Jurnal Kepariwisataan Indonesia 18 (2) (2024) 331–350
http://ejournal.kemenparekraf.go.id/index.php/jki
P-ISSN: 1907-9419, E-ISSN: 2685-9075, December 2024

BIG DATA ANALYTICS FOR FORECASTING TOURISM RECOVERY IN BALI ISLAND USING MULTIVARIATE TIME SERIES

Astrinariswari Rahmadian Prasetyo[1], Wahyu Calvin Frans Mariel[2], Geri Yesa Ermawan[3], Setia Pramana[4*]

[1,3,4]Statistical Computing Department, Politeknik Statistika STIS, Indonesia
[2]Directorate of Statistical Information System, BPS Statistics, Indonesia

Correspondence: setia.pramana@stis.ac.id

Article Info	Abstract
Keywords: big data, forecasting, multivariate time series, remote sensing, tourism demand. **Received:** March 10, 2024 **Approved:** October 8, 2024 **Published:** November 08, 2024	Bali is a famous tourist area and can significantly contribute to the Indonesian tourism sector. The COVID-19 pandemic has made Indonesian tourism, including Bali tourism, experience a decline. In March 2022, COVID-19 cases decreased, and the government began to relax some policies. The tourism sector is vital in economic recovery efforts after the COVID-19 pandemic. Therefore, it is necessary to identify tourism recovery to determine strategies and policies related to Indonesian tourism, especially in Bali. Multivariate time series forecasting of tourism demand can be used to identify tourism recovery using several significant data sources. The methods used are Vector Autoregressive (VAR), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The data used are the monthly official number of tourists, room occupancy rate, Google Trends, number of booking.com user reviews, and nighttime light intensity in Bali Province from January 2019 to December 2022. The results show that the best forecasting method is VAR, and modeling with multivariate time series forecasting can improve the performance of forecasting results. In addition, big data can be used as a source of supporting data that can provide better forecasting results, and the size of the dataset affects the selection of the best model. Furthermore, the descriptive and forecasting analysis results show that Bali tourism has experienced post-pandemic tourism recovery. The strategies and policies of the Bali government to restore Bali tourism faster are good enough.

How to cite:
Prasetyo, A. R., Mariel, W. C. F., Ermawan, G. Y., & Pramana, S. (2024). Big Data Analytics for Forecasting Tourism Recovery in Bali Island Using Multivariate Time Series. *Jurnal Kepariwisataan Indonesia: Jurnal Penelitian dan Pengembangan Kepariwisataan Indonesia, 18*(2), 331–350. https://doi.org/10.47608/jki.v18i22024.331–350.

INTRODUCTION

The tourism sector is one of the service-based sectors that plays a vital role in the Indonesian economy. It can be shown in 2021 when the tourism sector contributes to the national GDP by 4.2 percent (Yanwardhana, 2021). The tourism sector is also one of the sources of foreign exchange earnings. Based on data from BPS-Statistics Indonesia (2018) the tourism sector contributed USD 16.426 Billion (4.5% of GDP) to Indonesia's foreign exchange in 2018. Indonesian tourism has been recognized by tourists from various countries, especially Bali tourism. Tourism on the island of Bali is one of the tourism in Indonesia that can significantly contribute to Indonesia. Based on 2017-2021 data from BPS-Statistics of Bali Province (2021), most foreign tourists visiting Indonesia visit Bali. Bali Island is not only a destination for foreign tourists, but it is also a destination for domestic tourists. Indirectly, Bali Island already has its magnet to attract tourists to visit Bali Island through natural tourism, cultural tourism, and religious tourism.

The existence of the COVID-19 pandemic that entered Indonesia in early 2020 made Indonesia's tourism activities, including Bali Tourism, experience a decline (Pramana et al., 2022). The number of tourists, both foreign tourists and domestic tourists, has decreased significantly. Based on BPS data in 2020 and 2021, the growth of domestic tourist visits to Bali was negative. The COVID-19 pandemic has also caused the Gross Domestic Product (GDP) for the tourism sector and foreign exchange to decline.

Starting in March 2022, the number of COVID-19 cases has declined. The government began to relax some health protocol policies and people began to resume normal activities. The prolonged pandemic has encouraged tourists to travel. It can be an indication of the recovery of the tourism sector. The tourism sector has an essential role in recovering the economy after the COVID-19 pandemic, so tourism recovery needs to be identified. Tourism recovery can be identified by using tourism demand forecasting. Tourism demand is one measure of tourism based on goods and services used in a country as measured by the number of tourists (Friscintia & Alamsyah, 2019). Accurate forecasting can provide good results for the medium term and build tourism strategies, pricing policies, investment plans and strategies, and allocation of existing resources.

Nowadays, tourism demand forecasting cannot only rely on statistical data published by the government. The use of big data can be one of the data sources for forecasting tourism demand (Li et al., 2020). In addition, remote sensing data can be linked to tourism activities (Chang et al., 2022). The use of multisource data can strengthen the process of identifying tourism recovery (Bi et al., 2020; Liu et al., 2018; Zhang & Tian, 2022). Multivariate time series forecasting can be a solution for identifying tourism recovery in Indonesia. Research on tourism demand forecasting using multivariate time series and big data has been conducted in many other countries, but the Indonesian locus is still rare (Frechtling, 2001; Li et al., 2020; Song et al., 2019). These studies have also utilized deep learning technology, which still needs to be applied in research related to tourism demand forecasting in Indonesia. The utilization of deep learning technology in the field of tourism has been widely done to develop new and better models that can recommend the most suitable tourist activities/attractions (Cepeda-Pacheco & Domingo, 2022; Essien & Chukwukelu, 2022).

Research conducted by (Zhang & Tian, 2022) entitled "Forecast daily tourist volumes during the epidemic period using COVID-19 data, search engine data, and weather

Doi: 10.47608/jki.v18i22024.331–350

data". This study forecasts the volume of tourists during the COVID-19 pandemic using a hybrid model of Variational Mode Decomposition and Gated Recurrent Unit Networks (VMD-GRU) and then compares it with other methods. Other forecasting methods used in this study are Simple Random Walk, ARIMAX, Support Vector Regression (SVR), Artificial Neural Network (ANN), Long Short-Term Memory Networks (LSTM), Gated Recurrent Unit Networks (GRU), hybrid model Empirical Mode Decomposition and Gated Recurrent Unit Networks (EMD-GRU), hybrid model Ensemble Empirical Mode Decomposition and Gated Recurrent Unit Networks (EEMD-GRU), and hybrid model Variational Mode Decomposition and Long Short-Term Memory Networks (VMD-LSTM). The data used are the number of COVID-19 cases, search engine data, and weather data. The result of this study is that VMD is very suitable for predicting non-stationary traveler volumes and can significantly improve forecasting performance. In addition, the GRU network has the best performance in forecasting the volume of tourists and the use of multisource big data can improve the quality of forecasting results (Zhang & Tian, 2022).

Furthermore, a previous research study entitled "Daily Tourism Volume Forecasting for Tourist Attractions" was conducted by (Bi et al., 2020). This research analyses the daily tourism volume in the Jiuzhaigou and Huangshan mountain areas using the Long Short-Term Memory (LSTM) method. Other methods compared are the naïve method, ARIMAX, ANN, and SVR. In addition, this study was also conducted using several combinations of variables. The data used in this research are historical, search engines, and weather. The result of this research is that the LSTM method can be used for forecasting using historical, search engine, and weather data to produce solid data and forecast tourism volume. This research has shortcomings, among others, in forecasting using only three variables even though in real life, some other factors or predictors may have an influence on tourism volume (Bi et al., 2020).

