JOURNAL AFRICAIN DES SCIENCES
VOL. 3, 2 (2026) 97-102

https://jafrisci.org

ISSN (online) : 3107-0949 ; ISSN (print) : 3107-0930

OPEN ACCESS

RESEARCH ARTICLE

Article Info.:
Received : 13th November 2025
Accepted : 28th February 2026
Published : 04th March 2026

DEEP LEARNING-BASED FORECASTING OF FISCAL AND NON-FISCAL REVENUES FROM THE EXTRACTIVE SECTOR IN THE DEMOCRATIC REPUBLIC OF CONGO USING LSTM NEURAL NETWORK

James Greg Mpoyi Ntambwe [1,*], Eugene Mbuyi Mukendi [1]

https://doi.org/10.70237/jafrisci.2026.v3.i2.09

Résumé :

La République Démocratique du Congo (RDC), riche en ressources minérales, demeure fortement dépendante des recettes issues de son secteur extractif. Toutefois, la prévision des recettes fiscales et non fiscales reste complexe en raison de la volatilité des marchés, des changements réglementaires et de l'incertitude macroéconomique. Cette étude propose un modèle basé sur un réseau de neurones à mémoire à long et court terme (LSTM) afin de prédire ces recettes à partir de données mensuelles couvrant la période 2010–2023, collectées auprès de la Banque Centrale du Congo, de l'ITIE et des administrations fiscales nationales. Les résultats montrent des performances prédictives élevées, avec un RMSE de 38,82 (nRMSE ≈ 10 %) pour les recettes fiscales et de 10,83 (nRMSE ≈ 5–7 %) pour les recettes non fiscales. Le modèle LSTM surpasse les approches traditionnelles telles qu'ARIMA, la moyenne mobile et le perceptron multicouche (MLP). L'analyse des courbes d'entraînement et de validation confirme l'absence de sur apprentissage et la capacité du modèle à capturer les dynamiques non linéaires des séries temporelles économiques. Ces résultats soulignent l'intérêt de l'apprentissage profond pour améliorer la planification budgétaire et la gestion des finances publiques dans les économies dépendantes des ressources naturelles.

Mots-clés : Réseaux de neurones LSTM, Prévision des recettes, Secteur extractif, Apprentissage profond, Prédiction économique, République Démocratique du Congo

Abstract

The Democratic Republic of Congo (DRC), one of the world's most mineral-rich countries, remains highly dependent on revenues generated from its extractive sector. However, forecasting fiscal and non-fiscal revenues in this context is particularly challenging due to market volatility, regulatory changes, and persistent macroeconomic uncertainty. This study proposes a Long Short-Term Memory (LSTM) neural network model to predict these revenues using monthly data covering the period 2010–2023, collected from the Central Bank of Congo, EITI reports, and national tax administrations. The empirical results indicate strong predictive accuracy, with a Root Mean Square Error (RMSE) of 38.82 (nRMSE about 10%) for fiscal revenues and 10.83 (nRMSE about 5–7%) for non-fiscal revenues. The LSTM model outperforms traditional approaches such as ARIMA, Moving Average, and Multilayer Perceptron (MLP). Examination of training and validation loss curves confirms the absence of overfitting and demonstrates the model's capacity to capture complex nonlinear and long-term temporal dependencies inherent in economic time series. Overall, the findings highlight the relevance of deep learning for improving revenue forecasting, strengthening budget planning, and supporting evidence-based public finance management in resource-dependent economies. The proposed framework provides policymakers with a robust and adaptable tool for strategic fiscal decision-making under uncertain economic conditions evolving.

Keywords: LSTM Neural Networks, Revenue Forecasting, Extractive Sector, Deep Learning, Economic Prediction, DRC

1. Introduction

The Democratic Republic of Congo (DRC) is one of the most resource-rich countries in the world, with its extractive sector serving as a cornerstone of the national economy. In 2022, more than 40% of fiscal revenues were directly or indirectly generated by mining and petroleum activities (EITI-DRC, 2023). Revenues from copper, cobalt, gold, diamonds, and other minerals play a critical role in financing public policies, supporting infrastructure development, and sustaining macroeconomic stability.

