Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Forecasting Rice Prices Using the ARIMA Method: A Case Study in DKI Jakarta Province

Belsana Butar Butar[1], Aulya Giffari[2], Zhafirah Dwi Putri[3], Melisa Karisma[4], Wawan Kurniawan[5], M. Haddiel Fuad[6]
Information System, Bina Sarana Informatika University. Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia 10450

Article Info	ABSTRACT
Keywords: ARIMA, Python, Forecasting Rice Prices, DKI Jakarta, Error Metrics, Grid Search	Rice is a food structure with very important nutrients and is very much consumed by the people of Indonesia. The availability of rice in the country must be fulfilled because rice is a strategic and political food commodity. DKI Jakarta Province as the economic centre and capital city of the country has a strategic role in advancing the national economy, the movement of food commodity prices such as rice is often the centre of attention compared to other regions. Various factors can have an impact on the stability of rice prices and stocks, resulting in fluctuations in rice prices every month. For this reason, a programme is needed that can forecast the staple price of rice to illustrate the problem of food price instability in the future. In this research, we created a rice price forecasting programme using the Python programming language and Jupyter Notebook IDE by applying the ARIMA (Autoregressive Integrated Moving Average) time series method. The data used is a single data set containing rice prices in the period 2018 - 2023 with monthly granularity. The results of rice price forecasting are measured through 4 error metrics namely MSE, RMSE, MAE, and MAPE which show that the results are accurate with low error interpretation.
This is an open access article under the CC BY-NC license	Corresponding Author: Belsana Butar Butar Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia 10450. belsana.bbb@bsi.ac.id

INTRODUCTION

The fluctuation of food commodity prices, particularly rice in Jakarta, often attracts significant attention compared to other regions in Indonesia. As the capital city and economic hub, Jakarta serves as a reference point for national rice price trends. Any changes in rice prices in Jakarta can have a broad impact, influencing perceptions and market behavior in other regions. Given that rice is a staple food for the majority of Indonesians, even slight price fluctuations can directly affect household expenditures and economic stability.

According to data from the National Food Agency (https://badanpangan.go.id/), as of April 2024, the national average price of rice was recorded at fifteen thousand rupiahs per kilogram. In contrast, during the period from January to June 2024, the highest average price of premium rice in DKI Jakarta reached seventeen thousand rupiahs per kilogram (Fadhlurrahman, 2024). This indicates a significant price increase, which raises concerns among consumers and policymakers alike. The rising cost of rice can be attributed to various factors, including supply chain disruptions, inflation, seasonal demand fluctuations, and government policies on food security.

Despite the sharp increase in rice prices, consumers have no choice but to continue purchasing it, as rice is an essential part of the daily diet. Unlike other commodities, the demand for rice remains relatively inelastic, meaning that even with rising prices, consumption levels do not decrease significantly. This situation creates economic pressure on lower-income households, which may struggle to afford sufficient food supplies. As a result, price volatility in the rice market can contribute to broader socio-economic challenges, including food insecurity and decreased purchasing power.

Understanding and forecasting rice price movements are crucial to ensuring food affordability and market stability. By analyzing historical price trends, it is possible to predict future price fluctuations, allowing stakeholders to make informed decisions. Accurate price forecasting can benefit policymakers, traders, and consumers by providing insights into expected market conditions. It can also help the government implement effective interventions, such as price stabilization programs, subsidies, or import regulations, to protect consumers from excessive price spikes.

One of the most effective methods for forecasting time-series data, such as rice prices, is the ARIMA (AutoRegressive Integrated Moving Average) model. This statistical approach is widely used in economic and financial forecasting due to its ability to analyze past trends and predict future values. The ARIMA method considers historical price data to identify patterns and relationships, enabling a more accurate estimation of future price movements. By applying ARIMA to rice price forecasting, stakeholders can gain a clearer picture of potential price changes in the coming months or years.

The implementation of a forecasting system using ARIMA can provide early warnings about potential price hikes or declines. This information allows businesses and policymakers to take proactive measures, such as adjusting supply chain strategies, managing inventory levels, or implementing price control policies. Additionally, an effective forecasting model can support agricultural planning by helping farmers anticipate market demand and adjust production accordingly. This can contribute to a more stable and sustainable food supply chain.

