THE 1st INTERNATIONAL CONFERENCE ON DATA SCIENCE AND OFFICIAL STATISTICS
ICDSOS

"Strengthening the Contribution of Data Science and Official Statistics to the Society in the Distruption Era"

2021

Short-Term Forecasting of Air Travelers Outflows from Bali Using Web Search Data

W O Parma Dwi [1]

[1] Statistics Indonesia-BPS RI, Jl. Dr. Sutomo 6-8 Central Jakarta Indonesia

*Corresponding author's e-mail: prm.oktama@gmail.com

Abstract. Air travelers have become one of the strategic indicators in the transportation sector. The official data-released by Statistics Indonesia (BPS) for thirty days-lag, makes the condition of this indicator can't be known in real-time. By the utilization of web search data that has been briskly evolving in recent years, this study aims to explore the possibility of using web search data in performing short-term forecasting to know the general outlook of the indicator earlier. Based on this study, web search data and official statistics figures show a strong correlation and having similar movement patterns over time. The application of web search data as a predictor in time series modeling, especially on time series regression and autoregressive model (SARIMA and SARIMAX), turn out a predicted value that well-approach the actual value of the response variable. In addition, it is proven that the use of web search data can increase model accuracy. The analysis results using SARIMAX model shows that the number of air traveller's outflows from Bali in September and October 2021 will generally be higher than the number in August 2021. The increasing number of air travelers is thought due to a decrease in Covid-19 cases which has triggered the public's confidence in travelling about to rise again.

1. Introduction

The use of the internet as support for human life is growing rapidly nowadays. The internet has become a driving engine in the IoT (Internet of Things) and digitalization era that has disrupted almost all sectors of life. With the number of active users around 4.65 billion people all over the world, it brings a great implication for the use of search engines on the internet in supporting human daily activities. In line with the rapid use of search engines, web search data on the internet also get accelerated. One of the search engines that often be used is Google. Since Google has more than 90% of the global market share, it has become the common way for users to explore and access information. Google search performs a search for information according to the keywords entered by the user. The intensity of the web search data with certain keywords produces a large volume of data which is then stored in the Google database. Since 2004, Google has started to disseminate web-search intensity based on keyword volume called Google Trends Index (GT Index). The GT index presents an index that has been normalized in the range 1-100 and is available in various reporting periods ranging from daily to annual indexes. Choi and Varian stated that information described by the GT index often correlate with the phenomenon or trend of current activities which happens in the communities [1].

Many indicators are used to describe the state of the country's economy. In Indonesia, statistics on these indicators are mostly released by the Statistics Indonesia (BPS) with various time-lag depend on the business process of each survey. One of these strategic indicators that have a domino effect on the condition of other economic sectors is the number of air travelers. This indicator has an important contribution especially to the transport sector in general. According to BPS data, in the second quarter of 2021, air transport contributes a share of around 0.59% of Indonesia's GDP. In terms of its business process, this indicator is produced by BPS with a lag of approximately 30 days. With the lag from data collection up to the release, the condition of this sector in the current period can't be known.

By combining the two phenomena above, namely the problem of official statistics data that has a time lag in the release process, and the shift in people's consumption patterns to digital consumption preference in line with the rapid growth of internet usage - includes air travelers in managing their trips, this study aims to examine the possibility of using web search data, especially Google Trends Index (GT), in forecasting the movement of the official statistical data earlier. Furthermore, if there were enough evidence in which web search data could approach the actual data well, the short-term forecasting steps will be performed, which is estimating the number of air traveller's outflows from Bali during September and October 2021.

There have been many studies that use web search data, especially GT in time series modeling. The role of GT in predicting the number of travelers has been carried out by Choi and Varian. In his research, it was found that GT data was able to capture the movement of traveller data to Hong Kong as well as the data released by the National Statistics Officer [1]. Camacho stated that web search data applied on time series regression is appropriate in modeling the number of travelers who visit Spain [2]. In addition, Sangkon Park, Jungmin Lee, and Wonho Song also predict the number of tourist-inflows to South Korea using Google Trends Data. In his research, it was found that Google-compounded models perform better than the common-normal time-series models in terms of short-term forecasting accuracy [3].

