Leveraging Machine Learning for Inclusive Procurement: An Analysis of Economic Efficiency in the Dominican Republic's Public Sector

Luis M. Duran Cornelio[1]
Department of Project Management, International Ibero-American University, Arecibo, Puerto Rico, USA
The Americas Institute of Technology ITLA, La Caleta, Boca Chica, Dominican Republic
Email: luis.duran@doctorado.unib.org

Ernesto Bautista Thompson
Department of Project Management, International Ibero-American University, Campeche, México
Email: ernesto.bautista@unini.edu.mx

Abstract

This research undertakes a quantitative exploration of the impact of procurement modalities on economic efficiency and inclusive procurement within the Dominican Republic's public sector. Applying data science methodologies and machine learning techniques, this study analyzes a rich dataset of public procurement contracts spanning from 2018 to 2023. A focus is placed on understanding how different procurement modalities affect contract amounts and processing times—two key indicators of economic efficiency. Machine learning models, including Random Forest and Gradient Boosting, are employed to predict these essential metrics across different procurement modalities. The findings indicate a significant positive correlation between certain procurement modalities and economic efficiency. The machine learning models reveal that procurement strategies focused on local sourcing and transparent bidding processes tend to yield higher economic returns and more equitable outcomes. This research underscores the potential of quantitative analysis to inform procurement policies, fostering a more economically efficient and inclusive procurement ecosystem in the Dominican Republic's public sector. The insights garnered from this study could have significant implications for the design of more inclusive procurement strategies, contributing to socio-economic growth and corruption reduction in the Dominican Republic.

Keywords: Public Procurement, Machine Learning, Socio-Economic Impacts, Inclusive Procurement, Dominican Republic, Random Forest, Gradient Boosting

Introduction:

Public procurement, representing the structured process through which governments engage with the private sector for goods and services, serves as a potent mechanism for socio-economic planning, fostering social equity, and mitigating corruption. In the Dominican Republic, where public procurement accounts for 17% of total government expenditures

[1] Corresponding author

(Mercado, 2021; OECD, 2020), it becomes an essential lever for targeted socio-economic development and strategic planning.

Inclusive procurement, a socially responsible sourcing strategy, promotes participation from traditionally marginalized groups such as SMEs (Small, Medium, and Micro-sized Enterprises) and women-owned enterprises (Silva et al., 2021). By fostering competition and diversifying the supplier base, inclusive procurement can enhance economic efficiency, stimulate growth, and limit opportunities for corruption (Borsatto et al., 2020).

In the Dominican territory, SMEs constitute 98.4% of registered production and service units (Mercado, 2021). Despite their economic significance, they have often encountered barriers in securing government contracts, largely due to limited awareness of opportunities and financial and technical constraints (Carril et al., 2022). Furthermore, large companies, representing only 1.6% of registered units, contribute 61.7% of formal employment, (Mercado, 2021) highlighting the economic disparity between SMEs and large enterprise owners.

The General Directorate of Public Procurement (DGCP) is the governing body of the DR's public procurement system for goods, works, and services. The Dominican government, recognizing the importance of inclusive procurement, has enacted several initiatives. Law No. 340-06 and Decree No. 543-12 laid the groundwork for SMEs participation in government procurement, with mandates to all public institutions to set aside 20% of purchases for SMEs. These legislative efforts, coupled with DGCP's technology-driven initiatives like the procurement portal and open data portal, have enhanced SMEs' capacity to compete for government contracts. With these efforts, tangible impacts have been observed, with SMEs' share of government contracts increasing from 30% in 2017 to 50% in 2022, and women-owned SMEs' participation rising from 15% to 25% (Colman, 2020). These strides contribute to job creation and economic empowerment.

Despite these advances, the complexity of procurement processes calls for a more thorough and nuanced understanding. This research aims to quantitatively analyze the critical relationships between different procurement strategies and key performance indicators, focusing on operational efficiency and social inclusion within the Dominican Republic's public sector.

The main research questions guiding this study are:

i. How do different procurement modalities impact contract amounts and processing times in the Dominican Republic?
ii. How do these impacts influence the overall efficiency and inclusiveness of public procurement in the Dominican Republic?

To address these questions, the research will utilize data science exploratory analysis and two machine learning models, to transform data into actionable insights. This approach will enable us to scrutinize variations across different procurement modalities while providing new insights into a more efficient and equitable public procurement landscape. The research holds the potential to inform data-driven policy-making in procurement, thereby contributing to robust socio-economic planning and enhanced transparency in the Dominican Republic.

Literature Review:

The complex interplay between public procurement strategies, inclusive procurement, operational efficiency, and the application of quantitative models like machine learning forms the core of this interdisciplinary research in the domain of socio-economic planning. Public procurement modalities have evolved over the years, with a shift towards more competitive and transparent methods (Leão et al., n.d.). Traditional procurement modalities, such as open tendering, restricted tendering, and direct contracting, each have their advantages and disadvantages in terms of cost, time, and quality (Arrowsmith, 2003). Recent studies have begun to explore these modalities' impact on the efficiency and inclusiveness of public procurement (Borsatto et al., 2020; Leão et al., n.d.; Bernal et al., 2019).

Operational efficiency, a critical aspect of public procurement, has garnered significant attention in literature from the realms of management science and socio-economic planning. Efficiency is typically measured in terms of cost savings, time savings, and quality improvements (Fazekas & Blum, n.d.; Dimitri, 2013). Several studies have found that competitive procurement modalities tend to result in lower contract prices and shorter delivery times (Lamothe, n.d.), (Carril et al., 2022; Onur et al., 2012). However, others note that the relationship between procurement modality and efficiency is complex and influenced by factors like contract size, market competition, and the procuring entity's integrity (Dale-Clough, 2015; Loader, 2015).

