JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 7, No. 4. September 2025 DOI: https://doi.org/10.34288/jri.v7i4.388
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

FORECASTING THE HIGHEST NUMBER OF HOTEL VISITORS IN MOJOKERTO REGENCY USING ARIMA MODEL

Gading Putri Diniarti-[1], Rizka Hadiwiyanti-[2], Prasasti Karunia F. A-[3]

Information Systems Study / Faculty of Computer Science
Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia
gadingdnr@gmail.com-[1], rizkahadiwiyanti.si@upnjatim.ac.id-[2], prasasti.karunia.fasilkom@upnjatim.ac.id-[3]

Abstract

This study aims to forecast the number of hotel visitors in Mojokerto Regency using the Autoregressive Integrated Moving Average (ARIMA) model based on monthly data from 2022 to 2024 provided by the Department of Culture, Youth, Sports, and Tourism (Disbudporapar). The research focuses on three hotels with the highest number of visitors: Hotel Grand Whiz, Puri Indah Hotel, and Hotel Arrayana. The implementation was carried out using Python via the Google Colab platform, involving several analytical stages including data stationarity testing (ADF), differencing, identification of ARIMA parameters (p, d, q) using ACF and PACF plots, automatic model estimation with auto ARIMA, and residual diagnostics. Model performance was evaluated using MSE, RMSE, and MAPE. The results show that ARIMA performed best on Puri Indah Hotel data with a MAPE of 9.65%, indicating high accuracy, while performance was lowest for Hotel Arrayana with a MAPE of 32.31%. Visualization of the predictions revealed that ARIMA works effectively for stable patterns but is less adaptive to volatile trends. The implementation of ARIMA proves to be a useful tool in supporting data-driven decision-making for tourism planning and hotel operational strategy in Mojokerto Regency.

Keywords: auto ARIMA; time series forecasting; hotel visitor prediction; MSE; RMSE; MAPE

Abstrak

Penelitian ini bertujuan untuk meramalkan jumlah kunjungan hotel di Kabupaten Mojokerto menggunakan model Autoregressive Integrated Moving Average (ARIMA) berdasarkan data bulanan tahun 2022 hingga 2024 yang diperoleh dari Dinas Kebudayaan, Kepemudaan, Olahraga, dan Pariwisata (Disbudporapar). Fokus penelitian ditujukan pada tiga hotel dengan jumlah kunjungan tertinggi, yaitu Hotel Grand Whiz, Puri Indah Hotel, dan Hotel Arrayana. Implementasi model dilakukan menggunakan Python melalui platform Google Colab, dengan tahapan analisis yang mencakup pengujian stasioneritas data (ADF), proses differencing, identifikasi parameter ARIMA (p, d, q) melalui grafik ACF dan PACF, estimasi otomatis menggunakan auto ARIMA, serta evaluasi residual. Kinerja model diuji dengan metrik MSE, RMSE, dan MAPE. Hasil penelitian menunjukkan bahwa model ARIMA memberikan performa terbaik pada Puri Indah Hotel dengan nilai MAPE sebesar 9,65% (sangat akurat), dan performa terendah pada Hotel Arrayana dengan nilai MAPE sebesar 32,31% (wajar). Visualisasi hasil prediksi menunjukkan bahwa ARIMA efektif untuk pola data yang stabil, namun kurang adaptif terhadap tren yang fluktuatif. Penerapan model ARIMA terbukti dapat mendukung pengambilan keputusan berbasis data dalam perencanaan pariwisata dan strategi operasional hotel di Kabupaten Mojokerto.

Kata kunci: auto ARIMA; peramalan deret waktu; prediksi pengunjung hotel; MSE; RMSE; MAPE;

INTRODUCTION

The tourism sector is one of the strategic components in driving economic growth, both at the national and regional levels. In Mojokerto Regency, East Java Province, tourism contributes significantly to the increase of local revenue, the creation of job opportunities, and the development of supporting infrastructure. One of the key indicators reflecting the dynamics of this sector is the hotel occupancy rate, which also represents the intensity of tourist visits (Pratama, 2020).

According to data managed by the Department of Culture, Youth, Sports, and Tourism (Disbudporapar) of Mojokerto Regency, three hotels have consistently recorded the highest number of visitors: Hotel Grand Whiz, Hotel Avila, and Puri Srijaya Hotel. These hotels play a vital role as primary accommodation providers in strategic tourism areas and contribute significantly to the local socio-economic activities (Pariwisata et al., 2020). The following is a graphical representation of visitor numbers at the three most visited hotels managed by Disbudporapar during the period from 2022 to 2024.

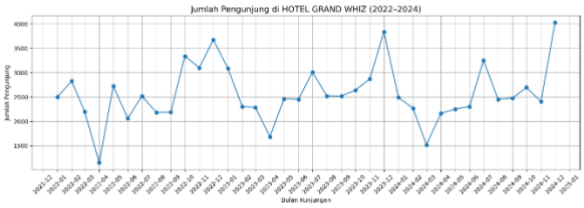


Figure 1. Visitor Number Chart of Hotel Grand Whiz

Figure 2. Visitor Number Chart of Puri Indah Hotel

Figure 3. Visitor Number Chart of Hotel Arrayana

Seasonal fluctuations in tourist numbers present a major challenge in resource management and hotel operational planning. Demand uncertainty, particularly during off-peak seasons, can lead to imbalances between service capacity and visitor needs. Therefore, a data-driven analytical approach is necessary to accurately forecast visitor numbers (Mazlan et al., 2023).

