International Journal of Advances in Data and Information Systems
Vol. 6, No. 1, April 2025, pp. 197~203
ISSN: 2721-3056, DOI: 10.59395/ijadis.v6i1.1356

Implementation of ARIMA for Prediction of Paddy Rice Production in Cisolok Sub-District, Sukabumi District

Rafi Abdul Mugni[1], Iwan Rizal Setiawan[2], Didik Indrayana[3]
[1,2,3] Faculty of Science and Technology, Universitas Muhammadiyah Sukabumi
[1]rafiabdulmugni757@gmail.com, [2]metalizer_5150@ummi.ac.id, [3]didik.ind@ummi.ac.id

Article Info

Article history:

Received Feb 19, 2025
Revised Mar 29, 2025
Accepted Apr 30, 2025

Keywords:

Prediction,
Rice Production,
Arima,
KDD,
Cisolok District

ABSTRACT

Indonesia as an agricultural country, agriculture, especially paddy production, plays an important role in food security. However, Cisolok District, Sukabumi Regency faces challenges in terms of effective rice production management. This study aims to improve the accuracy of rice production prediction in Cisolok District by implementing Arima. The methodology used is Knowledge Discovery in Databases (KDD), which includes data selection, data pre-processing, model selection, model training, and model evaluation. The data used include weather attributes and paddy production, which are collected from various related sources. The results of the study indicate that the model built with Arima provides accurate estimates and can help farmers and decision makers in planning and managing paddy production more efficiently. These findings are expected to increase paddy productivity in Cisolok District, Sukabumi Regency.

Corresponding Author:

Rafi Abdul Mugni,
Faculty of Science and Technology,
Universitas Muhammadiyah Sukabumi
Jl. R. Syamsudin, S.H., No. 50 Kota Sukabumi
Email: rafiabdulmugni757@gmail.com

1. INTRODUCTION

The majority of Indonesians live from agriculture, as Indonesia is an agrarian country. This is due to the fact that most of Indonesia is in the tropics, which means that the weather conditions, soil and other resources in each region have high potential for developing the agricultural sector. The utilization of agricultural resources is essential to increase agricultural productivity, so limited resources must be used as effectively as possible.

Rural agricultural development is an interrelated component. Rural life is greatly influenced by agriculture. However, the overall role of agriculture has not developed sufficiently to improve farmers' welfare as expected. As the rural economy develops, the role of the agricultural sector faces many challenges. To increase the contribution of the agricultural sector in the rural and national economy, a future development plan is needed that includes a policy agenda [1].

Rice is a very important agricultural commodity in Indonesia, as most of the population consumes rice. The demand for rice continues to increase along with population growth. To face this challenge, the government continues to strive to increase rice production through the Ministry of Agriculture by providing seed assistance to farmers, as high-quality seeds contribute to increased yields. Rice production plays a key role in the agricultural sector in Indonesia and has a major effect on food security. In addition, increasing rice production also brings benefits such as the development of water infrastructure, the application of modern technology in rice cultivation, and the improvement of the quality of human resources in the sector [2].

According to West Java Province Statistics (2021), the harvest area of Sukabumi District reached 93,371 ha in 2020 with a production of 521,459 tons of milled dry grain (MDG). The harvest area decreased to 89,510 ha in 2021 to 492,926 tons of MDG. Based on the above production achievements, the rice yield reached 301,133 tons in 2020 and decreased to 284,656 tons in 2021. Sukabumi District's rice planting target for 2022 has been increased by approximately 2,925 hectares. This will allow the planting area to reach 149,191 hectares and achieve surplus [3].

An increase in temperature leads to an increase in transpiration, which is followed by an increase in nutrient uptake. These two metabolic mechanisms, including photosynthesis, increase the capacity to produce new cells, which is the basis for tiller formation. Plants produce more tillers during high temperature stress to acclimatize to their environment and lower the air temperature [4].

Optimal soil moisture for rice plant growth can vary depending on the growth phase of the plant and the type of soil. In general, rice requires sufficient soil moisture but not too wet or too [5].

Evaporative pressure increases as air humidity decreases, which causes plants to lose more water through transpiration. As a result, rice plants are more susceptible to drought and water stress, which can inhibit vegetative growth, flower formation, and grain filling (Nurhidayat A et al., 2024) .

Rainfall greatly affects agricultural cultivation activities and farmland productivity. Rainfall is a limiting factor in agricultural cultivation activities and production. Several studies on the effect of climate change, especially changes in rainfall patterns on agricultural land productivity, have been conducted in various regions in Indonesia [7].

