The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025
"From Data to Impact: Leveraging Innovation in Data Science and Official Statistics for Data-Driven Policy Making and Global Competitiveness"

Forecasting Indonesian Monthly Rice Prices at Milling Level Using Google Trends and Official Statistics Data

I B P Swardanasuta[1], W A Sofa[1], S Muchlisoh[1], and A W Wijayanto[1,*]

[1] Politeknik Statistika STIS, Jakarta, Indonesia

*Corresponding author's email: ariewahyu@stis.ac.id

Abstract. Hunger is a very complex social issue to address. Alleviating hunger is closely related to achieving food security, which is a goal in realizing the second Sustainable Development Goals (SDGs), zero hunger. The most frequently consumed food commodity by the Indonesian population is rice, which has fluctuating prices in the market. Therefore, price forecasting is necessary so that the government can take preventive measures against rice price increases at certain times. Research on rice price forecasting using big data from Google Trends is still very rare in Indonesia, even though Google Trends has great potential to reflect the public's search popularity for certain keywords. Therefore, this study aims to forecast the monthly medium rice price in Indonesia at the milling level using exogenous variables of dried milled grain prices and the popularity index of related keywords on Google Trends. The forecasting is conducted using Seasonal Autoregressive Integrated Moving Average (SARIMA), SARIMA with Exogenous Variables (SARIMAX), and Extreme Gradient Boosting (XGBoost) models. The SARIMAX model has the best performance in forecasting rice prices, with a Root Mean Squared Error (RMSE) of 941.6933, Mean Absolute Error (MAE) of 817.9021, and Mean Absolute Percentage Error (MAPE) of 0.0620.
Keyword: Google Trends, Rice prices, SARIMA, SARIMAX, XGBoost.

1. Introduction

Food security is a target to be achieved by the Indonesian government with the aim of ensuring that all levels of society have easy access to food. Food is an essential need for humans because good quality food will improve the quality of human resources. In 2023, the prevalence of undernourishment in Indonesia was 8.53 percent [1]. The prevalence of undernourishment describes the percentage of the population whose source of energy consumption from food is not sufficient for normal activities. In other words, as much as eight percent of the Indonesian population consumes food that is not sufficient for daily activities. This condition is a concern for the government because one of the goals to be achieved in the Sustainable Development Goals (SDGs) is zero hunger [2]. This goal has been the focus of the Indonesian government for the past few years because hunger is closely related to the welfare of the population. The Medium-Term National Development Plan (RPJMN) 2020–2024 prioritizes food security and poverty alleviation [3]. The 2020 Global Hunger Index report noted that Indonesia ranked 70th out of 107 countries experiencing difficulties in meeting psychological needs related to food and nutrition, with an index score of 19.1 [4]. The government and the people of Indonesia must work together in alleviating hunger to achieve SDGs point 2.

The 3rd International Conference on Data Science and Official Statistics November 27 - 28, 2025

Hunger is a complex problem because it is caused by many factors and occurs in areas that are difficult to reach. Since 2015, hunger has become an alarming condition that has become more severe due to pandemics, conflict, social inequality, climate change, scarcity of food commodities, and higher food prices [5]. One of the factors that result in many people still going hungry is the high price of food so that people are unable to meet their food needs at high prices. The role of the government as a stakeholder who plays a role in monitoring food prices in the market. Rice is one of the food commodities with a high level of consumption by the Indonesian people and its price fluctuates greatly because it is influenced by regional, social and economic conditions. Over the past few years, rice prices in Indonesia have experienced price instability characterized by increasing and decreasing prices at certain times [6]. One solution to overcome the problem of fluctuating rice prices in Indonesia is to conduct forecasting to detect early the potential for an unnatural increase in national average rice prices so that the government can anticipate it with certain policies. The rice prices vary spatially across Indonesia, with generally higher prices in eastern regions of Indonesia due to climatic conditions [7].

Based on the problem of unstable food prices, especially rice, price forecasting can be done based on the past prices of these goods. In addition, the development of advanced technology has made it easier for people to obtain information. This advancement is often utilized by the public to find information about food prices through the internet. Through this information search, technology allows us to analyze search patterns on the internet using big data. The machine learning approach has been widely used in several previous studies to predict rice prices in order to overcome unpredictable crises [8]. Price fluctuations in agricultural commodities can be analyzed using machine learning to provide appropriate policy recommendations to stakeholders [9]. Forecasting the prices of staple foods using machine learning algorithms such as random forest and XGBoost algorithms excellent results, even better than traditional methods [10,11]. Beside that, machine learning models such as LSTM and XGBoost are effective at capturing dynamic patterns in agricultural commodity prices [12,13].

Google Trends is an application of big data that can be used to analyze people's search patterns. Google Trends has the potential to become an analytic tool used for forecasting and consideration in decision making [14]. In addition, Google Trends data has been used to forecast chili prices using official statistics and Google Trends data in the form of keywords related to chili prices that have a significant effect on chili prices [15].

Until now, research on rice price forecasting in Indonesia using big data is very rarely done even though it has enormous potential as a consideration for the government in planning policies related to rice prices. Therefore, this research aims to forecast the monthly price of Indonesian rice by utilizing Google Trends data and official statistics. This research aims to find the best model in forecasting the monthly price of Indonesian rice and forecasting the monthly price of Indonesian rice until June 2024. Therefore, the results of this study can be useful for the community to estimate rice expenditure and assist the government in making the right decisions.

