

# Machine Learning Based Modelling of Economic Growth and Quality of Governance: The MENA Region

nadia Farjallah (✉ [nadiafarjallah25@gmail.com](mailto:nadiafarjallah25@gmail.com))

IHEC Sousse: Institut des Hautes Etudes Commerciales de Sousse <https://orcid.org/0000-0003-1106-8011>

Asma Sghaier

IHEC Sousse: Institut des Hautes Etudes Commerciales de Sousse

---

## Research Article

**Keywords:** Random Forest, Machine Learning, Economic Growth, Government Effectiveness, Control of Corruption and Rule of Law

**Posted Date:** August 3rd, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1772256/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

Governance (which entails transparency, accountability, the rule of law and the presence of effective and legitimate institutions) is considered an essential factor in economic development. A large number of academic studies have attempted to identify and explain the influence of governance quality on economic growth, bearing on different theoretical perspective and using a panoply of estimation methods, including correlation and regression analyses. This study approaches the phenomenon from a predictive analytical perspective using contemporary Machine Learning techniques to uncover the most important predictors of economic growth in the MENA region in sample observed from 1996 to 2020. Random Forest algorithm was used with three machine learning models (Support Vector Machine, Boosted TREE, Linear Regression) to predict economic growth. The empirical results indicated that the predictions obtained using Random Forest were more accurate than those obtained by the other models. The results indicated that Government Effectiveness, Control of Corruption and Rule of Law are the most influential factors explaining economic growth.

## Introduction

The concept of good governance has been widely debated in the literature (Agere, 2000; Graham et al., 2003a ; Armstrong et al., 2005 ; Andrews, 2008 ; Bovaird and Löffler, 2009). Good governance is one of the public sector's management models (Armstrong et al., 2005). Generally, good governance has three dimensions: political, administrative and judicial. The first represents access to authority and the last two denote the exercise of authority (Kaufmann et al., 2010). Good governance means eliminating corruption, inefficiency, ill-administrative secrecy and bureaucracy to include accountability, transparency and governance, efficiency, exclusivity, fairness and responsiveness (Stoker, 1998; Graham et al., 2003a, b). Aside from these macro indicators, Mauro (1995), Kaufmann and Wei (1999), Méon and Sekkat (2005), Méon and Weill (2010), d'Agostino et al. (2016) and Prakash et al. (2019) found a negative effect of corruption on investment, efficiency and economic growth. In this regard, Etsy Daniel (2006) claimed that good governance is the process by which public institutions conduct public affairs, manage public resources and ensure the protection of human rights in a manner that is essentially free of abuse and corruption, while respecting the rule of law. Generally speaking, good governance means that the citizen and their security are ensured by law, which is guaranteed in particular by the independence of the judicial, fostering thus the "rule of law". Governance means also that public expenditure is managed in a correct and fair way by public institutions and that information must be accessible to all citizens. Democratic institutions and regimes have common characteristics that should promote their visions of cooperation: balance of power, multiparty system and free and periodic elections, an active civil society, free and independent media, and armed and security forces under the control of the nation's representatives. In a democratic society, participation or voice of citizens in the selection of a government is a prerequisite for any democracy. In democratic governments, ethical administrative practices have always been recognized as an essential tool for determining good governance and a crucial element to be incorporated in building citizen trust. Waheduzzaman (2010) explained that the World Bank and IMF consider participation as an important indicator of good governance. Accountability is another indicator of good governance that is a barrier to building good governance (Rahaman, 2009; Ray, 1999). Accountability refers to the extent to which one should be held accountable for one's actions to high authority or the public (Shafritz & Russell, 1997). The other two indicators of good governance are transparency and responsiveness (Griffin, 2010). Machine learning is a subset of artificial intelligence, which is frequently used when computing devices attempt to mimic human cognitive functions related to learning and problem-solving processes to achieve "optimal" results. Within the machine learning community, the development of methods that

enable and improve the machine learning process is often tested by model comparison methods. In other words, the goal of the process is to identify the machine learning model and its optimal set of parameters that achieve the highest unbiased predictive accuracy for a given problem and associated data set. In a manner parallel to traditional stochastic data modelling, algorithmic culture has been exploring and improving the predictive accuracy of machine learning models for decades (Huang et al.,2014). To the best of our knowledge and to date, no academic study has tried to discover the most important predictors of economic growth in the MENA region through the lens of modern data mining techniques. The main objective of this analytical study is to reveal the likely factors and their relative importance as predictors of economic growth with modern machine learning techniques. The rest of the paper is organized as follows. Section 2 summarizes the research methodology in terms of the machine learning models and methods used in the study. Section 3 includes briefly describes the data, and data pre-processing. Section 4 presents the modelling results, explains the predictive accuracy and the order of importance of the predictors of all types of models. The final section, Section 5, summarizes the study, its findings, and provides insights and implications of the findings.

