A Comparative Study of The Weighted High Order Fuzzy Time Series and ARIMA Method in Forecasting Tourist Visits

Istin Fitriana Aziza[1*], Siti Soraya[2], Adawiyah Asti Khalil[3], Ardiana Fatma Dewi[4], Annisa Ramadhan[5]

[1]Information Technology, Universitas Bumigora, Mataram, Indonesia
[2]Computer Science, Universitas Bumigora, Mataram, Indonesia
[3]Digital Business, Universitas Negeri Makassar, Makassar, Indonesia
[4]Data Science, UIN Sayyid Ali Rahmatullah Tulungagung, Tulungagung, Indonesia
[5]Digital Business, Telkom University, Surabaya, Indonesia

*Corresponding author: istin.fitriana@universitasbumigora.ac.id

Received: 12 Februari 2026
Revised: 13 July 2026
Accepted: 7 August 2026

ABSTRACT – The tourism sector is a strategic sector that plays a crucial role in driving regional economic growth. Tourism is a leading sector in West Nusa Tenggara Province, contributing significantly to regional income, job creation, and community welfare. The presence of leading tourist destinations such as the Mandalika Special Economic Zone, Mount Rinjani, and Gili makes West Nusa Tenggara one of the leading tourist destinations in Indonesia. Local governments, tourism businesses, and other relevant parties need information on future tourist visits to plan the provision of facilities and infrastructure, manage tourist destinations, promote tourism, and develop human resources. This study aims to forecast tourist visits to West Nusa Tenggara. The methods used in this study are the weighted high-order fuzzy time series (WHOFTS) and Autoregressive Integrated Moving Average (ARIMA) methods, and these two forecasting methods are compared. The results of this study showed that WHOFTS performs better than ARIMA, as indicated by the lower MAPE value (WHOFTS is 6.17% and ARIMA is 14.67%). The forecasting results will be useful for stakeholders, especially the government, in formulating policies. The WHOFTS method used in this study cannot be applied to data with long-term seasonal patterns. Suggestion that can be given to future researchers is develop the WHOFTS that can capture additional long-term seasonal patterns.

Keywords– Development; Long-term; Non-Stationary; Seasonal; Uncertainty.

I. INTRODUCTION

The tourism sector is strategic and plays a crucial role in driving regional economic growth. In West Nusa Tenggara Province, tourism is a leading sector that contributes significantly to regional income, job creation, and community welfare improvement. The presence of leading tourist destinations such as the Mandalika Special Economic Zone, Mount Rinjani, and Gili Trawangan makes West Nusa Tenggara one of the leading tourist destinations in Indonesia. Tourist arrivals are a main indicator used to measure the development of the tourism sector. Local governments, tourism businesses, and other relevant parties need information on future tourist arrivals to plan the provision of facilities and infrastructure, manage tourist destinations, promote tourism, and develop human resources. Therefore, a forecasting method capable of producing accurate predictions is needed to inform the development of tourism policies and strategies in the region.

Tourist visit data generally exhibit complex characteristics, such as trends, fluctuations, uncertainty, and recurring seasonal patterns. These seasonal patterns can be influenced by various factors, including school holidays, religious holidays, weather conditions, and the implementation of tourism activities and events. These data characteristics often make it difficult for conventional forecasting methods to capture the nonlinear relationships and uncertainties inherent in the data. Furthermore, changes in tourist numbers do not always follow a linear pattern, making classical statistical approaches less optimal in some circumstances.

Forecasting is a systematic process aimed at estimating future values based on historical observations, where the selection of an appropriate technique depends on the data characteristics, influencing factors, and underlying pattern structures [1]. In classical time series analysis, future values are modeled as functions of past observations without incorporating exogenous variables using methods such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing. However, numerous studies have demonstrated that these linear models have limitations in capturing nonlinear dynamics. For example, [2] reported that ARIMA performed poorly in modeling nonlinear water quality variables, yielding significantly lower accuracy than Multilayer Perceptron (MLP) and hybrid ARIMA–MLP models. Similarly, [3] showed that while ARIMA achieved the lowest forecasting error in specific COVID-19 datasets, the model performance remained highly data-dependent. Consequently, hybrid approaches such as ARIMA–LSTM and SARIMAX–LSTM have gained increasing attention because of their ability to integrate linear structures, nonlinear dynamics, and exogenous information, thereby improving predictive accuracy [4], [5], [6], [7]. In tourism forecasting, Exponential Smoothing performs well under stable and short-term conditions but requires modification or hybridization during structural disruptions such as the COVID-19 pandemic [8], [9].

©Department of Statistics, Institut Teknologi Sepuluh Nopember

One method that can be used to address these issues is the Fuzzy Time Series (FTS). The FTS method was first introduced by Song and Chissom and has been further developed by various researchers to improve the forecasting accuracy. The advantage of the FTS method lies in its ability to handle uncertain data and not requiring specific distributional assumptions. Beyond statistical and hybrid models, Fuzzy Time Series (FTS) represent an extension of classical time series methods by incorporating fuzzy set theory and linguistic variables, transforming numerical data into fuzzy sets to better handle uncertainty, vagueness, and nonlinearity [10]. The fuzzy logic system provides a framework for "counting" using linguistic variables instead of numerical values. This study differs from previous research in that it employs the Fuzzy Time Series approach grounded in fuzzy logic, whereas earlier studies primarily relied on statistical models or exponential smoothing techniques. However, the basic FTS method still faces significant challenges in handling non-stationary data; therefore, it is necessary to develop a model that is more adaptive to changes in data patterns over time [11]. Changes in trends in non-stationary data can cause the fuzzy sets formed by the FTS to no longer be representative, thus reducing the forecasting performance [12].

The High Order Fuzzy Time Series (HOFTS) method was developed to address the limitations of the Fuzzy Time Series method, which considers more than one previous period in the forecasting process. A recent study developed a high-order hesitant fuzzy time series forecasting method that can capture fuzzy information more richly than traditional FTS in representing complex uncertainty and randomness [13]. However, The High-Order FTS model still faces problems in the form of inconsistent forecasting rules, reduced utilization of data at higher orders, and sensitivity to the formation of fuzzy intervals [14]. Furthermore, the use of weights in fuzzy relationships resulted in the weighted high-order fuzzy time series (WHOFTS) method, which can provide different contributions at each lag, resulting in more accurate predictions [15]. The use of weights at each lag also allows the model to exert different influences based on the relevance of each historical period to the forecasted value.

