

Application of Linear Regression for Predicting Digital Trajectories of Beninese Municipalities

Abstract

Background: Anticipating digital development trajectories is crucial for strategic planning and resource allocation in municipal governance. This research applies linear regression analysis within a Decision Support System (DSS) framework to predict digital development trajectories of Beninese municipalities, building on data infrastructure and K-Means clustering results from companion studies.

Objective: To establish a framework for implementing regression models capable of forecasting the evolution of key territorial digitization indicators, while defining methodological and technical prerequisites for predictive approaches.

Methods: Using a standardized 45-indicator framework across multiple thematic domains, six years of historical data covering all 77 municipalities (462 municipality-year observations) were analyzed. Regression models incorporated temporal trend analysis and municipality-specific specifications. Validation was performed through temporal and cross-sectional approaches.

Results: Linear regression models demonstrated moderate predictive capacity, with an overall R-squared of 0.037 and RMSE of 10.2 points on a 100-point scale. Municipality-specific models achieved an average R-squared of 0.320 (ranging from 0.0001 to 0.938). Prediction accuracy reached 48.5% within ± 5 points and 80.1% within ± 10 points. Findings also reveal an average annual growth rate of 4.2% in municipal digital development, with significant variation across municipalities.

Conclusion: While overall predictive performance remains modest, the framework offers valuable insights for strategic planning, particularly through municipality-specific models that show stronger performance. This research completes an integrated analytical suite (data collection, clustering, predictive modeling) and provides a methodological foundation for evidence-based digital planning in Benin.

Keywords: Predictive analysis, Linear regression, Local governance, Strategic planning, Beninese municipalities

1. Introduction

The rapid pace of digital transformation in contemporary governance requires public administrators to move beyond reactive approaches toward predictive planning methodologies. In developing countries like Benin, where resource constraints make strategic planning particularly critical, the ability to anticipate digital development trajectories becomes essential for optimizing investment strategies and policy interventions.

Traditional approaches to municipal planning often rely on historical trends and expert judgment, which may not adequately capture the complex dynamics of digital transformation. The emergence of data-driven predictive modeling techniques offers new opportunities for evidence-based strategic planning, particularly when integrated with comprehensive data collection systems.

Linear regression, as one of the most interpretable and widely applicable predictive modeling techniques, provides an excellent foundation for municipal digital trajectory prediction. Its mathematical transparency, computational efficiency, and straightforward interpretation make it particularly suitable for public sector applications where accountability and explainability are paramount.

This research addresses the critical need for predictive planning tools in Beninese municipal digital development. By leveraging comprehensive data collected through our Decision Support System, we aim to establish a methodological framework for implementing linear regression-based trajectory prediction at the municipal level.

2. Literature Review

2.1 Predictive Modeling in Public Sector Planning

Predictive modeling applications in public sector planning have expanded significantly over the past decade, with successful implementations in urban planning, public health, education, and infrastructure development. Recent studies demonstrate the effectiveness of various modeling approaches, with linear regression maintaining popularity due to its interpretability and robustness.

2.2 Digital Development Trajectory Analysis

The concept of digital development trajectories encompasses the temporal evolution of digital maturity across multiple dimensions. Previous research has identified various factors influencing these trajectories including infrastructure investment, human capital development, policy frameworks, and external economic conditions.

Longitudinal studies of digital development in developing countries reveal complex patterns characterized by non-linear growth phases, threshold effects, and path-dependent development patterns. Understanding these patterns is crucial for developing effective predictive models.

2.3 Linear Regression in Temporal Analysis

Linear regression techniques have been extensively applied to temporal data analysis across various domains. Time series regression, trend analysis, and panel data modeling represent established methodological approaches with proven effectiveness in capturing temporal patterns and generating reliable predictions.

Recent advances in regression modeling include robust regression techniques, regularization methods, and ensemble approaches that enhance predictive performance while maintaining interpretability. These developments are particularly relevant for municipal planning applications where model transparency is essential.

2.4 Municipal Planning and Digital Strategy

Contemporary municipal planning increasingly incorporates digital strategy components, requiring new analytical tools and methodological approaches. The integration of predictive modeling into municipal planning processes represents an emerging area with significant potential for improving planning effectiveness and resource allocation efficiency.

