

Hybrid TBATS-Extreme Learning Machine Model for Forecasting the Number of Train Passengers in Java Excluding Jabodetabek

Salsa Agung Oktariana^{1*}, Achmad Efendi¹, Rahma Fitriani¹

¹Department Statistics, Brawijaya University, Indonesia

salsaoktariana@sudent.ub.ac.id

ABSTRACT

Article History:

Received : 27-04-2026

Revised : 18-05-2026

Accepted : 05-06-2026

Online : 01-10-2026

Keywords:

TBATS;

Extreme Learning

Machine;

Hybrid Model;

Non-Negative Least

Square.



Railway transportation is one of the most important modes of transport in Java, yet the COVID-19 pandemic caused major disruptions to passenger patterns that made the forecasting process considerably more challenging. Previous research has compared TBATS and Extreme Learning Machine (ELM) models separately, but no one has tried to combine the two in one hybrid model. This research fills this gap by building a hybrid TBATS-ELM model using Non-Negative Least Squares (NNLS) to determine the best combination weights between the two models. The data used consists of monthly railway passenger data in Java excluding Jabodetabek from January 2006 to December 2025, comprising 240 observations divided into training, validation, and testing sets. The combination weights were estimated on the validation data using NNLS to produce an optimal and interpretable combination. Test results show that the hybrid model reduced forecast error by 74.27% compared to the standalone TBATS model, with a final MAPE of 11.47% classified as accurate. These findings suggest that combining a seasonal statistical model with a machine learning approach can offer a more reliable solution when data experiences extreme structural disruptions such as those caused by the pandemic.



Crossref

<https://doi.org/10.31764/jtam.v10i4.39382>



This is an open access article under the [CC-BY-SA](#) license

A. INTRODUCTION

Trains are becoming an increasingly important mode of transportation in supporting the mobility of modern society because of their efficiency, capacity and punctuality (S. Zhao & Mi, 2019). The rail transportation sector plays a strategic role in the community mobility system on the island of Java, especially in corridors that connect large cities exclude the Jabodetabek area. This transportation serves medium and long distance trips that can compete with other modes of transportation. The ability to accurately predict the number of passengers is a fundamental requirement for transportation managers to support operational decision making, such as adding train sets, setting schedules and allocating resources (Nar & Arslankaya, 2022). Based on data from Badan Pusat Statistik (BPS) show that Java excluding Jabodetabek train passenger volume experienced a significant recovery after being impacted by the COVID-19 pandemic in 2020 until 2021, even recording an all time high record at the end of 2025 (Badan Pusat Statistik, 2025).

The ability to accurately predict the number of passengers is a fundamental requirement for transportation managers to support operational decision making, such as adding trains, setting schedules and allocating resources. However, train passenger data has quite complex characteristics, including a strong annual seasonal pattern with spikes in the holiday months (Eid and year end), non linear long term growth trends, as well as structural disruption due to the pandemic which resulted in a drastic decline in 2020 and gradual recovery after 2021 (Svetunkov & Boylan, 2020). Disturbances like this have been shown to have a serious impact on the performance of forecasting models based on time series data, especially in models that assume stability of data patterns over time (Hightower et al., 2024a). Conventional models such as ARIMA which have been widely used in transportation forecasting are often unable to capture non linear patterns and sudden structural changes well, so their accuracy decreases when applied to post-pandemic data (Lim & Zohren, 2021).

The TBATS (Trigonometric Seasonality, Box-Cox Transformation, ARMA Errors, Trend, and Seasonal Components) model introduced by Livera et al. (2011) is a state-space framework specifically designed to handle complex seasonal patterns through trigonometric Fourier representation (Perone, 2022). On the other hand, the Extreme Learning Machine (ELM) proposed by Huang et al. (2006) is a single layer artificial neural network (Single Hidden Layer Feedforward Network, SLFN) with randomly assigned input weights and analytically estimated output weights. ELM is known to have high training speed as well as good ability to capture non linear relationships in data (Sattar et al., 2019).

Zhao & Zhang (2022) conducted a study that directly compared the TBATS and ELM models in the context of time series forecasting, and found that both models showed different performance depending on the characteristics of the data they faced. However, this research only places TBATS and ELM as alternative models to be compared, not as combined models. Whereas, Kourentzes et al. (2019) show that combining multiple forecasting models generally leads to more stable and accurate results than relying on a single best model. Furthermore, Hightower et al. (2024) found that TBATS performance experienced a significant decline when applied to post-pandemic transportation data, because the structural changes that occurred could not be captured by the trend component of the state-space-based model. In another hand, Chen et al. (2022) shows that ELM applied in a hybrid framework is able to significantly improve forecasting accuracy compared to a single model, especially on data that contains non-linear patterns. This condition creates an unanswered research gap regarding how to combine the strengths of TBATS in capturing complex seasonal patterns with the adaptive capabilities of ELM in studying post-pandemic recovery patterns, so that the weaknesses of each model can be mutually compensated for.

This research uses a parallel hybrid framework, where TBATS and ELM produce predictions independently, then combine them using the Non-Negative Least Square (NNLS) method as developed by Lawson & Hanson (1955). The selection of NNLS as a weighting method is based on several considerations. Compared to regular OLS which can produce negative weights and is difficult to interpret practically, NNLS ensures that all weights are non-negative so that each model only contributes positively to the final prediction result (Clements, 2017). In contrast to equal weighting which treats all models with the same contribution without considering their individual performance, NNLS automatically suppresses the weight of lower performing models close to zero without requiring additional parameter tuning (Roccazzella et al., 2022). Segura-

Heras et al. (2022) also show that weighting strategies based on optimized validation data consistently produce better combination accuracy than weighting that does not consider performance on data the model has never seen. These advantages make NNLS a more stable and interpretable choice, especially when one of the models suffers from systematic bias due to structural distortions in the training data as occurred during the COVID-19 pandemic period in this research.

