14th ISCA 2024

Tea Price Prediction Using Hybrid Arima-BP Neural Network Model

Deden Supena[1*], Dian Purnomo Jati[2], Ekaningtyas Widiastuti[3]

[1*] University of Jenderal Soedirman, deden.supena@mhs.unsoed.ac.id, Indonesia
[2] University of Jenderal Soedirman, dian.jati@unsoed.ac.id, Indonesia
[3] University of Jenderal Soedirman, ekaningtyas.widiastuti@unsoed.ac.id, Indonesia
*corresponding author

ABSTRACT

This study examines the effectiveness of a hybrid ARIMA-BP Neural Network model in predicting global tea prices. The model integrates ARIMA to capture linear trends and Back Propagation Neural Network (BP NN) to address non-linear patterns, combining their strengths for improved forecasting accuracy. Using monthly tea price data from 2015 to 2022 obtained from Index Mundi, the study evaluates the hybrid model's performance against standalone ARIMA. Stationarity of the data was achieved through first-order differencing, and model selection was based on Akaike Information Criterion (AIC) and diagnostic checks. The hybrid model demonstrated superior predictive accuracy, achieving a Mean Absolute Percentage Error (MAPE) of 6.32%, compared to 12.79% for ARIMA alone. These results underscore the potential of hybrid models for volatile commodity markets, offering practical implications for risk management and decision-making in the tea industry. Stakeholders can leverage this model to anticipate price fluctuations, optimize operations, and enhance financial planning. Future research could explore hybrid approaches in other commodities and incorporate additional predictive variables.

Keywords: tea price prediction, ARIMA, back propagation neural network, hybrid model, forecasting, commodity markets.

---

1. Introduction

The global increase in tea consumption has positioned tea as a commodity with significant economic value (Wang et al., 2022). The demand is fueled by its unique flavor and widely recognized health benefits, such as antioxidant properties and disease prevention potential (Xiang et al., 2024). During the COVID-19 pandemic, tea gained popularity as a healthier alternative to other caffeinated beverages, aiming to boost immunity and overall health (Castellana et al., 2021). Innovations like mass-produced, convenient tea products have further expanded its global reach (Sinaga et al., 2023).

In 2023, the global tea market was valued at $276.3 billion, with over 7 billion tons produced and exports totaling $8.13 billion (Eximpedia, 2023). However, price volatility remains a critical challenge for stakeholders in the tea industry, impacting financial outcomes and operational planning (Siagian et al., 2023). Predicting tea prices has become essential for effective financial and strategic management.

This study explores the effectiveness of a hybrid prediction model combining ARIMA and Back Propagation Neural Network (BP NN) to address the limitations of linear and non-linear forecasting methods individually. The hybrid model leverages ARIMA for linear patterns and BP NN for non-linear trends, offering an innovative approach to enhance prediction accuracy.

2. Literature Review

2.1 Signaling Theory

Signaling theory provides insights into market dynamics where asymmetric information affects decision-making (Spence, 1973). In the context of tea prices, signaling explains how producers and consumers interpret price changes, influencing supply and demand dynamics. Studies have shown that price fluctuations often serve as signals about market conditions, which can lead to changes in production strategies or consumer behavior (Chen & Zhang, 2019).

2.2 Price Theory

Price theory examines factors influencing commodity prices, including supply, demand, and market behavior (Friedman, 1990). For tea, fluctuations arise from climatic conditions, production output, and global trade policies (Ahmed et al., 2014). Additionally, quality differentiation and consumer preferences play a significant role in shaping price trends (Yang et al., 2015).

2.3 Predictive Models

ARIMA models are effective for capturing linear patterns in time-series data, while BP Neural Networks excel at identifying complex non-linear relationships. Combining these methods into a hybrid model provides a more comprehensive tool for accurate forecasting. Previous research has demonstrated the hybrid model's utility in various domains, including agricultural commodities and financial markets (Dou et al., 2022; Visakha, 2023).

3. Research Methodology

3.1 Research Design

This study employs a mixed-method approach, integrating quantitative predictions with qualitative validation through stakeholder interviews. Historical monthly tea price data (2015-2022) from Index Mundi served as the primary dataset. The quantitative analysis focused on forecasting accuracy, while the qualitative component explored industry relevance and practical implications.

3.2 Data Characteristics

The dataset included monthly global tea prices from January 2015 to December 2022. An initial exploration revealed patterns of seasonality and trends, necessitating stationarity tests using the Augmented Dickey-Fuller method.

3.3 Hybrid Model Development

3.3.1 ARIMA Modeling:

- Identification of optimal parameters (p, d, q) based on Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots.
- Selection of the best-fit model using Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC).