Based on this description, research on the application of the Weighted High Order Fuzzy Time Series method to forecast tourist visits in the West Nusa Tenggara Province is crucial. The forecasting results are expected to provide more accurate information on the number of tourist visits in the future, thus providing a basis for more effective and sustainable planning, management, and development of the tourism sector in the province.

II. LITERATUR REVIEW

Fuzzy Time Series

Fuzzy Time Series is a method proposed by Song and Chissom based on fuzzy set theory and the concept of linguistic variables. Fuzzy Time Series is a method used to predict the value of a variable, where historical data are expressed in linguistic values. In classical FTS, fuzzy relations are formed using a previous period. If there is a universe of discourse, U, where U = $u_1, u_2, \dots, u_p$, then the fuzzy set A_i of U with a general membership function is expressed as

$$A_i = \frac{\mu_{A_i}(u_1)}{u_1} + \dots + \frac{\mu_{A_i}(u_p)}{u_p} \quad (1)$$

where μ_{A_i} is the membership function of fuzzy set A_i and is the membership degree from u_i to A_i, where $\mu_{A_i}(u_i) \in [0,1]$ and $1 \le i \le p$. The membership degree value is determined based on the following rules.

Rule 1: If the historical data X_t is included in u_i, then the membership degree value for u_i is 1 and u_{i+1} is 0.5. If u_i and u_{i+1} are not included, the value is declared as zero.

Rule 2: If the historical data X_t is included in u_i, then the membership degree value for ui is 1, and for u_{i-1} and u_{i+1}, it is 0.5. If ui, ui-1, and u_{i1} are not included, it is declared zero.

Rule 3: If the historical data X_t is included in u_p, then the membership degree value for u_p is 1, for u_{p-1} it is 0.5, and if u_i is not included in u_p and u_{p-1}, it is declared zero [16]. The steps for analysis using Fuzzy Time Series are as follows [17]

Universe of Discourse

The Universe of Discourse from historical data is determined using the following formula:

$$U = [X_{\min} - D_1, X_{\max} + D_2] \quad (2)$$

where $X_{\min}$ is the minimum value of the observed variable, $X_{\max}$ is the maximum value of the observed variable, and D_1 and D_2 are arbitrary positive numbers adjusted by the researcher.

2.3 Class and Interval Length

The universe of discourse, U, is divided into several interval classes with the same interval length for each class, as follows: The formation of interval classes uses Sturges' rule with the following formula:

$$K = 1 + 3{,}3\log(n) \quad (3)$$

where n is the number of observed data points. The next step was to determine the interval class length using the following formula:

$$I = \frac{R}{K} \quad (4)$$

where R is the difference between the minimum and maximum values of the observed data, and K is the number of interval classes.

2.4 Fuzzy Sets

Identify fuzzy sets A_1, A2,..., Ap, and fuzzify the historical data. The definitions of A_1, A_2,..., Ap in universe U are as follows:

$$\begin{aligned}
A_1 &= \frac{1}{u_1}+\frac{0,5}{u_2}+\frac{0}{u_3}+\ldots+\frac{0}{u_p}\\
A_2 &= \frac{0,5}{u_1}+\frac{1}{u_2}+\frac{0,5}{u_3}+\frac{0}{u_4}+\ldots+\frac{0}{u_p}\\
A_3 &= \frac{0}{u_1}+\frac{0,5}{u_2}+\frac{1}{u_3}+\frac{0}{u_4}+\ldots+\frac{0}{u_p}\\
&\vdots\\
A_p &= \frac{0}{u_1}+\frac{0}{u_2}+\frac{0}{u_3}+\frac{0}{u_4}+\ldots+\frac{0,5}{u_{p-1}}+\frac{1}{u_p}
\end{aligned} \tag{5}$$

Fuzzification

The first step in the fuzzy inference process is called fuzzification. After data input, the system determines the membership function values and categorizes the data according to predetermined intervals to transform numeric (non-fuzzy) variables into linguistic (fuzzy) variables.

Fuzzy Logical Relationship (FLR)

A Fuzzy Logical Relationship (FLR) was created based on the research data. The fuzzy logical relationship is represented by, where denotes the fuzzy set in the current state and denotes the fuzzy set in the subsequent state.

Fuzzy Logical Relationship Group (FLRG)

The obtained FLRs are classified into groups to form Fuzzy Logical Relationship Groups (FLRG). The FLRG is obtained by combining the values obtained through the FLR formation process. The Chen and Lee models are among the common approaches used to determine the sequence of FLRG formation. This grouping process differentiates between the two models. Because no weighting is used to determine relationships within a group in Chen's model, identical relationships are considered one. In Lee's model, identical relationships are not considered. According to Lee, these values must be calculated because they can affect the predicted value.

Defuzzification

The forecast value obtained from the midpoint of each interval in the FLRG formed in the previous stage is called defuzzification. Prediction was performed by observing the following rules:

Rule 1: If the fuzzification result for the t-th dataset is A_i and there is a fuzzy set that does not have a fuzzy logic relationship, for example, $A_i \rightarrow \emptyset$, where the maximum value of the membership function of A_i is in the interval u_i and the midpoint of u_i is m_i.

Rule 2: If the fuzzification result for the t-th dataset is A_i and there is only one FLR in the FLRG, for example, $A_{i=1} \rightarrow A_{i=2}$, where the maximum value of the membership function of A_i is in the interval u_i, and the midpoint of u_i is m_i.

Rule 3 : If the fuzzification result on the t-th data is A_i and A_i has several FLRs on FLRG, for example Ai $\rightarrow A_{i1}, A_{i2},\ldots, A_{ik}$, where A_i , A_{i1}, A_{i2} ,..., A_{ik} are fuzzy sets and the maximum value of the membership function of A_{i1}, A_{i2} ,..., A_{ik} is in the interval u_{i1}, u_{i2} ,..., u_{ik} and m_{i1}, m_{i2} ,..., m_{ik}.