Since photosynthesis only occurs in the presence of light and is carried out by the green pigment chlorophyll in chloroplasts, a cytoplasmic organelle, the speed of photosynthesis can be affected by the intensity of sunlight. The absorbed solar energy is converted into chemical energy, a chemical substance with high energy, which is also necessary for sugar production [8].

Evaporation is influenced by wind as it can move basaltic air in direct contact with permafrost to warmer places. As wind speed increases, more water vapor is lost, resulting in higher rainfall. Pollinating insects are more active when the wind is light, which helps pollinate flowers. However, when the wind is strong, pollinating insects tend to be absent, which reduces the chances of successful seed breeding [9].

The problem in Cisolok sub-district is the absence of implementation of dataset modeling using machine learning for the prediction of paddy rice production. To date, there has been no systematic effort to utilize machine learning technology to predict the factors that influence the prediction of paddy rice production in Cisolok sub-district, Sukabumi district. The lack of dataset modeling using machine learning can limit the potential in Sukabumi district to identify these patterns more precisely, so that intervention and corrective actions can be taken more precisely to increase the level of paddy rice production in Cisolok sub-district, Sukabumi district.

Comparison of algorithms based on the regression model obtained, 82.6% of rice productivity factors can be explained by production, harvest area, planting area, rainfall, and rainy days [10]. By using the SVM (*Support Vector Machine*) algorithm, the RMSE (*Root Mean Square Error*) value of 1.49071 was obtained from manual calculations, which had 11 training data and 9 test data, and had an accuracy rate of 44%. The results of both manual calculations and prediction models both get an RMSE (*Root Mean Square Error*) value of 1.49 with an accuracy of 44% [11]. Using the *Decision Tree* algorithm, the MAPE values for 1000 trials are 12.758%; 12.79%; 12.92%; and 13.13%; and the proportion of test data is 10%, 20%, 30%, and 40% [12]. To overcome these problems, one solution that can be implemented is to use the *Arima* method in modeling datasets related to paddy rice production in Cisolok District, Sukabumi Regency.

2. RESEARCH

In this research, the method used is Knowledge Discovery in Database (KDD) to analyze and predict rice production effectively and efficiently.

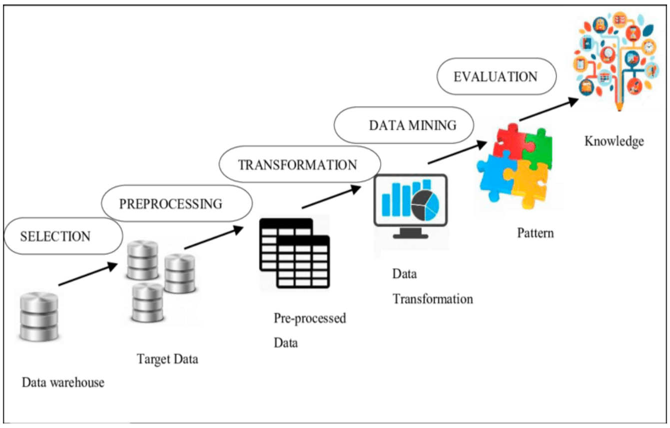

Figure 1. KDD method

The KDD process begins with data selection, which comes from an Excel file containing 60 months of rice production information along with supporting variables such as temperature, rainfall, humidity, air pressure, wind speed, and sunshine, as well as the type of fertilizer used. This data was chosen due to its relevance to rice production, which allowed this study to focus on the elements that are most significant in prediction.

Next, a data pre-processing stage is conducted to ensure the data is ready for use in the analysis. This step includes imputation of blank values using the averaging method, format adjustment for time series analysis, and data cleaning to address outliers or invalid values. Thus, the data becomes more consistent and can be analyzed further. Once the data is ready, the model selection stage is conducted, where the ARIMA (Autoregressive Integrated Moving Average) algorithm is chosen due to its ability to predict time series data relevant to the purpose of this study.

The last stage involves training the model using the training data for the first 48 months, followed by evaluating the model to measure its performance. This evaluation is done by comparing the prediction results with the test data using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), as well as correlation analysis to determine the factors that most affect rice production. This research uses Lenovo Ideapad Slim hardware with AMD Ryzen 5, Windows 11 operating system, and Python programming tools and language in Google Colab for model implementation.