2. Research Method

2.1. Data source

This study uses time series data obtained from secondary sources. The data used in this study has a monthly period from January 2013 to February 2024. The data consists of 134 rows, which represent data for each month. The sample size is sufficient for time series forecasting, as it covers over 11 years of data and exceeds the minimum recommendation of 50 observations suggested by Box and Jenkins [16]. Keyword data related to rice prices from Google Trends is taken using web scraping techniques and combined with taking directly from the Google Trends web. The official statistics data used are monthly rice prices and monthly grain prices sourced from the Statistics Indonesia, through the Rice Producer Price Survey at Rice Mills every year [17]. The variable descriptions used in this study are described in Table 1.

Table 1. Research variables

Variable	Description	Unit	Source
Rice price	Average monthly price of medium quality rice at milling level	Rp/kg	Statistics Indonesia
Milled dry grain price (GKG)	Average monthly milled dry grain price at farm level	Rp/kg	Statistics Indonesia
"beras" keyword	Search index of the keyword "beras" on Google Trends	-	Google Trends
"harga beras" keyword	Search index of the keyword "harga beras" on Google Trends	-	Google Trends
"sembako" keyword	Search index of the keyword "sembako" in Google Trends	-	Google Trends

2.2. *Analysis Method*

In analyzing monthly rice price data, several methods are needed that can handle price movements that change throughout the time period. The analysis methods used in this research are the classic time series models, namely SARIMA and SARIMAX and the machine learning model, namely XGBoost.

2.2.1. SARIMA

Seasonal autoregressive integrated moving average (SARIMA) is a time series model introduced by Box and Jenkins that is used on data with trends and seasonal components [18]. The data used in the SARIMA model must meet the assumption of stationarity. The SARIMA model is written with the notation ARIMA (p,d,q) $(P,D,Q)_S$ which follows equation (1).

$$\phi_p(B)\Phi_P(B^S)Y_t = \alpha + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \tag{1}$$

2.2.2. SARIMAX

SARIMA with exogenous regressors (SARIMAX) is an extension of SARIMA that includes exogenous variables in time series modeling [19]. The SARIMAX model is written with the notation ARIMA (p,d,q) $(P,D,Q)_S$ plus exogenous variables following equation (2).

$$\phi_p(B)\Phi_P(B^S)Y_t = \alpha + \beta_k X_{k,t} + \theta_q(B)\Theta_Q(B^S)\varepsilon_t \tag{2}$$

2.2.3. XGBoost

Extreme gradient boosting (XGBoost) is a machine learning model that builds new trees iteratively and combines the trees for forecasting results. XGBoost has feature importance to see the significance of variables in prediction [19].

2.3. *Model evaluation*

In evaluating the models used to find monthly rice price patterns, this study uses several evaluation metrics to compare several models obtained. The evaluation metrics used are the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) following equations (3), (4), and (5) [20,21].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \quad (y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{n} \quad |y_i - \hat{y}_i|}{n} \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \quad \left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

3. Result and Discussion

3.1. Overview of Indonesia's Monthly Rice Price Trends 2013-2024

During the period 2013-2021, Indonesia's monthly rice prices at the mill level have an increasing trend despite several price drops in certain months. Figure 1 shows that monthly rice prices from January 2013 to mid-2022 experienced a fairly constant price increase. Since August 2022, the price of rice has increased considerably compared to the previous month as indicated by the steeper line. This indicates that there are other factors that have caused the increase in rice prices since mid-2022 until now. Figure 1 also shows a seasonal pattern in rice prices with a high increase every year around October or November. This pattern is a consideration in the use of SARIMA and SARIMAX models due to the seasonal pattern in the data.

This study uses the monthly price of milled dry grain (MDG) at the farm level as a variable that is thought to affect the price of rice. During 2013-2024, the price of unhusked rice also increased like the price of rice. From January 2013 to mid-2021, the price of GKG experienced a fairly steady increase and experienced several decreases in certain months, as shown in Figure 1. Since August 2021, the price of GKG has increased dramatically from the previous period. This condition is in line with the increase in rice prices that occurred in the same period. The increase in the price of GKG at the farm level indicates that the price of rice production at the farm level has also increased, leading to an increase in the price of rice.