(Liu et al., 2018) conducted research titled "Big Data Analytics for Forecasting Tourism Destination Arrivals with the Applied Vector Autoregression Model". This study analyses the correlation of weather, temperature, weekends, and holidays with tourist destination arrivals and web search requests and then performs forecasting analysis. The method used is Vector Autoregression (VAR) and the data use daily ticket data, weather, temperature, calendar information, weekends and public holidays), and web search queries. This study found that weather does not correlate with the actual arrival of the cultural tourism destinations studied or with web search queries (Liu et al., 2018).

Research on Indonesian tourism using big data has been conducted. (Pramana et al., 2022) conducted one of the studies titled "Impact of COVID-19 pandemic on tourism in Indonesia". This research investigates the different impacts of the COVID-19 pandemic on the tourism industry. The method used is clustering. The data used in this study are room occupancy rate data published by BPS, Google mobility index data, the number of flights from the flight status website, the number of reviews from TripAdvisor, the number of reviews from Booking.com, and data from Google Trends. This study provides results that show that the COVID-19 pandemic impacts the tourism industry and each province has a different impact. The utilization of big data helps the government and industry to provide knowledge about tourism plans and needs, real-time visitors, rates, and recommendations for activities and tourist attractions (Pramana et al., 2022).

Meanwhile, tourism demand forecasting in Indonesian tourism has also been carried out. Research conducted by (Friscintia & Alamsyah, 2019) titled "Forecasting

Doi: 10.47608/jki.v18i22024.331–350

Tourism Demand in Indonesian Tourism with the Artificial Neural Network Backpropagation Method" created a model to forecast tourism demand in Indonesia accurately. The data used in this study are historical data on Gross Domestic Product (GDP), Customer Price Index (CPI), and exchange rates of visitor countries, which are used as variables that affect the number of foreign tourist arrivals. This study shows that the government can use ANN forecasting to anticipate the availability of tourism infrastructure and services. This research has limitations in only using three variables to forecast tourism demand. In forecasting tourism demand using ANN, the proper configuration process is needed to choose architecture, parameters, and delays to ensure the model can work optimally (Friscintia & Alamsyah, 2019).

Recovering tourism is not far from the influence of a strategic leadership perspective, where leaders are required to prepare strategies to develop technical capabilities (Fenitra et al., 2022). The leader in this context is the government, which must also provide financial support and always strive to improve the international image (Yang et al., 2024). Taking a promising approach to the tourism industry's micro, small, and medium enterprises can also accelerate the recovery process (Fenitra et al., 2022). Another strategic approach that can be taken is a strategic approach with public-private partnerships and collaboration to develop local tourism (Alcoriza & Policarpio, 2023). In the tourism recovery process, focusing on product development, marketing, and destination coordination is critical (Faeni et al., 2023).

The government needs to identify Indonesia's tourism recovery immediately to determine strategies and policies related to Indonesian tourism. Tourism recovery is an urgent process in every country that aims to boost the economy and revive the lives of its people (Lin et al., 2023). Therefore, this study identifies tourism recovery through tourism demand forecasting using multivariate time series data consisting of big data and available tourism-related secondary data. The use of multivariate time series data can provide more accurate forecasting. The locus of this research is Bali Province because Bali is the icon of tourism in Indonesia.

METHODOLOGY

The time series data used in the study is from January 2019 to December 2022, with a monthly period. Based on related research and the availability of existing data, five variables will be used for forecasting. The following is an explanation of the data used.

Number of tourists visits

This study uses data on the number of monthly foreign tourists from Ngurah Rai Airport and Benoa Port and the number of domestic tourists who go to Bali. Foreign tourists were obtained from the Directorate General of Immigration, while domestic tourists were obtained from the survey results of the Bali Provincial Tourism Office. This data has been published by BPS Bali Province. In this study, the data on the number of tourists visits is denoted by "H".

Room Occupancy Rate (ROR)

Room Occupancy Rate (ROR) is the ratio between the number of room nights used and the number of room nights available. This study uses ROR for all classes of star hotels.

ROR data is obtained from the results of the Monthly Hotel Survey (VHTS), BPS Bali Province. In this study, ROR data is denoted by "T".

Google Trends

Google Trends is an index that shows the volume of keywords entered by users in a region. In this research, several keywords related to Bali tourism will be taken and then made into a composite index using the first principal component analysis by considering the weight of each keyword. The category used is travel and the keywords used are Bali, tourist Bali, Bali tourism, hotel Bali, travel in Bali, holiday in Bali, Bali tourist place, and Bali Indonesia.

The keywords are made into one time series or commonly called a composite index using Principal Component Analysis (PCA) weighting. After obtaining the first principal component, the PCA weight calculation will be carried out. First, calculate the weight value for each keyword in the first component.

$$B = \frac{LF}{RLF} \times RSSL \qquad ...(1)$$

Where:
LF = loading factor
RLF = average loading factor in 1 component
RSSL = proportion of variance.

The second step is to calculate the contribution of each keyword.

$$b = \frac{B}{JB} \qquad ...(2)$$

Where:
B = weight,
JB = sum of weights.

The third step is to calculate the composite index as follows:

$$Composite\ Index = \sum b_i K_i \qquad ...(3)$$

Where:
b_i = contribution weight value,
K_i = keyword search index value (Pramana et al., 2022).

In this research, the composite index data from Google Trends is denoted by "S".

Number of Booking.com user reviews

Booking.com is one of the world's leading digital travel companies, available in 43 languages and offers over 28 million accommodation listings, including more than 6.6 million houses, apartments, and other unique places to stay (Booking.com, n.d.). Data collection using web scraping is used to collect the number of reviews from Booking.com

Doi: 10.47608/jki.v18i22024.331–350

from each hotel in Bali. In this study, only count the number of reviews in the column of the date of stay at the hotel. The steps taken to collect data on the number of reviews from Booking.com are as follows.

1. Scraping the accommodation data list. The data obtained from scraping this accommodation data list consists of hotel name, URL, aid, and province.
2. Scraping reviews. After getting the list of accommodations, the next step is scraping to extract the review information, as presented in Figure 2. The data collected consists of province, hotel name, reviewer name, reviewer origin, room type, review date, hotel stay date, review score, review reaction, good comments, and bad comments.
3. Data cleaning. The raw data obtained from the previous step is cleaned from duplicate data before being analyzed.

In this study, the data on the number of Booking.com user reviews is denoted by "B".

Intensity Night Time Light (NTL)

Night-Time Light Intensity data is obtained through cloud-free composite data from VIIRS Day/Night Band (DNB), which is then aggregated with regional boundaries using the zonal statistics feature. Areas that have high NTL intensity values can be identified as urbanized areas and maintain more tourism infrastructure and services (Chang et al., 2022; Devkota et al., 2019). NTL is currently also used for different purposes, such as poverty mapping, as it can be used as an approach to economic activity in an area (S. R. Putri et al., 2023; Ramadhan et al., 2023; Utami et al., 2023).

In this study, NTL intensity data is denoted by "N". The processed satellite images were then aggregated with the boundaries of the area. In this research, aggregation was done using the zonal statistics function available in QGIS.

Multivariate Time Series Forecasting

This research uses multivariate time series forecasting, one type of forecasting with more than one variable that changes over time, where the results can be more accurate than using univariate time series. Data analysis in this study was carried out with the help of RStudio software and Python. The research flow used is based on the research of Li et al. (2020) and Zhang and Tian (2022). The flow of this research is illustrated in Figure 1.