Inclusive procurement, conceived as a socially responsible sourcing strategy, plays a pivotal role in socioeconomic planning and development (Silva et al., 2021) and has emerged as a key strategy for achieving social and economic objectives (McCrudden, 2004; Dimand, 2022). Inclusive procurement, particularly relevant for SMEs, plays a crucial role in stimulating economic growth and reducing poverty (Ayyagari et al., 2007). Key theories emphasize public procurement as a tool for social inclusion and economic development (ibid.). However, weaknesses in initial information and unreliable supplier registries have posed challenges (Ho et al., 2010). Broader issues like discrimination in government regulations and unequal access to credit require further exploration (Williams-Elegbe, 2015).

Quantitative models assessing government initiatives aimed at fostering inclusive procurement are increasingly gaining attention but require further interdisciplinary research (Brooks et al., 2019), but their impact on the efficiency and inclusiveness of public procurement remains underexplored. While the Dominican Republic has made strides in public procurement efficiency and inclusiveness, the complex nature of procurement processes necessitates a more comprehensive understanding (Silva et al., 2021; Fazekas & Blum, 2021).

The application of machine learning and other quantitative methodologies in public procurement is an emerging and promising avenue in the field of socio-economic planning sciences. Techniques like Random Forest and Gradient Boosting have been used to predict contract prices and detect anomalies (Aldana et al., 2022; Garcia et al., 2021; Nai et al., 2022). Furthermore, the potential of machine learning to enhance procurement efficiency and transparency has been demonstrated, but its application in analyzing procurement modalities' impact is still nascent (Zheng et al., 2017).

Significant gaps in the field of public procurement include understanding the socio-cultural factors influencing procurement inclusiveness, exploring procurement modalities' impact, and analyzing government initiatives' effectiveness (McCue & Pitzer, 2000). This research contributes by quantitatively exploring the Dominican Republic's procurement modalities, analyzing their effects on economic efficiency and inclusiveness, and extending the existing literature through the application of machine learning.

The unique value of this study lies in focusing on the Dominican Republic's specific socio-political context, revealing disparities in procurement efficiency and inclusiveness, and offering insights for more inclusive procurement strategies to foster socio-economic growth and reduce corruption.

Experimental Methods:

The principal objective of this study is to investigate the influence of different procurement modalities on contract amounts and processing times in the Dominican Republic's public sector. This section outlines the dataset, machine learning algorithms, error metrics, and validation methods utilized.

Data Sourcing and Collection:

The primary data source for this research is a dataset from the Dominican Republic's procurement system, spanning 2018-2023. This dataset encompasses diverse contract details such as procurement modality, estimated and actual contract amounts, processing times, and provider information. Data were obtained directly from the DGCP's procurement system's database. During preprocessing, all this information was converted to suitable formats for the learning methods being evaluated. The output variables were the actual contract amounts and processing times. Secondary data sources were not utilized, yet future research could integrate supplementary data such as socio-economic indicators from national statistical databases.

Exploratory Data Analysis:

Initial explorations included examining summary statistics and visualizing distributions to identify patterns, outliers, and relationships. This guided the selection of machine learning methods and the preprocessing steps required.

Preprocessing Data:

The preprocessing stages for the data were as follows:

i. **Data Cleaning:** Removing duplicates and correcting inconsistencies.
ii. **Handling Missing Data:** Missing values were treated based on the nature of the data, e.g., missing categorical features were labeled 'unknown.'
iii. **Data Transformation:** Variables were transformed to better suit the analytical methods. Specifically, the target variable 'Amount contract real' was log-transformed to normalize the data and reduce the impact of outliers:
iv. **Feature Encoding:** Categorical variables were transformed using one-hot encoding, converting categories into binary columns.

Selection of Machine Learning Models:

Various machine learning algorithms were considered for this research, including Linear Regression, Support Vector Machines (SVM), and Neural Networks. Each of these models comes with its strengths and weaknesses, but the selection ultimately rested on several key criteria: the characteristics of the data, interpretability of the model, model complexity, and computational resources required. The choice to disregard other methods like SVM or Neural Networks was primarily driven by the requirement for interpretability. In public procurement settings, being able to interpret and explain model decisions is crucial for gaining stakeholder trust and for policy implementation. Additionally, the non-linear nature of the relationships in the data and considerations of computational efficiency also influenced the decision to focus on Random Forest and Gradient Boosting.

Random Forest: Random Forest was chosen for its ability to handle complex, high-dimensional data and its robustness to outliers. The model also can capture non-linear relationships, which is often important in socio-economic settings. In a Random Forest model, each decision tree is trained independently on a bootstrap sample of the data. The final prediction is an average (for regression) or majority vote (for classification) of the individual tree predictions. The Random Forest model combines predictions from multiple decision trees. For a regression task such as predicting contract amounts or processing times, the formula for the predicted output y would be the average of N tree outputs:

$$y = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$$

where $fi(x)$ is the prediction of the i^{th} decision tree for input x. Essentially, the final prediction is the average of the predictions made by all the trees in the forest.

Gradient Boosting: Gradient Boosting was selected for its strength in reducing bias and variance, handling missing data, and the ability to iteratively improve predictions. The model builds decision trees sequentially, each one correcting the errors of its predecessor. The final prediction is the sum of the predictions from all individual trees:

$$y = \sum_{i=1}^{n} \alpha_i f_i(x)$$

where α_i is the weight for the i^{th} decision tree and $f_i(x)$ is its prediction for input x. The weights ai, are learned during the training process to minimize the loss function.

Selection of Performance Metrics: The metrics chosen for model evaluation aim to reflect both the accuracy and interpretability of the model's predictions. Specifically, we employ Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Square Error (RRSE) as our primary evaluation metrics.

i. **Mean Absolute Error (MAE) and Root Mean Square Error (RMSE):** MAE and RMSE are common metrics for assessing the prediction accuracy of regression models. MAE is particularly useful as it provides a straightforward interpretation of the average absolute prediction error, making it easy to communicate to stakeholders. RMSE, on the other hand, penalizes larger errors more severely, making it a robust metric when large prediction errors are particularly undesirable. Both metrics are consistent with the model evaluation metrics discussed by Alavi & Naser (2021) emphasizing the need for simple yet informative metrics in studies involving complex procurement systems.

ii. **Relative Absolute Error (RAE) and Root Relative Square Error (RRSE):** RAE and RRSE are chosen for their ability to normalize the absolute and squared errors, respectively, by the errors of a naive prediction method, usually the mean of the observed values. These metrics offer a comparative assessment of the model's performance against a naive baseline, thereby providing an additional layer of interpretability to the model's predictive capabilities. The use of relative metrics is in line with the suggestions by Botchkarev (2019), which emphasizes the importance of context-sensitive metrics in complex predictive models, particularly in public sectors characterized by heterogeneous data and varied stakeholder interests.

