

# A machine learning approach in Python is used to forecast the number of train passengers using a fuzzy time series model

Solikhin<sup>1</sup>, Septia Lutfi<sup>1</sup>, Purnomo<sup>2</sup>, Hardiwinoto<sup>3</sup>

<sup>1</sup>Department of Informatics Engineering, STMIK Himsya, Semarang, Indonesia

<sup>2</sup>Department of Mechanical Engineering, Faculty of Engineering, Universitas Muhammadiyah Semarang, Semarang, Indonesia

<sup>3</sup>Department of Accounting, Faculty of Economics, Universitas Muhammadiyah Semarang, Semarang, Indonesia

## Article Info

### Article history:

Received Dec 20, 2021

Revised Jun 2, 2022

Accepted Jul 14, 2022

### Keywords:

Frequency-based partitioning

Machine learning

Prediction

Rate of change

Transportation public

## ABSTRACT

Train passenger forecasting assists in planning, resource use, and system management. forecasts rail ridership. Train passenger predictions help prevent stranded passengers and empty seats. Simulating rail transport requires a low-error model. We developed a fuzzy time series forecasting model. Using historical data was the goal. This concept predicts future railway passengers using Holt's double exponential smoothing (DES) and a fuzzy time series technique based on a rate-of-change algorithm. Holt's DES predicts the next period using a fuzzy time series and the rate of change. This method improves prediction accuracy by using event discretization. positive, since changing dynamics reveal trends and seasonality. It uses event discretization and machine-learning-optimized frequency partitioning. The suggested method is compared to existing train passenger forecasting methods. This study has a low average forecasting error and a mean squared error.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Solikhin

Department of Informatics Engineering, STMIK Himsya

Jl. Raya Karanganyar Tugu Km. 12 No. 58 Semarang, Indonesia

Email: solikhin@stmik-himsya.ac.id

## 1. INTRODUCTION

Time series forecasting is crucial in today's dynamic climate. Forecasting helps businesses plan. This helps management and administration make good judgments. Management information systems, rail transportation, economic forecasting, sales forecasting, budgeting analysis, and stock market analysis benefit from time series techniques. Real-time forecasting was difficult. Choosing an appropriate forecasting strategy is vital. Time series data has been forecast using moving average, auto regression, or a combination of these. In complicated systems like rail transport, the future value of time series data isn't governed by deterministic mathematical models. probabilistic In the last 20 years, time series forecasting has changed. Processing power and soft computing have improved this. Fuzzy logic, artificial neural networks (ANN), and evolutionary algorithms are time series alternatives. Create a trustworthy model for accurate forecasts.

The importance of passenger forecasting on trains, in the context of global transportation use, it is vital to improve services across diverse modes of transportation and their interconnections. Rail transportation will continue to be important in the future as long as its capacity is increased and its long-term, efficient, and safe service is kept up. Strengthening network interoperability and national security are required to enhance the railway market [1]. Rail services, both provision and use, are important for public transit. Developing countries need rail transport. Mass rail public transportation is affected by how well legal railway company resources are planned and used.

A railroad corporation must precisely forecast demand for public rail service to operate it well. This helps the organization deliver effective train travel. A new study says that train passengers could be used to figure out how the rail industry uses its resources. Add up and realize orders, as well as cancel requests, to determine arrivals [2], [3]. A railroad company's ability to reserve seats is determined by the number of passengers predicted at a specific departure time. On the other hand, supply may not always meet demand, which can cause long lines of people who want to get on the train but can't.

Many of a business's productive and effective operational decisions are based on accurate demand projections [3]-[6]. In order to improve operational efficiency and economic income, rail transportation business actors must make precise predictions about how many people will ride the train [6]. For the development of future rail transport expansion strategies, investments, and the effectiveness of the facilities [7], as well as for the development of local economies, the allocation of resources, and cutting costs [8], accurate transportation volume estimations are essential. The rail business utilizes it to decide whether to run more trains [9], how to distribute tickets [10], and ticket prices [11].

Accurate passenger forecasts are essential for train resource planning and distribution. Predicting future rail passenger numbers takes a lot of time and effort. Previous passenger surveys were erratic and time-consuming. The forecasting process and results must be clear to both rail service operators and decision-makers. Public transportation resource use and passenger service planning rarely use soft computing. Labs and case studies are forecasting methods. This doesn't reflect the rail system's complexity, so it's impractical. Despite complex time series data, modern forecasting is largely theoretical and statistical. Use a valid hybrid method to predict train passenger numbers. This machine learning-hybrid method should accurately estimate time series data.

Related work, Balabanov *et al.* [12] shows how an ANN trained with differential evolution (DE) can be used to predict time series. The neural network model is better at handling non-linear circumstances, but it takes too long to train. The anticipated number isn't accurate because the data is inconsistent and uncontrollable. FTS is a cutting-edge soft computing method. FTS can discover trends or cyclical portions in time series observations. Fuzzy logical links replace random and non-random functions in time series analysis.

Previous academics have employed the FTS model to solve forecasting issues. In [13], [14] used time series data and fuzzy set theory to get a rough estimate of how many students attend the university. As a measure of dependence, Song [15] utilized the auto-mean correlation function. Chen [16] looks at high-order FTS models and suggests ways to solve them with less math. Fuzzy forecasting systems have been developed by other researchers [17]. Chen and Tanuwijaya [18] suggests that FTS and automated clustering can be used to make a multivariate fuzzy forecasting method for TAIEX. Garg *et al.* [19] suggests optimizing an OWA-based sales model with a genetic algorithm. Yu [20] created a forecasting model using the iterative connection principle. The transition weights were used to calculate the weights in the expectation and class selection technique [21]. Chang and Huang [22] makes weighted fuzzy rules by counting each fuzzy relationship and giving it a weight with OWA operators.

Jiang *et al.* [23] says that FTS and advanced optimization algorithms can be used to predict how much tourism will come into China in the future. The forecast of visitor arrivals using machine learning and internet search indexes Sun and Huang [24] is another related problem. An FTS model with one optimization method is used in [25]. Then Silva [26] introduced the non-stationary FTS (NSFTS). The forecasting of stock indices can be made more accurate with the help of a novel weighted fuzzy trend time series algorithm [27]. Effective interval length [28], [29] influences how many fuzzy sets are produced, and has an impact on the generation of fuzzy [30], [31]. A new computational FTS model proposed by Garg [32] can be used to predict how often outpatient visits will happen. Solikhin *et al.* [33] used the FTS model and a statistical method called Holt's double exponential smoothing (DES) to predict the number of Indonesian train passengers. It uses the discretization approach to divide the interval into sub-intervals based on frequency ranking, from highest to lowest.