The rest of this paper is arranged as follows: In Section 2, there is a comprehensive description of the method used for analysis purposes in this paper. In Section 3, the result of the research where all general descriptions and parameter results are presented. At the end, in Section 4, there are some summary points of our study.

2. Method

2.1. Data source

There are two main data used in this study. Namely the official statistics on the number of air travelers and web search data. This study examines monthly data from January 2011 until August 2021. This period is chosen considering the adequacy of the research sample, the improvement of GT calculation methodology by Google since January 2011, and the process of digitalization on the aviation industry in Indonesia. In general, the details of data sources can be explained as follows

a. As response variable (Y_t) – Air travelers data is taken from monthly data on the number of airplane passengers, both domestic and international departure, who depart from Ngurah Rai International Airport collected by Statistics Indonesia.

b. As predictor variable (X_t) – The web search data is taken from a google trends index which is selected and processed with a certain queries' selection framework. The web search data in September and October 2021 is used to nowcast the number of air travelers' outflows in those periods.

In this paper, we look for queries related to travelers' data outflows by adopting and combining the queries' selection framework proposed by Mitra, P., Anirban, S., and Sohini [4] and Feng, Y, Guowen Li, Xiaolei Sun, Jianping Li [5]. Their construct includes the following steps:

Steps 1 : Define the seed queries. Following Choi and Varian, we start with the preliminary long-listed keyword deemed relevant for air travelers' information [1].

Steps 2 : Determine the related queries from seed queries using google correlate. Google correlate is a tool that can be found on Google Trend which provides a suggestive list of keywords with typically similar search patterns.

Steps 3 : Eliminating the duplicates queries and measure the correlation of all the clean-related queries with the dependent variables.

Steps 4 : Choose a query with the highest correlation corresponds to the dependent variable.

2.2. Analysis Method

The application of inference statistics was first done by using the time series regression model. This model is chosen because it is considered well enough in capturing the dynamics of time series data which is often influenced by lag at the previous time. Furthermore, as a benchmark for comparison of models, autoregressive modeling is also carried out which accommodates the seasonal pattern in the SARIMA model and its addition for exogenous variable on SARIMAX. These models will be evaluated with certain indicators to obtain the best model before being used for the forecasting process of air travelers' outflows in September and October 2021.

2.2.1. Pearson Coefficient Correlation. Pearson correlation analysis is used to determine the strength of the relationship between two theoretically interrelated variables. Determination of the correlation's magnitude can be done by calculating as follows: [6]

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{\{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\} - \{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}} \qquad (1)$$

Where,
r_{xy} = Pearson Coefficient Correlation
x_i = Predictor variable
y_i = Response variable
n = Amount of sample

2.2.2. Time Series Regression. Regression analysis is a data analysis technique used to examine two or more related variables. Time series regression consists of a dependent variable that is influenced by theoretically-known independent variables [7].

$$Y_t = \beta_1 X_{1,t} + \beta_1 X_{2,t} + \cdots + \beta_k X_{k,t} + \alpha_t \qquad (2)$$

Where,
Y_t = Response variable in t-period
$X_{k,t}$ = Predictor variable in t-period
β_k = Regression coefficient from predictors-k
α_t = *error terms*

2.2.3. SARIMA and SARIMAX model. The Autoregressive Integrated Moving Average (ARIMA) model that accommodates a seasonal effect is known as Seasonal Autoregressive Integrated Moving Average (SARIMA). Addition of variables run as predictors (exogenous variables) in the SARIMA model is known as the SARIMAX model [8]. The general form of SARIMA [(p,d,q) (P,D,Q)]^s with modifications of adding exogenous variables can be written as follows [9]:

$$\phi^*(B^s)\phi(B)(1-B)^d(1-B^s)^D y_t = \delta + \beta_1 X_{1,t} + \beta_1 X_{2,t} + \cdots + \beta_k X_{k,t} + \Theta^*(B^s)\Theta(B)\epsilon_t \qquad (3)$$