This research aims to estimate the TBATS model on training data, build an ELM model by selecting optimal hyperparameters via grid search, estimate hybrid combination weights via NNLS on validation data, and evaluate the performance of the three models on test data using MAPE, RMSE, and MAE metrics. The main contribution of the research lies in the application of the hybrid TBATS-ELM approach with NNLS-based weight estimation on data experiencing post pandemic structural distortions.

B. METHODS

This research is applied quantitative research using secondary data, where the main focus is the development and evaluation of the TBATS-ELM hybrid forecasting model based on historical train passenger data.

1. Data Sources and Sharing

The data used is monthly data on the number of Java train passengers excluding Jabodetabek which is sourced from *Badan Pusat Statistik* (BPS). Data coverage includes 240 monthly observations from January 2006 to December 2025. In accordance with the hybrid design used, the data is divided into three partitions as shown in Table 1.

Table 1. Data Partitioning

Section	Period	Number of Observations
Training Data	January 2006 – December 2021	192
Validation Data	January 2022 – December 2022	12
Testing Data	January 2023 – December 2025	36
Total	January 2006 – December 2025	240

This division follows the principle that evaluation data must be out-of-sample and never used in the training or model adjustment process. The following shows the seasonal pattern of train passenger numbers in Java (excluding Jabodetabek) (January 2006 – December 2025) and the distribution of monthly values, as shown in Figure 1.

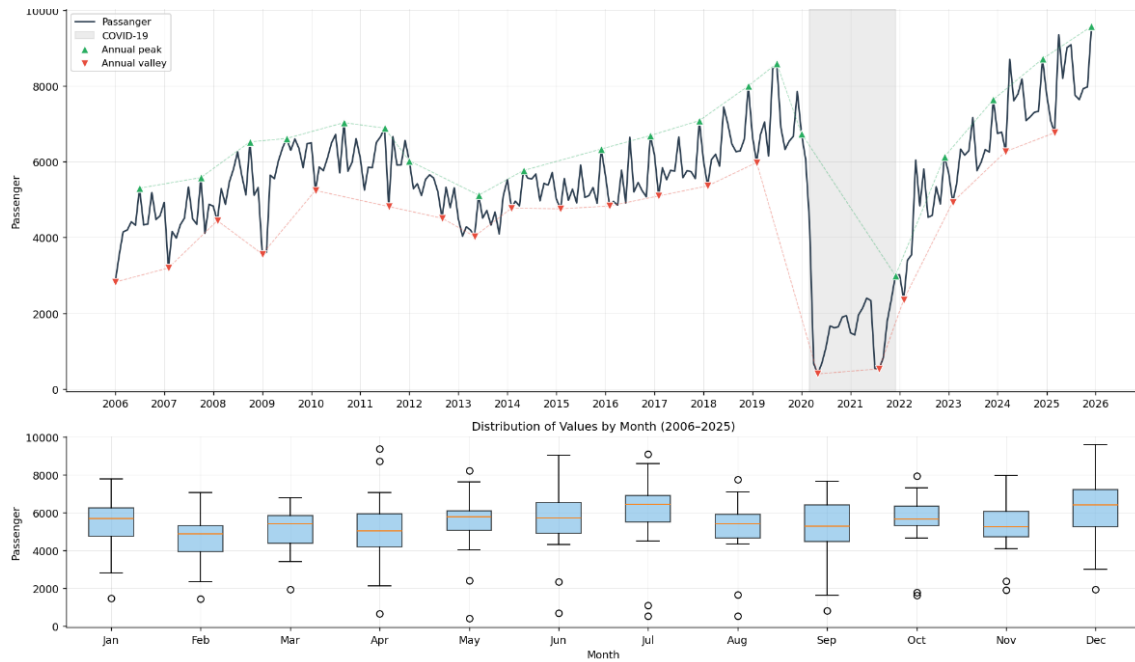


Figure 1. Seasonal Pattern of the Number of Railway Passengers in Java Exclude Jabodetabek (January 2006 – December 2025) and Monthly Value Distribution

Figure 1 shows the overall pattern of data throughout 2006 until 2025 along with the distribution of values per month. Visually, there is a consistent long-term growth trend before the pandemic, with seasonal fluctuations recurring every year. The highest value spikes consistently occur in July and December, in line with the Eid period and year-end holidays. In contrast, February routinely records the lowest values of the year. The pandemic disruption in 2020 appears as a deep dip on the charts, followed by a gentle recovery through 2022 and a sharp acceleration in 2023–2025 that exceeds all pre-pandemic levels, as shown in Figure 2.

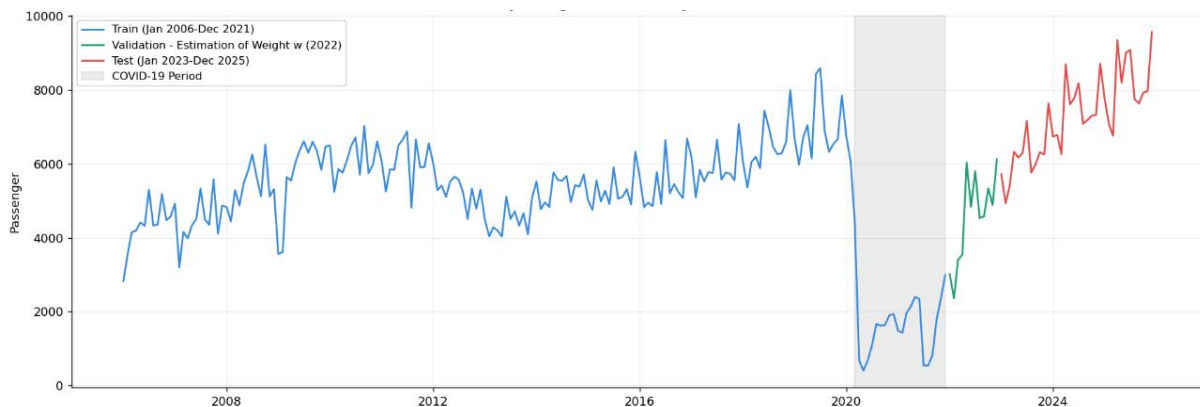


Figure 2. Data Partition Structure: Training (January 2006–December 2021), Validation (2022), and Testing (2023–2025)