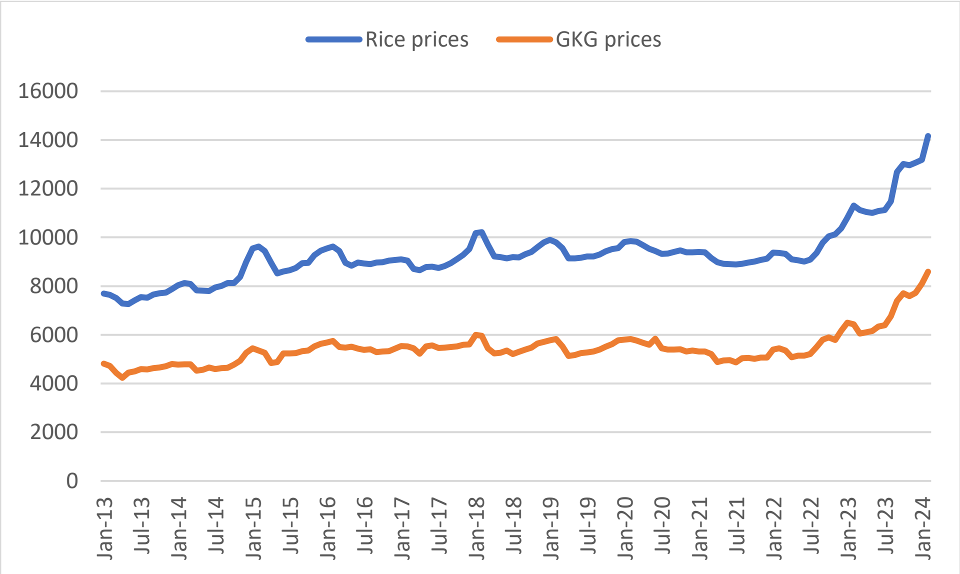

Figure 1. Monthly rice and milled dry grain prices trends, 2013-2024

To perform monthly rice price forecasting, this research uses the popularity factor of keyword searches related to rice prices from Google Trends which is included in big data. There are 3 keywords used in this research that are thought to have a strong correlation with rice prices, namely the keywords “beras”, “harga beras”, and “sembako”. The selection of these keywords is based on words related to rice prices in Indonesia. In addition, the selection of these keywords refers to research [15] which conducted chili price forecasting in Indonesia. The popularity of these three keywords has an increasing trend since 2013 as shown in Figure 2. This increasing trend is an initial assumption that these three keywords can help forecast rice prices with a big data approach. These keyword searches indicate that people respond to rising rice prices by searching for information related to rice prices on the internet. Of the three keywords used, all three have similarities, namely the highest search index in 2020 and 2021 which occurred during the Covid-19 pandemic. This search is related to the difficult economic conditions at that time due to rising food prices.

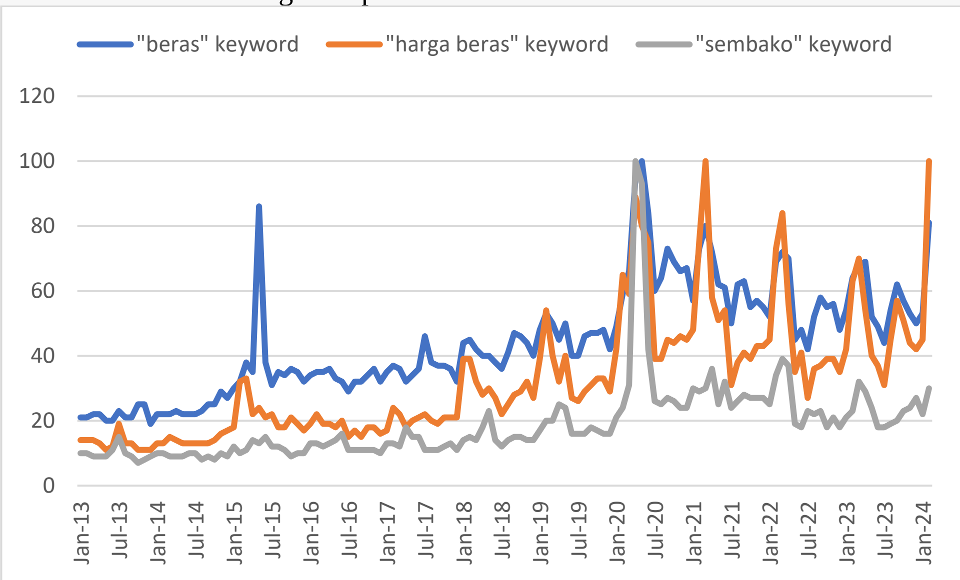

Figure 2. “beras”, “harga beras”, “sembako” keywords search trends, 2013-2024

The summary statistics in Table 2 shows the distribution of data for each variable used. The price of rice has a considerable difference between the minimum and maximum values, while the price of milled dry grain tends to be homogeneous when compared to the price of rice. In addition, the Google Trends keyword index used shows that the keywords “harga beras” and “sembako” have a smaller average index than “beras”.

3.2. Data preparation

Data that has been collected completely cannot be directly analyzed because some adjustments are needed. In this study, all variables used underwent a transformation process into natural logarithm form. The purpose of this data transformation is to avoid poor forecasting results because the size of each variable is quite spread. Data transformation using natural logarithms was also carried out in research [22] to predict carbon dioxide emissions because the variables used had different units of measurement.

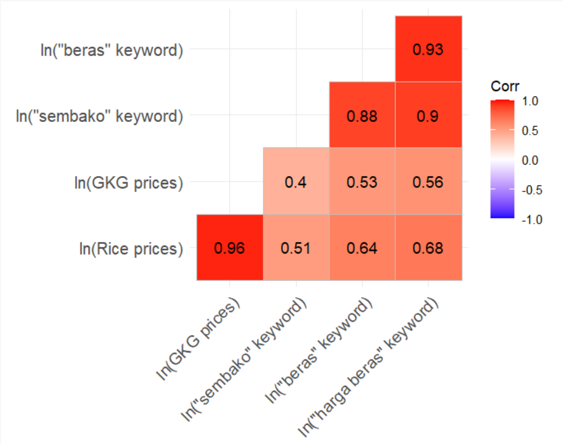

Figure 3. Correlation between transformed variables

The data transformation used did not change the overall pattern of the data and only made the data scale smaller and more uniform. The transformed data was then examined for correlations between variables to identify variable relationships. Figure 3 shows that all variables have a strong correlation with rice prices with correlation values above 0.5.

