E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss2.art2

EKSAKTA
Journal of Sciences and Data Analysis

Forecasting Hotel Occupancy Rates in Bali Province using the SARIMAX Method with Tourist Data as an Exogenous Variable

Ihsan Fathoni Amri[1,*], Yolan Triky[2], Kaia Raissa Akmalia[3], Abimanyu Arya Ramadhan[4], Abdul Ghufron[5], M. Al Haris[6]

[1,*]Program Studi Sains Data, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.
[2] Program Studi Statistika, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.
[3] Program Studi Statistika, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.
[4] Program Studi Statistika, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.
[5] Program Studi Statistika, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.
[6] Program Studi Statistika, FSTP, Universitas Muhammadiyah Semarang, Jl. Kedungmundu No.18, Semarang, 50273, Jawa Tengah, Indonesia.

Corresponding author : ihsanfathoni@unimus.ac.id

Received: 16 July 2024; Accepted: 27 August 2024; Published: 23 October 2024

Abstract: Tourism is a crucial economic sector in Bali, Indonesia. Sustainable tourism management requires an understanding of the dynamics between tourist numbers and hotel occupancy levels. This study uses the SARIMAX (Seasonal Autoregressive Integrated Moving Average) method to estimate between the two indicators and reveals a positive correlation between the two indicators. The SARIMAX model effectively captures seasonal patterns and external factors, providing accurate forecasts and supporting tourism management in Bali. Monthly data from 2010 to 2023 were analyzed. Accurate estimates can help tourism stakeholders in formulating appropriate management strategies to optimize the tourism sector. Implementing the right strategy can help ensure the preservation of the local environment and culture, as well as long-term economic benefits for Bali. From the data we use the SARIMAX (6,1,0) (1,1,0)12 model with an AIC value of 1920.553 and a MAPE value of 27%.

Keywords: Tourism forecasting, SARIMAX, tourist arrivals, hotel occupancy rates, sustainable tourism management.

Introduction

In the era of globalization, tourism has become a crucial sector for the economy of a country. Tourism is defined as the activity of traveling undertaken by individuals or groups from their place of origin to a destination for leisure, business, or other purposes. The province of Bali is renowned for its natural beauty and rich cultural heritage. The diverse tourist attractions, ranging from beaches to religious ceremonies, make Bali a popular destination.

Tourists are individuals who travel from their place of origin to a destination for purposes such as vacation, research, business, or other needs. The presence of tourists has a significant impact on the economy, society, and culture as it involves various aspects such as accommodation expenditure. In 2019, there were over 1.5 billion international tourist arrivals worldwide, with global tourist spending reaching approximately USD

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

1.7 trillion. Around 6.3 million tourists, with 4.2 million being international tourists and 2.1 million domestic tourists, visited Bali in 2022 [1].

Hotel occupancy rate is an important indicator in the hospitality industry, reflecting the percentage of occupied rooms compared to the total available rooms over a certain period. A high occupancy rate indicates strong demand and effective management, while a low occupancy rate suggests a decline in tourist interest or issues in marketing strategies.

In the tourism industry, the number of tourists and the hotel occupancy rate are two important interdependent and influential indicators. A high number of tourists can increase the hotel occupancy rate, while a good hotel occupancy rate indicates the attractiveness of a destination and the quality of services offered. In this context, developing appropriate methods to forecast the number of tourists and the hotel occupancy rate in Bali is crucial for effective tourism management, specifically by building a SARIMAX model. This model allows for identifying correlations and predicting the impact of changes between the two variables [2].

The SARIMA (Seasonal AutoRegressive Integrated Moving Average) method is used for forecasting complex time series data. This model combines the strengths of Autoregression (AR), differentiation, Integration (I), and Moving Average (MA) to capture complex seasonal patterns and trends. This capability results in far more accurate forecasts compared to simpler forecasting methods [3].

The SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) method is an advanced form of the ARIMA model that allows the inclusion of exogenous variables in time series forecasting. SARIMAX combines the components of Autoregressive (AR), Moving Average (MA), Integration (I), and Seasonal patterns (S) by incorporating external variables that can influence the predicted variable. The advantage of SARIMAX lies in its ability to capture seasonal patterns as well as the effects of exogenous variables, thereby increasing prediction accuracy [4].