Based on FTS research for forecasting challenges, look for tiny but significant signs beneath clear directives. This study proposes an innovative forecasting strategy to fix the problem. The proposed strategy uses data from multiple perspectives. Estimate and predict Indonesian train passenger numbers using FTS, which integrates a rate of change (RoC) algorithm and Holt's DES method. Holt's DES forecasts the next era based on the FTS, which forms the universe set. The best solution is to use machine learning to combine FTS modeling with Holt's DES modeling.

We altered the FTS model [32], [33]. We enhanced forecast accuracy by applying an event discretization approach and a new interval-splitting technique. This new technique is preferable to relying on historical facts. Discrete event processing reduces erroneous predictions by using past time series RoC data. RoC identifies seasonal tendencies. The new frequency-based partitioning works better with different data frequencies. The model stresses math. According to this study, including all of the above parameters

enhances forecast accuracy. Compare the suggested method to one that Solikhin *et al.* [33] used in a previous study that used the same data about railroad passengers.

## 2. METHOD

This section describes a fuzzy technique for time series forecasting. The algorithm's strength is its event discretization and frequency-based partitioning. "Partitioning by frequency at intervals" is used to weight the intervals depending on how often the rate of change occurs by discretizing time series occurrences in terms of data RoC and defining the rate of change universe of discourse. Fuzzy sets are defined by intervals that are divided into subintervals, and the rate of change is fuzzyfied to construct each  $X_i$  fuzzy set as the next defuzzification step. The fuzzy set  $X_i$  shows the linguistic value of the rate of change over time. Determine the rate of change by calculating the midpoint of the resultant interval using the triangle membership function. Lastly, a value is estimated based on the expected change rate.

### 2.1. Basic algorithm definitions

#### 2.1.1. Fuzzy set

A pair is an example of a fuzzy set  $(A, m)$  with  $A$  being a set and  $m$  being  $A \rightarrow |0,1|$ . The fuzzy set  $(A, m)$  is typically represented by  $\{(m(x_1)/x_1), \dots, (m(x_n)/x_n)\}$  for a finite set  $A = \{x_1, \dots, x_n\}$ .  $m(x)$  is the value of membership of  $x$  in each,  $x \in A$ .  $(A, m)$ . If  $m(x) = 0$ ,  $x$  is not included in the fuzzy set  $(A, m)$ ,  $x$  is included if  $m(x) = 1$ , and  $x$  is a fuzzy associate if  $m(x) = 0$ . The support of  $(A, m)$  is the set  $x \in A \mid m(x) > 0$ , and the kernel is the set  $x \in A \mid m(x) = 0$  [32].

#### 2.1.2. Time series

Observations are made sequentially in a predetermined order. A time series is represented in time domain analysis by the mathematical model  $G(t) = O(t) + R(t)$ , where  $O(t)$  denotes a logical or systematic part and  $R(t)$  represents an undetermined component. The two sections are interdependent and may have multiple parameters [32].

#### 2.1.3. Fuzzy time series

Let  $Y(t)$ , real-numbers subset  $(t = \dots, 0, 1, 2, \dots)$ , the discourse universe of 'fuzzy sets'  $f_j(t)$  are defined.  $Y$  if it is a set of  $f_1(t), f_2(t)$ , and so on  $(t)$  [32].

#### 2.1.4. Fuzzy relationship

Let  $F(t-1)$  equal  $A_i$  and  $F(t)$  equal  $A_j$ .  $A_i \rightarrow A_j$  it's on the left and  $A_j$  that which is on the right of a fuzzy logical relationships (FLRs) between two observations that follow each other in time,  $F(t)$  and  $F(t-1)$ . It's known as a "fuzzy logic" connection [32].

## 2.2. The fundamental notions of the algorithm

### 2.2.1. The procedure for discretizing events

FTS theory says discretization reduces discourse complexity. Linking events from different time periods helps prepare for numerical evaluation of the universe of speech. Forecasting approaches use time-series data [34]. Time series differences improve forecast accuracy. Differences can't be used to predict time series growth and decline. In our method, the discourse universe is the RoC from  $t$  to  $t+1$ .

It is possible to define the event discretization function  $RoC(t+1) = (X(t+1) - X(t)) \setminus X(t)$ , where  $X(t+1)$  is the value at the time  $t+1$  index and  $X(t)$  is the actual value at time  $t$  index, so that the value at time  $t$  index corresponds to an occurrence of the event at a later date in the future. From time  $t$  to time  $t+1$ , the RoC is defined as the change in value. As an example: As indicated in Table 1, the rate of change for the period 2012/02 is,  $(9515 - 10223) \setminus 10223$  which equals  $-6.93\%$ . The rate of change for the next year or month is determined similarly [32].

### 2.2.2. The procedure for frequency-based partitioning

After dividing the universe of speech into equal intervals  $(u_1, u_2, u_3, \dots, u_n)$ , then partition by frequency. In this section, we use and change a method for dividing frequency that has been described in previous publications [32], [33]. These are the alterations:

- Count how many rate of change frequencies fall within each interval.
- Divide the interval into numerous sub-intervals based on the total number of RoC frequencies. If the total number of RoC frequencies is 1 or 0, the interval is fixed or not partitioned.
- rep for the next interval in the same manner.

Table 2 illustrates how this works by displaying some example data for a variety of RoC frequencies. In Table 2, it shows that the intervals [-15, -10] and [-5, 0] have one rate of change data frequency, hence the fixed interval or not is separated into sub-intervals. The next interval is [-10, -5], which has a total rate of change data frequency of four and may be broken into four subintervals: [-10, -8.75]; [-8.75, -7.5]; [-7.5, -6.25]; and [-6.25, -5]. The interval [0, 5] will be separated into three sub-intervals [0, 1.67], [1.67, 3.33], and [3.33, 5].