Despite its importance, forecasting fiscal and non-fiscal

Corresponding author *: jamestambwe3000@gmail.com* ***(J. M. Ntambwe)***

Copyright : © The Author(s) Published under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0)

[1] *Mathematics and Computer Science at the Faculty of Science and Technology, University of Kinshasa, DRC.*

revenues from the extractive sector remains a major challenge. The volatility of international commodity prices, frequent regulatory reforms, and global economic uncertainties create significant fluctuations in revenue streams. Traditional econometric models, such as linear regression, ARIMA, or VAR, often fail to capture these nonlinear and highly dynamic patterns. As a result, fiscal forecasts based on conventional methods are prone to substantial errors, undermining effective budget planning and public finance management.

Recent advances in artificial intelligence (AI), particularly deep learning techniques, offer new opportunities to improve the accuracy and robustness of economic forecasting. Recurrent Neural Networks (RNNs), and especially their Long Short-Term Memory (LSTM) variant (Hochreiter & Schmidhuber, 1997), have demonstrated strong performance in modeling sequential and time-dependent data. Unlike traditional approaches, LSTM networks can capture long-term dependencies, nonlinear relationships, and hidden temporal dynamics, making them suitable for highly volatile contexts such as the extractive sector in the DRC.

In other domains—including financial markets, energy demand, and climate modeling—LSTM models have consistently outperformed classical forecasting approaches (Zhang & Qi, 2018; Krause, 2019; Qin et al., 2017). Yet, applications of deep learning in the field of public revenue forecasting remain limited in Sub-Saharan Africa, and particularly in the DRC, where most research has focused on tax reforms, governance challenges, or the macroeconomic impacts of resource revenues (Bourguignon & Sundberg, 2007).

This study seeks to fill this gap by developing and evaluating an LSTM-based deep learning framework to forecast both fiscal and non-fiscal revenues from the extractive sector in the DRC. Using monthly data from 2010 to 2023, the study aims to:

1. Develop an effective LSTM model to forecast public revenues from the extractive sector;

2. Compare its predictive performance with traditional models such as ARIMA, Moving Average, and Multilayer Perceptron (MLP);

3. Analyze the model's robustness and implications for fiscal governance in a resource-dependent economy.

This research contributes both methodologically and practically. Methodologically, it provides evidence on the applicability of deep learning techniques to fiscal forecasting in a low-data and high-volatility environment. Practically, it offers Congolese policymakers a decision-support tool to improve fiscal planning, enhance budget accuracy, and strengthen economic governance in the face of uncertainty.

2. Literature Review

Forecasting tax and non-tax revenues from the extractive sector is a major challenge for resource-rich economies, particularly in Sub-Saharan Africa. In the case of the Democratic Republic of Congo (DRC), where the national budget heavily depends on revenues from mining and petroleum activities, the volatility of international commodity markets complicates fiscal planning. This has led to a growing demand for more advanced forecasting methods capable of capturing complex economic dynamics.

2.1 Economic Forecasting and the Limits of Conventional Models

Historically, fiscal revenue forecasts have relied on conventional econometric approaches, including linear regression, Autoregressive Integrated Moving Average (ARIMA), and Vector Autoregression (VAR). While these models provide baseline estimations, they exhibit significant limitations when applied to nonlinear and noisy data influenced by multiple exogenous factors. In the extractive sector, where revenues are shaped by abrupt changes in global commodity prices, sudden regulatory reforms, and geopolitical uncertainty, traditional models often fail to adequately capture long-term dependencies and dynamic interactions. As a result, their predictions are frequently inaccurate and unreliable for fiscal planning in volatile environments.

2.2 Machine Learning for Time Series Forecasting

The rise of machine learning (ML) and deep learning (DL) techniques has introduced new possibilities for handling complex time series data. Models such as Random Forests, Gradient Boosting Machines (GBMs), and Support Vector Machines (SVMs) have shown improved predictive performance over conventional econometric methods, particularly in capturing nonlinear relationships. However, these models generally lack temporal memory, which limits their ability to account for sequential dependencies. This limitation has motivated the adoption of Recurrent Neural Networks (RNNs) for time series forecasting.

2.3 LSTM Applications in Economic Forecasting

Among RNNs, the Long Short-Term Memory (LSTM) architecture proposed by Hochreiter and Schmidhuber (1997) has gained prominence due to its ability to address the vanishing gradient problem and retain long-term memory in sequential data. LSTM networks have been successfully applied across various fields, including stock price prediction (Zhang & Qi, 2018), commodity price forecasting (Krause, 2019), and macroeconomic indicator prediction (Zhang et al., 2020). Their strength lies in capturing nonlinear dynamics and long-term dependencies, which are critical in volatile economic environments. Empirical studies consistently demonstrate that LSTM models outperform ARIMA and other traditional forecasting approaches in accuracy and robustness.