In the context of tourism in Bali, the SARIMAX method is used to forecast the hotel occupancy rate (Y) based on the number of tourists (X) visiting. This study found that the number of tourists has a positive correlation with the hotel occupancy rate. The application of the SARIMAX model to this data produced the coefficient of the exogenous variable (number of tourists) to show the percentage increase in the number of tourists contributing to the hotel occupancy rate. The accuracy of the SARIMAX model, measured by the Mean Absolute Percentage Error (MAPE), indicates a low prediction error and high model reliability [5].

Considering the complexity of the interaction between the number of tourists and the hotel occupancy rate, this study aims to develop a holistic analytical approach to effectively forecast and manage tourism development in Bali. This approach not only provides strategic insights for the government and tourism industry managers but also has the potential to enhance data-driven decision-making in responding to dynamic changes in the economic and social environment.

Research Methodology

2.1 Data Collection

Data collection was obtained from the official news of BPS Bali (https://bali.bps.go.id/) by taking hotel occupancy rate data from the Hotel Occupancy Rate in Bali Province and tourist data from the Number of Tourists to Bali for the period 2018-2023 on a monthly basis, using a total of 72 data points.

2.2 Research Variables

The variables studied in this case are the hotel occupancy rate as the dependent variable and the number of tourists as the independent variable. The research variables are presented in the following table:

Table 1. Research Variables

Notation	Variable
Y	Hotel Occupancy Rate
X	Number of Tourists

2.3 SARIMA

The ARIMA model can be extended to handle the seasonal factors of time series data, written with the notation ARIMA (p, d, q) (P, D, Q)^s, thus called the SARIMA model [6]. The formula is explained as follows:

$$Ø(β)\ Ø_p(\beta^s)(1-β)^d(1-β^s)\ Yt = Ø_q(β)\theta_q(β)\ \theta_Q(\beta^s)\varepsilon t \tag{1}$$

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

2.4 SARIMAX

The SARIMAX model is an extension of SARIMA by including exogenous variables. In this model, the factors influencing the variable Y at time t are not only a function of Y itself over different times but are also influenced by other variables X at time t [7].

In general, the form of the SARIMAX (p, d, q) (P, D, Q) s model is as follows:

$$\emptyset_p\ (\beta)\ \emptyset_p(\beta^s)(1-\beta)^d\ (1-\beta^s)\ ^D \mathrm{Yt} = \emptyset_q(\beta)\theta_q(\beta)\ \theta_Q(\beta^s)\varepsilon t + \ \alpha_1 \mathrm{X}_1\ + \alpha_2 \mathrm{X}_2\ + \ldots.\ \alpha_k \mathrm{X}_k \tag{2}$$

Where: $Xk, t =$ the additional variable or the k-th exogenous variable at time t.

2.5 Model Evaluation

In building the forecasting model, the data is divided into training data and test data. The training data is used to build the forecasting model, while the test data is used to evaluate the prediction results. The parameter used to measure the goodness of this model is MAPE (Mean Absolute Percentage Error), In the case of this study, it uses MAPE because to measure the relative error in percentages, this metric is independent of the data scale. This allows comparison of model performance on datasets of different scales. However, it is important to remember that MAPE also has weaknesses, such as problems with very small actual values, which can cause percentage errors to be very large or infinite. Therefore, its use must be adjusted to the characteristics of the existing data. Here's the formula:

$$MAPE = 100\%\ \Sigma t\ yt - \hat{y}t\ n \tag{3}$$

2.6 Preprocessing

This study uses the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) method. It begins with data preprocessing which includes data collection, data cleaning, attribute selection, data transformation, and data merging or replicating data from multiple datasets to create a new dataset [8].

Data is divided into train data and test data from the entire dataset. Time series decomposition is carried out to observe trends, seasonality, and residual patterns. The Augmented Dickey-Fuller (ADF) test is used to test stationarity, and the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to assess inter-variable correlations. The best SARIMA model is selected based on the lowest AIC value [4].

Diagnostic testing is conducted to ensure the model is free from residual traces indicating white noise. Diagnostic testing and parameter estimation are performed with the SARIMAX model to predict hotel occupancy rates in Bali Province. Prediction performance is evaluated using the Mean Percentage Absolute Error (MAPE) method to assess whether the forecast results are satisfactory [9].

In the context of inferential statistics, a confidence interval is used to describe population attributes such as mean and standard deviation [10]. The following flowchart illustrates the SARIMAX model procedure in Figure 1.