Table 1. RoC calculation example

Time	Time series data	RoC (%)
2012/01	10223	
2012/02	9515	-6.93
2012/03	10787	13.37
2012/04	9926	-7.98

Table 2. Frequency of the rate of change at each interval

Intervals	Number of RoC
[-15, -10]	1
[-10, -5]	4
[-5, 0]	1
[0, 5]	3

**2.2.3. Using a triangular membership function, we define a fuzzy set**

The triangle membership function generates an interval in which  $A_j=1, 2, 3, 4, \dots, n$  is used to build a fuzzy set. Take the average of the values in the interval to compute the expected rate of change in value. Then, use the membership function of the triangle [35] in (1) to estimate the RoC of the data.

$$t_j = \begin{cases} \frac{1+0.5}{\frac{1}{a_1} + \frac{0.5}{a_2}} & , \text{if } j = 1, \\ \frac{0.5+1+0.5}{\frac{0.5}{a_{j-1}} + \frac{1}{a_j} + \frac{0.5}{a_{j+1}}} & , \text{if } 2 \leq j \leq n - 2, \\ \frac{0.5+1}{\frac{0.5}{a_{n-1}} + \frac{1}{a_n}} & , \text{if } j = n. \end{cases} \tag{1}$$

Where  $a_{(j-1)}$ ,  $a_j$  and  $a_{(j+1)}$  denote the mean of the fuzzy intervals corresponding to  $x_{(j-1)}$ ,  $x_j$ , and  $x_{(j+1)}$ , respectively. The  $t_j$  function forecasts rate of change (FRoC) in the actual data history from one period to the next. Based on the forecast value  $t_j \rightarrow F(t)$ , compute the forecasting result, where  $x_{(t-1)}$  denotes the location of the data at  $t-1$ .

**2.2.4. Calculate the forecast for the next time period (t+1)**

Combining approaches is done using Holt's DES approach. Using only linear equations, Holt's DES method is often used in business and economics to predict trends in a set of time series data [36]. Introduction [37] discusses a set of forecasting methods that use exponential smoothing. In business, it is common to use forecasting models from the family of equations called "exponential smoothing" [38]. The exponential smoothing is extended to a trend time series by the DES [39]. As stated in (2) to (6), calculate a prediction for the future time period  $t+1$ :

$$S'_t = \alpha X_t + (1 - \alpha)(S'_{t-1} + t_{t-1}) \tag{2}$$

$$t_t = \beta(S'_t - S'_{t-1}) + (1 - \beta)t_{t-1} \tag{3}$$

$$F_{t+m} = S'_t + t_{tm} \tag{4}$$

$$S'_1 = X_1 \tag{5}$$

$$t_1 = ((x_2 - x_1) + (x_4 - x_3))/2 \tag{6}$$

Data at the point in time  $t$  is represented by  $X_t$ , single-value smoothing  $S'_t$  is denoted here, the smoothing trend's value is denoted by the symbol  $t_t$ , alpha and beta are smoothing parameters ranging from 0 to 1, the forecast value is  $F_{(t+m)}$  and  $m$  is the future period.

**3. RESULTS AND DISCUSSION**

**3.1. Passengers on trains: an experimental investigation**

In this experiment, we use the Central Statistics Agency's dataset on the change in Indonesian rail passengers between January 2006 and December 2019 [33] to test our model. Figure 1 shows a graphical

representation of the dataset. The proposed algorithm's stepwise calculation is shown below. The following are the steps for using a FTS with a RoC technique to model the prediction of the number of railway passengers. This is a very early stage in the process of determining the data for the number of train passengers over time, that is  $X=x_1, x_2, x_3, \dots, x_n$ . Using the discretization process and (7), the following basic procedure can be used to figure out the RoC of the time series data.

$$RoC_{(t+1)} = \frac{(X_{(t+1)} - X_{(t)})}{X_{(t)}} * 100 \tag{7}$$

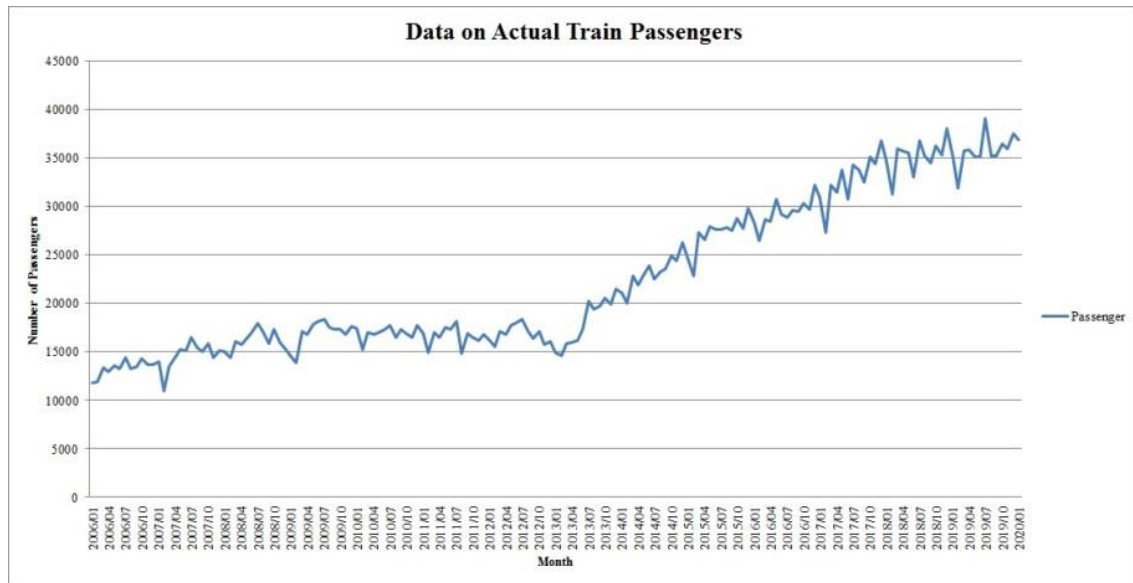


Figure 1. Train passenger dataset graph

Step 1: define the scope of the discourse universe U

Determine the scope of the conversational universe depending on the value of the rate of change obtained. To accomplish this, you'll need to go through numerous stages, among which [33]:

- a. Determining the lowest level (LL) and upper level (UL) based on the rate of change. As a result, (8) can be used to determine U.

$$U = [LL - D_1, UL + D_2] \tag{8}$$

From the rate of change data, we get  $LL=-21.43$  and  $UL=23.53$ . Define U as  $[-23, 25]$ , where  $D_1$  equals 1.57 and  $D_2$  equals 1.47.  $D_1$  and  $D_2$  are positive numbers that aid in the definition of the U universe set.