While the papers by Alavi & Naser (2021) and Botchkarev (2019) present a broad range of metrics, including but not limited to precision, recall, and F1-score, these metrics are more aligned with classification problems and thus were not considered suitable for this study focusing on regression models.

Analytical Methods and Validation:

Both Random Forest and Gradient Boosting models were trained to predict real contract amounts and processing times, using an 80-20 train-test split. The performance of the models was assessed using various metrics:

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

$$\text{Relative Absolute Error (RAE)} = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{\sum_{i=1}^{n}\left|y_i - mean(y)\right|}$$

$$\text{Root Relative Square Error (RRSE)} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - mean(y)\right)^2}}$$

$$\text{Pearson correlation coefficient } (r) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

In these formulas:

y_i is the actual value for each procurement contract, either as contract amount or processing time.

$\hat{y}$ is the model's predicted value for each contract, also either as contract amount or processing time.

n is the total number of contracts in the test set, used for model evaluation.

The Pearson correlation coefficient denoted by *r* measures the linear relationship between two variables, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation). This coefficient will be used to evaluate the agreement between the two models.

To ensure a robust assessment, 80% of the contracts (train set) were used for training, and the remaining 20% (test set) served as validation. This validation framework was maintained throughout the experiments to ensure consistency and reliability in the findings.

This methodology section outlines the rigorous process followed in selecting the most appropriate machine learning models for this study, offering a transparent and systematic approach to understanding public procurement in the Dominican Republic.

Results and Findings:

The quantitative analysis of the Dominican Republic's public procurement data forms the crux of this investigation, revealing pivotal insights into the socio-economic influence of various procurement strategies. This section presents the research's essential discoveries, focusing on the impact of different procurement modalities on contract amounts and processing times—two primary indicators of economic efficiency.

Exploratory Data Analysis: The graphical representation of procurement modalities showed a preponderance of 'Purchases below the Threshold', 'Minor Purchases', and 'Exception Processes'.

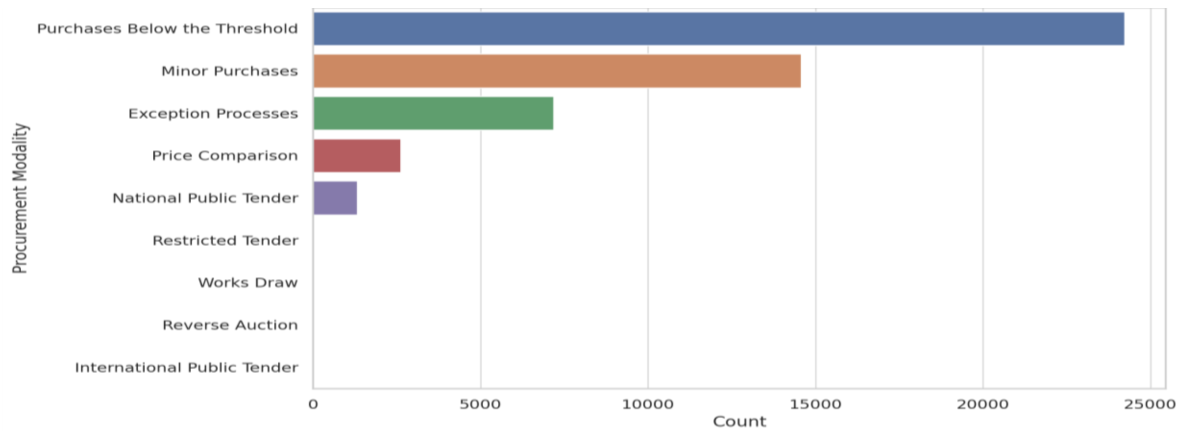

Figure 1: Distribution of Procurement Modalities

Next, the variance between the estimated and actual contract amounts was analyzed, uncovering potential insights into initial estimation accuracy. The positively skewed histogram of contract variance reflected occasional significant deviations between actual and estimated amounts. The histogram plots display the distribution of both estimated and actual contract amounts. Both distributions appear to be positively skewed, indicating that there are a few contracts with exceptionally high amounts that are causing a long tail on the right side of the distribution.

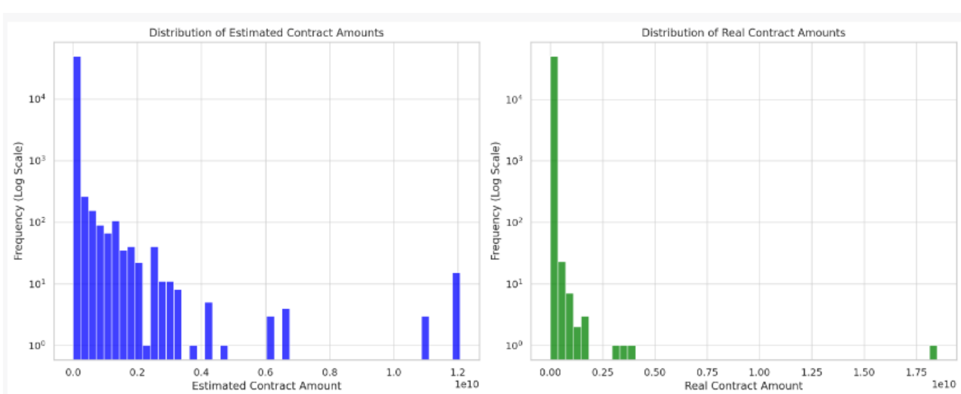

Figure 2: Distribution of Estimated vs. Real Contract Amounts

An examination of the 'Time days processing' variable depicted a processing time distribution, with many procurements processed swiftly but some taking considerably longer. The distribution of contract statuses, primarily 'Active' and 'Closed,' added another layer of understanding. The histogram of processing times shows that much procurement is processed within a relatively short time frame, but there is a long tail indicating that some procurements take much longer. This is reflected in the positive skew of the distribution.

