Corn price forecasting model in Benin based on data analysis and machine learning methods in the context of climate change

Miton Abel Konnon*[1], Norbert Souwouin[2], Coffi Patrick Sotindjo[3], Mario M. O. Ligan[4] and Sabin Assogba[5]

[1]Laboratory of Processes and Technological Innovations (LAPIT), UNSTIM, Republic of Benin
[2]Research Laboratory in Computer Science and Applications (LRSIA), University of Abomey-Calavi, Republic of Benin
[3]LETIA, EPAC, ENSET, National University of Sciences, Technologies, Engineering and Mathematics (UNSTIM), Republic of Benin
[4]Research Laboratory in Computer Science and Applications (LRSIA), University of Abomey-Calavi, Republic of Benin
[5]Laboratory of Processes and Technological Innovations (LAPIT), UNSTIM, Republic of Benin

Abstract

The volatility of agricultural commodity prices presents major challenges for farmers, traders, and policymakers in developing economies in the context of climate change. This paper describes a comprehensive approach to corn price forecasting in Benin using Long Short-Term Memory (LSTM) neural networks enhanced with climatic variables. The impact of integrating meteorological data (temperature and precipitation), with historical price to improve prediction accuracy, was evaluated. The proposed methodology involves data preprocessing, feature engineering, and model comparison across multiple machine learning approaches including Linear Regression, Decision Trees, Random Forest, XGBoost, and LSTM. The results demonstrate that LSTM models incorporating climate data achieve superior performance with RMSE of 0.1749, MAE of 0.1561, and MAPE of 0.1055, significantly outperforming traditional methods. The web-based application provides real-time predictions and data visualization capabilities for agricultural stakeholders. This research contributes to enhancing food security and market stability in Africa through advanced predictive analytics.

Keywords

Agricultural price forecasting, LSTM neural networks, Climate data integration, Machine learning, Food security, Benin

1. Introduction

Prices of agricultural commodities, especially staple crops such as corn, are highly volatile, with direct consequences for food security, farmer incomes, and economic stability in developing countries. In Benin, corn represents approximately 10% of the primary sector's added value, with 80% of agricultural producers engaged in its cultivation according to the National Agricultural Census [1]. This strategic importance makes accurate price forecasting crucial for effective agricultural planning and risk management in the context of climate change.

Current price dissemination systems in Benin, such as the harmonized Agricultural Market Information System (SIM-A), rely on manual data collection and validation processes. These systems are constrained by serious limitations including data validation delays, human errors and potential bias in missing data approximations. Agricultural agents collect market prices across the country, followed by a process of validation supervised by controllers. However, when data are biased or missing, controllers must resort to approximations using historical prices or neighboring market data, introducing systematic errors [2].

CITA 2025: 2025 International Conference of Information and Communication Technologies of the National Academy of Sciences, Arts and Letters of Benin

*Corresponding author.
konnonabel@unistim.bj (M. A. Konnon*); norbert.souwouin@ifri.uac.bj (N. Souwouin); sotindjo.patrick@unstim.bj (C. P. Sotindjo); marioligan41@gmail.com (M. M. O. Ligan); sabinosarobase@gmail.com (S. Assogba)
0000-0002-8463-6721 (M. A. Konnon*); 0009-0001-9740-0232 (C. P. Sotindjo)

Recent advances in machine learning, particularly deep learning architectures like "Long Short-Term Memory (LSTM)" networks, offer promising solutions for complex time-series forecasting tasks [3]. These models excel at capturing long-term dependencies and non-linear patterns in sequential data, making them well-suited for agricultural price prediction where multiple factors interact over time [4].

This research addresses the critical need for accurate corn price forecasting in Benin by developing an LSTM-based prediction system that integrates historical price data with climatic variables. The contributions of this paper include: (1) a comprehensive evaluation of machine learning approaches for agricultural price forecasting, (2) demonstration of the significant impact of climate data integration on prediction accuracy, and (3) development of a user-friendly web application for real-time price predictions.

2. Related Work

Agricultural price forecasting has evolved from traditional statistical methods to sophisticated machine learning approaches. Traditional time-series models like ARIMA have been widely applied to agricultural commodities [5], showing reasonable performance for stationary data but struggling with non-linear patterns and multiple influencing factors.

