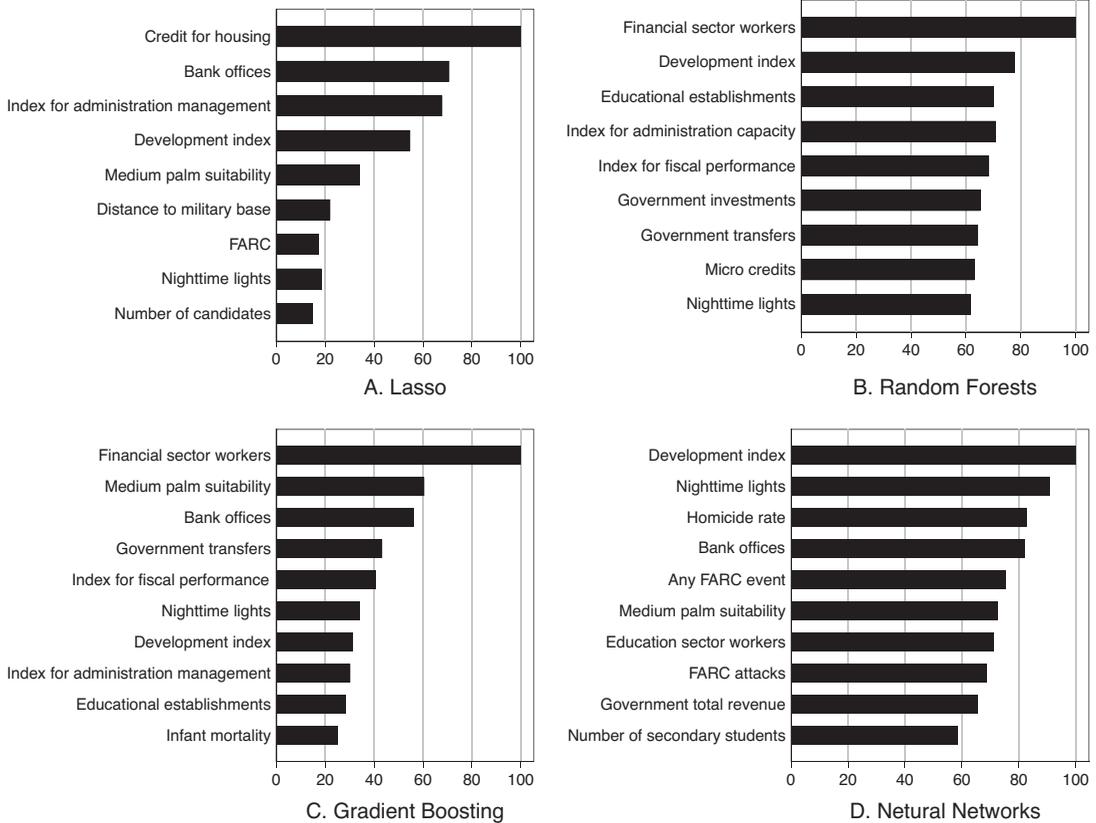


Figure 3. Covariates importance. This figure presents the relative importance of covariates as described in Section 3.4.

public servants from the Ministry of Transportation in 2009 to be awarded the construction of a major highway in the country.¹⁴

Finally, we would like to underscore one of the main limitations of our analysis. As we said before, the exercise presented here is predictive, but not causal. This fact means that these types of tools inform in what kind of places acts of corruption are more likely to occur (e.g., depending on the development of the financial sector). However, the models hardly shed light on the type of reforms or interventions (in that sector) that would help control corruption. To answer this type of questions, causal inference tools such as randomized controlled trials or quasi-experimental methods are still useful, some of which are strengthened by machine learning, such as the doubly robust models proposed by Belloni et al. (2014).

5. Conclusions

In this article, we propose the use of artificial intelligence tools to predict where rulers are more likely to commit acts of corruption. We apply these methods to the Colombian case, exploiting the fact that municipal mayors manage a significant fraction of public resources and are frequently involved in corruption scandals. Using information from prosecutions conducted by the Office of the Inspector General against these mayors, we trained four canonical machine learning algorithms, and ensembled

¹⁴ See <https://colombiareports.com/amp/colombias-prosecution-under-fire-for-absolving-ceo-of-corrupt-banking-giant/> (last accessed August 2, 2022).

their predictions, to forecast where there is a greater risk of corruption. The performance of our models is good and allows us to understand what are the features of municipalities that have the greatest predictive power to anticipate where will misconduct occur. Surprisingly, variables associated with the financial sector have the greatest weight, features related to the public sector play a secondary role, while characteristics associated with armed conflict, illicit activities, and dependence on natural resources have the least predictive power.

From a public policy perspective, we consider these tools to be particularly useful. Anti-corruption audits, carried out by independent agencies, have proven to be efficient in curbing this phenomenon (Olken, 2007; Ferraz and Finan, 2008, 2011) and in improving economic performance (Colonnelli and Prem, 2022). But audits are expensive and therefore a scarce resource. Its use must be optimized so that its effectiveness is the highest. The use of artificial intelligence tools, as illustrated in this article, helps to fulfill this purpose, especially during crises such as pandemics, wars, and natural disasters, in which governments must spend a lot and in a short time, which creates opportunities for corruption.

Acknowledgments. We thank Misión de Observación Electoral, Contraloría General de la República, and Luis Martínez for sharing with us the data used in this project. Erika Corzo and Andrés Rivera provided excellent research assistance. We also thank seminar participants at the World Bank and University of Pennsylvania.

Funding Statement. This work received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing Interests. The authors declare no competing interests exist.

