Forecasting Hotel Occupancy Rate in Riau Province Using ARIMA and ARIMAX

F A Rizalde[1,2], S Mulyani[3], N Bachtiar[3]
[1] Statistics Study Program, Department of Mathematics
[2] Faculty of Mathematics and Natural Science, University of Riau
[3] Regional Account and Statistics Analysis Department, BPS-Statistics of Riau Province

*Corresponding author's e-mail: fadlika.arsy3039@student.unri.ac.id

Abstract. Hotel Occupancy Rate is one of the important leading indicators for calculating the Accommodation Sub-Category of Gross Regional Domestic Product (GRDP). By the extreme decline of the Hotel Occupancy Rate data due to COVID-19 and the unavailability of current data to counting GRDP quarterly, the Hotel Occupancy Rate prediction needs to do with the appropriate forecasting method. The authors use data from Google Trends as an additional variable in predicting the Hotel Occupancy Rate using the ARIMAX model and then compares it with the ARIMA model. The results showed that the ARIMAX model had better accuracy than ARIMA, with a MAPE value of 9.64 percent and an RMSE of 4.21 percent. This research concluded that if there is no change in government policy related to social restrictions until the end of the year, the ARIMAX model predicts the December 2021 Hotel Occupancy Rate of 38.59 percent.

1. Introduction

From March 2020 until now, all regions in Indonesia have been hit by the COVID-19 pandemic, including Riau Province. As of July 31 in 2021, there were 97,123 confirmed COVID-19 cases in Riau Province, as shown in Figure 1[1]. With a high spike in cases, the government implemented a community activity restriction policy to prevent the spread of COVID-19. However, the policy has disrupted the social and economic activities of the community. This has resulted in Gross Regional Domestic Product (GRDP) decreasing in various industries, including the hotel industry.

Badan Pusat Statistik (BPS) calculates GRDP quarterly in all industries. The hotel industry in GRDP is recorded in the Accommodation Sub-Category. To calculate this sub-category, we need one of the leading indicators, the Hotel Occupancy Rate. BPS collects Hotel Occupancy Rate data with coverage of star hotels with a release schedule one month after the data collection month ends. During the COVID-19 pandemic, this indicator experienced an extreme decline. This condition is in line with the decrease in tourist visits in several regions[2]. In Figure 2, we can see that the Hotel Occupancy Rate decreased drastically in April 2020. This is due to Riau Provincial Government issued a policy that refers to the Minister of Transportation Regulation No. 25 of 2020, which is closing access to airports and seaports for passenger departures/arrivals from and to neighboring countries (Malaysia and Singapore). This policy has resulted in a decrease in tourist visits and stays at the hotel. In line with the loosening of the policy on restricting community activities, the Hotel Occupancy Rate indicator is slowly starting to increase even though it has not been fully normal until today.

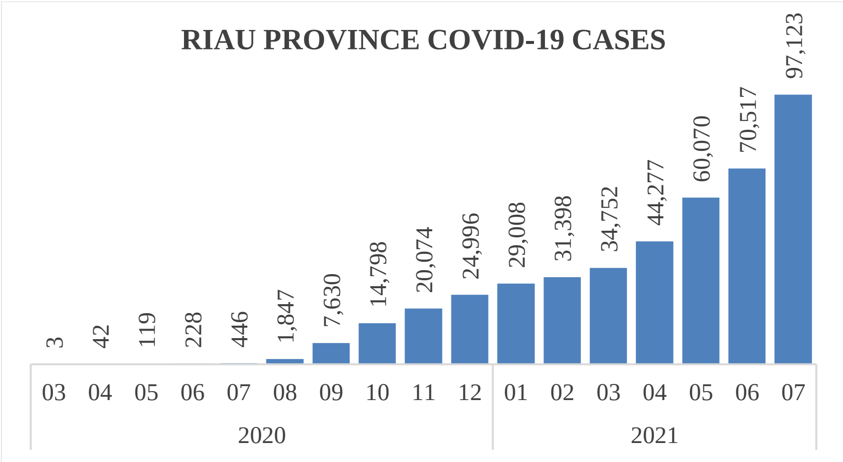

Figure 1. Riau Province Covid-19 Cases.

