A machine learning approach to find the determinants of Peruvian coca illegal crops

Débora Belén Cipriano Romero[a*], Yadira Gina Melo Estrella[a], María Isabel Zambrano Laureano[a], Rubén Ángel Ruiz Parejas[b] and Jimmy Alberth Deza Quispe[c]

[a]*Graduate student, Computer and Systems Engineering Program, Faculty of Engineering, Universidad Continental, Huancayo, Junín, Peru*
[b]*Full-Time Professor, Computer and Systems Engineering Program, Faculty of Engineering, Universidad Continental, Huancayo, Junín, Peru*
[c]*Assistant researcher, Business Program, Faculty of Management Sciences, Universidad Continental, Huancayo, Junín, Peru*

CHRONICLE	ABSTRACT
	The current study analyzed the determinants of the Peruvian coca illegal plantations in the period 2003-2019. Hence, the DEVIDA database variables were gathered at first. Then, a machine learning-based technique is employed to select the most relevant variables for the study. That technique, Lasso, selected as accurate variables eradication of coca plantations and pasta base. Both OLS and VAR are employed to analyze the relevance of the selected variables. OLS finds that eradication was negatively related to the dependent variable. Nonetheless, pb confiscation had a positive relationship with illegal coca crops. Furthermore, VAR encounters that only pb confiscation affected the dependent variable. Supplementary tests are carried to ensure the accuracy of the results. In consequence, it is concluded that eradication policies by themselves were not enough to discourage the coca plantations. Farmers should get instruction about alternative crops and financial help. Furthermore, it has been claimed that pb confiscation generates scarcity of the drug, which elevates its price. Thus, coca farmers are more motivated to plant coca because of the higher prices. Therefore, as long as the international demand, which is disposed to pay high prices, the coca illegal crops and its illicit products will exist.

1. Introduction

The coca leaf is a controversial plant. For many people and authorities, it is directly linked with cocaine. Efforts have been carried to destroy coca plantations. One of the leaders of this effort is the Drug Enforcement Administration which implements measures to decrease drugs. Nixon´s war declaration in the 1970s created this office to control both plants and substances both in the U.S. and in allied countries like Perú (Daniels, 2006). Nixon´s war declaration in the 1970s created this office to control both plants and substances both in the U.S. and in allied countries like Perú (Daniels, 2006). Hence, in the 1982, Peru establish the Programa de Erradicación del Proyecto Especial de Control y Reducción de Cultivos Ilegales en el Alto Huallaga or CORAH (Rojas & Parra, 2018). That office controls and reduces coca plantations to decrease their main product, cocaine. Actions initially have been focused on the Alto Huallaga, where is a high-productive coca zone. However, the coca business expanded to the Valle de los Ríos Apurímac, Mantaro y Ene, VRAEM, which has been an issue for the government since 2014 (Zevallos, 2017). Hence, the Comisión Nacional para el Desarrollo y Vida sin Drogas,

DEVIDA, have been proposed by the Peruvian government (DEVIDA, 2017). Also, Peruvian law punishes heavily for engaging in the drugs business (Soberón, 2010).

Although efforts have been made to eradicate the coca crops, it still survives (Zevallos, 2017). Attractive international prices, corruption, and weak policies may be the reasons why the drugs business has not disappeared (Thompson & Uggen, 2012). Also, the coca leaf has been almost sacred for Peruvians. Incas used to call it sacred leaf because they employed it in religious ceremonies and social relationships (Palacios, 2003). The traditional way to consume the coca leaf is by "chacchar" or chewing. The Andean people consume coca to get energy for hard labor (Palacios, 2003). Furthermore, the coca leaf is employed in medicine locally and internationally (Zuleta & Daza, 2018). In consequence, the Peruvian government has created the Empresa Nacional de Hoja de Coca, ENACO. This company buys and resells coca leaf to national and international companies for legal uses (Zevallos, 2017). However, there are considerable gaps in amount and price between ENACO and Narcos (Glave & Rosemberg, 2005). Also, in the towns, people usually exchange coca leaf without any control. Furthermore, coca is a wealth resource for farmers in certain zones (Castillo, 2012). The anti-drugs policy has been focused on eradicating illegal coca plantations, controlling chemicals and cocaine ingredients. Moreover, it heavily punishes the transport and exchange of coca leaf, cocaine components, and cocaine. Further, the authorities encourage farmers to put alternative plantations like coffee or fruits (Ramírez et al., 2005). However, those efforts have not been enough to eradicate and control drug traffic (Ortiz, 2009). Cocaine traffic is a multibillionaire business that is valued at more than 100 billion dollars (McDermott et al., 2021). Drugs dealers like Pablo Escobar became infamous for the high quality of their cocaine. Sadly, the Peruvian coca leaf is a key element for one of the most valued cocaine in the world. Peruvian pasta base, cocaine hydrochloride, and pure cocaine are exported to the USA, Europe, and Asia at US$ 1000 per kilo (DEVIDA, 2020). Once they arrive at their destiny, they can cost more than 57 times their original price (McDermott et al., 2021). Consequently, the economic benefits are too high for farmers and drug dealers. However, dirt wealth is not the only consequence of cocaine traffic. It generates violence and death to innocent people from fights between bands, human traffic, and other harmful consequences (National Institute on Drug Abuse, 2016).