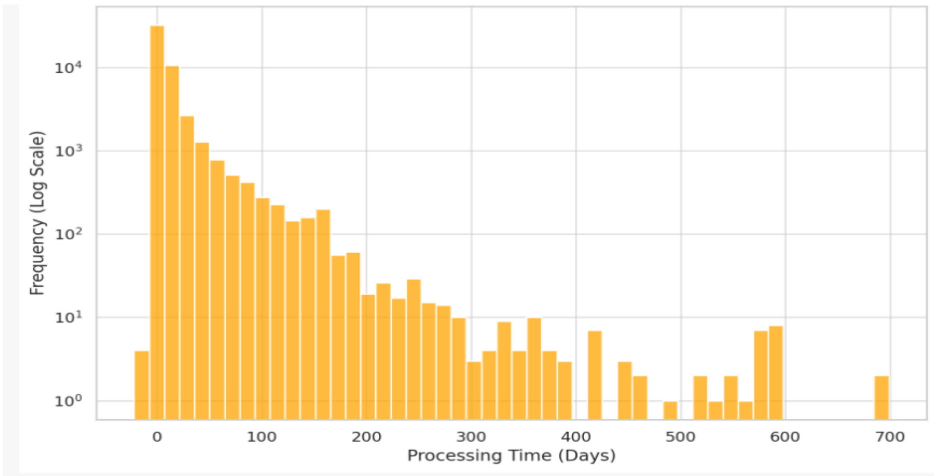

Figure 3: Distribution of Processing Times

The bar plot shows the distribution of contract statuses in the sample dataset. Most contracts are 'Active', followed by 'Closed'. A smaller number of contracts are 'Suspended' or 'Cancelled'.

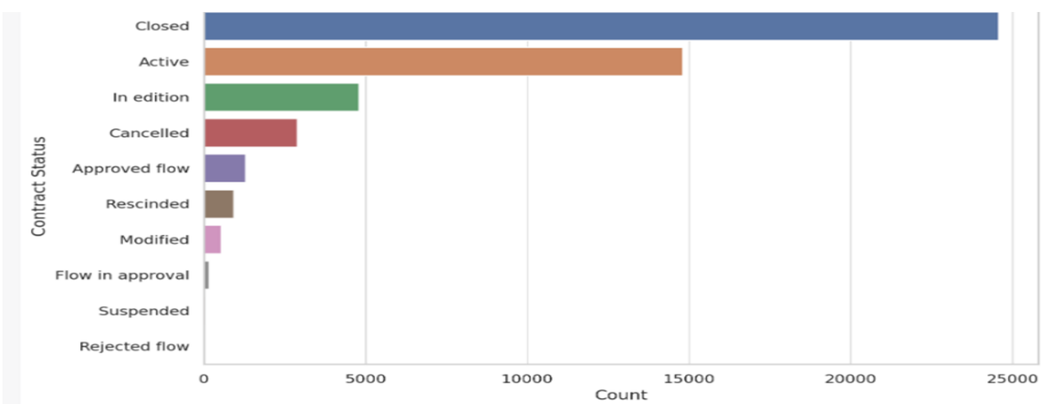

Figure 3: Distribution of Contract Status

Figure 5 shows a correlation heatmap of key variables. The correlation between procurement modality and both estimated and real contract amounts is minimal, indicating that the type of procurement modality might not have a strong linear relationship with the contract amounts. In contrast, there is a significant positive correlation between the estimated and real contract amounts, suggesting that the estimations are generally aligned with the actual contract values. The processing time does not show a strong correlation with other variables, indicating that it might be influenced by factors not captured in this heatmap. Additionally, variables such as contract status and provider classification do not exhibit strong correlations with other key variables.

The heatmap reveals subtle insights into the relationships between procurement modalities and key indicators of economic efficiency, such as contract amounts and processing times. However, it also suggests that these relationships may be more complex and not fully captured by simple linear correlations. The findings point to the multifaceted nature of public procurement, emphasizing the need for more comprehensive analytical techniques to understand the underlying patterns and dynamics.

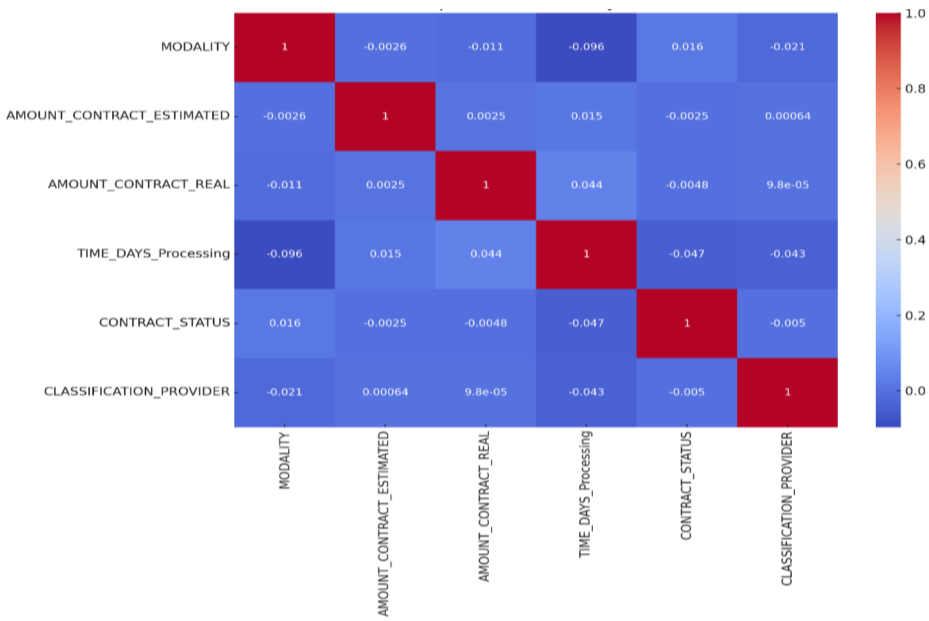

Figure 4: Heatmap Correlation of Key Variables

Figure 6 shows the distribution of contract amounts across different procurement modalities. It highlights variations, with some modalities demonstrating higher median contract values. Substantial disparities were observed across different procurement modalities, indicating the critical role of procurement modality selection in determining contract amounts.