Where,
$\phi(B)$ = AR *non seasonal*
$\phi^*(B^s)$ = AR *seasonal*
$\Theta(B)$ = MA non-seasonal
$\Theta^*(B^s)$ = MA seasonal
$(1-B)^d$ = *Differencing non seasonal*
$(1-B^s)^D$ = *Differencing seasonal*
$X_{k,t}$ = Exogenous variable-k on period-t
β_k = parameter of exogenous variable-k
δ = *intercept*

2.2.4. Model Selection Indicators. The model selection is done by evaluating the model through two accuracy measures, namely Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Hyndman and Koehler state that RMSE is often used as a model evaluation method because it can equalize the different scales of the data used. The RMSE formula can be written as follows [10]:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \widehat{Y_t})^2}{n}} \tag{4}$$

Mean Absolute Error (MAE) states the average absolute value of the deviation on the forecasting results against the actual data value. MAE can be used to indicate how feasible a model is to use in forecasting. The smaller the MAE of a model, the better the forecast value produced in approaching the actual observed value.

$$MAE = \frac{1}{n}\,|Y_t - \hat{Y}_t| \tag{5}$$

Where,

Y_t	= Actual data
$\widehat{Y_t}$	= Predicted Value
n	= Amount of sample

3. Results and Discussion

3.1. Seed Queries

Seed Queries means all the long-pre listed keywords deemed relevant to the dependent variable. As indicated in the previous section, there were 10 different pre listed keywords were analysed in the initial stage. Those keywords are considered as all first-thought and related aspects when travelers experienced their travel from or into Bali, includes airport's name, tourism places, special cuisine, and regency's name. The final set of initial keywords selected are: 'bali', 'denpasar', 'seminyak', 'ngurah rai', 'kuta', 'ayam betutu, 'nusa penida', 'jimbaran', 'uluwatu', and 'sanur'.

3.2. Related Queries

In this process, the seed queries' scope is extended by google correlate, which is defined as all automatic-suggested queries by google that are mostly connected with the seed queries. There were almost 76 unique queries produced by this step after filtering out the duplicates. In short, this process completing the frame of all-listed queries related to travel in Bali.

3.3. Selection of Queries

After we get the frame by combining the seed and all related queries, then we calculate the correlation of each query toward the official statistics number of air travelers' outflows from Bali. Here are five queries with the highest correlation with the dependent variable:

Table 1. Five queries keyword with strongest correlation value toward dependent variable.

Queries Keyword	Correlation Coeff.
seminyak	0,9353
sanur	0,9056
bali	0,8974
nusa dua	0,8954
jimbaran	0,8697

Table 1 shows us the queries related to tourism places that playing significant roles and fulfilling the top five keywords that mostly correlate with the outflows of people from Bali. By this finding, it can be inferred that in general, people spent their time during at Bali is aimed at tourisms activity, knowing that Bali famous for their tourism's attractiveness ad their special culture that can't be found in any other places. Web Search Queries with the keyword "Seminyak" has the strongest correlation from the other alternatives. By this result, the set data of google trend index from this query is used as a proxy of web search data that roles as independent variables in modeling process.