## Methodology

### 2.1. Machine Learning Model

Machine learning is a subset of artificial intelligence, which is frequently used when computing devices that attempt to mimic human cognitive functions related to learning and problem-solving processes to achieve "optimal" results. Within the machine learning community, the development of methods that enable and improve the machine learning process is often tested by model comparison methods. In other words, the goal of the process is to identify the machine learning model and its optimal set of parameters that achieve the highest unbiased predictive accuracy for a given problem and the associated data set. In a manner parallel to traditional stochastic data modelling, algorithmic culture has been exploring and improving the predictive accuracy of machine learning models for decades (Huang et al., 2014).

**2.2.1. Linear Regression (LR)** represents the gold standard technique in the field of machine learning. The ordinary least squares (OLS) method is commonly used to estimate intercept and slope regression parameters. The model can be expressed as follows:

$$\hat{y}_i = \beta_0 + \sum_{i=1}^p \beta_i X_i + \epsilon_i$$

1

where Y is the output variable,  $X_i$  is the independent variable, and  $\beta_i$  is the parameter estimated through OLS regression. The limitation of LR is that it only clarifies the interaction between the mean input variables and output variables. Nevertheless, the functional form of the link between the target variable Y and predictor X can be obtained in a flexible manner through advanced machine learning models.

**2.2.2. Random Forest Algorithm** is of the bagging [1] type predictor algorithms consisting of a set of small trees. It is characterized by its superior performance and robust results. It works well with large data sets and efficiently handles imperfect data while maintaining a high level of accuracy. Even if the generalization error occurs in a random drill tree, it depends on the predictive strength of the model's decision trees. Ensemble models are

expected to provide robust models with high predictive accuracy over individual classifiers (Svetnik et al., 2003).

At each node of the forest, the input variables are randomly selected such that the nodes are divided according to criteria of internal impurities when growing trees in the forest (Breiman, 2001).

A random forest consists of an arbitrary number (set) of single trees, or their responses are combined (averaged) to obtain an estimate of the dependent variable (regression). The use of ensembles of trees can lead to a significant improvement in prediction accuracy (i.e., better ability to predict new data cases). The response of each tree depends on a set of independently chosen predictive values (with replacement) and with the same distribution for all trees in the forest, which is a subset of the predictive values in the original data set. The votes of all trees are aggregated and successful applications of the Random Forest algorithm are reported in several domains such as e-commerce, finance, sports and medicine (Sharda et al., 2017). In the social sciences, Random Forest has attracted attention because of its high efficiency, ease of execution and low computational costs. This learning method was developed by Breiman (2001). The samples bootstrap[2] obtained from each tree were used to train the entire learning (Babar et al. 2020). The model can be expressed as follows:

$$\hat{y}_i = \frac{1}{Q} \sum_{k=1}^Q h_k(X)$$

2

Where  $h(x)$  is a set of  $k$ -th random learner trees and  $X$  is the vector of input variables. A new training set is generated by replacing the original data for each regression tree constructed. To improve the predictive ability of the model, the hyperparameters of the model must be adjusted using a validation dataset

### 1.2.3. Vast Margin Separator Regression (SVM)

Support vector machines are a set of supervised learning techniques for solving regression problems. The SVM is a learning system using a high-dimensional feature space. The SVM regression was proposed by Vapnik et al. (1997). The regression model produced by SVM depends only on a subset of training data because the cost function for building the model ignores any training data that is close to the model's prediction.

The idea is to transpose the data into another higher dimensional space and then apply the linear algorithm on the projected data.  $K$  is a positive definite kernel:

$$K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \text{ alors } : k(x, y) = \langle \phi(x), \phi(y) \rangle$$

With  $\phi : \mathbb{R}^d \rightarrow \{\text{Hilbert space}\}$ ,  $x \rightarrow \phi(x)$  is a transformation of  $\mathbb{R}^d$  to a Hilbert space  $\{\text{Hilbert space}\}$ , which we do not need to explain: any algorithm that uses only scalar products between the data samples can all be applied in the space  $H$  via the scalar product  $k(x, y) = \langle \phi(x), \phi(y) \rangle$ .