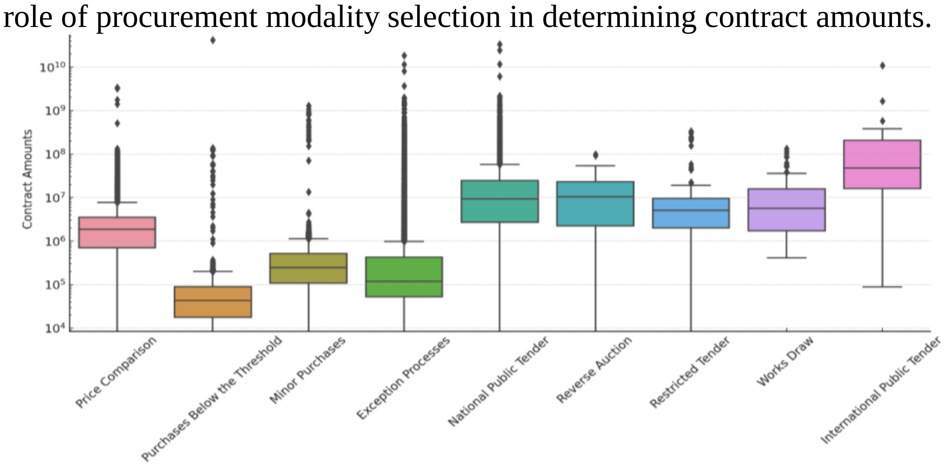

Figure 5: Distribution of Contract Amounts Across Procurement Modalities

The main findings for each modality are explained below:

i. **Exception Processes:** This modality exhibits a broad range of contract amounts with some noticeable outliers. The distribution suggests that exception processes may involve both small and large contracts, reflecting the diverse nature of this procurement method.
ii. **International Public Tender:** Similar to exception processes, international public tenders display a wide distribution, indicating that they may encompass various contract sizes, potentially reflecting international engagement in procurement.
iii. **Minor Purchases:** This modality shows a more concentrated distribution towards lower contract amounts, in line with the nature of minor purchases.
iv. **National Public Tender:** Exhibiting a wide range, national public tenders may include both routine and significant contracts, reflecting the national scope of this procurement method.
v. **Price Comparison:** The distribution here is relatively narrow with a few outliers, suggesting that price comparison often involves contracts within a specific range.
vi. **Purchases below the Threshold:** This modality appears to focus on lower contract amounts, aligning with its nature of handling purchases below certain thresholds.
vii. **Restricted Tender:** A wide distribution with higher median values suggests that restricted tenders may often involve more substantial contracts.
viii. **Reverse Auction:** This modality shows a concentrated distribution, indicating a specific range of contract amounts typically handled through reverse auctions.
ix. **Works Draw:** This modality exhibits a broad range, including some higher-value contracts, reflecting its application in various works-related procurements.

Machine Learning Modeling for Predicting Contract Amount and Processing Time:

In this study, two ensemble learning methods, Random Forest, and Gradient Boosting, were employed to predict contract amount and processing time within the public procurement domain in the Dominican Republic. Both Random Forest and Gradient Boosting were employed to predict contract amounts. For processing contract amount prediction, the Gradient Boosting model demonstrated slightly better performance with an RMSE of 126,235,128.56 compared to Random Forest's RMSE of 140,175,774.52. For processing time prediction, Random Forest achieved an RMSE of 17.83 days, outperforming Gradient Boosting, which had an RMSE of 19.93 days. The results reveal substantial disparities in procurement modalities, suggesting the choice of modality significantly impacts both efficiency and inclusiveness.

Comparative Visualization of Predictions:

While both models demonstrate reasonable performance in predicting processing time and contract amount, the Gradient Boosting model seems to exhibit slightly better alignment with actual values, particularly for contract amount prediction.

Time Prediction: In Figure 6 both models exhibit a noticeable alignment between actual and predicted processing times. The Gradient Boosting model seems to capture the pattern slightly better, with a denser clustering along the alignment line.

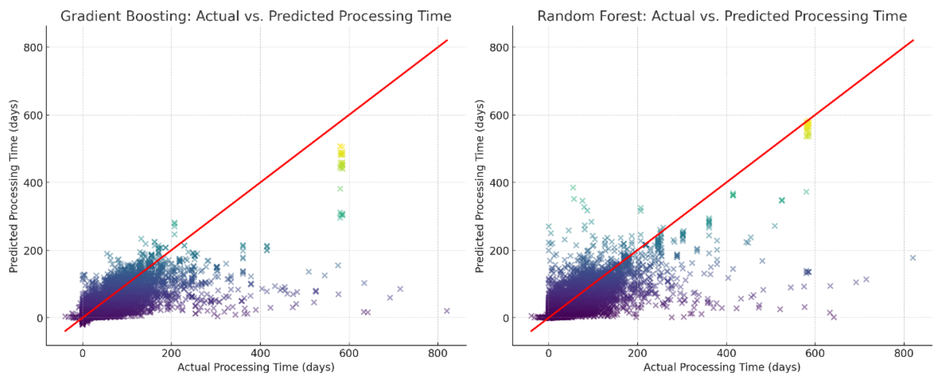

Figure 7: Comparative visualization of prediction models: Processing Time

Contract Amount Prediction: In Figure 8 the scatter plots for contract amount prediction show more dispersion, indicating a more complex relationship. The Gradient Boosting model appears to have a slightly tighter alignment with the red line, indicating better predictive accuracy.

