Journal of Engineering Research and Education
Volume 15, 2023 [27-36]

Enhancing Vegetable Price Forecasting Accuracy: A Hybrid SARIMA-DWT-GANN Approach

Kar Yan Choong[1], Suhizaz Sudin[1,2,*], Rafikha Aliana A. Raof[1,2] and Rhui Jaan Ong[1]

[1]Faculty of Electronic Engineering and Technology. Universiti Malaysia Perlis, Pauh Putra Campus, Malaysia
[2]Sports Engineering Research Center, Universiti Malaysia Perlis, Pauh Putra Campus, Malaysia

ABSTRACT

In the context of Malaysia's agricultural sector, the development of a robust vegetable price forecasting model holds paramount importance. The pricing dynamics of vegetables can significantly affect various stakeholders, including farmers, distributors, and consumers. The agricultural sector in Malaysia encounters persistent challenges such as supply-demand imbalances, seasonal variations, and market uncertainties, which can lead to income disparities for farmers and disruption in supply chains. Addressing these issues requires accurate and timely predictions of vegetable prices to enhance planning, resource allocation, and decision-making. This study introduces an innovative SARIMA-DWT-GANN hybrid model for vegetable price forecasting. By fusing the strengths of traditional time series modeling with the capabilities of neural networks, the proposed model offers a comprehensive solution to capture both linear and non-linear price patterns. The results demonstrate the superiority of the SARIMA-DWT-GANN model over the individual SARIMA model, as evident from correlation coefficients that closely approach unity and p-values confirming statistical significance. The model's ability to predict price changes has significant implications for making informed decisions throughout the agricultural supply chain. This research provides a robust forecasting tool that not only enhances market efficiency and profitability but also offers a promising solution to address the challenges in Malaysia's agricultural sector.

Keywords: Discrete Wavelet Transform, Forecasting, Genetic Algorithm, Neural Network, SARIMA

1. INTRODUCTION

The agriculture sector in Malaysia, predominantly centered around oil palm cultivation, faces multifaceted challenges. These challenges include the underemphasized broader agricultural domain, limited digitalization, an aging farming population, shifting consumer preferences, and vulnerabilities exposed by the COVID-19 pandemic, such as the disposal of vegetables due to supply chain disruptions [1], [2]. These issues highlight the need for innovative solutions to revitalize the sector.

Addressing these challenges, this research focuses on enhancing vegetable price forecasting accuracy. The proposed approach combines the power of Seasonal Autoregressive Integrated Moving Average (SARIMA) with the Discrete Wavelet Transform (DWT) and a Genetic Algorithm-based Neural Network (GANN). This hybrid methodology aims to overcome the limitations of SARIMA's linear forecasting by capturing complex non-linear components in price time series data.

*Corresponding author: suhizaz@unimap.edu.my

The importance of a vegetable price forecasting model in addressing agriculture-related problems is underscored. Such a model empowers stakeholders with insights for proactive planning, efficient resource allocation, and informed decision-making. SARIMA, a popular forecasting model, is highlighted for its ability to handle seasonal patterns but criticized for its limitation in addressing non-linear relationships [3].

To address these shortcomings, the research integrates DWT and GANN. The DWT dissects time series data into frequency components, revealing intricate patterns, while GANN optimizes neural network parameters using genetic algorithms for improved non-linear forecasting. The hybrid approach promises a comprehensive solution, enhancing accuracy and resilience in the agricultural sector.

Subsequent sections delve into the literature review, methodology, and results of the SARIMA-DWT-GANN approach. The research strives to contribute to the modernization of Malaysia's agriculture sector by embracing technology and innovation.

2. LITERATURE REVIEW

The literature review delves into a collection of studies focusing on agricultural price forecasting methods, specifically with an emphasis on vegetable prices. Researchers have employed various techniques to analyze and predict price behavior, aiming to assist stakeholders in making informed decisions.