After performing the transformation process, the data is ready to be used for the modeling process. Before modeling, the data is first divided into training and testing data. Data from January 2013 to June 2023 became training data, while data from July 2023 to February 2024 became testing data. Training data is used to build a forecasting model and testing data is used to evaluate the performance of the model in forecasting.

Table 2. Summary statistics of variables used

	Harga beras	Harga GKG	"beras" keyword	"harga beras" keyword	"sembako" keyword
Minimum	7262	4232	19	11	7
1st quartile	8848	5067	32	18	11
Median	9160	5385	42	29	15
Mean	9297	5446	45	33	19
3rd quartile	9548	5616	55	42	24
Maximum	14162	8591	100	100	100

The summary statistics in Table 2 shows the distribution of data for each variable used. Rice prices vary considerably between the minimum and maximum values, while milled dry grain prices tend to be more homogeneous compared to rice prices. In addition, the Google Trends keyword index used shows that the keywords “harga beras” and “sembako” have a lower average index than “beras”.

3.3. Stationarity test

Modeling rice price forecasting with classical time series models is done using SARIMA and SARIMAX. The data used in this model must meet stationarity assumptions to avoid spurious regression. Stationarity testing is done with graphs and the Augmented Dickey-Fuller test (ADF test).

The ACF plot of rice prices in Figure 4 shows that rice prices are not yet stationary at the level as they are exponentially decreasing. In addition, the PACF plot of rice prices in Figure 8 shows that there is an exponential decline in lags 12, 24, and 36 which indicates that there is a seasonal pattern in the data. The stationarity test results using the ADF test in Table 3 shows that the rice price variable is not yet stationary at the level because the *p-value* < 0.05. The exogenous variables of milled dry grain price and the keyword "beras" are also not stationary at the level because the *p-value* < 0.05.

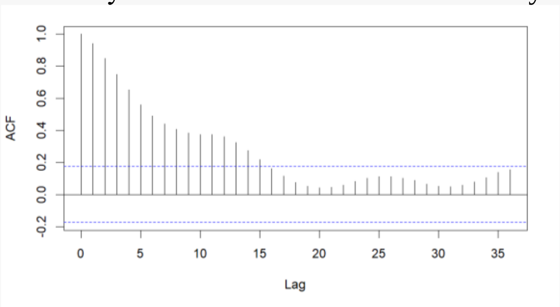

Figure 4. ACF plot of rice price

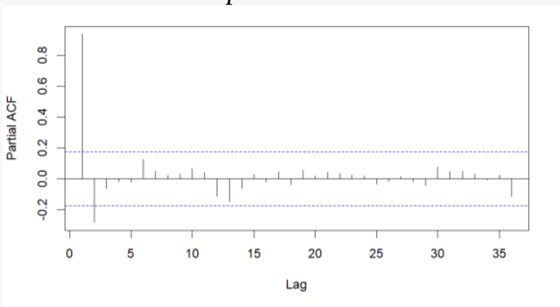

Figure 5. PACF plot of rice price

Since the data at the level has not met the assumption of stationarity, further differencing is performed on the data to achieve stationarity. The results of the stationarity test on the first difference produce a *p-value* < 0.05 for each variable so that the data has met the assumption of stationarity in the first difference. Therefore, forecasting modeling uses first difference data that is already stationary.

Table 3. Stationarity test results

Variable	*p-value* ADF test (prob.)	
	Level	First difference
Rice prices	0.2375	< 0.01
GKG prices	0.4165	< 0.01
"beras" keyword	0.3768	< 0.01
"harga beras" keyword	0.0123	< 0.01
"sembako" keyword	0.0487	< 0.01

3.4. SARIMA modeling

By using data that has been stationary in the first difference, modeling is done with SARIMA. Parameter determination in the model is done by trying several possible models to produce the best model. Forecasting modeling is done using training data that has been determined at the data preparation stage. The best SARIMA model in forecasting monthly rice prices is ARIMA (2,1,0) $(2,0,0)_{12}$.

Table 4. SARIMA parameter estimation results

	Estimate	Std. Error	z value	*p-value*
AR(1)	0.5330	0.0938	5.6835	0.0000
AR(2)	-0.1623	0.0912	-1.7801	0.0751
SAR(1)	0.3059	0.0939	3.2563	0.0011
SAR(2)	0.2250	0.0984	2.2876	0.0222

The parameter estimation results of the ARIMA (2,1,0) $(2,0,0)_{12}$ model are shown in Table 4. Based on the four estimated parameters, the results show that AR (1), SAR (1), and SAR (2) are significant at the 5% significance level. This result indicates that the monthly Indonesian medium rice price at the milling level has a seasonal pattern so all seasonal parameters are significant at $\alpha = 5\%$.

Table 5. SARIMA residual assumption testing results

Test	*p-value*
Normality (Kolmogorov-Smirnov)	0.5487
Ljung-Box	0.8961

In forecasting time series data, there are assumptions that must be met to state that the forecasting model can be used. The residuals generated by the model must be normally distributed and fulfill the white noise assumption [23]. Testing the normality of residuals using the Kolmogorov-Smirnov test written in Table 5 with a p-value of 0.5487 which indicates that the residuals are normally distributed because the *p-value* > 5%. For testing the white noise assumption using the Ljung-Box test with a *p-value* of 0.8961 which indicates that the residuals have met the white noise assumption because the *p-value* > 5%. Thus, the ARIMA (2,1,0) $(2,0,0)_{12}$ model can be used because it has met all the assumptions and is the best SARIMA model for forecasting rice prices.