Weighted High Order Fuzzy Time Series (WHOFTS)

In first-order FTS (classical FTS), the predicted value is influenced by only one previous data point. This limitation prevents the optimal utilization of longer historical information. Therefore, High-Order Fuzzy Time Series (HOFTS) models were developed, namely models that use more than one previous period as the basis for forming fuzzy logic relationships. For example, in the third-order model, the prediction for period t is determined by the conditions in periods t−1, t−2, and t−3. In the WHOFTS method, the terms antecedent and consequent are known, where the antecedent is the previous condition (2nd order) and the consequent is the condition that occurs afterwards. This approach can capture more complex temporal patterns and improve forecasting accuracy. Higher-order fuzzy logic relationships can be expressed as

$$X_t = f\left(X_{t-1}, X_{t-2}, \ldots, X_{t-n}\right) \tag{6}$$

where *n* indicating the model order (Zhang and Wang, 2023) [18]

Weighted High-Order Fuzzy Time Series (WHOFTS) is an extension of HOFTS, in which weights are added to each fuzzy logic relationship. Weights are used to represent the importance or frequency of occurrence of a particular fuzzy relationship, allowing more relevant historical information to have a greater influence on the prediction process. In this method, each high-order FLRG is assigned a weight based on the frequency of occurrence of fuzzy relationships, probability of occurrence of fuzzy states, or a specific weighting scheme determined by the researcher. The advantages of the WHOFTS over the classical HOFTS are improved prediction accuracy, better handling of nonlinear data, and more optimal utilization of historical information [19]. The WHOFTS general equation is as follows:

$$X_t = \omega_1 X_{t-1} + \omega_2 X_{t-2} + \ldots + \omega_n X_{t-n} \tag{7}$$

with

$$\omega_i = \frac{f_i}{\sum_{j=1}^{n} f_j} \tag{8}$$

with

with ω_i is the i-th consequent weight , f_i is frequency of occurrence of i-th consequent in a WFLRG, n is the number of different consequents contained in WFLRG, and $\sum_{j=1}^{n} f_j$ is total frequency of all consequents in WFLRG

Mathematically, defuzzification (prediction) in the WHOFTS can be obtained using a weighted average:

$$F_t = \frac{\sum_{i=1}^{n} \omega_i m_i}{\sum_{i=1}^{n} \omega_i} \tag{9}$$

with

with F_t is predicted value at t-th period, ω_i is i-th consequent weight, m_i is midpoint of the i-th fuzzy consequent interval, n is the number of consequents differs in FLRG

Autoregressive Integrated Moving Average (ARIMA) Box Jenkins

The Box-Jenkins method is a time-series data application method popularized by George-Box and Gwilyn in 1970, combining the autoregressive and moving average approaches. Box-Jenkins recommends that the stationare of a time series data can be determined by differentiating one or more difference processes using the ARIMA model [20]. ARIMA is a method capable of solving various problems in forecasting time-series data, including forecasting the number of tourist visits in West Nusa Tenggara by examining past visit patterns [21].

The ARIMA model is composed of three models: Autoregressive (AR), Moving Average (MA), and Autoregressive and Moving Average (ARMA), which are preceded by checking for stationarity of the data.

If there is an order d (d ≥ 1) in a forecasting process X_t, then a homogeneous non-stationary model, known as the ARIMA (p,d,q) model (16). In general, the ARIMA model can be expressed as in Equation (1) [22]:

$$\emptyset_p(B)(1-B)^d X_t = \theta_0 + \theta_q(B)_{a_t} \tag{10}$$

In the next section, the parameter θ_0 shows the different roles of d = 0 and $d \geq 0$. When d = 0 implies that the data is stationare, such as: $\theta_0 = \mu\,(1 - \theta_1 - \cdots - \theta_q)$. However, if d ≥ 0, it is known as a deterministic trend. This is rarely used or even omitted in the forecasting process.

Plotting the data and identifying the formation of the ARIMA model are necessary. This is related to the stationarity of the data, which is a requirement of the ARIMA model, both stationary in mean and variance [23]. If the data are not stationary in variance, a transformation is needed, and if the data are not stationary in mean, differencing is needed. When the stationary condition is met, the model is identified by examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots [24]. Next, residual white noise assumption testing (diagnostic checking) was performed on several tentative models. The Ljung-Box diagnostic check is used, followed by checking that the formed model meets the residual normal distribution requirement [25]. The following equation is used for the residual white noise assumption test:

$$Q^* = n(n+2)\sum_{k=1}^{p} \frac{\hat{\rho}_k^2}{n-k} \tag{11}$$

with the hypothesis:

$H_0\ :\ \rho_1 = \rho_2 = \cdots = \rho_k = 0$ (residual white noise)

H_1 : at least one $\rho_k \neq 0$, for k=1,2,3,…,n (residual is not white noise)

Equation (10) explains that at a significance level of 5%, if the Q^* value is greater than the $\chi^2_{[q;K-p-q]}$ critical value or the p-value < significance level, then the decision is made to reject H_0, which means that the residual is not white noise.

In forecasting the characteristics of future data, this process is carried out by dividing the data into in-sample and out-sample data. Out-sample data are used to predict or validate the accuracy of data in a forecasting process, so that the model that appears is the best model from the in-sample data [26]. In this study, the best model selection measures used were AIC. The calculation formula for AIC is given in Equation 12.

$$AIC = n\ x\ ln\left(\frac{SSE}{n}\right) + 2k \tag{12}$$

©Department of Statistics, Institut Teknologi Sepuluh Nopember

where n is the number of historical data points, SSE is the sum of squared errors, and k is the number of parameters [27].

A. Accuration Measurement

The inaccuracy of the predicted results can be calculated using the Mean Absolute Percentage Error (MAPE) with the following formula:

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{X_t - F_t}{X_t}\right|}{n} . 100\% \tag{13}$$

where X_t is the historical data for period t, and F_t is the predicted data [28].

III. METODOLOGI

The data used in this study were monthly tourist visit data from January 2021 to December 2024. The data in this study is sourced from the Central Statistics Agency of West Nusa Tenggara through the website https://NTB.bps.go.id/. The method used is the weighted high-order fuzzy time series (WHOFTS) and ARIMA. The steps for completing the research are as follows:

A. The Weighted High Order Fuzzy Time Series (WHOTS) Method

The analysis steps using the WHOFTS method are as follows:

1. Input historical data
2. Determine the universe of discourse
3. Dividing the data into intervals using a specific method
4. Forming fuzzy set
5. Fuzzifying historical data
6. Determining the order of the fuzzy time series
7. Forming a High-Order Fuzzy Logical Relationship (HOFLR).
8. Calculating the weights for each fuzzy relationship
9. Forming a Weighted Fuzzy Logical Relationship Group (WFLRG)
10. Performing defuzzification
11. Accuracy Measurement
12. Generating forecast values

The analysis steps using the WHOFTS method are described in the flowchart in Figure 1.