Given the critical role of rice in Indonesian society, developing a reliable price forecasting model is essential for addressing food price instability. The integration of technology in price prediction can enhance market transparency and efficiency, reducing the uncertainty faced by consumers and businesses. Moreover, digital tools can make price forecasting accessible to a wider audience, including small-scale traders, farmers, and policymakers, who can use the data to make strategic decisions.

The rising rice prices in Jakarta and across Indonesia highlight the need for an accurate and effective forecasting system. By leveraging statistical methods such as ARIMA, it is possible to identify trends, anticipate price changes, and develop strategies to mitigate economic impacts. A well-implemented forecasting model can provide valuable insights for

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

policymakers and consumers, helping to ensure food security and market stability in the future.

METHODS

This study begins by identifying economic issues, with a primary focus on rice prices as a fundamental necessity for society. A comprehensive literature review follows, incorporating academic journal articles, books, statistical reports, and other credible sources to establish a strong theoretical foundation and identify key concepts relevant to the research. The data utilized in this study consists of historical rice price records sourced from the official website of the Central Bureau of Statistics (BPS) for DKI Jakarta Province, covering the period from 2018 to 2023. This dataset falls under the category of secondary data.

After collecting the data, it is compiled into a single file using Microsoft Excel, then converted into CSV format before being uploaded to a Jupyter Notebook repository. The subsequent stage involves data processing, which includes installing the necessary Python libraries in Jupyter Notebook, verifying data types, conducting seasonal analysis, and assessing stationarity using the adfuller function.

The ARIMA model is applied through several stages, beginning with model identification to determine the AR(p), I(d), and MA(q) parameters. This process utilizes the auto_arima function with the Grid Search technique. Additionally, residual autocorrelation tests are performed, supported by visualizations of the Autocorrelation Function (ACF; q) and Partial Autocorrelation Function (PACF; p) plots. To forecast rice prices for the years 2024 to 2026, the accuracy of predictions is assessed using four key error metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

RESULTS AND DISCUSSION

Data Preparation and Collection

The initial stage involves preparing the raw data that will be analyzed.

Bulan *Month*	Setra
(1)	(5)
Januari/*January*	13.302,99
Pebruari/*February*	13.369,11
Maret/*March*	13.362,84
April/*April*	13.365,65
Mei/*May*	13.361,27
Juni/*June*	13.234,29
Juli/*July*	13.234,29
Agustus/*August*	13.240,35
September/*September*	13.238,98
Oktober/*October*	13.227,57
Nopember/*November*	13.226,07
Desember/*December*	13.220,07
2019	**13.281,96**

Figure 1. Table of Retail Rice Prices in DKI Jakarta

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

The data we collected from the annual publication documents on the official BPS Jakarta website (https://jakarta.bps.go.id/publication/) was converted into CSV format and then uploaded to the Jupyter Notebook repository.

