Contents lists available at https://citedness.com/index.php/jdsi

Data Science Insights

Journal Page is available to https://citedness.com/index.php/jdsi

Research article

Forecasting Model using Fuzzy Time Series for Tourist Arrivals in Langkawi

Nur Fazliana Rahim [1,*], *Mahmod Othman* [2], *Abdullah Husin* [3]

[1] *Centre for Pre-University Studies, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*
[2,3] *Department Information System, Universitas Islam Indragiri, Indonesia*
email: [1,*] rnfazliana@unimas.my, [2]mahmod.othman@unisi.ac.id, [3]abdullah@unisi.ac.id
** Correspondence*

ARTICLE INFO

Article history:
Received 01 22, 2023
Revised 03 11, 2023
Accepted 05 06, 2023
Available online 08 21, 2023

Keywords:
Forecasting Tourist Arrivals;
Fuzzy Time Series

Please cite this article in IEEE style as:
N.F. Rahim, M. Othman, and A. Abdullah, " Forecasting Model using Fuzzy Time Series for Tourist Arrivals in Langkawi, vol. 1, no. 1, pp. XX-XXX, 2023.

ABSTRACT

In several applications, fuzzy time series forecasting was utilized to generate predictions about the future value of variables that were of interest. This study focuses on predicting how many tourists will visit Langkawi since a precise estimate of tourism demand would enable the government to decide whether to raise or lower the money allocated to the sector in the future. To be more precise, this study attempts to choose the best model that may be applied to forecast visitors to Langkawi and assist the public and private sectors in managing tourism-related preparations. The data collection contains monthly data from January 2009 to December 2010 and was directly extracted from the Langkawi Development Authority (LADA) website. When estimating visitor arrivals to Langkawi, the suggested fuzzy time series' accuracy was compared to that of the earlier technique. The experimental findings in this study demonstrated that the Fuzzy Time Series approach can anticipate more accurately. The results of this study could serve as inspiration for the public and private sectors to take action to bring more tourists to Langkawi, make their stay pleasant and pleasurable, and improve the possibility that they would visit again and again in the future.

Correspondence:
Nur Fazliana Rahim
Centre for Pre-University Studies,
Universiti Malaysia Sarawak, 94300
Kota Samarahan, Sarawak,
Malaysia

1. Introduction

Tourism industry has become a prominent sector that gives an impact on development of the country's economy that can be taken into consideration. Most of the country earnings and job opportunity offers by the tourism industry. One of the popular attentions on the tourist industry, where the potential for tourism includes nature, beaches, and islands that are large enough to be frequently referred to as "green tourism" and "blue tourism." [1]. The government has also made numerous large-scale initiatives to increase security, comfort, and cleanliness so that infrastructure and facilities can be developed that can be enjoyed not just by tourists but also by the entire population, which is estimated to be approximately 30 million [2].

In Malaysia, one of the most popular destinations chosen by the travellers is the Langkawi Island, located in the state of Kedah. In such cases, the Malaysian government need to pay their attention on the growth of the tourist's number that arrive and stay in Langkawi, since it tends to increase the country's potential income [3]. Thus, forecasting the tourism demand is crucially needed and the suitable forecasting model that can validly predict the tourism demand will help the government and private sectors properly build tourist infrastructures and give the best services to the tourists. Tourism industry is found to be an interesting matter to deal with Fuzzy Time Series forecasting.

In the previous study, [4] was conducted a study to forecast wheat production by using the Fuzzy Time Series method to assist the government in managing the wheat yield and storage space. The results showed that the forecasted value is close to the actual data in the wheat production data. Additionally, in the other study [5] an estimation of Taiwan's export trade using an ARIMA and fuzzy time series model was performed. In order to find the best forecasting model that can support and implement economic policies and business policies, particularly in the international export trade, both forecasting methods were examined. According to the findings of the forecasting, the Fuzzy Time Series model performed better than the ARIMA model by having a smaller average

error and being able to predict the Taiwan export trade. Furthermore, it was also proven by [6], fuzzy time series forecasting method able to generate more accurate future values of Crude Palm Oil Prices.

When dealing with tourism industry, the government and private sectors need to invest properly and effectively, which sometimes can cause a shortage of infrastructures and services. This action could minimize the profits gained and dissatisfied the tourists and sometimes it may cause an oversupply of infrastructures and services, and this action could bring loss in the total revenue [7]. Therefore, forecasting is required to assist in their decision-making. Researchers have employed a variety of forecasting methods to predict tourism demand, however there is no data to support the best forecasting technique that can be used to predict tourism demand. Therefore, this study suggested using a fuzzy time series technique to forecast how many tourists will visit Langkawi between January and December 2022.