Doi: 10.47608/jki.v18i22024.331–350

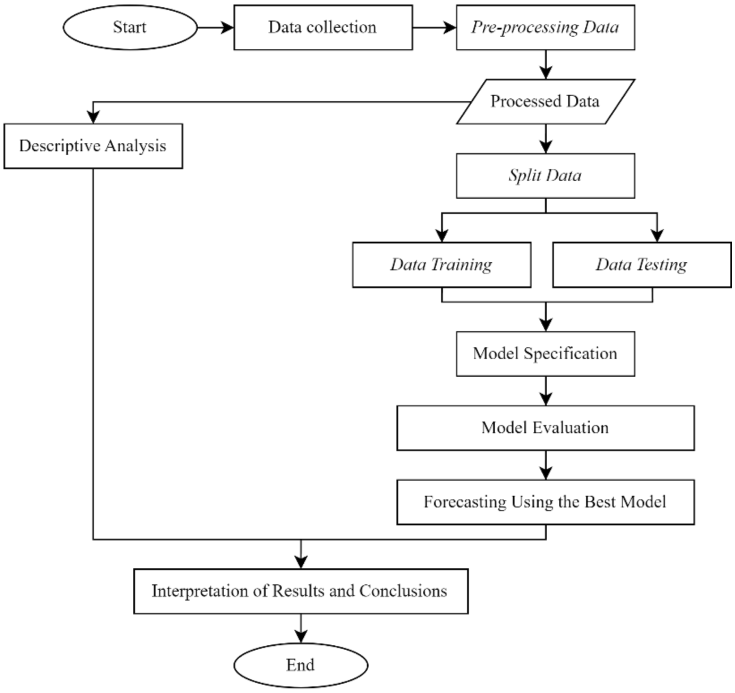

Figure 1. Research flow chart
Source: Li et al., 2020; Zhang & Tian, 2022

Step 1: Data Preprocessing

This research uses multivariate time series data to convert each variable into time series data with a monthly period. In the Booking.com user review data, data validation has previously been carried out, including data checking and data updating. Then, the number of reviews data will be aggregated to calculate the number of reviews per month. The Google Trends composite index data needs to be aggregated from per week to per month. After all variables are in the form of time series data, all variables are made into one file and then ready for further analysis.

Step 2: Descriptive Analysis

In this study, descriptive analysis was carried out to see the condition of Bali tourism on each variable in three periods, namely the period before the COVID-19 pandemic, the period during the COVID-19 pandemic, and the transition period to the endemic (Table 1).

Table 1. Date categories according to the chronology of events

Time Period	Event Name	Event Date
I	The period before the COVID-19 Pandemic	January 2019 – February 2020
II	During the COVID-19 pandemic	March 2020 – March 2022
III	Transitional period toward endemic	April 2022 – December 2022

Source: Biro Komunikasi dan Informasi Publik, 2022; Nuraini, 2020

Step 3: Forecasting Model Analysis

Previous research is the basis for choosing the method to be used. The method used in modeling to predict tourism demand using multivariate time series data is as follows.

Doi: 10.47608/jki.v18i22024.331–350

Vector Autoregressive (VAR)

The VAR model is a method for forecasting time series data of more than one variable that assumes all variables in the model are endogenous variables (Widarjono, 2013). The VAR model can be used to explain simultaneous variables that have an influence on each other. In other words, this VAR is two-way influencing each other.

Long Short Term Memory (LSTM)

LSTM is a Recurrent Neural Network (RNN) capable of learning short-term and long-term dependencies (Zhao et al., 2017). LSTM has three gates: forget gates, which decide what information to remove from the cell; input gates, which decide what new information to store in the cell; and output gates, which decide what to produce.

Gated Recurrent Units (GRU)

GRU can be a variation of LSTM, but its internal structure is simpler because GRU has fewer calculations required to update the hidden states, making training time faster (Gulli & Pal, 2017). GRU has two gates, which are update gates and reset gates.

Support Vector Regression (SVR)

Support Vector Regression (SVR) is the application of a Support Vector Machine (SVM) to solve regression cases (Jändel, 2010). The output of SVR is both real and continuous. SVR is able to overcome overfitting. Another advantage compared to other regression methods is that SVR can handle nonlinear regression problems efficiently.

Step 4: Identify Tourism Recovery

After model evaluation, the best model will be used for further forecasting analysis to identify whether there is a recovery in tourism. The data used is the overall data from January 2019 to December 2022.

FINDINGS AND DISCUSSION

Number of tourists visits

Figure 2 shows that before the COVID-19 pandemic, the number of tourists coming to Bali, both domestic and foreign, increased during June-August and December-January. This is because these months are high seasons in Bali. During the COVID-19 pandemic, the number of tourists coming to Bali, both domestic and foreign tourists, experienced a very significant decline. Even for foreign tourists from June 2021 to January 2022, only a few foreign tourists were visiting Bali. This is because the number of COVID-19 cases increased that month, so the international entrance to Bali was closed (Paludi, 2022). The number of tourists decreased by 74.07 percent compared to the period before the pandemic. During the transition period to the COVID-19 endemic, the number of tourists coming to Bali gradually increased. This can be an indication of tourism recovery in Bali.

Doi: 10.47608/jki.v18i22024.331–350

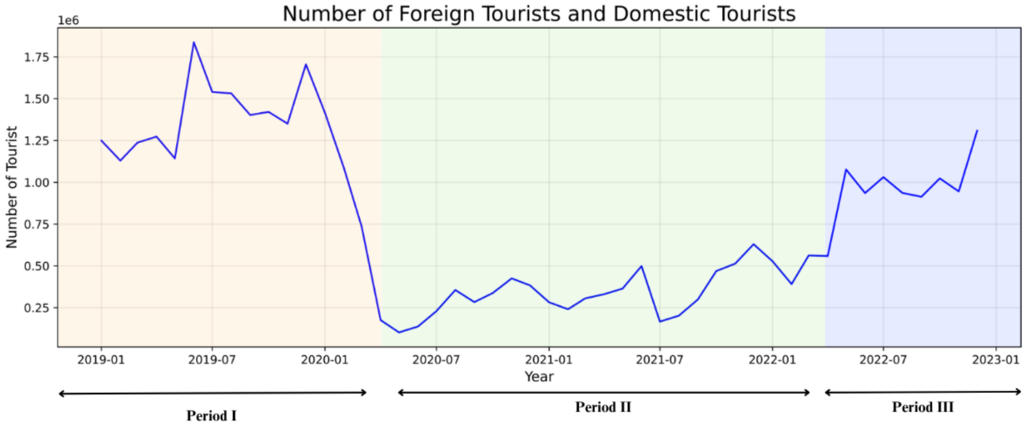

Figure 2. The number of foreign tourists and domestic tourists (in millions) visiting Bali, 2019-2022
Source: Author’s analysis, 2023

Room Occupancy Rate (ROR)

The Bali Hotel Room Occupancy Rate (ROR) for all hotel classes began to experience a significant decline in March 2020, as shown in Figure 3. Compared to March 2019, which reached 55.43 percent, the room occupancy rate in March 2020 decreased by 45.84 percent. Even during the COVID-19 pandemic, the ROR reached its lowest point in May and June 2020 when the hotel ROR value in Bali for all classes was only 2.07 percent. The ROR value decreased by 79.81 percent compared to the pre-pandemic period and can be categorized as a low ROR value (small). A small ROR value can mean that accommodation in the area is less attractive to visitors and vice versa. A high ROR value can mean that visitors consider accommodation in the area desirable. However, during the COVID-19 pandemic, a small ROR value does not mean that the area is in less demand by visitors; somewhat, the number of tourists visiting Bali has decreased. Bali is one of the provinces that experienced the most significant decline in Room Occupancy Rate (ROR) values (Pramana et al., 2022). During the transition period to the COVID-19 endemic, starting in March 2022, the ROR value increased again.

