The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00024 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Implementation of the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) Algorithm for Rice Price Prediction.

Ezra Sasqia Syahna[1] **Zara Yunizar**[2] **Zahratul Fitri**[3]

[1,2,3] Universitas Malikussaleh
[1,2,3] ezra.200170143@mhs.unimal.ac.id zarayunizar@unimal.ac.id , zahratulfitri@unimal.ac.id
* Correspondence: Ezra Sasqia Syahna

Abstract

This study implements the Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) model to predict rice prices (gabah) based on historical data from 2020 to 2024. Utilizing data obtained from Investing.com, the research integrates key external variables such as temperature, fertilizer prices, and production levels to enhance prediction accuracy. The methodology comprises systematic steps, including data collection, processing, and model evaluation, employing metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess performance. The findings reveal a strong correlation between the predicted and actual market prices, particularly in the closing price category, which achieved a MAPE of 1.354%. The evaluation metrics further confirmed the model's robustness, with the closing price exhibiting the lowest MSE of 299,629.64 and RMSE of 547.38. Although the highest price category showed a slightly higher MAPE of 2.007%, all categories remained under the acceptable threshold of 2%, indicating satisfactory prediction accuracy. In conclusion, the SARIMAX model demonstrates significant effectiveness in rice price forecasting, providing valuable insights for stakeholders in the agricultural market. The implementation within a web application facilitates real-time predictions, supporting informed decision-making and enhancing market strategies.
Keywords: SARIMAX; rice prices; prediction model; MAPE; agricultural market; time series analysis.

Introduction

Agriculture plays a pivotal role in the economy of Indonesia, where a significant portion of the population relies on farming for their livelihood. Among agricultural commodities, rice (gabah) is fundamental to food security in the nation, serving as the primary staple food for millions of Indonesians. The stability of rice production and distribution directly impacts food availability, pricing, and ultimately the socio-economic stability of communities across the country. As highlighted by Dhamira and Irham [1], fluctuations in rice prices can have profound implications not only for farmers but also for consumers and policymakers. In recent years, rice prices have exhibited considerable volatility due to a myriad of factors. These include climatic changes that affect crop yields, government policies related to import and export regulations, fluctuations in global market conditions, and unexpected events such as pest infestations or natural disasters. As a result, stakeholders throughout the rice supply chain, from farmers to end consumers, face significant risks associated with price instability [2].Understanding these dynamics is crucial for developing effective strategies to mitigate risks and enhance the overall resilience of the agricultural sector. Traditional methods of price forecasting often rely on fundamental analyses, focusing on macroeconomic conditions and supply-demand relationships. However, these methods can sometimes fall short in providing timely and actionable insights. Technical analysis, which studies past price movements to predict future trends, offers an alternative approach. Unlike fundamental analysis, technical analysis leverages historical price data and trading volumes to identify patterns and potential future price movements. This methodology allows for the identification of trends, support and resistance levels, and chart patterns that can aid decision-making in a fluctuating market environment. In the context of rice price prediction, the Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) model stands out as a powerful tool. SARIMAX integrates historical data with external variables such as weather conditions and government policies to enhance forecasting accuracy. By capturing seasonal patterns and trends while accounting for external influences, SARIMAX enables a more nuanced understanding of the factors driving rice prices. This approach not only aids farmers in planning their production but also supports policymakers in formulating effective agricultural strategies. Previous studies, such as those conducted by Nasirudin and Dzikrullah [3],have demonstrated the effectiveness of

SARIMAX in predicting prices of other agricultural commodities, such as chili peppers. These studies underscore the importance of incorporating exogenous variables into predictive models to achieve more accurate forecasting results. Building upon this foundation, the current research aims to implement the SARIMAX algorithm to develop a robust rice price prediction model. This research seeks to contribute to the academic literature on agricultural price forecasting and provide valuable insights for stakeholders involved in rice production and trade. By enhancing the predictive accuracy of rice prices, the research endeavors to support better decision-making in a complex and dynamic agricultural landscape, ultimately contributing to improved food security and economic stability in Indonesia.