2. Research Methods

To verify the efficacy of the proposed method, data on monthly tourist arrivals in Langkawi for the year 2022 was gathered. The data was obtained from the Langkawi Development Authority website. The Fuzzy time series model was used to forecast the number of tourists arriving in Langkawi based on the data. The universal set was first defined in this research (universe of discourse). The universe discourse U is defined as [Dmin–D1, Dmax+ D2], where Dmin is the minimum number of tourists and Dmax is the maximum number of tourists respectively. While D1 and D2 are two appropriate positive values selected to make sure the data are inside the range of discourse. Next, intervals of n equivalent intervals with a length of L were constructed to divide the discourse U. The interval was then divided again to obtain four equal-length sub-intervals depending on the rank provided in Table 1, where the highest number of tourists is divided. Subsequently, the intervals with the second largest and third largest number of tourists are divided into three and two sub-intervals of equal length respectively. While the length of the intervals is unchanged for the interval with the fourth-highest number of tourists. If the interval has no distributed data, it was ignored.

Table 1 Sub-Intervals Partitions

	Ranking				
Frequency	First	Second	Third	Fourth	Fifth
Sub-Interval	4	3	2	Unchanged	Unchanged

By establishing the membership functions in corresponding partitions, this research continues. The definition of the fuzzy set Ai is based on the re-divided intervals and fuzzification of historical tourist data, where fuzzy set Ai denotes a linguistic value of the output represented by a fuzzy set and $1 \leqslant i \leqslant 5$. For example, A1 = very few, A2 = few, A3 = moderate, A4 = many, A5 = too many. For simplicity, the membership values of fuzzy set Ai either are 0, 0.5, or 1 where $1 \leqslant i \leqslant 5$. The linguistic values of the tourist after fuzzifying the historical tourist data are A1, A2, ..., A5. To create a fuzzy time series, the crisp values were converted into fuzzy sets using the fuzzifying process. If the value of D located at μi, then it belongs to fuzzy sets Ai. All D must be classified into corresponding fuzzy set.

The fuzzy logical relationships were then constructed using the data from the fuzzified tourists. The fuzzy logical relationship "$A_i \rightarrow A_j$" denotes "IF the fuzzified value of year n – 1 is A_j, THEN fuzzified tourists of year n is A_k. Figure 1 depicted the fuzzy set relationship.

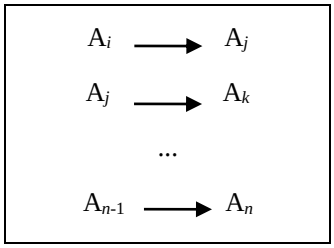

Figure 1. Fuzzy Set Relationship

Since there can only be one relationship of a given kind, it doesn't matter if the same relationships exist more than once. To continue this research, fuzzy logical relationship groups are being derived (FLRGs). Based on the same fuzzy numbers on the left-hand sides of fuzzy logical relationships, the derivation of fuzzy relationships can be grouped into fuzzy logical relationships. The fuzzy logical relationship group is shown as in Eq. (1).

$$\left.\begin{matrix} A_i & A_{j1} \\ A_i & A_{j2} \end{matrix}\right\} \quad A_i \rightarrow A_{j1}, A_{j2}, \ldots, \tag{1}$$

The Fuzzy set A_i of linguistic term number tourist is expressed as in Eq. (2).

$$A_i = F_{A1(u1)/u1} + F_{A2(u2)/u2} + \ldots + F_{Ai(un)/un,} \tag{2}$$

where u_i *(i = 1, 2, ..., n)* is the element and the number below 1 is the membership of u_i to *A (i = 1,2, ..., n).* The general rules for determining the degree of the membership historical data belongs to u_i are expressed as Eq. (3).

$$A_i = \sum_{j=1}^{n} \frac{u_{ij}}{u_{ij}} \tag{3}$$

where u_{ij} is membership degree of u_i belonging to A_i which defined as Eq. (4).

$$u_{ij} = 0.5 \quad \begin{cases} i = j \\ j = i - 1 \quad , \quad i = 1 \\ 0 \quad , \quad otherwise \end{cases} \tag{4}$$

The sketch follows rules as below:

Rule 1: If the historical data belongs to u_i, then the membership degree is 1 of u_i, 0.5 for u_2 and otherwise is 0.

Rule 2: If the historical data belongs to u_i, $1 < i < n$ then the membership degree is 1 of u_i, 0.5 for u_{i-1} and u_{i+1} and otherwise is 0.

Rule 3: If the historical data belongs to u_n, then the membership degree is 1 of u_n, 0.5 for u_{i-1} and otherwise is 0.