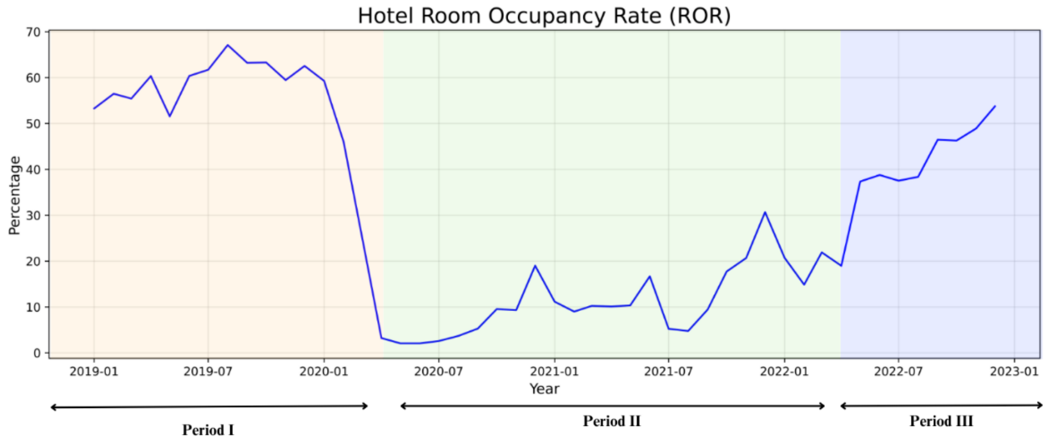

Figure 3. Room Occupancy Rate (ROR) of hotels in Bali, 2019-2022
Source: Author’s analysis, 2023

Doi: 10.47608/jki.v18i22024.331–350

Google Trends

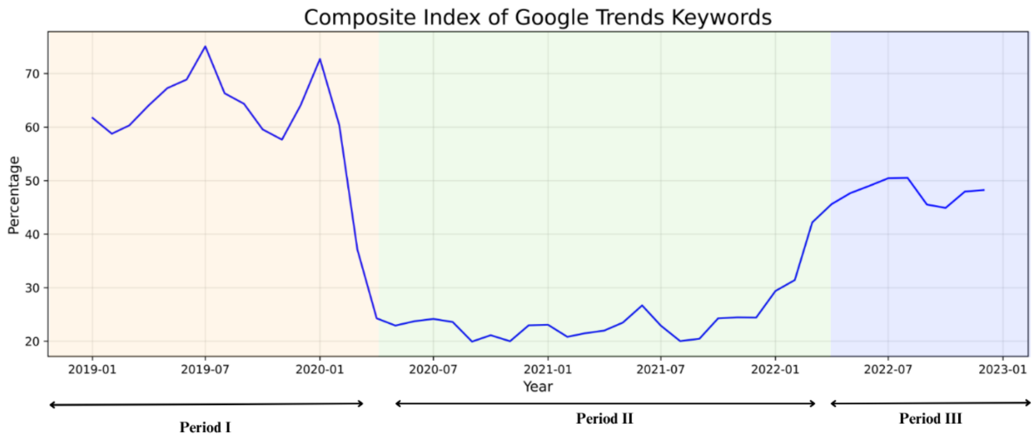

Figure 4. Composite index of multiple keywords on Google Trends, 2019-2022
Source: Author's analysis, 2023

In the pre-pandemic period, the value of the Google Trends composite index (presented in Figure 4) increased in the vacation months of June-August and the end of the year in December-January. Then, March 2020 experienced a significant decline, and during the transition period towards the endemic, it showed a gradual increase. When viewed from the monthly average during the COVID-19 pandemic, there was a decrease of 61.67 percent compared to the period before the COVID-19 pandemic. The decline in Google searches for keywords related to Bali tourism is due to restrictions on community activities and mobility, which allow people to do all their activities from home. It has an impact on reducing the intensity of people's search for things related to activities outside the home, such as vacations. During the transition period to the COVID-19 endemic, this composite index experienced a gradual increase. The prolonged pandemic has encouraged tourists to travel to Bali, so Google search for keywords related to Bali tourism has increased again.

Number of reviews from Booking.com

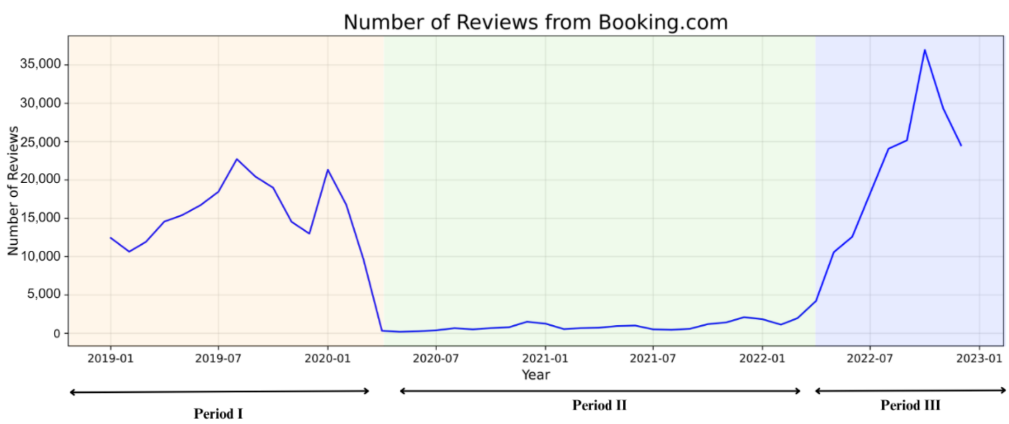

Figure 5. Number of user reviews of Booking.com hotels in Bali, 2019-2022
Source: Author's analysis, 2023

When compared to other provinces in Indonesia, Bali Province has a larger number of hotel accommodations. This number of Booking.com user reviews, as shown in Figure

Doi: 10.47608/jki.v18i22024.331–350

5, includes both international and domestic reviews. Most reviewers come from Australia, domestic tourists, France, and Germany. This is in line with the results of BPS publications regarding the number of foreign tourists visiting Bali. Australian tourists dominate foreign tourists who come to Bali (BPS, 2022). In the pre-pandemic period, there was an increase in the number of Booking.com user reviews in July-August and at the end of the year (December-January).

In March 2020, there was a very significant decline. Compared to the period before the COVID-19 pandemic, the average number of reviews per month during the COVID-19 pandemic decreased by 92.3 percent. The lowest value for the number of Booking.com user reviews occurred in May 2020, with only 192 reviews. During the COVID-19 pandemic, the number of reviews per month of visitors from Indonesia has mostly stayed the same when compared to the number of reviews per month of visitors from abroad. In February 2022, the number of Booking.com user reviews started to increase gradually. This gradual increase in the number of reviews can be interpreted as a recovery in Bali tourism.

Night Time Lights Intensity

Of the five variables shown previously, only the Night Time-Light Intensity variable has a different pattern (see Figure 6). The graph shows that the COVID-19 pandemic does not influence the NTL intensity value. NTL intensity values tend to provide a similar pattern from year to year or what is commonly referred to as a seasonal pattern. At the beginning of the COVID-19 pandemic, almost all districts or cities in Bali Province experienced a decrease in community activity. However, after the implementation of PPKM in early 2021, some areas in Bali Province experienced an increase in activity (Putro, 2022).