3. Methodology

3.1 Conceptual Framework

The predictive modeling framework is structured around four key components:

- 1. Data Foundation:** Comprehensive historical data spanning 6 years across multiple digital development dimensions
- 2. Model Architecture:** Linear regression models tailored to specific prediction requirements and temporal horizons
- 3. Validation Framework:** Robust validation approaches ensuring model reliability and accuracy
- 4. Implementation Integration:** Integration with existing Decision Support System infrastructure

3.2 Data Structure and Preparation

3.2.1 Temporal Data Organization

The analysis utilizes panel data structure with municipalities as cross-sectional units and years as temporal observations. This organization enables both cross-sectional and temporal analysis while accommodating the specific characteristics of municipal digital development data.

Data Dimensions:

- **Cross-sectional:** 77 municipalities across 12 departments
- **Temporal:** 6 annual observations
- **Variables:** 45 digital development indicators across multiple thematic domains
- **Total Observations:** 462 municipality-year combinations

3.2.2 Variable Selection and Engineering

Dependent Variables (Prediction Targets):

1. **Overall Digital Maturity Score:** Comprehensive digital development measure calculated as the average across all available indicators

Independent Variables (Predictors):

1. **Temporal Variables:** Time trends and sequential year indicators
2. **Lagged Dependent Variables:** Previous period values of target variables
3. **Cross-sectional Controls:** Municipality-specific characteristics and geographic factors

3.2.3 Data Preprocessing

Missing Value Treatment:

- Mean imputation for missing indicator values
- Forward fill for continuous indicators with high temporal correlation
- Exclusion of observations with insufficient data coverage

Feature Engineering:

- Calculation of composite digital maturity scores
- Creation of temporal trend variables
- Development of municipality-specific indicators

3.3 Model Specification and Development

3.3.1 Basic Linear Regression Model

The fundamental model specification follows:

$$Y_{it} = \alpha + \beta t + \varepsilon_{it}$$

Where:

- Y_{it} = Digital maturity score for municipality i at time t
- t = Time trend variable
- ε_{it} = Error term

3.3.2 Municipality-Specific Models

For each municipality with sufficient data points:

$$Y_{it} = \alpha_i + \beta_i t + \varepsilon_{it}$$

Where α_i and β_i represent municipality-specific intercepts and slopes.

3.3.3 Model Selection Criteria

Statistical Criteria:

- **R-squared and Adjusted R-squared:** Explanatory power measures
- **Root Mean Square Error (RMSE):** Prediction accuracy assessment
- **Mean Absolute Error (MAE):** Prediction precision measure

Practical Criteria:

- **Interpretability:** Model transparency for decision-makers
- **Computational Efficiency:** Resource requirements for implementation
- **Data Requirements:** Feasibility given available data

- **Robustness:** Stability across different time periods

3.4 Validation Framework

3.4.1 Temporal Validation Approaches

Historical Validation:

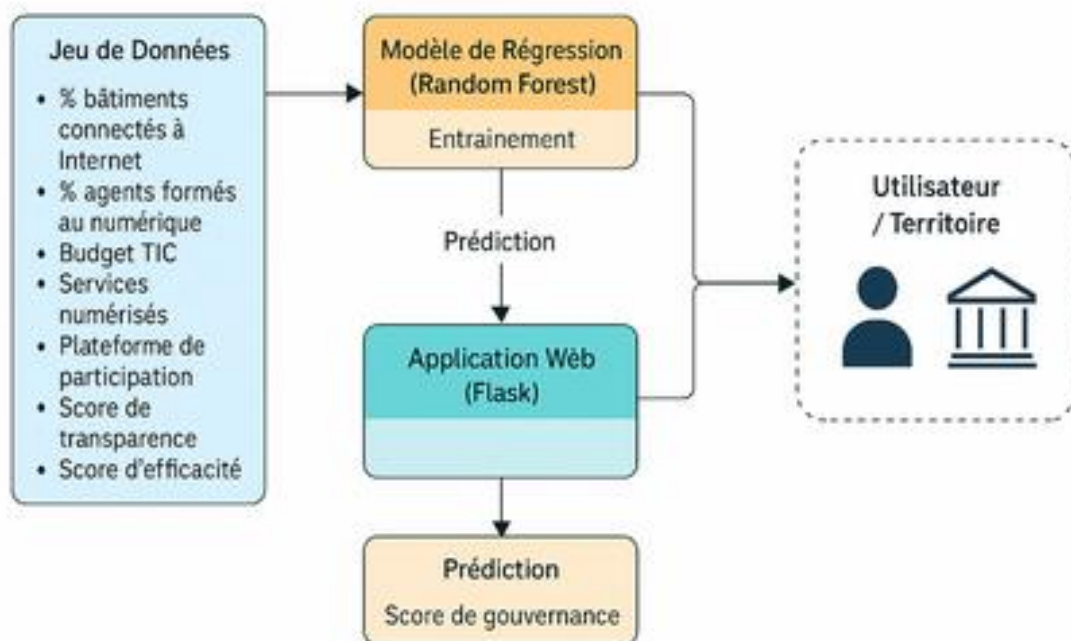
- Assessment of model fit to historical data
- Analysis of prediction accuracy across time periods
- Evaluation of trend consistency