Figure 2 displays the chronological data division structure for the purposes of hybrid TBATS-ELM modeling. The blue color shows the training data (1 January 2006 until 31 December 2021, 192 observations) used for TBATS parameter estimation and ELM training. The green segment is validation data (January until December 2022, 12 observations) which is used specifically to estimate NNLS weights. The red segment is test data (January 2023 until

December 2025, 36 observations) which does not touch the training or calibration process at all, used purely for the final out-of-sample evaluation. Note that all training data covers the COVID-19 pandemic period marked in gray, which is the main source of distortion in the TBATS trend component.

The period 2023 to 2025 was chosen as test data because this period represents the recovery phase after the COVID-19 pandemic, namely conditions where passenger patterns are starting to grow again but with different characteristics from the pre-pandemic period. Testing the model in this period provides a more realistic picture of the model's ability to deal with data that has undergone structural changes, so that the validity of the resulting forecast conditions is stronger and more relevant to the needs for transportation operational decision making in the recovery period.

2. TBATS MODEL

TBATS is a state space forecasting framework developed to handle data with complex seasonal patterns (Munim, 2022). The general model of TBATS is expressed as:

$$y_t(\omega) = \ell_{t-1} + \phi b_{t-1} + \sum_j s_{jt} - m_j + d_t \quad (1)$$

where $y_t(\omega)$ is the data value after Box-Cox transformation with parameter ω , ℓ_t is the level component, b_t is the trend component, ϕ is the trend damping coefficient ($0 < \phi \leq 1$), s_{jt} is the j -th seasonal component, and d_t is the error component that follows the ARMA(p, q) process.

The seasonal components are represented using Fourier harmonic pairs (2), parameter updating is performed through the Kalman Filter algorithm (3), and model selection is based on the Akaike Information Criterion (4).

$$s_{jt} = \sum_{k=1}^K k_j \left[\alpha_{jk} \cos\left(\frac{2\pi kt}{m_j}\right) + b_{jk} \sin\left(\frac{2\pi kt}{m_j}\right) \right] \quad (2)$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha e_t, \quad b_t = \phi b_{t-1} + \beta e_t, \quad \gamma_{jk} \text{ updated at each step} \quad (3)$$

$$AIC = -2 \ln L(\theta) + 2k \quad (4)$$

3. EXTREME LEARNING MACHINE (ELM) MODEL

ELM is an artificial neural network with one layer (Single Hidden Layer Feedforward Network, SLFN) which estimates output weights analytically via the Moore-Penrose pseudoinverse matrix (Huang et al., 2006). For N training samples $\{(x_i, y_i)\}_{i=1}^N$, the ELM model with L hidden neurons is defined as (de Mattos Neto et al., 2021):

$$\hat{y}_i = \sum_{j=1}^L \beta_j g(w_j \cdot x_i + b_j), \quad i = 1, 2, \dots, N \quad (5)$$

where $w_j \in \mathbb{R}^n$ is the input weight vector to the j -th hidden neuron, b_j is the bias, $g(\cdot)$ is the activation function, and β_j is the output weight. The sigmoid activation function used in this research is:

$$g(x) = \frac{1}{(1+e^{-x})} \quad (6)$$

The input weights w_j and bias b_j are randomly assigned with a uniform distribution on $[-1, 1]$ and are not updated during training. In matrix form, the ELM model can be written as:

$$H\beta = Y \quad (7)$$

where $H \in \mathbb{R}^{N \times L}$ is a hidden layer matrix with elements $H_{ij} = g(w_j \cdot x_i + b_j)$, $\beta \in \mathbb{R}^L$ is the output weight vector, and $Y \in \mathbb{R}^N$ is the target vector. Estimated output weights are obtained through:

$$\hat{\beta} = H^+Y = (H^T H)^{-1} H^T Y \quad (8)$$

where H^+ is the Moore-Penrose pseudoinverse matrix of H . Before training, the data is normalized using Min-Max Scaling:

$$x'_i = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (9)$$

with x_{min} and x_{max} calculated exclusively from training data to prevent information leakage (data leakage). In this research, $x_{min} = 399$ and $x_{max} = 9589$ (in thousands of passengers). The training data is then converted to supervised learning format using a lag window approach with lag length ℓ , so that the input vector for the t -th observation is $x_t = y_{t-1}, y_{t-2}, \dots, y_{t-\ell}$.

Optimal hyperparameter selection was carried out via grid search on validation data with a total of 180 combinations, including the number of hidden neurons $\{5, 10, 20, 30, 50\}$, lag $\{3, 6, 12\}$, activation function $\{\text{sigmoid}, \text{relu}, \text{tanh}\}$, and seed $\{7, 42, 99, 123\}$. The best model is selected based on the smallest MAPE value on the validation data, not on the training data, so that the selection process is completely out of sample and is not affected by the fitting performance on the data used to train the model.

This approach also functions as an overfitting control mechanism. Because ELM uses randomly assigned input weights and only trains output weights analytically, the risk of overfitting in ELM is relatively lower than conventional artificial neural networks. However, selecting a design using validation data that is completely separate from the training data ensures that the selected model has good generalization capabilities, rather than just memorizing training data patterns (Cerqueira et al., 2020). The use of several different seed values also aims to ensure the stability of the results against variations in random weight initialization. The following is the raw data before normalization, as shown in Figure 3.