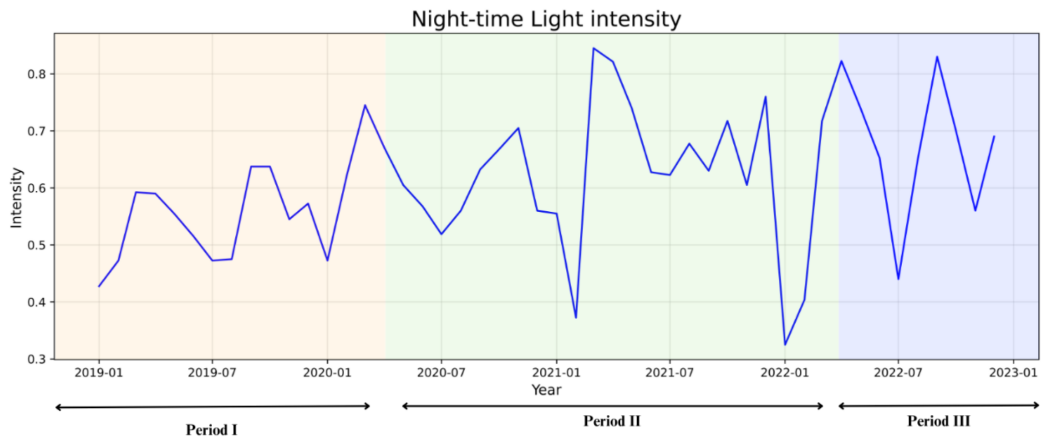

Figure 6. Night Time-Light (NTL) Intensity Bali Province, 2019-2022
Source: Author's analysis, 2023

Best Forecasting Model Analysis

Modeling using VAR starts with a stationarity test of the data and the determination of the optimal lag. Then, modeling will be done based on the best lag. Forecasting using SVR is done with Radial Basic Function (RBF) kernel; gamma 0.5; C 10; and epsilon 0.05. Epoch 250 and batch size 32 were used for forecasting with deep learning model specifications. Researchers determine the best model by comparing the model evaluation of each method and the combination of variables used. The result is presented in Figure 7.

Doi: 10.47608/jki.v18i22024.331–350

By doing this comparative analysis, we can find out how multisource data improves the performance of forecasting results.

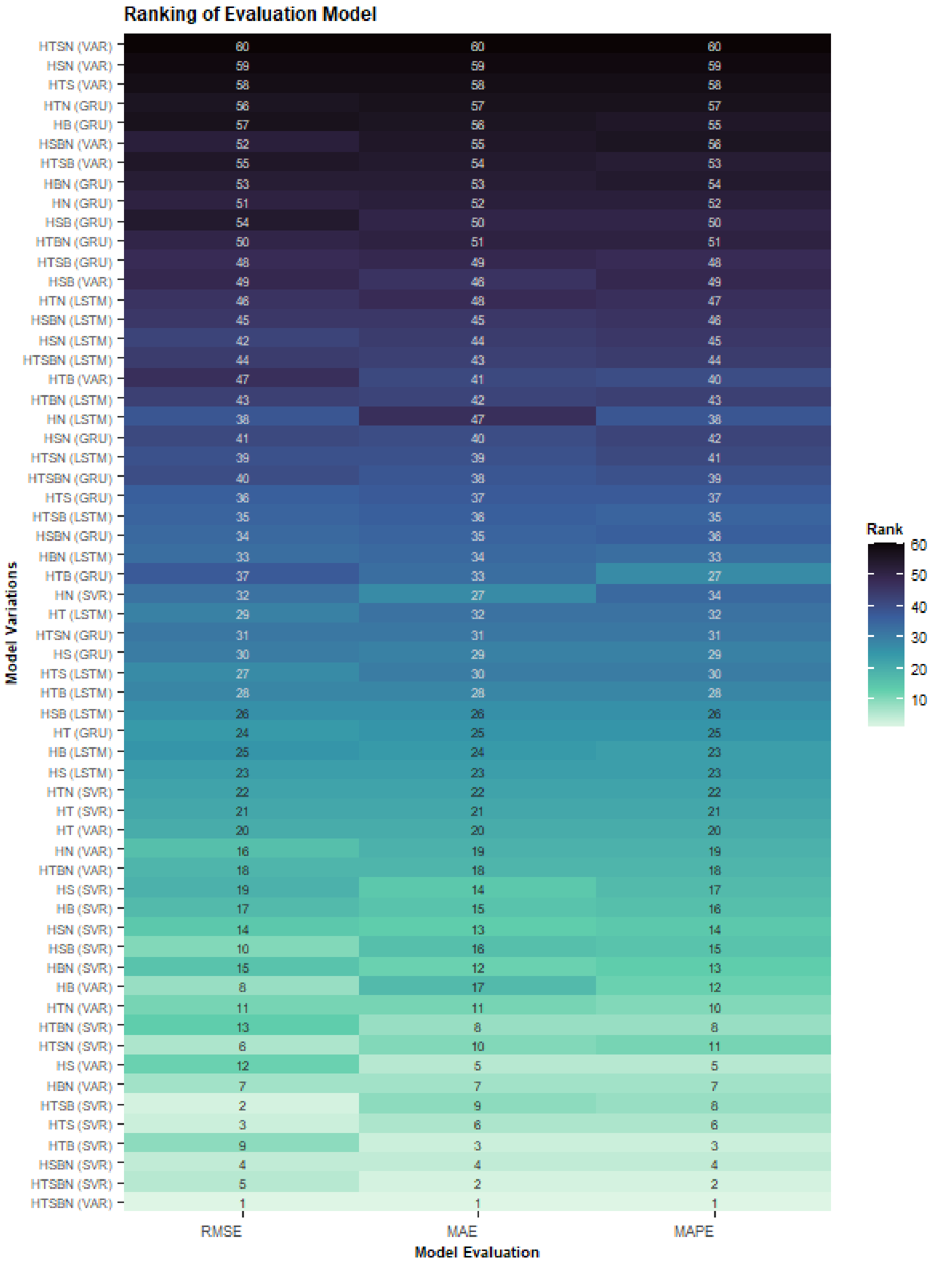

Figure 7. Ranking of Evaluation Model
Source: Author's analysis, 2023

The results of the three model evaluations above show that the best forecasting model is a model with a combination of HTSBN variables using the Vector Autoregressive (VAR) method. The HTSBN model with the VAR method has a model evaluation value of RMSE 283,699.9; MAE 220,987.4; and MAPE 0.219 (this forecasting model can be categorized as good enough or feasible to use). Visualization of the ranking results of the model evaluation can be seen in Figure 7.

Doi: 10.47608/jki.v18i22024.331–350

Forecasting tourism demand using multivariate time series where the data used consists of a combination of BPS data with data sourced from big data has been able to improve the performance of forecasting results. Previous research shows that integrating various data sources has the advantage of improving the results of tourism demand forecasting (Li et al., 2020). The selection of appropriate variables for multivariate time series forecasting is also able to improve the performance of forecasting results. In addition, the size of the dataset used in this study also plays a vital role in determining the best model. The size of the dataset used in this study can be categorized as small because there are only 48 data points in this study. The use of machine learning can provide better performance if the dataset used is large (Cerqueira et al., 2019). For small datasets, classical methods show better performance. However, as the dataset size increases, the machine learning method can outperform the classical method. This is in line with the results of this study, which show that the best model is the one using the VAR method.

The model uses a combination of big data variables and BPS data. Big data can already be used as an alternative data source for forecasting tourism demand. Big data has been used as a supporting data source for official statistics produced by BPS Statistics Indonesia, which is integrated with survey or administrative data available at BPS (Faris & Pramana, 2021).

Identify Tourism Recovery

Based on the model evaluation, the best model obtained is a model with a combination of HTSBN variables using the VAR method. Forecasting is carried out using all data to identify tourism recovery (January 2019-December 2022) using the best method and variable combination.

Before forecasting modeling, a stationary test is conducted using the Augmented Dickey-Fuller (ADF) test. Based on the ADF test results, at the level α=0.05, only the NTL intensity variable is stationary, while the other variables are not stationary. Then, the data is standardized and differenced. After standardizing and differencing, all variables are stationary. Because the data is stationary, the next step can be continued, namely testing the lag length.