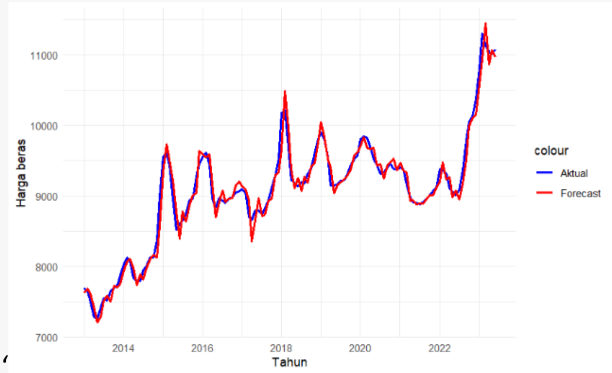

Figure 6. Plot of actual and forecast data of SARIMA rice price

The ARIMA (2,1,0) $(2,0,0)_{12}$ model is the best SARIMA model from several candidate models. The comparison between the actual rice price in the training data and the forecast results are visualized in Figure 6. The plot shows that the SARIMA model can forecast rice prices well, which is indicated by the training and forecast line patterns are almost the same.

3.5. SARIMAX modeling

The SARIMAX model is a time series model extension of SARIMA, but SARIMAX incorporates exogenous variables into the model to perform forecasting. SARIMAX modeling is done by estimating the model parameters by trying several possibilities to get the best model. In forecasting monthly rice prices in this study, the SARIMAX model with the notation ARIMA (0,1,1) $(2,0,0)_{12}$ is the best model.

Table 6. SARIMAX parameter estimation results

	Estimate	**Std. Error**	***z-value***	***p-value***
MA(1)	0.4135	0.0833	4.9631	0.0000
SAR(1)	0.1610	0.0949	1.6960	0.0899
SAR(2)	0.3039	0.0971	3.1286	0.0018
GKG price	0.2975	0.0522	5.7011	0.0000
“beras” keyword	-0.0105	0.0076	-1.3847	0.1661
“harga beras” keyword	0.0040	0.0077	0.5277	0.5977

"sembako" keyword	0.0136	0.0059	2.3250	0.0201

Table 6 shows the parameter estimation results for ARIMA(0,1,1) $(2,0,0)_{12}$. From the estimation results, the search keyword "beras" has a negative influence on the price of rice. This illustrates that when people's interest in searching for the keyword "beras" decreases, the price of rice increases. It also provides information that when the popularity of searching for "beras" decreases, the price of rice gets lower. This condition can occur when people are busy searching for rice information on the internet when the price of rice drops. However, "harga beras" and "sembako" keyword searches have a positive influence on the price of rice, which means that when the popularity of "harga beras" and "sembako" increases, the price of rice also increases, and vice versa. Based on Table 6, the parameter SAR(2), the coefficient of milled dry grain price, and the keyword "sembako" have a significant effect on the monthly rice price at the 5% significance level.

Table 7. SARIMAX residual assumption testing results

Test	*p-value*
Normality (Kolmogorov-Smirnov)	0.0547
Ljung-Box	0.7083

After obtaining the best SARIMAX model, a residual assumption check must be performed. Based on Table 7, testing the normality of residuals using the Kolmogorov-Smirnov test has a *p-value* of 0.0547 so that the residuals are normally distributed because the *p-value* > 5%. For testing the white noise assumption using the Ljung-Box test with a *p-value* of 0.7083 so that the residuals have met the white noise assumption because the *p-value* > 5%.

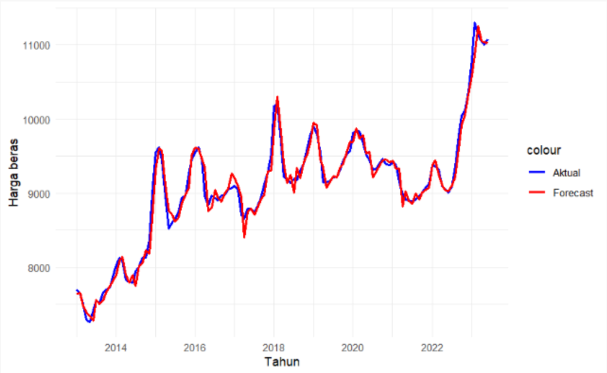

Figure 7. Plot of actual and forecast data of SARIMAX rice price

Figure 7 provides a visualization of the comparison of actual data with the forecast results of the SARIMAX model with ARIMA(0,1,1) $(2,0,0)_{12}$. The plot shows the similarity between the actual rice prices and the forecasting results. This indicates that the SARIMAX model using the Google Trends keyword exogenous variable has a good influence on the performance of the resulting model.

3.6. XGBoost modeling

Modeling with XGBoost requires some hyperparameter adjustments to obtain better forecasting results. Hyperparameter selection for modeling is done by tuning the hyperparameters first. The best hyperparameter tuning results are presented in Table 8.