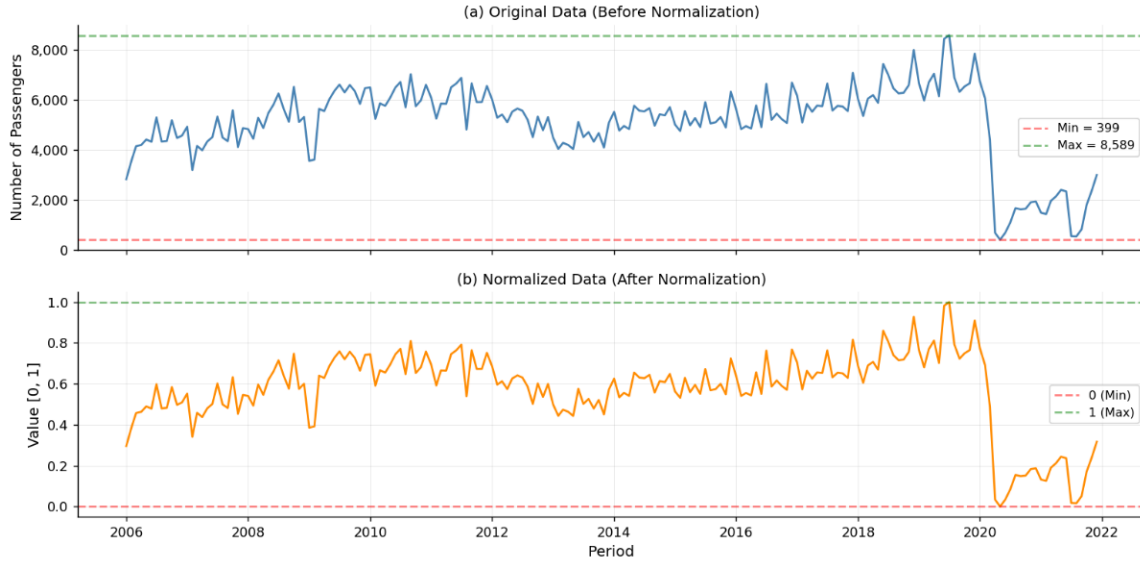


Figure 3. Original Data Before Normalization (top) and Normalized Data in [0, 1] (bottom) with Min = 399, Max = 8,589 thousand passengers

The training data is then converted to supervised learning format using a lag window with lag length ℓ , so that the input for the t -th observation is $x_t = (y_{t-1}, y_{t-2}, \dots, y_{t-\ell})$. Optimal hyperparameter selection is carried out via grid search on validation data (180 combinations): $n_{hidden} \in \{5, 10, 20, 30, 50\}$, $lag \in \{3, 6, 12\}$, $activation \in \{sigmoid, relu, tanh\}$, $seed \in \{7, 42, 99, 123\}$.

4. Estimating Combination Weights Using Non-Negative Least Squares (NNLS)

The TBATS-ELM hybrid model is defined as a weighted linear combination of the forecasts of both models (Clements, 2017):

$$\hat{y}_{hybrid,t} = w \cdot \hat{y}_{TBATS,t} + (1 - w) \cdot \hat{y}_{ELM,t} \quad (10)$$

with constraints $w \in [0, 1]$. The optimal weight w is estimated on validation data using the Non-Negative Least Squares (NNLS) algorithm from (Lawson & Hanson, 1995) which minimizes:

$$\min_w \|y_{val} - [\hat{y}_{TBATS,val} \ \hat{y}_{ELM,val}]w\|_2^2 \quad (11)$$

The NNLS approach was chosen because it produces weights that are interpretatively meaningful, where a value of zero in one of the weights indicates that the model does not contribute optimally to the combination without requiring negative value weights which are practically difficult to interpret. The performance of the combination is evaluated using the coefficient of determination R^2 as follows:

$$R^2 = 1 - \frac{\sum_t y_t - \hat{y}_{hybrid,t}^2}{\sum_t (y_t - \bar{y})^2} \quad (12)$$

5. Accuracy Metric

Model performance was evaluated using three metrics. Mean Absolute Percentage Error (MAPE) is used as the main metric because it is unit free (Houssein et al., 2025):

$$MAPE = \left(\frac{1}{n} \right) \frac{\sum_t |y_t - \hat{y}_t|}{y_t} \times 100\% \quad (13)$$

Root Mean Squared Error (RMSE) is sensitive to large errors:

$$RMSE = \sqrt{\left(\frac{1}{n} \right) \sum_t (y_t - \hat{y}_t)^2} \quad (14)$$

Mean Absolute Error (MAE) is a measure of the average error in the original units of data:

$$MAE = \left(\frac{1}{n} \right) \sum_t |y_t - \hat{y}_t| \quad (15)$$

Accuracy classification based on MAPE values refers to Mahsup et al. (2024), where below 10% is categorized as very accurate, 10–20% accurate, 20–50% worthy, and above 50% not accurate.

C. RESULTS AND DISCUSSION

1. Analysis of Data Characteristics

Railway passenger data in Java excluding Jabodetabek for the period 2006–2025 show a long term upward trend accompanied by a consistent annual seasonal pattern. Based on Figure 1, peak passengers consistently occur in April (Eid) and December (end of year holidays), while the troughs occur in January until February and September until October. The monthly value distribution observed in the boxplot confirms that July and December have the highest medians, while February has the lowest median.

Significant structural disruption was clearly visible in the March 2020 until December 2021 period due to the COVID-19 pandemic, where passenger volume dropped drastically to a low of 399 thousand in April 2020. A gradual recovery occurred starting in early 2022 and continued until it reached a record high of 9,568 thousand passengers in December 2025, this surpasses the peak pre-pandemic level of 8,589 thousand in July 2019.