Previous studies have shown that the ARIMA model consistently delivers more accurate predictions compared to other methods such as Prophet and LSTM. For example, in a study comparing these three algorithms for predicting amusement park ticket sales, ARIMA(1,0,1) achieved the best performance with an MAE of 478,887 and an RMSE of 762,009, outperforming Prophet and LSTM, which had higher prediction errors (Rizkya et al., 2023). Another study on library visitor forecasting reported that ARIMA yielded an RMSE of 26.17 and a MAPE of 22%, compared to LSTM which produced an RMSE of 35.59 and a MAPE of 25% (Umam, 2023). Similarly, the application of ARIMA to forecast international tourist arrivals at Ngurah Rai Airport in 2024 showed an MSE of 1899.60, reinforcing ARIMA's relevance in tourism forecasting contexts (Cristanto & Mailoa et al., 2023).

The Autoregressive Integrated Moving Average (ARIMA) model is one of the most widely used time series forecasting methods in quantitative research due to its capability to capture trends and seasonal fluctuations. It combines autoregressive (AR), differencing (I) to ensure stationarity, and moving average (MA) components, making it reliable for predicting nonlinear and seasonal data patterns (Petropoulos et al., 2022).

The auto ARIMA approach is used in this study to determine model parameters efficiently and accurately. This method utilizes an automatic search algorithm to evaluate various combinations of ARIMA parameters (p, d, q) without requiring a complex manual exploration process (As et al., 2017). The main advantage of auto ARIMA lies in its ability to adjust parameters adaptively to the characteristics of the actual data. This process includes detecting the need for differencing to achieve stationarity and selecting the best model based on evaluation criteria values such as AIC and BIC (Fazrina et al., 2024). Research by Shedriko & Firdaus (2024) comparing manual ARIMA and Auto ARIMA in the context of Indonesian export prediction shows that Auto ARIMA produces a model (1,1,0) that is less significant (Prob(H) = 0.34), compared to the manual model (1,1,1) which is significant and provides a lower error value (MAE = 0.06 vs. 0.36; RMSE = 0.07 vs. 0.41). However, Auto ARIMA is still considered useful for model initialization or as a baseline in the early stages of time series data exploration(Firdaus et al., 2024).

This study aims to implement the Auto ARIMA model to forecast visitor numbers at the three hotels with the highest occupancy rates in Mojokerto Regency, namely Hotel Grand Whiz, Hotel Avila, and Puri Srijaya Hotel. The forecasting results are expected to serve as a basis for hotel operational strategy development and support data-driven decision-making by Disbudporapar Mojokerto, particularly in efforts to enhance efficiency, service effectiveness, and competitiveness in the regional hotel industry.

RESEARCH METHODS

This research method is systematically structured as illustrated in Figure 4, which depicts the stages of the process from data collection to result representation. Each stage is designed to ensure that the implementation of the ARIMA model proceeds in a structured manner and produces accurate forecasts. The process begins with the collection of historical data on hotel visitor numbers, followed by a pre-processing stage to ensure data quality before model application. Subsequently, the Auto ARIMA method is implemented, followed by prediction result evaluation and presentation in a form that is informative and easily understood by stakeholders.

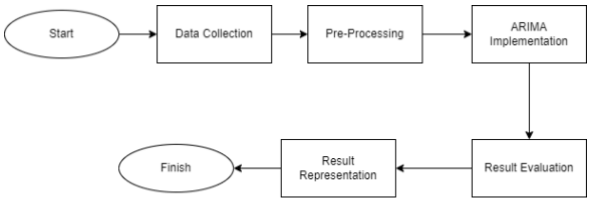


Figure 4. Research Method Flowchart

Data Collection

This study utilizes secondary data in the form of Microsoft Excel spreadsheets obtained from the Department of Culture, Youth, Sports, and Tourism (Disbudporapar) of Mojokerto Regency. The dataset includes monthly visitor numbers at 17 hotels managed by Disbudporapar, covering the observation period from January 2022 to December 2024. Among these are Hotel Grand Whiz, Hotel Avila, and Puri Srijaya Hotel, which consistently record the highest number of visitors. The dependent variable in this study is the number of visitors, while the independent variables are the monthly and yearly visit indices. The time series characteristics of the data make it suitable for analysis using the ARIMA (Autoregressive Integrated Moving Average) forecasting method.