3.4. General Description of Data

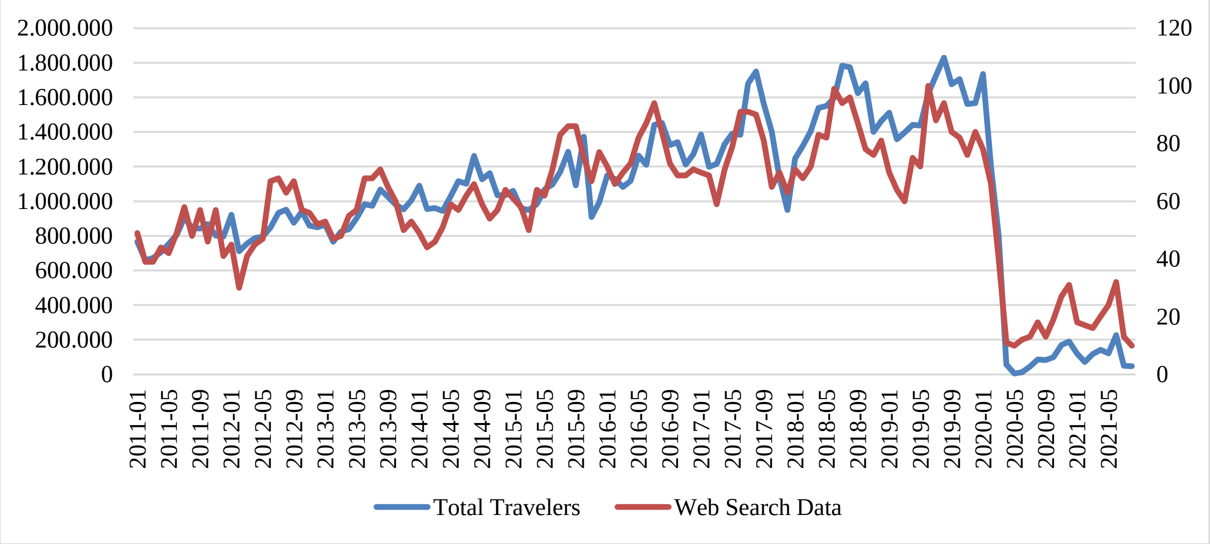

Figure 1. The pattern's comparison of official statistics number and web search data.

From the plot in Figure 1, it can be shown that official statistics' number released by Statistics Indonesia (BPS) and web search data produced with google trends have a similar pattern in general, including their fluctuation and moving direction. Besides, the correlation coefficient which reaches 0,94 shows a strong correlation between those two variables quantitatively. These findings support the idea that a web search data, which can be obtained relatively faster rather than the official statistics figures, could be used as an appropriate proxy to know the condition of indicators earlier.

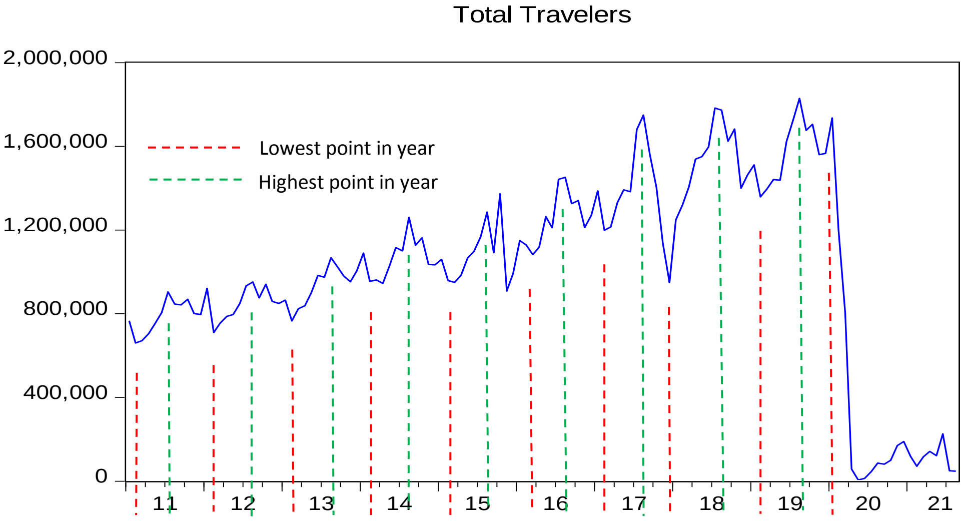

Figure 2. Seasonal patterns of air traveler's outflows from Bali during 2011-2021 period.

From Figure 2 above, it can be seen that the trend of air travelers experiences a monthly-seasonal pattern, where the highest peak in the number of passengers in each year usually occurs during the mid-year holiday around August-September, and the lowest number happen often in January to March.

This phenomenon indicates that the figures last year (t-12) is still an important point to note in describing the fluctuation of variable value in the current period. Besides, the last two months figures (t-1, t-2) also must be considered since a common truth of time series data that often affected by the lag figures. Those are two concerns that the author considers when creating the statistical model, whether using time series regression, SARIMA and SARIMAX.