The optimum lag is the lag that has the smallest AIC, BIC, FPE, and HQIC values. Based on optimum lag test results, the criteria that indicate the optimum lag is at lag 6 because it has the smallest AIC, FPE, and HQIC values. By using lag 6, VAR modeling will be done for forecasting.

Then, to check whether the model is sufficiently able to explain the variance and pattern in the time series, the residual serial correlation test is used using the Durbin Watson test. Based on the Durbin Watson test results, the serial correlation looks quite good because the resulting value of each variable is close to 2, so it can be concluded that there is no significant serial correlation. The VAR model equation obtained from this study for the number of tourists is as follows.

$$H_t = 0{,}67T_{t-1} - 0{,}66B_{t-1} - 0{,}289N_{t-4} \qquad \text{...(4)}$$

The number of tourists with ROR has a positive one-way relationship. ROR is influential at lag 1 because it has a probability value that is smaller than α (0.05). In addition, the number of tourists has a negative one-way relationship with the number of

Doi: 10.47608/jki.v18i22024.331–350

Booking.com user reviews and NTL intensity. The number of Booking.com user reviews was affected at lag 1 and NTL intensity was affected at lag 4.

After analyzing the VAR model, VAR model forecasting will be carried out. Forecasting is done for the next 12 months. The results of this forecasting are used to identify the progress of Bali's tourism recovery.

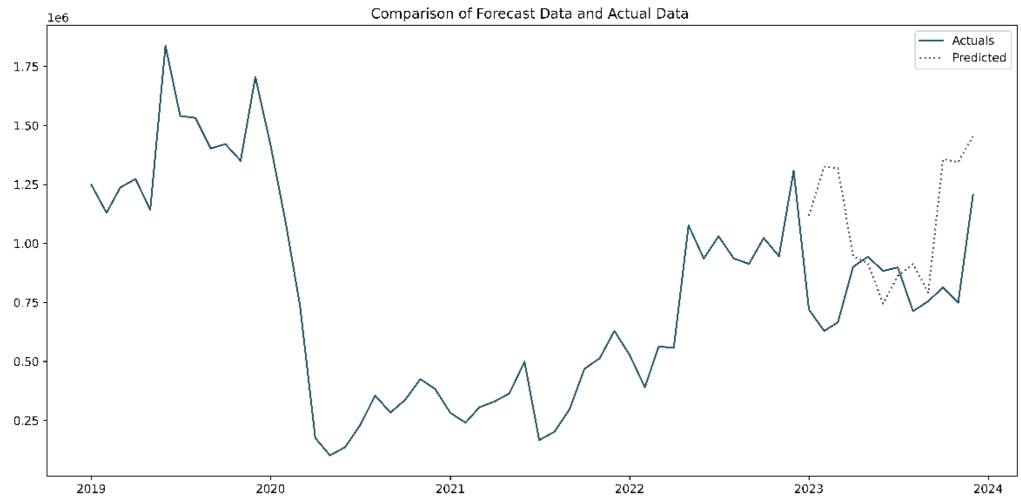

Figure 8. Tourism demand forecasting results and the actual data
Source: Author's analysis, 2023

Based on Figure 8, Bali tourism's condition has generally experienced recovery. However, Bali tourism cannot be said to have recovered one hundred percent like the conditions before the COVID-19 pandemic. Based on previous research, Bali Province is a province whose tourism has yet to be able to bounce back quickly when compared to other provinces. In other words, Bali's tourism recovery still needs to catch up to other provinces (Pramana et al., 2022). Therefore, the government has made many policies and strategies to overcome this. Rather than refining the existing model, systematic change is needed to cope with another extraordinary event and build a stronger tourism system (Gössling & Higham, 2021; Rosenbloom et al., 2020). Although not fully recovered, the hard work of the government and the Balinese people to restore Bali's tourism conditions after the COVID-19 pandemic has produced quite good results. The most essential component of tourism recovery is the tourists, who must be willing to go on holiday and recreate during the COVID-19 pandemic (Gössling & Schweiggart, 2022).

The forecasting results graph shows that in January 2023, the number of tourists decreased. Then, in February and March 2023, the number of tourists visiting Bali began to increase again. The Bali Tourism Office has compiled tourist, cultural, MICE (Meeting, Incentive, Convention, and Exhibition), and sports events spread across the districts or cities of Bali Province. This is one form of strategy carried out by the government to restore Bali tourism. One of the MICE events held in February 2023 was the ASEAN+3 meeting (ASEAN countries together with Japan, Korea, and China) held at the Bali Nusa Dua Convention Center (BNDCC). The increase in the number of tourists coming to Bali in March 2023 is due to the long Nyepi holiday (K. Putri, 2023). Nyepi Day falls on Wednesday, so employees can apply for leave for Monday and Tuesday so that employees get five days off starting from Saturday.

Doi: 10.47608/jki.v18i22024.331–350

For the forecasting results for April to June 2023, the number of tourists has decreased again. Many problems exist regarding foreign tourists visiting Bali when associated with the current field conditions. Many foreign tourists have violated both the law and cultural ethics in Bali, so from January to May 2023, there have been 123 foreign tourists deported from Bali (Winata, 2023). The easy access to Bali and the lack of strict regulations make the list of violations committed by foreign tourists more and more. Therefore, the Balinese government needs to evaluate and tighten tourism regulations in Bali so that tourists who come on vacation to Bali can feel comfortable and safe. Tightened tourism regulations and entry access can lead to a decrease in the number of tourists coming to Bali.

In the tourism demand forecasting graph, the number of tourists rose again in July 2023. This is because July is the vacation season for school children in Indonesia. In addition, Australia, as a country that dominates foreign tourist visits to Bali, is also on winter vacation. This increase in the number of tourists lasts until August 2023. However, in September 2023, there was a slight decrease in tourists. In October 2023, the forecasting results of the number of tourists visiting Bali increased again until the peak at the end of 2023. Based on the forecasting results, in December 2023, the number of tourists visiting Bali will be more than that in December 2022. The increase in the number of tourists visiting Bali in December is because there are long holidays in this month (Christmas, New Year, and school holidays). In addition, countries such as the United States, South Korea, Japan, countries in Europe, and surrounding countries are on winter vacation and Australia is on summer vacation.

However, after being compared with the actual data on the number of foreign and domestic tourists in 2023, for the months of January to April 2023, the amount remained the same. Based on forecast data, for January-April 2023, there was an increase. However, in reality, there was a decrease. Based on data from BPS, domestic tourists still dominate. The number of foreign tourists from January to April is half that of domestic tourists. So, tourism recovery is slower from January to April. This is because there are still COVID-19 regulations from outside regarding going abroad, especially for holidays. In May 2023, WHO stated that the pandemic had ended (WHO, 2023). The emergence of this statement was followed by an increase in the number of tourists in Bali. Tourism recovery in Bali is starting to improve.

The Bali government's strategy and policy in restoring Bali tourism faster is appropriate for now. The Bali government still has to improve Bali's image in the international arena. There is a need for a review of tourism regulations in Bali by the Balinese government and regulations on access to Bali to realize quality, safe, and comfortable Bali tourism. Anticipatory planning and management are essential and can contribute to tourism recovery (Gössling & Schweiggart, 2022). Not only the government but also market players, technology innovators, and the workforce working in the tourism industry are participating in the tourism recovery program (Sharma et al., 2021).

CONCLUSION

Before the COVID-19 pandemic, the variables of the number of tourists, Room Occupancy Rate (ROR), a composite index of several Google Trends keywords, and the number of Booking.com user reviews increased in June-August and December-January.

Doi: 10.47608/jki.v18i22024.331–350

During the COVID-19 pandemic, these four variables experienced a significant decline. Meanwhile, during the transition to the COVID-19 endemic, it has shown a gradual increase. This can mean that Bali tourism has been recovering little by little. The Night Time Light (NTL) intensity variable has a value that is not much different from the three periods (before the pandemic, during the pandemic, and during the transition to the COVID-19 endemic).