In summary, the coca leaf has benefits, but also many negative effects. Then, if there were not coca leaves available to drug producers, there would not exist drugs. Since drug producers employ huge quantities of coca leaves to produce cocaine, what are the variables that affect the coca leaf's quantity destined for drug production? The current study will seek to answer that question by employing a machine learning tool.

2. Literature Review

2.1 Previous Studies

Sandoval et al. (2009) analyzed the determinant of the cocaine offer in Colombia. That study employed variables like cocaine price, air aspersions, cocaine confiscations, manual eradication, and destroyed labs. He found that destroyed cocaine labs, cocaine confiscation, and prices were important determinants for cocaine traffic. Both destroyed labs and confiscation harmed cocaine traffic, but the price was a real enhancer for it. Pedroni & Verdugo (2011) analyzed the unrecorded and formalized economic activity impact on illicit coca production. They have found that illicit coca production harms formal production. However, the incomes of the formal sector do increase at the same time. Also, they encountered that government stimulus for legal coca production harms it. Moreno-Sanchez et al. (2003)analyzed the coca leaf eradication policy in Colombia. They found that instead of reducing coca plantations, they increased. They encountered that coca farmers compensated for the destruction of crops by cultivating more extensively. The main reason for farmers to insist on coca plantation was the price. However, they encountered that financial stimulus given to farmers for alternative plantations had a positive effect on the eradication of coca crops. Reyes (2014) estimated the causal effect of forced eradication on coca cultivation in some Colombian cities. Although they found a negative impact on coca production, it was less than 1% of eradication per 1% in coca cultivation. Listerman (2014) encountered that it was necessary to implement an alternative crop program in the places where the government eradicates coca leaves. The impact of coca plantations on protected areas was documented by Garcia-Yi (2015). That study found that farmers took advantage of policies that protected natural areas. They encountered that the main reason to do such activity was economic.

2.2 Variables definition

2.2.1 Illegal coca leaf crops

The current study will reference it as *crops*. Those crops grew significantly in the analyzed period (Devida, 2020). Since they are not controlled, coca leaves are employed in drug elaboration. In Peru, the crops are placed principally in the Alto Huallaga and VRAEM. Actually, with the help of international agencies, many policies have been taken to vanish these crops (DEVIDA, 2017).

2.2.2 Eradication policies

These policies are focused on eliminating the illegal coca leaf plantations (Lévano, 2019). Both the Peruvian police and army participate in activities related to eradicating illegal coca leaf crops. As in many parts of the world, it is common that

those activities exacerbate violence since farmers pay groups to protect their crops (Cubides, 2015). Eradicated coca crops data do not provide information about supplemental activities to prevent coca replanting. Hence, it is impossible to know if those activities provide tools to farmers for abandoning coca plantations (Quispe, 2019). In the current study, it will be referred to as *eradication*.

2.2.3 Pasta base confiscation

Pasta base is employed as an ingredient for more refined cocaine or directly consumed (Pascale et al., 2014). Confiscation of this product has become an element of the Peruvian antidrug policy (DEVIDA, 2017). Often, the confiscation takes place at airports or roads (DEVIDA, 2017). In the current study, the pasta base confiscation will be referred to only as *pb confiscation.*

3. Materials and Methods

3.1 Theoretical approach

3.1.1 Ordinary least square

The OLS regression is expressed:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{1}$$

where $y, x_1, \ldots x_k$ represent the observed variables, u is the error, and $\beta_0 + \beta_1 + \beta_2 \ldots \beta_k$ are the estimated parameters (Wooldrige, 2010). The OLS model must meet the following conditions:

The first condition is linearity. It checks the linear relationship between the independent variables and the dependent variable (Burton, 2020). The F test is employed to ensure that relationship. Furthermore, variable error terms should not be correlated (Burton, 2020). The Durbin-Watson test analyzes that. Moreover, variable residuals need to follow a normal distribution, which is a precondition for the parametric statistical analysis (Fonti, 2017). Also, errors normally distributed ensure the maximum efficiency in the OLS results. The next condition is called homoscedasticity. The variable variance must be constant around the OLS error variance (Burton, 2020). It avoids OLS estimator inefficiency and standard errors miscalculations (Yang et al., 2019). The final condition, multicollinearity, means that independent variables do not have to be perfectly related (Burton, 2020). Hence, OLS results become stable and clear. With the help of the Variance Inflation Factor test, multicollinearity was checked.