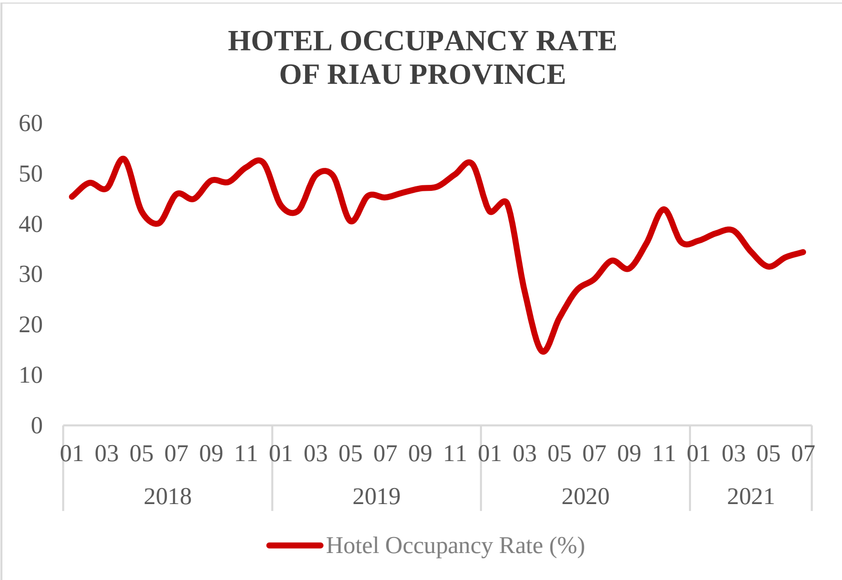

Figure 2. Hotel Occupancy Rate of Riau Province

The COVID-19 pandemic has caused the Hotel Occupancy Rate indicator to experience an extreme decline. Meanwhile, the need for availability of the latest data of this indicator for GRDP calculation has a month lag releases. Based on this issue, a forecasting method is needed to get the latest prediction numbers and predict extreme data conditions. There are several studies related to forecasting Hotel Occupancy Rates. Rahma (2021)[3] was predicting West Sumatra Hotel Occupancy Rate using ARIMA. Silalahi (2015)[4] predicts Hotel Occupancy Rate by comparing the ARIMA method with the Transfer Function. Ramadiani, et al.(2019)[5] were predicting Hotel Occupancy Rate using Double Exponential Smoothing in East Kalimantan.

In forecasting, the time series method widely used is Autoregressive Integrated Moving Average (ARIMA). ARIMA is a suitable statistical method for predicting a variable quickly, simply, cheaply, and accurately cause it requires only one variable. ARIMA is also known as the Box-Jenkins time series method that consists of three models that are: Autoregressive (AR), Integration (I), and Moving Average (MA)[6]. The development of the ARIMA model using additional variables is known as the Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX). The ARIMAX model can be an option if we suspect that other variables can help explain the model characterized by increasing accuracy values[7].

A study using the ARIMAX model was conducted by Ayuningtyas & Wirawati (2019)[8]. In their study, ARIMAX is more accurate than the time series regression method in predicting the Hotel Occupancy Rate in Indonesia, as evidenced by the smaller RMSE and MAPE values. They use the GT Index as an exogenous variable because it is related to the current activity that is happening and can help predict the next period's data release. Based on this research, we use the GT Index to forecast the Hotel Occupancy Rate in one of the regions, Riau Province. We will compare the ARIMA forecasting method with ARIMAX to determine the best model for forecasting the Hotel Occupancy Rate of Riau Province.

2. Methodology

2.1 Data

The data used in this study is the Riau Province Hotel Occupancy Rate as a response variable and GT Index data as an additional variable or exogenous variable:

1. Hotel Occupancy Rate is the level of efficiency and productivity of a hotel in a certain place in a certain period[9]. Hotel Occupancy Rate is a comparison of the number of room nights used with the number of available room nights. If the Hotel Occupancy Rate is high, then the hotel accommodation in an area is in high demand by visitors. On the contrary, if the Hotel Occupancy Rate is low, then the accommodation in an area is less attractive to visitors[10]. Hotel Occupancy Rate is one of the leading indicators because it will determine the direction and magnitude of the Accommodation Sub-Category. The Hotel Occupancy Rate period used in this study is January 2018 to July 2021.
2. Google Trends was first released in 2006. It is available in a web search statistics graph that displays the popularity of search topics over a certain period that can show by city, region, or period. Google trends data have characteristics that are part of the characteristics of big data[11]. The GT index was obtained from google trends web[12]. The number of keywords that appear is displayed in the form of an index with a value range of 0-100. A value of 100 indicates that the search was at its peak of popularity in the selected search period. In this study, GT Index data was used for the period January 2018 to July 2021. Before we decide the right keywords as additional variables to predict Hotel Occupancy Rate, we carried out several stages of collecting Index data Google Trends**:**
   a. Gather some keywords related to accommodation services.
   b. Adjust the search data set by selecting the location of Riau Province.
   c. Compare the search interest rate of selected keywords.
   d. Testing the correlation between Hotel Occupancy Rate and selected keywords.

Selected keywords are obtained from keywords with the highest index. The index value reflects the search interest in a particular topic[13]. Searches that have a high relation with a keyword will have also has a high search interest. To support the selection of keywords, the authors tested the level of correlation with the main variable, Hotel Occupancy Rate.