3.5. Time series regression analysis

Time series regression is done by examining the two model-combination between dependent and independent variables as follows:

Model 1:

This model is built with data of air travelers' outflows in last two months and seasonal effect (t-12) as its independent variables.

$$\boldsymbol{log\ (travelers_t)} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} * \mathbf{log}(\boldsymbol{travelers_{t-1}}) + \boldsymbol{\beta_2} * \mathbf{log}(\boldsymbol{travelers_{t-2}}) + \boldsymbol{\beta_3} * \mathbf{log}(\boldsymbol{travelers_{t-12}}) + \boldsymbol{\varepsilon_t} \quad (6)$$

Table 2. The first equation of air traveler's outflows from Bali with time series regression

Variable	Coeff.	Std. Error	t-Statistic	Prob.
(Intercept)	0,571769	0,686875	0,832421	0,4069
$\log(travelers_{t-1})$	1,298886	0,087060	14,91945	0,0000***
$\log(travelers_{t-2})$	-0,403383	0,090627	-4,451024	0,0000***
$\log(travelers_{t-12})$	0,060542	0,046103	1,313170	0,1918

Sig.	: 1% (***) 5%(**) 10%(*)	Prob. (F-Statistic)	: 0,0000
R-squared	: 0,8743	SIC	: 1,0534
Adjusted R^2 model	: 0,8710	AIC	: 0,9585

The results from Table 2 can be written in equation scheme as follows:

$$\log(travelers_t) = 0{,}5718 + 1{,}2989\ \log(travelers_{t-1}) - 0{,}4034\ \log(travelers_{t-2}) + 0{,}0605 * \log(travelers_{t-12}) + \varepsilon_t \quad (7)$$

Model 2:

This model is built with data of air travelers in last two months, seasonal effect of dependent variables (t-12), web search data in previous month (t-1), and web search data in the current period (t) as its independent variable.

$$log\ (travelers_t) = \beta_0 + \beta_1 * \log(travelers_{t-1}) + \beta_2 * \log(travelers_{t-2}) + \beta_3 * \log(travelers_{t-12}) + \beta_4 * \log(webindex_{t-1}) + \beta_5 * \log(webindex_t) + \varepsilon_t \quad (8)$$

Table 3. Second equation of air traveler's outflows from Bali with time series regression

Variable	Coeff.	Std. Error	t-Statistic	Prob.
(Intercept)	3,357107	0,607610	5,525103	0,0000***
$\log(travelers_{t-1})$	0,822272	0,082552	9,960687	0,0000***
$\log(travelers_{t-2})$	-0,328082	0,058644	-5,594506	0,0000***
$\log(travelers_{t-12})$	-0,006668	0,030142	-0,221205	0,8253
$\log(webdata_{t-1})$	-0,426276	0,159095	-2,679376	0,0085***
$\log(webdata_t)$	1,326606	0,110068	12,05265	0,0000***

Sig	: 1% (***) 5%(**) 10%(*)	Prob. (F-Statistic)	: 0,0000

R-squared	: 0,9498	SIC	: 0,2167
Adjusted R^2 model	: 0,9476	AIC	: 0,0742

The results from Table 3 can be written in equation scheme as follows:

$$log\ (travelers_t) = 3{,}3571 + 0{,}8222 * \log(travelers_{t-1}) - 0{,}3281 * \log(travelers_{t-2}) - 0{,}0067 * \log(travelers_{t-12}) - 0{,}4262 * \log(webdata_{t-1}) + 1{,}3267 * \log(webdata_t) + \varepsilon_t \quad (9)$$

3.6. SARIMA and SARIMAX

SARIMA is an autoregressive model that accommodates seasonal effect in data. When SARIMA model is combined with an exogenous variable as a predictor it will form SARIMAX. These two models then compared to search the best-fitted model with the actual data. The analysis of SARIMA model begins by observing the ACF and PACF plot of response variable in stationary phases.