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

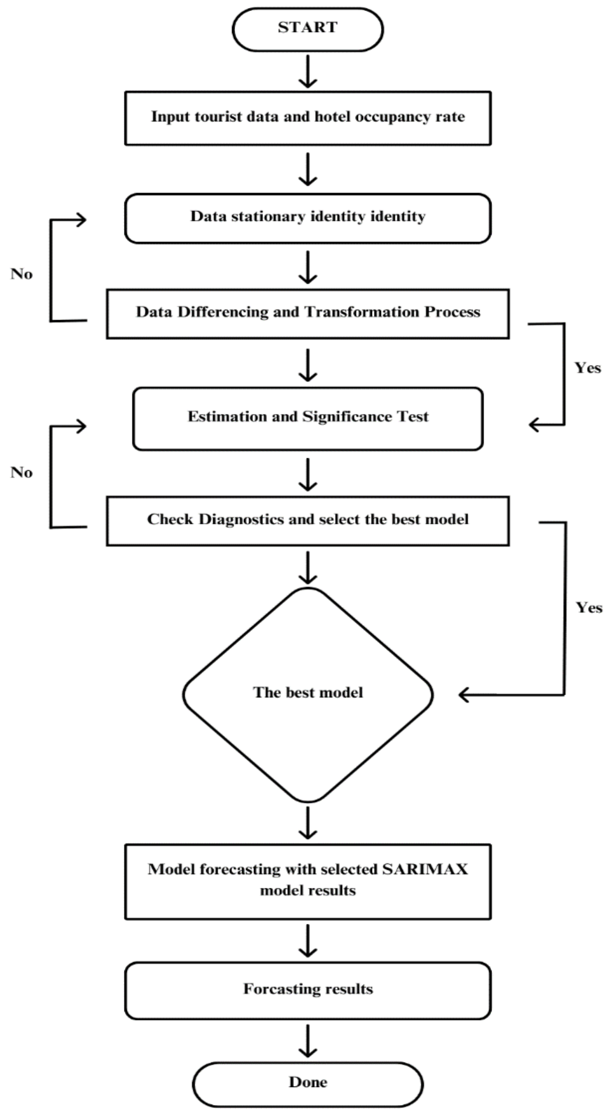


Figure 1. SARIMAX Procedure Flowchart

Based on Figure 1, the SARIMAX method procedure is explained by ensuring data stationarity with the Augmented Dickey-Fuller Test and performing differencing if the data is not stationary. Data is divided into training data (90%) and test data (10%) for model evaluation. Model testing is continued with the normality test (Shapiro-Wilk test) to ensure residuals are normally distributed. The White Noise test (Ljung-Box test) is used to ensure the model is free from patterns and additional information in the residuals. The correlation test measures the relationship between the number of tourists (Exogenous) and the hotel occupancy rate (Endogenous).

Research Results and Discussion

3.1 Descriptive Analysis

Descriptive analysis is a statistical method used to describe or summarize data. The main objective of descriptive analysis is to provide an overview of the data that will be tested. The following presents some elements of the descriptive analysis for the variables Number of Tourists (Y) and Hotel Occupancy Rate (X) [11].

E-ISSN: 2720-9326
P-ISSN: 2716-0459

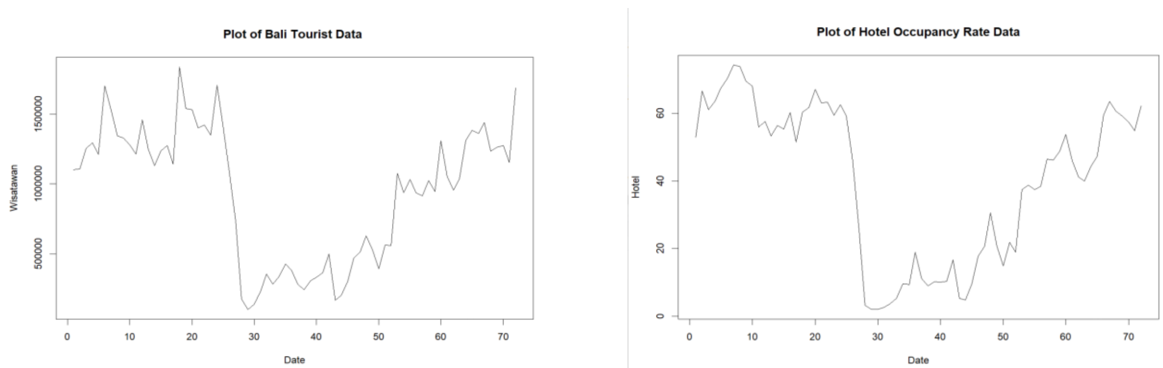
DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Table 2. Descriptive Analysis