3.3.2 BP Neural Network:

- Residuals from ARIMA predictions served as input data.
- Network structure included one hidden layer, optimized for minimal prediction error.

3.3.3 Integration:

- Final predictions combined ARIMA forecasts and BP NN residual adjustments using the formula:

  $yt = Lt + Nt$

  Where Lt represents ARIMA predictions and Nt represents BP NN residual adjustments.

3.4 Evaluation Metrics

Prediction accuracy was assessed using Mean Absolute Percentage Error (MAPE):

$$MAPE = \sum \frac{\frac{|Actual\ Forecast|}{Actual} \times 100\%}{n}$$

4. Results

4.1 Data Characteristics

Time-series analysis indicated the presence of both trends and seasonality. Stationarity tests confirmed the necessity of first-order differencing to stabilize the series.

Null Hypothesis: TEA has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.477857	0.5378
Test critical values:	1% level	-3.546099	
	5% level	-2.911730	
	10% level	-2.593551	

*MacKinnon (1996) one-sided p-values.

Figure 1. Augmented Dickey-Fuller method

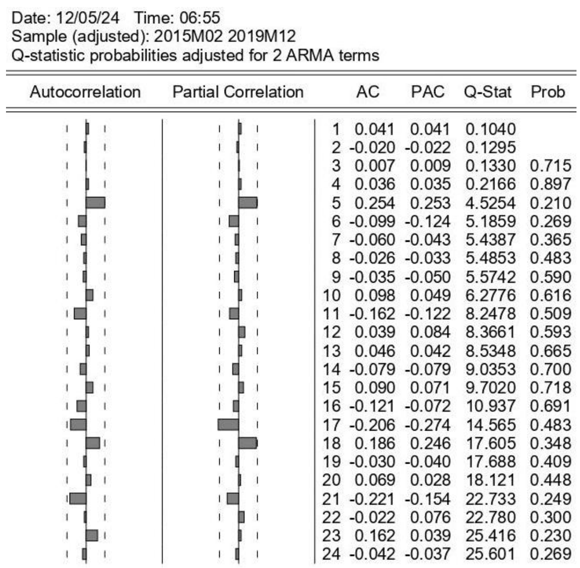
4.2 ARIMA Model

The optimal model, ARIMA(10,1,9), was selected based on the lowest AIC and SBC values. Diagnostic checks confirmed the absence of autocorrelation in residuals, validating the model's suitability.

Table 1. ARIMA Model Parameters

Model	AIC	SIC	Significance
ARIMA (9,1,9)	-0.977489	-0.836639	

ARIMA (9,1,10)	-1.141287	-1.000437	
ARIMA (10,1,9)	-1.189471	-1.048621	Significant


Date: 12/05/24 Time: 06:55
Sample (adjusted): 2015M02 2019M12
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.041	0.041	0.1040	
		2	-0.020	-0.022	0.1295	
		3	0.007	0.009	0.1330	0.715
		4	0.036	0.035	0.2166	0.897
		5	0.254	0.253	4.5254	0.210
		6	-0.099	-0.124	5.1859	0.269
		7	-0.060	-0.043	5.4387	0.365
		8	-0.026	-0.033	5.4853	0.483
		9	-0.035	-0.050	5.5742	0.590
		10	0.098	0.049	6.2776	0.616
		11	-0.162	-0.122	8.2478	0.509
		12	0.039	0.084	8.3661	0.593
		13	0.046	0.042	8.5348	0.665
		14	-0.079	-0.079	9.0353	0.700
		15	0.090	0.071	9.7020	0.718
		16	-0.121	-0.072	10.937	0.691
		17	-0.206	-0.274	14.565	0.483
		18	0.186	0.246	17.605	0.348
		19	-0.030	-0.040	17.688	0.409
		20	0.069	0.028	18.121	0.448
		21	-0.221	-0.154	22.733	0.249
		22	-0.022	0.076	22.780	0.300
		23	0.162	0.039	25.416	0.230
		24	-0.042	-0.037	25.601	0.269

Figure 2. ARIMA(10,1,9) Residual Randomness Diagnostic Test

4.3 BP Neural Network

Training the BP NN on ARIMA residuals improved accuracy by capturing non-linear patterns. The network achieved convergence with a learning rate of 0.01 and 500 epochs.

4.4 Hybrid Model Performance

The hybrid model significantly outperformed standalone ARIMA, reducing the MAPE from 12.79% to 6.32%. This improvement highlights the synergy between linear and non-linear modeling approaches.