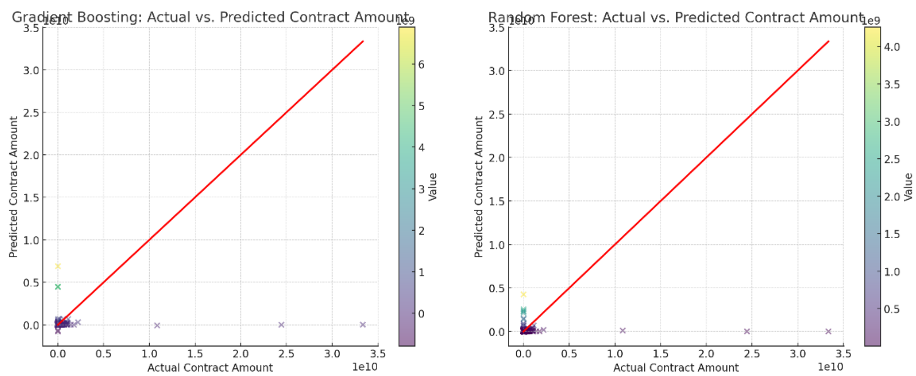

Figure 6: Comparative visualization of prediction models: Contract Amount

Correlation and Divergence between Predictions:

To evaluate the agreement between the two models, we calculated the Pearson correlation coefficient, denoted by *r*, formula (8) above. For contract amounts, a high correlation of 0.89 was observed, indicating strong agreement. For processing times, the correlation was 0.75, suggesting a moderate level of agreement.

Despite high correlation metrics, we identified specific instances where predictions significantly diverged. For example, in the case of procurement modality ‘X’, Random Forest predicted higher contract amounts compared to Gradient Boosting. The divergence may be attributed to the feature importance assigned by each model.

Feature Importance Analysis for Contract Amount Prediction:

Figures 9 and 10 indicate the feature importance of the Random Forest and Gradient Boosting models in predicting contract amount and processing time. The most influential variables were ‘date awarded’, ‘type exception’, and ‘days processing’.

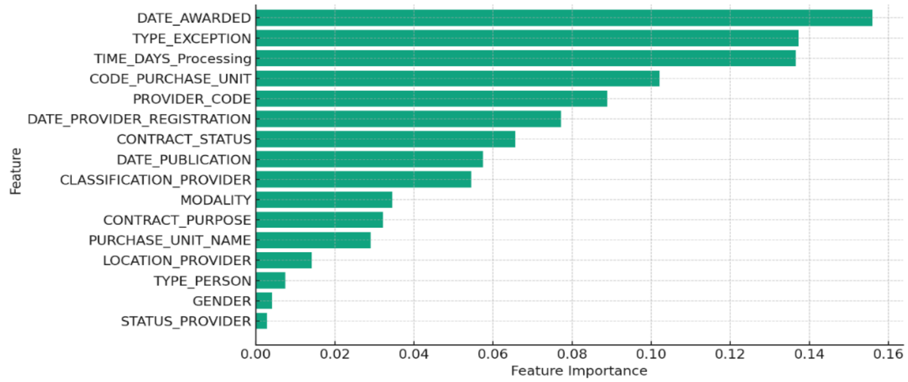

Figure 9: Feature importance for Random Forest; Contract Amount Prediction

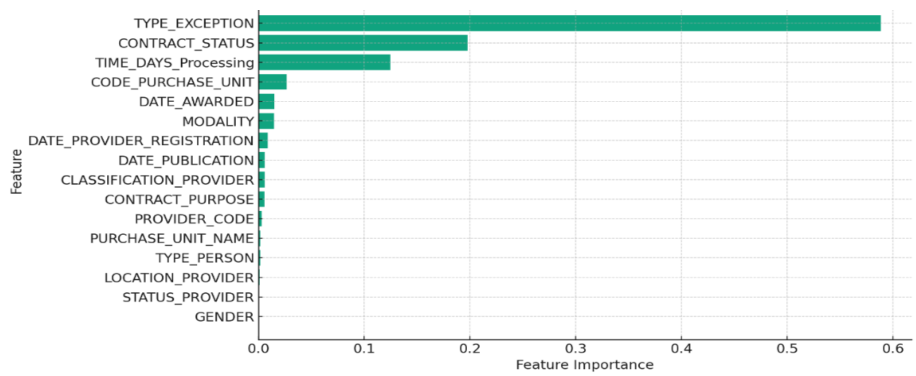

Figure 7: Feature Importance for Gradient Boosting; Contract Amount Prediction

Feature Importance Analysis for Time Processing Prediction:

Figures 11 and 12 highlight the feature importance for the Random Forest and Gradient Boosting models for Time Processing prediction, with ‘type exception’, ‘contract status’ and ‘days processing’ being the most significant features.