3. RESULTS AND DISCUSSION

3.1 Data Selection

This study used data from sixty months of rice production, which included rice production variables in kilograms as well as eleven supporting factors, including environmental factors including temperature, rainfall, humidity, air pressure, wind speed, and sun exposure. The following fertilizers were used: UREA, SP-36, ZA, NPK, and organic fertilizer. These data were collected from the Sukabumi District Agriculture Office and BMKG Pelabuhanratu, and compiled in a format suitable for analysis. As a result of previous literature research, the selection of variables showed that these components have a significant impact on rice production yield. Here are the top 5 data

No.	Month	Year	Temperature	Rainfall	Sunlight	Pressure	Humidity	Wind Speed
1	January	2020	27.6	11.3	41	1004.0	96	6.5
2	February	2020	27.8	13.0	43	1004.3	97	7.2
3	March	2020	27.7	22.9	45	1004.1	95	7.0
4	April	2020						

UREA	SP-36	ZA	NPK	ORGANIC	Production
224.723	59.072	5.200	642.500	4.500	2082
224.723	41.107	4.500	620.000	3.800	1600
314.612	163.737	4.300	675.000	4.300	1697
359.557	72.510	4.800	630.500	3.900	1646

3.2 Data Pre-processing

To ensure that the data used in the model has no missing values, the averaging method is used to fill in the blank values in the numerical column. This is very important for proper analysis and prediction. Addition of Time Identification Column: To provide consecutive time numbers for use in analysis and visualization, an additional column "Month to" was added. The processed data was then divided into two groups: training data for the first forty-eight months and test data for the last twelve months.

The Augmented Dickey-Fuller (ADF) test results show a statistical value of -6.022, with a p-value of 1.48e-07, which is smaller than 0.05. This indicates that the data is stable and ready to be used in ARIMA modeling without additional differencing.

3.3 Model Selection

The best parameter used in the ARIMA model is ARIMA(0, 2, 2), with a differencing level of (d d) 2, the number of autoregressive lags (p p) 0, and the number of moving average lags (q q). This model is designed to handle data that requires two differencing times to achieve stationarity, without an autoregressive component but with two lags for the Moving Average component.

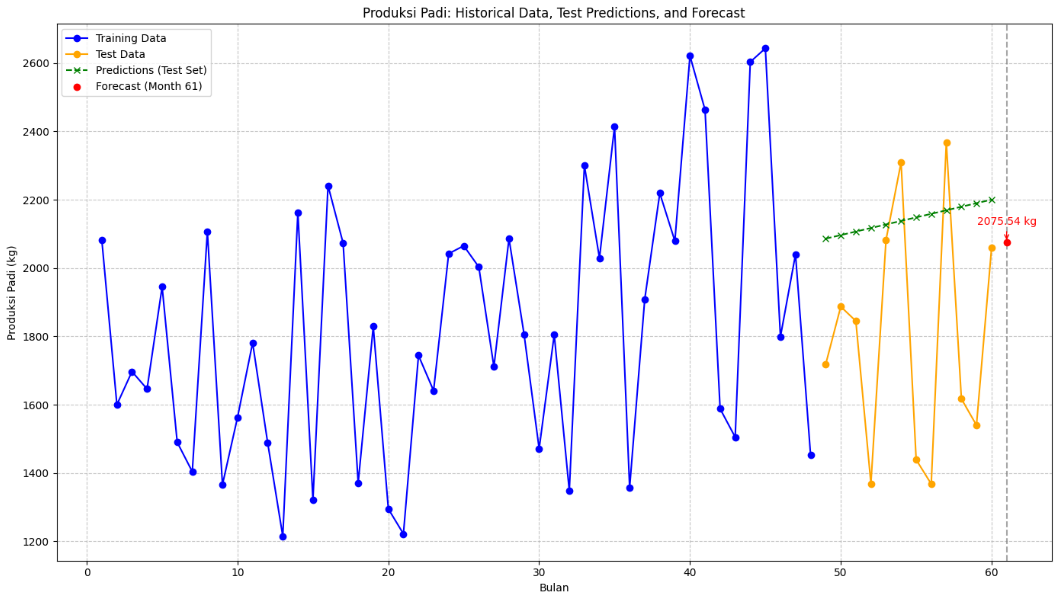

Figure 2. Prediction Chart

3.4 Model Training

An Akaike Information Criterion (AIC) value of 690.832 and a Log Likelihood of -342.416 were found in the training results of the ARIMA (0, 2, 2) model, indicating that the model has a good fit. Statistically, the moving average (MA) parameters are significant: ma.L1 = -1.9932 (p-value < 0.000) and ma.L2 = 0.9933 (p-value < 0.001). In addition, the residual variable (sigma^2) of 133.100 is significant. The Q value = 0.27 (p-value = 0.60) from the Ljung-Box diagnostic test indicates that there is no significant autocorrelation in the residuals. The Jarque-Bera test, with a p-value of 0.29, indicates that the residuals are close to a normal distribution, and the kurtosis of 1.87 and skewness of 0.02 indicate that the residual distribution is symmetric. Thus, the ARIMA (0, 2, 2) model can be considered a feasible prediction.