Literature Review

2.1 Rice (Gabah)

Rice is a staple food that holds immense significance in ensuring food security, especially in countries like Indonesia, where a large portion of the population depends on it for their daily sustenance. The production and price of rice (gabah) are influenced by various factors, including climate conditions, government policies, and market dynamics. As pointed out by Simanjuntak [4], the challenges faced by the rice sector are compounded by the need to balance two conflicting goals: ensuring profitable prices for farmers while keeping rice affordable for consumers. A survey conducted by the Central Statistics Agency (BPS) indicated a decline in rice harvesting area from January to December 2023, reaching approximately 10.21 million hectares—a decrease of 2.29 percent from 2022. The peak harvest period in 2023 remained consistent with the previous year, occurring in March, when the harvested area was recorded at 1.65 million hectares. This fluctuation in harvesting area highlights the volatility within the rice market and the various challenges faced by stakeholders [5]

2.2 Price Trends of Rice

The market for rice is characterized by its dependence on seasonal production and its vulnerability to weather disruptions and pest infestations, leading to unpredictable supply levels. Farmers often possess minimal surplus production capacity, limited storage capabilities, and an urgent need for cash flow, which diminishes their bargaining power in rice trading. These dynamics can result in significant price volatility, making it essential for stakeholders to have access to reliable price predictions. As demonstrated in previous research, the integration of technical analysis with predictive modeling can provide valuable insights into price movements. For instance, a study by Nasirudin [3] focused on modeling chili prices in Indonesia using SARIMAX, incorporating factors such as rainfall and inflation, which resulted in a Mean Absolute Percentage Error (MAPE) of 6.889%. This finding underscores the effectiveness of incorporating external variables to enhance forecasting accuracy.

2.3 SARIMAX Algorithm

The Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) model is an extension of the ARIMA model, designed to handle time series data that exhibit both seasonality and external influences. SARIMAX incorporates seasonal patterns while allowing for the inclusion of exogenous variables that may impact the dependent variable. According to Rochayati (2023), the SARIMAX model considers that the effect on the dependent variable (Y) at a specific time (t) is influenced not only by its historical behavior but also by other independent variables present at that same time.[3]

The SARIMAX model is characterized by the following components:

a. AR (p): The autoregressive part of the model, which indicates the number of lagged observations included.
b. I (d): The integrated part, representing the number of times the data has been differenced to achieve stationarity.
c. MA (q): The moving average part, which represents the number of lagged forecast errors in the prediction equation.
d. Seasonal components are represented by P, D, Q, and m, where m signifies the length of the seasonal cycle.

The mathematical representation of the SARIMAX model can be expressed as follows:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where:
Y_t = the value of the time series at time t
ϕ = autoregressive coefficients
θ = moving average coefficients
ϵ = error term at time t

The equation highlights how the current value of the series (Y_t) is influenced by its past values and the past errors in prediction, enabling the model to capture both trends and seasonality effectively. In this research, the SARIMAX model will be utilized to analyze historical rice price data and provide accurate predictions. By integrating external variables such as weather conditions and government policies, the model aims to enhance forecasting reliability, ultimately assisting stakeholders in making informed decisions regarding rice management and sales.[6]

Materials & Methods

3.1 Research Location and Timeframe

This research was conducted as part of the final project requirements for the Informatics Engineering program at Universitas Malikussaleh. The study was carried out online from January 2023 until its completion. Historical data was obtained from Investing.com to predict rice prices (gabah) for the period of 2020 to 2024. The key input data for this research included opening prices, highest prices, lowest prices, closing prices, trading volumes, and price changes, forming the foundation for the subsequent analysis.

3.2 Research Steps

To ensure a comprehensive understanding of the research subject, the author implemented a structured and systematic methodology. Each step of the research process was meticulously designed. The methodology involved several key phases: First, data collection was initiated, focusing on gathering data that forms the core of this research. This phase included both primary and secondary data, with a specific emphasis on obtaining relevant historical data on rice prices. The primary data consisted of historical prices sourced from Investing.com, covering the specified period.

Following data collection, the research proceeded to the data processing methods phase. Various basic statistical methods were utilized to analyze the data, including calculations of the mean, mode, median, and percentage. Software tools such as Excel and Google Spreadsheet facilitated the processing and analysis of this data.

The next phase involved system design, where Context Diagrams, Data Flow Diagrams (DFD), and Entity-Relationship Diagrams (ERD) were created. This design phase was crucial for guiding the application development process, allowing for a structured approach to programming. In the system implementation phase, the application was developed using programming languages such as HTML, Python, and JavaScript. This step marked the transition from design to practical application, where the theoretical framework was translated into a functioning system.

Finally, the research encompassed a system testing phase, which involved rigorous testing and debugging procedures to ensure that the application operated as intended. This phase included data preparation, where historical time series data on rice was cleaned, and checked for stationarity, along with exploratory data analysis to assess potential AR and MA parameters.