Training data  $D = \left\{ (x_i, y_i), i = 1, \dots, n \right\}$ , ou  $x_i \in X, y_i \in \mathbb{R}$ . The goal, given a new observation  $x$ , is to predict the value of the associated  $y$ . For this regression ( $\epsilon - SV$ ), we look for a function  $f : X \rightarrow \mathbb{R}$ , flat as  $|f(x_i) - y_i| < \epsilon$ , for all  $i$ . We consider the projections of the learning points in the arrival space  $\{\text{Hilbert space}\}$  and we look for solutions of the form:

$$f(x) = \langle w, \phi(x) \rangle + b$$

The flattening condition is identified by the minimum  $\|w\|^2 = \langle w, w \rangle$ . The regression  $\epsilon$ -SV is identified using the cost function  $\epsilon$ -insensitive :

$$\rho(y - \langle w, \phi(x) \rangle) = \begin{cases} 0 & \text{if } |y - \langle w, \phi(x) \rangle| \leq \epsilon \\ |y - \langle w, \phi(x) \rangle| - \epsilon & \text{if } |y - \langle w, \phi(x) \rangle| > \epsilon \end{cases}$$

Moreover, we introduce the <easing variables>  $\xi_i$  and  $\xi_i^*$  are plotted in Fig. 2. The optimization problem is:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad \text{s.t.} \quad \langle w, \phi(x_i) \rangle - b \leq \epsilon + \xi_i, \quad \langle w, \phi(x_i) \rangle - b \geq \epsilon - \xi_i^*, \quad \xi_i, \xi_i^* \geq 0, \quad i=1, \dots, n$$

With :

$C > 0$ , allows to choose the balance point between the flattening of the solution and the acceptance of errors beyond  $\epsilon$ -SVM for the regression.

The Lagrange multipliers we get are:

$$\min_{\alpha, \alpha^*} \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_j) (\alpha_i^* - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^n (\alpha_i + \alpha_i^*) \epsilon - \sum_{i=1}^n (y_i (\alpha_i - \alpha_i^*)) \quad \text{s.t.} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \leq C, \quad i,j=1, \dots, n$$

Inside of  $\epsilon$ -SVM, the learning points have  $\alpha_i = \alpha_i^* = 0$ . The dots have  $\alpha_i, \alpha_i^* \neq 0$  are called support vectors. With  $w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i)$

The regression function is:  $f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b$

**2.2.4. Boosted Trees algorithm** (*stochastic gradient boosting trees*) comes from the application of boosting methods to regression trees (GC&RT). The general idea is to compute a series of (very) simple decision trees, where each consecutive decision tree is constructed to predict the residuals of the previous tree's prediction. This method is a complete implementation of the stochastic gradient boosting method.

The implementation of these powerful algorithms can be used for regression problems, with continuous and/or categorical predictors. Among the regression procedures derived from Friedman (1999), TreeBoost has the best overall performance and was considered the method of choice. In this study, we compare different levels of randomization in terms of performance on the 100 target functions for two different error distributions. Hundreds of datasets  $\{(y_i, x_i)\}_{i=1}^N$  have been generated:

$$\widehat{Y}_i = F(x_i) + \epsilon_i$$

With  $\{F\}^{\{*\}}\left(\{x\}_{\{i\}}\right)$  represents each of the 100 randomly generated target functions. For the first study, the errors  $\{\epsilon\}_{\{i\}}$  were generated from a Gaussian distribution with zero mean, and variance adjusted so that:

$$E\left|\epsilon\right| = E\left|F\left(x\right) - \text{median}_{\{x\}}\left\{F\left(x\right)\right\}\right|$$

5

## 2.2. Testing and evaluating - cross-validation

For these methods, we used k-fold cross-validation to randomly divide the data into k number of mutually exclusive subsets for “training” and “testing” sets. The folds (k-1) of the data are used to build the model and the remaining fold is used to test the model. Delen et al. (2012) proved that a single random assignment can potentially lead to heterogeneous subsets of data which, in turn, would produce biased results. For this reason, we used five rounds (k = 10) of cross-validation on the entire dataset. In each round of the 5-fold cross-validation, the model is trained in all except one of the folds and tested on the excluded fold, which is the test subset for that specific round. Finally, the average of the results of the five rounds is compiled for the final analysis. Olson and Delen (2008) report that the use of stratified cross-validation tends to decrease bias compared to regular cross-validation. According to Delen et al. (2012), overall accuracy is measured using the average of each individual k accuracy measurement.

$$CV = \frac{1}{K} \sum_{i=1}^K A_{\{i\}}$$

5

## 2.3. Model Performance Comparison

To evaluate and compare the prediction performance of the four machine learning algorithms, we chose to use the mean squared error (MSE), which is defined as the mean for each individual of the deviation test basis squared between the prediction of the model to be tested and the true output value. For each model, we will evaluate the following quantity:

$$\{MSE\}_{\{j\}} = \frac{1}{\{n\}_{\{test\}}} \sum_{i=1}^{\{n\}} \left( \{y\}_{\{i\}} - \widehat{\{y\}}_{\{i,j\}} \right)^2$$

6

Where :

- - m : the number of models to test.
- - n : the number of individuals in the initial base.
- -  $\{n\}_{\{test\}}$  : the number of individuals in the test database.
- -  $\{y\}_{\{i\}}$  : the actual output variable of the individual.
- -  $\widehat{\{y\}}_{\{i,j\}}$  : the output variable of individual i predicted by model j.