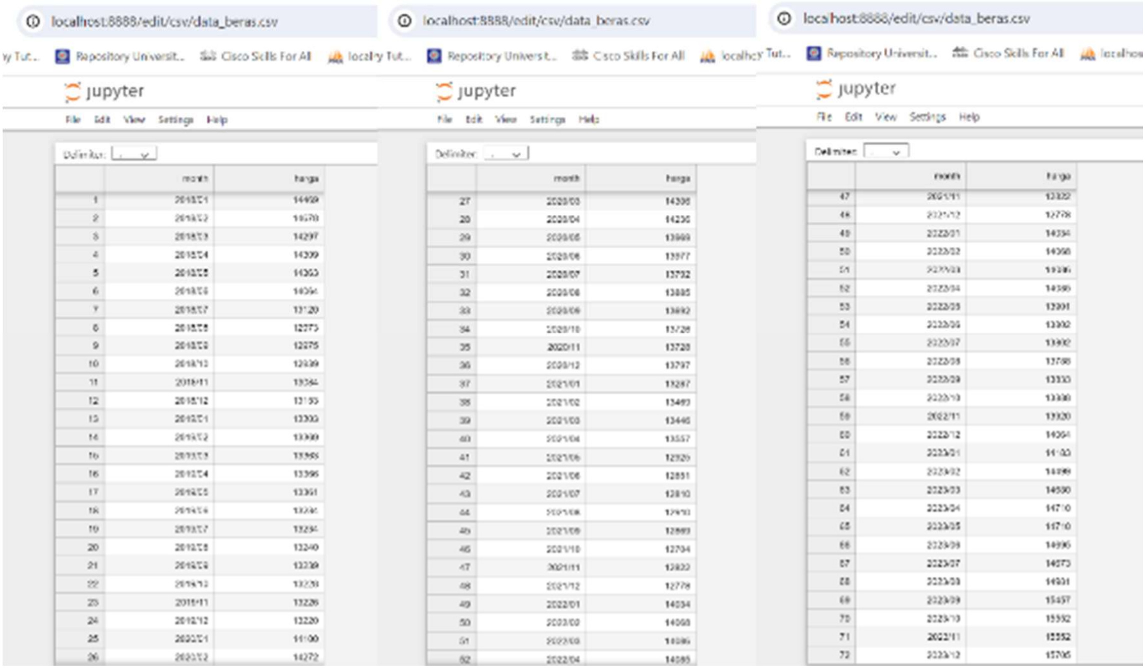

Figure 2. Price Data of Setra Rice (2018-2023)

Data Preprocessing

Before processing the data in the model, several libraries need to be installed in the Python program running via Jupyter Notebook. These include PMDARIMA, Pandas, Numpy, Matplotlib, Statsmodels, Seaborn, Scikit-Learn, Warnings, and Itertools. The data type used in the historical column for forecasting is the 'price' column, formatted as an integer and free of missing values. The results below indicate that the 'price' column is correctly formatted as an integer, with a 'non-null' status, meaning there are no empty rows (missing values).

```
[5]: import pandas as pd

# Memuat data dari CSV
file path = 'data beras.csv'  # Ganti dengan path sesuai lokasi file CSV
df = pd.read_csv(file_path)

# Memeriksa info data
print(df.info())

# Memeriksa jumlah missing values
missing_values = df.isnull().sum()
print("Jumlah missing values per kolom:")
print(missing_values)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72 entries, 0 to 71
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   month   72 non-null     object
 1   harga   72 non-null     int64
dtypes: int64(1), object(1)
memory usage: 1.3+ KB
None
Jumlah missing values per kolom:
month    0
harga    0
dtype: int64
```

Figure 3. Checking for Missing Values in the Data

Seasonal and Stationarity Testing

The retail rice price data in CSV format will be plotted annually to determine whether it exhibits a seasonal pattern. Next, the data will be tested using the adfuller function to check

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

for stationarity. If the results indicate that the data is not yet stationary, a differencing process will be applied to eliminate trends or seasonal patterns.

Differencing is the process of subtracting the current value from the previous value in a time series to measure the consistency of differences across each lag. A lag represents the time gap between data points. The ARIMA method requires the raw data to be stationary before it can be used for forecasting. The results show that the data does not exhibit a clear seasonal pattern, as its fluctuations appear random and vary significantly from one period to another.

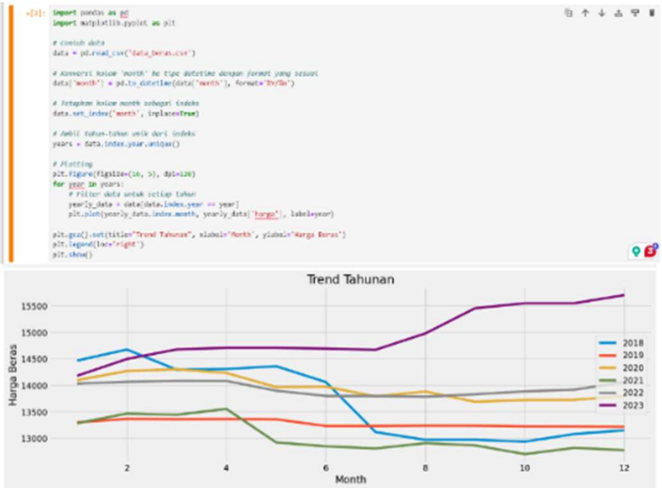

Figure 4. Seasonal Data Testing

Once the data is confirmed to have no seasonal patterns, it proceeds to the stationarity test using the adfuller function from the Statsmodels library. This involves applying the differencing operation with the Akaike Information Criterion (AIC) parameter and evaluating the p-value as a benchmark. If the p-value is less than 0.05, the data is considered stationary at the initial level (without differencing). However, if the p-value exceeds 0.05, the data is not yet stationary at the initial level, requiring differencing to be applied. The differencing process is performed up to a maximum of two levels. The results indicate that stationarity is achieved at the first differencing level.