3.5 Model Evaluation

The results of evaluating the performance of the ARIMA model with test data show an acceptable level of prediction accuracy. The Mean Square Error (MSE) value is 230356.64, the Mean Square Error (RMSE) value is 479.95, and the Absolute Mean Error (MAE) value is 404.62. This is the average value of the absolute error between the actual and predicted values. The results of this evaluation show that the ARIMA model can provide a good prediction of rice production with a relatively small error rate. In projecting data production results for the future, these evaluation values support the use of the model.

Table 1. Prediction Table

Month to	Actual Production	Predicted Production
49	1718.0	2085.919580
50	1887.0	2096.295430
51	1845.0	2106.671280
52	1369.0	2117.047129
53	2082.0	2127.422979
54	2309.0	2137.798829
55	1439.0	2148.174678
56	1368.0	2158.550528
57	2368.0	2168.926377
58	1617.0	2179.302227
59	1540.0	2189.678077
60	2059.0	2200.053926
61	NaN	2075.543731

The prediction results show that rice production is projected to be 2075.54 kg in the 61st month. In the prediction result graph, the historical data (training data) is shown in blue, the test data is shown in orange, and the prediction for the test data is shown with a green dotted line. On the graph, the red dot indicates the 61st month prediction of 2075.54 kg, which is followed by an annotation line to help understand the results. The transition between the historical data and the 61st month prediction is shown by the gray vertical line. This graph provides a clear picture of how the ARIMA model works to study past patterns of rice production as well as the ability to predict future values.

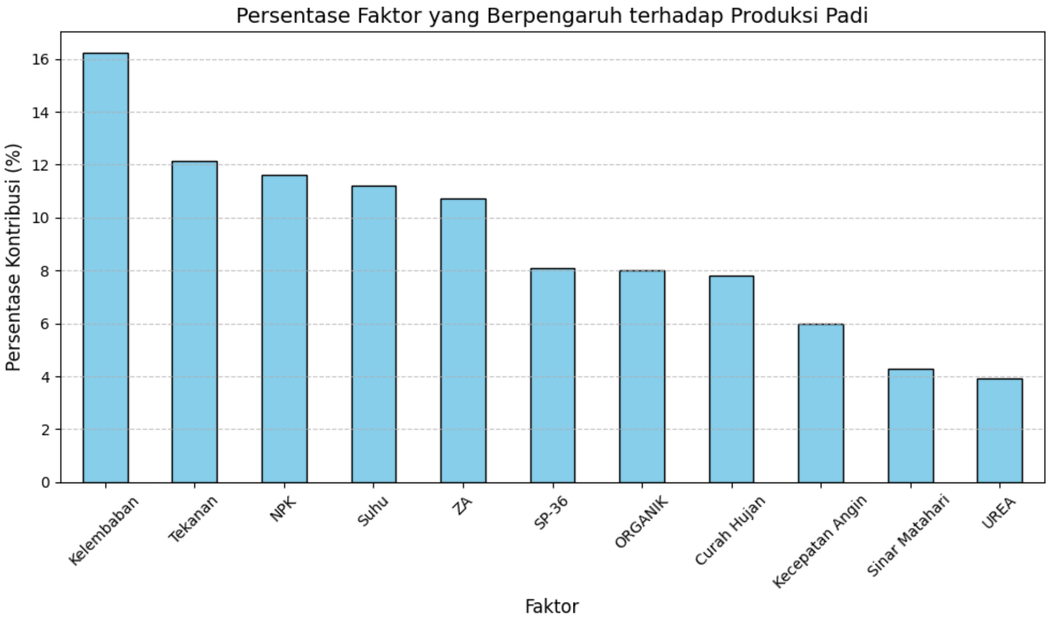

Figure 3. Percentage Diagram of Influential Factors

Moisture is the most significant factor affecting rice production with a contribution of 16.21%. Although there is a negative correlation, it shows that too much moisture can affect yield, so moisture management is very important.

In addition, atmospheric pressure contributes 12.13% to rice production, as it aids plant growth and increases yield.

4. CONCLUSION

This study utilizes an ARIMA model to analyze and predict rice production based on 60 months of data. For immediate use in modeling, the stationarity test results show that the data is stationary. The Akaike Information Criterion (AIC) value of 690.832, which indicates a high degree of fit, suggests that the ARIMA(0, 2, 2) model is the best. The model training results show that the Moving Average (MA) parameter is statistically significant. In addition, the diagnostic results show that the model residuals have no autocorrelation and have a more normal distribution.