Table 8. XGBoost hyperparameter tuning results

Parameter	Value
nrounds	157
max depth	20
eta	0.1
gamma	0
colsample bytree	1
min child weight	1
Subsample	1

The parameter tuning results are used in creating the XGBoost model to forecast monthly rice prices. Based on Figure 8, the XGBoost model has a very good performance for forecasting the training data as seen from the actual data lines and forecasts that are very similar so that the resulting forecasting error is also small. This can happen because the XGBoost model learns the data pattern more deeply so that the model building process is also longer than SARIMA and SARIMAX. The forecasting results with XGBoost obtained in this study are also in line with the results of research [24] which found that XGBoost has excellent performance to find patterns of electricity consumption and prices in depth with faster processing time than other machine learning methods.

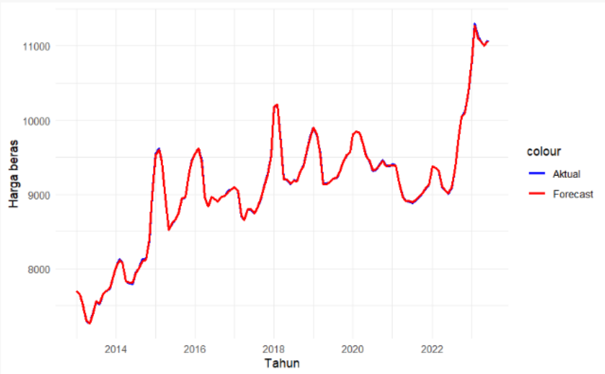

Figure 8. Plot of actual and forecast data of XGBoost rice price

One of the most important features of XGBoost is the importance feature, which can be used to see which variables have the most influence on the target variable. Figure 9 shows that the price of milled dry grain is the most important variable in forecasting rice prices. The keyword “harga beras” is the second most important variable in forecasting because “harga beras” is the most suitable keyword for rice price forecasting as evidenced by the XGBoost model.

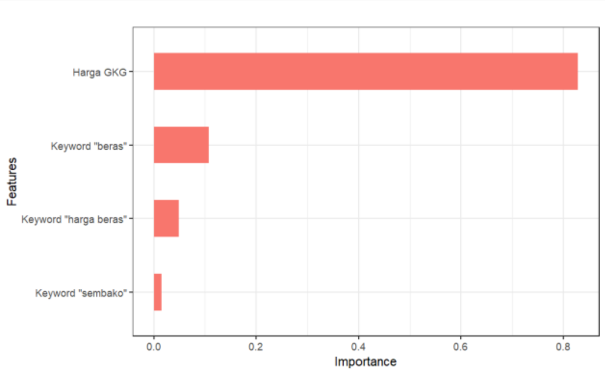

Figure 9. Feature importance of XGBoost model

3.7. Best model selection

The next step in this research after obtaining the three existing models is to evaluate the performance of the three models in forecasting the testing data. Table 9 provides information on the performance of the three models for forecasting rice prices using the RMSE, MAE, and MAPE evaluation metrics. The best model for forecasting is the model that has the smallest value in all three evaluation metrics. The SARIMAX model is the best model compared to the other two models because it has the lowest value for all three evaluation metrics, with RMSE of 0.0738, MAE of 0.0646, and MAPE of 0.0068. Thus, the SARIMAX model was chosen as the best model to forecast Indonesia's monthly rice prices at the mill level.

The selection of the best model resulted in the SARIMAX model with the notation ARIMA (0,1,1) $(2,0,0)_{12}$ being the best model. The parameter estimation results of the SARIMAX model explain that the price of milled dry grain and the keyword “sembako” are variables that have a significant effect on the price of medium rice at the milling level. The price of milled dry grain has a very significant influence on the price of medium rice at the milling level. This result is in line with previous research which found that changes in rice prices at the milling level can be significantly predicted by grain prices at the farm level [25]. High grain prices at the farm level indicate that rice production costs are also high. Rice production is closely related to rice maintenance, such as fertilizer application. This is information that can be provided to the government to implement the right policy in providing subsidies to farmers in buying fertilizers or other rice production materials. Thus, the price of production at the farm level can be suppressed, which leads to a decrease in the price of grain so that the price of rice at the milling level is not too high.

Table 9. Evaluation of model performance on testing data

Model	RMSE	MAE	MAPE
ARIMA (2,1,0) $(2,0,0)_{12}$ [SARIMA]	1296.9970	1130.0288	0.0857
ARIMA (0,1,1) $(2,0,0)_{12}$ [SARIMAX]	941.6933	817.9021	0.0620
XGBoost	2088.4154	1918.5934	0.1470

Based on the results, the SARIMAX model outperformed XGBoost on testing data. These results are consistent with Thejovathi et al. [26], which were due to non-stationary data at the level and the SARIMA model being able to capture seasonal patterns and trends in the data. Meanwhile, XGBoost makes more varied predictions and is less effective at capturing the temporality of the data. The results of this study are also consistent with Obaidat et al. [27], which found that the SARIMAX model is better at forecasting demand in the dairy industry. SARIMAX is better because it can capture seasonal demand well and the integration of exogenous variables improves the accuracy of forecasting while maintaining interpretability [27]. The bias-variance tradeoff produced by the SARIMAX model is superior to flexible machine learning models such as XGBoost, which are sensitive to variance [28]. However, the results of this study are not in line with Hossain et al. [29], which found that machine learning models are better than SARIMA because the amount of data is still relatively small for machine learning models, resulting in the XGBoost model being unable to capture the overall data patterns properly.