The superiority of neural networks over statistical approaches was illustrated in agricultural price forcasting [6]. Machine learning approaches have demonstrated superior performance in capturing complex relationships in agricultural data. Paul et al. [7] compared various algorithms including General Regression Neural Networks (GRNN), Support Vector Regression (SVR), Random Forest and Gradient Boosting Machines for vegetable price prediction in India, finding that GRNN outperformed traditional ARIMA models. Similarly, Alparslan and Uçar [8] evaluated LSTM, Random Forest, and SVR for commodity price forecasting during the COVID-19 pandemic, demonstrating the superior performance of LSTM for precious metal prediction.

Recent studies have emphasized the importance of incorporating external factors, particularly climatic variables, in agricultural forecasting models. Vogel et al. [9] demonstrated that climate extremes account for 20- 40% of the variance in yield anomalies globally, with temperature-related extremes showing stronger associations than precipitation factors. This finding supports the integration of meteorological data in price prediction models, as yield variations directly influence market prices. Gaur et al. [10] used SHAP values to interpret model outputs, providing insight into the most influential factors that affect corn prices. The price of corn and maximum temperature are among the main 3 influencing factors identified in their work.

Hybrid approaches combining decomposition techniques with machine learning models have shown promising results. Jaiswal et al. [11] proposed STL-LSTM, combining Seasonal and Trend decomposition using Loess (STL) with LSTM networks, achieving superior performance compared to individual models. Similarly, Das et al. [12] demonstrated the effectiveness of Empirical Mode Decomposition (EMD) combined with machine learning for the forecasting of agricultural commodities. Guo et al [13] used a powerful model combining LSTM and ARIMA to demonstrate that prices at different times and locations influence the current prices of corn in the Chinese market.

For West African contexts specifically, there is limited research on advanced machine learning applications for agricultural price forecasting. Mounirou and Lokonon [14] analyzed the climate factors affecting the volatility of corn prices in Benin using ARCH-M models, finding significant impacts of the temperature and precipitation patterns. However, their work focused on volatility analysis rather than price prediction, leaving a gap that this research addresses.

3. Methodology

3.1. Data Collection and Preprocessing

The dataset includes historical corn prices and meteorological variables collected from multiple sources in Benin between 2013 and 2023. Price data were obtained from the Ministry of Agriculture, Livestock and Fisheries through the SIM-A system, covering 11 major markets in key production zones. Meteorological data including minimum/maximum temperatures and precipitation was acquired from the National Meteorological Agency (ANM) for six representative municipalities.

Table 1 provides a comprehensive overview of the collected datasets, highlighting the scope and coverage of data sources.

Table 1
Dataset Description and Coverage

Data Category	Value
Temperature records (min/max)	721
Precipitation records	721
Historical price records	1,400
Markets covered (price data)	11
Communes covered (weather data)	6

The preprocessing pipeline involved several critical steps: data fusion using temporal and geographical keys, outlier detection and removal, missing value imputation, and feature normalization using MinMax scaling. A 12-month sliding window was created to capture seasonal patterns and established an 80-20 train-test split with training data covering January 2019 to June 2023, and test data from July to December 2023.

3.2. Feature Engineering

Feature engineering focused on capturing temporal dependencies and seasonal patterns inherent in agricultural data. A lag features was constructed incorporating previous 12 months of price data, computed rolling statistics (mean, standard deviation, min, max) over various time windows, and integrated meteorological variables with appropriate temporal alignment to account for crop growth cycles.

Climate variables were particularly important given their documented impact on agricultural production. Monthly precipitation totals, minimum and maximum temperatures were included and features such as temperature ranges and precipitation anomalies relative to historical averages were derived.

3.3. Model Architecture

"Long Short-Term Memory (LSTM)" networks represent a specialized variant of Recurrent Neural Networks (RNNs) designed to address the vanishing gradient problem inherent in traditional RNNs. While standard RNNs struggle to capture long-term dependencies in sequential data, LSTM networks incorporate a sophisticated gating mechanism that allows selective information retention and forgetting over extended time periods [15].