Hernández, Puente, & Gómez (2019) used SARIMA for apple price behavior, achieving a 2% relative prediction error [4]. Mutwiri (2019) analyzed tomato prices in Kenya with SARIMA, finding it optimal for Nairobi County [5]. Vibas & Raqueño (2019) explored fruit and vegetable prices with ARIMA, SARIMA, and ANIMAX models [6]. Rathnayake, Razmy, & Alibuhtto (2019) used ARIMA and SARIMA for green chili and tomato price analysis [7]. Reddy (2018) proposed SARIMA for tomato price forecasting during harvest, revealing limitations [8].

Comparing forecasting methods, the review highlights the strengths and limitations of SARIMA, linear regression (LR), ARIMA, and exponential smoothing. While LR offers simplicity, it struggles with complex agricultural dynamics [9]–[11]. ARIMA excels in various economic predictions but faces challenges with agricultural seasonality [12]–[16]. Exponential smoothing captures trends and seasonality but falls short in intricate patterns [17]–[21]. SARIMA stands out due to its ability to handle seasonality and trends.

Studies showcase the potential of data mining techniques and neural networks, like back-propagation neural networks (BPNN), for accurate price predictions. Combining these approaches with SARIMA emerges as a solution to address both linear and non-linear aspects of time series data. Additionally, studies highlight the effectiveness of combining genetic algorithms (GA) with neural networks in various fields [22], [23]. The combination approach amplifies prediction accuracy by navigating complex solution spaces.

Combination forecasting methods gain traction, bridging mathematics, econometrics, and computer science for economic forecasting. The forecast of vegetable prices is crucial for both farmers and citizens. Incorporating combination methods, various research introduces innovative hybrid approaches. For instance, combining discrete wavelet transform (DWT) and artificial neural networks (ANN) enhances load forecasting accuracy [24]. Combining ARIMA with neural networks yields improved predictive accuracy [25]–[27]. Fusion of time series decomposition-based models and machine learning models enhances financial time series forecasting.

These studies collectively underscore the value of SARIMA, combination forecasting methods, and the potential of hybrid approaches to advance accurate vegetable price forecasting, addressing both linear and non-linear complexities in the agricultural context.

3. METHODOLOGY

The implementation of the proposed hybrid model for forecasting vegetable prices was accomplished by coding in Python within the Spyder development environment. Figure 1 visually depicts the hybrid forecasting methodology, illustrating the integration of SARIMA, DWT, and GANN models. This graphic outlines the roadmap for the methodology, integrated with its practical implementation. It captures the essence of the methodology, showcasing the synergy among techniques to enhance price forecasts in agriculture.

The methodology begins with the input data, which is the monthly vegetable price data from 2010 to 2021 provided by the Federal Agricultural Marketing Authority (FAMA) Malaysia. Next step followed with data preparation and preprocessing, where the vegetable price data is transformed into time series, inspected, and split into training and test sets. The hybrid forecasting method involves three main procedures: SARIMA modeling, DWT, and GANN. A SARIMA model is constructed for linear forecasting from 2019 to 2021, its steps involve identifying, estimating, diagnosing, and linear forecasting. The SARIMA model's residuals are then decomposed using DWT into sub-frequencies, which are input to the GANN model for non-linear forecasting. The GANN output, representing non-linear components, is reconstructed using inverse DWT. The final forecast is obtained by combining the linear forecast from SARIMA with the non-linear forecast from GANN.