Next, the historical data was forecasted. This study proposed heuristic rules and outlined as:

Rule 1: If there is one-to-one relationship say $A_j \rightarrow A_k$ and the highest degree belongs to A_k occur in interval u_k, then forecast output of *F(t)* equal to midpoint of u_k.

Rule 2: If A_j is empty and interval A_j has the highest degree belongs to u_j, then the forecast output equal to midpoint of u_j.

Rule 3: If there exist one-to-many relationship group of A_j, say $A_j \rightarrow A_1, A_2, \ldots An$ and the highest degree belongs occur at set $u_1, u_2, \ldots, u_n$ then forecast output is computed as the average midpoint $m_1, m_2, \ldots, m_n$ of $u_1, u_2, \ldots, u_n$ this equation can be expressed as Eq. (5).

$$\frac{m1 + m2 + m3}{n} \tag{5}$$

Finally, the Mean Square Error (MSE) was calculated using Eq. (6).

$$MSE = \frac{(\text{forecasting data} - \text{actual data})^2}{n} \tag{6}$$

3. Results and Discussion

The findings of the analysis carried out for this study are discussed in this section. The historical data was fuzzify by classifying all the data in a group of A_1, A_2, A_3, A_4, and A_5 based on the intervals. Table 2 below shows the fuzzification of the actual tourist arrivals in Langkawi.

Table 2. Fuzzified the actual tourist arrivals in Langkawi

Month	Total	Fuzzified
January	206,500	A3
February	175,309	A2
March	192,072	A2
April	108,322	A1
May	237,935	A3
June	250,407	A4
July	235,197	A3
August	210,739	A3
September	255,423	A4
October	228,511	A3
November	177,241	A2
December	303,949	A5

The relationships were then established using historical data that had been fuzzified. Figure 2 below shows fuzzy logical relationships, where the fuzzy logical relationship $A_j \rightarrow A_k$ means "If the tourist arrivals of month i is A_j, then that of month $i+1$ is A_k", where A_j is called the current month of the tourist arrivals, and A_k is called the next month of the tourist arrivals. Repeated relationships are counted only once.

$\boldsymbol{A_6 \rightarrow A_6}$	$\boldsymbol{A_4 \rightarrow A_3}$	$\boldsymbol{A_2 \rightarrow A_2}$	$\boldsymbol{A_1 \rightarrow A_1}$
$\boldsymbol{A_1 \rightarrow A_1}$	$A_1 \rightarrow A_1$	$A_1 \rightarrow A_1$	$\boldsymbol{A_1 \rightarrow A_1}$
$\boldsymbol{A_1 \rightarrow A_1}$	$A_1 \rightarrow A_1$	$A_3 \rightarrow A_2$	$\boldsymbol{A_5 \rightarrow A_1}$

Figure 2. Fuzzy Logical Relationships

The right-hand sides are combined if the same fuzzy set relates to many sets. A complete overview of the relationship groups obtained from Table 3 is as below:

Table 3. Fuzzy logical relationship groups

Group	Fuzzy Logical Relationships
Group 1	$A_1 \rightarrow A_1$
Group 2	$A_2 \rightarrow A_2$
Group 3	$A_3 \rightarrow A_2$
Group 4	$A_4 \rightarrow A_2$
Group 5	$A_5 \rightarrow A_1$
Group 6	$A_6 \rightarrow A_6$

This research continues with defuzzifying the forecasted output by assuming that the fuzzified tourist arrivals of $F(t-1)$ isA_j, then forecasted output of $F(t)$ is carried out by the following principles:

(1) If there exists one-to-one relationship in the relationship group ofA_j, say $A_j \rightarrow A_k$, and the highest degree of belongingness of A_k occurs at intervalu_k, then the forecasted output of $F(t)$ equals the midpoint ofu_k.
(2) If A_j is empty,$A_j \rightarrow \varphi$, and the interval where A_j has the highest degree of belongingness is u_j, then the forecasted output equals the midpoint of u_j.
(3) If there exist a one-to-many relationship in the relationship group ofA_j, say$A_j \rightarrow A_1, A_2, \ldots, A_n$, and the highest degrees of belongingness occur at set$u_1, u_2, \ldots, u_n$, then the forecast output is computed as the

average of the midpoints $m_1, m_2, \ldots, m_n$ of $u_1, u_2, \ldots, u_n$. This equation of forecasted tourist arrivals can be expressed as Eq. (7).

$$i + 1 = \frac{m_1 + m_2 + \ldots + m_n}{n} \quad (7)$$

Table 4 below summarized the forecast value for tourist arrivals in Langkawi from January until December 2022.