We are interested in the model which has  $\{MSE\}_{\{j\}}: \min \left\{ \left( \{MSE\}_{\{j\}} \right) \right\}_{\{1 \leq j \leq m\}}$ . The model that minimizes the MSE error appears to be the model that best predicts the dependent variable.

[1] Breiman, L. (1996). Bagging Predictors, Machine Learning, 26 (2),123-140.

[2] The principle of the bootstrap is to draw randomly and with replacement observations from the starting base, and to repeat this mechanism several times to obtain a new sample.

## Database

The empirical study consists of 18 countries (Appendix 1) in the MENA (Middle East and North Africa) region. The study covers the period from 1996 to 2020. Table 1 describes the variables in detail, their definitions, notations and sources. The relationship between Economic Growth (LGDP), Corruption (CORR), Government Effectiveness (EG), Political Stability and Absence of Violence/Terrorism (SP), Regulatory Quality (QR), Rule of Law (ED), and Voice and accountability (VR) are expressed as:

$$LPIB = f(\text{CORR, EG, SP, QR, ED, VR})$$

Table 2 summarizes the statistics for the studied variables. EG and LPIB show the highest mean and standard deviation for all countries. The higher standard deviation also implies that the EG and LPIB series are more volatile than the other variables. The normality test is estimated by the Shapiro-Wilk statistic (the results prove that the variables are not normally distributed at the 5% threshold). Moreover, machine learning algorithms are likely to extract deeply hidden knowledge in large datasets involving multiple types of input variables that are not necessarily normally distributed. (Delen et al., 2012 ; Sharda et al., 2017).

Table 1  
Data description

The variables	Definition	Sources
Economic growth GDP (constant 2010 US\$)	GDP at purchaser prices is the sum of the gross value added of all resident producers in an economy plus any product taxes and minus any subsidies not included in the value of the products. Its data is in constant 2010 US dollars. Dollar amounts for GDP are converted from local currencies using official 2010 exchange rates.	<a href="https://donnees.banquemondiale.org/">https://donnees.banquemondiale.org/</a>
Corruption (CORR)	captures perceptions of the extent to which public power is exercised for private gain, including petty and grand forms of corruption, as well as the "capture" of the state by elites and private interests. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>
Government effectiveness (EG)	captures perceptions of the quality of public services, the quality of the civil service and its degree of independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government commitment to these policies. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>
Political stability and absence of violence/terrorism (SP)	measures perceptions of the likelihood of political instability and/or politically motivated violence, including terrorism. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>
Regulatory quality (QR)	captures perceptions of government's ability to formulate and implement sound policies and regulations that enable and promote private sector development. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>
rule of law (ED)	captures perceptions of the extent to which agents trust and obey societal rules, and in particular the quality of contract enforcement, property rights, police and courts, as well as the likelihood of crime and of violence. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>

The variables	Definition	Sources
Voice and Accountability (VR)	captures perceptions of the extent to which a country's citizens can participate in the selection of their government, as well as freedom of expression, freedom of association and freedom of the media. The estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from around - 2.5 to 2.5.	<a href="https://databank.banquemondiale.org/">https://databank.banquemondiale.org/</a>

Table 2  
Descriptive Statistics

Variables	Observations	Mean	Min	Max	Std Dev	Shapiro-Wilk stat	P-value
Economic growth GDP (constant 2010 US\$)	450	10.91489	10.11461	11.83160	0.434191	0.95676	0.0000
Corruption (CORR)	450	-0.23527	-1.71257	1.56719	0.770349	0.98017	0.0001
Government effectiveness (EG)	450	-0.18063	-2.30766	1.50927	0.803437	0.98865	0.00145
Political stability and absence of violence/terrorism (SP)	450	-0.58416	-3.18080	1.22362	1.035486	0.97564	0.0000
rule of law (ED)	450	-0.21391	-2.09213	1.27893	0.794410	0.96074	0.0000
Voice and Accountability (VR)	450	-0.93667	-2.05034	0.78666	0.610573	0.94774	0.0000
Regulatory quality (QR)	450	-0.28381	-2.34709	1.31674	0.86889	0.97065	0.0000

## Results And Discussions

We constructed artificial intelligence models to predict economic growth (LPIB). To achieve our goal, we used four machine learning approaches namely Random Forest, Support Vector Machine, Boosted TREE and Linear Regression. Training and testing data were split in a ratio of 70:30; 70% of data was used for training the model and 30% for examining the efficiency and accuracy of the model. To assess the performance of forecasting models, we compare the error rate of the models. Table 3 presents the error rate values for the test data of all models. Random Forest proves the best predictive performance, with minimum error term than Support Vector Machine, Boosted TREE and Linear Regression. This implies that Random Forest significantly improves the