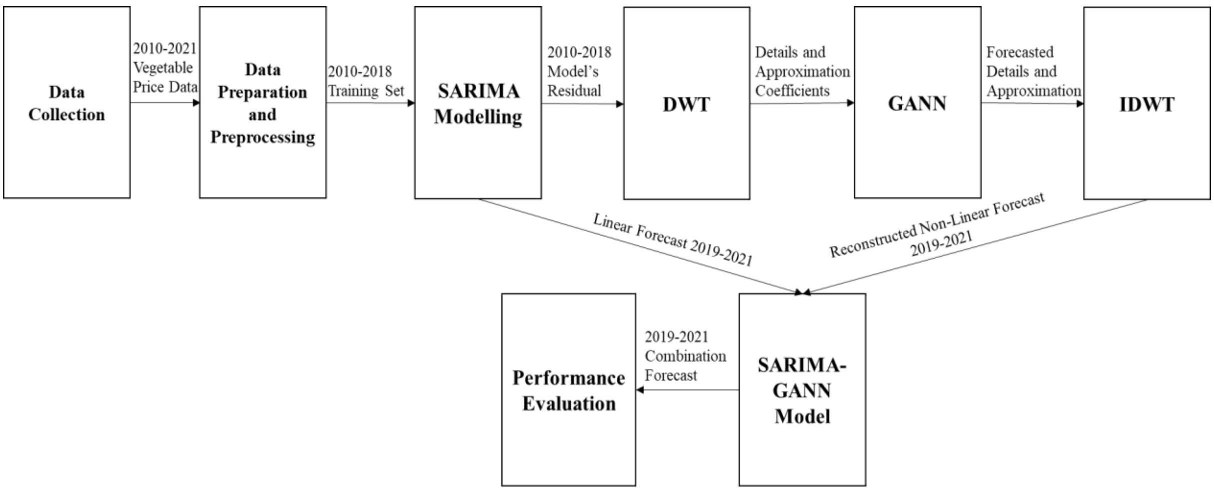


Figure 1 Block Diagram of the Proposed SARIMA-DWT-GANN Model

The forthcoming subsections will comprehensively delve into the intricate intricacies of executing the hybrid model. These sections will meticulously unfold each step, elucidating the sequence of operations involved in amalgamating the strengths of SARIMA, DWT, and GANN models.

3.1 Data Collection

The data collection process commenced with engagement efforts at the Malaysian Agricultural Research and Development Institute (MARDI) Sintok, where a proposal presentation was conducted in pursuit of a pertinent vegetable price dataset. However, MARDI's focus on research activities rendered them unable to provide the dataset.

Consequently, attention shifted to the Federal Agricultural Marketing Authority (FAMA) Kedah, resulting in a collaborative effort that led to the acquisition of a five-year dataset spanning 2015 to 2019. Despite its hard copy format, the dataset was transcribed into Excel. Yet, its limitations of only 60 observations became evident for forecasting.

Acknowledging the need for a more substantial dataset, the endeavor progressed to FAMA headquarters. Persistence paid off, securing access to a comprehensive dataset spanning 2010 to 2021 in *.xlsx* format. This acquisition proved transformative, providing a larger and contextually relevant dataset that invigorated subsequent analytical pursuits.

3.2 Data Preparation and Preprocessing

The process of preparing and pre-processing the data started by converting the monthly vegetable price data provided by FAMA Malaysia from .xlsx to .csv format. After importing the data into Python, a quick check for missing values and attributes was done using the *pandas* library. Some data was found to be incomplete for certain vegetable types between 2018 and 2021.

The galangal *(lengkuas)* with complete data set is chosen to ensure reliability. The time format was adjusted from integers to a time series format to enable time-based analysis. This data was then plotted using *matplotlib* to visualize price trends and check if the data showed any patterns. The dataset covered 144 observations spanning 12 years, which were divided into a training set of 108 observations and a test set of 36 observations for SARIMA modeling.

3.3 Seasonal Autoregressive Integrated Moving Average (SARIMA) Modeling

The SARIMA model involves several steps to predict future outcomes using time series data. It begins with identifying SARIMA parameters (*p, d, q, P, D, Q, m*) as shown in Figure 2 and estimating them. Then, the model's adequacy is checked through diagnosis. Next, the model is utilized for forecasting, which involves converting non-stationary data to stationary, testing stationarity, and potentially applying differencing.