- b. Using (9), count how many intervals there are [40]:

$$B = 1 + 3.3 * \log(n) \tag{9}$$

$$B=1+3.3*\log(167)=8.3350 \approx 8$$

where n is the number of RoC

- c. In (10) is used to calculate the length of the interval.

$$P = (UL - LL)/B \tag{10}$$

$$P=(25-(-23))/8=6$$

So the same intervals are obtained as a result of this process, and they are as follows:  $u_1=[-23, -17]$ ,  $u_2=[-17, -11]$ ,  $u_3=[-11, -5]$ ,  $u_4=[-5, 1]$ ,  $u_5=[1, 7]$ ,  $u_6=[7, 13]$ ,  $u_7=[13, 19]$ , and  $u_8=[19, 25]$ .

Step 2: frequency-based partitioning

After dividing the discourse universe U into equal intervals, such as  $u_1, u_2, u_3, \dots, u_n$ . For partitioning based on frequency. Calculate the quantity of the rate of change that has dropped in each interval; determine the frequency of intervals depending on the suitable amount of the rate of change, as shown in Table 3. With each frequency, divide the interval into many sub-intervals. In this section, we modify the proposals [32], [33]. Carry on with the same procedure for the next interval division.

Table 3. Frequency of the rate of change

Intervals	Number of RoC
[-23.00 , -17.00]	2
[-17.00 , -11.00]	3
[-11.00 , -5.00]	23
[-5.00 , 1.00]	65
[1.00 , 7.00]	40
[7.00 , 13.00]	25
[13.00 , 19.00]	6
[19.00 , 25.00]	3

The first interval [-23, -17] is divided into two subintervals [-23, -20] and [-20, -17]. The second interval [-17, -11] has three subintervals: [-17, -15], [-15, -13], and [-13, -11]. The third interval [11, -5] is subsequently subdivided into twenty-three subintervals. There are sixty-five sub-intervals within the fourth interval [-5, 1]. There are forty subintervals within the fifth interval [1, 7]. There are twenty-five subintervals within the sixth interval [7, 13]. There are six sub-intervals within the seventh interval [13, 19]. The final interval [19, 25] consists of three subintervals: [19, 21], [21, 23], and [23, 25]. The next step is to calculate the mean value of the subinterval, as presented in Table 4 (in appendix).

Step 3: the rate of change value is defuzzified

Use the triangle membership function in (1) to figure out the rate of change in the data for this session.

Step 4: estimate the value of the predicted data based on the projected discovery RoC.

The F(t) forecasting data is determined by the results of the FRoC. In (11) is used to determine the value of F(t).

$$F(t) = (FRoC/100 * x_{(t-1)}) + x_{(t-1)} \tag{11}$$

where  $x_{(t-1)}$  equals the actual data up to t-1.

Step 5: determine the following time period's forecast

While the traditional forecasting method of Holt's DES was used for the t+1 prediction using the parameters  $\alpha=0.377001$  and  $\beta=0.007219$ . When using (2) to (6), the smoothing value for December, 2019 was 36766, while the smoothing value for the trend was 145. The projected value for January 2020 is  $36766+145*1=36911$ , according to (4). Using (11), following this, the projected results for January 2020 are:  $F_{(Jan\ 2020)}=(-1.44/100*37463)+37463=36924$ . Figure 2 shows the final prediction results.

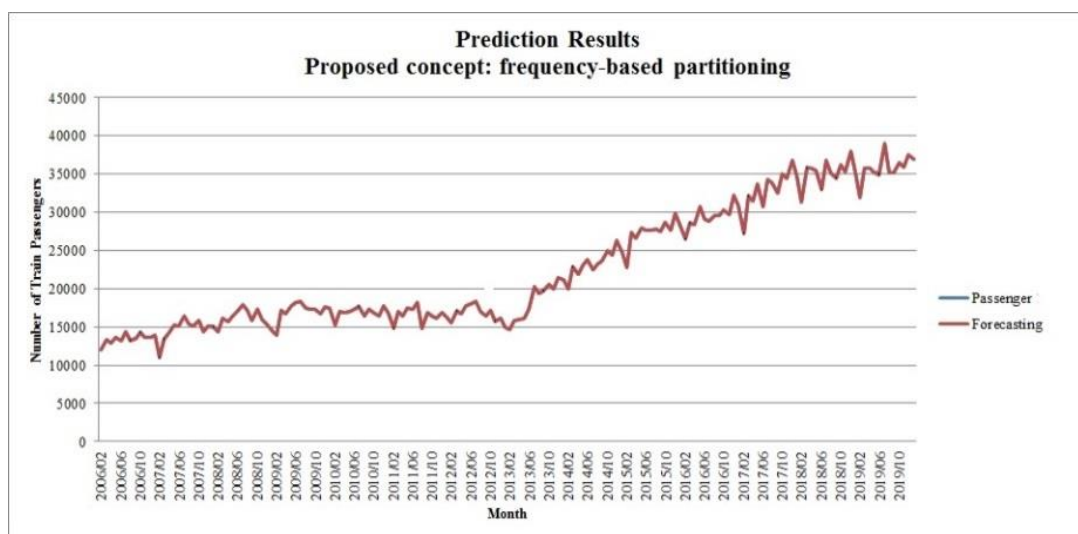


Figure 2. Prediction graph based on the proposed concept

**3.2. Performance comparison and evaluation**

On the same rail passenger data, we compare our proposed model's performance in forecasting passenger numbers to that of the previous model [33]. We also use statistics on train passengers from the same year for verification. In Table 5, average forecasting error rate (AFER) and mean square error (MSE), as calculated by (12) and (13) [32], compare the proposed model's performance with that of existing models. Higher forecasting accuracy is measured by lower MSE and AFER values.

$$AFER = (\sum_{n=1}^t (|A_t - F_t| / A_t)) / n * 100\% \tag{12}$$

$$MSE = \sum_{n=1}^t (A_t - F_t)^2 / n \tag{13}$$

Table 5. Comparison results

Proposer	The concept that was used	AFER (%)	MSE
The current proposed model	Frequency-based partitioning	0.078	935
Solikhin [33]	Ranking-based partition	0.89	131325

Table 5 shows that the results of our AFER and MSE are better than those of the previous model, which used the same data and case study. The suggested model combines discretization event function, frequency-based partition, and optimization with an integrated and innovative Python machine learning technique to maintain consistency and predictability. By basing the discourse universe on historical static data and weighting the intervals based on RoC and frequency, the disadvantage is reduced. Past and future values affect the FTS's current value.