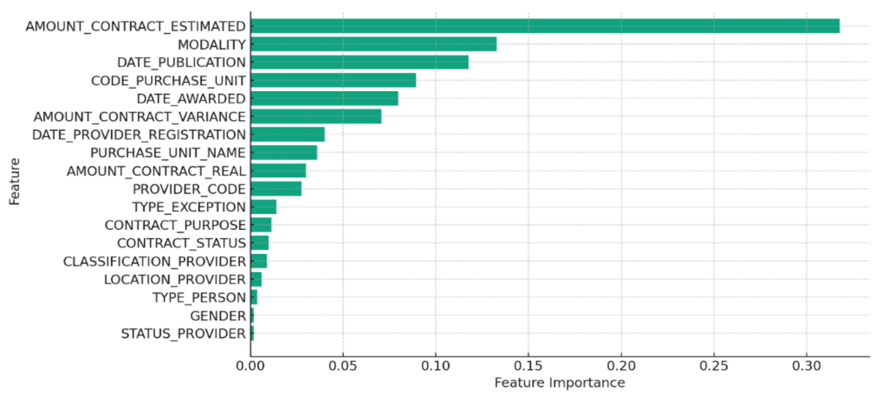

Figure 8: Feature Importance for Random Forest; Processing Time Prediction

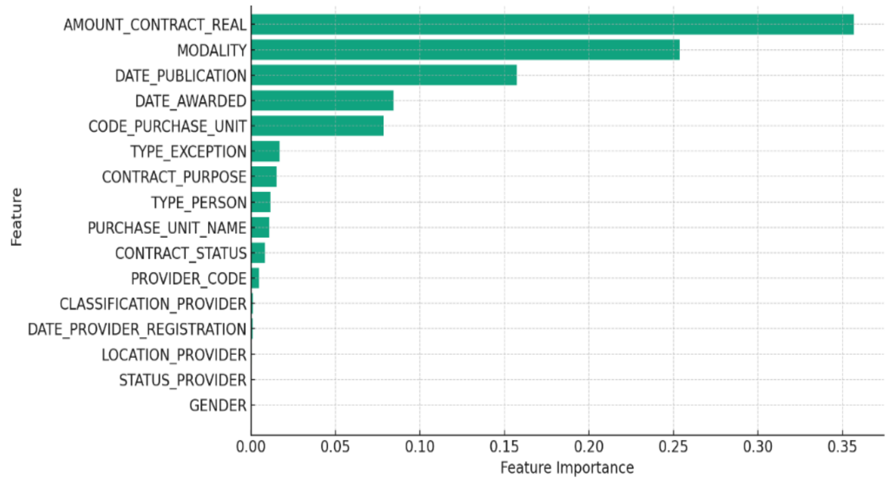

Figure 12: Feature Importance for Gradient Boosting; Processing Time Prediction

Discussion and Recommendations:

This interdisciplinary research aimed to quantitatively analyze the socio-economic planning implications of inclusive procurement strategies within the public sector of the Dominican Republic. We sought to understand how different procurement modalities affect contract amounts and processing times, which are key indicators of economic efficiency. Employing machine learning as part of a quantitative analytical framework, we derived critical insights that contribute to socio-economic planning literature.

As this study shows, procurement modality is a significant factor in determining contract amounts. This aligns with existing literature that identifies procurement methods as a crucial element in shaping the outcomes of procurement processes. However, our research extends these findings by quantifying this impact within the specific context of the Dominican Republic, using a data-driven approach.

The research has both practical and theoretical implications for the domain of inclusive public procurement. On a practical level, our research suggests that policymakers and procurement officials could potentially influence contract amounts and processing times by selecting appropriate procurement modalities. This could contribute to economic efficiency and help maximize the socio-economic benefits of public procurement. Theoretically, our research enriches the existing body of socio-economic planning literature by providing empirical evidence on how procurement strategies affect operational efficiency and social inclusion.

Our research, however, is not without its limitations. First, the data used in this study may not capture all factors that influence contract amounts and processing times. Future research could aim to incorporate more comprehensive data, including socio-economic indicators and measures of corruption. Second, our analytical method, while robust, only explains about 44% of the variance in the real contract amounts. This suggests that there are other significant factors not captured in our model. Third, the socio-political context of the Dominican Republic presents unique challenges that may affect the generalizability of our findings. Subsequent studies could extend this research to diverse socio-political settings, further enhancing its applicability to global socio-economic planning challenges.

There are several opportunities for future research. One potential avenue is to explore other machine learning models or techniques that could improve the predictive power of the model. Another possibility is to examine the impact of inclusive procurement strategies on other outcomes of interest, such as contract quality or provider diversity. Finally, future research could aim to develop a more comprehensive theoretical framework that integrates inclusive procurement strategies, socio-economic outcomes, and corruption.

In summary, this research provides valuable insights into the socio-economic impact of inclusive procurement strategies in the Dominican Republic. In summary, our findings underscore the potential of employing quantitative methodologies, such as data science techniques, in informing data-driven socio-economic planning and policy optimization. By leveraging these insights, we can contribute to socio-economic growth and the fight against corruption in the Dominican Republic and beyond.

Conclusions:

This interdisciplinary research employed quantitative methodologies, specifically data science and machine learning techniques, to rigorously scrutinize public procurement strategies in the context of socio-economic planning within the Dominican Republic’s public sector. Analyzing a comprehensive dataset of public procurement contracts from 2018 to 2023, the study revealed significant disparities in contract amounts and processing times, serving as key performance indicators that inform operational efficiency and social inclusion across diverse procurement strategies.

Crucially, the application of Random Forest and Gradient Boosting algorithms was pivotal in uncovering these insights. These machine learning models allowed for the efficient handling of high-dimensional data and afforded the capability to capture non-linear relationships, which traditional statistical methods might overlook. The algorithms’ feature importance scores provided valuable information on the variables that most significantly impact procurement efficiency and inclusiveness. The utilization of these advanced techniques not only underscores the methodological rigor of the study but also strengthens the case for the adoption of data science methodologies in public policy research.

The difference in contract amounts and processing times can be attributed to various socio-economic and political factors that are unique to the Dominican Republic. For example, the long-standing issues of corruption and favoritism in the country’s public procurement system have often favored less competitive and less efficient modalities like direct contracting.

The findings add nuanced empirical insights to the interdisciplinary field of socio-economic planning, particularly focusing on public procurement in the Dominican Republic’s unique socio-political context. Practically, these insights offer a data-driven framework for policymakers engaged in socio-economic planning to optimize public procurement systems. For instance, integrating predictive models into decision-making tools can enable real-time optimization of procurement modalities based on predicted efficiency and inclusivity outcomes.

Despite its strengths, the study has limitations, including the inability to capture qualitative metrics like stakeholder satisfaction or the quality of deliverables. Also, the unique cultural and political landscape of the Dominican Republic could introduce variables that the study’s machine-learning models did not account for.

Subsequent studies could enrich this research by integrating qualitative metrics and delving deeper into the roles of individual stakeholders within the socio-economic planning ecosystem, thus providing a more comprehensive view. The application of causal inference models could provide a more comprehensive understanding of the direct and indirect impacts of different procurement modalities on various socio-economic aspects.