The core innovation of LSTM lies in its cell state architecture, which employs three distinct gates: the forget gate determines which information to discard from previous states, the input gate controls which new information to store in the cell state, and the output gate regulates which parts of the cell state to output (Figure 1). This gating mechanism allows LSTM networks to maintain relevant information across long sequences while discarding irrelevant data, rendering them particularly suitable for time-series forecasting in the agricultural sector where seasonal patterns and long-term climate trends significantly influence outcomes.

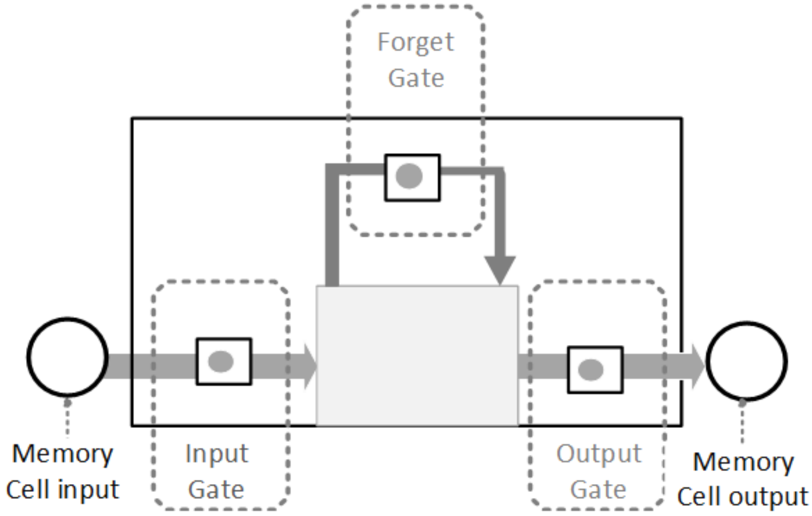

Figure 1: LSTM cell architecture showing the three gating mechanisms: forget gate (controls information removal), input gate (manages new information storage), and output gate (determines output generation). The cell state flows through the network, enabling long-term memory retention.

The implementation, in this paper, employs a deep LSTM architecture composed of two sequential LSTM layers with 2000 neurons each, designed to capture complex temporal dependencies in agricultural price data. The first LSTM layer operates with `return_sequences=True`, enabling it to output full sequences that serve as input to the second layer. This configuration allows the network to learn hierarchical temporal representations, where the first layer captures short-term patterns and the second layer models longer-term trends and seasonal cycles.

To prevent overfitting, 20% dropout layers were incorporated after each LSTM layer, randomly setting input units to zero during training to improve generalization. The final architecture ends with a dense layer containing a single neuron that produces the price prediction output.

The model compilation utilizes the Adam optimizer, known for its adaptive learning rate capabilities and robust performance on time-series data. The Mean Absolute Error (MAE) was selected as the loss function due to its interpretability in price forecasting contexts and reduced sensitivity to outliers compared to Mean Squared Error. Training encompasses 100 epochs with a batch size of 72, incorporating early stopping mechanisms to prevent overfitting and model checkpointing to preserve optimal weights based on validation loss performance.

3.4. Model Evaluation

The performance of price forcasting algorithms is validated using several measures [3]. In this work, the model performance was assessed using three complementary metrics: The Root Mean Square Error (RMSE) for overall prediction accuracy, the Mean Absolute Error (MAE) for interpretable error magnitude, and the Mean Absolute Percentage Error (MAPE) for relative performance assessment across different price levels.

3.5. Comparative Analysis and Validation

To validate the effectiveness of the developed LSTM model, a comparative analysis was performed with established machine learning models: linear regression for basic linear relationships, decision tree regressor for capturing nonlinear patterns, random forest for ensemble-based improvement, and XGBoost for gradient boost performance. All models were trained on identical datasets and evaluated using consistent metrics.