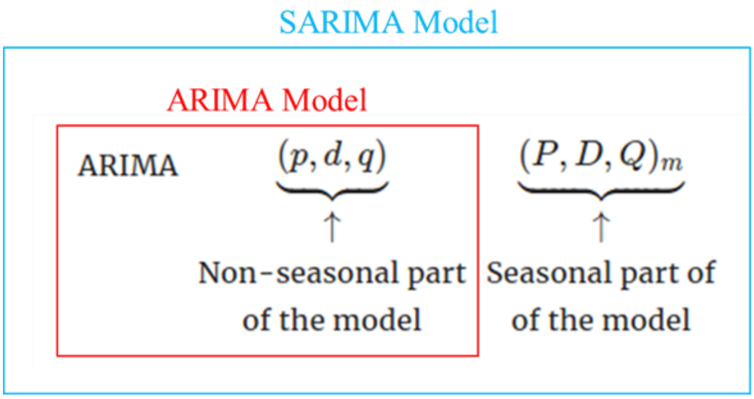


Figure 2 Notation of SARIMA Model

ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots help determine component orders while grid search is used for parameter optimization with the model evaluation relying on AIC (Akaike Information Criterion) values. Once the SARIMA model is selected, it is fitted with training data and diagnosed. The diagnostic plots assess the residuals of the best-fitted SARIMA model. The chosen SARIMA model is then used for forecasting future outcomes in the test set linearly. This forecasting process ensures accurate predictions for further analysis.

3.4 Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) decomposes SARIMA residual errors into high and low-frequency components, which are "details" and "approximation" coefficients, using the "Haar" wavelet filter. The choice of decomposition level, influenced by data size, balances information extraction and data integrity. A decomposition level of 2 is selected to retain insight without excessive smoothing.

Components like A_2, D_1, and D_2 are obtained for level 2 using *wavedec()*, then fed into a neural network with a genetic algorithm for robust forecasting, effectively integrating wavelet-transformed SARIMA residuals into the process.

3.5 Genetic Algorithm based Neural Network (GANN) Modeling

The steps of operation of the modified GANN for this research are illustrated in Figure 3 starting with the initialization of the population, construction of recurrent neural network (RNN), fitness evaluation, selection, crossover, mutation, replacement, and stopping criterion.

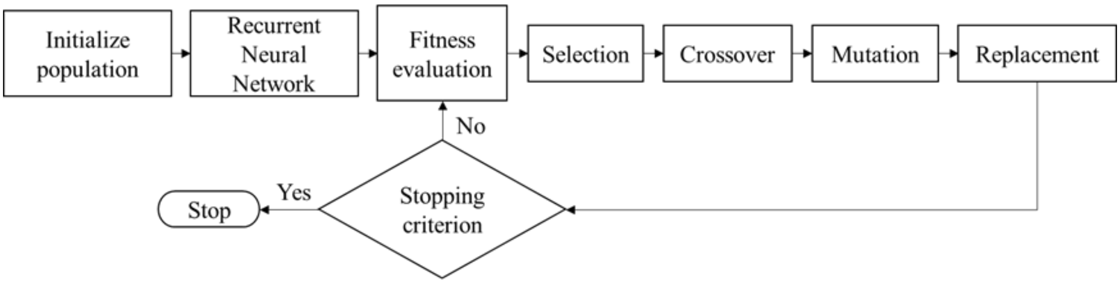


Figure 3 Genetic Algorithm based Neural Network (GANN) Model

The steps of operation for the GANN model as illustrated in Figure 3 in this research, can be summarized as follows:

1. **Start:** Begin with initializing a population of potential solutions and set it to 20, where each solution is represented by weight and bias matrices for a neural network.
2. **Build RNN:** Create a neural network with a recurrent structure (RNN). This RNN considers a certain number of previous data points to capture patterns in the time series data.
3. **Evaluate Fitness:** Measure the fitness of each solution by comparing RNN predictions to actual data. A lower mean squared error (MSE) indicates better fitness.
4. **Select Parents:** Use a roulette wheel method to pick individuals from the population as potential parents for creating new solutions.
5. **Combine Parents:** Combine pairs of parents using arithmetic crossover, creating new offspring with combined genetic information.
6. **Introduce Variation:** Apply mutation to a couple of the offspring, adding a touch of randomness to explore new solutions.
7. **Create a Mating Pool:** Combine offspring and parents to form a mating pool for the next generation.
8. **Stop if Overfitting:** If the validation error increases for three consecutive generations, stop to prevent overfitting.
9. **Best Model:** Select the best-performing model from the generation with the lowest validation error.
10. **Non-linear forecast with Test Data:** Do non-linear forecasting for the test set

3.6 Inverse Discrete Wavelet Transform (IDWT)

The forecasted detail (D_1, D_2) coefficients from the GANN model are combined with the forecasted approximation (A_2) coefficients to reconstruct the forecasted time series using the IDWT. The details add back the high-frequency variations, while the approximation restores the overall trend. The IDWT process yields a reconstructed time series that represents the non-linear forecast of the original data. This final forecast integrates the predictions for both high-frequency variations and overall trends, providing a comprehensive view of the future behavior of the time series.

3.7 SARIMA-GANN Model

By adding the linear forecast from SARIMA to the non-linear forecast from the GANN model, its output is the final forecast of the vegetable price time series. After that, the *matplotlib* library is then used to plot the actual and forecasted vegetable prices from the test set, visually evaluating the model's performance. The next step involves using accuracy metrics and statistical tests for a thorough performance evaluation of the hybrid model.

3.8 Performance Evaluation

In the evaluation phase, the hybrid SARIMA-GANN model is assessed using accuracy measures like *MAE*, *MAPE*, *RMSE*, and the *Pearson correlation coefficient* (r). Its performance is compared to the SARIMA individual model. The goal is to see how well the hybrid model improves forecasting accuracy compared to the baseline SARIMA, considering seasonality.

The evaluation involves formulas for metrics like r, *MAE*, *MAPE*, and *RMSE*, which help gauge the accuracy of the forecasts as follows:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2}$$

$$MAPE\% = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

The n denotes the number of observations to be forecasted, the $\bar{y}$ signifies the mean of the actual values (y_i), and the $\bar{\hat{y}}$ indicates the mean of the predicted values ($\hat{y}$). To determine if one model significantly outperforms another, the Diebold-Mariano (DM) test is used. The DM test hypothesis assesses whether one forecasting model significantly outperforms another in terms of accuracy.

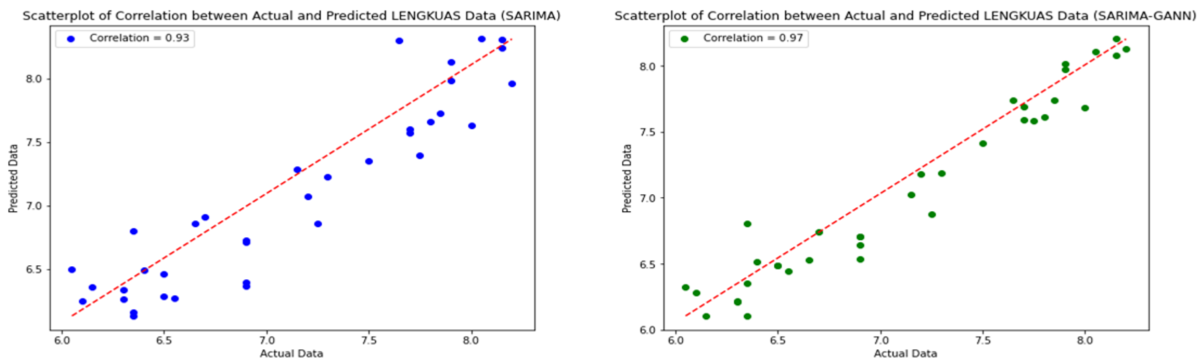
4. RESULTS

The results analysis tabulated in Table 1 reveals that the SARIMA-DWT-GANN hybrid model exhibited the highest correlation coefficients compared to the individual SARIMA model, indicating a strong linear relationship between predicted and observed values. Scatterplots demonstrated a robust positive linear relationship between actual and predicted values for the galangal (*lengkuas*) price using SARIMA and SARIMA-GANN models, with respective correlation coefficients of 0.93 and 0.97 as illustrated in Figure 4.