2.2 Research steps

There are several stages of research to forecasting by comparing the ARIMA and ARIMAX models, as follows:

1. Plot time-series data
2. Stationarity test with Augmented Dickey-Fuller (ADF) test. Differencing the data while it is not stationary
3. Identify the ARIMA model from the PACF and ACF plot
4. From the ARIMA model formed, a significant parameter test is conducted and the selection of the best criteria with the smallest AIC value and diagnostic model checks on residual.
5. Next, for ARIMAX modeling, do a stationarity test for the exogenous variables used. Differencing the data while it is not stationary
6. Perform ARIMAX modeling, using the best ARIMA model and stationary exogenous variables
7. Perform significant parameter tests and model diagnostic checks on the residuals from the model ARIMAX.
8. Forecasting with ARIMA and ARIMAX model

9. Check forecasting accuracy using RMSE and MAPE to determine the best model in the identification of forecast results
10. Conclude the results of the analysis.

2.3 Pearson Correlation (Product Moment)

We need additional variables when forecasting with the ARIMAX model. In this study, we use the GT Index. A correlation test between Hotel Occupancy Rate and selected keywords is needed to know the relationship between two variables. Correlation analysis can see the strength and the direction between two variables[14]. The positive correlation indicates the relationship of variables is moving in the same direction, while the negative correlation indicates the relationship of variables is not moving in the same direction. One of the correlation coefficients used is the Pearson Correlation (Product Moment), where the data used in this correlation is at least an interval scale. A strong relationship between two variables is achieved if the Pearson correlation value is between 0.61-0.80 and the correlation will be very strong if the correlation coefficient is 0.81-1.00[14].

2.4 Forecasting

Forecasting is a method for predicting some future event. Forecasting can be in the short, medium, and long term. Short-term forecasting can predict daily, weekly, or monthly for the period to come. The medium-term can predict for a period of one to two years to come. The long term can predict more than two years. Forecasting with short and medium-term needs to look at the historical data patterns of the previous period and identify the model. Statistic methods can use in conducting short and medium-term forecast[15].

The time series analysis method is an analytical technique in forecasting. One of the methods is the Autoregressive Integrated Moving Average (ARIMA). The ARIMA method is a more appropriate method to be used in short-term time forecasting, where the data used must be stationary [16]. The development of the ARIMA model using additional variables is known as the Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) method. The ARIMAX model can be used, if it is suspected that other variables can help explain the model by increasing the value of accuracy[7].

2.5 Autoregressive Integrated Moving Average (ARIMA)

In theory, the ARIMA model is one of the best in forecasting time series data. This stage in forecasting modeling includes a pair of models that meet, estimate parameters, and verifies the model[17]. ARIMA modeling is composed of three time-series models that are: autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA). The model can be used if the time series data has met the assumption of stationarity.

When the data is not stationary, it is necessary to do differencing to make the non-stationary data stationary[18]. So the general model of ARIMA consists of order p as AR, order d as differencing, and q as order MA. So the equation model of ARIMA (p, d, q) is as follows [19]:

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) a_t \tag{1}$$

where:

$(1-B)^d$	= differencing
$\phi_p(B) = \left(1-\phi_1 B - \cdots - \phi_p B^p\right)$	= the parameters of AR
$\theta_q(B) = \left(1-\theta_1 B - \cdots - \theta_q B^q\right)$	= the parameters of MA
a_t	= residual at time-*t*

After the data used is stationary, we can continue to identify the ARIMA model. We used plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine the possible order of ARIMA models[6]. PACF identifies the AR model and ACF identifies the MA model.

2.6 Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)

ARIMAX is a forecasting model where it is the expansion of the ARIMA model[20]. ARIMAX uses additional variables in forecasting. Lingga, et al.(2019)[7] forecasted tourist arrivals to tourist attractions with the GT Index using the ARIMAX model, where the parameters were obtained from the ARIMA model parameters that had fulfilled all the assumptions tests. In general, the form of the ARIMAX model (p,d,q) is,

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B)a_t + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_p X_{p,t} \quad (2)$$

where:

$(1-B)^d Z_t$	= differencing
$\phi_p(B) = \left(1-\phi_1 B - \cdots - \phi_p B^p\right)$	= the parameters of AR
$\theta_q(B) = \left(1-\ \theta_1 B - \cdots - \theta_q B^q\right)$	= the parameters of MA
a_t	= residual at time-t
$X_{p,t}$	= exogenous variables
β_p	= coeficients of exogenous variables

The steps taken in the identification of the ARIMAX model are as follows[7]:
1. Determine additional variables as external variable/exogenous variables
2. Perform ARIMA modeling on the main variables used
3. From the ARIMA model that has been selected, proceed to the ARIMAX modeling stage.