Figure 5. Data Stationarity Test

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

Model Identification and Selection

Once the data is confirmed to be stationary, the AR(p), I(d), and MA(q) model parameters are identified using the auto_arima function. The model with the lowest Akaike Information Criterion (AIC) value is selected as the best model. The results from auto_arima indicate that the optimal model is ARIMA(0,1,0).

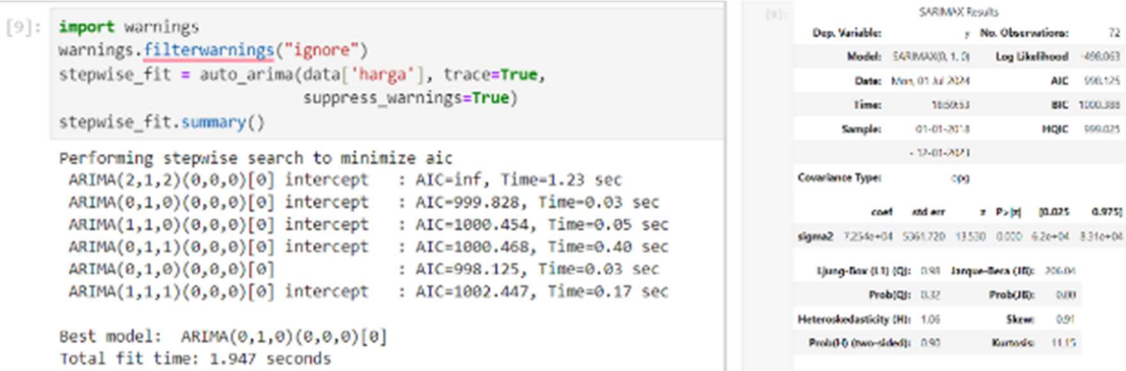

Figure 6. Model Identification Using Auto-ARIMA

The initial model interpretation is suboptimal because both the p and q values are 0, meaning there are no lag terms available as a reference for forecasting results. To improve the model, parameter tuning is performed using the Grid Search technique. In this process, the data is split into two sets: 80% as the training set (57 observations) and 20% as the test set (15 observations). The model is validated using the Walk-Forward Validation method, with Mean Squared Error (MSE) as the evaluation metric to determine the best model. The optimal combination of (p, d, q) parameters found is (4, 2, 4), where the differencing level selected by the script is 2.

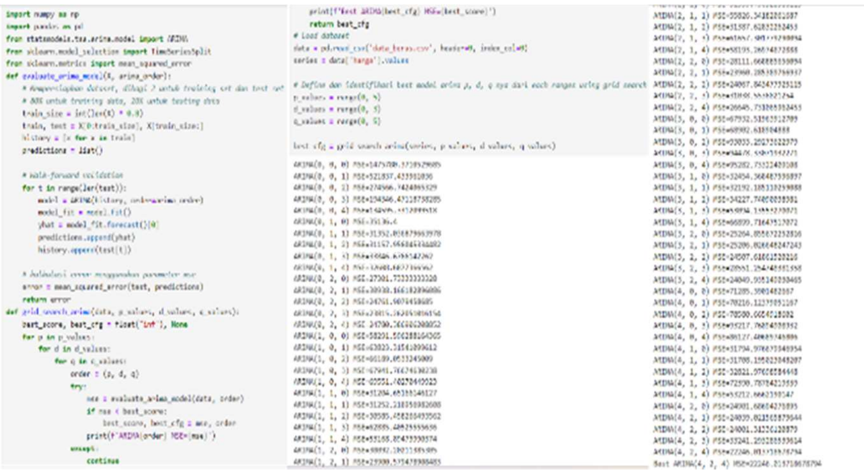

Figure 7. Tuning Grid Search

Next, the model fitting process is performed. The selected model (4, 2, 4) is applied to the training set and fitted using evaluation criteria such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Hannah-Quinn Information Criterion (HQIC), where lower values indicate better model performance. Additionally, the model is assessed using the Ljung-Box Test, Jarque-Bera Test, Heteroskedasticity Test, as well as skewness and kurtosis measurements to evaluate residual normality and model adequacy.