2. TBATS Model Estimation and Diagnostic Results

The TBATS model selected based on the AIC criterion is $TBATS(-, 0.9431, 0, 0)\{12^5\}$, which indicates: no Box-Cox transformation ($\omega = -$), a damped trend with $\phi = 0.9431$, no ARMA components ($p = q = 0$), and five pairs of Fourier harmonics for the seasonal period $m = 12$. The main parameters of the estimation results are presented in Table 2.

Table 2. Parameter Estimates of the TBATS(-, 0.9431, 0, 0){12⁵} Model

Parameter	Value	Information
Box-Cox Transformation (ω)	-	Not applied
Trend Component	Present	-
Alpha (α)	0.8034	Smoothing level coefficient
Beta (β)	-0.0571	Smoothing trend coefficient
Phi (ϕ)	0.9431	Damped parameter
Gamma 1 (γ_1)	≈ 0	Seasonal smoothing cos
Gamma 2 (γ_2)	≈ 0	Seasonal smoothing sin
ARMA error (p, q)	(0, 0)	There is no ARMA component
Harmonic Fourier (K)	5	For $m = 12$
AIC	3,526.83	Model selection criteria
$\text{Sigma}^2(\sigma^2)$	413,593.00	Innovation variance

The value $\phi = 0.9431$ (< 1) indicates that the model estimates long term growth that is slowly slowing down (damped trend). The relatively high coefficient $\alpha = 0.8034$ indicates that the model assigns substantial weight to recent observations, making it responsive to changes in the data level. Gamma (γ) values close to zero indicate a relatively stable seasonal pattern throughout the sample.

TBATS residual diagnostics are presented in Table 3. The Ljung–Box test at lag 6 ($Q = 4.45$, $p = 0.6163$) and lag 12 ($Q = 13.45$, $p = 0.3374$) fails to reject the null hypothesis of no autocorrelation, whereas at lag 24 the null is rejected ($Q = 37.17$, $p = 0.0421$). The Jarque-Bera test rejects H_0 of residual normality ($JB = 381.86$, $p < 0.001$), and the ARCH–LM test reveals conditional heteroskedasticity at lag 6 ($LM = 16.09$, $p = 0.0133$) and lag 12 ($LM = 21.56$, $p = 0.0428$). These deviations from the residual assumptions, especially normality and homoskedasticity, are consistent with the pandemic period, which led to extreme residual values. This is one of the reasons for combining TBATS with the more adaptive ELM model.

Table 3. Residual Diagnostic Results of the TBATS Model

Test	Lag	Statistics	p-value	Conclusion
Ljung-Box	6	$Q = 4.45$	0.6163	Fail to reject H_0
Ljung-Box	12	$Q = 13.45$	0.3374	Fail to reject H_0
Ljung-Box	24	$Q = 37.17$	0.0421	Reject H_0
Jarque-Bera	—	$JB = 381.86$	< 0.001	Reject H_0
ARCH-LM	6	$LM = 16.09$	0.0133	Reject H_0
ARCH-LM	12	$LM = 21.56$	0.0428	Reject H_0

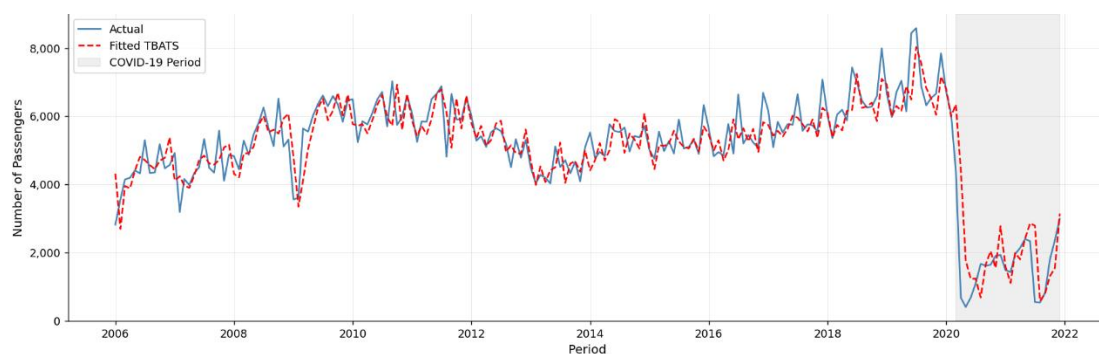
**Figure 4.** Actual vs Fitted TBATS Values on Training Data (January 2006–December 2021) with MAPE = 15.79%

Figure 4 displays a comparison of actual and fitted TBATS values for all training data (January 2006 until December 2021). In general, the fitted line follows the actual fluctuations quite well in the 2006 until 2019 period, with a training MAPE of 15.79%. However, in the 2020 until 2021 pandemic period the TBATS model had difficulty adjusting for the drastic decline and non-linear initial recovery. The fitted patterns tend to be smoother than the actual sharp fluctuations, which is an inherent characteristic of smoothing-based state-space models.

3. Grid Search Results and Optimal ELM Configuration

Grid search of 180 hyperparameter combinations produces an optimal ELM architecture with $n_{hidden} = 5$, lag = 12, sigmoid activation function, and seed = 42, with a validation MAPE of 10.73%. A summary of the 15 best combinations is presented in Table 4.