Analysis	Value	
	Average Hotel Occupancy	Average Number of Tourists
Mean	944109	40.28
Median	1081024	46.37
Maximum	1837393	74,40
Minimun	101984	2,07
Standard Deviation	181714,44	23,24

Based on Table 2, it is known that the average value of the Y variable is 944,109, with a median of 1,081,024, a maximum value of 1,837,393, a minimum value of 101,984, and a standard deviation of 181,714.44. In addition, the average value of the X variable is 40.28, with a median of 46.37, a maximum value of 74.40, a minimum value of 2.07, and a standard deviation of 23.24. Both indicators have a correlation value of 0.956, indicating that the two variables have a strong correlation.

The first step in SARIMA modeling is to test the stationarity of the data, both in terms of mean and variance. The stationarity test of the data can be seen using the ACF (Autocorrelation Function) Plot.



(a) (b)

Figure 2.(a) Plot of Bali Tourist Data (b) Plot of Hotel Occupancy Rate Data

Based on Figures 2.(a) and 2.(b), it can be seen that the pattern of Tourist data and Hotel Occupancy Rate data in Bali Province from May 2020 to January 2022 is non-stationary, due to the presence of seasonal patterns. Seasonal patterns can be identified from data showing repeated fluctuations up and down at regular intervals.

3.2 Data Stationarity Test

From Figures 1 and 2, it can be seen that the data is non-stationary and therefore needs differencing. Further testing for data stationarity is carried out using the Augmented Dickey-Fuller (ADF) test. Before conducting the Augmented Dickey-Fuller (ADF) test to assess the stationarity of the data, it is crucial to establish the hypotheses and the criteria for decision-making. The ADF test is commonly used to determine whether a time series is stationary, which means that its mean, variance, and autocovariance remain constant over time.

Hypotheses:

H_0 : The data is non-stationary in mean, implying that there is a unit root present.

H_1 : The data is stationary in mean, implying that there is no unit root present.

α : A significance level (α) of 0.05, which means that there is a 5% risk of rejecting the null hypothesis when it is actually true.

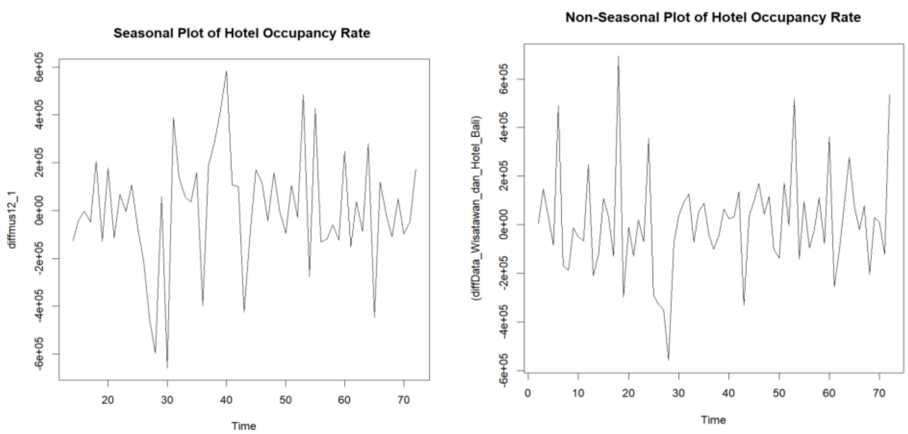
E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Table 3. ADF Test Results

Test	*P-value*	Remark
ADF Test	0.01	Stasionary

Based on Table 3, the results of the ADF test results show a P-value of 0.01 < 0.05, indicating that the data is stationary.