The proposed SARIMA-DWT-GANN model showcased superior performance in terms of Mean Absolute Percentage Error (MAPE), achieving a low MAPE of 1.9632% for galangal (lengkuas) tabulated in Table 1, implying a high forecasting accuracy of 98.04%. Additionally, both Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics yielded small values close to zero, reflecting the proposed model's effective forecasting performance.

Table 1 Accuracy Metrics of Galangal *(Lengkuas)* using SARIMA and SARIMA-DWT-GANN

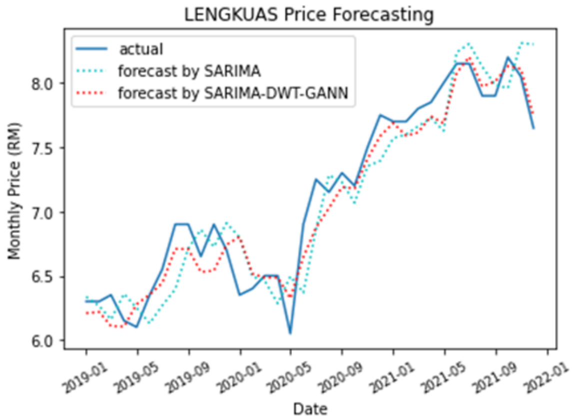
Accuracy Metrics	**SARIMA**	**SARIMA-DWT-GANN**
Pearson Correlation Coefficients r (p-value)	0.9293 (2.8230E-16)	0.9708 (1.1573E-22)
MAE	0.2219	0.1368
MAPE	3.1496	1.9632
RMSE	0.2669	0.175



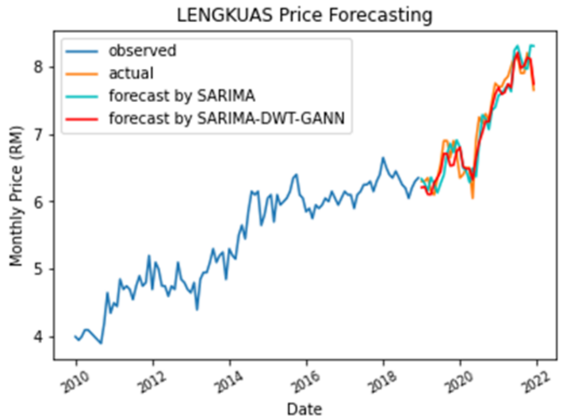
(a) SARIMA (b) SARIMA-DWT-GANN

Figure 4 Pearson Correlation Coefficients of Galangal (Lengkuas) using SARIMA and SARIMA-DWT-GANN

Statistically significant p-values (below 0.05) which were recorded as 0.0049 and a negative Diebold-Mariano (DM) test statistic of -2.8105 for galangal (*lengkuas*) further validated the superiority of the SARIMA-DWT-GANN model in comparison to the SARIMA. The hybrid approach leveraged the SARIMA model's linear forecasting and the DWT-GANN model's non-linear forecasting to achieve accurate and comprehensive forecasts, as evidenced by the final forecast plot and trend alignment in Figure 5.



(a) Plot without training set



(b) Plot with training set

Figure 5 Forecasted Galangal *(Lengkuas)* using SARIMA and SARIMA-DWT-GANN

5. CONCLUSION

In conclusion, the study introduced a new SARIMA-DWT-GANN hybrid approach for forecasting vegetable prices. The results showcased that this hybrid model performed remarkably better compared to standalone SARIMA model. The SARIMA-DWT-GANN model demonstrated a strong ability to capture both linear and non-linear relationships, leading to higher correlation coefficients and lower forecasting errors. Notably, it excelled in forecasting the galangal (lengkuas) price with an accuracy of 98.04%. The model's hybrid nature was its strength, as it effectively combined the strengths of SARIMA's linear forecasting and GANN's non-linear prediction. This research underscores the potential of the SARIMA-GANN approach in accurate time series forecasting across various domains.