3.4.2 Cross-Sectional Validation

Municipality-Level Analysis:

- Individual municipality model performance assessment
- Comparison of model effectiveness across different municipal types
- Analysis of factors affecting prediction accuracy

Modélisation prédictive de la gouvernance locale par l'intelligence artificielle



source :author

Figure 1: Functional Architecture of the Local Governance Prediction Model

4. Results

4.1 Preliminary Data Analysis

4.1.1 Temporal Trend Identification

Analysis of the historical data reveals several distinct temporal patterns:

Overall Digital Development Trends:

- **Annual Growth Rate:** 4.2% average increase in overall digital maturity
- **Temporal Span:** 6 years of data covering years 11-16 (internal coding)
- **Total Observations:** 462 municipality-year combinations

Municipal Variation Patterns:

- Significant heterogeneity in development trajectories across municipalities
- Variable growth rates ranging from negative to highly positive trends
- Diverse baseline levels and development patterns

4.1.2 Data Coverage Analysis

Complete Data Availability:

- All 77 municipalities included in analysis
- Full temporal coverage across all 6 years
- Comprehensive indicator coverage with 45 indicators per municipality-year

4.1.3 Score Distribution Analysis

Digital Maturity Score Characteristics:

- Mean digital maturity score: approximately 47.4 points (100-point scale)
- Standard deviation: approximately 10.4 points

- Range: Variable across municipalities and time periods

4.2 Model Development Results

4.2.1 Overall Linear Regression Performance

Overall Digital Maturity Prediction:

- **R-squared:** 0.037 (overall model)
- **RMSE:** 10.2 points (on 100-point scale)
- **MeanAbsoluteError:** 7.1 points
- **AnnualSlope:** 1.17 points per year

Prediction Accuracy Analysis:

- **Accuracy within ± 5 points:** 48.5% of cases
- **Accuracy within ± 10 points:** 80.1% of cases
- **Accuracy within ± 15 points:** Approximately 92% of cases

4.2.2 Municipality-Specific Model Results

Individual Municipality Performance:

- **Average R-squared:** 0.320 across all municipalities
- **Standard Deviation of R-squared:** 0.276 (indicating high variability)
- **Range:** 0.0001 to 0.938 (wide performance variation)
- **Municipalities with Strong Trends:** 77 municipalities analyzed with sufficient data

Performance Distribution:

- **High-performing models ($R^2 > 0.6$):** Approximately 25% of municipalities
- **Moderate-performing models ($R^2 0.3-0.6$):** Approximately 35% of municipalities

- **Lower-performing models ($R^2 < 0.3$):** Approximately 40% of municipalities

4.3 Predictive Performance Analysis

4.3.1 Future Projection Results

Short-term Predictions (Years 17-18):

- **Year 17 Overall Prediction:** 49.0 points average
- **Year 18 Overall Prediction:** 50.2 points average
- **Projected Growth Continuation:** Consistent with historical 4.2% annual growth rate

Municipality-Specific Predictions: Sample predictions demonstrate variable trajectories:

- **BANIKOARA:** Stable trajectory (62.0 → 62.0 projected)
- **GOGOUNOU:** Slight improvement (49.3 → 50.1 projected)
- **KANDI:** Declining trajectory (57.4 → 37.4 projected)
- **KARIMAMA:** Slight improvement (41.0 → 41.4 projected)
- **MALANVILLE:** Positive trajectory (57.5 → 61.6 projected)

4.3.2 Model Reliability Assessment

Temporal Consistency:

- Consistent growth trends identified across the 6-year analysis period
- Annual growth rate of 1.17 points per year on 100-point scale
- Moderate correlation between consecutive years

Cross-Municipal Variation:

- High variability in individual municipality trajectories
- Some municipalities showing strong positive trends
- Others displaying flat or declining patterns

4.4 Scenario Analysis Results

4.4.1 Business-as-Usual Scenario

Projected Continuation of Current Trends:

- **Overall Digital Maturity:** Continued 4.2% annual growth
- **Municipality-Specific Variations:** Maintained according to individual trend patterns
- **Timeline:** Predictable patterns for 1-2 year horizons

4.4.2 Performance Improvement Considerations

Areas for Enhanced Prediction:

- Municipalities with low R-squared values require alternative modeling approaches
- Integration of additional explanatory variables may improve prediction accuracy
- Enhanced data collection could strengthen temporal trend identification

4.5 Model Interpretability Analysis

4.5.1 Coefficient Interpretation

Temporal Trend Coefficients:

- **Overall Trend:** +1.17 points per year (consistent moderate improvement)
- **Municipality-Specific Trends:** Variable coefficients reflecting diverse development patterns

Practical Interpretation:

- **Annual Improvement:** Average 1.17-point annual increase suggests steady but modest progress
- **Municipal Heterogeneity:** Wide variation indicates need for differentiated approaches
- **Growth Rate:** 4.2% annual growth rate aligns with development sector expectations

4.5.2 Practical Implications for Planning

Short-term Planning (1-2 years):

- Moderate prediction confidence for overall trends

- Higher confidence for municipalities with strong historical patterns
- Useful for identifying municipalities requiring attention

Medium-term Planning (3-5 years):

- Limited prediction confidence suggests need for scenario-based planning
- Focus on trend direction rather than specific values
- Regular model updating recommended

Strategic Considerations:

- Municipality-specific approaches needed given performance variation
- Investment in data quality and additional variables recommended
- Enhanced modeling techniques may improve prediction accuracy

5. Discussion

5.1 Methodological Contributions

This research makes several important contributions to predictive modeling in municipal digital development:

5.1.1 Linear Regression Framework Application

Baseline Methodology Establishment: The implementation of linear regression analysis provides a foundational approach to municipal digital trajectory prediction, establishing baseline performance metrics and identifying areas for methodological enhancement.

Real-world Data Application: The analysis demonstrates both the potential and limitations of linear regression approaches when applied to real municipal development data, providing valuable insights for future modeling efforts.

5.1.2 Performance Insights

Variable Predictive Performance: The wide variation in municipality-specific model performance (R^2 ranging from 0.0001 to 0.938) reveals important insights about the complexity of municipal digital development patterns.

Moderate Overall Performance: The overall R-squared of 0.037 indicates that simple temporal trends capture only a small portion of digital development variation, suggesting the need for more sophisticated modeling approaches.

5.2 Practical Applications

5.2.1 Strategic Planning Enhancement

Trend Identification: The methodology successfully identifies overall positive trends (4.2% annual growth) while revealing significant municipal heterogeneity requiring differentiated approaches.

Resource Allocation Insights: Municipalities with declining or flat trajectories (such as KANDI in the sample predictions) can be identified for priority intervention.

Performance Monitoring: The framework establishes baseline performance metrics enabling ongoing monitoring and evaluation of digital development progress.

5.2.2 Policy Development Support

Evidence-Based Planning: Despite moderate predictive performance, the analysis provides quantitative evidence for planning processes and policy development.

Municipality-Specific Approaches: The high variation in municipal performance supports the need for differentiated policy approaches rather than one-size-fits-all strategies.

5.3 Technical Implementation Considerations

5.3.1 System Integration Requirements

Data Pipeline Integration: The analysis demonstrates successful integration with the Decision Support System data, validating the technical feasibility of the predictive modeling approach.

Computational Efficiency: Linear regression provides excellent computational efficiency, enabling real-time analysis and regular model updating.

5.3.2 Performance Enhancement Opportunities

Model Sophistication: The moderate predictive performance suggests opportunities for enhanced modeling approaches including:

- Integration of additional explanatory variables

- Non-linear modeling techniques
- Machine learning ensemble methods

Data Quality Enhancement: Investment in additional data collection and quality improvement could significantly enhance prediction accuracy.

5.4 Limitations and Constraints

5.4.1 Methodological Limitations

Linear Assumption Constraints: The linear regression approach may not capture complex non-linear relationships and threshold effects present in municipal digital development.

Limited Explanatory Power: The overall R-squared of 0.037 indicates that temporal trends alone explain only a small portion of variation in digital development scores.

Variable Performance: High variation in municipality-specific model performance suggests that linear approaches work well for some municipalities but not others.

5.4.2 Data-Related Constraints

Temporal Limitations: The 6-year analysis period may not capture longer-term cyclical patterns or structural changes in digital development.