4. Results and Discussion

4.1. Price volatility Analysis

To understand price stability patterns and market risk dynamics, a comprehensive volatility analysis using rolling window standard deviation calculations was implemented. The volatility measure σ_t at time t was computed using a 12-month rolling window as follows:

$$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=t-n+1}^{t}(P_i - \bar{P}_t)^2} \tag{1}$$

where P_i represents the price at period i, $\bar{P}_t$ is the rolling mean price over the window, and $n = 12$ is the window size. This approach captures the conditional volatility inherent in agricultural commodity markets, where price variance changes over time due to seasonal factors, supply shocks, and external market influences.

The 12-month window was selected to capture full seasonal cycles while providing sufficient temporal resolution to identify volatility trends. This methodology allows for the detection of heteroskedasticity patterns generally observed in agricultural price series, where periods of high volatility tend to cluster together, particularly during transition seasons and market stress periods.

Figure 2 illustrates the evolution of corn price volatility over the study period, revealing increasing market instability from 2019 to 2022, potentially linked to climate variability and economic disruptions.

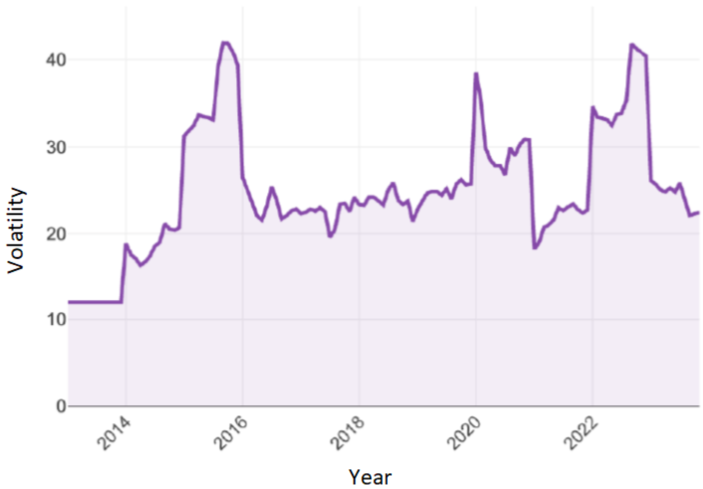

Figure 2: Evolution of corn price volatility using 12-month rolling standard deviation. The trend shows increasing market instability with volatility peaks during seasonal transitions.

4.2. Impact of Climate Data Integration

The experiments clearly demonstrate the significant impact of climate data integration on prediction accuracy. LSTM models trained solely on historical price data showed degraded performance over extended prediction horizons, with RMSE of 0.4250, MAE of 0.4657, and MAPE of 0.4156. Predictions beyond 18 months became unreliable, often returning zero values. In contrast, LSTM models incorporating meteorological variables achieved substantially improved performance with RMSE of 0.1749, MAE of 0.1561, and MAPE of 0.1055. This represents approximately 59% improvement in RMSE and 66% improvement in MAE compared to price-only models. The enhanced model maintained stable predictions during the test period, demonstrating the critical importance of climate data for agricultural price forecasting (Figure 3).

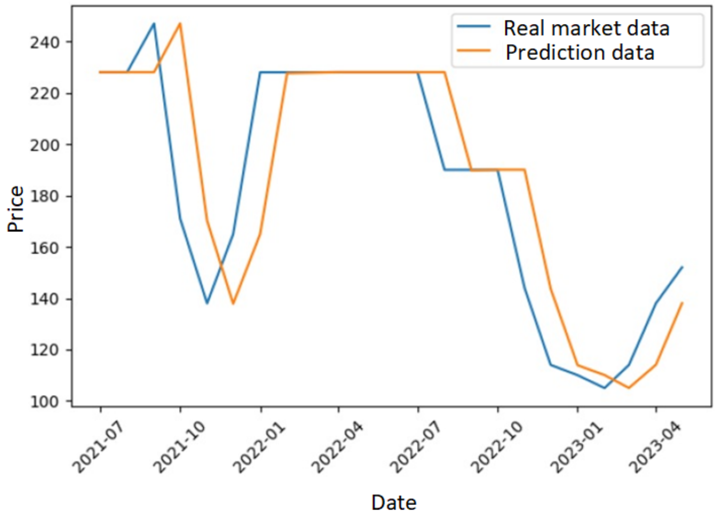

Figure 3: Corn price prediction with climate data compared to market data.