ACKNOWLEDGEMENTS

The authors extend their heartfelt appreciation to the Federal Agricultural Marketing Authority (FAMA) Malaysia for generously providing the historical monthly vegetable price data. This research received fund support from a grant awarded under the Fundamental Research Grant Scheme (FRGS/1/2020/ICT06/UNIMAP/02/1) by the Ministry of Higher Education Malaysia.

REFERENCES

[1] K. Waiho *et al.*, "Potential impacts of COVID-19 on the aquaculture sector of Malaysia and its coping strategies," *Aquac Rep*, vol. 18, p. 100450, 2020, doi: https://doi.org/10.1016/j.aqrep.2020.100450.

[2] M. F. Nasir and M. A. Jamlos, "Smart Farming Monitoring Towards IR 4.0 Implementation," *Journal of Engineering Research and Education*, vol. 14, pp. 97–103, 2022.

[3] T. Zhang, Y. Wu, Y. Chen, T. Li, and X. Ren, "Collaborative Energy Price Computing Based on Sarima-Ann and Asymmetric Stackelberg Games," *Symmetry (Basel)*, vol. 15, no. 2, p. 443, Feb. 2023, doi: 10.3390/sym15020443.

[4] J. Hernández, G. Puente, and A. Gómez, "Analysis of the price of the apple using a SARIMA model," *Rev Mex De Cienc Agric*, vol. 10, pp. 225–237, 2019, doi: 10.29312/remexca.v10i2.509.

[5] R. M. Mutwiri, "Forecasting of Tomatoes Wholesale Prices of Nairobi in Kenya : Time Series Analysis Using Sarima Model," *International Journal of Statistical Distributions and Applications*, vol. 5, no. 3, pp. 46–53, 2019, doi: 10.11648/j.ijsd.20190503.11.

[6] V. M. Vibas and A. R. Raqueño, "A Mathematical Model for Estimating Retail Price Movements of Basic Fruit and Vegetable Commodities Using Time Series Analysis," *International Journal of Advance Study and Research Work*, vol. 2, 2019, doi: 10.5281/zenodo.3333529.

[7] R. M. Y. L. Rathnayake, A. M. Razmy, and M. C. Alibuhtto, "A Comparison of Price Fluctuations for the Green Chilli and Tomato in Dambulla Dedicated," in *5th Annual Science Research Sessions, South Eastern university of Sri Lanka*, 2019.

[8] A. A. Reddy, "Price Forecasting of Tomatoes," *International Journal of Vegetable Science*, vol. 25, pp. 1–9, Jul. 2018, doi: 10.1080/19315260.2018.1495674.

[9] W. Xu, H. Peng, X. Zeng, F. Zhou, X. Tian, and X. Peng, "A hybrid modelling method for time series forecasting based on a linear regression model and deep learning," *Applied Intelligence*, vol. 49, no. 8, pp. 3002–3015, 2019, doi: 10.1007/s10489-019-01426-3.

[10] Y. Xiao and Z. Jin, "The Forecast Research of Linear Regression Forecast Model in National Economy," *OAlib*, vol. 08, no. 08, pp. 1–17, 2021, doi: 10.4236/oalib.1107797.

[11] A. I. Harsapranata, D. Manongga, and S. Wijono, "LINER REGRESSION ANALYSIS IN FORECASTING HYDROPONIC PLANT NUTRIENT NEEDS," *INFOKUM*, vol. 10, no. 5, pp. 222–226, 2022, Accessed: Feb. 08, 2023. [Online]. Available: http://infor.seaninstitute.org/index.php/infokum/article/view/890

[12] M. R. Abonazel and A. I. Abd-Elftah, "Forecasting Egyptian GDP using ARIMA models," *Reports on Economics and Finance*, vol. 5, no. 1, pp. 35–47, 2019, doi: 10.12988/ref.2019.81023.