Table 4. The Forecast Value for Tourist Arrivals in Langkawi from January until December 2022

Month	Total	Forecast Value
January	206,500	206 325
February	175,309	174 264
March	192,072	192 012
April	108,322	109 254
May	237,935	236 944
June	250,407	250 100
July	235,197	236 598
August	210,739	211 125
September	255,423	255 036
October	228,511	227 968
November	177,241	176 348
December	303,949	303 756

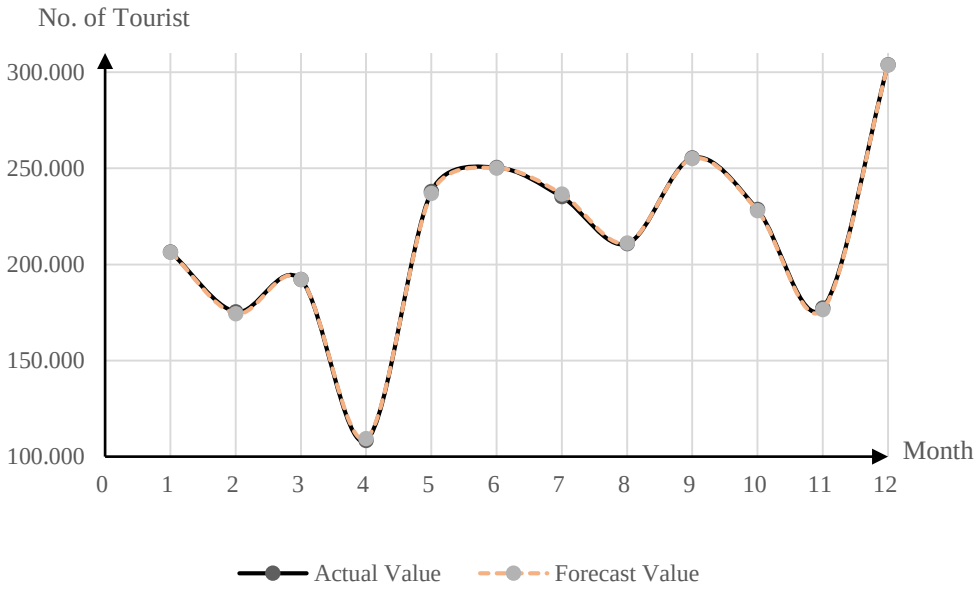

Figure 3. The tourist arrivals in Langkawi from January until December 2022

Referring to the Figure 3, it shows that, from January until December 2022, the tourist arrivals in Langkawi tend to inconsistent but at the end of the year 2022, the tourist arrivals in Langkawi was increase rapidly. When comparing the actual value and the forecasting value, the number of tourist arrivals is almost similar. This mean that, the fuzzy time series method used to forecast the number of tourist arrivals was suitable and best fit with the data.

4. Conclusion

This study's goal is to provide a method for utilizing fuzzy time series to improve forecasting accuracy. It should be clearly stated that the forecasted numbers are based solely on historical data presented numerically, without the use of any extra information. Based on the monthly data taken from Langkawi Development Authority (LADA) website from January until December 2022, it can be concluded that this method able to generate future data value and will help the government and private sectors properly build tourist infrastructures and give the best services to the tourists with the known of future tourist arrivals to their place.

References

[1] M. Abd Hamid and S. Mohd Isa, "Exploring the sustainable tourism practices among tour operators in Malaysia," *J. Sustain. Sci. Manag*, vol. 15, pp. 68–80.

[2] A. Nurhuda, "Benchmarking and Exploring Educational Tourism in Malaysia," *J. English Lang. Teaching, Lit. Cult.*, vol. 2, no. 1, pp. 1–11.

[3] A. K. Patwary, "Examining environmentally responsible behaviour, environmental beliefs and conservation commitment of tourists: a path towards responsible consumption and production in tourism," *Environ. Sci. Pollut. Res.*, vol. 30, no. 3, pp. 5815–5824.

[4] S. Kumar and N. Kumar, "Fuzzy time series-based method for wheat production forecasting," *Int. J. Comput. Appl.*, vol. 44, no. 12, pp. 5–10.

[5] P. Arumugam and V. Anithakumari, "Fuzzy time series method for forecasting Taiwan export data," *Int. J. Eng. Trends Technol.*, vol. 4, no. 8, pp. 3342–3347.

[6] N. F. Rahim, M. Othman, R. Sokkalingam, and E. A. Kadir, "Forecasting crude palm oil prices using fuzzy rule-based time series method," *IEEE Access*, vol. 6, pp. 32216–32224.

[7] H. L. Wu, "Opportunity and Business Policies and its Impact on Tourism in Malaysia," *Soc. Sci. J. Adv. Res.*, vol. 3, no. 1, pp. 10–15.