3.3 Research Methods

The methodology included testing historical data as a measurement tool for predicting rice prices. This involved a combination of literature and data collection methods. The library method was employed to gather relevant literature, including books and scientific journals, to build a solid foundation for the research. This literature review encompassed previous studies focusing on similar thematic issues, reinforcing the theoretical basis for the analysis.

Additionally, a thorough literature study was conducted to gather essential theoretical frameworks regarding the implementation of the SARIMAX algorithm for predicting rice prices. This process involved reviewing various sources such as scholarly articles and online resources, focusing on time series analysis, the SARIMAX model, and its application in commodity price prediction. The literature review aimed to critically analyze past research, synthesize findings, and identify external factors that may influence rice prices.

3.4 System Requirements Analysis

The analysis of system requirements aimed to acquire a comprehensive understanding of the actual needs for the system design and development. This analysis served as a key reference for the subsequent stages of system design.

Key components of the analysis included data input, which involved collecting historical rice price data from 2020-2024, capturing price variability such as opening prices, highest prices, lowest prices, and trading volumes. Additionally, real-time data access was identified as crucial for enhancing the accuracy of future price predictions.

Furthermore, the algorithm implementation focused on applying the SARIMAX algorithm with optimized parameters to predict rice prices based on historical data. Validation processes were established to ensure the model's reliability and accuracy. The user interface design emphasized presenting historical and predicted rice prices in a user-friendly manner, ensuring easy navigation for accessing analysis features and reports.

In summary, this research methodology provided a structured approach to the investigation of rice price prediction, utilizing the SARIMAX model and a rigorous analytical framework to address the complexities of agricultural price forecasting.

Rice Futures Data (April 30, 2024 - May 29, 2024)

This table displays the historical data for international rice futures prices sourced from Investing.com. The data is arranged chronologically from the earliest date to the most recent, providing insights into the price trends and market behavior over the specified period. Each entry includes key metrics relevant to trading, which are essential for analyzing price movements and forecasting future trends using models like SARIMAX.

Table 1 Rice Price Sample Data

Date	Last Price	Opening Price	Highest Price	Lowest Price
29/05/2024	$ 17.82	$ 18.04	$ 18.04	$ 17.81
28/05/2024	$ 18.09	$ 18.46	$ 18.60	$ 18.05
24/05/2024	$ 18.34	$ 18.46	$ 18.49	$ 18.30
23/05/2024	$ 18.49	$ 18.82	$ 19.01	$ 18.38

22/05/2024	$ 18.75	$ 18.58	$ 18.80	$ 18.57
21/05/2024	$ 18.59	$ 18.83	$ 18.84	$ 18.57
20/05/2024	$ 18.85	$ 18.77	$ 18.87	$ 18.70
17/05/2024	$ 18.76	$ 19.09	$ 19.15	$ 18.73
16/05/2024	$ 19.10	$ 19.04	$ 19.18	$ 18.83
15/05/2024	$ 19.11	$ 18.90	$ 19.16	$ 18.73
14/05/2024	$ 18.72	$ 18.65	$ 18.93	$ 18.56
13/05/2024	$ 18.71	$ 19.25	$ 19.25	$ 19.25
10/05/2024	$ 19.26	$ 18.76	$ 18.84	$ 18.76
09/05/2024	$ 18.69	$ 18.44	$ 19.01	$ 18.44
08/05/2024	$ 18.32	$ 18.40	$ 18.48	$ 18.40
07/05/2024	$ 18.30	$ 18.41	$ 18.41	$ 18.41
06/05/2024	$ 18.41	$ 18.40	$ 18.42	$ 18.18
03/05/2024	$ 18.74	$ 18.65	$ 18.65	$ 18.65
02/05/2024	$ 18.17	$ 18.45	$ 18.45	$ 18.30
01/05/2024	$ 18.97	$ 19.09	$ 19.17	$ 19.09
30/04/2024	$ 18.89	$ 19.11	$ 19.15	$ 18.85