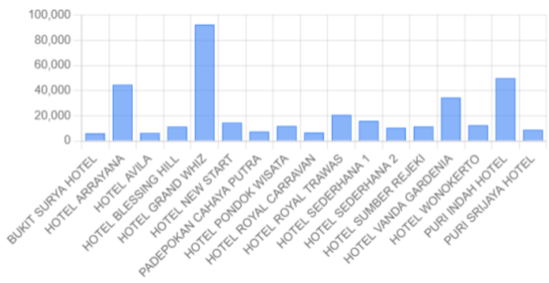


Figure 5. Comparison Chart of Hotel Visitor Numbers

Pre-Processing

The data pre-processing stage was conducted to ensure the dataset was ready for analysis using the ARIMA method. The raw data obtained from Disbudporapar was cleaned and reformatted using Microsoft Excel to align with a time series structure. The data was organized into a vertical table consisting of columns for hotel name, number of visitors, and time indices (month and year of visit) (Suhendra et al., 2023). In this study, the dependent variable is the number of hotel visitors, while the independent variable is the visit time, represented by the month and year. This step aimed to simplify the data structure and ensure that each observation had a consistent time reference, as required for time series modeling.

ARIMA Implementation

The implementation of the ARIMA method in this study was carried out using the Python programming language on the Google Colab platform. The implementation process was conducted systematically through the following stages:

1. Stationarity Testing

The initial step involved testing the stationarity of the data to ensure that the mean and variance remained constant over time. This was performed using the Augmented Dickey-Fuller (ADF) statistical test. The results of this test serve as the basis for determining whether the data can be used directly or require transformation (Wahyu ngestisari, 2020).

2. Differencing

If the ADF test indicated that the data were non-stationary, differencing was applied. This process aims to eliminate long-term trends or fluctuations by subtracting the current data value from its previous value, thereby transforming the data into a stationary series.

3. ARIMA Identification

After the data is confirmed to be stationary, the next step is to identify the ARIMA model parameters (p, d, q) using the Auto ARIMA approach. In its internal process, Auto ARIMA still considers the autocorrelation pattern (ACF) and partial autocorrelation (PACF) to estimate the structure of the relationship between data, which is automatically used to compile the initial configuration and validate the most optimal model (Cherrly & Somya, 2023).

4. ARIMA Parameter Estimation

Parameter estimation was performed using the auto ARIMA approach, which automatically selects the optimal combination of p, d, and q values based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood (Fazrina et al., 2024). The estimation was conducted using the Maximum Likelihood Estimation (MLE) method to maximize the likelihood of the observed data given the model (Dewi & Sutarman, 2022).

5. Diagnostic Checking

Diagnostic evaluation of the model residuals was conducted to ensure the model's validity. This included analysis of ACF and PACF plots of the residuals, white noise testing using the Ljung-Box test, and normality testing. If any assumption violations were detected, such as non-random residuals or non-normal distribution, model adjustments were applied to achieve more accurate and reliable results.

Result Evaluation

The evaluation of results was conducted to assess the accuracy of the ARIMA model in forecasting hotel visitor numbers managed by the Department of Culture, Youth, Sports, and Tourism (Disbudporapar) of Mojokerto Regency. The assessment utilized three key evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). MSE calculates the average of the squared differences between actual and predicted values, while RMSE presents this error in the same unit as the original data, making it more interpretable. MAPE, on the other hand, measures the percentage of error relative to the actual values and is commonly used to evaluate forecasting accuracy. A MAPE value below 10% is considered very accurate, 10%–20% is accurate, 20%–50% is reasonable, and above 50% is considered inaccurate. These three metrics are applied in combination to provide a comprehensive evaluation of the model's performance (Febiola et al., 2024). The lower the MSE, RMSE, and MAPE values, the higher the accuracy of the ARIMA model.

These differences are not only caused by the internal structure of the data, but are also influenced by a number of external factors that are not included in the modeling. Some of these include seasonal activities or local events such as national holidays, regional festivals, or government activities that can cause a sudden spike in the number of visitors. In addition, hotel location factors, accessibility to the city center or tourist attractions, promotional strategies, and changes in management or pricing policies can also affect the fluctuation of the number of guests. Even unexpected external conditions, such as natural disasters or changes in tourism sector regulations, also have an impact on hotel visit patterns.

Result Representation

The forecasting results of hotel visitor numbers are presented in the form of visualizations. This result representation displays a comparison between actual data and predicted values, both numerically and visually. Time series charts are used to illustrate the monthly fluctuation patterns in visitor numbers. Such visual presentation is essential to facilitate the interpretation process and support data-driven decision-making. As noted by (Pipit Muliyah et al., 2020), interactive and informative result representations can enhance clarity in understanding data patterns and support analytical processes in the tourism sector. Accordingly, the ARIMA model's prediction results can serve as a strategic reference for Disbudporapar and hotel management in planning promotions, managing capacity, and optimizing service delivery.

RESULTS AND DISCUSSION

Pengujian Stasioner

Stationarity testing was conducted to ensure that the data had a constant mean and variance over time, which is a critical prerequisite for applying the ARIMA model. The stationarity test was performed using the Augmented Dickey-Fuller (ADF) method, with a p-value threshold of less than 0.05 serving as an indicator of stationarity (Vean & Witanti, 2024).

	ADF Statistic	p-value	Stasioner
HOTEL GRAND WHIZ	-4.468468	0.000224	True
PURI INDAH HOTEL	-3.558021	0.006616	True
HOTEL ARRAYANA	-2.405288	0.14025	False

Figure 6. Stationarity Test Result

Based on the ADF test results, two out of the three hotels analyzed exhibited stationary data characteristics. Hotel Grand Whiz and Puri Indah Hotel had p-values of 0.000224 and 0.006616, respectively, both below the 0.05 significance threshold, thus fulfilling the stationarity assumption. In contrast, Hotel Arrayana had a p-value of 0.14025, exceeding the threshold, indicating that the data were non-stationary and required differencing before being used in the ARIMA modeling process.