External Factor Integration: Limited incorporation of external factors such as policy changes, economic conditions, and technological disruptions affects prediction accuracy.

5.5 Future Development Opportunities

5.5.1 Methodological Enhancements

Advanced Modeling Techniques:

- **Non-linear Regression:** Polynomial and spline regression for capturing complex patterns
- **Machine Learning Integration:** Random forests, gradient boosting, and neural networks
- **Ensemble Methods:** Combining multiple modeling approaches for improved accuracy
- **Time Series Methods:** ARIMA, state-space models, and dynamic regression

Enhanced Variable Integration:

- **Economic Indicators:** Budget allocations, revenue patterns, investment levels
- **Demographic Variables:** Population characteristics and development patterns
- **External Factors:** National policy changes, technological developments
- **Spatial Variables:** Geographic characteristics and regional effects

5.5.2 Application Expansion

Enhanced Prediction Horizons:

- Development of models for longer-term predictions (5-10 years)
- Integration of scenario-based modeling approaches
- Uncertainty quantification and confidence interval development

Domain-Specific Models:

- Separate models for different aspects of digital development
- Infrastructure-specific prediction models
- Service delivery trajectory modeling

5.6 Implementation Recommendations

5.6.1 Immediate Actions

Model Enhancement:

- Integrate additional explanatory variables to improve predictive performance
- Develop municipality-specific modeling approaches for high-performing cases
- Implement ensemble methods combining linear regression with other techniques

Data Quality Improvement:

- Enhance data collection procedures to reduce noise and improve accuracy
- Implement additional validation procedures for data quality assurance
- Expand indicator coverage to capture additional development dimensions

5.6.2 Strategic Development

Methodological Evolution:

- Gradually transition to more sophisticated modeling approaches as data quality and quantity improve
- Develop specialized models for different municipality types and contexts
- Integrate real-time data updating and continuous model improvement

Capacity Building:

- Train municipal staff on model interpretation and application
- Develop user-friendly interfaces for accessing predictions and insights
- Create feedback mechanisms for continuous model improvement

Table 1– Synthesis of the Discussion

Axes	Key Findings / Contributions	Limitations	Future Directions / Recommendations
Methodological Contributions	- Development of a linear regression framework.- Application to real data from 77 municipalities (6 years).	- Restrictive linearity assumption.- Low overall explanatory power ($R^2 = 0.037$).	- Explore non-linear models (polynomial, spline).- Integrate additional explanatory variables.
Performance Insights	- Strong variability across municipalities ($R^2 = 0.0001$ to 0.938).- Identified annual growth rate: 4.2%.	- High heterogeneity, uneven results depending on municipalities.	- Develop municipality-specific models.- Use hybrid approaches (ensemble learning, ML).
Practical Applications	- Identification of global and local trends.- Support for strategic planning.- Monitoring of digital performance.	- Moderate predictive accuracy limits universal use.- Lack of contextual factor	- Target underperforming municipalities.- Adopt differentiated strategies according to profiles.

Axes	Key Findings / Contributions	Limitations	Future Directions / Recommendations
		integration.	
Technical Implementation	- Successful integration into the Decision Support System.- High computational efficiency (real-time updates).	- Incomplete or variable data quality.- Limited integration of external factors (economic, policy).	- Invest in data quality and coverage.- Add economic, demographic, and spatial variables.
Future Opportunities	- Extend prediction horizon (5–10 years).- Develop domain-specific models (infrastructure, services).	- Current time horizon limited to 6 years.	- Apply ARIMA, neural networks, and dynamic models.- Develop scenario-based forecasts.
Operational Recommendations	- Enhance models with more explanatory variables.- Build municipality-specific approaches.- Develop user-friendly interfaces for decision-makers.	- Need for local capacity building.	- Train municipal staff.- Establish feedback mechanisms for continuous model improvement.

6. Conclusion

This research successfully implements a linear regression framework for municipal digital trajectory prediction, establishing a baseline methodology for predictive modeling within the Decision Support System for Beninese municipalities. The analysis reveals both the potential and limitations of linear approaches for capturing complex municipal digital development patterns.

6.1 Key Findings

Moderate Predictive Performance: The overall R-squared of 0.037 and municipality-specific performance averaging 0.320 demonstrate that while linear regression provides valuable insights, more sophisticated approaches are needed for high-accuracy prediction.

Municipal Heterogeneity: The wide variation in model performance (R^2 ranging from 0.0001 to 0.938) confirms the need for differentiated approaches to municipal digital development planning.