Table 4. Fifteen Best Combinations of ELM Hyperparameters based on MAPE Validation

Rank	n_{hidden}	Activation	n_{lag}	Seed	MAPE Val (%)
1*	5	sigmoid	12	42	10,73
2	5	sigmoid	12	123	13,64
3	20	tanh	3	42	13,85
4	10	tanh	6	99	13,93
5	20	relu	12	42	14,04
6	50	sigmoid	6	123	14,38
7	30	sigmoid	3	99	14,50
8	5	sigmoid	6	99	14,56
9	5	tanh	6	99	14,61
10	20	relu	6	123	14,64
11	20	tanh	6	42	14,68
12	20	sigmoid	6	42	14,91
13	20	tanh	3	123	14,91
14	20	relu	3	123	14,91
15	5	tanh	12	42	14,95

The best configuration (*) is characterized by a validation MAPE value of 10.73%, significantly superior to the second combination (13.64%). The optimal lag coinciding with the seasonal period of the data (lag = 12) is consistent with the dominant annual seasonal structure of the rail passenger data. The OLS results for the ELM output weights (β) with a 12 input, 5 hidden, and 1 output architecture are presented in Table 5. The R^2 value of 0.6931, along with an F-statistic of 98.79 ($p < 0.001$), indicates that the ELM model is statistically significant overall. All five hidden neurons are significant at the 0.05 level, with the fourth neuron showing the largest contribution ($\beta_4 = -3.2420$, $t = -9.5515$, $p < 0.001$).

Table 5. OLS Statistics ELM Output Weights (β)

Neuron	$\hat{\beta}$	SE($\hat{\beta}$)	t-stat	p-value	Sig.
4	-3.2420	0.3394	-9.5515	< 0.001	***
1	1.1085	0.1620	6.8410	< 0.001	***
5	0.4240	0.0913	4.6469	< 0.001	***
2	1.5324	0.3498	4.3810	< 0.001	***
3	0.2890	0.1177	2.4555	0.0150	*

Notes: *** $p < 0.001$; * $p < 0.05$. $R^2 = 0.6931$; Adj. $R^2 = 0.6842$; $F(5.174) = 98.79$ ($p < 0.001$).

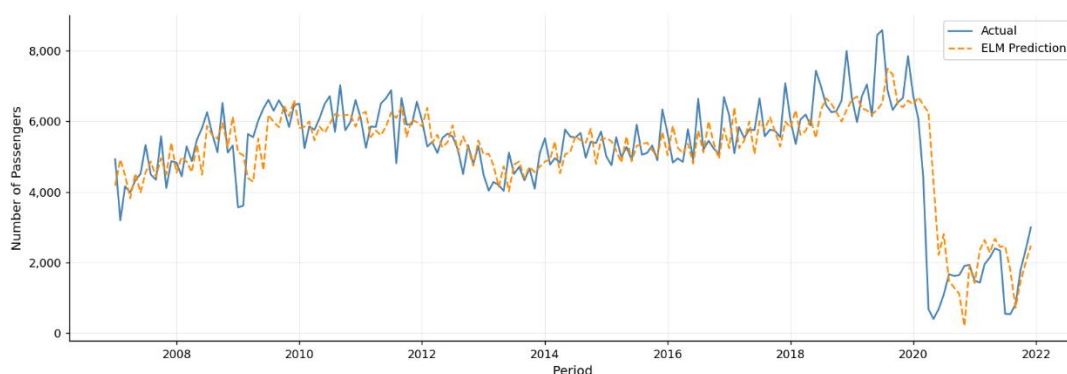


Figure 5. ELM Actual vs Predicted Values on Training Data (January 2006–December 2021) with MAPE = 25.75%

Figure 5 displays a comparison of the actual and predicted ELM values on the training data. In contrast to TBATS, the ELM prediction line shows greater variability and follows short-term fluctuations more aggressively. In the normal period 2006 until 2019, ELM was able to capture seasonal peaks and troughs quite well, although sometimes several months late. In the 2020 until 2021 pandemic period, ELM was more responsive in following sharp declines compared to TBATS, but ELM's trained MAPE (25.75%) was higher than TBATS (15.79%) because ELM learned from noise more on normal data before the pandemic.

4. Hybrid Weight Estimation via NNLS

Estimation of NNLS combination weights on 12 validation data observations (January until December 2022) produces $\hat{w}_{TBATS} = 0$ and $\hat{w}_{ELM} = 1$. Statistically, the value of $t = 0$ with $p = 1.000$ indicates that the TBATS weight is not significantly different from zero. The hybrid R^2 on validation data is 0.6905. The condition $w_{TBATS} = 0$ indicates that the hybrid model formed is identical to the single ELM in the test period, but this does not necessarily reduce the value of the hybrid framework built. NNLS is designed to produce optimal weights based on actual performance on validation data, not based on the assumption that both models must always contribute. When one of the models suffers from severe systematic bias, then pressing the model weight to zero is the most rational decision both mathematically and empirically (Diebold & Shin, 2019). Claeskens et al. (2016) and (Chan & Pauwels, 2018) theoretically assert that the weighted combination should produce accuracy that is no worse than its single best component, but this superiority generally appears when both models contribute positively, but this condition is not met because TBATS suffers from structural bias due to pandemic distortions. Thus, the NNLS-based parallel hybrid actually shows its methodological superiority, when one of the components is uninformative, the weights are filtered automatically without manual intervention.

This result reflects the empirical fact that on validation data, ELM predictions are much closer to actual values than TBATS. MAPE TBATS on validation data reached 32.26% compared to 10.73% for ELM. Therefore, the NNLS algorithm optimally allocates all weights to ELM. This condition can be explained by the impact in the TBATS model caused by the inclusion of the COVID-19 pandemic period in the training data. The sharp decline in 2020 until 2021 affects the estimation of the trend and level components, leading the model to systematically produce forecasts that fall below the actual values during the recovery period. In contrast, ELM with a

lag-window architecture is able to learn recovery patterns more adaptively from recent historical data.

5. Accuracy Evaluation on Test Data

The performance of the three models on test data for the period January 2023 until December 2025 (36 observations) is presented in Table 6.

Table 6. Performance Comparison of TBATS, ELM, and Hybrid TBATS-ELM on Test Data

Metric	TBATS	ELM	Hybrid TBATS-ELM	Improve (%)
MAPE (%)	44.58	11.47	11.47	74.27%
RMSE	3,393.19	1,055.17	1,055.17	68.90%
MAE	3,285.04	866.09	866.09	73.64%
Category *	Worthy	Accurate	Accurate	—

* Classification based on Lewis (1982): MAPE < 10% = Very Accurate;
10–20% = Accurate; 20–50% = Worthy; > 50% = Not accurate.