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

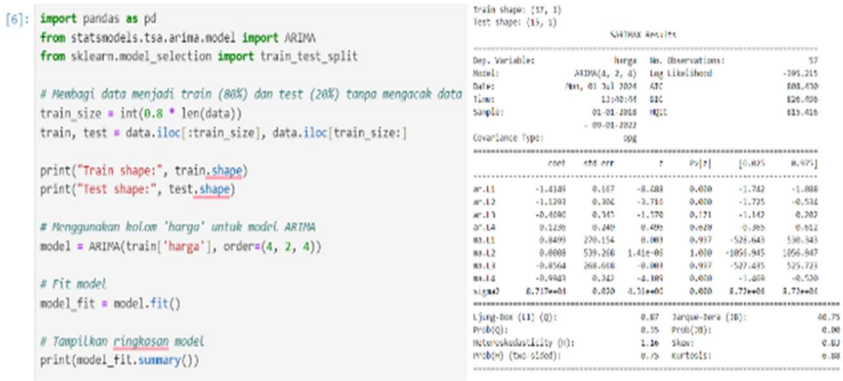

Figure 8. Model Fitting Process

The p-values from each statistical test show varying results. The Ljung-Box test indicates a good probability as it is greater than 0.05, suggesting that the residuals are not significantly autocorrelated. However, the Jarque-Bera test yields a probability of less than 0.05, indicating that the residuals do not follow a normal distribution. Meanwhile, the Heteroskedasticity test shows a good result, with a probability greater than 0.05, suggesting that the variance of the residuals remains stable over time.

Residual Autocorrelation Test and ACF & PACF Plots

Next, a residual autocorrelation test is conducted to determine whether the residuals exhibit sharp differences, potentially leading to outliers (significant residual spikes) in the graphical visualization. This test also assesses whether residuals are correlated across different time periods.

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are displayed in the form of correlograms to manually examine the correlation between the current data and previous data at specific lag intervals. These plots help in understanding the dependency structure of the time series and validating the model's effectiveness in capturing underlying patterns.

Figure 9. Residual Autocorrelation Test and ACF & PACF Plots

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

It can be observed that the residual outliers, when handled using the mean, are still not normally distributed. Therefore, further adjustments using the median are necessary. The results, illustrated in the histogram and Q-Q plot, show that the residual points tend to align more closely with the normal curve. However, the lag values in the ACF and PACF plots cannot be used as references for determining the (p, d, q) model parameters, as the model has already been optimized using the Grid Search tuning method earlier in the analysis.