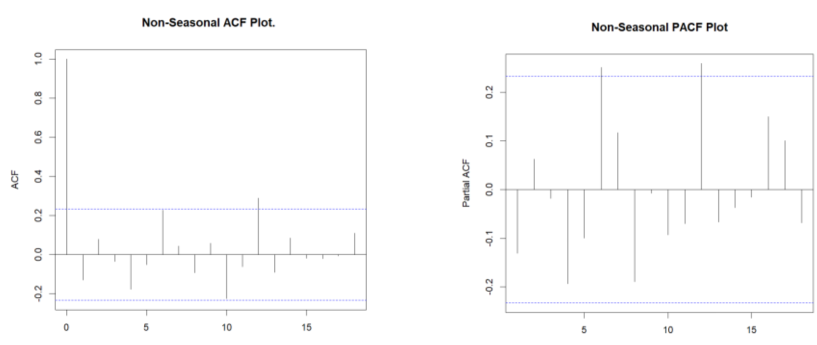
(a) (b)

Figure 3. (a) Seasonal Plot of Hotel Occupancy Rate (b) Non-Seasonal Plot of Hotel Occupancy Rate

Based on Figures 3.(a) and 3.(b), the plots show a clear fluctuation pattern that appears more stable after the first differencing process. Therefore, these two datasets have met the stationarity requirements necessary for further analysis. The differencing process applied successfully removed the non-stationary components from the initial data, allowing for more accurate analysis and clearer interpretation.

3.3 Model Identification

Model identification from the data can be done by plotting the differenced tourist data and hotel occupancy rate data in Bali Province into ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots [12]. The following are the ACF and PACF plots:



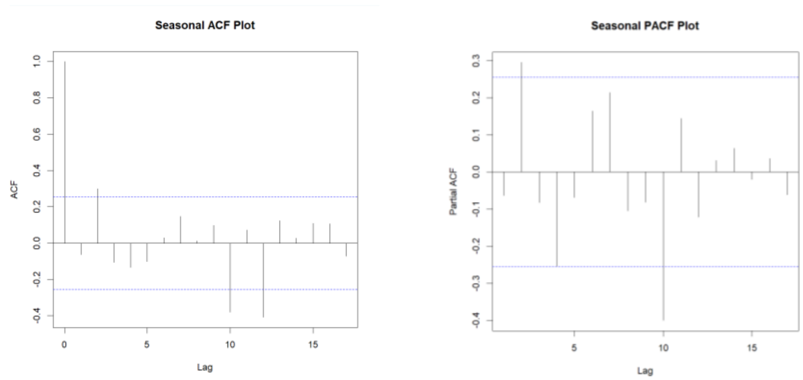
(a) (b)

Figure 4. (a) Non-Seasonal ACF Plot. (b) Non-Seasonal PACF Plot

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Based on Figures 4.(a) and 4.(b), it can be seen that at lag 0, there is a cut-off pattern where the lag falls outside the confidence line. The order q in the Moving Average (MA) model is identified as q (MA) 0, while the order p in the Auto-Regressive (AR) model is identified as p (AR) 6 with one differencing, d = 1.



(a) (b)

Figure 5. (a) Seasonal ACF Plot. (b) Seasonal PACF Plot

Based on Figures 5.(a) and 5.(b), the order P (SAR) is identified as P (SAR) 2, indicating a seasonal cut-off pattern where lag 2 falls outside the confidence line. The order Q in the Seasonal Moving Average (SMA) model is also identified as SMA (2) 12, where lag 2 falls outside the confidence line. Based on the above explanation, the preliminary SARIMA model is obtained as follows:

Table 4. Preliminary SARIMA Model with AIC

	SARIMA	
Model	Orde	AIC
1	(6,1,0)(2,1,2)	420.58
2	(6,1,0)(2,1,1)	420.95
3	(6,1,0)(2,1,0)	422.62
.....	.....	.....
40	(2,1,0)(0,1,1)	413.11
.....	.....	.....
48	(1,1,0)(0,1,1)	413.74

Based on Table 4, it can be seen from various model tests that the 40th model has the smallest AIC value of 413.11. Therefore, the SARIMA (2,1,0) $(0,1,1)^{12}$ model is the best model and is suitable for forecasting the next period. The best model formed is as follows:

$$Ø(\beta)\, Ø_p(\beta^s)(1-\beta)^d(1-\beta^s)\, Yt = Ø_q(\beta)\theta_q(\beta)\, \Theta_Q(\beta^s)\varepsilon t$$
$$(1\text{-}0{,}17968\, \beta\,) - 0{,}21798\, \beta^2\, (1\text{-}\beta)\, Yt = (\, 1\text{-}0{,}70008\, \beta^{12})\, \varepsilon t$$

3.4 Model Diagnostic Test

The SARIMA model diagnostic test is used to examine whether the selected model meets the residual assumptions. The following are the test results for the SARIMA (2,1,0) $(0,1,1)^{12}$ model.