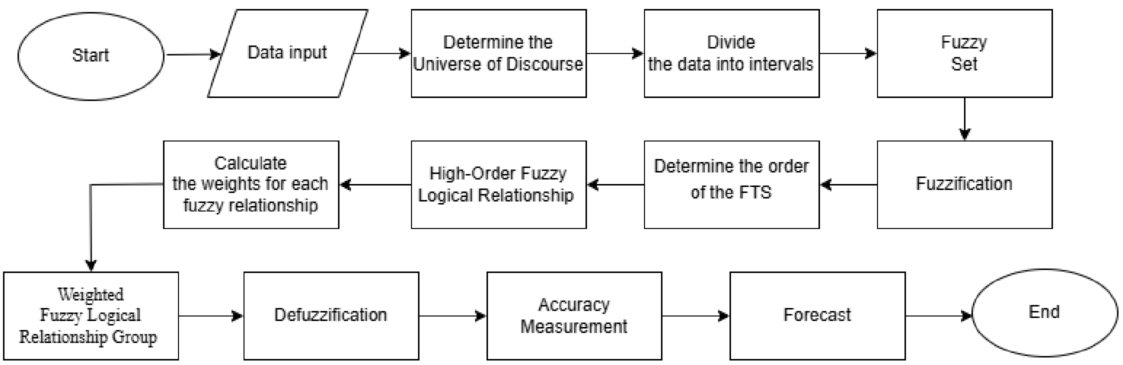

Figure 1. Research Flowchart using WHOFTS Method

B. ARIMA Method

The analysis steps using the ARIMA method are summarized as follows:

1. Input historical data
2. Stationarity of the mean and variance testing
3. Identify the model using ACF and PACF plot
4. Diagnostic Checking (residual white noise)
5. Select the best model based on the AIC
6. Accuracy meausrement
7. Forecast using the best model

The analysis steps using the ARIMA method are shown in Figure 2.

DOI: 10.12962/j27213862.v9i2.9187
©Department of Statistics, Institut Teknologi Sepuluh Nopember

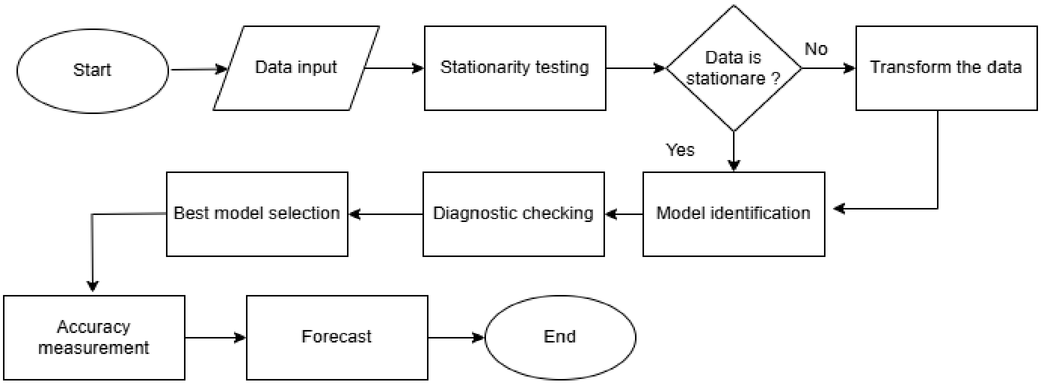

Figure 2. Research Flowchart using ARIMA Method

IV. RESULT AND DISCUSSION

Overview of Tourist Visits in West Nusa Tenggara from January 2021 to December 2024. Figure 3 shows a plot of the number of tourist visits to West Nusa Tenggara from January 2021 to December 2024.

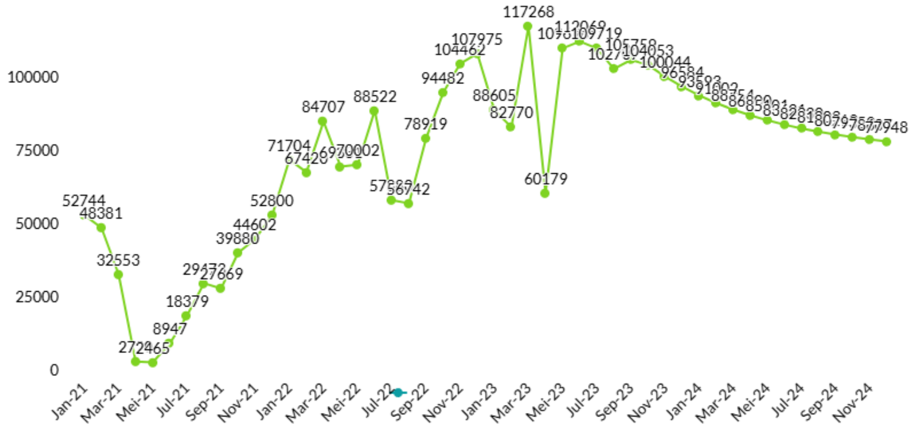

Figure 3. Plot of Tourist Visit Data in West Nusa Tenggara in January 2021 to December 2024

Figure 3 shows that the number of tourist visits to West Nusa Tenggara is volatile. In the early to mid-2021 period, there was a significant decrease, followed by a significant increase from the mid-2021 to mid-2023 period. This increase is thought to be caused by the reopening of international borders after the COVID-19 pandemic. From mid-2023 to the end of 2024, the number of tourist visits decreased. However, the number of visits during that period was still higher than that during 2021 and 2022.

A. The Implementation of the Weighted High Order Fuzzy Time Series (WHOFTS) Method

The weighted high-order fuzzy time series (WHOFTS) method was used to forecast the number of tourist visits in West Nusa Tenggara. The steps taken to forecast the number of tourist visits in West Nusa Tenggara using the WHOFTS method are as follows:

1. The Formation of the Universe of Discourse

The first step is to sort the data, where the minimum value (the least number of tourist visits) in January 2021 to December 2024 is 2,465 people (May 2021), while the maximum value (the highest number of tourist visits) is 117,268 people (March 2023); thus, the difference between the minimum and maximum number of tourist visits (range) is 114,803 people. Based on Equation (2), where the values D_1 and D_2 are positive random numbers, $D_1 = 0$ and $D_2 = 0$, so that the Universe of Discourse formed is:

U = [2645-0, 117268+0] = [2645, 117268].

2. Determination of class and interval length

Before determining the interval length, the number of interval classes must be determined using the Sturges formula based on the formula in Equation (3), so that the number of interval classes is:

$k = 1 + 3{,}3 \log(n) = 1 + 3{,}3 \log(48) = 6{,}54 \approx 7$

Therefore, the number of interval classes obtained was seven. The next step is to determine the length of the class interval based on the formula in Equation (4). The number of interval classes is generated as follows:

$I = R/k$ = 114,803/7 = 16,400

Therefore, the length of the class interval was 16,400. Table 1 presents the lower, upper, and midpoint for each interval class.