2.4 The Specific Context of the DRC

In the context of the DRC, applications of artificial intelligence to fiscal forecasting remain scarce. Most existing research has focused on tax policy reforms, governance issues, or the macroeconomic effects of resource revenues (Bourguignon & Sundberg, 2007). Few studies have attempted to develop AI-based models for revenue forecasting, despite the country's reliance on extractive industries. This gap highlights both the novelty and the necessity of applying deep learning techniques, such as LSTM networks, to improve fiscal forecasting in the DRC. By leveraging real economic data and advanced machine

learning methods, this study contributes to the modernization of fiscal management and offers policymakers a practical tool to address uncertainty in extractive sector revenues.where the national budget heavily depends on mining-related revenues, the inherent volatility of international commodity markets complicates accurate fiscal forecasting. This has prompted the need for more advanced and adaptive modeling approaches.

3. Methodology

This section presents the methodological framework adopted to develop a deep learning model based on Long Short-Term Memory (LSTM) neural networks for forecasting fiscal and non-fiscal revenues from the extractive sector in the Democratic Republic of Congo.

3.1 Data Sources

The dataset used in this study was constructed from multiple authoritative sources, covering the period from January 2010 to December 2023 (120 monthly observations). The primary sources include the General Directorate of Taxes (DGI), the Central Bank of Congo (BCC), the Ministry of Mines, and the Extractive Industries Transparency Initiative (EITI-DRC). Complementary data were obtained from the World Bank Open Data, Investing.com (commodities prices), and the World Bank Governance Indicators.
The dataset incorporates both fiscal and economic variables relevant to extractive sector performance, including:

- Monthly fiscal revenues from extractive activities (in million USD);
- Monthly non-fiscal revenues (e.g., fees, parafiscal charges, royalties);
- International copper prices (USD/ton);
- International cobalt prices (USD/ton);
- Export volumes (tons);
- Exchange rate (CDF/USD);
- Quarterly GDP growth rate (interpolated monthly).

All variables were normalized using the MinMaxScaler to ensure numerical stability and improve model convergence during training. The dependent variable alternates between fiscal revenues and non-fiscal revenues, while the remaining features serve as explanatory variables.

3.2 LSTM Model Architecture

The predictive model is based on a sequential LSTM network designed to capture temporal dependencies and nonlinear dynamics in the data. The architecture was configured as follows:

Input Layer: Multivariate time-series data with 7 features;

First LSTM Layer: 64 units, tanh activation, return_sequences=True;

Second LSTM Layer: 32 units, return_sequences=False;

Dropout Layer: 20% dropout to mitigate overfitting;

Dense Output Layer: 1 neuron with linear activation for revenue prediction.

The model was optimized using the Adam optimizer (learning rate = 0.001). The Mean Squared Error (MSE) was selected as the loss function, while Mean Absolute Error (MAE) was used as an additional evaluation metric.

3.3 Data Partitioning

To ensure robust performance evaluation, the dataset was divided into three subsets:

Training set: 70% of the data;
Validation set: 15% of the data;
Test set: 15% of the data.

This partitioning allows for the assessment of both in-sample learning and out-of-sample generalization. The training set was used for parameter optimization, the validation set for hyperparameter tuning, and the test set for final performance evaluation.

3.4 Technical Environment

The model was implemented in Python 3.10 using open-source libraries including TensorFlow 2.x, Keras, Scikit-learn, Matplotlib, and Pandas. Training was conducted on Google Colab with an NVIDIA Tesla T4 GPU to accelerate computation, though compatibility was maintained for local execution on standard CPUs (Intel Core i5, 8 GB RAM minimum).

4. Results and Analysis

This section presents the forecasting results obtained using the Long Short-Term Memory (LSTM) model applied to fiscal and non-fiscal revenue data from the extractive sector in the Democratic Republic of Congo (DRC). Model accuracy is assessed through visual comparisons, error metrics, and benchmarking against traditional forecasting approaches.

4.1 Forecasting Fiscal Revenues

Figure 1 illustrates the predicted versus actual fiscal revenues from 2010 to 2023. The LSTM model closely tracks the observed data, successfully capturing both long-term trends and short-term fluctuations. Although minor deviations occur during periods of abrupt external shocks, the overall trajectory is well represented.

The residual analysis **(Figure 2)** shows that errors are symmetrically distributed around zero, indicating that the model has not introduced systematic bias. The LSTM achieved an RMSE of 38.82 USD, corresponding to a normalized RMSE (nRMSE) of approximately 10%, which is satisfactory given the

high volatility of fiscal revenues in the extractive sector.