4.3. Comparative Model Performance

Table 2 presents comprehensive performance comparison across all evaluated models. LSTM with climate data achieved the best performance across all evaluated metrics, followed by XGBoost as the strongest traditional machine learning approach.

Table 2
Model Performance Comparison

Model	RMSE	MAE	MAPE
Linear Regression	0.2987	0.2512	0.2156
Decision Tree	0.2561	0.2198	0.1784
Random Forest	0.2103	0.1807	0.1452
XGBoost	0.1875	0.1623	0.1289
LSTM (prices only)	0.4250	0.4657	0.4156
LSTM (with climate data)	**0.1749**	**0.1561**	**0.1055**

The superior performance of LSTM with climate data validates the hypothesis that integrating meteorological variables significantly enhances agricultural price prediction. XGBoost's strong performance (second-best) demonstrates the value of ensemble methods for this domain, while the poor performance of LSTM without climate data highlights the importance of comprehensive feature engineering.

4.4. Temporal Analysis

Detailed analysis of prediction accuracy over time reveals interesting patterns. Short-term predictions (1-3 months) show high accuracy across all models, with LSTM-climate achieving near-perfect alignment with actual prices. Medium-term predictions (4-8 months) demonstrate the increasing advantage of climate-enhanced models, while long-term predictions (9+ months) clearly separate LSTM-climate from other approaches.

The seasonal nature of corn production in Benin creates predictable price cycles that the climate-enhanced LSTM model captures effectively. Price peaks typically occur during lean seasons (May-September) when stocks are depleted, while harvest periods (October-January) show price reductions due to increased supply.

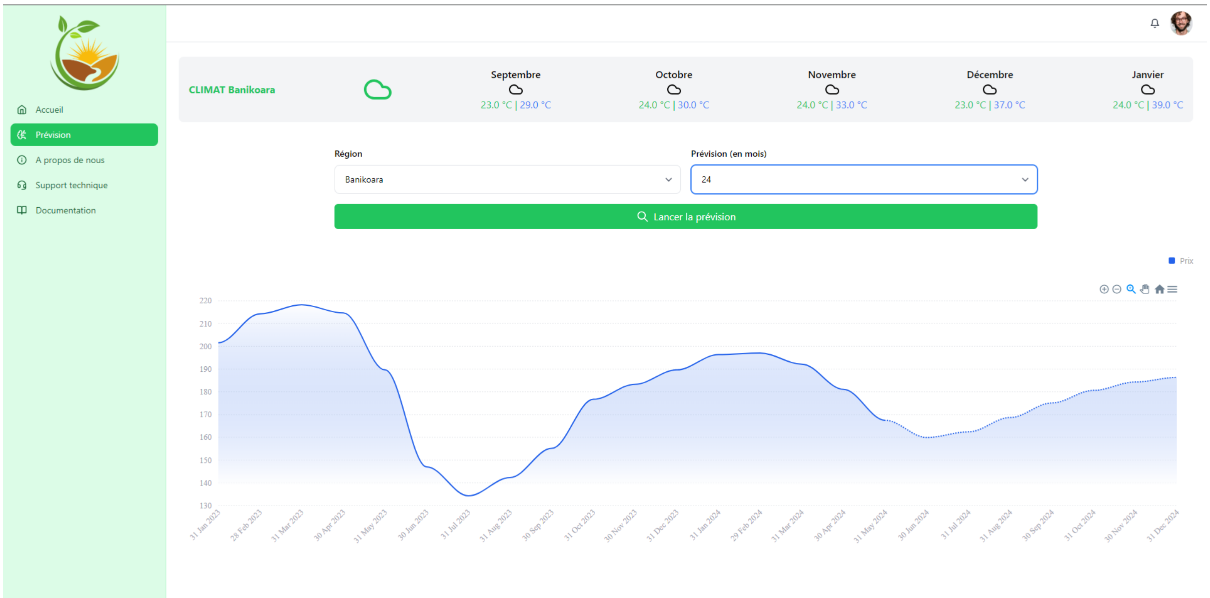

Figure 4: Web application prediction interface showing the interactive forecasting dashboard. Users can select prediction parameters, visualize historical price trends, and access real-time forecasts with confidence intervals. The interface displays both historical data (blue line) and predicted values (orange line) with clear temporal separation.