©Department of Statistics, Institut Teknologi Sepuluh Nopember

Table 1. Lower Limit, Upper Limit, and Midpoint for Each Interval Class

Interval Class	Lower Limit	Upper Limit	Midpoint
A1	2,465	18,865	10,665,2
A2	18,865	35,266	27,065,6
A3	35,266	51,666	43,466,1
A4	51,666	68,067	59,866,5
A5	68,067	84,467	76,266,9
A6	84,467	100,868	92,667,4
A7	100,868	117,268	109,067,8

3. Fuzzification

The fuzzification process is determined based on the upper and lower bound values presented in Table 1 according to the number of interval classes formed, and the fuzzification results are presented in Table 2.

Table 2. Fuzzification Results of the Number of Tourist Visits in West Nusa Tenggara for in January 2021 – December 2024

Period	The Number of Tourist Visits	Fuzzification	Period	Number of Tourist Visits	Fuzzification
Jan-21	52,744	A_4	Jan-23	88,605	A_6
Feb-21	48,381	A_3	Feb-23	82,770	A_5
Mar-21	32,553	A_2	Mar-23	117,268	A_7
Apr-21	2,720	A_1	Apr-23	60,179	A_4
May-21	2,465	A_1	May-23	109,696	A_7
Jun-21	8,947	A_1	Jun-23	112,069	A_7
Jul-21	18,379	A_1	Jul-23	109,719	A_7
Aug-21	29,473	A_2	Aug-23	102,707	A_7
Sep-21	27,669	A_2	Sep-23	105,759	A_7
Oct-21	39,880	A_3	Oct-23	104,053	A_7
Nov-21	44,602	A_3	Nov-23	100,044	A_6
Dec-21	52,800	A_4	Dec-23	96,584	A_6
Jan-22	71,704	A_5	Jan-24	93,593	A_6
Feb-22	67,420	A_4	Feb-24	91,002	A_6
Mar-22	84,707	A_6	Mar-24	88,754	A_6
Apr-22	69,331	A_5	Apr-24	86,800	A_6
May-22	70,002	A_5	May-24	85,101	A_6
Jun-22	88,522	A_6	Jun-24	83,621	A_5
Jul-22	57,880	A_4	Jul-24	82,330	A_5
Aug-22	56,742	A_4	Aug-24	81,203	A_5
Sep-22	78,919	A_5	Sep-24	80,219	A_5
Oct-22	94,482	A_6	Oct-24	79,359	A_5
Nov-22	104,462	A_7	Nov-24	78,607	A_5
Dec-22	107,975	A_7	Dec-24	77,948	A_5

4. Fuzzy Logical Relationship (FLR)

This study used the weighted high-order fuzzy time series method with order 2, where the prediction is influenced by the previous two periods. The FLR results are presented in Table 3.

 ©Department of Statistics, Institut Teknologi Sepuluh Nopember

Table 3. Fuzzy Logical Relationship (FLR) Order 2 Results

t-2	t-1	t	FLR
A_4	A_3	A_2	$(A_4, A_3) \rightarrow A_2$
A_3	A_2	A_1	$(A_3, A_2) \rightarrow A_1$
A_2	A_1	A_1	$(A_2, A_1) \rightarrow A_1$
A_1	A_1	A_1	$(A_1, A_1) \rightarrow A_1$
A_1	A_1	A_2	$(A_1, A_1) \rightarrow A_2$
⋮	⋮	⋮	⋮
A_5	A_5	A_5	$(A_5, A_5) \rightarrow A_5$
A_5	A_5	A_5	$(A_5, A_5) \rightarrow A_5$
A_5	A_5	A_5	$(A_5, A_5) \rightarrow A_5$

5. Weighted Fuzzy Logical Relationship Group (WFLRG)

The formation of the WFLRG is carried out by grouping fuzzy sets that have the same antecedent based on Table 3 and then grouping them into one group. The WFLRG formed is shown in Table 4.

Table 4. Results of the Weighted Fuzzy Logical Relationship Group (WFLRG)

Grup	WFLRG
1	$(A_4, A_3) \rightarrow A_2$
2	$(A_3, A_2) \rightarrow A_1$
3	$(A_2, A_1) \rightarrow A_1$
4	$(A_1, A_1) \rightarrow A_1, A_2$
5	$(A_1, A_2) \rightarrow A_2$
6	$(A_2, A_2) \rightarrow A_2, A_3$
7	$(A_2, A_3) \rightarrow A_3$
8	$(A_3, A_3) \rightarrow A_4$
9	$(A_3, A_4) \rightarrow A_5$
10	$(A_4, A_5) \rightarrow A_4, A_5, A_6$
11	$(A_5, A_4) \rightarrow A_5$
12	$(A_5, A_5) \rightarrow A_5, A_6, A_5, A_5, A_5, A_5, A_5, A_5$
13	$(A_5, A_6) \rightarrow A_4, A_7$
14	$(A_6, A_4) \rightarrow A_4$
15	$(A_4, A_4) \rightarrow A_5$
16	$(A_6, A_7) \rightarrow A_7$
17	$(A_7, A_7) \rightarrow A_6, A_7, A_7, A_7, A_7, A_6$
18	$(A_7, A_6) \rightarrow A_5, A_6$
19	$(A_6, A_5) \rightarrow A_7, A_5$
20	$(A_5, A_7) \rightarrow A_4$
21	$(A_7, A_4) \rightarrow A_7$
22	$(A_4, A_7) \rightarrow A_7$
23	$(A_6, A_6) \rightarrow A_6, A_6, A_6, A_6, A_5$

6. Weighting Process

The weight is calculated by dividing the frequency of occurrence of a consequent by the total frequency of all consequents in the same WFLRG. Therefore, the consequences that appear more frequently have a greater weight and exert a more dominant influence on the prediction results. The weight calculation formula is presented in Equation 8. The weight calculation results are presented in Table 5.