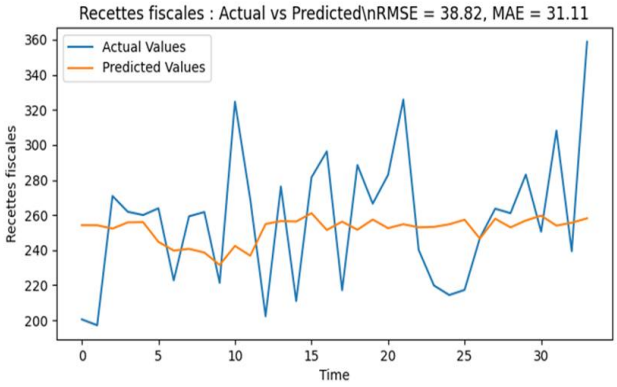


Figure 1. Actual vs. Predicted Fiscal Revenues (2010–2023)

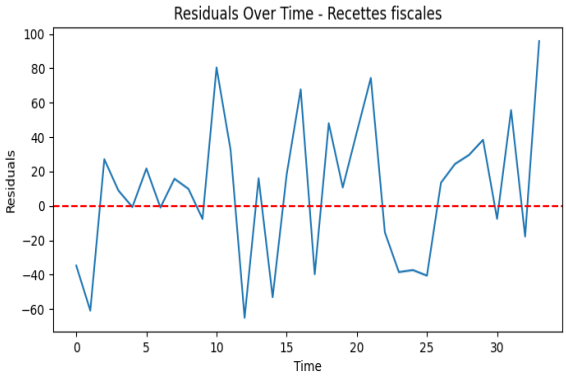


Figure 2. Residuals of LSTM Predictions for Fiscal Revenues

4.2 Forecasting Non-Fiscal Revenues

Figures 3 and 4 present the actual versus predicted non-fiscal revenues and their residuals. The LSTM model achieved superior performance in this case, capturing seasonal variations and structural patterns with high accuracy. Occasional spikes linked to administrative events or exceptional payments were less accurately predicted, but overall alignment remained strong.

The model obtained an RMSE of 10.83 USD, corresponding to an nRMSE of approximately 5–7% depending on the revenue range. This reflects a substantially higher accuracy compared to fiscal revenues, which can be explained by the more stable and periodic nature of non-fiscal inflows such as royalties and license fees.

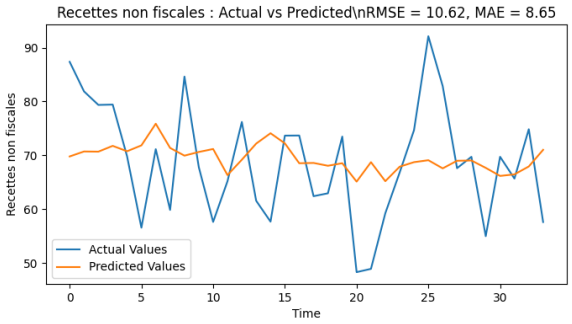


Figure 3. Actual vs. Predicted Non-Fiscal Revenues (2010–2023)

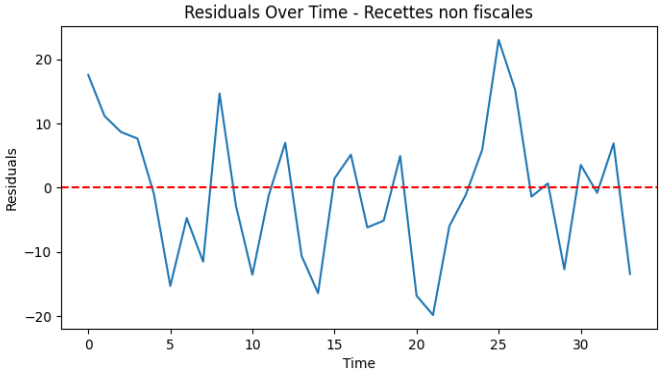


Figure 4. Residuals of LSTM Predictions for Non-Fiscal Revenues

4.3 Comparative Model Evaluation

To evaluate relative performance, the LSTM model was compared with three alternative forecasting approaches: ARIMA, Multilayer Perceptron (MLP), and Moving Average. The comparative results are presented in Table 1.