**4. CONCLUSION**

A new technique has the best accuracy and least MSE based on past forecasts. Our approach accurately predicts train passenger counts in every region. The model helps railroads plan, allocate, and manage resources. The technique can help training organizations judge. This strategy could boost rail transit efficiency. Future models will use more trustworthy methods and time-series data.

**ACKNOWLEDGEMENTS**

The authors are thankful that this research was paid for by the Indonesian Ministry of Education, Culture, Research, and Technology. In addition, the writers would want to convey their gratitude to the editorial team and reviewers as a result of their recommendations and remarks that are constructive and which have assisted us in improving the presentation of our work.

**APPENDIX**

Table 4. Distribution of frequencies, fuzzy sets, and the value in the middle

Fuzzy sets	Intervals	Middle Value	Fuzzy sets	Intervals	Middle Value	Fuzzy sets	Intervals	Middle Value
A <sub>1</sub>	[-23, -20]	-21.5	A <sub>57</sub>	[-2.42, -2.32]	-2.37	A <sub>113</sub>	[3.85, 4]	3.93
A <sub>2</sub>	[-20, -17]	-18.5	A <sub>58</sub>	[-2.32, -2.23]	-2.28	A <sub>114</sub>	[4, 4.15]	4.08
A <sub>3</sub>	[-17, -15]	-16	A <sub>59</sub>	[-2.23, -2.14]	-2.18	A <sub>115</sub>	[4.15, 4.3]	4.23
A <sub>4</sub>	[-15, -13]	-14	A <sub>60</sub>	[-2.14, -2.05]	-2.09	A <sub>116</sub>	[4.3, 4.45]	4.38
A <sub>5</sub>	[-13, -11]	-12	A <sub>61</sub>	[-2.05, -1.95]	-2	A <sub>117</sub>	[4.45, 4.6]	4.53
A <sub>6</sub>	[-11, -10.74]	-10.87	A <sub>62</sub>	[-1.95, -1.86]	-1.91	A <sub>118</sub>	[4.6, 4.75]	4.68
A <sub>7</sub>	[-10.74, -10.48]	-10.61	A <sub>63</sub>	[-1.86, -1.77]	-1.82	A <sub>119</sub>	[4.75, 4.9]	4.83
A <sub>8</sub>	[-10.48, -10.22]	-10.35	A <sub>64</sub>	[-1.77, -1.68]	-1.72	A <sub>120</sub>	[4.9, 5.05]	4.98
A <sub>9</sub>	[-10.22, -9.96]	-10.09	A <sub>65</sub>	[-1.68, -1.58]	-1.63	A <sub>121</sub>	[5.05, 5.2]	5.13
A <sub>10</sub>	[-9.96, -9.7]	-9.83	A <sub>66</sub>	[-1.58, -1.49]	-1.54	A <sub>122</sub>	[5.2, 5.35]	5.28
A <sub>11</sub>	[-9.7, -9.43]	-9.57	A <sub>67</sub>	[-1.49, -1.4]	-1.45	A <sub>123</sub>	[5.35, 5.5]	5.43
A <sub>12</sub>	[-9.43, -9.17]	-9.3	A <sub>68</sub>	[-1.4, -1.31]	-1.35	A <sub>124</sub>	[5.5, 5.65]	5.58
A <sub>13</sub>	[-9.17, -8.91]	-9.04	A <sub>69</sub>	[-1.31, -1.22]	-1.26	A <sub>125</sub>	[5.65, 5.8]	5.73
A <sub>14</sub>	[-8.91, -8.65]	-8.78	A <sub>70</sub>	[-1.22, -1.12]	-1.17	A <sub>126</sub>	[5.8, 5.95]	5.88
A <sub>15</sub>	[-8.65, -8.39]	-8.52	A <sub>71</sub>	[-1.12, -1.03]	-1.08	A <sub>127</sub>	[5.95, 6.1]	6.03
A <sub>16</sub>	[-8.39, -8.13]	-8.26	A <sub>72</sub>	[-1.03, -0.94]	-0.98	A <sub>128</sub>	[6.1, 6.25]	6.18
A <sub>17</sub>	[-8.13, -7.87]	-8	A <sub>73</sub>	[-0.94, -0.85]	-0.89	A <sub>129</sub>	[6.25, 6.4]	6.33
A <sub>18</sub>	[-7.87, -7.61]	-7.74	A <sub>74</sub>	[-0.85, -0.75]	-0.8	A <sub>130</sub>	[6.4, 6.55]	6.48
A <sub>19</sub>	[-7.61, -7.35]	-7.48	A <sub>75</sub>	[-0.75, -0.66]	-0.71	A <sub>131</sub>	[6.55, 6.7]	6.63

Table 4. Distribution of frequencies, fuzzy sets, and the value in the middle (continue)