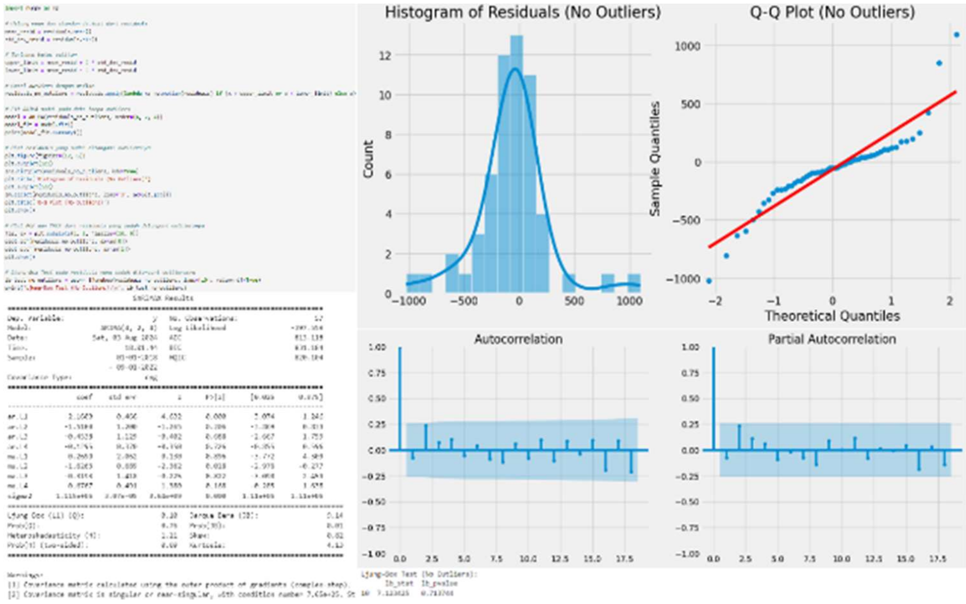

Figure 10. Outlier Handling

Forecasting Results

The next step involves the forecasting process and evaluating the accuracy of the results. Previously, the data was divided into training and testing sets, where the training set was used for model training before being implemented on the testing set. The accuracy of the forecast is then assessed using error metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) by comparing the actual 2023 data with the forecasted data for the same year.

Ff the results show a significant discrepancy, the accuracy is low, whereas if the forecast closely matches the actual data, the accuracy is high. The forecasting results for 2024, 2025, and 2026 indicate that rice prices in DKI Jakarta Province will continue to rise, with the confidence interval for predictions widening over time.

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

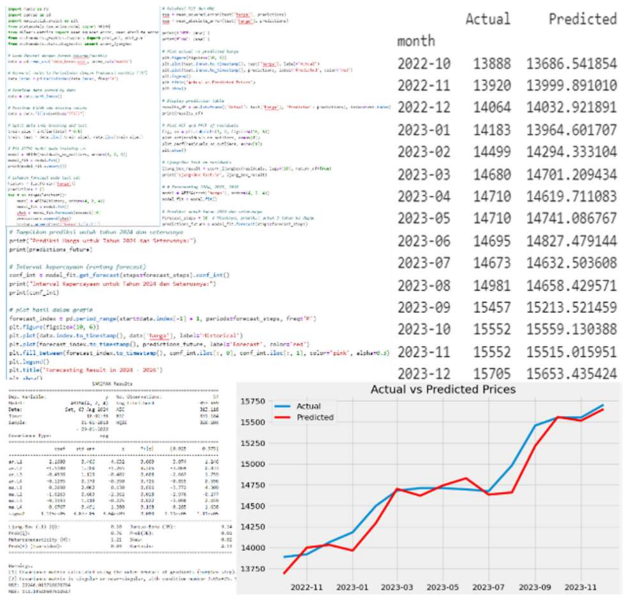

Figure 11. Actual and Forecasted Prices for 2023

Accuracy Testing

The final step in the forecasting process is accuracy testing using four error metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The lower the MSE and RMSE values, the better the model's performance. Similarly, MAE follows the same principle as RMSE and MSE. A MAPE value is considered accurate if it is less than 10%.

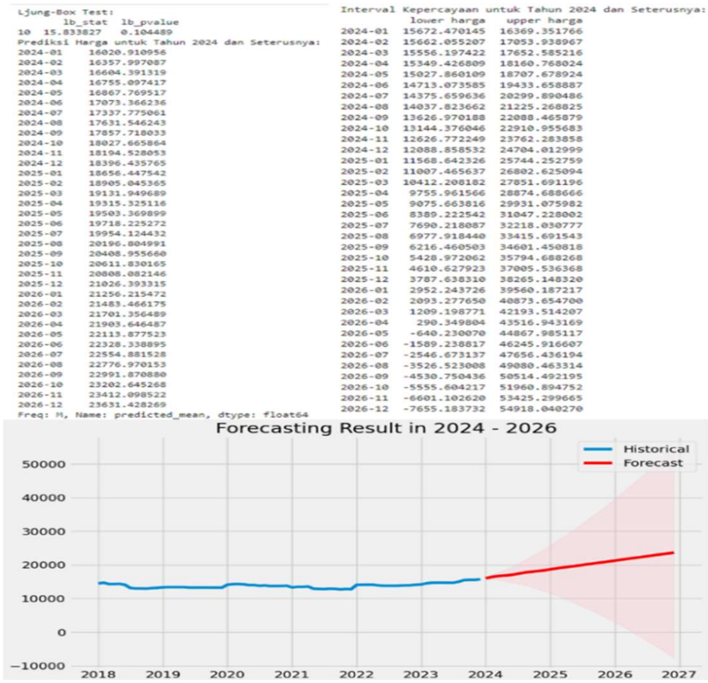

Figure 12. Forecasting Results for 2024 – 2026

Jurnal Multidisiplin Sahombu
Volume 5, Number 02, 2025, DOI 10.58471/jms.v5i02
ESSN 2809-8587 (Online)
https://ejournal.seaninstitute.or.id/index.php/JMS/index