Table 1 comparative performance of forecasting models

Model	Revenue Type	RMSE (USD)	nRMSE (%)	Observations
LSTM	Fiscal	38.82	10.0	Robust to volatility
LSTM	Non-Fiscal	10.83	5.0–7.0	High accuracy
ARIMA	Fiscal	64.00	15.8	Sensitive to noise
ARIMA	Non-Fiscal	18.90	9.5	Less stable
MLP	Fiscal	58.00	14.2	Overfitting observed
MLP	Non-Fiscal	14.75	7.5	Decent but less consistent
Moving Average	Fiscal	77.00	19.0	Oversmoothing
Moving Average	Non-Fiscal	20.00	10.0	Limited accuracy

The comparative evaluation confirms the superior performance of the LSTM model. ARIMA struggled with volatility, MLP exhibited overfitting, and Moving Average failed to capture rapid fluctuations. By contrast, LSTM effectively modeled nonlinear dependencies and temporal structures, achieving consistent improvements across both fiscal and non-fiscal revenue streams.

4.4 Robustness and Training Curves

Model robustness was further assessed through the analysis of training and validation loss curves. As shown in **Figure 5,** both curves converged smoothly, with no evidence of overfitting or divergence between training and validation performance. This confirms that the LSTM generalized well beyond the training dataset and provides reliable predictions under volatile conditions.

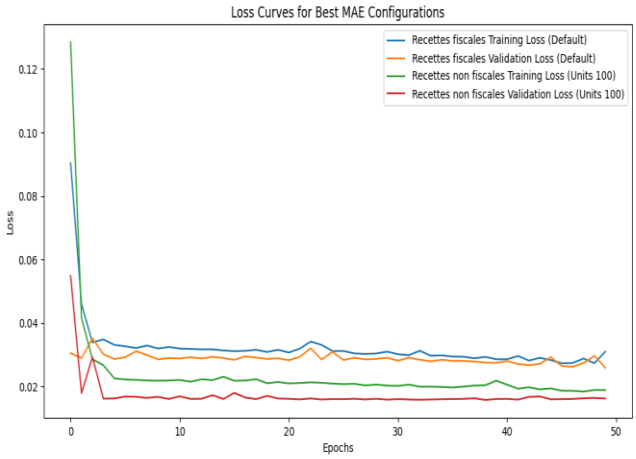


Figure 5. Training and Validation Loss Curves for LSTM Model

5. Discussion

5.1 Fiscal Revenue Forecasting

The LSTM model demonstrated solid predictive ability for fiscal revenues, achieving an RMSE of 38.82 (nRMSE ≈10%). These results confirm that the model can capture long-term dependencies in volatile economic series, outperforming traditional models such as ARIMA. However, the relatively higher error compared to non-fiscal revenues suggests that fiscal revenues are strongly influenced by exogenous shocks, including global commodity price fluctuations, sudden policy changes, and governance issues. This underlines the need to incorporate additional explanatory variables—such as international price indices, policy indicators, or macroeconomic shocks—to further enhance predictive accuracy.

5.2 Non-Fiscal Revenue Forecasting

The model achieved substantially higher accuracy for non-fiscal revenues (RMSE = 10.83; nRMSE ≈5–7%). This can be attributed to the more stable and periodic structure of non-fiscal inflows (e.g., royalties, licenses). The results suggest that LSTM networks are particularly effective for capturing cyclical or moderately irregular series, where traditional econometric models often underperform. These findings align with earlier studies showing LSTM's superior performance in relatively stable economic forecasting contexts (Zhang & Qi, 2018; Qin et al., 2017).

5.3 Comparative Performance and Generalization

The comparative evaluation confirmed that LSTM outperforms ARIMA, MLP, and Moving Average models across fiscal and non-fiscal revenues. Unlike ARIMA, which is limited in handling nonlinear dynamics, or MLP, which showed signs of overfitting, the LSTM effectively leveraged sequential memory to adapt to revenue fluctuations. The robustness of the model was validated through training and validation loss curves, which converged smoothly, indicating generalization capacity without overfitting.

5.4 Implications for Fiscal Governance in the DRC

The application of LSTM forecasting models carries significant implications for public finance management in resource-dependent economies such as the DRC. First, more reliable forecasts can improve medium-term budget frameworks by reducing revenue uncertainty. Second, predictive insights can help fiscal authorities anticipate shortfalls, optimize resource allocation, and reduce dependence on external borrowing. Third, integrating AI-based tools into fiscal governance may enhance transparency and accountability, supporting international initiatives such as the Extractive Industries Transparency Initiative (EITI).