©Department of Statistics, Institut Teknologi Sepuluh Nopember

Tabel 5. The Weight Calculation Results

Antecedent	Consequent	Frequency	Weight	Midpoint
(A4,A3)	A2	1	1,0	26285,71
(A3,A2)	A1	1	1,0	9428,57
(A2,A1)	A1	1	1,0	9428,57
(A1,A1)	A1	1	0,5	9428,57
(A1,A1)	A2	1	0,5	26285,71
(A1,A2)	A2	1	1,0	26285,71
(A2,A2)	A2	1	0,5	26285,71
(A2,A2)	A3	1	0,5	43142,86
(A2,A3)	A3	1	1,0	43142,86
(A3,A3)	A4	1	1,0	60000,00
(A3,A4)	A5	1	1,0	76857,14
(A4,A5)	A4	1	0,3	60000,00
(A4,A5)	A5	1	0,3	76857,14
(A4,A5)	A6	1	0,3	93714,29
(A5,A4)	A5	1	1,0	76857,14
(A5,A5)	A5	7	0,9	76857,14
(A5,A5)	A6	1	0,1	93714,29
(A5,A6)	A4	1	0,5	60000,00
(A5,A6)	A7	1	0,5	110571,40
(A6,A4)	A4	1	1,0	60000,00
(A4,A4)	A5	1	1,0	76857,14
(A6,A7)	A7	1	1,0	110571,4
(A7,A7)	A6	2	0,3	93714,29
(A7,A7)	A7	4	0,7	110571,4
(A7,A6)	A5	1	0,5	76857,14
(A7,A6)	A6	1	0,5	93714,29
(A6,A5)	A5	1	0,5	76857,14
(A6,A5)	A7	1	0,5	110571,4
(A5,A7)	A4	1	1,0	60000,00
(A7,A4)	A7	1	1,0	110571,4
(A4,A7)	A7	1	1,0	110571,4
(A6,A6)	A5	1	0,2	76857,14
(A6,A6)	A6	4	0,8	93714,29

7. Defuzzification

The defuzzification process in the WHOFTS is the stage of converting the prediction results in the form of fuzzy sets into numeric values. Defuzzification was performed using the weighted average of the consequent midpoints for each WFLRG. The defuzzification calculation is performed using the formula in Equation 9, so that defuzzification is generated from the WFLRG in Table 6.

Table 6. WFLRG Defuzzification Results

Period	The Number of Tourist Visits	State	Defuzzification (Predict)
Jan-21	52744	A4	
Feb-21	48381	A3	
Mar-21	32553	A2	26285,71
Apr-21	2720	A1	9428,571
May-21	2465	A1	9428,571

©Department of Statistics, Institut Teknologi Sepuluh Nopember

Jun-21	8947	A1	17857,14
Jul-21	18379	A2	17857,14
Aug-21	29473	A2	26285,71
Sep-21	27669	A2	34714,29
Oct-21	39880	A3	34714,29
Nov-21	44602	A3	43142,86
Dec-21	52800	A4	60000,00
Jan-22	71704	A5	76857,14
Feb-22	67420	A4	76857,14
Mar-22	84707	A5	76857,14
Apr-22	69331	A5	76857,14
May-22	70002	A5	78964,29
Jun-22	88522	A6	78964,29
Jul-22	57880	A4	85285,71
Aug-22	56742	A4	60000,00
Sep-22	78919	A5	76857,14
Oct-22	94482	A6	76857,14
Nov-22	104462	A7	85285,71
Dec-22	107975	A7	110571,40
Jan-23	88605	A6	104952,40
Feb-23	82770	A5	85285,71
Mar-23	117268	A7	93714,29
Apr-23	60179	A4	60000,00
May-23	109696	A7	110571,40
Jun-23	112069	A7	110571,40
Jul-23	109719	A7	104952,40
Aug-23	102707	A7	104952,40
Sep-23	105759	A7	104952,40
Oct-23	104052	A7	104952,40
Nov-23	100043	A6	104952,40
Dec-23	9658	A6	85285,71
Jan-24	93593	A6	90342,86
Feb-24	91001	A6	90342,86
Mar-24	88753	A6	90342,86
Apr-24	86800	A6	90342,86
May-24	85100	A5	90342,86
Jun-24	83620	A5	93714,29
Jul-24	82329	A5	78964,29
Aug-24	81203	A5	78964,29
Sep-24	80219	A5	78964,29
Oct-24	79359	A5	78964,29
Nov-24	78606	A5	78964,29
Dec-24	77947	A5	78964,29

B. The Implementation of Autoregressive Integrated Moving Average (ARIMA) Box Jenkins Method

The Box-Jenkins method is a time series data application method popularized by George-Box and Gwilyn in 1970, combining the Autoregressive and Moving Average approaches. Box-Jenkins recommends that the stationare of a time series data can be determined by differentiating one or more difference processes using the ARIMA model.

Stationarity occurred if the data did not experience increases or decreases (data fluctuated around an average value and remained constant over time). Data can be used for forecasting if its mean and variance are stationary. The results of the stationarity test of the variance on the number of tourist visits in West Nusa Tenggara are shown in Table 7.

©Department of Statistics, Institut Teknologi Sepuluh Nopember

Table 7. Stationarity Test in Terms of Variance

Variable	Rounded value
The Number of Tourist Visit	1.000

Table 7 shows that the rounded value obtained is 1.000, which indicates that the data are stationary in terms of variance. The next step is the stationarity test of the mean using the Augmented Dickey-Fuller (ADF) test. The results of the stationarity test of the mean on tourist visits in West Nusa Tenggara using the ADF test are shown in Table 8.

Table 8. Stationarity Test in Terms of Mean (Level)

Variable	t - statistic	p-value
The Number of Tourist Visit	-0.173	0.618

H_0 : there is a unit root (not stationare in terms of mean)
H_1 : there is no a unit root (stationare in terms of mean)

Based on Table 8, the p-value is 0.618, which is greater than the significance level (0.05); therefore, we fail to reject H_0 and conclude that there is a unit root (not stationary in terms of mean). Therefore, differencing must be performed. The results of the stationarity test of the mean for the differenced data using the ADF test are shown in Table 9.

Table 9. Stationarity Test in Terms of Mean (Differenced Data)

Variable	t - statistic	p-value
The Number of Tourist Visit	-2.476	0.017

H_0 : there is a unit root (not stationary in terms of mean)
H_1 : there is no a unit root (stationary in terms of mean)

Table 9 shows that the p-value is 0.017, which is less than the significance level (0.05); therefore, we reject H_0 and conclude that there is no unit root (stationary in terms of mean) for differenced data. The next step is model identification using the Autocorrelation Function (ACF) plot to identify the Moving Average (MA) order and the Partial Autocorrelation Function (PACF) plot to identify the autoregressive (AR) order. The ACF and PACF plots are shown in Figure 4.