Fuzzy sets	Intervals	Middle Value	Fuzzy sets	Intervals	Middle Value	Fuzzy sets	Intervals	Middle Value
A <sub>20</sub>	[-7.35, -7.09]	-7.22	A <sub>76</sub>	[-0.66, -0.57]	-0.62	A <sub>132</sub>	[6.7, 6.85]	6.78
A <sub>21</sub>	[-7.09, -6.83]	-6.96	A <sub>77</sub>	[-0.57, -0.48]	-0.52	A <sub>133</sub>	[6.85, 7]	6.93
A <sub>22</sub>	[-6.83, -6.57]	-6.7	A <sub>78</sub>	[-0.48, -0.38]	-0.43	A <sub>134</sub>	[7, 7.24]	7.12
A <sub>23</sub>	[-6.57, -6.3]	-6.43	A <sub>79</sub>	[-0.38, -0.29]	-0.34	A <sub>135</sub>	[7.24, 7.48]	7.36
A <sub>24</sub>	[-6.3, -6.04]	-6.17	A <sub>80</sub>	[-0.29, -0.2]	-0.25	A <sub>136</sub>	[7.48, 7.72]	7.6
A <sub>25</sub>	[-6.04, -5.78]	-5.91	A <sub>81</sub>	[-0.2, -0.11]	-0.15	A <sub>137</sub>	[7.72, 7.96]	7.84
A <sub>26</sub>	[-5.78, -5.52]	-5.65	A <sub>82</sub>	[-0.11, -0.02]	-0.06	A <sub>138</sub>	[7.96, 8.2]	8.08
A <sub>27</sub>	[-5.52, -5.26]	-5.39	A <sub>83</sub>	[-0.02, 0.08]	0.03	A <sub>139</sub>	[8.2, 8.44]	8.32
A <sub>28</sub>	[-5.26, -5]	-5.13	A <sub>84</sub>	[0.08, 0.17]	0.12	A <sub>140</sub>	[8.44, 8.68]	8.56
A <sub>29</sub>	[-5, -4.91]	-4.95	A <sub>85</sub>	[0.17, 0.26]	0.22	A <sub>141</sub>	[8.68, 8.92]	8.8
A <sub>30</sub>	[-4.91, -4.82]	-4.86	A <sub>86</sub>	[0.26, 0.35]	0.31	A <sub>142</sub>	[8.92, 9.16]	9.04
A <sub>31</sub>	[-4.82, -4.72]	-4.77	A <sub>87</sub>	[0.35, 0.45]	0.4	A <sub>143</sub>	[9.16, 9.4]	9.28
A <sub>32</sub>	[-4.72, -4.63]	-4.68	A <sub>88</sub>	[0.45, 0.54]	0.49	A <sub>144</sub>	[9.4, 9.64]	9.52
A <sub>33</sub>	[-4.63, -4.54]	-4.58	A <sub>89</sub>	[0.54, 0.63]	0.58	A <sub>145</sub>	[9.64, 9.88]	9.76
A <sub>34</sub>	[-4.54, -4.45]	-4.49	A <sub>90</sub>	[0.63, 0.72]	0.68	A <sub>146</sub>	[9.88, 10.12]	10
A <sub>35</sub>	[-4.45, -4.35]	-4.4	A <sub>91</sub>	[0.72, 0.82]	0.77	A <sub>147</sub>	[10.12, 10.36]	10.24
A <sub>36</sub>	[-4.35, -4.26]	-4.31	A <sub>92</sub>	[0.82, 0.91]	0.86	A <sub>148</sub>	[10.36, 10.6]	10.48
A <sub>37</sub>	[-4.26, -4.17]	-4.22	A <sub>93</sub>	[0.91, 1]	0.95	A <sub>149</sub>	[10.6, 10.84]	10.72
A <sub>38</sub>	[-4.17, -4.08]	-4.12	A <sub>94</sub>	[1, 1.15]	1.08	A <sub>150</sub>	[10.84, 11.08]	10.96
A <sub>39</sub>	[-4.08, -3.98]	-4.03	A <sub>95</sub>	[1.15, 1.3]	1.23	A <sub>151</sub>	[11.08, 11.32]	11.2
A <sub>40</sub>	[-3.98, -3.89]	-3.94	A <sub>96</sub>	[1.3, 1.45]	1.38	A <sub>152</sub>	[11.32, 11.56]	11.44
A <sub>41</sub>	[-3.89, -3.8]	-3.85	A <sub>97</sub>	[1.45, 1.6]	1.53	A <sub>153</sub>	[11.56, 11.8]	11.68
A <sub>42</sub>	[-3.8, -3.71]	-3.75	A <sub>98</sub>	[1.6, 1.75]	1.68	A <sub>154</sub>	[11.8, 12.04]	11.92
A <sub>43</sub>	[-3.71, -3.62]	-3.66	A <sub>99</sub>	[1.75, 1.9]	1.83	A <sub>155</sub>	[12.04, 12.28]	12.16
A <sub>44</sub>	[-3.62, -3.52]	-3.57	A <sub>100</sub>	[1.9, 2.05]	1.98	A <sub>156</sub>	[12.28, 12.52]	12.4
A <sub>45</sub>	[-3.52, -3.43]	-3.48	A <sub>101</sub>	[2.05, 2.2]	2.13	A <sub>157</sub>	[12.52, 12.76]	12.64
A <sub>46</sub>	[-3.43, -3.34]	-3.38	A <sub>102</sub>	[2.2, 2.35]	2.28	A <sub>158</sub>	[12.76, 13]	12.88
A <sub>47</sub>	[-3.34, -3.25]	-3.29	A <sub>103</sub>	[2.35, 2.5]	2.43	A <sub>159</sub>	[13, 14]	13.5
A <sub>48</sub>	[-3.25, -3.15]	-3.2	A <sub>104</sub>	[2.5, 2.65]	2.58	A <sub>160</sub>	[14, 15]	14.5
A <sub>49</sub>	[-3.15, -3.06]	-3.11	A <sub>105</sub>	[2.65, 2.8]	2.73	A <sub>161</sub>	[15, 16]	15.5
A <sub>50</sub>	[-3.06, -2.97]	-3.02	A <sub>106</sub>	[2.8, 2.95]	2.88	A <sub>162</sub>	[16, 17]	16.5
A <sub>51</sub>	[-2.97, -2.88]	-2.92	A <sub>107</sub>	[2.95, 3.1]	3.03	A <sub>163</sub>	[17, 18]	17.5
A <sub>52</sub>	[-2.88, -2.78]	-2.92	A <sub>108</sub>	[3.1, 3.25]	3.18	A <sub>164</sub>	[18, 19]	18.5
A <sub>53</sub>	[-2.78, -2.69]	-2.83	A <sub>109</sub>	[3.25, 3.4]	3.33	A <sub>165</sub>	[19, 21]	20
A <sub>54</sub>	[-2.69, -2.6]	-2.74	A <sub>110</sub>	[3.4, 3.55]	3.48	A <sub>166</sub>	[21, 23]	22
A <sub>55</sub>	[-2.6, -2.51]	-2.65	A <sub>111</sub>	[3.55, 3.7]	3.63	A <sub>167</sub>	[23, 25]	24
A <sub>56</sub>	[-2.51, -2.42]	-2.46	A <sub>112</sub>	[3.7, 3.85]	3.78			