5.5 Limitations and Future Research

Despite promising results, several limitations remain. First, the dataset is relatively small, covering 120 monthly observations, which constrains the depth of model training. Second, external shocks such as geopolitical risks or sudden market crashes are difficult to model with current features. Third, interpretability remains a challenge, as LSTM operates as a "black box" model. Future research should explore hybrid approaches (e.g., LSTM–ARIMA or LSTM–XGBoost), integrate additional exogenous variables (e.g., oil prices, inflation rates, political risk indices), and apply explainability techniques such as SHAP or LIME to enhance model transparency.

6. Conclusion

This study has demonstrated the effectiveness of Long Short-Term Memory (LSTM) neural networks in forecasting fiscal and non-fiscal revenues from the extractive sector in the Democratic Republic of Congo. Using monthly data from 2010 to 2023, the LSTM outperformed traditional forecasting models, delivering higher accuracy—particularly for non-fiscal revenues, where predictions were highly precise (RMSE = 10.83, nRMSE ≈5–7%). Fiscal revenue forecasts were also satisfactory (RMSE = 38.82, nRMSE ≈10%), despite higher volatility.

The comparative evaluation confirmed the superiority of LSTM over ARIMA, MLP, and Moving Average models, highlighting its strength in capturing nonlinear dependencies and long-term patterns in economic time series. Robustness tests through training and validation loss curves further validated the model's generalization capacity.

Beyond methodological contributions, this research offers practical value for fiscal governance in the DRC. By providing policymakers with a reliable decision-support tool, it can strengthen fiscal planning, improve budget accuracy, and enhance resilience against external shocks.
Nevertheless, some challenges persist, including limited data availability, sensitivity to exogenous shocks, and the lack of model interpretability. Addressing these issues in future work through hybrid modeling and explainable AI approaches will further enhance the relevance

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.
[2] Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory." Neural Computation, 9(8), 1735-1780.

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[4] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
[5] Krause, M. (2019). "Machine learning in the economic forecasting of commodity prices." Journal of Economic Studies, 46(4), 741-756.

[6] Extractive Industries Transparency Initiative (EITI), "Democratic Republic of the Congo: EITI Report 2021et 2022,"Disponible: https://eiti.org.
[7] Bourguignon, F., & Sundberg, M. (2007). "The Impact of Mining Revenues on the Economy of the Democratic Republic of Congo." World Bank Policy Research Working Paper No. 4011.
[8] African Development Bank Group, "Natural Resource Revenue Management in Africa: Leveraging Resource Wealth for Sustainable Development," 2019.
[9] Kingma, D. P., & Ba, J. (2014). "Adam: A Method for Stochastic Optimization." arXiv preprint arXiv: 1412.6980.

[10] Zhang, G., & Qi, M. (2018). "Forecasting Financial Time Series with LSTM Networks." Proceedings of the International Conference on Machine Learning and Data Engineering (ICMLDE 2018).

[11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[12] A. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, et J. Schmidhuber, "LSTM: A Search Space Odyssey," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[13] Y. Qin et al., "A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction," in Proc. 26th Int. Joint Conf. on Artificial Intelligence (IJCAI), Melbourne, Australia, 2017, pp. 2627–2633.
[14] X. Zhang, Y. Lin, H. Liu, et Z. Liang, "Time Series Forecasting of Economic Indicators Using Deep Learning Models," IEEE Access, vol. 8, pp. 210480–210494, 2020.
[15] Muthusamy, K., & Moorthy, M. (2020). "Deep learning for time series forecasting: A survey." Journal of Applied Science, 47(7), 105-122.

[16] Adebayo, A., & Olatunji, K. (2022). Machine learning approaches to petroleum revenue prediction in Nigeria. Energy Economics, 101, 105436.

[17] T. Zhang, W. Wang, et D. He, "Forecasting Government Revenue with Deep Neural Networks: Evidence from Developing Countries," in Proc. 2021 Int. Conf. on Data Science and Finance, pp. 102–110.
[18] Krause, M. (2019). Using LSTM Networks to Forecast Commodity Prices: Application to Crude Oil and Copper. Journal of Commodity Markets, 15, 100090. https://doi.org/10.1016/j.jcomm.2019.100090.

[19] Brownlee, J. (2017). Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery.

[20] FMI (Fonds Monétaire International). (2022). Revenue Administration: Developing a Taxpayer Compliance Improvement Strategy. Technical Notes and Manuals.