With a Mean Squared Error (MSE) of 230356.64, Root Mean Squared Error (RMSE) of 479.95, and Mean Absolute Error (MAE) of 404.62, the model evaluation shows a relatively small error rate on the test data. In addition, the model predicts rice production in the 61st month to be 2075.54 kg. By looking at the prediction results, you can have a clear understanding of past production patterns, test data predictions, and predictions for the future. Therefore, the ARIMA(0, 2, 2) model is considered feasible to be used to forecast rice paddy production.

The results of the study allow for the following recommendations. Given that climate and fertilizer factors greatly affect rice production, it is necessary to add other factors such as pesticide data, irrigation, and rice varieties. To optimize data processing, a lot of data is needed such as hundreds or even thousands so that the model used can provide the best performance in processing data. Similar research can be conducted in other places to measure how well the model adapts to data with various characteristics.

REFERENCES

[1] Abdurrozzaq Hasibuan, Suhela Putri Nasution, Fitri Amja Yani, Henni Adlini Hasibuan, and Nyak Firzah, “Strategi Peningkatan Usaha Tani Padi Sawah Untuk Meningkatkan Perekonomian Masyarakat Desa,” *ABDIKAN J. Pengabdi. Masy. Bid. Sains dan Teknol.*, vol. 1, no. 4, pp. 477–490, 2022, doi: 10.55123/abdikan.v1i4.1095.

[2] M. Bagas Septyono, A. Diki Prasetyo, B. Riza Pratama, A. Puspita Sari, and U. Veteran Jawa Timur, “Penerapan Metode Mamdani Untuk Prediksi Produksi Padi di Pulau Sumatra,” *Semin. Nas. Inform. Bela Negara*, vol. 3, pp. 2747–0563, 2023.

[3] F. Perdhana and I. Noviana, “Keragaan Agronomis Vub Padi Spesifik Lokasi Di Kabupaten Sukabumi,” *Pros. Semin. Nas. Has. Penelit. Agribisnis VI*, vol. 6, no. 1, pp. 241–247, 2022.

[4] H. Husna, B. Bakhtiar, and C. N. Ichsan, “Pengaruh Suhu, Pemupukan K dan N terhadap Pertumbuhan Tanaman Padi Inpari 30 (Oryza sativa L.),” *J. Ilm. Mhs. Pertan.*, vol. 6, no. 4, pp. 81–90, 2021, doi: 10.17969/jimfp.v6i4.18715.

[5] M. Habib Rifai, N. Vera, N. Sekar, S. Dewi, and R. R. Narfandi, “Prototipe Alat Pengukur Kelembaban Tanah Berbasis Sensor Media Tanaman Padi,” *J. Elektron. List. dan Teknol. Inf. Terap.*, vol. 5, no. 1, pp. 16–21, 2023.

[6] et al. Nurhidayat, “Aldy+Nurhidayat+1181,” vol. 2, no. 2, pp. 111–117, 2024.

[7] F. Azizah, S. Suwarsito, and E. Sarjanti, “Pengaruh Pola Curah Hujan Terhadap Produktivitas Padi di Kecamatan Bukateja Kabupaten Purbalingga,” *Sainteks*, vol. 18, no. 1, p. 1, 2021, doi: 10.30595/sainteks.v18i1.10567.

[8] S. Mayasari, Sudiarti, and Yushardi, “Analisis Hubungan Intensitas Panas Energi Matahari dengan Proses Fotosintesis pada Tanaman Padi,” *Mekanikal,InovasidanTeknologi*, vol. 9, no. 1, pp. 2477–5029, 2023.

[9] Henny Puspita Sari and S. K. Sari, “Analisis Neraca Air Untuk Tanaman Padi Dalam Upaya Adaptasi Dan Mitigasi Bencana Terhadap Kelompok Rentan,” *Ekasakti J. Penelit. Pengabdi.*, vol. 3, no. 2, pp. 156–170, 2023.

[10] D. Wulandari and R. Rumini, “Pemodelan dan Prediksi Produksi Padi Menggunakan Regresi Linear,” *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 12, no. 4, 2023, doi:

10.30591/smartcomp.v12i4.5905.

[11] R. Kurniawan, A. Halim, and H. Melisa, “Prediksi Hasil Panen Pertanian Salak di Daerah Tapanuli Selatan Menggunakan Algoritma SVM (Support Vector Machine),” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 2, pp. 903–912, 2023, doi: 10.30865/klik.v4i2.1246.

[12] A. T. Nurani, A. Setiawan, and B. Susanto, “Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma,” *J. Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.