## REFERENCES

- [1] M. Menéndez, C. Martínez, G. Sanz, and J. M. Benitez, "Development of a Smart Framework Based on Knowledge to Support Infrastructure Maintenance Decisions in Railway Corridors," *Transportation Research Procedia*, vol. 14, pp. 1987-1995, 2016, doi: 10.1016/j.trpro.2016.05.166.
- [2] Y.-S. Zhang, E.-J. Yao, and X. Sun, "Impact analysis of HSR fare discount strategy on HSR share in Beijing-Shanghai transportation corridor," *Transportation Letters*, vol. 9, no. 4, 215-227, 2016, doi: 10.1080/19427867.2016.1253635.
- [3] T.-H. Tsai, "Self-evolutionary sibling models to forecast railway arrivals using reservation data," *Engineering Applications of Artificial Intelligence*, vol. 96, Nov. 2020, doi: 10.1016/j.engappai.2020.103960.
- [4] D. Wu, H. Song, and S. Shen, "New developments in tourism and hotel demand modeling and forecasting," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 1, pp. 507-529, Jan. 2017, doi: 10.1108/IJCHM-05-2015-0249.
- [5] T. Webb, Z. Schwartz, Z. Xiang, and N. Singal, "Revenue management forecasting: The resiliency of advanced booking methods given dynamic booking windows," *International Journal of Hospitality Management*, vol. 89, August 2020, doi: 10.1016/j.ijhm.2020.102590.
- [6] Y. Wang, Z. Lin, L. Wang, H. Wang, and J. Zhang, "Prediction of Railway Passenger Volume including Waiting Passengers without Tickets," *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2020, pp. 712-716, doi: 10.1109/ITAIC49862.2020.9338840.
- [7] W. q. Tian, P. Zhao, and K. Qiao, "Railway Freight Volume Forecast Based on GRA-WD-WNN," *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*, 2019, pp. 84-88, doi: 10.1109/ICITE.2019.8880238.
- [8] X. Fuquan and Z. Yi, "Research on Forecast Model and Algorithm of Train Passenger Sending Volume," *2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 2019, pp. 807-811, doi: 10.1109/ICMTMA.2019.00182.
- [9] T. Widiyaningtyas, Muladi and A. Qonita, "Use of ARIMA Method To Predict The Number of Train Passenger In Malang City," *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*, 2019, pp. 359-364, doi: 10.1109/ICAIIIT.2019.8834663.
- [10] J. Qin, W. Qu, X. Wu, and Y. Zeng, "Differential Pricing Strategies of High Speed Railway Based on Prospect Theory: An Empirical Study from China," *Sustainability*, vol. 11, no. 14, pp. 1-17, July 2019, doi: 10.3390/su11143804.
- [11] W. Huang and B. Shuai, "A Methodology for Calculating the Passenger Comfort Benefits of Railway Travel," *Journal of Modern Transportation*, vol. 26, no. 2, pp. 107-118, Feb. 2018, doi: 10.1007/s40534-018-0157-y.
- [12] T. Balabanov, I. Zankinski, and N. Dobrinkova, "Time series prediction by artificial neural networks and differential evolution in distributed environment," *In International Conference on Large-Scale Scientific Computing*, Springer, Berlin, Heidelberg, June 2011, pp. 198-205, doi: 10.1007/978-3-642-29843-1\_22.