In summary, this study serves as a foundational contribution to the field of socio-economic planning sciences, particularly in the domain of public procurement in emerging economies like the Dominican Republic. The advanced data science techniques employed here offer a blueprint for data-driven policymaking, with the potential to significantly contribute to sustainable socio-economic development and corruption reduction in the Dominican Republic.

Data Availability: The processed data supporting the findings of this study are available from the corresponding author upon request. Raw data were obtained from the General Directorate of Public Procurement of the Dominican Republic and are publicly accessible via their open data platform: https://datosabiertos.dgcp.gob.do/opendata/tablas

Conflicts of Interest: The authors declare no conflicts of interest regarding the publication of this paper.

References

Naser, M. Z., & Alavi, A. H. (2021). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*, 1-19.

Aldana, A., Falcón-Cortés, A. & Larralde, H. (2022). A machine learning model to identify corruption in Mexico's public procurement contracts. Retrieved March 3, 2023, from https://arxiv.org/abs/2211.01478

Arrowsmith, S. (2003). Government Procurement in the WTO. Retrieved March 3, 2023, from https://lccn.loc.gov/2002030128

Ayyagari, M., Beck, T. & Demirguc-Kunt, A. (2007). Small and Medium Enterprises across the Globe. *Small Business Economics*, 29, 415-434.

Bernal, R., San-Jose, L. & Retolaza, J. L. (2019). Improvement actions for a more social and sustainable public procurement: A Delphi analysis. *Sustainability, 11*(15), 4069.

Borsatto, R. S., Altieri, M. A., Duval, H. C., & Perez-Cassarino, J. (2020). Public procurement as strategy to foster organic transition: insights from the Brazilian experience. *Renewable Agriculture and Food Systems, 35*(6), 688-696.

Botchkarev, A. (2019). A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management, 14*, 45-76.

Brooks, J., Commandeur, D., & Vera, E. (2019). Inclusive procurement and transparency--connecting smallholder farmers to school feeding. *Gates Open Res, 3*(495).

Carril, R., Lira, A. G. & Walker, M. S. (2022). Competition under incomplete contracts and the design of procurement policies. *Universitat Pompeu Fabra, Department of Economics and Business.*

Colman, R. (2020). *Open Contracting Partnership.* Retrieved March 3, 2023, from https://www.open-contracting.org/2020/09/23/women-win-one-in-four-contracts-in-the-dominican-republic-thanks-to-inclusive-procurement-reforms

Dale-Clough, L. (2015). Public procurement of innovation and local authority procurement: procurement modes and framework conditions in three European cities. *Innovation: The European Journal of Social Science Research, 28*(3), 220-242.

Dimand, A. M. (2022). Determinants of local government innovation: the case of green public procurement in the United States. *International Journal of Public Sector Management, 35*(5), 584-602.

Dimitri, N. (2013). "Best value for money" in procurement. *Journal of Public Procurement, 13*(2), 149-175.

Fazekas, M. & Blum, J. R. (2021). Improving public procurement outcomes. Retrieved March 3, 2023, from https://documents.worldbank.org/curated/en/656521623167062285/pdf/Improving-Public-Procurement-Outcomes-Review-of-Tools-and-the-State-of-the-Evidence-Base.pdf

Garcia, R.M. J., Rodríguez, M. V., Aranguren, A., Santana, H. R., Sierra, A. B., & Zelaia, J. A. (2021). Award price estimator for public procurement auctions using machine learning algorithms: case study with tenders from Spain. *Studies in informatics and control.*

Ho, W., Xu, X. & Dey, P. K., 2010. Multi-criteria decision-making approaches for supplier evaluation and selection: A literature review.. *European Journal of Operational Research, 202*(1), 16-24.

Lamothe, S. (n.d.). How competitive is "competitive" procurement in the social services? *The American Review of Public Administration, 45*(5), 584-606.

Leão, R., Ijatuyi, E. J., & Goulao, L. F. (n.d.). How Public Procurement Mechanisms Can Be Used as a Tool for Developing Pro-Poor Food Value Chains: From Entry Points to Interventions. *Sustainability, 15*(12).

Loader, K. (2015). SME suppliers and the challenge of public procurement: Evidence revealed by a UK government online feedback facility. *Journal of Purchasing and Supply Management, 21*(2), 103-112.

McCrudden, C. (2004). Using public procurement to achieve social outcomes. *In Natural Resources Forum, 28*(4), 257-267.

McCue, C. P., & Pitzer, J. T. (2000). Centralized vs. decentralized purchasing: Current trends in governmental procurement practices. *ournal of Public Budgeting, Accounting & Financial Management, 12*(3), 400-420.

Mercado. (2021). Obtenido de SMEs in the Dominican Republic: Retrieved March 3, 2023 from https://www.revistamercado.do/economia/empresas-micro-pequenas-medianas-y-grandes-que-tantas-hay-en-el-mercadooo

Nai, R., Sulis, E., & Meo, R. (2022). Public procurement fraud detection and artificial intelligence techniques: a literature review. *In Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*, (p. 1-13).

OECD. (2020). *Government at a Glance: Latin America and the Caribbean 2020.* https://doi.org/10.1787/13130fbb-en

Onur, İ., Özcan, R., & Taş, B. K. (2012). Public procurement auctions and competition in Turkey. *Review of industrial organization, 40*, 207-223.

Silva, O. N., Navarro, E. R., Montoya, E. O., De Peña, N. I., & Farroñán, E. V. (2021). The Inter-American Development Bank: purpose, results and challenges in Latin America and the Caribbean. *Universidad y Sociedad, 13*(2), 495-503.

Williams-Elegbe, S. (2015). A Comparative analysis of public procurement reforms in Africa: Challenges and prospects. *Procurement Law Journal, 1*, 11-32.

Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An overview of blockchain technology: Architecture, consensus, and future trends. *IEEE international congress on big data (BigData congress).*