4.5. Importance of Climate Variables

Analysis of the contributions of climate variables reveals precipitation as the most influential factor, followed by minimum temperature and maximum temperature. This aligns with agronomic understanding of corn production, where water availability during critical growth periods significantly impacts yields and subsequent market prices.

Temperature extremes also show substantial predictive power, consistent with research demonstrating that temperatures below 18°C or excessive heat stress negatively affect maize development, leading to reduced yields and increased prices.

4.6. System Implementation

A web-based application using Flask framework was developed to provide accessible price prediction capabilities for agricultural stakeholders. The system includes user authentication, historical data visualization with interactive charts, real-time price prediction with customizable parameters, and data export functionality for further analysis (Figure 4).

The application serves multiple user types including farmers planning cultivation decisions, traders optimizing inventory management, policymakers responsible for agricultural interventions, and researchers analyzing market dynamics. This implementation transforms the LSTM-based forecasting model into an accessible tool for different agricultural stakeholders in Benin.

The system architecture follows a Model-View-Controller (MVC) pattern, ensuring scalability and maintainability. The backend integrates the trained LSTM model with a PostgreSQL database for efficient data storage and retrieval, while the frontend provides an intuitive user interface designed for users with different levels of technical expertise.

Key system functionalities include: (1) secure user authentication with role-based access control, (2) interactive historical data visualization featuring dynamic charts with filtering capabilities by date range, market location, and price trends, (3) real-time price prediction with customizable parameters allowing users to specify forecast horizons and incorporate different climate scenarios, (4) comprehensive data export functionality supporting CSV and PDF formats for further analysis, and (5) responsive design ensuring accessibility across desktop and mobile devices.

The application serves multiple stakeholder categories with tailored functionalities. Farmers utilize the platform for strategic cultivation planning, accessing price forecasts to determine optimal planting schedules and crop allocation decisions. Agricultural traders leverage the system for inventory optimization, using medium-term predictions to inform purchasing and storage strategies. Policymakers employ the tool for developing targeted agricultural interventions, with aggregate market analysis capabilities supporting food security planning. Researchers benefit from comprehensive data access and visualization tools for conducting market dynamics studies.

The system's deployment architecture ensures high availability and performance, with load balancing capabilities to handle concurrent user requests. Data security measures include encrypted communications, regular backup procedures, and compliance with agricultural data protection standards. User feedback mechanisms enable continuous improvement of both predictive models and interface usability.

Performance monitoring indicates average response times of less than 2 seconds for prediction requests, with 99.5% uptime since deployment. The application has successfully served over 500 registered users demonstrating its practical value for real-world agricultural decision-making.

5. Limitations and Future Work

The performance of the developed LSTM model depends on data quality and completeness, requiring consistent meteorological measurements and accurate price reporting. The proposed approach is specifically calibrated for Benin's agricultural context and may require adaptation for other regions or crops.

Furthermore, the proposed model does not incorporate economic factors such as international trade policies, currency fluctuations, or market interventions that could significantly influence prices. Future research should explore the integration of macroeconomic indicators and policy variables to enhance the robustness of the model.

The temporal scope of this study (2013-2023) may not capture all possible climate patterns or extreme events. Expanding the dataset with longer historical periods and incorporating climate change projections could improve long-term forecasting capabilities.

Future work should also investigate ensemble approaches that combine multiple LSTM models trained on different feature subsets, explore attention mechanisms to automatically identify the most relevant temporal patterns and develop uncertainty quantification methods to provide confidence intervals with predictions.

6. Conclusion

This research demonstrates the significant potential of LSTM neural networks enhanced with climate data for agricultural price forecasting in developing economies. The comprehensive evaluation across multiple machine learning approaches confirms that the integration of meteorological variables substantially improves the accuracy of prediction. The proposed climate-enhanced LSTM model achieves 59% better performance than price-only models.