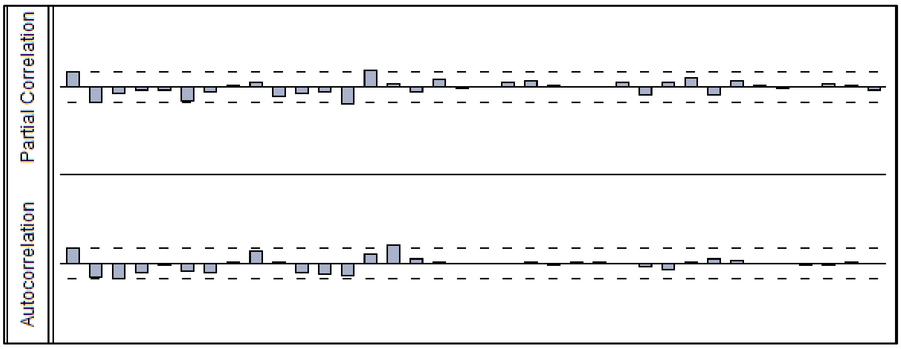

Figure 3. ACF and PACF Plot from the number of air travelers' outflows in 1st diff.

From Figure 3 above, ACF and PACF play significant in the 1st and 2nd lag and they perform a cut-off in the next lag indicate the tentative model of AR (1) or AR (2) and MA (1) or MA(2). Besides, there is a seasonal plot of lag multiple of 12 so then it forms a seasonal plot of SAR (12). By considering the most minimum AIC Value, the tentative SARIMA model that has been chosen: SARIMA $(1,1,1)\ (1,1,0)^{12}$

Table 4. Equation of SARIMA models

Variable	Coeff.	Std. Error	t-statistic	Prob.
(Intercept)	13.45056	0.776239	17.32787	0.0000***
AR(1)	0.843566	0.048758	17.30111	0.0000***
SAR(12)	0.100856	0.145348	0.693892	0.4891
MA(1)	0.617759	0.056327	10.96738	0.0000***

Sig.	: 1% (***) 5%(**) 10%(*)	Prob. (F-Statistic)	: 0,0000
R-squared	: 0,8814	SIC	: 0,9473
Adjusted R^2 model	: 0,8775	AIC	: 0,8359

Considering the SARIMA model in equation (3), the results from Table 4 can be written in equation scheme as follows:

$$y_t = 13{,}4506 + 0{,}8436 y_{t-1} + 0{,}1009 y_{t-12} + 0{,}9149\, y_{t-13} + 0{,}8436 y_{t-14} + 0{,}01009 y_{t-25} + 0{,}0851 y_{t-26} + \varepsilon_t - 0{,}6178\, \varepsilon_{t-1} \quad (10)$$

Table 5. Equation of SARIMAX models

Variable	Coeff.	Std. Error	t-statistic	Prob.
(Intercept)	6.924799	0.391691	17.67925	0.0000***
$\log(webindex)$	1.667697	0.107381	15.53070	0.0000***
$\mathrm{AR}(1)$	0.524146	0.142036	3.690236	0.0003***
$\mathrm{SAR}(12)$	0.265193	0.102355	2.590899	0.0107**
$\mathrm{MA}(1)$	0.077832	0.176049	0.442104	0.6592

Sig.	: 1% $^{(***)}$ 5%$^{(**)}$ 10%$^{(*)}$	Prob. (F-Statistic)	: 0,0000
R-squared	: 0,9279	SIC	: 0,4763
Adjusted R^2 model	: 0,9249	AIC	: 0,3426

Considering the modified SARIMAX model in equation (3), the results from Table 5 can be written in equation scheme as follows:

$$y_t = 6{,}9247 + 1{,}6677X_t + \ 0{,}5246y_{t-1} + 0{,}2652y_{t-12} + 0{,}8610\ y_{t-13} \\ + 0{,}5246y_{t-14}\ \ 0{,}2652y_{t-25} + \ 0{,}1390y_{t-26} + \ \varepsilon_t - 0{,}0778\ \varepsilon_{t-1} \tag{11}$$

3.7. Model Comparison

After dealing with model fitting, the next step is choosing the best-fitted model by comparing the models quantitively with some accuracy indicators. Some indicators that are used can be listed as follows: Adjusted R-squared, F-Statistics, AIC, SIC. Besides, in order to know the suitability of models for forecasting purposes, the RMSE, MAPE, and White Noise Assumption test also be conducted.