a. White Noise Autocorrelation Test

Hypothesis

H_0 : Residuals are not autocorrelated
H_1 : Residuals are autocorrelated
α : 0.05

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Table 5. White Noise Assumption Test

Test	P-value	Description
Ljung-Box	0,746	White Noise

Based on Table 5, the Ljung-Box test results show a P-value of 0.746 > 0.05, which means $H0$ is accepted. Thus, it can be concluded that the SARIMA (2,1,0) $(0,1,1)^{12}$ model meets the white noise assumption. This means that there is no autocorrelation in the residuals, and the non-autocorrelation assumption is fulfilled.

b. Normality Assumption Test

Hypothesis

H_0 : Residuals are normally distributed

H_1 : Residuals are not normally distributed

α : 0.05

Table 6. Normality Assumption Test

Test	P-value	Description
Kolmogorov-Smirnov	0.3533	Normally Distributed

Based on Table 6, the Kolmogorov-Smirnov test results show a P-value of 0.3533 > 0.05, which means $H0$ is accepted. Thus, it can be concluded that the SARIMA (2,1,0) $(0,1,1)^{12}$ model meets the normality assumption. Since the SARIMA (2,1,0) $(0,1,1)^{12}$ model meets all assumptions, it is suitable for forecasting the hotel occupancy rate in Bali Province in the future with MAPE (Mean Absolute Percentage Error).

Table 7. MAPE Value Test

Test	P-value	Description
MAPE	34.53949	Fairly Accurate

Based on Table 7, a MAPE value of 34.5% was obtained due to fluctuations in hotel occupancy rates caused by COVID-19, indicating that the pandemic has caused drastic changes in the data that are difficult for existing models to predict. Therefore, there is a need for more adaptive and sophisticated approaches to improve prediction accuracy [13, 14].

Based on Table 4, the interim SARIMA results with AIC will be used as a determinant for the interim SARIMAX model by incorporating exogenous variables into the model. The interim SARIMAX model is as follows:

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Table 8. Interim SARIMAX results with AIC Model

SARIMAX		
Model	Orde	AIC
1	(6,1,0)(2,1,2)	1924.39
2	(6,1,0)(2,1,1)	1923.68
3	(6,1,0)(2,1,0)	1922.46
4	(6,1,0)(1,1,2)	1922.4
5	(6,1,0)(1,1,1)	1922.49
6	(6,1,0)(1,1,0)	1920.553
.....	.....	.....
48	(1,1,0)(0,1,1)	1936.9

Based on Table 8, it can be seen that the interim SARIMAX model with the smallest AIC value is SARIMAX (6,1,0) $(1,1,0)^{12}$, with an AIC value of 1920.553. Thus, the SARIMAX (6,1,0) $(1,1,0)^{12}$ model can be further tested.

The best-formed model is as follows:

$$\emptyset_p\,(\beta)\,\emptyset_p(\beta^s)(1-\beta)^d\,(1-\beta^s)\,^D Yt = \emptyset_q(\beta)\theta_q(\beta)\,\theta_Q(\beta^s)\varepsilon t + \alpha_1 X_1 + \alpha_2 X_2 + \dots \alpha_k X_k$$

$$(1+1{,}00316\beta)\,(1+0{,}59964\beta^2)\,(1+0{,}45936\beta^3)\,(1+0{,}61851\beta^4)\,(1+0{,}63273\beta^5)\,(1+0{,}26467\beta^6)\,(10{,}36892\beta^{12}) = \varepsilon t$$

3.5 Significance Test of Parameters

Table 9. Significance Test of Parameters for SARIMAX (6,1,0) $(1,1,0)^{12}$

Coefficient	Estimate Value	P-Value
ar_1	-1.00316	< 2.2e-16
ar_2	-0.59964	0.0001993
ar_3	-0.45936	0.0031149
ar_4	-0.61851	2.429e-05
ar_5	-0.63273	1.928e-05
ar_6	-0.26467	0.0261731
sar_1	0.36892	0.0067105

Based on Table 9, all coefficients in the SARIMAX model are significant because the P-Value < 0.05, indicating that each parameter has a significant contribution to the model and helps explain the variability of the observed data.