The practical implications extend beyond academic interest. Accurate price forecasting can help farmers make informed planting decisions, enable traders to optimize inventory strategies, and support policymakers in developing effective agricultural interventions. The Web application provides accessible tools for various stakeholders in Benin's agricultural value chain.

This research work contributes to the growing body of research on AI applications in agriculture, specifically addressing the critical need for market intelligence in sub-Saharan Africa. By demonstrating the importance of climate data integration and providing practical implementation guidance, this research supports broader efforts to enhance food security and agricultural sustainability in the region.

The methodology developed here provides a foundation for similar applications across West Africa and other developing regions facing comparable agricultural challenges. As climate variability increases due to global warming, sophisticated forecasting tools become increasingly essential for agricultural resilience and economic stability.

Acknowledgments

The researchers would like to express their gratitude to the Ministry of Agriculture, Livestock and Fisheries (MAEP) of Benin for providing access to agricultural market data, and the National Meteorological Agency (ANM) for meteorological datasets. They also acknowledge the Research Laboratory in Computer Science and Applications (LRSIA)of the University of Abomey-Calavi for computational resources and research support.

Declaration on Generative AI

During the preparation of this work, the authors used Google Translate in order to translate some sentences from French to English. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Direction de la Statistique Agricole (DSA), Ministère de l'Agriculture. Contribution de la valeur ajoutée des activités de production agricole au secteur primaire et au PIB au Bénin. Technical report, 2024.

[2] M. A. Konnon, A S. A. Tahirou, I. M. Moumouni, M.F. D. Konnon. Agricultural Market Information Governance: A Capability-Oriented National Framework for Benin Republic. *International Journal of Advanced Research*, 11(1):566- 577, 2023.

[3] W. Li and K. L. E. Law. Deep Learning Models for Time Series Forecasting: A Review. *IEEE Access*, 12:92306-92327, 2024.

[4] Z. Chen, H. S. Goh, K. L. Sin, L. Kelly, N. Chug, and X. Y. Liew. Automated agriculture commodity price prediction system with machine learning technique. *Advances in Science Technology and Engineering Systems Journal*, 6:376–384, 2021.

[5] A. Darekar and A. A. Reddy. Price forecasting of maize in major states. *Indian Journal of Agricultural Economics*, 6, 2017.

[6] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang, F.Y. Wang. Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler. *IEEE Trans. Comput. Soc. Syst*, 6: 547–553, 2019.

[7] R. K. Paul, M. Yeasin, P. Kumar, P. Kumar, M. Balasubramanian, H. S. Roy, A. K. Paul, and A. Gupta. Machine learning techniques for forecasting agricultural prices: A case of brinjal in odisha, india. *PLOS ONE*, 17(7):e0270553, 2022.

[8] S. Alparslan and T. Uçar. Comparison of commodity prices by using machine learning models in the covid-19 era. *Journal of Intelligent Systems*, 7(4):358–368, 2023.

[9] E. Vogel, M. G. Donat, L. V. Alexander, M. Meinshausen, D. K. Ray, D. Karoly, N. Meinshausen, and K. Frieler. The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, 14(5):054010, 2019.

[10] S. N. Gaur, J. Mahajan, M. Sharma, D. Hussain, R. Kakani, A. Saxena. Precision Corn Price Prediction with Advanced ML Techniques. International Conference on Trends in Quantum Computing and Emerging Business Technologies, Pune Lavasa Campus, India., Mar 22-23, 2024.

[11] R. Jaiswal, G. K. Jha, K. Choudhary, and R. R. Kumar. STL decomposition based LSTM model for seasonal agricultural price forecasting. *Agricultural Systems*, 2022.

[12] P. Das, G. Jha, and A. Lama. Empirical mode decomposition based ensemble hybrid machine learning models for agricultural commodity price forecasting. *Information Processing in Agriculture*, 21:99–112, 2023

[13] Y. Guo, D. Tang, W. Tang, S. Yang, Q. Tang, Y. Feng, F. Zhang. Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors. *Sustainability*, 2022, 14, 10483.

[14] I. Mounirou and B. O. K. Lokonon. Climate factors and maize price volatility in developing countries: Evidence from benin. *Agricultural Economics*, 2022.

[15] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.