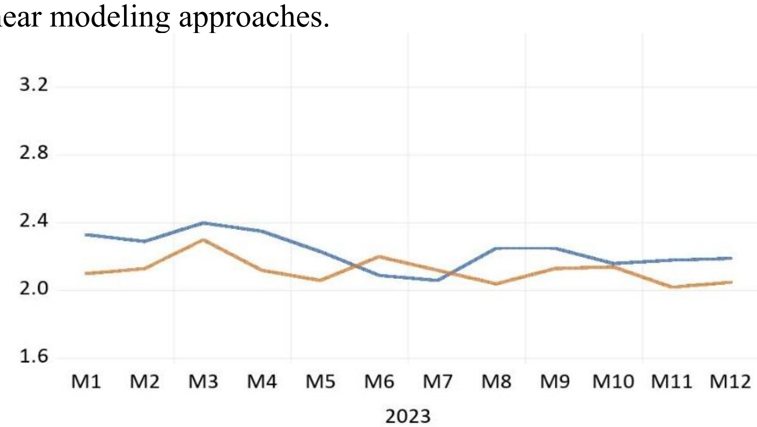


Figure 3. Actual Price Plot with Hybrid Prediction

5. Discussion

The findings underscore the hybrid model's potential for addressing complex forecasting challenges. By combining ARIMA and BP NN, this approach leverages the strengths of both methods, making it particularly effective for volatile markets like tea.Practical implications include better risk management and strategic planning for stakeholders. For instance, accurate price forecasts can inform inventory management, pricing strategies, and investment decisions.

6. Conclusion

This study confirms the effectiveness of the Hybrid ARIMA-BP NN model in predicting tea prices with higher accuracy. The model offers a valuable tool for industry stakeholders to mitigate financial risks and improve strategic planning. Future research could explore hybrid applications in other commodities and incorporate additional variables for enhanced accuracy.

References

Adlini, M. N., Dinda, A. H., Yulinda, S., Chotimah, O., & Merliyana, S. J. (2022). Metode Penelitian Kualitatif Studi Pustaka. *Edumaspul: Jurnal Pendidikan*, *6*(1), 974–980. https://doi.org/10.33487/edumaspul.v6i1.3394

Ahmed, S., Stepp, J. R., Orians, C., Griffin, T., Matyas, C., Robbat, A., Cash, S., Xue, D., Long, C., & Unachukwu, U. (2014). Effects of extreme climate events on tea (Camellia sinensis) functional quality validate indigenous farmer knowledge and sensory preferences in tropical China. *PloS One*, *9*(10), e109126.

Ardheta, P. A., & Sina, H. R. (2020). Pengaruh Capital Adequacy Ratio, Dana Pihak Ketiga, Non Performing Financing dan Pembiayaan Murabahah Terhadap Profitabilitas. *Jurnal Akuntansi Dan Manajemen*, *17*(02), 32–38. https://doi.org/10.36406/jam.v17i02.328

Balova, S., Firsova, I., & Osipova, I. (2019). *Market leading marketing concepts in the management of consumer behaviour on the energy market*. *318*(Icseal), 354–360. https://doi.org/10.2991/icseal-19.2019.55

Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025–1044. https://doi.org/10.1287/mnsc.2018.3253

Bloch, H. (2020). Price Theory, Historically Considered: Smith, Ricardo, Marshall and Beyond. *History of Economics Review*, *75*(1), 50–73. https://doi.org/10.1080/10370196.2020.1745439

Browning, E. K., & Zupan, M. A. (2020). *Microeconomics: Theory and applications*. John Wiley & Sons.

Burtnyak, I., & Rusyn, R. (2023). *Удк 519.86:336.76 Іван Буртняк*. 55–60.

Candra, S. F., & Irmeilyana, I. (2024). Model Regresi Data Panel pada Pengaruh Faktor Curah Hujan Terhadap Produksi Kopi di Provinsi Sumatera Selatan Tahun 2014-2021. *Jurnal Penelitian Sains*, *26*(1), 30–39.

Castellana, F., De Nucci, S., De Pergola, G., Di Chito, M., Lisco, G., Triggiani, V., Sardone, R., & Zupo, R. (2021). Trends in coffee and tea consumption during the covid-19 pandemic. *Foods*, *10*(10), 1–10. https://doi.org/10.3390/foods10102458

Che, C., Huang, Z., Li, C., Zheng, H., & Tian, X. (2024). Integrating generative AI into financial market prediction for improved decision making. *Applied and Computational Engineering*, *64*(1), 155–161. https://doi.org/10.54254/2755-2721/64/20241376