[13] M. Mustapa, R. R. Ponnusamy, and H. M. Kang, "Forecasting prices of fish and vegetable using web scraped price micro data," *International Journal of Recent Technology and Engineering*, vol. 7, no. 5, pp. 251–256, 2019.

[14] N. Novkovic, L. Drinic, S. Mihajlovic, N. Vukelic, and D. Ivanisevic, "Price Parities For Vegetables in Serbia - Analysis and Forecasting," *Serbian Journal of Agricultural Sciences*, vol. 68, no. 3–4, pp. 51–59, 2019, doi: 10.2478/contagri-2019-0009.

[15] V. Priyanga, T. P. Lazarus, S. Mathew, and B. Joseph, "Forecasting coconut oil price using auto regressive integrated moving average (ARIMA) model," vol. 8, no. 3, pp. 2164–2169, 2019.

[16] K. Islam and A. Raza, "Forecasting Crime Using ARIMA Model." 2020.

[17] N. Nurhamidah, N. Nusyirwan, and A. Faisol, "FORECASTING SEASONAL TIME SERIES DATA USING THE HOLT-WINTERS EXPONENTIAL SMOOTHING METHOD OF ADDITIVE MODELS," *Jurnal Matematika Integratif*, vol. 16, no. 2, p. 151, Dec. 2020, doi: 10.24198/jmi.v16.n2.29293.151-157.

[18] V. Shah, "A comparative study of univariate time-series methods for sales forecasting," *International Journal of Business and Data Analytics*, 2020.

[19] S. Lima, A. M. Gonçalves, and M. Costa, "Time series forecasting using Holt-Winters exponential smoothing: An application to economic data," *AIP Conf Proc*, vol. 2186, no. December, 2019, doi: 10.1063/1.5137999.

[20] R. Rossetti, "Forecasting the sales of console games for the Italian market," *Econometrics*, vol. 23, no. 3, pp. 76–88, 2019, doi: 10.15611/eada.2019.3.07.

[21] G. A. N. Pongdatu and Y. H. Putra, "Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing," *IOP Conf Ser Mater Sci Eng*, vol. 407, no. 1, 2018, doi: 10.1088/1757-899X/407/1/012153.

[22] X. Wang and Y. Li, "Facial Recognition System Based on Genetic Algorithm Improved ROI-KNN Convolutional Neural Network," *Appl Bionics Biomech*, vol. 2022, p. 7976856, 2022, doi: 10.1155/2022/7976856.

[23] Y. L. Liu, E. C. Nisa, Y. Der Kuan, W. J. Luo, and C. C. Feng, "Combining Deep Neural Network with Genetic Algorithm for Axial Flow Fan Design and Development," *Processes*, vol. 11, no. 1, Jan. 2023, doi: 10.3390/pr11010122.

[24] M. A. Awal and M. M. Abdullah Al Mamun, "Long Term Meteorological Drought Forecasting for North-western Region of Bangladesh Using Wavelet Artificial Neural Network," *Revista Brasileira de Meteorologia*, vol. 37, no. 4, pp. 453–465, 2022, doi: 10.1590/0102-77863740058.

[25] X. Zhang and R. Li, "A Novel Decomposition and Combination Technique for Forecasting Monthly Electricity Consumption," *Front Energy Res*, vol. 9, Dec. 2021, doi: 10.3389/fenrg.2021.792358.

[26] E. Yao, L. Zhang, X. Li, and X. Yun, "Traffic Forecasting of Back Servers Based on ARIMA-LSTM-CF Hybrid Model," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, Dec. 2023, doi: 10.1007/s44196-023-00232-7.

[27] J. Duan, Y. Gong, J. Luo, and Z. Zhao, "Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-36620-4.