2.7 Stationarity and differencing

In forecasting, the data time series used must be able to satisfy the assumption of stationarity to use forecasting models. To be able to test whether the data used have met the assumption of stationarity, we use Augmented Dickey-Fuller (ADF) [20]. The hypothesis of the test is,

H_0: there is a unit root in the data (data is not stationary)
H_1: there is no unit root in the data (data is stationary)

The ADF test is formulated as $t_{count} = \frac{\widehat{\emptyset}-1}{sd(\widehat{\emptyset})}$ where $\widehat{\emptyset}$ is the estimate for the AR parameter and $sd(\widehat{\emptyset})$ is the standard deviation of the AR parameter estimate. We reject H_0 if the p-value < α (significant level) used is 0.05. When the data used does not meet the assumption of stationarity, it is necessary to do differencing until the data is close to stationary.

2.8 Parameter significant test

To find out the formed model can be used, the parameter significance test is carried out with the hypothesis:

H_0: Parameters in the model are not significant
H_1: Parameters in the model are significant

Conclusion criteria, H_0 is rejected if p-value < α (significant level) used is 0.05. It can be concluded that the parameters in the model are significant. If there are many possible models formed and all parameters are significant, we can use the minimum value of Akaike Information Criteria (AIC) as the best model selection criteria[20]. The AIC calculation formula is given by equation (3),

$$AIC = -2\,log(maximum\ likelihood) + 2k \quad (3)$$

Where k is $p+q+1$ if the intercept is constant or $p+q$ if the intercept is not constant.

2.9 Model diagnostics test

To identify whether the obtained model is appropriate for use, it needs to do test the suitability of the model which includes residual autocorrelation and residual normality tests.

2.9.1 Residual autocorrelation test

To test the residual autocorrelation, the Ljung-Box test statistic can be used, with the hypothesis:

$H_0: \rho_1 = \rho_2 = \cdots = \rho_m = 0$ (there is no autocorrelation until the residue until the *m-th* lag)
H_1: at least one $\rho_i \neq 0$ for i=1,2, …, m (at least one lag residue has autocorrelation)

The Ljung-Box test is formulated as $Q = T(T+2)\sum_{k=1}^{m} \frac{\tilde{\rho}_k^2}{T-K}$, where T is the amount of data, k as the number of lags, $\tilde{\rho}_k$ shows autocorrelation of the *k-th* sample lag and the maximum tested m lag. Conclusion criteria H_0 is accepted if the p-value > α (significant level) used is 0.05.

2.9.2 Residual normality test

To test the normality of the residuals in the model, one of them can use the Shapiro-Wilk test, with the hypothesis:

H_0: residual in a normal distributed model
H_1: residuals in the model are not normal distributed

Conclusion criteria, H_0 is accepted if p-value > α (significant level) used is 0.05.

2.10 Forecasting accuracy

To determine the best forecasting of the ARIMA and ARIMAX models, we compare the value of Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The best model is the one with the smallest RMSE and MAPE values. RMSE is the root of the sum of the difference between the actual value and the forecast value then it is divided by the amount of data[21]. MAPE states the percentage error of forecasting results. MAPE value <10% indicates that the model is better to be used by a forecasting model [22]. The RMSE calculation formula is given by equation (4)

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2}{T}} \tag{4}$$

And the MAPE calculation formula is given by equation (5)[15]:

$$\text{MAPE} = \frac{\sum_{t=1}^{T}|(y_t - \bar{y}_t)/y_t|}{T} \times 100 \tag{5}$$

Where:
y_t = actual value of t-th data
$\bar{y}_t$ = forecast value of t-th data
T = time period

3. Results and discussion

Based on Figure 3, it can be seen the value of Hotel Occupancy Rate Hotels in Riau experienced an extreme decline in April 2020. The reason was that the Riau Provincial government issued a policy to close the airport and seaport access for passenger departures/arrivals from and to neighboring countries (Malaysia and Singapore). The Hotel Occupancy Rate figure began to increase slowly. That is in line with the easing of government policies regarding social restrictions.

From the results of the visualization of the data plot, it can be seen that the data pattern does not match the stationary pattern, the form of data not being at a constant mean (mean). Proving the data is stationary or not, a unit root test can be performed Augmented Dickey-Fuller (ADF).