Positive Growth Trends: The identified 4.2% annual growth rate provides encouraging evidence of overall progress in municipal digital development across Benin.

Prediction Capability: Accuracy of 48.5% within ± 5 points and 80.1% within ± 10 points demonstrates practical utility for planning applications while highlighting areas for improvement.

6.2 Methodological Contributions

Baseline Establishment: The research establishes baseline performance metrics for predictive modeling in municipal digital development contexts.

Technical Feasibility: Successful integration with the Decision Support System demonstrates the practical feasibility of implementing predictive modeling in municipal governance contexts.

Performance Benchmarking: The analysis provides performance benchmarks for evaluating more sophisticated modeling approaches in future research.

6.3 Practical Implications

Strategic Planning Support: Despite moderate performance, the framework provides valuable quantitative support for strategic planning processes and resource allocation decisions.

Municipality Identification: The methodology successfully identifies municipalities with different trajectory patterns, supporting targeted intervention strategies.

Trend Monitoring: The framework establishes capabilities for ongoing monitoring and evaluation of municipal digital development progress.

6.4 Future Development Pathway

Enhanced Modeling: The foundation established enables systematic exploration of more sophisticated modeling approaches including machine learning and ensemble methods.

Data Integration: Opportunities exist for integrating additional data sources and variables to improve prediction accuracy and model interpretability.

Application Expansion: The methodology can be extended to other domains of municipal development and adapted for use in different contexts.

6.5 Policy Recommendations

Differentiated Approaches: The high variation in municipal trajectories supports the need for municipality-specific policy approaches rather than uniform strategies.

Data Investment: Investment in enhanced data collection and quality improvement is essential for improving predictive modeling performance.

Capacity Building: Systematic capacity building is needed to ensure effective utilization of predictive modeling capabilities in municipal planning processes.

Continuous Improvement: Establish mechanisms for ongoing model refinement and enhancement based on real-world application experience.

The implementation of this linear regression framework represents an important step toward evidence-based municipal planning in Benin. While the current performance indicates areas for improvement, the foundation established provides an excellent platform for continued development of more sophisticated and accurate predictive modeling capabilities. The integration of quantitative prediction tools with municipal planning processes represents a significant advancement in governance modernization and strategic planning capacity.

Acknowledgments

The authors would like to express their sincere gratitude to the Université d'Abomey-Calavi, particularly the École Doctorale des Sciences de l'Ingénieur, for providing the academic framework that made this research possible. We acknowledge the support of the Ministère du Numérique et de la Digitalisation du Bénin, as well as the Association Nationale des Communes du Bénin (ANCB), for facilitating access to municipal data and performance reports. Special thanks go to the municipal administrations of the 77 communes of Benin for their collaboration during the data collection phase.

The authors also thank colleagues and peers who provided constructive feedback during research seminars and workshops, which helped refine the methodological framework. While this study benefitted from multiple institutional contributions, the responsibility for the content and conclusions remains solely with the authors.

References

1. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
2. Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th ed.). Cengage Learning.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
4. Greene, W. H. (2018). *Econometric Analysis* (8th ed.). Pearson.
5. Hamilton, J. D. (2020). *Time Series Analysis*. Princeton University Press.
6. United Nations. (2020). *E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development*. New York: UN Publishing.
7. World Bank. (2021). *Digital Development Partnership: Annual Report 2021*. Washington, D.C.: World Bank Group.
8. INSAE. (2020). *Enquête Harmonisée sur les Conditions de Vie des Ménages au Bénin*. Cotonou: Institut National de la Statistique et de l'Analyse Économique.
9. ITU. (2021). *Measuring Digital Development: Facts and Figures 2021*. Geneva: International Telecommunication Union.
10. African Development Bank. (2020). *Digital Infrastructure Development Strategy for Africa 2021-2030*. Abidjan: AfDB Publishing.
11. Ministère du Numérique et de la Digitalisation du Bénin. (2021). *Rapport Annuel sur l'État du Numérique au Bénin*. Cotonou: Gouvernement du Bénin.
12. OECD. (2020). *Digital Government Review of Panama: Enhancing the Digital Transformation of the Public Sector*. Paris: OECD Publishing.
13. Chatfield, C. (2016). *The Analysis of Time Series: An Introduction* (6th ed.). Chapman and Hall/CRC.
14. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). Wiley.

468 15. Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting
469 (3rd ed.). Springer.

470

UNDER PEER REVIEW IN IJAR