Table 6. Comparison of the model accuracy

Model	Adj. R-squared	AIC	SIC	RMSE	MAE	*White Noise*
Regression (1st Model)	0,8710	0,9584	1,0534	1,1386	0,9714	No
Regression (2nd Model)	0,9475	0,0742	0,2167	0,3282	0,2487	No
SARIMA (1,1,1)(1,1,0)^{12}	0,8775	0,8359	0,9473	1,0738	0,7717	No
SARIMAX (1,1,1)(1,1,0)^{12}	0,9249	0,3426	0,4763	0,3396	0,2347	Yes

Table 6 shows us about the comparison between all analysis method that is used in this research. As shown above, the model which includes web search data as a predictor mostly has better accuracy in terms of Lower Mean Square Error compared to other models which exclude it. This finding is in line with the research conclusion done by Collison and D'Amuri, in which models that accommodate web search data produce much better accuracy than the standard time-series models that don't accommodate it [11,12].

By observing the results above, it can be concluded that the SARIMAX model shows the best accuracy compared to the other models. This model also has fulfilled the white noise assumption, so then it convinces us that the predicted value from the models can approach the actual data remarkably good. Therefore, this model is appropriate to be used in conducting short-term forecasting on air traveler's outflows from Bali during September and October 2021.

3.8. Short-term forecasting of the number air travelers' outflows from Bali

Before performing the short-term forecasting into the dependent variable, we must evaluate whether the predicted value has a similar pattern and fluctuation compared to the actual data.

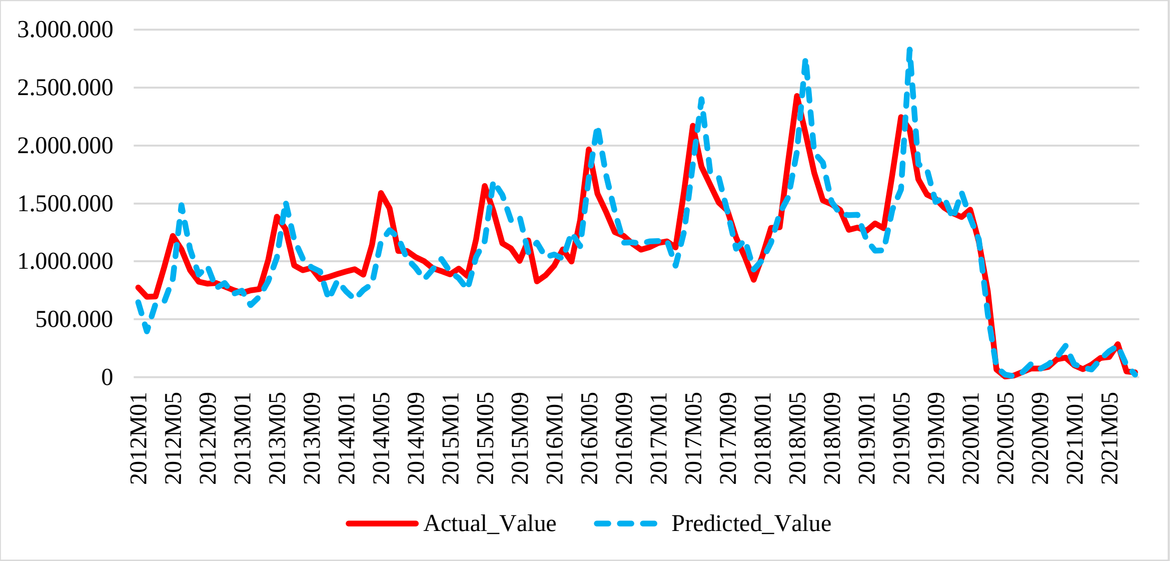

Figure 4. Comparison between the actual data and predicted value.