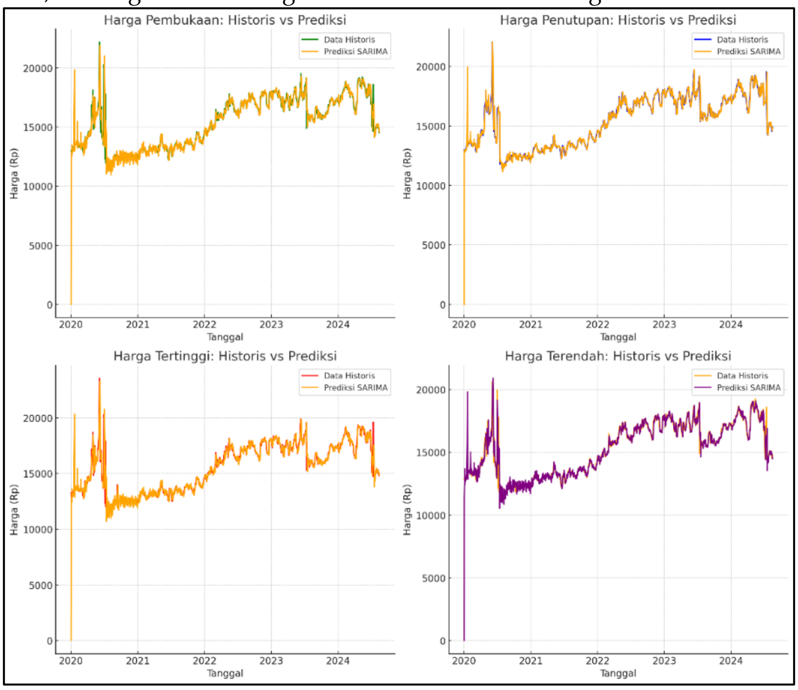
Explanation of the Table Format

Date: The date when the price data was recorded, formatted as DD/MM/YYYY.
Last Price: The closing price of rice on the specified date.
Opening Price: The price at which rice started trading on that date.
Highest Price: The highest price reached during the trading session.
Lowest Price: The lowest price recorded during the trading session.

This format allows for a clear and concise presentation of rice futures data, facilitating analysis of price trends over time. The organization of the data in this manner is particularly useful for applying statistical models to forecast future price movements, providing essential insights for stakeholders in the agricultural market.

Results and Discussion

The implementation of the Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) model was designed to predict rice prices (gabah) effectively. This model utilizes historical data alongside exogenous variables, such as temperature, fertilizer prices, and production levels, to enhance the accuracy of its predictions. The integration of the SARIMAX algorithm into a web-based system allows users to view real-time price predictions based on relevant external factors, offering valuable insights for stakeholders in the agricultural market.



Picture 1 Prediction Result

Opening Price Analysis

The analysis of opening prices reveals a close correlation between the SARIMAX prediction line and historical data trends. For instance, on May 29, 2024, both the predicted and actual opening price were $18.04, demonstrating the model's effectiveness. However, minor fluctuations are observed on certain days, particularly during volatile trading periods. This discrepancy may arise from the inherent volatility of opening prices, which are often influenced by market sentiment. Despite these variations, the SARIMAX model demonstrates strong capabilities in identifying seasonal patterns and long-term trends, although the predictions may not always reflect actual prices due to external market influences.

Closing Price Analysis

The predictions for closing prices show a robust alignment with historical data. On May 29, 2024, the predicted closing price of $17.82 matched the actual closing price precisely. The model effectively tracks the movement of closing prices, even during instances of volatility, such as observed on May 23, 2024. The stability of closing prices, compared to opening prices, allows SARIMAX to capture seasonal patterns more accurately. However, minor deviations can occur during periods of sudden market activity, which the model may not fully anticipate.

Highest Price Analysis

When examining the predictions for the highest prices, the SARIMAX model again demonstrates its capability to capture seasonal trends and long-term price movements effectively. For example, on May 28, 2024, the predicted highest price was $18.60, which corresponded with the actual price. Nonetheless, the model occasionally struggles with short-term extreme fluctuations, reflecting the speculative nature of market sentiment. On days with significant market activity, such as May 10, 2024, the model may fail to predict sharp price spikes, illustrating the challenges of forecasting in a volatile market environment.

Lowest Price Analysis

The predictions for the lowest prices are generally well-aligned with historical data, indicating the model's effectiveness in identifying downward trends. On May 29, 2024, the predicted lowest price of $17.81 matched the actual lowest price precisely. The SARIMAX model successfully captures seasonal patterns, although minor discrepancies can arise during highly volatile periods. For example, on May 21, 2024, both the predicted and actual lowest prices were $18.57, demonstrating the model's reliability in forecasting market conditions. However, for sudden fluctuations, further adjustments to the model may be necessary to address extreme volatility.