3.1.2 LASSO

Least Absolute Shrinkage and Selection Operator or Lasso is a machine learning method. It seeks the optimal independent variables (Fonti, 2017). Lasso is classified as a supervised machine learning approach (Benvenuto et al., 2018)). Moreover, Lasso mixes wrapping and filtering methods (Fonti, 2017). Therefore, Lasso classifies and erases selected coefficients to zero and lowers the prediction error (Tibshirani, 1996). The advantages of Lasso are the equilibrium between variance and bias; and the building of an optimal regression model, hence, erasing non-necessary regressors (Fonti, 2017). It is expressed as:

$$Min = [(||Y - X\beta||_2^2)/n] \text{ subject to } \sum_{j=1}^{k} ||\beta||_1 < t \tag{2}$$

where $t >= 0$, is the upper limit of the coefficients sum. It also can be written as:

$$\hat{\beta}(\lambda) = argmin_\beta \left(\frac{||Y - X\beta||_2^2}{n} + \lambda ||\beta||_1 \right) \tag{3}$$

where $|Y - X\beta||_2^2$ is equivalent to $\sum_{i=0}^{n}(Y_i - (X\beta)_i)^2$; and $||\beta||_1$ can be expressed as $\sum_{j=1}^{k} |\beta_j|$ y $\lambda >= 0$. The λ –lambda-parameter controls the penalty force; hence, the longer λ, the longer the constriction. Furthermore, λ and its upper bound, t, are negatively related. Consequently, in case the upper bound, t, goes to infinity, λ becomes 0. Inversely, if t is 0, λ tends to infinity; thus, coefficients become zero (Fonti, 2017). When coefficients are zero $[\hat{\beta}(\lambda) = 0]$, they are erased. Otherwise, they are kept. Therefore, the independent variables are provided.

The λ selection techniques most times are three. They are cross-validation or CV, adaptive selection or AS, and Minimum Bayes criterion or MinBIC. Due to the data length, only CV and AS were employed for this research. Lasso CV begins with the entire sample divided into two. They are known as training and testing sub-samples (Reitermanová, 2010). It is employed to ensure a consistent estimation of models´ performance. The CV splits the sample into 10 folds. One fold is chosen, and a linear regression is built in the other folds (Stata, 2019). Next, regression coefficients predict the chosen fold data. With that data, the other folds are estimated. Therefore, 10 average squared errors or MSE are obtained. The CV needs those ten MSE on average (Stata, 2019).. Only when the minimum λ is found, the CV function stops. Thus, the λ with the highest

prediction power and the littlest MSE is chosen. AV employs the CV approach to select p, but more frequencies are needed. Hence, multiple Lassos are performed at the same time. As stated before, coefficients with zero values are deleted after each *LASSO,* and the others are preserved. However, they get penalty weights to make them become zero in the next Lasso. Hence, only coefficients with enough data are always preserved. The p with the best prediction power and the least error is chosen (Stata, 2019).

3.1.3 Vector Autoregressive Model

This is a method employed to analyze the relationship in a set of variables. This is a combination of many autoregressive models characterized for forming a vector between the studied variables (Hossain & Kamruzzaman, 2015). This model is suggested to be applied in time series data because it describes the relationship between variables at different times (Papanicolas & McGuire, 2011). It is expressed as:

$$y = \Gamma_1 X_{t-1} + \varepsilon_t. \tag{4}$$

There are three important assumptions when employing the model (Stock & Watson, 2001). They have to be stationary at the first difference, data error must be distributed normally, and third variables must keep independence. To check the first assumption, the Unit Root test is employed. Error normal distribution is evaluated through the Jarque Bera test. Finally, autocorrelation analysis is by employing the Lagrange test. Moreover, Lütkepohl (2007) states that VAR is useful to analyze several variables at the same time. Also, it does not allow segregation between the dependent variable and independent variables. Hence, a general form to express VAR is:

$$y_t = c + \varphi(B)y_t + \varepsilon_t$$
$$y_t = c + (\varphi_1 B + \varphi_1 B^2 + \cdots + \varphi_p B^p)y_t + \varepsilon_t \tag{5}$$
$$\varepsilon_t \sim N(0, \Sigma)$$

3.1.4 Granger Causality Test

This test analyzes the determination of the cause and effect relationship between the studied variables (Niyazbekova et al., 2016). The test is made of the following bivariate regressions:

$$V_{s,t} = \alpha_1 + \sum \beta_i V_{s,t-1} + \sum \rho_i V_{M,t-1} + \varepsilon_t \tag{6}$$
$$V_{M,t} = \alpha_2 + \sum \theta_i V_{s,t-1} + \sum \kappa_i V_{M,t-1} + \varepsilon_{2t} \tag{7}$$

According to Niyazbekova et al. (2016), these possible situations can be presented: causality of V_M to V_S, V_s to V_M, bidirectional causality, and no causality. To find the determinants of illegal coca leaf production, a set of variables have been harnessed. They have been gathered from the DEVIDA web page. Those variables were confiscated coca leaves, cocaine ingredients, eradicated coca crops, confiscated cocaine, confiscated pb, cocaine price, coca leaf price, and ENACO acquired coca leaves. After that, the Lasso technique was employed to select the best determinants of illegal coca crops. Next, OLS, VAR, and Granger tools have been employed. Supplemental tests were employed to guarantee the accuracy of the employed tools.