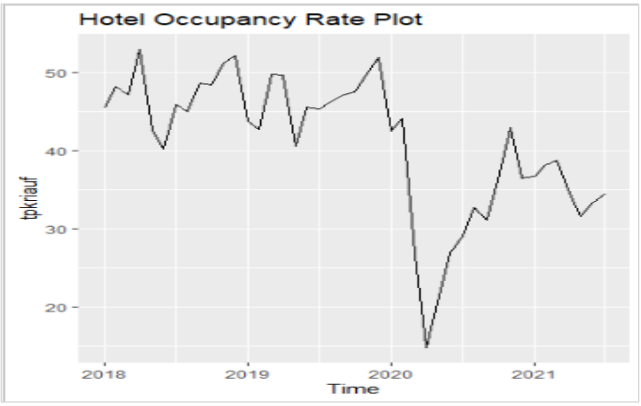

Figure 3. Hotel Occupancy Rate Plot

3.1 Stationarity test

From the results of the ADF test, the Riau Hotel Occupancy Rate data showed the p-value (0.266) > alpha (0.05). That means that H_0 is accepted, it can be concluded that the data Hotel Occupancy Rate in Riau is not stationary. Therefore, it is necessary to do differencing to stationary the data.

3.2 Differencing

From the results of the ADF test on data that has been differencing first-order, it shows that the p-value (0.05573) > 0.05. That means that the data is not stationary. We need to do the second-order differencing. The test results showed a p-value (0.01) < 0.05 and H_0 is rejected. Based on this result, we can continue the ARIMA identification model with the second-order differencing data.

3.3 ARIMA model identification

The ACF and PACF plots of the second-order differencing data can be used to determine the ARIMA model formed. The ACF and PACF plots are given in Figure 4:

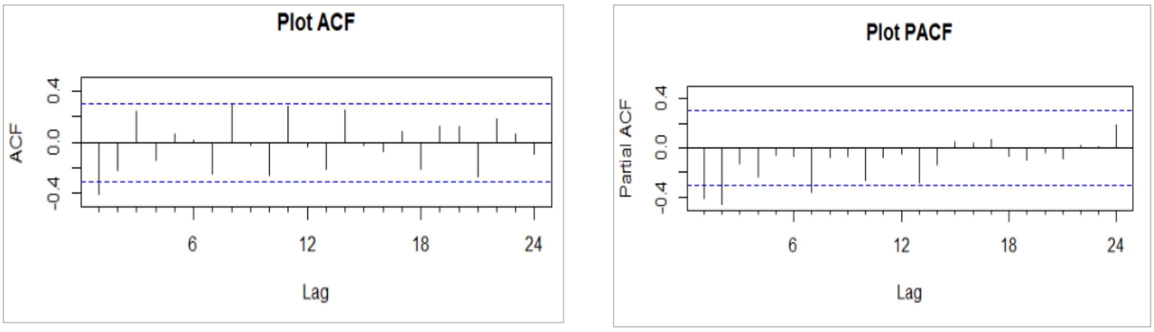

Figure 4. ACF and PACF plots

The ACF plot shows the lag that crosses the lower limit line on the first lag, meaning there is a process of moving average order 1 or MA(1). The results The PACF plot shows the lag that crosses the lower boundary line on the first and second lags, meaning that there is a process of autoregressive order 1 and 2 or AR (1) and AR (2). Due to the second-order differencing, we define d = 2. Then, the ARIMA model that has been selected is a mixed model of ARIMA (p, d, q). ARIMA models that may be formed are ARIMA (1,2,1) and ARIMA (2,2,1).

3.4 Parameter significant test

Based on the results of the significant parameter test, the ARIMA (1,2,1) and ARIMA (2,2,1) models showed that all parameters were significant. It can be indicated by using the p-value of all parameters that are smaller than 0.05. The AIC values are 308.53 and 292.67, respectively. Based on the results of the significant parameter test, both models have significantly overall parameters. We found the smallest AIC value is given using the ARIMA model (2,2,1). Then the model that will be used for the next stage of analysis, the parameter coefficient values, and significant parameters are given in table 1.

Table 1. ARIMA (2,2,1) Model Parameter Estimation Results

AR (1)		AR (2)		MA (1)	
Est.	**Pr (>|z|)**	**Est.**	**Pr (>|z|)**	**Est.**	**Pr (>|z|)**
-0.894	< 0.000	-0.617	< 0.000	-0.999	< 0.000

3.5 ARIMA model diagnostic test

Using the ARIMA model (2,2,1), the results of the diagnostic test model are as follows:

a. Residual Autocorrelation Test (White Noise)
   The test results show the p-value > 0.05, then the model residual of the Riau Hotel Occupancy Rate data has no autocorrelation (white noise).
b. Residual Normality Test
   Based on the results of the normality test with the Shapiro-Wilk test, was found the p-value > 0.05. It was concluded the residual data had a normal distribution.

Residual autocorrelation test values and residual normality of the ARIMA model are given in table 2.