Author Contributions. Conceptualization: J.G., M.P., J.V.; Data analysis: J.G., M.P., J.V.; Data curation: J.G., M.P., J.V.; Methodology: J.G., M.P., J.V.; Writing: J.G., M.P., J.V.

Data Availability Statement. The replication data that support the findings of this study will be available at <https://osf.io/vfdhj/>.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/dap.2022.35>.

References

- Belloni A, Chernozhukov V and Hansen C** (2014) High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28, 29–50.
- Bertrand M, Djankov S, Hanna R and Mullainathan S** (2007) Obtaining a driver’s license in India: An experimental approach to studying corruption. *The Quarterly Journal of Economics* 122, 1639–1676.
- Cetina C, Garay L, Salcedo-Albaran E and Vanegas S** (2020) “La analítica de redes como herramienta de integridad: el caso de la Procuraduría General de la Nación en Colombia,” Policy Brief No. 22 CAF.
- Colonnelli E, Gallego J and Prem M** (2022) What Predicts Corruption? In *A Modern Guide to the Economics of Crime*, P. Buonanno, P. Vanin and J. Vargas (eds). Cheltenham: Edward Elgar Publishing.
- Colonnelli E, Lagaras S, Ponticelli J, Prem M and Tsoutsoura M** (2022) Revealing corruption: Firm and worker level evidence from Brazil. *Journal of Financial Economics* 143, 1097–1119.
- Colonnelli A and Prem M** (2022) Corruption and firms. *Review of Economics Studies* 89, 695–732.
- Colonnelli E, Prem M and Teso E** (2020b) Patronage and selection in public sector organizations. *American Economic Review* 110, 3071–3099.
- Cooray, A. and F. Schneider** (2018). Does corruption throw sand into or grease the wheels of financial sector development? *Public Choice*, 177, p.p. 111–133.
- De Blasio G, D’Ignazio A and Letta M** (2020) “Predicting Corruption Crimes with Machine Learning. A Study for the Italian Municipalities,” Working Paper.
- Decarolis F and Giorgiantonio C** (2020) “Corruption red flags in public procurement: new evidence from Italian calls for tenders,” Working Paper.
- Ferraz C and Finan F** (2008) Exposing corrupt politicians: The effects of Brazil’s publicly released audits on electoral outcomes. *The Quarterly Journal of Economics* 123, 703–745.
- Ferraz C and Finan F** (2011) “Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance,” Working Paper.
- Fisman R and Golden M** (2017) *Corruption. What Everyone Needs to Know*. New York: Oxford University Press.
- Freund Y, Schapire R and Abe N** (1999) A short introduction to boosting. *Journal-Japanese Society for Artificial Intelligence* 14, 1612.
- Friedman J, Hastie T and Tibshirani R** (2001) *The Elements of Statistical Learning, Vol. 1, Springer Series in Statistics*. New York, NY: Springer.

- Friedman J, Hastie T, Tibshirani R.** (2000) Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics* 28, 337–407.
- Gallego J, Li C and Wantchekon L** (2020a) *A Theory of Broker-Mediated Clientelism*. Mimeo, Bogota, Colombia.
- Gallego J, Prem M and Vargas J** (2020b) *Corruption in the Times of Pandemia*. Mimeo, Bogota, Colombia.
- Gallego J, Prem M and Vargas J** (2021a) Pandemic Corruption: Insights from Latin America. In “*Procurement in Focus: Rules, Discretion, and Emergencies*”. Eds O. Bandiera, E. Bosio, and G. Spagnolo London: CEPR Press.
- Gallego J, Rivero G and Martínez J** (2021b) Preventing rather than punishing: An early warning model of malfeasance in public procurement. *International Journal of Forecasting* 37, 360–377.
- Lakkaraju H, Kleinberg J, Leskovec J, Ludwig J and Mullainathan S** (2017) “The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 275–284.
- Marsland S** (2011) *Machine Learning: An Algorithmic Perspective*. Boca Raton: Chapman and Hall/CRC.
- Martinez L** (2019) *Sources of Revenue and Government Performance: Evidence from Colombia*. Mimeo, Chicago, U.S.
- Olken B** (2007) Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy* 115, 200–249.
- Olken B and Pande R** (2012) Corruption in developing countries. *Annual Review of Economics* 4, 479–509.
- Polley EC, Rose S and Van der Laan MJ** (2011) Super learning. In *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, pp. 43–66.
- Robinson J and Verdier T** (2013) The political economy of clientelism. *Scandinavian Journal of Economics* 115, 260–291.
- Rose-Ackerman S and Palifka BJ** (2016) *Corruption and Government: Causes, Consequences, and Reform*. New York: Cambridge University Press.
- Salles M and Delles D** (2020) Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly* 37, 101407.
- Sequeira S and Djankov S** (2014) Corruption and firm behavior: Evidence from African ports. *Journal of International Economics* 94, 277–294.
- Shaxson N** (2007) Oil, corruption and the resource curse. *International Affairs* 83, 1123–1140.
- Svensson J** (2005) Eight questions about corruption. *Journal of Economic Perspectives* 19, 19–42.
- Tibshirani R** (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.
- Transparencia por Colombia** (2017) “Índice de Transparencia Municipal. Resultados 2015-abril 2016,” Tech. rep., Corporacion Transparencia por Colombia.
- Transparencia por Colombia** (2019): “Así se mueve la corrupci'on. Radiografía de los hechos de corrupci'on en Colombia 2016–2018.” Tech. rep., Transparencia por Colombia.
- Van der Laan MJ, Polley EC and Hubbard AE** (2007) Super learner. *Statistical Applications in Genetics and Molecular Biology* 6, 1–23.