4. Results

Table 1 shows the descriptive statistics of the variables employed. Along with Fig. 1, it is observed that coca leaf's crops had a maximum volume in 2011. In the whole period, there were 51 thousand hectares of coca leaf plantations. The eradication of illegal coca leaf plantations had a peak in 2015, but it decreased in the next few years. On average, there were about 18 thousand hectares of illegal plantations. The pb confiscation reached a peak in 2012, with a huge drop in the following year. The lowest ever recorded confiscation was 3 tons which were too far from the average confiscation of more than 11 tons.

Table 1
Descriptive Statistics

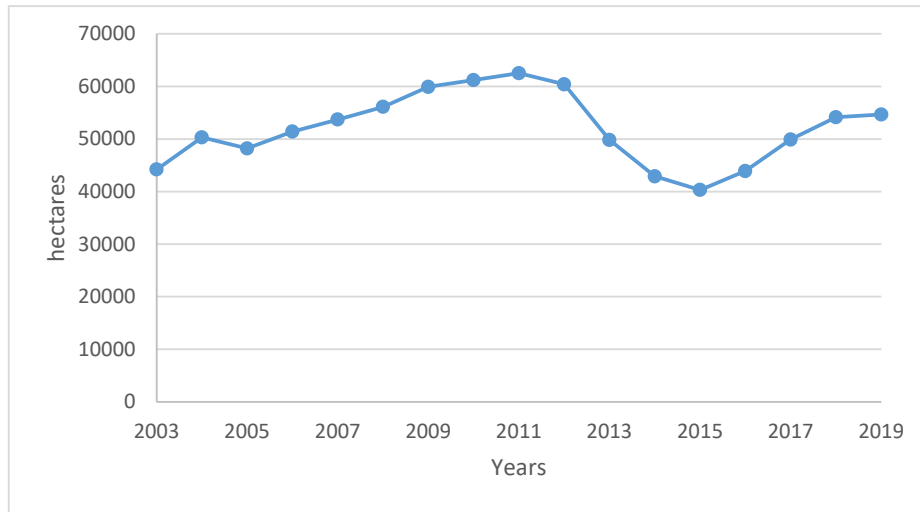| Measures	Variables		
	Crops*	Eradication*	Pb confiscation**
Mean	51969.94	18396.38	11005.59
Median	51400.00	12688.00	11375.00
Max	62500.00	35868.00	19700.00
Min	40300.00	10025.00	3199.00
Standard deviation	6764.97	8926.30	4616.11
		* hectares	** kilograms

Fig. 1. Coca crops evolution
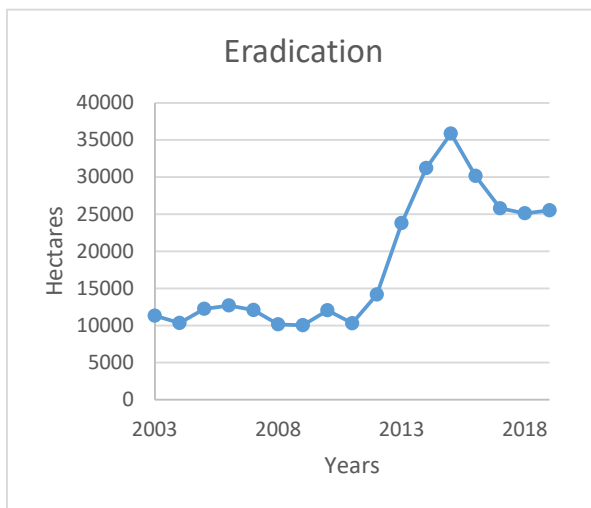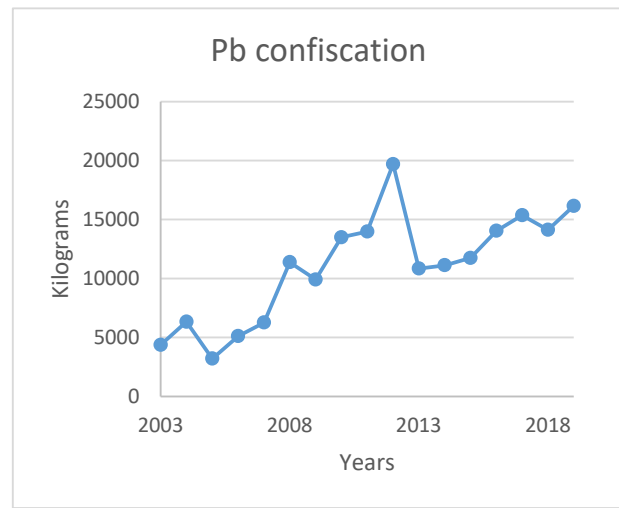