Differencing

Hotel Arrayana was found to be non-stationary based on the initial ADF test results; therefore, a differencing process was applied to eliminate trends and render the data stationary. The transformation results are shown in Figure 6, where the ADF value after differencing improved to -2.96 with a p-value of 0.0392. This value is lower

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.388
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

than the 10% critical value (-2.64) but slightly above the 5% critical threshold (-2.99), indicating that the data can be considered stationary at the 10% confidence level. Thus, the differencing process successfully stabilized the data structure and fulfilled the fundamental assumptions required for ARIMA modeling.

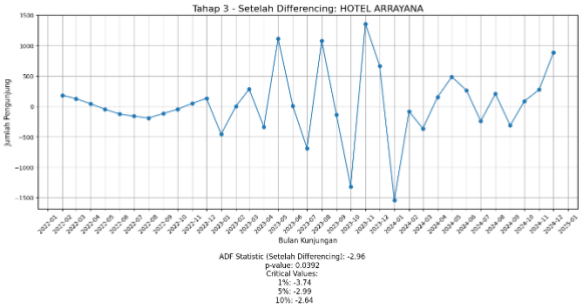


Figure 7. Differencing Result of Hotel Arrayana

ARIMA Identification (ACF and PACF)

The ARIMA model identification stage was carried out by analyzing the patterns in the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for each hotel's data. This analysis aims to determine the appropriate order of the AR (p) and MA (q) parameters to be used in the model, with the identification process automated through the use of auto ARIMA, which determines the optimal parameters based on ACF and PACF patterns (Insti, 2024).

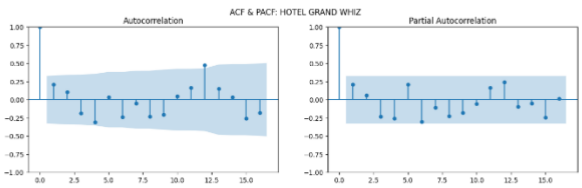


Figure 8. ACF and PACF Plots of Hotel Grand Whiz

Hotel Grand Whiz, the ACF plot shows a gradual and insignificant decline over several lags, while the PACF plot exhibits a significant spike only at lag one. This pattern indicates that a simple autoregressive structure such as AR(1), or even a stationary model without AR or MA components (ARIMA(0,0,0)), may be appropriate.

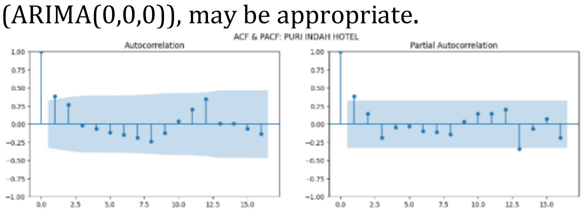


Figure 9. ACF and PACF Plots of Puri Indah Hotel

Puri Indah Hotel, both ACF and PACF plots show a significant spike at lag one followed by minor fluctuations in subsequent lags. This suggests a potential AR(1) or MA(1) model structure, although a low-order combination such as ARIMA(1,0,1) should also be considered to account for the observed variation.

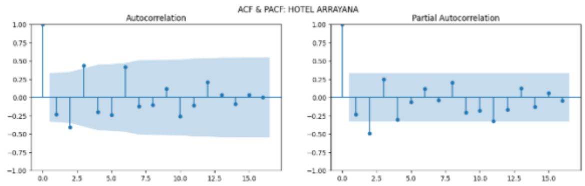
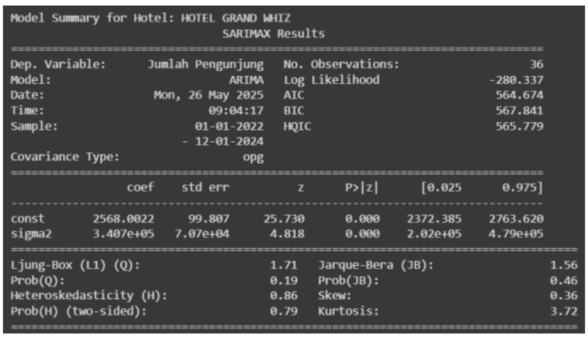


Figure 10. ACF and PACF Plots of Hotel Arrayana

Hotel Arrayana, the ACF plot displays a strong spike at the first lag followed by a sharp decline, while the PACF plot does not reveal any pronounced pattern. This implies a tendency toward a moving average structure, such as MA(1), and supports the use of an ARIMA(0,1,1) model, particularly given that the data had previously undergone differencing to achieve stationarity.