Table 2. ARIMA (2,2,1) Model Diagnostic Test Results

Residual Autocorrelation		Residual Normality	
Chi-Squared	**P-value**	**W**	**P-value**
29.312	0.2086	0.97482	0.4879

3.6 Selection of Keywords for Exogenous Variables

After selecting the location set for Riau Province, keyword information related to accommodation services was collected from January 2018-July 2021 and several popular keywords were obtained: "hotel", "akomodasi", "booking", " penginapan", "homestay". The graph of the popularity of the five keywords can be seen in Figure 5,

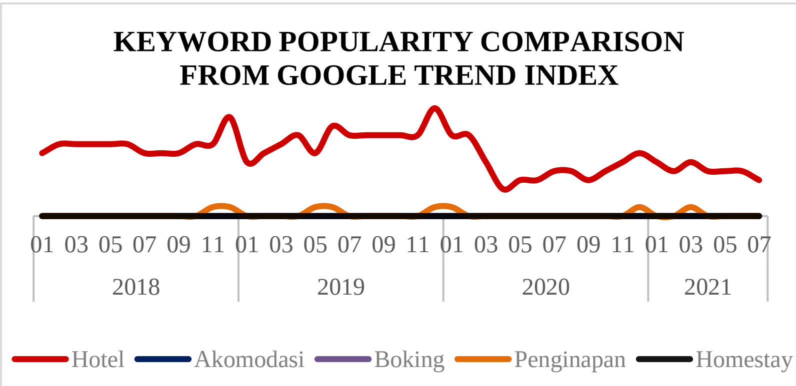

Figure 5. Keyword Popularity Comparison From Google Trend Index

From the graph, it can be seen that the keyword “hotel” has the highest popularity. Furthermore, to test whether the keyword "hotel" has a strong relationship with the Hotel Occupancy Rate variable, a correlation analysis was carried out with the five selected keywords with the Hotel Occupancy Rate data. The results of the correlation analysis with Pearson are given in table 3.

Table 3. Correlation Analysis Using Pearson Correlation

	TPK (%)	hotel	akomodasi	boking	penginapan	homestay
TPK (%)	1.0000					
hotel	0.8233	1.0000				
akomodasi	-0.1303	-0.1390	1.0000			
boking	0.5746	0.6625	-0.0538	1.0000		
penginapan	0.4627	0.5298	-0.0085	0.4874	1.0000	
homestay	0.3202	0.4961	0.0785	0.3897	0.2886	1.0000

From table 3, we know that the keyword "hotel" has a correlation value with the Hotel Occupancy Rate data of 0.8233. This value can be considered as a very strong correlation. Thus, this keyword is used as an exogenous variable in ARIMAX modeling.

3.7 ARIMAX model identification

ARIMA (2,2,1) has been fulfilled all assumptions that used to build ARIMAX (2,2,1) modeling with the addition of an exogenous variable, namely the GT Index with the keyword "hotel" in Riau Province. Before ARIMAX modeling, the GT Index as an exogenous variable must meet the assumption of stationarity. ARIMAX modeling is possible if the differencing response variables and exogenous variables are in the same order. Since the Hotel Occupancy Rate differencing is stationary on the second order, then the second-order differencing is performed for the GT index data. The results of the differencing test for the GT index exogenous variable have met the stationarity test where the p-value (0.01) < 0.05.

Furthermore, the GT index can be used as an exogenous variable in ARIMAX modeling (2,2,1). After testing the significance of the parameters, we found that the ARIMAX model parameters (2,2,1) were significant in the model shown in table 4 (p-value < 0.05). The ARIMAX (2,2,1) model can be continued to the next analysis stage.

Table 4. ARIMAX (2,2,1) Model Parameter Estimation Results

AR (1)		AR (2)		MA (1)		Var Eksogen	
Est.	**Pr (>|z|)**	**Est.**	**Pr (>|z|)**	**Est.**	**Pr (>|z|)**	**Est.**	**Pr (>|z|)**
-0.997	< 0.000	-0.629	< 0.000	-0.999	< 0.000	0.216	0.00083

3.8 ARIMAX model diagnostic test

Based on the ARIMAX model that has met the significant parameter assumptions, then the residual assumption test is carried out by testing the residual autocorrelation (white noise) and residual normality. Using the ARIMAX (2,2,1) results of the test model, the diagnostic model results are as follows:

a. Residual Autocorrelation Test (White Noise)
   Based on the L-Jung Box test, the p-value result is > 0.05. That means that the residuals in the Riau Hotel Occupancy Rate and Google trend data have no autocorrelation (white noise).
b. Residual Normality Test
   Based on the Shapiro-Wilk test, the p-value result is > 0.05. That means that the residual data of Riau Hotel Occupancy Rate and Google trend are normal distribution.

The residual autocorrelation test values and residual normality of the ARIMAX model are given in table 5. The ARIMAX model (2,2,1) has satisfied the test and then can predict the value of the Hotel Occupancy Rate in the five next months.