Fig. 2. Eradication

Fig. 3. Pb confiscation

Table 2
Lasso estimation

LASSO technique	ID	Lambda	Number of non-zero coefficients	Out of sample R2	Cross validation mean prediction error
Cross Validation	1	4105.561	0	0.090	4.70E+07
	26	401.117	5	0.764	1.02E+07
	*27	365.483	5	0.766	1.01E+07
	28	333.015	5	0.764	1.02E+07
	30	276.475	5	0.755	1.05E+07
Adaptative	32	18751.820	0	0.168	5.03E+07
	63	1048.378	2	0.772	9.82E+06
	*64	955.243	2	0.773	9.80E+06
	65	870.382	2	0.772	9.83E+06
	109	14.519	5	0.740	1.12E+07

Table 2 depicts the Lasso estimation. It showed that the Cross-validation technique selected the id 27. In the adaptive method, it was 64. Both of them had the highest out-of-sample R2 and the lowest mean prediction error.

Table 3
Lasso post-estimation

Lambda technique selection	Mean Squared Error	R2
Cross-validation	6742177.000	0.844
*Adaptive lasso	4185586.000	0.903

Table 4

Selected variables

Variables	Cross-validation	Adaptive lasso
Eradication	√	√
Pb confiscation	√	√
Coca leaf confiscation	√	
Price pb	√	
Cocaine confiscation	√	
Constant	√	√

Table 3 portrays the selected method. It was chosen the adaptive method since it had the lowest means squared error and the highest R2. Hence, the chosen variables were eradication and pb confiscation, as showed in Table 4.

Table 5

OLS Regression

Variables	Coefficient	Standard Error	t	p>t	95% confidence interval	
Eradication	-0.70	0.09	-7.97	0.00	-0.89	-0.51
Pb confiscation	1.08	0.17	6.36	0.00	0.71	1.44
Constant	52959.09	2033.09	26.05	0.00	48598.55	57319.62
F	37.72			0.00		
R2	0.84					

The OLS regression portrayed in Table 5 shows that eradication had a negative relationship with the illegal crops. Nonetheless, pb confiscation had a positive relationship with the dependent variable. The coefficient of determination was 84%, and the significance of the F test gave confidence about the building of the model.

Table 6 depicts the assumption test necessary to ensure that the OLS model is correctly built. The autocorrelation test rejected the alternative hypothesis of autocorrelation. The distribution test showed that data fulfilled the normal residual distribution. Moreover, the homoscedasticity test showed that data did not present homoscedasticity issues. The VIF test proved that there were no multicollinearity issues. Finally, the specification test shows that the regression model has been correctly specified.

Table 6

OLS Assumption tests

Autocorrelation test		
chi2	df	p>chi2
0.11	1.00	0.74
Distribution test			
Skewness	Kurtosis	chi2	p>chi2
0.27	0.58	1.72	0.42

Homoscedasticity test	
chi2	p>chi2
0.47	0.49

Multicollinearity test		
Variable	VIF	1/VIF
Pb confiscation	1.19	0.84
Eradication	1.19	0.84
Mean	1.19	

Specification test	
F	p>F
0.94	0.4534

In consequence, the suggested model provided by Lasso fulfills the OLS assumption. Hence, an accurate model has been built to expose the determinants of coca illegal crops. After that, it is necessary to probe if there was a cause-effect relationship between the selected variables. Hence the Johansen test was employed.

Table 7

Lag selection test

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-407.83				6.20E+21	58.69	58.68	58.83
1	-382.28	51.10	9.00	0.00	6.10E+20	56.33	56.27	56.87
2	-359.94	44.68	9.00	0.00	1.20E+20	54.42	54.33	55.38
3	-342.18	35.517*	9.00	0.00	7.4e+19*	53.1681*	53.0413*	54.5375*

According to Table 7, the number of suggested lags was 3. This is because the majority of parameters arrived at that number. In consequence, three lags have been employed in the next tests.

Table 8
Stationary test

	Variable	Test statistic	1% c. value	5% c. value	10% c. value	p>z
At level	Crops	-1.88	-4.38	-3.60	-3.24	0.33
First difference		-3.78	-3.75	-3.00	-2.63	0.00
At level	Eradication	-2.08	-3.75	-3.00	-2.63	0.25
First difference		-3.20	-3.75	-3.00	-2.63	0.02
At level	Pb confiscation	-1.51	-3.75	-3.00	-2.63	0.53
First difference		-5.95	-3.75	-3.00	-2.63	0.00

All the selected variables had root issues at the level. Nonetheless, they were stationary at the first level. Hence, it meets the assumption detailed in the methodology part.

Table 9
Johansen Test

Maximum rank	Parameters	LL	Eigenvalue	Trace statistic	5% critical value
0	10	-207.5243	.	6.7443*	15.41
1	13	-204.8234	0.32012	1.3426	3.76
2	14	-204.1521	0.09145		

The Johansen test shows that the Var is the most accurate model since the trace statistic was lower than the critical value, according to Table 9.