accuracy of the prediction method, reaching an overall accuracy of 99.54%. Linear Regression was the second-best model with an overall accuracy of 93.89%, followed by Boosted TREE with 93.89% and Support Vector Machine with 86.89%. Like other machine learning techniques, the Random Forest method remains a "black box" method, i.e. one cannot really visualize the decision tree allowing to obtain the final prediction. This type of models offers great predictive performance, is versatile and detects interactions without a parametric form having to be specified (Hamza and Larocque 2005). Graphs (1) explain the mechanism of the Random Forest algorithm to avoid overfitting. Generally speaking, the root mean square error of the training and test data will decrease. Graph 1 shows that the regression error rate of the training sample is 0.075 while that of the test sample is 0.055. To assess the most influential factors on economic growth in the MENA region, we use the random forest-based predictor importance technique. Chart 2 shows that Government Effectiveness (100%), Control of Corruption (82.57%) and Rule of Law (80.36%) are the most important predictors influencing economic growth. In contrast, Voice and Accountability (80.13%), Regulatory Quality (70.40%) and Political Stability and Absence of Violence/Terrorism (64.24%) are the least important predictors that influence economic growth.

Table 3  
Comparison of forecasts of economic growth (GDP)

	<b>Error rate</b>
<b>Random Forest</b>	<b>0,005406</b>
<b>Support Vector Machine</b>	0,131014
<b>Boosted TREE</b>	0,065730
<b>Linear Regression</b>	0,060107

## Conclusion

In this study, we explored the most important dimensions of governance quality that affect economic growth in 18 MENA countries observed from 1996 to 2020. A large number of academic studies have attempted to identify and explain the effects, causes and potential consequences of governance quality on economic growth, primarily through lenses using correlation and regression-based statistical analyses. This study approached the phenomenon from a predictive analytical perspective using contemporary machine learning techniques to uncover the most important governance quality perception predictors that affect economic growth. The cross-validation results highlighted Random Forest as the most accurate method for predicting economic growth. Interestingly, given the notable predictive accuracy of our results, we highlight government effectiveness as the most relevant factor influencing economic growth. This implies that the quality of public services, civil service and its degree of independence vis-à-vis political pressures are the main factors that affect economic growth in the MENA region.

## Declarations

Conflict of Interest : the author declares that he has no known competing financial interests or personal relationships which might appear to influence the work reported in this article.

## References

1. Agere, S., 2000. Promoting Good Governance: Principles, Practices and Perspectives, vol.11. Commonwealth secretariat, pp. 147 ISBN 0-85092-629-7.
2. Armstrong, A., Jia, X., Totikidis, V., 2005. Parallels in private and public Sector governance. GovNet Annual Conference, Contemporary Issues in Governance, 28-30 Nov 2005 (Accessed 17 April 2018).
3. Andrews, M., 2008. The good governance agenda: beyond indicators without theory. *Oxford Dev. Stud.*36(4),379–407.
4. Armstrong, A., Jia, X., Totikidis, V., 2005. Parallels in private and public Sector governance. GovNet Annual Conference, Contemporary Issues in Governance, 28-30 Nov 2005 (Accessed 17 April 2018).
5. Bovaird, T., Löffler, E. (Eds.), 2009. Public Management and Governance. Taylor & Francis ISBN 0-203-88409-4 Master e-book ISBN.
6. Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting*, 28, 543–552. <https://doi.org/10.1016/j.ijforecast.2011.05.002>.
7. d'Agostino, G., Dunne, J. P., & Pieroni, L. (2016). Government spending, corruption and economic growth. *World Development*, 84, 190–205.
8. Esty Daniel, C., 2006. Good governance at the supranational scale: globalizing administrative law. *Yale Law J.* 115, 1490–1562.
9. Graham, J., Amos, B., Plumptre, T., 2003a. Principles for Good Governance in the 21<sup>st</sup> Century. Policy Brief No. 15. Available at: (Accessed 09 February 16). Institute of Governance, Canada.
10. Graham, J., Plumptre, T.W., Amos, B., 2003b. Principles for Good Governance in the 21<sup>st</sup> Century. Policy Brief No.15. Institute on governance, Ottawa, pp. P.8.. (Accessed 16 April 2018).
11. Huang, X., Shi, L., & Suykens, J. A. (2014). Support vector machine classifier with pinball loss. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, 984–997.
12. Sharda, R., Delen, D., & Turban, E. (2017). *Business intelligence, analytics, and data science : A managerial perspective* (4th ed.). London : Pearson.
13. V. Vapnik, S. Golowich and A. Smola, (1997), "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing", in M. Mozer, M. Jordan, and T. Petsche (eds.), *Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA.
14. Hamza, Mounir, and Denis Larocque. 2005. "An Empirical Comparison of Ensemble Methods Based on Classification Trees." *Journal of Statistical Computation and Simulation* 75 (8): 629–43. <https://doi.org/10.1080/00949650410001729472>.