The evaluation results show that TBATS produces a MAPE of 44.58% (worthy), while ELM and the hybrid model both achieve a MAPE of 11.47% (accurate). The MAPE reduction of 74.27% from TBATS to hybrid is a very significant increase. In terms of RMSE, hybrid is 68.90% ahead of TBATS (1,055.17 vs 3,393.19), and in terms of MAE it is 73.64% ahead (866.09 vs 3,285.04). While it should be noted that overall the hybrid model produces good accuracy, there are several periods where the forecast error is still relatively large, especially in peak months such as April 2024 and December 2025. This indicates that post pandemic growth patterns in the test data exceeded levels anticipated by models trained through December 2021. This exceedance is inherently difficult to predict by any model without recent data reflecting this accelerated recovery, as shown in Figure 6, Figure 7, and Figure 8.

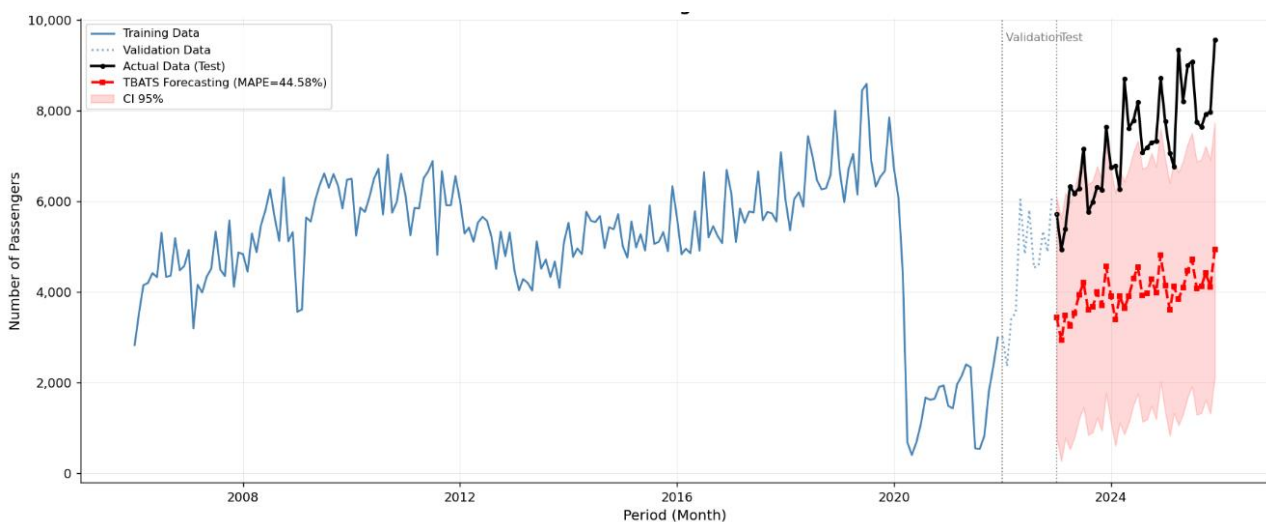


Figure 6. TBATS Forecasting vs Actual Data on Test Data (January 2023–December 2025) with MAPE = 44.58%