Table 10
VAR

	Variable	Coefficient	Standard Error	z	P>z	95% confidence interval	
Crops	Crops	0.36	0.57	0.64	0.52	-0.75	1.47
	Eradication	0.26	0.44	0.58	0.56	-0.61	1.12
	Pb confiscation	-1.25	0.67	-1.87	0.06	-2.57	0.06
	Constant	42556.81	30593.79	1.39	0.16	-17405.91	102519.50
Eradication	Crops	0.20	0.44	0.45	0.65	-0.66	1.06
	Eradication	0.39	0.34	1.14	0.25	-0.28	1.06
	Pb confiscation	1.34	0.52	2.57	0.01	0.32	2.36
	Constant	-10369.40	23716.30	-0.44	0.66	-56852.50	36113.70
Pb confiscation	Crops	0.74	0.27	2.78	0.01	0.22	1.26
	Eradication	0.67	0.21	3.21	0.00	0.26	1.08
	Pb confiscation	-0.52	0.32	-1.64	0.10	-1.14	0.10
	Constant	-32044.89	14404.88	-2.22	0.03	-60277.94	-3811.85

Table 10 shows the Var results. There is a possible cointegration between the dependent variable and pb confiscation. Moreover, eradication seemed to have a cointegration with pb confiscation. Analogously, pb confiscation was cointegrated with crops and eradication. Hence, it can be claimed that there was a short-term relationship between pb confiscations with crops and eradication. The causality test in Table 11 shows that pb confiscation affected crops. Also, it affected eradication. Finally, crops and eradication both affected pb confiscation.

Table 11
Granger causality test

Equation	Excluded	chi2	df	prob>chi2
Crops	Eradication	0.33964	1	0.56
	Pb confiscation	3.4885	1	0.062
	All	7.0876	2	0.029
Eradication	Crops	0.20085	1	0.654
	Pb confiscation	6.6011	1	0.01
	All	33.758	2	0
Bp confiscation	Crops	7.7525	1	0.005
	Eradication	10.322	1	0.001
	All	10.526	2	0.005

Table 12
Supplementary tests

Distribution test			
Equation	chi2	df	p>chi2
Crop	0.378	2	0.82793
Eradication	0.551	2	0.75912
Pb confiscation	0.686	2	0.70948
All	1.615	6	0.95147
Autocorrelation test			
Lag	chi2	df	p>chi2
1	12.9193	9	0.16629
2	14.2644	9	0.11322
3	10.898	9	0.28276

Table 12 shows the test that proved that both VAR and Granger analyses were correctly specified; thus, they are valid. The distribution test shows that all the studied variable distributions were normally distributed. Also, the autocorrelation test showed that the data employed at the third lab did not have any autocorrelation issue.

5. Discussion

The current study found that the Lasso technique selected pb confiscation and eradication to explain the coca leaf's crops. In consequence, the OLS regression found that eradication had a negative relationship with the dependent variable, while the pb confiscation had a positive one. Hence, eradication policies have an important impact on diminishing the coca loaf plantations destined for illegal activities. However, pb confiscation seemed to increase the coca plantations. Sandoval et al. (2009) also found that eradication policies had an important impact on the cocaine supply, but that study found a negative impact of those variables. Nonetheless, Moreno-Sanchez et al. (2003) found that eradication policies increased coca plantations. Listerman (2014) claimed that the lack of alternative crops and financial aid move farmers to coca crops. Also, Garcia-Yi (2015) encountered that coca farmers are moved to cultivate coca leaves because of economic benefits. Pedroni & Verdugo (2011) stated that pb economic benefits move coca farmers to produce it.

The subsequent analysis employing VAR and Granger test gave the study a better understanding of the phenomena. Var found a light cointegration between crops and pb confiscation at 10% of significance. Also, eradication and pb confiscation seemed to be cointegrated at 5% of significance. Moreover, crops and eradication seemed to be cointegrated with pb confiscation. The Granger analysis is more explicit to detail the relationships between the variables. Pb confiscation affected crops, while eradication policies did not have any at all. Moreover, pb confiscation affected eradication policies. Also, those relationships were inverse because crops and eradication had effects on pb confiscation. This result did not match with the findings of Reyes (2014) since that study found a causal effect of eradication on coca leaf production.

Hence, eradication itself is not enough to fight against coca plantations. It must be accompanied by alternative crops policies and economic stimulus, as previous studies stated. Furthermore, the effect of pb confiscation on coca leaf crop might be due to the market demand.