Table 5. ARIMAX (2,2,1) Model Diagnostic Test Results

Residual Autocorrelation		Residual Normality	
Chi-Squared	**P-value**	**W**	**P-value**
28.856	0.2256	0.95714	0.1248

3.9 Forecasting

Using the ARIMA (2,2,1) and ARIMAX (2,2,1) models, the forecasting results for August-December 2021 are obtained and presented in the table below:

Table 6. Result of Forecasting

Period	ARIMA (2,2,1)	ARIMAX (2,2,1)
August 2021	33.80580	39.43796
September 2021	33.31278	38.65795
October 2021	33.08515	38.97804
November 2021	32.84031	38.81707
December 2021	32.55239	38.59013

Based on table 6, we can see that the forecasting results Hotel Occupancy Rate in Riau Province using the ARIMA model (2,2,1) shows the percentage of Hotel Occupancy Rate decreased from August 2021 by 33.81 percent to 32.55 percent in December 2021. Using the ARIMAX model (2,2,1), the percentage of Hotel Occupancy Rate increased in August 2021 by 39.43 percent but then decrease in September 2021 by 38.66 percent. This number continues to decrease by December 2021 to 38.59 percent. This value can be described by the plot of the results for forecasting the Hotel Occupancy Rate in Riau Province using the ARIMA (2,2,1) and ARIMAX (2,2,1) models in the following figure:

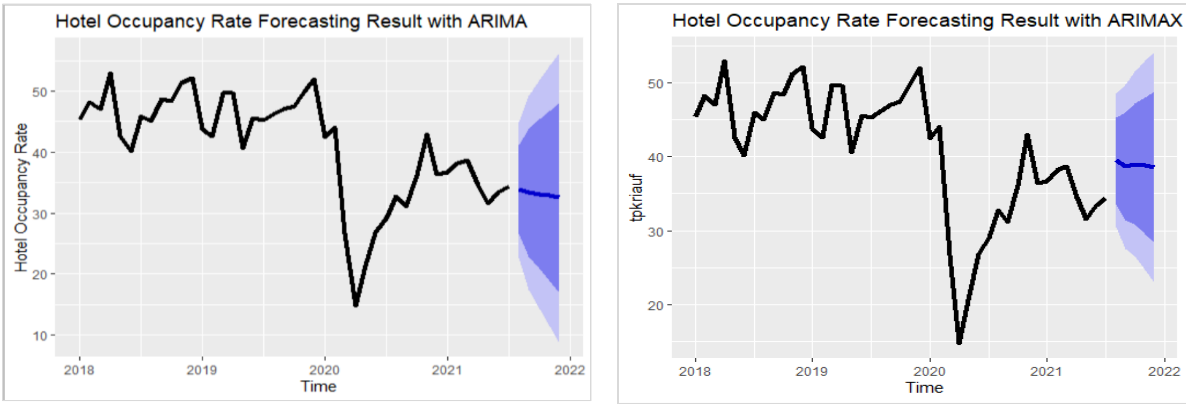

Figure 6. ARIMA and ARIMAX Model Forecasting Results

From the results of forecasting accuracy based on RMSE and MAPE values, ARIMA (2,2,1) and ARIMAX (2,2,1) models are given in table 7:

Table 7. Comparison Accuracy Value

Model	MAPE	RMSE
ARIMA (2,2,1)	11.05945	5.251983
ARIMAX (2,2,1)	9.636742	4.209775

Based on the results, it can be seen RMSE and MAPE from ARIMAX models (2,2,1) with Index GT (as an exogenous variable) with the keyword "hotel" in Riau Province has a smaller value than the ARIMA model (2,2,1). That informs that the best method for predicting the Hotel Occupancy Rate of Riau Province from August to December 2021 is ARIMAX (2,2,1) because it increases forecasting accuracy with a MAPE value of 9.64 percent and an RMSE value of 4.21 percent. These results are in line with research by Lingga, et al.(2019)[7] which predicts tourist arrivals to tourist attractions with the GT Index using ARIMAX. In their study, it was shown that the ARIMAX is a better model than the ARIMA because it can increase accuracy.

4. Conclusion

ARIMAX is an ARIMA model development that can be used properly to forecast the Hotel Occupancy Rate. By utilizing big data that has been provided by Google, such as Google Trends, using the GT Index with the right keywords can improve modeling accuracy. Based on the forecasting results, can be concluded that the ARIMAX model has a better forecasting ability than ARIMA by a MAPE value of 9.64 percent and an RMSE of 4.21 percent. From the results of this study, we suggest using the ARIMAX method to fulfill the availability of other current indicators. ARIMAX can be applied using other exogenous variables while still considering the high correlation with other indicators that need to be predicted.