The best model for forecasting tourism demand is a model with a combination of variables of the number of tourists, ROR, Google Trends, number of reviews, and NTL intensity (HTSBN) using the Vector Autoregressive (VAR) method. This best model is influenced by the size of the dataset used. Since the dataset used is small, the classical method shows better performance. Modeling using variables that are only sourced from big data has been able to create a forecasting model that is as good as a forecasting model that uses variables only from BPS data. However, the model evaluation value will be even better if the forecasting model uses variables from big data and BPS data. Therefore, big data can already be used as a supporting data source for forecasting tourism demand.

The forecasting results of tourism demand show an increase, so it can be concluded that Bali has experienced a tourism recovery. However, Bali's tourism recovery cannot be said to have recovered to its pre-pandemic condition. For now, the strategies and policies of the Bali government to restore Bali tourism faster are good enough. In the future, the government can organize international events that invite international tourists to Bali. In addition, the government needs to support all businesses in the tourism sector, such as UMKM. The government needs to make improvements, maintenance, and renewal regarding infrastructure and tourism facilities.

This research has shown that the best forecasting method, namely the model using the VAR method and modeling with multivariate time series forecasting, has improved the performance of forecasting results. In addition, this research also shows that big data can be used as a source of supporting data that can provide better forecasting results.

This study has limitations, namely the limited data available so that only five variables are used (number of tourists, Room Occupancy Rate (ROR), Google Trends, number of Booking.com user reviews, and NTL), so the results are still not optimal. Therefore, future research can add other variables related to tourism demand, such as the number of holidays, weather, GDP, consumption price index, rupiah exchange rate, and the number of reviews from other travel sites such as TripAdvisor. Adding data periods used for forecasting according to data availability can also be a recommendation for future research. In terms of methods, future research can use XGBoost, Decision Tree Regression, or variations of methods using a combination of deep learning and VAR to improve forecasting accuracy. Increased forecasting accuracy can provide accurate forecasting results so that the government can precisely determine strategies and policies. The Bali government's strategy and policy in restoring Bali tourism faster is appropriate for now. The Bali government still has to improve Bali's image in the international arena. There is a need for a review of tourism regulations in Bali by the Balinese government and regulations on access to Bali to realize quality, safe, and comfortable Bali tourism.

Doi: 10.47608/jki.v18i22024.331–350

ACKNOWLEDGEMENT

The authors gratefully acknowledge research funding from Politeknik Statistika STIS.

REFERENCES

Alcoriza, G. B., & Policarpio, J. E. A. (2023). Post-Disaster Tourism: Building Resilient and Sustainable Communities. In J. Kumar, G. E. Bayram, & A. Sharma (Eds.), *Resilient and Sustainable Destinations After Disaster: Challenges and Strategies* (pp. 97–107). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80382-021-720231008

Badan Pusat Statistik. (2018). *Jumlah Devisa Sektor Pariwisata*. https://www.bps.go.id/id/statistics-table/2/MTE2MCMy/jumlah-devisa-sektor-pariwisata.html

Badan Pusat Statistik. (2022). *Statistik Kunjungan Wisatawan Mancanegara 2021*. https://www.bps.go.id/id/publication/2022/04/28/d79faad2c263388e94e160ee/statistik-kunjungan-wisatawan-mancanegara-2021.html

Badan Pusat Statistik Provinsi Bali. (2021). *Banyaknya Wisatawan Mancanegara Bulanan ke Bali Menurut Pintu Masuk*. https://bali.bps.go.id/id/statistics-table/2/MTA2IzI=/banyaknya-wisatawan-mancanegara-bulanan-ke-bali-menurut-pintu-masuk--orang-.html

Bi, J.-W., Liu, Y., & Li, H. (2020). Daily Tourism Volume Forecasting for Tourist Attractions. *Annals of Tourism Research*, *83*, 102923. https://doi.org/10.1016/J.ANNALS.2020.102923

Biro Komunikasi dan Informasi Publik. (2022, May 20). *Prokes Dilonggarkan, Kemenhub Terbitkan SE Perjalanan Domestik dan Internasional Terbaru*. Kementerian Perhubungan. https://dephub.go.id/post/read/prokes-dilonggarkan,-kemenhub-terbitkan-se-perjalanan-domestik-dan-internasional-terbaru

Booking.com. (n.d.). *Tentang Booking.com™*. Booking.Com. Retrieved May 25, 2023, from https://www.booking.com/content/about.id.html

Cepeda-Pacheco, J. C., & Domingo, M. C. (2022). Deep Learning and Internet of Things for Tourist Attraction Recommendations in Smart Cities. *Neural Computing and Applications*, *34*(10), 7691–7709. https://doi.org/10.1007/S00521-021-06872-0

Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters. *ArXiv*. https://arxiv.org/abs/1909.13316

Chang, P., Pang, X., He, X., Zhu, Y., & Zhou, C. (2022). Exploring the Spatial Relationship between Nighttime Light and Tourism Economy: Evidence from 31 Provinces in China. *Sustainability*, *14*(12), 7350. https://doi.org/10.3390/SU14127350

Devkota, B., Miyazaki, H., Witayangkurn, A., & Kim, S. M. (2019). Using Volunteered Geographic Information and Nighttime Light Remote Sensing Data to Identify Tourism Areas of Interest. *Sustainability*, *11*(17), 4718. https://doi.org/10.3390/SU11174718

Essien, A., & Chukwukelu, G. (2022). Deep Learning in Hospitality and Tourism: A Research Framework Agenda for Future Research. *International Journal of Contemporary Hospitality Management*, *34*(12), 4480–4515. https://doi.org/10.1108/IJCHM-09-2021-1176

Faeni, D. P., Faeni, R. P., Riyadh, H. A., & Yuliansyah, Y. (2023). The COVID-19 Pandemic Impact on the Global Tourism Industry SMEs: A Human Capital Development Perspective. *Review of International Business and Strategy*, *33*(2), 317–327. https://doi.org/10.1108/RIBS-08-2021-0116

Faris, M., & Pramana, S. (2021). Preparing a Big Data Utilization Communication Strategy. *Journal of Physics: Conference Series*, *1863*, 012077. https://doi.org/10.1088/1742-6596/1863/1/012077

Fenitra, R. M., Abbas, A., Ekowati, D., & Suhairidi, F. (2022). Strategic Intent and Strategic Leadership: A Review Perspective for Post-COVID-19 Tourism and Hospitality Industry

Recovery. In P. Mohanty, A. Sharma, J. Kennel, & A. Hassan (Eds.), *The Emerald Handbook of Destination Recovery in Tourism and Hospitality* (pp. 23–44). Emerald Publishing Limited. https://doi.org/10.1108/978-1-80262-073-320221003

Frechtling, D. C. (2001). *Forecasting Tourism Demand: Methods and Strategies* (1st ed.). Routledge. https://doi.org/10.4324/9780080494968

Friscintia, P. B. A., & Alamsyah, A. (2019). Peramalan Tourism Demand Pada Pariwisata Indonesia Dengan Metode Artificial Neural Network Backpropagation. *E-Proceeding of Management*, *6*(1), 503–509. https://openlibrary.telkomuniversity.ac.id/pustaka/files/148875/jurnal_eproc/peramalan-tourism-demand-pada-pariwisata-indonesia-dengan-metode-artificial-neural-network-backpropagation.pdf

Gössling, S., & Higham, J. (2021). The Low-Carbon Imperative: Destination Management under Urgent Climate Change. *Journal of Travel Research*, *60*(6), 1167–1179. https://doi.org/10.1177/0047287520933679

Gössling, S., & Schweiggart, N. (2022). Two Years of COVID-19 and Tourism: What We Learned, and What We Should Have Learned. *Journal of Sustainable Tourism*, *30*(4), 915–931. https://doi.org/10.1080/09669582.2022.2029872

Gulli, A., & Pal, S. (2017). *Deep Learning with Keras: Implement Neural Networks with Keras on Theano and TensorFlow*. Packt Publishing.