CONCLUSION

The forecast for Setra rice prices in DKI Jakarta Province predicts a steady monthly increase from 2024 to 2026. The ARIMA model applied demonstrates high forecasting accuracy, as validated by accuracy testing using four error metrics, all of which fall within a low range. The results include an MSE value of 22,246, an RMSE value of 149.15, an MAE value of 114.18, and a MAPE value of 0.78% (with an ideal accuracy threshold below 10%). The RMSE and MAE indicate that the error margin or deviation between the forecasted model and actual data is only within the range of 100–150 rupiahs. The model successfully stationarized the data, eliminated outliers in the residual distribution, and applied the optimal (p, d, q) parameters obtained through the Grid Search tuning process. While rice price data exhibits a specific pattern, additional parameters are needed for more precise identification. Future research can explore this aspect further by incorporating additional variables and utilizing more complex forecasting methods.

REFERENCE

Anandyani, A. R., Astutik, D. K. A., Bariroh, N., & Indrasetianigsih, A. (2021). Prediksi Rata-Rata Harga Beras Yang Dijual Oleh Pedagang Besar (Grosir) Menggunakan Metode Arima Box Jenkins. Teknosains: Media Informasi Sains Dan Teknologi, 15(2), 151. https://doi.org/10.24252/teknosains.v15i2.17721

BPS. (2022). Provinsi DKI Jakarta Dalam Angka DKI Jakarta Province In Figures 2022. 81–81.

Elsa Fitri, A., Nabila, D., Melenia, D., Guntoro Aji, L., Salwa Azmah, N., & Fadilah, S. (2021). Peramalan Harga Beras Pada Pasar Tradisional Di Indonesia Dengan Menggunakan Model Arima. DwijenAGRO, 11(1), 12–16. http://ejournal.undwi.ac.id/index.php/dwijenagro/article/view/1079%0Ahttp://ejournal.undwi.ac.id/index.php/dwijenagro/article/download/1079/950

Fadhlurrahman, I. (2024). Harga Beras Premium di DKI Jakarta Tiga Bulan Terakhir Turun 5,15%. https://databoks.katadata.co.id/datapublish/2024/06/07/harga-beras-premium-di-dki-jakarta-tiga-bulan-terakhir-turun-515

Fauzani, S. P., & Rahmi, D. (2023). Penerapan Metode ARIMA Dalam Peramalan Harga Produksi Karet di Provinsi Riau. Jurnal Teknologi Dan Manajemen Industri Terapan, 2(4), 269–277. https://doi.org/10.55826/tmit.v2i4.283

Lestiyanto Wibisono, Arman, & Taufikurohman, M. R. (2022). Analisis Ketersediaan Stok Beras Terhadap Harga Beras Di Dki Jakarta (Studi Kasus: Pasar Induk Beras Cipinang). Jurnal Bioindustri, 4(2), 149–163. https://doi.org/10.31326/jbio.v4i2.1007

Natasya, Musdalifah, S., & Andri. (2021). Prediksi Harga Beras Di Tingkat Perdagangan Besar Indonesia Menggunakan Algoritma Backpropagation. Jurnal Ilmiah Matematika Dan Terapan, 18(2), 148–159. https://doi.org/10.22487/2540766x.2021.v18.i2.15688

Naya, F. P., Berlianti, S. S., Parcha, N., & Kayla, A. (2024). Peramalan harga beras indonesia menggunakan metode arima. 6(2), 184–193.

Shidiq, B. G. A., Furqon, M. T., & Muflikhah, L. (2022). Prediksi Harga Beras menggunakan Metode Least Square. Pengembangan Teknologi Informasi Dan Ilmu Komputer, 6(3), 1149–1154.