6. Conclusion

This study had two parts. The first one was focused on employing a machine learning technique to select variables. After that, the current study employed both OLS and VAR to validate the Lasso variable selection. OLS provided a positive relationship of eradication policies on coca crops. Also, it noticed a negative relationship between pb cultivation and crops. VAR found that only pb confiscation affected the coca crops, but at the same time, it encountered a causal effect between pb confiscation and eradication. In consequence, it is possible to state that eradication policies by themselves did not have a strong effect on lowering the coca crops. As previous authors stated, eradication policies should be accompanied by supplemental programs like credit stimulus and education on alternative crops. In Perú, those programs were focused on coffee plantations, but coffee is very susceptible to plagues and inconsistent prices (Gonzales, 2004). Hence, some former coca farmers returned to the coca plantations (Ibanez, 2012). In consequence, it is also necessary to industrialize other plantations away from coffee. Pineapple products, exotic plants beverages, and high-quality natural products can be alternatives for farmers. Again, it is not possible if credits are not available for farmers. Here, land formalization is a key factor.

Both OLS and VAR found that pb confiscation affected coca crops. OLS regression noticed that it was negatively related to the dependent variable possible because of economic factors. Here, rich countries pay exorbitant sums for illegal coca products (). When pb confiscation arises, scarcity is generated. Therefore, it increases the prices of the most refined products. In consequence, farmers increase their coca production to meet the new demand (Bruner & Johnson, 2014)

The previous results show that coca production can be lowered by eradication, but it needs alternative programs to convince farmers that there are legal ways to get economic benefits. However, pb confiscation might have the undesirable effect of stimulating coca production, as explained before. Hence, illegal coca production will exist as long as there are people who require it. In consequence, the best way to tackle coca illegal production is awareness and education about the bad effects of drugs.

Conflict of interest

The authors state that they do not have any conflicts of interest.

References

Benvenuto, F., Piana, M., Campi, C., & Massone, A. M. (2018). A Hybrid Supervised/Unsupervised Machine Learning Approach to Solar Flare Prediction. *The Astrophysical Journal*, *853*(1), 90. https://doi.org/10.3847/1538-4357/aaa23c

Bruner, N. R., & Johnson, M. W. (2014). Demand curves for hypothetical cocaine in cocaine-dependent individuals.

Psychopharmacology, *231*(5), 889–897. https://doi.org/10.1007/s00213-013-3312-5

Burton, A. (2020). OLS (Linear) Regression. In *The Encyclopedia of Research Methods and Statistical Techniques in Criminology and Criminal Justice*. Wiley.

Castillo, M. (2012). La Economía de la coca. In *Las relaciones internacionales de la pobreza en América Latina y el Caribe*. CLACSO.

Cubides, O. (2015). La violencia del narcotráfico en los países de mayor producción de coca: los casos de Perú y Colombia. *Papel Político*, *19*(2), 657. https://doi.org/10.11144/javeriana.papo19-2.vnpm

Daniels, A. (2006). EE UU y la guerra contra las drogas en Latinoamérica. *Política Exterior*, *20*(112), 131–140. http://www.jstor.org/stable/20645955

Devida. (2020). *Informe sobre la demanda de hoja de coca para fines tradicionales e industriales*. 214. https://cdn.www.gob.pe/uploads/document/file/1348630/Informe - Demanda Hoja de Coca.pdf

DEVIDA. (2017). Estrategia Nacional de Lucha Contra las Drogas 2017-2021. In *Devida*. https://www.devida.gob.pe/documents/20182/314196/Estrategia_FINAL_castellano2.pdf

DEVIDA. (2020). *Monitoreo de precios de hoja de coca y derivados cocaínicos en zonas estratégicas de intervención*.

Fonti, V. (2017). Feature Selection using LASSO. *VU Amsterdam*, 1–26. https://doi.org/10.1109/ACCESS.2017.2696365

Garcia-Yi, J. (2015). Drugs and protected areas: Coca cultivation and social acceptance of bahuaja-Sonene national park in Peru. *Sustainability (Switzerland)*, *7*(6), 7806–7832. https://doi.org/10.3390/su7067806

Glave, M., & Rosemberg, C. (2005). *La comercialización de hoja de coca en el Perú: análisis del mercado formal*.

Gonzales, A. (2004). Dealing With Risk: Small-Scale Coffee Production Systems In Mexico. *Perspectivas Latinoamericanas*, *1*(1), 1–39.

Hossain, A., & Kamruzzaman, A. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. *Chinese Business Review*, *14*(6). https://doi.org/10.17265/1537-1506/2015.06.001

Ibanez, M. (2012). *Who crops coca and why? The case of Colombian farmers*. August, 0–46.

Lévano, L. (2019). Instituto de Gobierno y de Gestión Pública. *Revista*, *N°4*, 0–47.

Listerman, J. (2014). Factors that Most Influence Success or Failure in Illicit Crop Reduction and Drug Supply Control. *ProQuest Dissertations and Theses*, 146.

Lütkepohl, H. (2007). Vector Autoregressions. *A Companion to Theoretical Econometrics*, *September*, 678–699. https://doi.org/10.1002/9780470996249.ch33

McDermott, J., Bargent, J., Douwe, den H., & Ramírez, M. F. (2021). *The cocaine pipeline* (Issue February). The Global Iniciative Against Transnational Organized Crime.