Jändel, M. (2010). A Neural Support Vector Machine. *Neural Networks*, *23*(5), 607–613. https://doi.org/10.1016/J.NEUNET.2010.01.002

Li, H., Hu, M., & Li, G. (2020). Forecasting Tourism Demand with Multisource Big Data. *Annals of Tourism Research*, *83*, 102912. https://doi.org/10.1016/J.ANNALS.2020.102912

Lin, W.-C., Wu, C. K., Le, T. K. T., & Nguyen, N. A. (2023). Assessment of Vietnam Tourism Recovery Strategies after COVID-19 Using Multi-Criteria Decision-Making Approach. *Sustainability*, *15*(13), 10047. https://doi.org/10.3390/SU151310047

Liu, Y.-Y., Tseng, F.-M., & Tseng, Y.-H. (2018). Big Data Analytics for Forecasting Tourism Destination Arrivals with the Applied Vector Autoregression Model. *Technological Forecasting and Social Change*, *130*, 123–134. https://doi.org/10.1016/J.TECHFORE.2018.01.018

Nuraini, R. (2020, March 2). *Indonesia.go.id - Kasus Covid-19 Pertama, Masyarakat Jangan Panik*. Indonesia.Go.Id. https://indonesia.go.id/narasi/indonesia-dalam-angka/ekonomi/kasus-covid-19-pertama-masyarakat-jangan-panik

Paludi, S. (2022). Setahun Pandemi COVID-19 dan Dampaknya Terhadap Industri Pariwisata Indonesia. *Equilibrium: Jurnal Penelitian Pendidikan Dan Ekonomi*, *19*(01), 49–60. https://doi.org/10.25134/EQUI.V19I01.4337

Pramana, S., Paramartha, D. Y., Ermawan, G. Y., Deli, N. F., & Srimulyani, W. (2022). Impact of COVID-19 Pandemic on Tourism in Indonesia. *Current Issues in Tourism*, *25*(15), 2422–2442. https://doi.org/10.1080/13683500.2021.1968803

Putri, K. (2023, March 20). Dispar Bali Prediksi Okupansi Hotel Naik 50 Persen Saat Nyepi. *DetikBali*. https://www.detik.com/bali/berita/d-6628370/dispar-bali-prediksi-okupansi-hotel-naik-50-persen-saat-nyepi

Putri, S. R., Wijayanto, A. W., & Pramana, S. (2023). Multi-Source Satellite Imagery and Point of Interest Data for Poverty Mapping in East Java, Indonesia: Machine Learning and Deep Learning Approaches. *Remote Sensing Applications: Society and Environment*, *29*, 100889. https://doi.org/10.1016/J.RSASE.2022.100889

Putro, D. H. (2022). *Penggunaan Pengineraan Jauh Untuk Mengamati Perubahan Aktivitas Manusia di Provinsi Bali Selama Pandemi*. Politeknik Statistika STIS.

Ramadhan, R., Wijayanto, A. W., & Pramana, S. (2023). Geospatial Big Data Approaches to Estimate Granular Level Poverty Distribution in East Java, Indonesia using Machine Learning

Doi: 10.47608/jki.v18i22024.331–350

and Deep Learning Regressions. In W. Marsisno, R. Kurniawan, I. M. Arcana, S. Muchlisoh, J. R. H. Sitorus, R. Yuliana, & S. M. Berliana (Eds.), *Proceedings of The International Conference on Data Science and Official Statistics (ICDSOS)* (Vol. 2023, Issue 1, pp. 186–200). Politeknik Statistika STIS. https://doi.org/10.34123/ICDSOS.V2023I1.359

Rosenbloom, D., Markard, J., Geels, F. W., & Fuenfschilling, L. (2020). Why Carbon Pricing is Not Sufficient to Mitigate Climate Change—and How “Sustainability Transition Policy” Can Help. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(16), 8664–8668. https://doi.org/10.1073/PNAS.2004093117

Sharma, G. D., Thomas, A., & Paul, J. (2021). Reviving Tourism Industry Post-COVID-19: A Resilience-Based Framework. *Tourism Management Perspectives*, *37*, 100786. https://doi.org/10.1016/J.TMP.2020.100786

Song, H., Qiu, R. T. R., & Park, J. (2019). A Review of Research on Tourism Demand Forecasting: Launching the Annals of Tourism Research Curated Collection on Tourism Demand Forecasting. *Annals of Tourism Research*, *75*, 338–362. https://doi.org/10.1016/J.ANNALS.2018.12.001

Utami, N. A., Wijayanto, A. W., Pramana, S., & Astuti, E. T. (2023). Spatially Granular Poverty Index (SGPI) for Urban Poverty Mapping in Jakarta Metropolitan Area (JMA): A Remote Sensing Satellite Imageries and Geospatial Big Data Approach. *Earth Science Informatics*, *16*(4), 3531–3544. https://doi.org/10.1007/S12145-023-01084-7

WHO. (2023, May 5). *Statement on the Fifteenth Meeting of the IHR (2005) Emergency Committee on the COVID-19 Pandemic*. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic?adgroupsurvey=

Widarjono, A. (2013). *Ekonometrika: Pengantar dan Aplikasinya* (4th ed.). UPP STIM YKPN.

Winata, I. N. (2023, May 26). Jalan Terjal Menuju Pariwisata Berkualitas. *Bali Post*. https://www.balipost.com/news/2023/05/26/341043/Jalan-Terjal-Menuju-Pariwisata-Berkualitas.html

Yang, L., Li, X., & Hernández-Lara, A. B. (2024). Tourism and COVID-19 in China: Recovery and Resilience Strategies of Main Chinese Tourism Cities. *International Journal of Tourism Cities*, *10*(2), 387–404. https://doi.org/10.1108/IJTC-04-2022-0084

Yanwardhana, E. (2021, December 22). Sandi Uno: Bali Penyumbang Devisa Terbesar di Pariwisata RI. *CNBC Indonesia*. https://www.cnbcindonesia.com/news/20211222193439-4-301297/sandi-uno-bali-penyumbang-devisa-terbesar-di-pariwisata-ri

Zhang, C., & Tian, Y.-X. (2022). Forecast Daily Tourist Volumes During the Epidemic Period Using COVID-19 Data, Search Engine Data and Weather Data. *Expert Systems with Applications*, *210*, 118505. https://doi.org/10.1016/J.ESWA.2022.118505

Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast. *IET Intelligent Transport Systems*, *11*(2), 68–75. https://doi.org/10.1049/IET-ITS.2016.0208

Doi: 10.47608/jki.v18i22024.331–350

Jurnal Kepariwisataan Indonesia
Volume 18 Issue 2 2024 page 331–350

AUTHOR PROFILE

Astrinariswari Rahmadian Prasetyo, S.Tr.Stat.
She is IT specialist focused on data science. Graduated from Politeknik Statistika STIS, Jakarta.

Geri Yesa Ermawan S.ST.
He is IT specialist with experience in data collection, processing, and analytics.

Wahyu Calvin Frans Mariel, S.ST.
He is data specialist with experience in data collection, processing, analytics, and communication.

Setia Pramana PhD.
ORCHID ID 0000-0002-8590-1451. He is Lecturer and researcher focus on big data analytics for public policies.

Doi: 10.47608/jki.v18i22024.331–350