- [13] Q. Song and B. S. Chissom, "Fuzzy Time Series and its Models," *Fuzzy Sets and Systems*, vol. 54, no. 3, pp. 269-277, March 1993, doi: 10.1016/0165-0114(93)90372-O.
- [14] Q. Song and B. S. Chissom, "Forecasting Enrollments with Fuzzy Time Series: Part II," *Fuzzy Sets and Systems*, vol. 62, no. 1, pp. 1-8, Feb. 1994, pp. 1-8, doi: 10.1016/0165-0114(94)90067-1.
- [15] Q. Song, "A Note on Fuzzy Time Series Model Selection with Sample Autocorrelation Functions," *Cybernetics and Systems*, vol. 34, no. 2, pp. 93-107, 2003, doi: 10.1080/01969720302867.
- [16] S. M. Chen, "Forecasting Enrollments based on High Order Fuzzy Time Series," *Cybernetics and Systems*, vol. 33, no. 1, pp. 1-16, 2002, doi: 10.1080/019697202753306479.
- [17] K. Huang, "Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series," *Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 387-394, Nov. 2001, doi: 10.1016/S0165-0114(00)00057-9.
- [18] S. M. Chen and K. Tanuwijaya, "Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10594-10605, 2011, doi: 10.1016/j.eswa.2011.02.098.
- [19] B. Garg, M. M. S. Beg and A. Q. Ansari, "Employing genetic algorithm to optimize OWA-fuzzy forecasting model," *2011 Third World Congress on Nature and Biologically Inspired Computing*, 2011, pp. 285-290, doi: 10.1109/NaBIC.2011.6089609.
- [20] H.-K. Yu, "Weighted Fuzzy Time Series Models for TAIEX Forecasting," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3-4, pp. 609-624, April 2005, doi: 10.1016/j.physa.2004.11.006.
- [21] C.-H. Cheng, J.-W. Wang, and C.-H. Li, "Forecasting the Number of Outpatient Visits using a New Fuzzy Time Series based on Weighted-Transitional Matrix," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2568-2575, May 2008, doi: 10.1016/j.eswa.2007.04.007.
- [22] J. Chang and Y. Huang, "A weighted fuzzy time series model based on adoptive OWA operators," *2011 International Conference on Uncertainty Reasoning and Knowledge Engineering*, 2011, pp. 94-97, doi: 10.1109/URKE.2011.6007849.
- [23] P. Jiang, H. Yang, R. Li, and C. Li, "Inbound tourism demand forecasting framework based on fuzzy time series and advanced optimization algorithm," *Applied Soft Computing*, vol. 92, July 2020, doi: 10.1016/j.asoc.2020.106320.
- [24] S. Sun, Y. Wei, K.-L. Tsui, and S. Wang, "Forecasting tourist arrivals with machine learning and internet search index," *Tourism Management*, vol. 70, pp. 1-10, Feb. 2019, doi: 10.1016/j.tourman.2018.07.010.
- [25] O. C. Yolcu and F. Alpaslan, "Prediction of TAIEX based on hybrid fuzzy time series model with single optimization process," *Applied Soft Computing*, vol. 66, pp. 18-33, May 2018, doi: 10.1016/j.asoc.2018.02.007.
- [26] P. C. de Lima e Silva, C. A. S. Junior, M. A. Alves, R. Silvae, M. W. Cohen, and F. G. Guimarães, "Forecasting in non-stationary environments with fuzzy time series," *Applied Soft Computing*, vol. 97, part b, pp. 1-22, Dec. 2020, doi: 10.1016/j.asoc.2020.106825.
- [27] A. Rubio, J. D. Bermúdez, and E. Vercher, "Improving stock index forecasts by using a new weighted fuzzy-trend time series method," *Expert Systems with Applications*, vol. 76, pp. 12-20, Jun. 2017, doi: 10.1016/j.eswa.2017.01.049.
- [28] P. Singh, "An efficient method for forecasting using fuzzy time series," *In Emerging research on applied fuzzy sets and intuitionistic fuzzy matrices*, IGI Global, 2017, pp. 287-304, doi: 10.4018/978-1-5225-0914-1.ch013.
- [29] S. M. Chen and B. D. H. Phuong, "Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors," *Knowledge-Based Systems*, vol. 118, pp. 204-216, 2017, doi: 10.1016/j.knsys.2016.11.019.
- [30] R. G. del Campo, L. Garmendia, J. Recasens and J. Montero, "Hesitant fuzzy sets and relations using lists," *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2017, pp. 1-6, doi: 10.1109/FUZZ-IEEE.2017.8015516.
- [31] S.-H. Cheng, S.-M. Chen, and W.-S. Jian, "Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures," *Information Sciences*, vol. 327, pp. 272-287, January 2016, doi: 10.1016/j.ins.2015.08.024.
- [32] B. Garg, M. M. S. Beg and A. Q. Ansari, "A new computational fuzzy time series model to forecast number of outpatient visits," *2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, 2012, pp. 1-6, doi: 10.1109/NAFIPS.2012.6290977.
- [33] S. Solikhin, S. Lutfi, P. Purnomo, and H. Hardiwinoto, "Prediction of passenger train using fuzzy time series and percentage change methods," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3007-3018, December 2021, doi: 10.11591/eei.v10i6.2822.
- [34] T. A. Jilani, S. M. A. Burney, and C. Ardil, "Multivariate High Order Fuzzy Time Series Forecasting for Car Road Accidents," *World Academy of Science, Engineering and Technology, Open Science Index 18, International Journal of Computer and Information Engineering*, vol. 2 no. 6, pp. 2038-2043, 2008.
- [35] T. A. Jilani, S. M. A. Burney, and C. Ardil, "Fuzzy Metric Approach for Fuzzy Time Series Forecasting Based on Frequency Density Based Partitioning," *World Academy of Science, Engineering and Technology, Open Science Index 43, International Journal of Computer and Information Engineering*, vol. 4, no. 7, pp. 1194-1199, 2010.
- [36] M. G. Chung and S. K. Kim, "Efficient Jitter Compensation Using Double Exponential Smoothing," *Information Sciences*, vol. 227, pp. 83-89, April 2013, doi: 10.1016/j.ins.2012.12.008.
- [37] L. Wu, S. Liu, and Y. Yang, "Grey Double Exponential Smoothing Model and Its Application on Pigprice Forecasting in China," *Applied Soft Computing*, vol. 39, pp. 117-123, February 2016, doi: 10.1016/j.asoc.2015.09.054.
- [38] D. Barrow, N. Kourentzes, R. Sandberg, and J. Niklewski, "Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning," *Expert Systems with Applications*, vol. 160, pp. 1-14, December 2020, doi: 10.1016/j.eswa.2020.113637.
- [39] C. C. Holt, "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages," *International Journal of Forecasting*, vol. 20, no. 1, pp. 5-10, January-March 2004, doi: 10.1016/j.ijforecast.2003.09.015.
- [40] H. A. Sturges, "The Choice of a Class Interval," *Journal of the American Statistical Association*, vol. 21, no. 153, pp. 65-66, 1926, doi: 10.1080/01621459.1926.10502161.





**BIOGRAPHIES OF AUTHORS**

**Solikhin**     is employed at an educational institution, STMIK Himsya, Semarang, Indonesia. He holds the position of Senior Lecturer in the school's Department of Informatics Engineering. He attended AKI University for his undergraduate studies and Diponegoro University for his graduate work. Both institutions awarded him bachelor's and master's degrees. His areas of expertise and areas of research interest include computer science, business intelligence, decision support systems, and information systems. He has also explored research interests in these areas. He can be contacted at email: iingshalihin@gmail.com.







**Septia Lutfi**     is employed at an educational institution, STMIK Himsya, Semarang, Indonesia. He holds the position of Senior Lecturer in the school's Department of Informatics Engineering. He attended AKI University for his undergraduate studies and Dian Nuswantoro University for his graduate work. Both institutions awarded him bachelor's and master's degrees. His areas of expertise and areas of research interest include computer science, computer engineering, and informatics, specifically in the areas of software engineering, programming, and data engineering. He can be contacted at email: septialutfi@stmik-himsya.ac.id.



**Purnomo**     is a Professor at the Faculty of Engineering, Universitas Muhammadiyah Semarang, Indonesia. His academic credentials include a bachelor's degree (Diponegoro University), a master's degree (Gadjah Mada University), and a doctorate degree (Brawijaya University). He specializes in and does most of his research in mechanical engineering and technology, applied engineering, engineering science and technology, and materials science and engineering. He can be contacted at email: purnomo@unimus.ac.id.



**Hardiwinoto**     is a Senior Lecturer at the Faculty of Economics, Universitas Muhammadiyah, Semarang, Indonesia. Diponegoro University awarded him his bachelor's, master's, and doctorate degrees. All of his professional and academic interests are related to economics, such as modern business, economics, finance, and accounting. He can be contacted at email: hardiwinoto@unimus.ac.id.