3.6 Diagnostic Test of the Model

The diagnostic test for the SARIMAX model is used to check whether the selected model meets the residual assumptions. The following are the test results for the SARIMAX (6,1,0) $(1,1,0)^{12}$ model:

Table 10. White Noise Assumption Test

Test	P-value	Description
Ljung-Box	0.9093	White Noise

Based on Table 10, the Ljung-Box test yielded a P-value (0.9093) > 0.05, meaning H0 is accepted. Thus, it can be concluded that the SARIMAX (6,1,0) $(1,1,0)^{12}$ model meets the white noise assumption. Next, a residual normal distribution test on the best model was conducted with the following results:

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

Table 11. Normality Assumption Test

Test	P-value	Description
Kolmogorov-Smirnov	0.7933	Normally Distributed

Based on Table 11, the Kolmogorov-Smirnov test yielded a P-value (0.7933) > 0.05, meaning H0 is accepted. Thus, it can be concluded that the SARIMAX (6,1,0) $(1,1,0)^{12}$ model meets the normality assumption. Since the SARIMAX (6,1,0) $(1,1,0)^{12}$ model meets all assumptions, it is suitable for forecasting hotel occupancy rates by adding the tourist variable in Bali Province for the future with MAPE (Mean Absolute Percentage Error).

Table 12. MAPE Value Test

Test	P-value	Description
MAPE	27.12293	Fairly Accurate

Based on Table 12, a MAPE value of 27% was obtained. If the MAPE value is greater than 20%, the SARIMAX (6,1,0) $(1,1,0)^{12}$ model is categorized as fairly accurate in forecasting hotel occupancy rates and tourist numbers.

3.7 Forecasting Data

Forecasting is the process of predicting or estimating future events based on historical data and current trend analysis. In this context, forecasting hotel occupancy data against tourist numbers is the process of predicting future hotel occupancy rates based on historical tourist data. This process is crucial for the hospitality and tourism industry to plan operations, manage room inventory, and set effective marketing strategies [15].

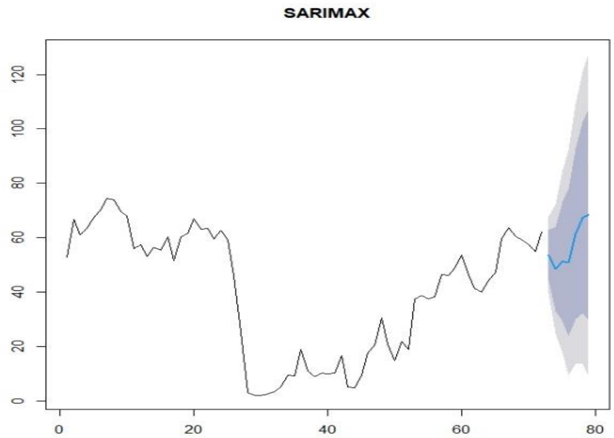
Forecasting is conducted using the SARIMAX method because the SARIMAX (6,1,0) (1,1,0)12 model has an AIC value of 1920.553 and a MAPE value of 27% compared to the SARIMA (2,1,0) (0,1,1)12 model, which has an AIC value of 413.11 and a MAPE value of 34.5%.

Tabel 13. Forecast SARIMAX

Point	Forecast	95% Confidence Interval	
		Lower	Upper
73	53.56238	39.443387	67.68137
74	48.50923	24.981888	72.03657
75	51.23270	17.705539	84.75986
76	50.81183	9.473723	92.14994
77	61.32183	13.538271	109.10538
78	67.35491	13.771115	120.93870
79	68.48706	9.335416	127.63870

Based on Table 13, the prediction values with a 95% confidence interval and the forecast plot from the SARIMAX model are as follows.

Figure 6. SARIMAX Forecast Plot



Conclusion

From the results of this study, a positive relationship between the number of tourists and hotel occupancy rates in Bali was established. The SARIMAX method with the smallest AIC value is the SARIMAX (6,1,0) $(1,1,0)^{12}$ model, with an AIC value of 1920.553 and a MAPE of 27%. Meanwhile, the best model from the SARIMA (2,1,0) $(0,1,1)^{12}$ model has the smallest AIC value of 413.11 and a MAPE of 34.5%. Therefore, it can be concluded that the suitable method to use is the SARIMAX method, and the best-formed model is as follows:

The prediction of hotel occupancy rates can be achieved with high accuracy, reaching a MAPE of 27%. The 27% MAPE due to fluctuations in hotel occupancy rates caused by COVID-19 indicates that the pandemic has caused drastic changes in the data that are difficult for existing models to predict. Therefore, there is a need for more adaptive and sophisticated approaches to improve prediction accuracy. This study also highlights the importance of monitoring and analyzing data in tourism destination management and the potential use of advanced analytical techniques to support better decision-making.