15. Kaufman, D., Kraay, A., & Mastruzzi, M. (2010). The worldwide governance indicators: Methodology and analytical issues. 2010 World Bank Policy Research Working Paper, No.5430.
- 16.
17. Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3), 681–712.
18. Kaufmann, D., and Wei, C. J. 1999. Does “Grease Money” Speed Up the Wheels of Commerce? Working Paper 7093. National Bureau of Economic Research.
19. Méon, P.-G., & Sekkat, K. (2005). Does corruption grease or sand the wheels of growth? *Public Choice*, 122(1), 69–97.
20. Méon, P. G., & Weill, L. (2010). Is corruption an efficient grease? *World development*, 38(3), 244–259.
21. Rahaman, M.M.,2009. Parliament and good governance : A Bangladesh persepticve.Japanies Journal of Political.9,39-62.
22. Rav,B.,1999. Good governance, administrative reform and socio-economic realities : A South Pacific perspective. *International Journal of Social Economics*.26,354-369.
23. Prakash, N., Rockmore, M., & Uppal, Y. (2019). Do criminally accused politicians affect economic outcomes? Evidence from India. *Journal of Development Economics*, 141 102370.
24. Stoker, G., 1998. Governance as theory: five propositions. *Int. Soc. Sci. J.* 155 (42), 17–28.
25. Waheduzzaman.,2010. Value of people's participation for good governance in developing countries. *Transforming Government: People, Process and Policy*.4,386–402.

## Appendix 1

### The studied MENA countries

United Arab Emirates	Bahrain
Algeria	Egypt
Iran	Iraq
Israel	Jordan
Kuwait	Lebanon
Libya	Morocco
Oman	Qatar
Saudi Arabia	Syrian Arab Republic
Tunisia	Yemen