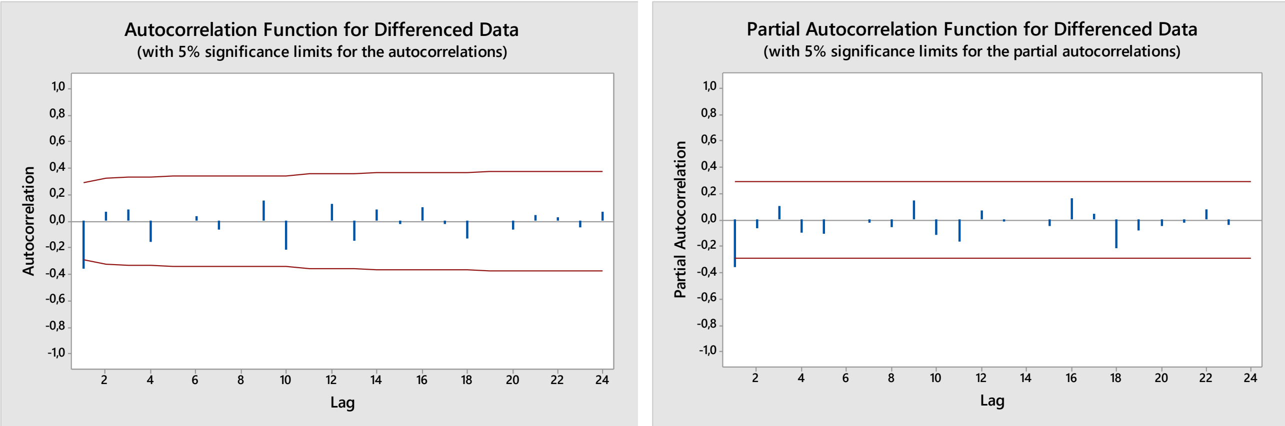

Figure 4. ACF and PACF Plot for Identify The Moving Average and Autoregressive Order

Based on Figure 4, The ACF plot shows a cutoff at lag 1, indicating that the Moving Average order is (q=1) or MA (1), and the PACF plot shows a cutoff at lag 1, indicating that the autoregressive order is (p=1) or AR(1), and it is known that the differencing has been carried out once (d = 1). So, it can concluded that the possible model is ARIMA (1,1,1). Once the preliminary model estimation process is obtained, the next step is parameter estimation. The results of the parameter estimation are presented in Table 10.

Table 10. Parameter Estimation for Each Tentative Model

Tentative Model	Parameter	t-statistic	P-value
ARIMA (1,1,0)	AR (1)	-2.580	0.013
ARIMA (1,1,1)	AR (1)	-0.620	0.538
	MA (1)	0.310	0.761

H_0 : parameter is not significant
H_1 : parameter is significant

Table 10 shows that in the ARIMA (1,1,0) model, the p-value of parameter AR(1) is 0.013. It is less than the significance level (α = 0.05); therefore, we reject H_0 and conclude that parameter AR (1) in the ARIMA (1,1,0) model is significant. ARIMA (1,1,1) had two parameter, that are AR (1) and MA (1). The p-values of parameters AR(1) and MA

(1) were 0.538 and 0.761, respectively. Both are greater than the significance level (α = 0.05); therefore, we failed to reject H_0 and concluded that parameter AR (1) and MA (1) in the ARIMA (1,1,1) model is not significant.

A diagnostic check was used to ensure that the model met the white noise assumption. A model is considered good if the error values are random, indicating that the residuals have no specific pattern (i.e., white noise). Diagnostic checking using the Ljung-Box test. The results of the Ljung-Box test are shown in Table 11.

Table 11. Diagnostic Check Result

Model		Lag		
		12	24	36
ARIMA(1,1,0)	Chi-Square	6.6	12.7	15.9
	P-Value	0.830	0.958	0.998

H_0 : residuals have no specific pattern (white noise)
H_1 : residuals have specific pattern (white noise not fulfilled)

Table 11 shows that the p-values for each lag of 12, 24, and 36 are 0.830, 0.958, and 0.998, respectively. These are greater than the significance level (0.05); therefore, we failed to reject H_0 and concluded that the residuals have no specific pattern (white noise). ARIMA (1,1,0) had a significant parameter and met the white noise assumption. So, we can concluded that the best model is ARIMA (1,1,0). This model was used in the forecasting process.

C. Accuracy Measurement

The Mean Absolute Percentage Error (MAPE) is used to measure the inaccuracy of forecasting. MAPE calculates the average percentage error between the predicted and actual values. This research used two approaches, the WHOFTS and ARIMA methods, to forecast the number of tourist visits in West Nusa Tenggara. The MAPE for both methods is shown in Table 12.

Table 12. MAPE Value of the WHOFTS and ARIMA Method

Criterion	The WHOFTS	ARIMA
MAPE	6.17%	14.67%

Based on Table 12, the MAPE value for the WHOFTS is 6.17%, and for ARIMA, it is 14.67%. The findings of this research indicate that the WHOFTS method is better than ARIMA in forecasting tourist visits, which is in line with the research conducted by [29], where FTS has a better accuracy (including smaller MAPE) than ARIMA in the case of tourist visits.

D. Forecasting using the Best Method

Forecasting was performed using the WHOFTS method. Forecasting for the 49th (January 2025) period based on the last state of the actual data (period December 48, 2024). The midpoint was calculated based on the WFLRG Defuzzification in Table 6. The midpoint value becomes the forecast value (numeric forecast) for the subsequent period. Based on the weighting and midpoints in Table 5, the results of the forecasting using the WHOFTS method are presented in Table 13.

Table 13. Forecasting Results using the WHOFTS Method

Period	Antecedent	Forecasting state	Forecast
January 2025	(A_5, A_5)	A_5	78,964.285
February 2025	(A_5, A_5)	A_5	78,964.285
March 2025	(A_5, A_5)	A_5	78,964.285
April 2025	(A_5, A_5)	A_5	78,964.285
May 2025	(A_5, A_5)	A_5	78,964.285

The state for the next period (49-th period or January 2025) is A5. Because this step produced A5 again, the steps will repeat with the same list, so the forecast remains 78,964.285 until 53-th period (May 2025).

V. Conclusion and Suggestion

Based on the results of the research, it can be concluded that the best method for forecasting tourist visits in West Nusa Tenggara is the WHOFTS, with an accuracy level of 93.83%, and the forecast remains 78,964.285 in the 49-th until 53-th period. The classical WHOFTS tends to perform better on volatile and nonlinear data but is difficult to capture long-term seasonal patterns or strong trends without additional modifications. Suggestion that can be given to future researchers is develop the WHOFTS that can capture additional long-term seasonal patterns or strong trends.

Acknowledgment

The authors are very grateful to the LPPM team of Bumigora University as well as the resource persons of the scientific article writing training, which made it possible for this study to be completed. Also, to the fellow authors involved, so that this manuscript could be published.