Model Evaluation

The evaluation of the SARIMAX model's accuracy in predicting rice prices (gabah) involved measuring three critical metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide insight into the model's predictive performance across different price categories.
The results of these evaluations are summarized in the following table:

Table 2 Model Evaluation

Category	MSE	RMSE	MAPE (%)
Opening Price	488,800.58	699.14	1.999
Closing Price	299,629.64	547.38	1.354
Highest Price	506,112.62	711.42	2.007
Lowest Price	433,995.39	658.78	1.901

Evaluation Metrics

The Mean Squared Error (MSE) measures the average squared difference between the actual and predicted values, with lower values indicating better predictive accuracy. In this study, the closing price category exhibited the lowest MSE at 299,629.64, highlighting the model's strong performance in this area. The Root Mean Squared Error (RMSE) is the square root of MSE, providing the error measurement in the same units as the original data. Here, the closing price again showed the best performance with an RMSE of 547.38. Lastly, the Mean Absolute Percentage Error (MAPE) reflects the average percentage error in predictions, with values below 2% considered acceptable for accuracy. The model achieved a MAPE of 1.354% for closing prices, while the highest price category had a MAPE of 2.007%, which is still within an acceptable range.

Overall Conclusions

The SARIMAX model demonstrated commendable accuracy across all four price categories—opening, closing, highest, and lowest prices. The model's strongest performance was noted in the closing price category, where it recorded the lowest MAPE of 1.354%. In terms of RMSE and MSE, the opening price showed a RMSE of 699.14 and an MSE of 488,800.58, indicating satisfactory performance in capturing daily price fluctuations.
The model effectively captured seasonal patterns, particularly in closing prices, which are generally more stable. The SARIMAX algorithm successfully integrated seasonal parameters (P,D,Q,s) = (1,1,1,12) to account for these patterns,

ensuring that the predictions remained consistent with the historical trends observed in the data. Furthermore, the implementation of the SARIMAX model within the web application has proven to be efficient, allowing users to access real-time predictions based on reliable historical data from sources like Investing.com.

In summary, the evaluation of the SARIMAX model affirms its robustness in predicting rice prices, providing stakeholders with reliable insights for informed decision-making in agricultural markets.

Conclusions

The SARIMAX model demonstrated strong predictive capabilities in forecasting rice prices (gabah) across various categories, including opening, closing, highest, and lowest prices. The model's predictions closely aligned with actual market prices, particularly for closing prices, which had a Mean Absolute Percentage Error (MAPE) of **1.354%**. This indicates that the model was able to capture the underlying trends effectively, despite some fluctuations due to market volatility. For instance, the predicted closing price on May 29, 2024, matched the actual price, showcasing the model's reliability.In terms of evaluation metrics, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) further confirmed the model's performance. The closing price category achieved the lowest MSE of **299,629.64** and RMSE of **547.38**, underscoring its accuracy in reflecting true market conditions. Conversely, the highest price category showed a higher MAPE of **2.007%**, indicating slightly less precision in its predictions. Overall, the SARIMAX model effectively balanced accuracy and reliability, making it a valuable tool for stakeholders in the agricultural market to inform their decision-making processes.

References

[1] A. Dhamira and I. Irham, “the Impact of Climatic Factors Towards Rice Production in Indonesia,” *Agro Ekon.*, vol. 31, no. 1, 2020, doi: 10.22146/ae.55153.

[2] A. W. Putra, J. Supriatna, R. H. Koestoer, and T. E. B. Soesilo, “Differences in local rice price volatility, climate, and macroeconomic determinants in the indonesian market,” *Sustain.*, vol. 13, no. 8, 2021, doi: 10.3390/su13084465.

[3] Faris Nasirudin and Abdullah Ahmad dzikrullah, “Pemodelan Harga Cabai Indonesia dengan Metode Seasonal ARIMAX,” *J. Stat. dan Apl.*, vol. 7, no. 1, pp. 105–115, 2023, doi: 10.21009/jsa.07110.

[4] Romauli Simanjuntak, HotmanTuah Purba, and Marojaan Candro Sitorus, “ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PENDAPATAN USAHATANI PADI SAWAH (Studi Kasus di Kelurahan Tong Marimbun, Kecamatan Siantar Marimbun, Kota Pematang Siantar),” *J. Agrilink*, vol. 3, no. 1, pp. 44–52, 2021, doi: 10.36985/jak.v3i1.210.

[5] B. P. S. Indonesia, “Luas Panen dan Produksi Padi di Indonesia 2023 (Angka Sementara),” 2023. [Online]. Available: https://www.bps.go.id/id/pressrelease/2023/10/16/2037/luas-panen-dan-produksi-padi-di-indonesia-2023--angka-sementara-html

[6] N. S. Arunraj, D. Ahrens, and M. Fernandes, “Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry,” *Int. J. Oper. Res. Inf. Syst.*, vol. 7, no. 2, pp. 1–21, 2016, doi: 10.4018/ijoris.2016040101.