The search keyword variable “sembako” also has a significant effect on the price of rice. This shows that the big data obtained through Google Trends has a significant relationship and influence on the price of rice in Indonesia. The coefficient for the search “sembako” is positive so that the higher popularity of basic food searches on Google leads to higher rice prices in Indonesia. This condition occurs because rice is part of the basic necessities that are most often sought after and consumed by the public. When searches about basic necessities on Google increase, there are signs that the price of rice will increase. People may search for information about basic necessities when there is social or economic turmoil that leads to an increase in the price of rice. Therefore, Google Trends can be a new source of data to analyze the movement of food commodity prices in Indonesia. In this case, big data becomes a new door for the source of evaluation of Indonesian government policies.

The advantage of using Google Trends is that it provides real-time search data, allowing the government to monitor monthly rice price developments even before official data is released. Google search trends among the public can serve as early indicators of economic trends [30]. The government can automatically collect Google Trends index data every month, thereby reducing its reliance on conventional surveys to produce official statistics. The SARIMA or SARIMAX model can be used by the government in the future to forecast data with a small number of observations and seasonal patterns. Meanwhile, machine learning models such as XGBoost are recommended for use when the number of observations is very large to extract hidden patterns in the data.

4. Conclusion

Rice is the main food commodity in Indonesia and its price fluctuates every month. The movement of rice prices must be monitored to maintain the stability of food prices so that people do not have difficulty in obtaining rice in order to achieve the SDGs goal of zero hunger. This research uses Google Trends data and official statistics to forecast Indonesia's monthly medium rice prices at the mill level. The models used for forecasting are SARIMA, SARIMAX, and XGBoost. The three forecasting models used provide good results in terms of evaluation metrics. To evaluate the model performance, the forecasting results on the testing data are measured by the RMSE, MAE, and MAPE metrics. Based on the three evaluation metrics used, the SARIMAX model with the notation ARIMA $(0,1,1)\ (2,0,0)_{12}$ which uses the exogenous variable of milled dry grain price as well as the Google Trends index for keyword searches “harga beras”, “beras”, and “sembako”. The SARIMAX model has an RMSE of 941.6933, MAE of 817.9021, and MAPE of 0.0620.

Based on the SARIMAX model selected as the best model, the exogenous variables of milled dry grain price and Google Trends index for the keyword search “sembako” have a significant effect on the monthly Indonesian medium rice price at the milling level. The government as a relevant stakeholder can provide assistance in the form of fertilizer subsidies or other rice production materials to reduce the price of milled dry grain at the farm level so that rice prices can be controlled. In addition, the big data source from Google Trends can help to do rice price forecasting which is an opportunity for further research in forecasting the prices of other food commodities so that it can provide appropriate input to the government to implement policies that support the achievement of zero hunger goals.

Acknowledgement

Thank you to Google for providing easy access to Google Trends data and thank you to everyone who helped complete this research.