 ©Department of Statistics, Institut Teknologi Sepuluh Nopember

REFERENCES

[1] Hyndman, R. J., & Athanasopoulos, G," *Forecasting: Principles and Practice."* 3rd ed. 2021.
[2] B. Propagation, "A Hybrid ARIMA And BP Neural Network Model To Forecast Industrial Water Use Efficiency Xinghua Fan," vol. 1012, pp. 1042–1046. 2014
[3] O. Postavaru, S. R. Anton, and A. Toma, "Science Direct COVID-19 Pandemic and Chaos Theory," *Math. Comput. Simul.*, vol. 181, pp. 138–149. 2021.
[4] Q. Askari and I. Younas, "Improved Political Optimizer for Complex Landscapes and Engineering Optimization Problems," *Expert Syst. Appl.*, vol. 182, no. May, p. 115178. 2021.
[5] Y. Zhou, C. Li, L. Chen, and T. Huang, "Global Exponential Stability of Memristive Cohen-Grossberg Neural Networks with Mixed Delays and Impulse Time Window," *Neurocomputing*, 2017.
[6] Y. Huang and R. Startz, "Improved Recession Dating Using Stock Market Volatility," *Int. J. Forecast.* 2019.
[7] A. B. Casado-díaz *et al.*, "Tourists ' Water Conservation Behavior in Hotels : The Role of Gender of Gender," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–21, 2020.
[8] H. Wu, H. Long, and J. Jiang, "Mixed-Order Fuzzy Time Series Forecast," *Mathematics*, vol. 13, no. 11, pp. 1–15, 2025.
[9] W. Haifeng, H. Yaqi, L. Jinxia, and W. Hongxu, "The Forecast Study of Sanya Tourism Income Based on Fuzzy Time Series Forecasting Model F5," in Advances in Wireless Communications and Applications: Wireless Technology: Intelligent Network Technologies, Smart Services and Applications, Proceedings of 3rd ICWCA 2019, Springer , pp. 213–219. 2020.
[10] K. K. Gupta and S. Saxena, "Probabilistic Fuzzy Time Series Forecasting Method Based on Cumulative Probability Based Discretization Approach and Particle Swarm Optimization," *SN Comput. Sci.*, vol. 6, no. 7, pp. 1–18, 2025.
[11] Wang, B., & Liu, X," Adaptive Non-Stationary Fuzzy Time Series Forecasting With Bayesian Networks," *Sensors, 25*(5), 1628. 2025.
[12] Zhan, T., He, Y., Deng, Y., & Li, Z," Differential Convolutional Fuzzy Time Series Forecasting," IEEE Transactions on Fuzzy Systems. 2023
[13] Gupta, K. K., & Kumar, S," K-Means Clustering Based High Order Weighted Probabilistic Fuzzy Time Series Forecasting Method," Cybernetics and Systems, 54(2), 197–219. 2023.
[14] Ortiz Arroyo, D," A Fuzzy Time Series-Based Model Using Particle Swarm Optimization and Weighted Rules,". 2023.
[15] García, J., et al, "Applying Fuzzy Time Series For Developing Forecasting Electricity Demand Models," Mathematics, *11*(17), 3667. 2023.
[16] Wang, X., Li, Y., & Chen, Z," Forecasting Using Fuzzy Time Series: An Improved Model with Adaptive Intervals and Dynamic Weights," Journal of Forecasting, 43*(2)*, 345–362. 2024.
[17] Lucas, P. O., Orang, O., Silva, P. C. L., & Mendes, E. M. A. M," A Tutorial On Fuzzy Time Series Forecasting Models: Recent Advances And Challenges," Learning and Nonlinear Models, 19(2), 29–50. 2022
[18] Zhang, L., Li, Y., & Wang, X," A Novel High Order Hesitant Fuzzy Time Series Forecasting By Using Mean Aggregated Membership Value With Support Vector Machine," Information Sciences, 626, 494–523. 2023.
[19] Wang, B., Liu, X., Chi, M., & Li, Y," Bayesian Network Based Probabilistic Weighted High-Order Fuzzy Time Series Forecasting," Expert Systems with Applications, 237, 121430. 2024
[20] G. Omulo and E. M. Kumeh, "Technological Forecasting & Social Change Farmer-to-Farmer Digital Network as a Strategy to Strengthen Agricultural Performance in Kenya : A Research Note on ' Wefarm ' Platform," *Technol. Forecast. Soc. Chang.*, vol. 158, no. May, p. 120120, 2020.
[21] H. Ratu, P. Negara, R. Aprianto, S. Soraya, and M. Indrawati, "A Hybrid Modified Orca Predation Algorithm with Backpropagation Neural Network Mathematical Modeling Approach for Time Series Forecasting : Comparative Analysis of Inflation and Export Data". 2025.
[22] A. A. Rizal, S. Soraya, and M. Tajuddin, "Sequence to Sequence Analysis with Long Short Term Memory for Tourist Arrivals Prediction," J. Phys. Conf. Ser., vol. 1211, no. 1, 2019.
[23] M. Bufalo, G. Orlando, and R. La Sapienza, "Improved Tourism Demand Forecasting with CIR Model : a Case Study of Disrupted Data Patterns in Italy," vol. 79, no. 2, pp. 445–464, 2025.
[24] D. Didiharyono and M. Syukri, "Forecasting With Arima Model in Anticipating Open Unemployment Rates in South Sulawesi," Int. J. Sci. Technol. Res., vol. 9, no. 3, pp. 3838–3841, 2020.
[25] Luo, L,"Statistical Model Validation Through White Noise Hypothesis Testing In Regression Analysis And ARIMA Models," Theoretical and Natural Science, *50*(1), 99–104. 2024.
[26] D. Didiharyono and M. Syukri, "Forecasting With Arima Model in Anticipating Open Unemployment Rates in South Sulawesi," Int. J. Sci. Technol. *Res.*, vol. 9, no. 3, pp. 3838–3841, 2020.
[27] Amaefula, C. G," Comparative Analysis Of Information Criteria With A Forecast-Based Criterion For Optimal ARIMA Model Identification: Empirical Evidence Using Naira–Franc Exchange Rate," Asian Journal of Pure and Applied Mathematics, *5*(1), 123–133. 2023.
[28] Siddique, M. A. B., et al," Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: Methodology, Evaluation, and Case Study in SAARC countries," Scientific Reports, *12*, 17907. 2022.
[29] M. A. Jafridin *et al.*, "Comparison of Fuzzy Time Series and ARIMA to Forecast Tourist Arrivals to Homestay in Pahang," *J.* Comput. Res. Innov., vol. 6, no. 4, pp. 80–89, 2021.

© 2026 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (http://creativecommons.org/licenses/by-sa/4.0/).