ARIMA Parameter Estimation

Parameter estimation for the ARIMA model was carried out following the identification stage, using the Maximum Likelihood Estimation (MLE) approach. The estimation results revealed different model configurations for each hotel, based on the structure of their historical data. To streamline this process and minimize manual trial-and-error in model selection, the *auto ARIMA* method was implemented. This approach automatically determines the best-fitting ARIMA model by evaluating combinations of parameters using model selection criteria such as AIC and BIC, while also leveraging MLE for parameter estimation (Vaibhava Lakshmi & Radha, 2021).


```
Model Summary for Hotel: HOTEL GRAND WHIZ
                               SARIMAX Results
==============================================================================
Dep. Variable:      Jumlah Pengunjung   No. Observations:                   36
Model:                          ARIMA   Log Likelihood                -280.337
Date:                Mon, 26 May 2025   AIC                            564.674
Time:                        09:04:17   BIC                            567.841
Sample:                    01-01-2022   HQIC                           565.779
                         - 12-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       2568.0022     99.807     25.730      0.000    2372.385    2763.620
sigma2      3.407e+05   7.07e+04      4.818      0.000    2.02e+05    4.79e+05
===================================================================================
Ljung-Box (L1) (Q):                   1.71   Jarque-Bera (JB):                 1.56
Prob(Q):                              0.19   Prob(JB):                         0.46
Heteroskedasticity (H):               0.86   Skew:                             0.36
Prob(H) (two-sided):                  0.79   Kurtosis:                         3.72
===================================================================================
```
Figure 11. Parameter Estimation Result of Hotel Grand Whiz

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Hotel Grand Whiz, the resulting model was ARIMA(0,0,0) with a constant value of 2568.0022, which was statistically significant (p-value < 0.01). This indicates that the data are stationary without strong autoregressive or moving average components. The AIC and BIC values were 566.674 and 567.814, respectively, suggesting a simple and non-complex model. The Ljung-Box test (p = 0.29) and heteroskedasticity test (p = 0.86) confirmed that the residuals were uncorrelated and homoskedastic, supporting the model's validity.

The formation of a very simple ARIMA(0,0,0) model may be attributed to the consistently stable visitation pattern at Hotel Grand Whiz over the observed period. The absence of trends, seasonality, or significant fluctuations in the time series data indicates that the number of visitors remained relatively constant from month to month. As a result, the model does not require any autoregressive or moving average terms to account for temporal dependencies. This simplicity reflects the sufficiency of a constant term alone to represent the data accurately. A brief discussion of this aspect highlights how certain hotel datasets with minimal variance may yield highly parsimonious models without compromising predictive accuracy.

```
Model Summary for Hotel: PURI INDAH HOTEL
                               SARIMAX Results
==============================================================================
Dep. Variable:      Jumlah Pengunjung   No. Observations:                   36
Model:                 ARIMA(1, 0, 0)   Log Likelihood                -236.828
Date:                Mon, 26 May 2025   AIC                            479.656
Time:                        09:04:17   BIC                            484.407
Sample:                    01-01-2022   HQIC                           481.314
                         - 12-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       1391.4544     61.726     22.543      0.000    1270.474    1512.435
ar.L1          0.4003      0.142      2.826      0.005       0.123       0.678
sigma2      3.016e+04   6012.272      5.017      0.000    1.84e+04    4.19e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):                19.68
Prob(Q):                              0.75   Prob(JB):                         0.00
Heteroskedasticity (H):               0.91   Skew:                            -1.31
Prob(H) (two-sided):                  0.87   Kurtosis:                         5.51
===================================================================================
```

Figure 12. Parameter Estimation Result of Puri Indah Hotel

Puri Indah Hotel, the optimal model was ARIMA(1,0,0), with an AR(1) parameter of 0.4034, significant at the 95% confidence level (p = 0.005). This model indicates a dependency on the previous observation. The AIC (479.656) and BIC (484.407) values demonstrate the model's efficiency with moderate complexity. Residual diagnostic tests indicated that the model met the fundamental assumptions, although the Jarque-Bera test suggested a slight deviation from normality.

```
Model Summary for Hotel: HOTEL ARRAYANA
                               SARIMAX Results
==============================================================================
Dep. Variable:      Jumlah Pengunjung   No. Observations:                   36
Model:                 ARIMA(0, 0, 1)   Log Likelihood                -270.849
Date:                Mon, 26 May 2025   AIC                            547.698
Time:                        09:04:17   BIC                            552.449
Sample:                    01-01-2022   HQIC                           549.356
                         - 12-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const       1255.9964    142.646      8.805      0.000     976.416    1535.577
ma.L1          0.6644      0.161      4.125      0.000       0.349       0.980
sigma2      1.974e+05   4.77e+04      4.140      0.000    1.04e+05    2.91e+05
===================================================================================
Ljung-Box (L1) (Q):                   0.42   Jarque-Bera (JB):                 3.81
Prob(Q):                              0.52   Prob(JB):                         0.15
Heteroskedasticity (H):               7.44   Skew:                             0.76
Prob(H) (two-sided):                  0.00   Kurtosis:                         3.50
===================================================================================
```

Figure 13. Parameter Estimation Result of Hotel Arrayana

Hotel Arrayana, the appropriate model was ARIMA(0,1,1), with an MA(1) parameter of -0.9664, which was highly significant (p < 0.01). This model was selected after differencing to address the non-stationary nature of the data. The AIC and BIC values were 547.698 and 552.449, respectively, indicating a good model fit with reasonable complexity. However, the heteroskedasticity test (p = 0.00) indicated the presence of variance instability in the residuals, suggesting the need for further analysis or potential model refinement.