References

[1] Statistics Indonesia. Prevalensi Ketidakcukupan Konsumsi Pangan (Persen), 2023 2024. https://www.bps.go.id/id/statistics-table/2/MTQ3MyMy/prevalensi-ketidakcukupan-konsumsi-pangan--persen-.html (accessed March 12, 2024).
[2] Goal 2: Zero Hunger - United Nations Sustainable Development n.d. https://www.un.org/sustainabledevelopment/hunger/ (accessed October 15, 2025).
[3] The Ministry of National Development Planning/Bappenas. Rancangan Teknokratik Rencana Pembangunan Jangka Menengah Nasional. 2019.
[4] Manurung ET, Maratno SFE, Permatasari P, Rahman AB, Qisthi R, Manurung EM. Do Village Allocation Funds Contribute towards Alleviating Hunger among the Local Community (SDG#2)? An Insight from Indonesia. Economies 2022, Vol 10, Page 155 2022;10:155. https://doi.org/10.3390/ECONOMIES10070155.
[5] United Nations. Goal 2: Zero Hunger 2023.
[6] Sukiyono K, Rosdiana R. Pendugaan Model Peramalan Harga Beras Pada Tingkat Grosir. Jurnal AGRISEP 2018;17:23–30. https://doi.org/10.31186/jagrisep.17.1.23-30.
[7] Putra AW, Supriatna J, Koestoer RH, Soesilo TEB. Differences in Local Rice Price Volatility, Climate, and Macroeconomic Determinants in the Indonesian Market. Sustainability 2021, Vol 13, Page 4465 2021;13:4465. https://doi.org/10.3390/SU13084465.
[8] Rathod S, Chitikela G, Bandumula N, Ondrasek G, Ravichandran S, Sundaram RM. Modeling and Forecasting of Rice Prices in India during the COVID-19 Lockdown Using Machine Learning Approaches. Agronomy 2022, Vol 12, Page 2133 2022;12:2133. https://doi.org/10.3390/AGRONOMY12092133.
[9] Patil A, Shah D, Shah A, Kotecha R, Somaiya KJ. Forecasting Prices of Agricultural Commodities using Machine Learning for Global Food Security: Towards Sustainable Development Goal 2. International Journal of Engineering Trends and Technology 2023;71:277–91. https://doi.org/10.14445/22315381/IJETT-V71I12P226.
[10] Theofilou A, Nastis SA, Michailidis A, Bournaris T, Mattas K. Predicting Prices of Staple Crops Using Machine Learning: A Systematic Review of Studies on Wheat, Corn, and Rice. Sustainability 2025, Vol 17, Page 5456 2025;17:5456. https://doi.org/10.3390/SU17125456.
[11] Swain NR, Choudhary V, Silswal S, Shastri AS, Spanogiannopoulos S. Crop Yield Prediction and Price Forecasting Using Machine Learning. Lecture Notes in Networks and Systems 2025;1201 LNNS:357–81. https://doi.org/10.1007/978-981-97-9839-1_22.
[12] Sari M, Duran S, Kutlu • Huseyin, Bulent Guloglu •, Atik Z. Various optimized machine learning techniques to predict agricultural commodity prices n.d. https://doi.org/10.1007/s00521-024-09679-x.
[13] Guo Y, Tang D, Tang W, Yang S, Tang Q, Feng Y, et al. Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors. Sustainability 2022, Vol 14, Page 10483 2022;14:10483. https://doi.org/10.3390/SU141710483.
[14] Silva ES, Hassani H, Madsen DØ, Gee L. Googling fashion: Forecasting fashion consumer behaviour using Google Trends. Soc Sci 2019;8. https://doi.org/10.3390/socsci8040111.
[15] Nasirudin F, Dzikrullah AA. Pemodelan Harga Cabai Indonesia dengan Metode Seasonal ARIMAX. Jurnal Statistika Dan Aplikasinya 2023;7.
[16] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 5th ed. New Jersey: John Wiley & Sons, Inc.; 2016.
[17] Statistics Indonesia. Statistics of Rice Producer Price at Huller Level 2024. Jakarta: 2025.
[18] Yang M, Zhai G. Measurement and Influencing Factors of Economic Resilience over a Long Duration of COVID-19: A Case Study of the Yangtze River Delta, China. Land (Basel) 2024;13:175. https://doi.org/10.3390/land13020175.
[19] Benitez IB, Ibañez JA, Lumabad CIIID, Cañete JM, Principe JA. Day-Ahead Hourly Solar Photovoltaic Output Forecasting Using SARIMAX, Long Short-Term Memory, and Extreme Gradient Boosting: Case of the Philippines. Energies (Basel) 2023;16. https://doi.org/10.3390/en16237823.
[20] González-Sopeña JM, Pakrashi V, Ghosh B. An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renewable and Sustainable Energy Reviews 2021;138:110515. https://doi.org/10.1016/J.RSER.2020.110515.
[21] Piotrowski P, Rutyna I, Baczyński D, Kopyt M. Evaluation Metrics for Wind Power Forecasts: A Comprehensive Review and Statistical Analysis of Errors. Energies 2022, Vol 15, Page 9657 2022;15:9657. https://doi.org/10.3390/EN15249657.
[22] Rehman A, Ma H, Ahmad M, Irfan M, Traore O, Chandio AA. Towards environmental Sustainability: Devolving the influence of carbon dioxide emission to population growth, climate change, Forestry, livestock and crops production in Pakistan. Ecol Indic 2021;125. https://doi.org/10.1016/j.ecolind.2021.107460.
[23] Christie G, Hatidja D, Tumilaar R. Penerapan Metode SARIMA dalam Model Intervensi Fungsi Step untuk Memprediksi Jumlah Pegunjung Objek Wisata Londa (Application of the SARIMA Method in the Step Function Intervention to Predict the Number of Visitors at Londa Tourism Object). JURNAL ILMIAH SAINS 2022;22:96. https://doi.org/10.35799/jis.v22i2.40961.
[24] Parizad B, Ranjbarzadeh H, Jamali A, Khayyam H. An Intelligent Hybrid Machine Learning Model for Sustainable Forecasting of Home Energy Demand and Electricity Price. Sustainability 2024;16:2328. https://doi.org/10.3390/su16062328.
[25] Andiojaya A. The Transmission of Paddy Grain Price to Rice Price: Review of Directions, Magnitude and Duration of Changes. 2021.

[26] Thejovathi M, Rao DrMVPCS, Priyadharsini EJ, Siddi S, Karthik B, Abbas SH. Optimizing Product Demand Forecasting with Hybrid Machine Learning and Time Series Models: A Comparative Analysis of XGBoost and SARIMA. SSRN Electronic Journal 2024. https://doi.org/10.2139/SSRN.5076161.
[27] Obaidat M, Almomani HA, Mallouhy R, AlMotari A, Al Meanazel OT. A Hybrid Machine Learning Framework for Daily Dairy Demand Forecasting: Integrating SARIMAX and XGBoost for Seasonal Production Optimization. IEEE Access 2025;13:162668–80. https://doi.org/10.1109/ACCESS.2025.3610316.
[28] Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice, 2nd Edition. Melbourne, Australia: OTexts; 2018.
[29] Hossain ML, Shams SMN, Ullah SM. Time-series and deep learning approaches for renewable energy forecasting in Dhaka: a comparative study of ARIMA, SARIMA, and LSTM models. Discover Sustainability 2025;6:1–25. https://doi.org/10.1007/S43621-025-01733-5/TABLES/8.
[30] Eichenauer VZ, Indergand R, Martínez IZ, Sax C. Obtaining consistent time series from Google Trends. Econ Inq 2022;60:694–705. https://doi.org/10.1111/ECIN.13049.