In general, the plot of predicted value has a similar pattern toward the actual value of the data, includes the moving direction and fluctuation within the years. By these results the model is said to be powerful in performing prediction toward the number of air travelers' outflows from Bali.

Table 7. Short-term forecasting of number air traveler's outflows from Bali in Sept and Oct 2021

Period	**Predicted Value**	***% y-o-y***	***% m-t-m***
September	114,662	28,73%	139,73%
October	263,527	164,43%	129,83%

The result of short-term forecasting is shown in Table 7. From the table above, it can be inferred that in September and October 2021 there will be an increase in the number of air travelers from Bali. This phenomenon could be understood because in July and August there was a tight policy from the central government about the implementation of Emergency PPKM to reduce the increasing spread of Covid-19 cases. While the case number of Covid-19 has been fallen and the number of new cases could be controlled well, it triggers people to travel more, therefore encourages both *m-t-m and y-o-y's* percentage to get a positive value.

4. Conclusion

In this study, it is shown that the use of web search data, especially with the google trends index, is appropriate to use in forming a model on the number of air traveller's outflows from Bali. It can be concluded based on the movement of those two variables which have similar patterns and fluctuation over time. By the availability of web search data which tends to be faster than official statistics figures released by Statistics Indonesia (BPS), it can be used as an alternative data source to give a general outlook on the number of air travelers outflows from Bali in the current period. Based on the results of the study, tourism-related queries still dominating queries that play significantly with people activity in Bali. In the analysis method, it can also be proven that the use of the web search data in the time series model can increase the model's accuracy. By using the SARIMAX model that has been formed, it is predicted that the number of air travelers outflows from Bali rise around 114,662 people in September and 263,527 people in October 2021. The increase both in *m-t-m* and *y-o-y's* trend compared to the previous month is expected to be the impact of de-restriction of PPKM due to the decrease of the new and active cases of Covid-19. Therefore, it is thought that the confidence and trust of public in traveling start to rise again.

References

[1] Choi, H., and Varian, H. 2012 Predicting the present with Google Trends Economic Record, 88(1), Issue s1, pages 2-9.

[2] Camacho, Maximo. & Pacce, Jose Matias 2017 Forecasting Travelers In Spains With Google's Search Volume Indices *SAGE Journals*

[3] Sangkon Park, Jungmin Lee & Wonho Song 2017 Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data, *Journal of Travel & Tourism Marketing*, 34:3, 357-368.

[4] Mitra, P., Anirban, S., and Sohini, C 2017 *Nowcasting Real Estate Activity In India Using Google Trend Data*. Reserve Bank of India Occasional Papers, 38 No. 1&2, 2017.

[5] Feng, Y, Guowen Li, Xiaolei Sun, Jianping Li 2019 Forecasting the Number of Inbound Tourist with Google Trends. *7th Int. Conf. on Information Technology and Quantitative Management (ITQM)*. China.

[6] Makridakis, S., S. Wheelwright, R. Hyndman, and Y. Chang 1998 *Forecasting Methods and Applications* 3rd ed. New York: John Wiley & Sons.

[7] Draper, N. R., & Smith, H. 1998 *Applied Regression Analysis* (3rd ed.) Canada: John Willey & Sons, Inc

[8] Arunraj, N., Ahrens, D., and Fernandes, M 2016 Application of SARIMAX Model to forecast Daily Sales in Food Retail Industry *International Journal of Operation Research and Information Systems.7*(1), 1-21.

[9] Montgomerry, et al 2008 *Introduction to Time Series Analysis and Forecasting*. Amerika: John Wiley & Sons Inc.

[10] Hyndman, R.J., & Koehler, A.B. 2006 Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*, 679-688.

[11] Collison, J.A. 2019 *Performance Assesment of Google Index in Forecasting Car Sales in Argentina* (Argetina: Universidad de Buenos Aires)

[12] D'Amuri, Francesco and Marcucci, Juri 2012 *The predictive power of Google searches in forecasting unemployment* No 891 Temi di discussione (Economic working papers) Bank of Italy, Economic Research and International Relations Area.