The limited adaptability of the ARIMA(0,1,1) model to the visitor data of Hotel Arrayana is caused by the irregular fluctuation patterns within the time series. The data exhibit dynamic changes that involve sudden increases or decreases in visitor numbers, which cannot be effectively captured by the linear structure of the ARIMA model. Additionally, the presence of variance instability, as indicated by the heteroskedasticity test, confirms that the model does not fully account for the varying dispersion in the data across time.

Diagnostic Checking

The residuals of the ARIMA models were evaluated to ensure they met the fundamental statistical assumptions, namely randomness (white noise), absence of autocorrelation, and approximate normal distribution. This diagnostic step is critical to validate the adequacy and reliability of the selected model. The evaluation process included visual inspection of residual plots to detect any obvious non-random patterns, analysis of autocorrelation through the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to confirm the absence of serial correlation, and assessment of distributional properties using histograms overlaid with Kernel Density Estimation (KDE) curves. Additionally, statistical tests such as the Ljung-Box Q-test were employed to verify the independence of

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.388
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

residuals. These combined methods help ensure that the residuals behave as white noise, confirming that the model has captured the underlying structure in the data adequately (Dar et al., 2024).

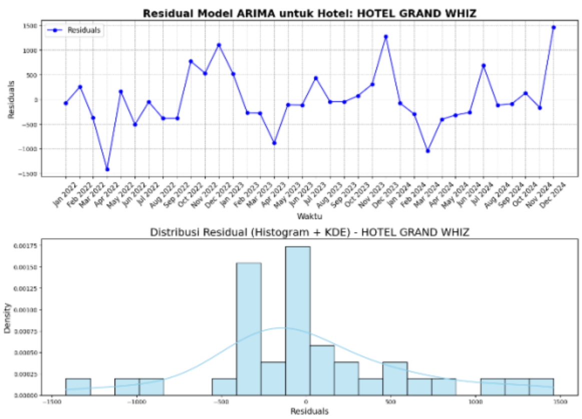


Figure 14. Diagnostic Checking Hotel Grand Whiz

Hotel Grand Whiz, the residual plot showed considerable fluctuations without a discernible pattern. The histogram of the residuals approximated a normal distribution, although some outliers were observed. This suggests that the model captured most of the underlying patterns in the data, although some variability remained unexplained, as also indicated by the kurtosis and skewness values from the parameter estimation stage.

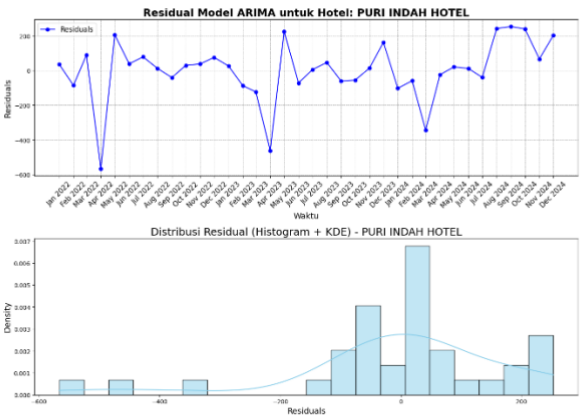


Figure 15. Diagnostic Checking Puri Indah Hotel

Puri Indah Hotel, the residuals were relatively evenly distributed around zero and did not exhibit any systematic patterns. The distribution was nearly symmetrical, albeit slightly skewed to the right. These characteristics indicate that the ARIMA(1,0,0) model implemented for this hotel effectively captured the dynamics of the historical data and produced stable forecasts. The absence of recurring patterns or autocorrelation in the residuals supports the validity of the model.

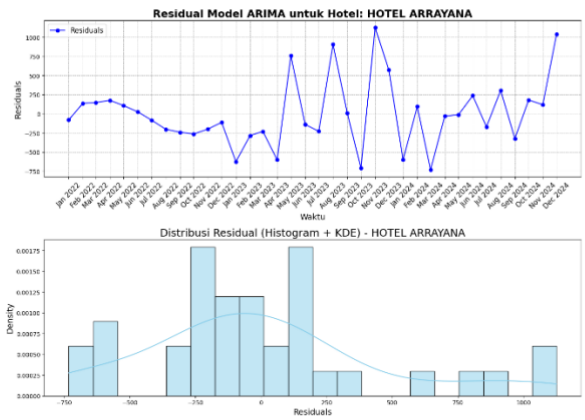


Figure 16. Diagnostic Checking Hotel Arrayana

Hotel Arrayana, the residuals showed higher fluctuations, and the histogram indicated a non-normal distribution with a bimodal shape. This suggests the presence of seasonal or structural components in the data that were not fully captured by the ARIMA(0,1,1) model. Although no strong autocorrelation was observed in the residuals, the deviation from normality and the presence of extreme values imply that further model exploration or the inclusion of external variables may be necessary to improve forecasting accuracy for data with more complex dynamics.

ARIMA Model Evaluation

The evaluation of the ARIMA model's performance was conducted using three primary accuracy metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics quantify the deviation between the actual observed values and the predicted values generated by the model. MSE and RMSE indicate the magnitude of prediction errors, where lower values signify better model performance. RMSE is especially interpretable as it retains the same unit as the original data, allowing direct comparison with actual values (Si, 2022).