References

[1] COVID-19 RT. Riau Tanggap COVID-19. Satgas Covid-19 Provinsi Riau 2021. https://corona.riau.go.id/.
[2] Premiere I. Kunjungan Wisman September Anjlok 5,94% Secara Bulanan 2020. https://www.indopremier.com.
[3] Hariyanti R. Peramalan Tingkat Penghunian Kamar Hotel Berbintang di Provinsi Sumatera Barat Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA). Universitas Negeri Padang, 2021.
[4] Silalahi MD. Peramalan Tingkat Penghunian Kamar Hotel Bintang Menggunakan Metode Arima Dan Fungsi Transfer. 2015.
[5] Ramadiani, Wardani N, Kridalaksana AH, Jundillah ML, Azainil. Forecasting the Hotel Room Reservation Rate in East Kalimantan Using Double Exponential Smoothing. Fourth Int. Conf. Informatics Comput., 2019.
[6] El-Mallah E, Elsharkawy S. Time-Series Modeling and Short Term Prediction of Annual Temperature Trend on Coast Libya Using the Box-Jenkins ARIMA Model. Adv Res 2016;6:1–11. https://doi.org/10.9734/air/2016/24175.
[7] Lingga NN, Rohmawati AA. Pemodelan dan Peramalan Kedatangan Wisatawan ke Tempat Wisata dengan Google Trends Menggunakan Metode Variasi Kalender ARIMAX. Proceeding Eng., vol. 8, 2021, p. 3361–72.
[8] Ayuningtyas I, Wirawati I. Nowcasting Tingkat Penghunian Kamar Hotel Menggunakan Google Trends (Nowcasting Occupancy Rate Of Hotel Room With Google Trends). Semin. Nas. Off. Stat. 2019, 2019, p. 338–43.
[9] Badan Pusat Statistik. Tingkat Penghunian Kamar Hotel. 2020.
[10] Badan Pusat Statistik. Tingkat Penghunian Kamar (TPK) Hotel. SiRuSa n.d. https://sirusa.bps.go.id/sirusa/index.php/indikator/60.
[11] Askitas N. Google search activity data and breaking trends. IZA World Labor 2015:1–10. https://doi.org/10.15185/izawol.206.
[12] Google. Google Trends n.d. https://trends.google.co.id/trends/?geo=ID.
[13] Google. Bantuan Trends n.d. https://support.google.com/trends/answer/4365533?hl=id.
[14] Hidayanti AA, Prathama BD, Wardah S, Pendidikan J, Agung D, Mataram K, et al. Analisis Korelasi Pearson Dalam Menentukan Hubungan Kualitas Produk, Pelayanan, Lokasi Dan Kepuasan Terhadap Loyalitas Pada Pelanggan Rumah Nutrisi Herbalife Mataram. J Innov Knowl 2021;1.
[15] Montgomery DC, Jennings CL, Kulachi M. *Introduction to Time Series Analysis and Forecasting*. New Jersey: John Wiley & Sons; 2007.
[16] Yunita T. Peramalan Jumlah Penggunaan Kuota Internet Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA). J Math Theory Appl 2019;1:16–22.
[17] Anokye R, Acheampong E, Owusu I, Isaac Obeng E. Time series analysis of malaria in Kumasi:

Using ARIMA models to forecast future incidence. Cogent Soc Sci 2018;4:1–13. https://doi.org/10.1080/23311886.2018.1461544.
[18] Cryer JD, Chan K-S. *Time Series Analysis With Applications in R*. Springer; 2008. https://doi.org/10.1016/0377-2217(85)90052-9.
[19] Wei WWS. *Time Series Analysis: Univariate and Multivariate Methods*. 2006. https://doi.org/10.2307/1269015.
[20] Siswanti TE, Yanti TS, Statistika P, Matematika F, Alam P. Pemodelan ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variable). Pros Stat 2020;6:113–8.
[21] Budiman H. Analisis Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu Lintas Jangka Pendek. Systemic J. 2016;2:19–24. https://doi.org/10.29080/systemic.v2i1.103.
[22] Chang PC, Wang YW, Liu CH. The development of a weighted evolving fuzzy neural network for PCB sales forecasting. Expert Syst Appl 2007;32:86–96. https://doi.org/10.1016/j.eswa.2005.11.021.

5. Acknowledgments

The author would like to thank Mr.Misfaruddin, Head of BPS Riau Province, has allowed the author to do practical work at BPS Riau Province. Thank you to the Practical Work Advisory Lecturer, Mrs. Anne Mudya Yolanda who has provided input and direction to the author in carrying out work practices. Thank you to Mr.Urip Widiyantoro, Coordinator Regional Account and Statistics Analysis gave many directions during the author's working experience. I don't forget to thank Mrs.Sri Mulyani and Mrs.Nelayesiana Bachtiar, Sub-Coordinator at Regional Account and Statistics Analysis, who supported and guided in the completion of this article, and thanks to all employees of BPS Riau Province.