Moreno-Sanchez, R., Kraybill, D. S., & Thompson, S. R. (2003). An econometric analysis of coca eradication policy in Colombia. *World Development*, *31*(2), 375–383. https://doi.org/10.1016/S0305-750X(02)00192-4

National Institute on Drug Abuse. (2016). *Cocaine Research Report. May*, 1–29. https://www.drugabuse.gov/download/1141/cocaine-research-report.pdf?v=e3cddc61770c865659ddfee9d082b8b3

Niyazbekova, S. U., Grekov, I. E., & Blokhina, T. K. (2016). The influence of macroeconomic factors to the dynamics of stock exchange in the Republic of Kazakhstan. *Economy of Region*, *12*(4), 1263–1273. https://doi.org/10.17059/2016-4-26

Ortiz, A. (2009). Producción de Cocaína: Simulación económica de algunos efectos de política en los países productores. *Anales Científicos*, *70*(3), 80–84.

Palacios, E. (2003). Problema de la coca: producción, consumo y control. *Pensamiento Crítico*.

Papanicolas, I., & McGuire, A. (2011). Using a vector autoregression framework to measure the quality of English NHS hospitals. In *English* (Issue May). http://www2.lse.ac.uk/LSEHealthAndSocialCare/LSEHealth/documents/LSEHealthWorkingPaperSeries.aspx

Pascale, A., Hynes, M., Cumsille, F., & Bares, C. (2014). Consumo de pasta base de cocaína en América del Sur: revisión de los aspectos epidemiológicos y médico-toxicológicos. In *Comisión Interamericana para el Control del Abuso de Drogas. Organización de los Estados Americanos.* https://www.redalyc.org/revista.oa?id=778

Pedroni, P., & Verdugo, C. (2011). The Relationship Between Illicit Coca Production and Formal Economic Activity in Peru. *IMF Working Papers*, *11*(182), 1. https://doi.org/10.5089/9781462310289.001

Quispe, J. (2019). Impacto social del programa Devida en el desarrollo agropecuario y ambiental en la cuenca del Inambari, Puno Perú. *Revista de Investigaciones Altoandinas - Journal of High Andean Research*, *21*(1), 29–48. https://doi.org/10.18271/ria.2019.443

Ramírez, E., Bernal, M. E., León, G., Serna, M., Julian, C., & Gómez, R. C. J. (2005). Análisis Económico Del Cultivo De La Coca En Colombia. *Ideas Ambientales*, *2*, 142–156.

Reitermanová, Z. (2010). Data splitting. *WDS'10 Proceedings of Contributed Papers, Part I*, 31–36.

Reyes, L. C. (2014). Estimating the causal effect of forced eradication on coca cultivation in Colombian municipalities. *World Development*, *61*(33478), 70–84. https://doi.org/10.1016/j.worlddev.2014.03.024

Rojas, L., & Parra, D. (2018). Procesos de sustitución y erradicación de cultivos ilícitos a nivel mundial: un punto de partida para Colombia. *Khoka Alternativa*. http://administrativos.cultura.gob.pe/intranet/dpcn/consulta.jsp?pagina=11

Sandoval, L. E., López, Á., & Cárdenas, C. (2009). *Determinantes y características de la oferta de cocaina determinants and characteristics of cocaine supply Abstract Em Colômbia (1989-2006)*. *XVII*(2), 199–208.

Soberón, R. (2010). *Reformas a las leyes de drogas en América Latina Documento de trabajo , capítulo México Leyes de Drogas y Cárceles en América Latina*.

Stata. (2019). *Stata Lasso Reference Manual Release 16*. https://www.stata.com/manuals/lasso.pdf

Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, *15*(4), 101–115. https://doi.org/10.1257/jep.15.4.101

Thompson, M., & Uggen, C. (2012). Dealers, Thieves, and the Common Determinants of Drug and Nondrug Illegal Earnings. *Criminology*, *50*(4), 1057–1087. https://doi.org/10.1111/j.1745-9125.2012.00286.x

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statiscal Society*, *58*(1), 267–288.

Wooldrige, J. (2010). *Econometric Analysus of Cross Sectin and Panel Data* (Second). The MIT Press.

Yang, K., Tu, J., & Chen, T. (2019). Homoscedasticity: An overlooked critical assumption for linear regression. *General Psychiatry*, *32*, 1–5. https://doi.org/10.1136/gpsych-2019-100148

Zevallos, N. (2017). Dinámicas locales en torno al cultivo de hoja de coca: elementos para el estudio desde el mercado ilegal de la cocaína. *Revista de Ciencia Política y Gobierno*, *4*(7), 9–29. https://doi.org/10.18800/rcpg.201701.001

Zuleta, P., & Daza, D. (2018). Revisión sistemática de artículos científicos de uso medicinal, nutricional y agroindustrial de la hoja de coca y sus derivados. *DDHH Elemental*, 1–19.