## Figures

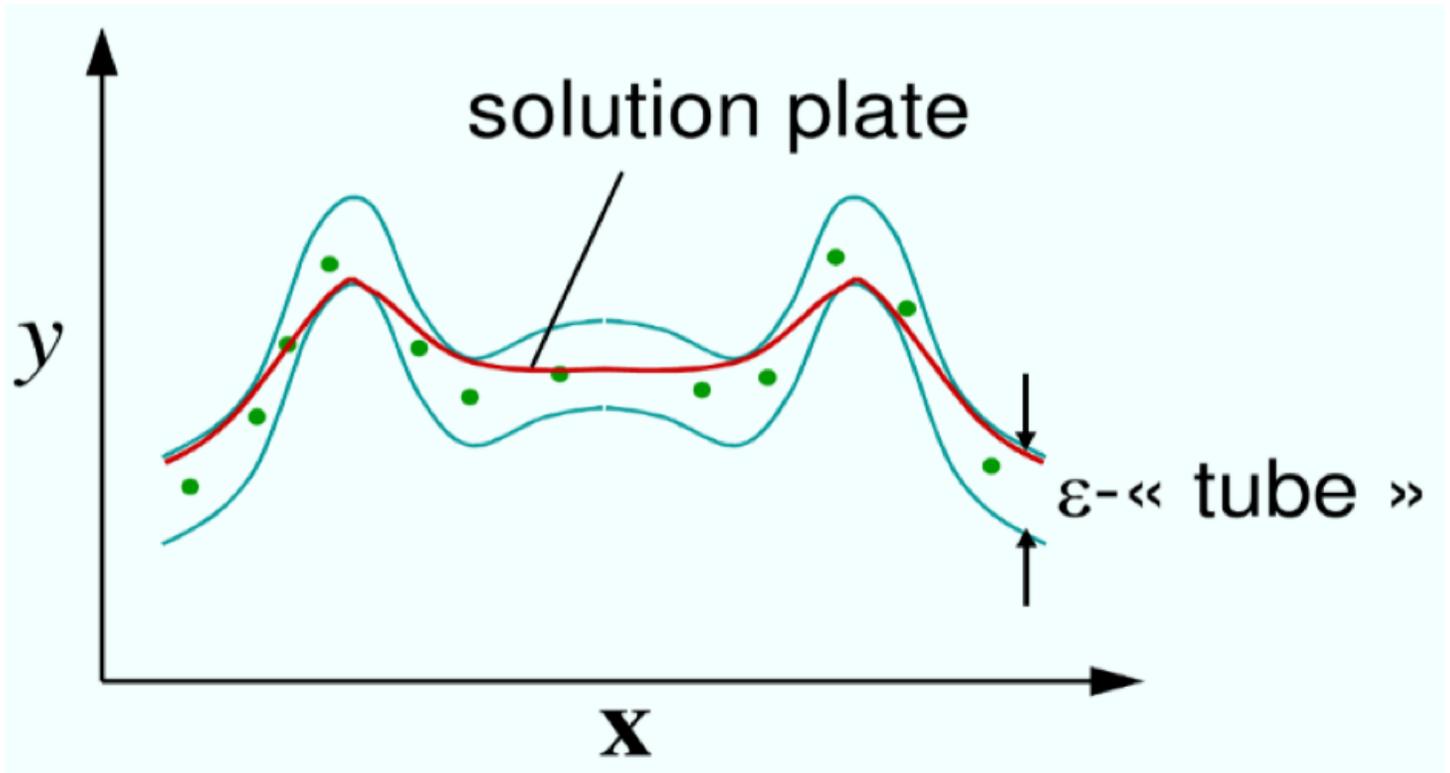


Figure 1

SVM for regression

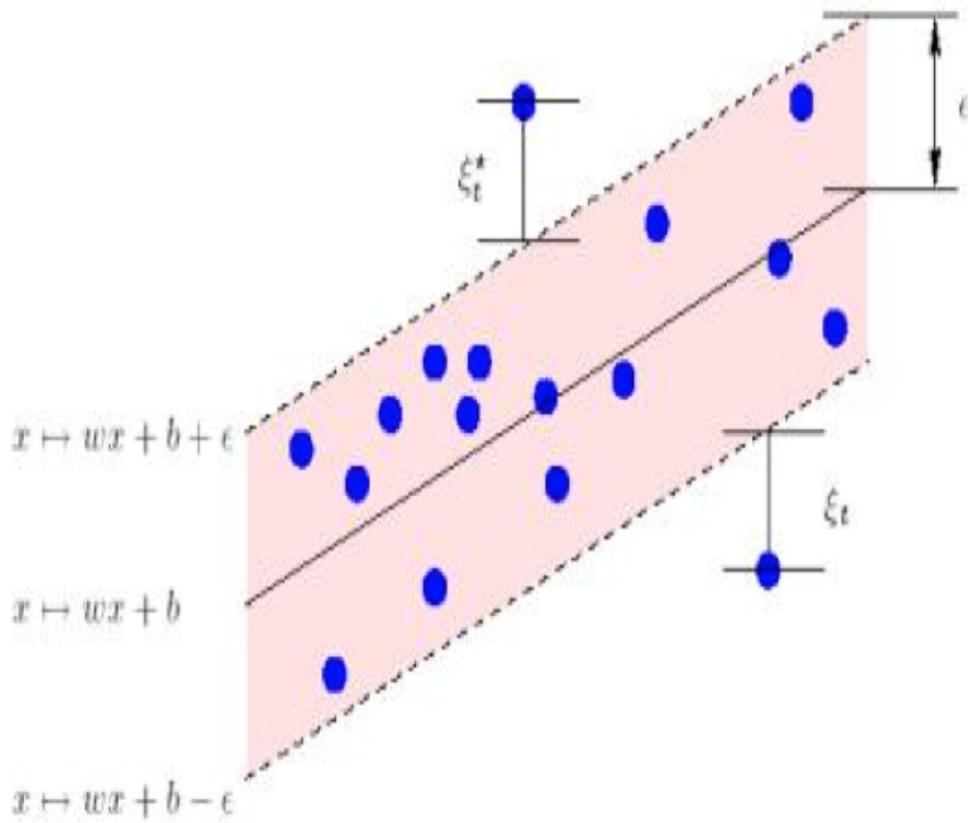


Figure 2

the relaxation variables