	Hotel	MSE	RMSE	MAPE
0	HOTEL GRAND WHIZ	339,896.44	583.01	18.44%
1	HOTEL ARRAYANA	197,595.44	444.52	32.31%
2	PURI INDAH HOTEL	30,169.62	173.69	9.65%

Figure 17. Result ARIMA Model Evaluation

The ARIMA model demonstrated the highest accuracy for Puri Indah Hotel, with a MAPE

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.388

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

of 9.65%, which falls into the "very accurate" category. In contrast, Hotel Arrayana showed the lowest accuracy, with a MAPE of 32.31%, indicating a "reasonable" level of prediction. Meanwhile, Hotel Grand Whiz achieved a MAPE of 18.44%, categorized as "accurate," although the relatively high RMSE and MSE values suggest the presence of significant data fluctuations.

Result Forecasting

The ARIMA model was applied to forecast hotel visitor numbers for the next six months for each of the three selected hotels. The prediction results are visualized using time series plots that illustrate the historical data, fitted values from the model, and future forecasts, accompanied by 95% confidence intervals. These visualizations serve as a diagnostic and interpretive tool, allowing for a clear comparison between actual historical trends and the model's projected values. If the forecast line follows the trajectory of historical patterns and remains within reasonable proximity of past fluctuations, it suggests that the model has effectively captured the underlying data structure. The confidence intervals indicate the expected range of future observations, helping decision-makers to understand the uncertainty surrounding the forecasts. Narrow intervals suggest higher model certainty, while wider bands indicate greater variability or volatility in the data.

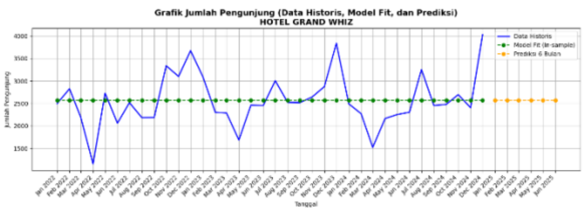


Figure 18. Forecasting Result of Hotel Grand Whiz

Hotel Grand Whiz, the ARIMA model forecasts a relatively stable pattern, with predicted values remaining close to the historical average. Although the predictions fall within a reasonable range, the model appears insufficient in capturing the actual fluctuations observed in the data. This suggests that the model may be too simplistic and does not fully reflect the underlying dynamics of visitor patterns.

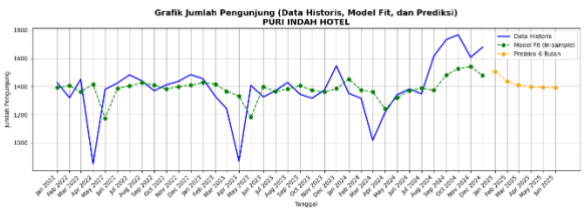


Figure 19. Forecasting Result of Puri Indah Hotel

Puri Indah Hotel, the forecast indicates a gradual downward trend over the next six months. The model successfully captures the seasonal trends present in the historical data, as evidenced by the alignment between the model fit and actual observations. The forecast intervals are relatively narrow, indicating model stability in representing the data structure.

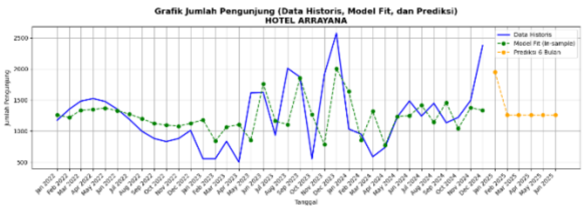


Figure 20. Forecasting Result of Hotel Arrayana

Hotel Arrayana, the forecast reveals a significant decline in visitor numbers. The predicted pattern does not fully reflect the fluctuating nature of the historical data, suggesting that the model lacks adaptability to high variability and may fail to capture unaccounted seasonal components. Therefore, further exploration of alternative models or more advanced approaches is recommended to improve forecasting accuracy for data with complex dynamics.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study successfully implemented the ARIMA model to forecast hotel visitor numbers in Mojokerto Regency using data from the Department of Culture, Youth, Sports, and Tourism (Disbudporapar) for the period 2022 to 2024. The results show that ARIMA produced varying levels of forecasting accuracy, with the highest performance achieved for Puri Indah Hotel (MAPE: 9.65%) and the lowest for Hotel Arrayana (MAPE: 32.31%). The analytical process including stationarity testing, parameter identification, model estimation, and accuracy evaluation demonstrated that ARIMA effectively captures historical patterns, especially in more stable datasets. The use of automated ARIMA (auto ARIMA) further enhanced accuracy and consistency in selecting optimal model configurations across different hotel datasets. Overall, the model offers a valuable tool for hotel operational planning and strategic decision-making in the regional tourism sector.

Suggestion

Future research is recommended to develop predictive models capable of handling high variability and complex seasonal components, such as SARIMA or hybrid models like ARIMA-LSTM. In addition, incorporating external variables such as seasonality, national holidays, and local tourism events may enhance forecasting accuracy. The development of an interactive web-based forecasting system is also suggested to support real-time decision-making by Disbudporapar and hotel management.