Recommendations

The analysis results can be used as a reference for relevant institutions in decision-making, such as data fluctuations that cause drastic changes that are difficult for existing models to predict.

Acknowledgements

We express our gratitude to Almighty God for His blessings and grace, which enabled us to complete this article. We also thank Mr. Ihsan Fathoni Amri, S.Si., M.Stat., as the supervising lecturer, for providing guidance and direction throughout this research. We greatly appreciate the cooperation and dedication of the entire team who contributed. We also extend our thanks to our families and friends who have always provided support and motivation. We hope this article can be useful and serve as a reference for future research.

References

[1] S. J. Page, J. Connell, Tourism: A Modern Synthesis, 5th ed., 2021.
[2] D. Kallioras, A. Petrou, Seasonal and business cycle variability in hotel occupancy rates: Evidence from Greece, Tourism Management Perspectives 27 (2018) 87-95.
[3] A. Aghelpour, M. H. Asgari, M. A. Sharifi, Temperature prediction using SARIMA model and hybrid neural networks: A comparative study, Environmental Modelling & Software 114 (2019) 104655.
[4] R. J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., OTexts, 2018.

E-ISSN: 2720-9326
P-ISSN: 2716-0459

DOI : 10.20885/EKSAKTA.vol5.iss2.art2

[5] A. G. Assaf, M. G. Tsionas, Forecasting international tourist arrivals using time series models: Evidence from a panel data analysis, Tourism Management 74 (2019) 199-218.
[6] F. I. Durrah, Y. Yulia, T. P. Parhusip, A. Rusyana, Peramalan Jumlah Penumpang dengan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA), Journal of Data Analysis 1 (2018) 1-11.
[7] S. Dahlia, D. Rahmi, Peramalan Produksi Tanaman Biofarmaka di Provinsi Riau dengan Metode Sarimax, Jurnal Pendidikan Ilmiah Transformatif 7 (2023).
[8] I. R. Julianto, I. Indwiarti, A. A. Rohmawati, Prediksi Jumlah Kunjungan Wisatawan Di Jawa Barat Dengan Model ARIMAX Dan SARIMAX Menggunakan Data Google Trends, eProceedings of Engineering 8 (2021).
[9] R. A. M. Yusuf, T. S. Yanti, Perbandingan Metode Seasonal Autoregressive Integreted Moving Average (SARIMA) dan Metode Fuzzy Time Series untuk Model Peramalan Jumlah Wisatawan Mancanegara di Bali, Prosiding Statistika (2021) 597–605.
[10] A. A. Maulana, H. Rosalina, Implementasi Metode Sarimax untuk Prediksi Curah Hujan Jangka Pendek di Pagerageung, Tasikmalaya, Jurnal Sumber Daya Air 20 (2024) 39-50.
[11] M. Tobing, Pengaruh Jumlah Obyek Wisata, Tingkat Penghunian Kamar, Dan Jumlah Kunjungan Wisatawan Terhadap Pendapatan Asli Daerah Kabupaten Simalungun, Jurnal Ekuilnomi 3 (2021) 127-139.
[12] Z. Zalila, Peramalan Jumlah Kunjungan Wisatawan Mancanegara ke Provinsi Jawa Tengah dengan Metode Seasonal Autoregressive Integrated Moving Average with Exogenous Input (SARIMAX), Doctoral dissertation, Muhammadiyah University, Semarang, 2019.
[13] S. K. Prilistya, A. E. Permanasari, S. Fauziati, The effect of the COVID-19 pandemic and Google trends on the forecasting of international tourist arrivals in Indonesia, 2021 IEEE Region 10 Symposium (TENSYMP) (2021) 1–8.
[14] N. D. A. Amrita, M. M. Handayani, L. Erynayati, Pengaruh pandemi covid-19 terhadap pariwisata bali, Jurnal Manajemen Dan Bisnis Equilibrium 7 (2021) 246–257.
[15] N. Permatasari, Penggunaan indeks google trend dalam peramalan jumlah pengunjung taman rekreasi selecta tahun 2020, Seminar Nasional Official Statistics 1 (2021) 1019-102.