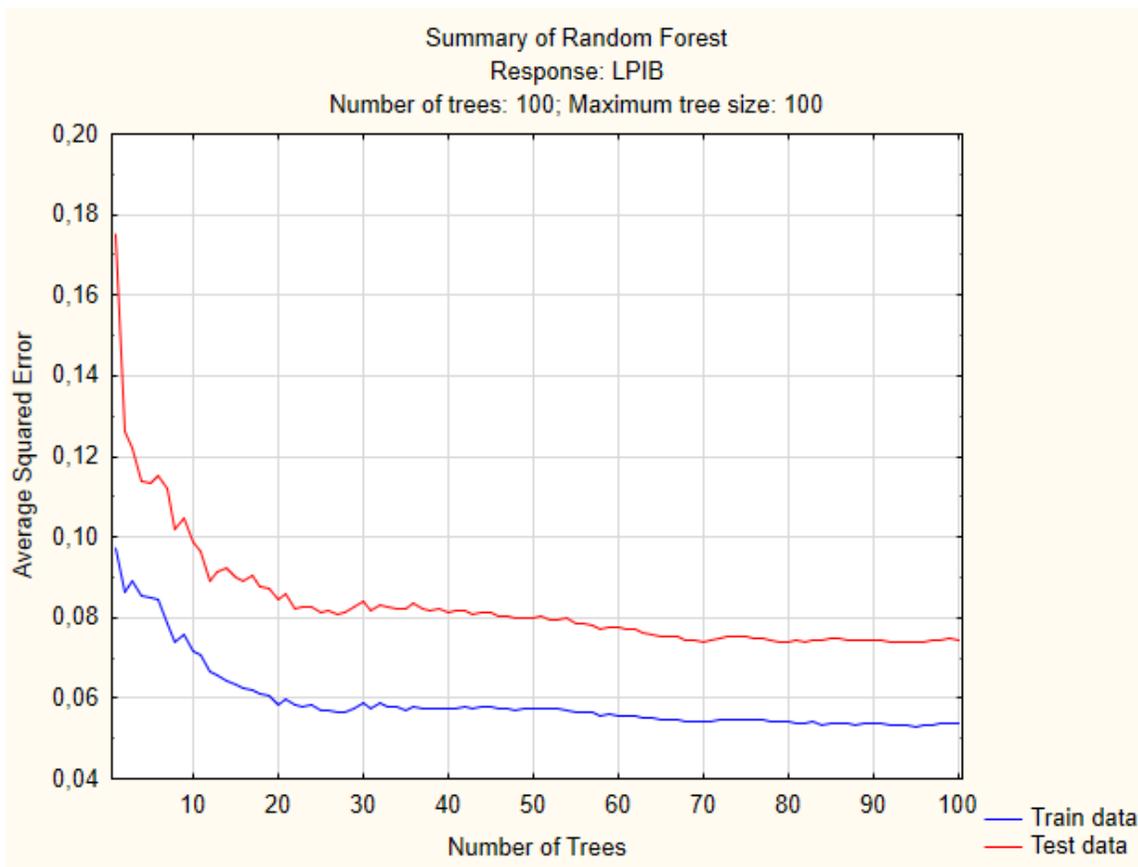


Figure 3

Figure 1: Summary of random forest response.

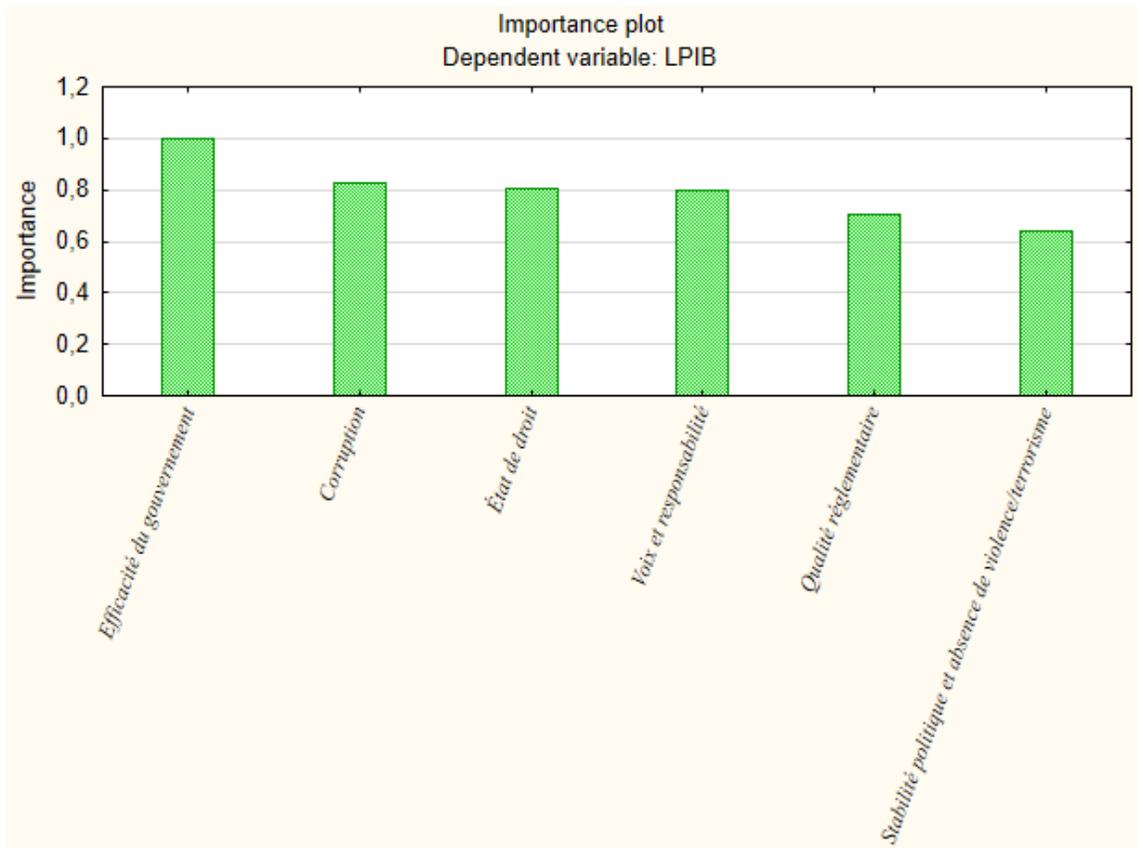


Figure 4

Figure 2 : Importance of variables