REFERENCES

As, M., Wibowo, S. S., & Sophia, E. (2017). *PERAMALAN JUMLAH MAHASISWA BARU DENGAN MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)*. *2*(3), 20–33.

Cherrly, A., & Somya, R. (2023). Prediksi Penjualan Tiket Wisata Taman Bermain Menggunakan Metode ARIMA. *Techno.Com*, *22*(2), 312–322. https://doi.org/10.33633/tc.v22i2.7950

Dar, A. A., Jain, A., Malhotra, M., Farooqi, A. R., Albalawi, O., Khan, M. S., & Hiba. (2024). Time Series analysis with ARIMA for historical stock data and future projections. *Soft Computing*, 1-12.

Dewi, M. S., & Sutarman, S. (2022). Estimasi Parameter Model Arch Menggunakan Metode Bayes. *Journal of Fundamental Mathematics and Applications (JFMA)*, *5*(2), 162–175. https://doi.org/10.14710/jfma.v5i2.15029

Fazrina, N. A., Wijaya, D. R., Hadie, B. A., & Budiwati, S. D. (2024). Poverty Level Prediction Based on Time Series Data using Auto Arima. *BIO Web of Conferences*, *144*. https://doi.org/10.1051/bioconf/202414404008

Febiola, A., Dewi, A., Fazarin, F. M., Ramadhani, F., Khaffi, M. A., Akbar, R., & Dalimunthe, D. Y. (2024). Perbandingan Metode ARIMA dan SARIMA Dalam Peramalan Jumlah Penumpang Bandara Provinsi Kepulauan Bangka Belitung. *Jambura Journal of Mathematics*, *6*(2), 160–168. https://doi.org/10.37905/jjom.v6i2.25081

Firdaus, M., Studi, P., & Informatika, T. (2024). *PERBANDINGAN MODEL ARIMA MANUAL DAN AUTO DALAM*. *9*(2), 147–155.

Insti, S. L. (2024). *Development of an ARIMA Model to Predict the Monthly Price of Bitcoin in USD*. *July 2021*.

Jumlah, P., Wisatawan, K., …… M., Cristanto, M., & Mailoa, E. (2023). *Peramalan Jumlah Kedatangan Wisatawan Mancanegara dengan Menggunakan Model ARIMA*.

Mazlan, N., Izham, S. S. B., & Ahmadun, M. B. (2023). Muslim Friendly Hotels As Islamic Tourism Products: Issues and Challenges. *International Journal of Academic Research in Business and Social Sciences*, *13*(2), 241–255. https://doi.org/10.6007/ijarbss/v13-i2/16344

Pariwisata, D., Olahraga, K. D. A. N., & Pengantar, K. (2020). *Dinas pariwisata, kepemudaan dan olahraga kabupaten mojokerto*. 1–53.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., … Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, *38*(3), 705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001

Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, T. (2020). 済無No Title No Title No Title. *Journal GEEJ*, *7*(2).

Pratama, I. G. S. (2020). The impact of tourism development on the economic, cultural and environmental aspects of local communities. *International Research Journal of Management, IT and Social Sciences*, *7*(1), 31–36. https://doi.org/10.21744/irjmis.v7n1.819

Rizkya, D., Roosaputri, H., & Dewi, C. (2023). Perbandingan Algoritma ARIMA, Prophet, dan LSTM dalam Prediksi Penjualan Tiket Wisata Taman Hiburan (Studi Kasus: Saloka Theme Park). *Jurnal Penerapan Sistem Infomatika (Komputer & Manajemen)*, *4*(3), 507–517.

Si, Y. (2022). Using ARIMA model to analyse and predict bitcoin price. *DCP Business and Management*, (34), 1210-1216.

Suhendra, C. D., Marini, L. F., & Sarungallo, A. (2023). Prediksi Mahasiswa Baru Universitas Papua Menggunakan Autoregressive Integrated Moving Average. *Jurnal Informatika*, *10*(2), 163–172. https://doi.org/10.31294/inf.v10i2.16637

Umam, K. (2023). MIND (Multimedia Artificial Intelligent Networking Database Perbandingan Metode ARIMA dan LSTM pada Prediksi Jumlah Pengunjung Perpustakaan. *Journal MIND Journal | ISSN*, *8*(2), 119–129. https://doi.org/10.26760/mindjournal.v8i2.119-129

Vaibhava Lakshmi, R., & Radha, S. (2021). Time series forecasting for the adobe software company's stock prices using ARIMA (BOX-JENKIN') model. *Journal of Physics: Conference Series*, *2115*(1). https://doi.org/10.1088/1742-6596/2115/1/012044

Vean, A. C., & Witanti, A. (2024). Prediksi Kurs Rupiah Terhadap Euro Dan Pound Menggunakan Arima. *Dinamika Rekayasa*, *20*(1), 39–47. https://doi.org/10.20884/1.dinarek.2024.20.1.24

Wahyu ngestisari. (2020). The Perbandingan Metode ARIMA dan Jaringan Syaraf Tiruan untuk Peramalan Harga Beras. *Indonesian Journal of Data and Science*, *1*(3), 96–107. https://doi.org/10.33096/ijodas.v1i3.18