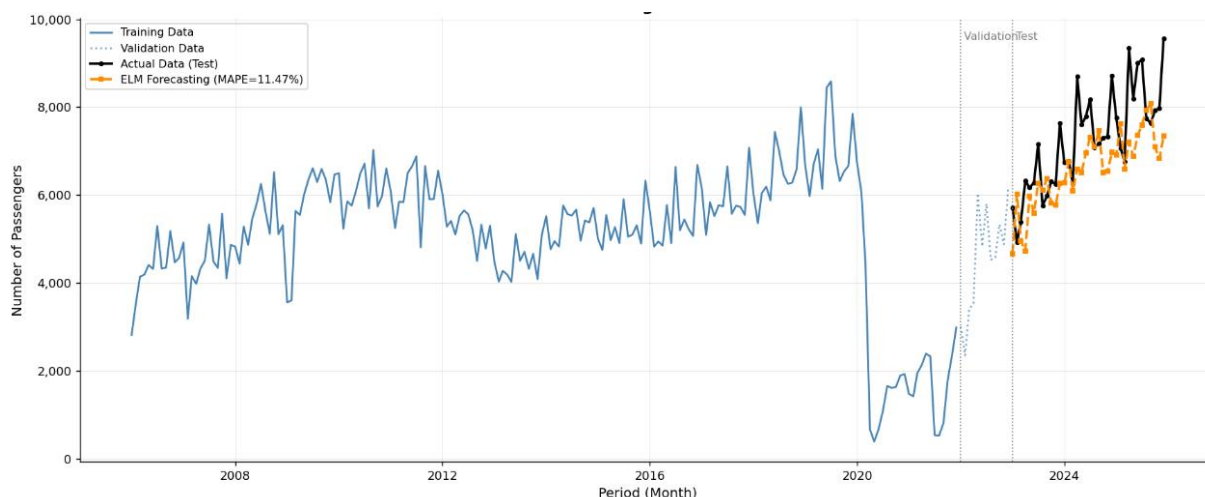


Figure 7. ELM Forecasting vs Actual Data on Test Data with MAPE = 11.47%

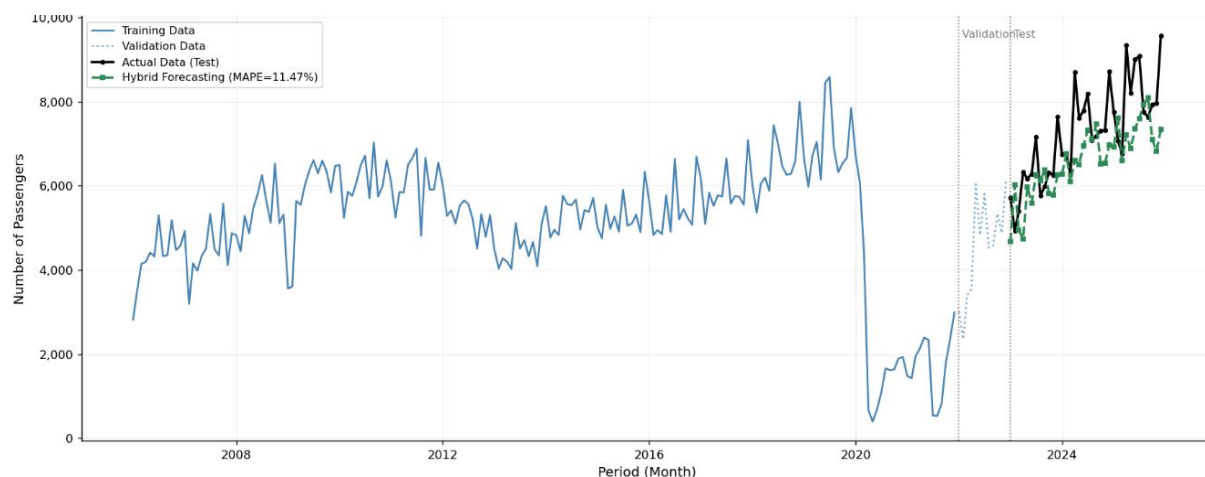


Figure 8. Hybrid TBATS-ELM Forecasting vs Actual Data on Test Data with MAPE = 11.47%

Figure 6 shows that TBATS forecasts are systematically below actual values throughout the entire 36 month test period, with confidence intervals that widen drastically over time and actual values frequently exceeding even the upper bound, confirming the structural nature of the underestimation. In contrast, Figure 7 shows that ELM follows the actual pattern considerably better, though underestimation still occurs at extreme seasonal peaks such as April 2024 and December 2025, where actual values exceeded the historical maximum of the training data. Figure 8 shows that the hybrid forecast is identical to ELM given the NNLS weight $w_{TBATS} = 0$, with errors distributed evenly without a systematic pattern across the test period.

The systematic failure of TBATS in the test period is due to the fundamental limitations of state-space based models when dealing with permanent structural changes. When the training data contains extreme disturbances such as a drastic downturn due to a pandemic, the trend and level components in TBATS are estimated with biases that then affect all future prediction results. This condition is in accordance with research by Hightower et al. (2024) who show that all historically based time series models, including TBATS, experience significant performance degradation on post-pandemic transportation data. This suggests that for data that experiences

permanent structural breaks, smoothing based statistical models require special intervention before estimation, or can be replaced by more adaptive approaches such as machine learning.

The findings in this research show that ELM tends to produce superior performance compared to state space-based statistical models. These results strengthen evidence from previous studies examining transportation time series forecasting. Zhu et al. (2019) showed that neural network-based approaches in a deep learning framework show better performance compared to traditional statistical approaches, especially on data with nonlinear patterns with sudden structural changes. Hightower et al. (2024) comparing seven time series methods, including TBATS on post-covid public transportation ridership data reported that the performance of all models decreased significantly in the post-disruption period compared to the pre-pandemic period, indicating the limitations of TBATS in adapting to sudden changes in structural patterns, similar to the behavior observed in this study. Cerqueira et al. (2020) also showed based on observations that instance-based learning methods that use historical lag patterns such as ELM, maintain more stable accuracy because the data distribution shifts between training and test periods, a condition that directly occurs in the context of the COVID-19 pandemic in this study.

On the other side, the results of NNLS weight estimation that optimizes all contributions to ELM need to be examined again from the perspective of the combined forecasting literature. Claeskens et al. (2016) and Chan & Pauwels (2018) theoretically assert that the weighted combination should produce accuracy that is no worse than the single best component, but that superiority is generally seen when both models contribute positively, this condition is not met because TBATS suffers from systematic bias due to pandemic disruption. This is consistent with the argument Diebold & Shin (2019) that the combination weights obtained through regression with constraints will naturally suppress the weights of the less informative components to close to zero. As for the related limitations of ELM in capturing extreme seasonal peaks of 2024 to 2025, Lim & Zohren (2021) noted that data driven models including ELM are inherently limited in extrapolating beyond the range of their training data, which is consistent with the underestimation observed at extreme seasonal peaks in April 2024 and December 2025 in this study, where actual values exceeded the previously recorded historical maximum.

D. CONCLUSION AND SUGGESTIONS

This research successfully builds a hybrid TBATS-ELM model with NNLS-based weight estimation for forecasting railway passenger volume in Java excluding Jabodetabek. The result show that the hybrid model achieves a test MAPE of 11.47%, reducing error by 74.27% compared to the single TBATS model. The NNLS weight estimation produces $w_{TBATS} = 0$, and $w_{ELM} = 1$, indicating that ELM dominates the combination due to the structural bias accumulated by TBATS during the pandemic period.

This research provides several important contributions, including proving that the NNLS-based parallel hybrid framework is able to adapt automatically when one of the components experiences systematic failure due to structural breaks, shows that NNLS weighting is transparent and can be interpreted directly from validation performance, and represents the first study to combine TBATS and ELM within a hybrid framework for transportation data. The main limitations of this research are the condition $w_{TBATS} = 0$ which causes the full potential

of the combination of the two models to not be realized, as well as the absence of explicit intervention correction for the pandemic period in the estimation process.

Based on the limitations found, several further research directions can be proposed. Applying intervention corrections to pandemic data before TBATS estimation has the potential to restore TBATS performance so that an even hybrid combination in both models can be implemented. Developing ELM with regularization is also recommended to increase the model's robustness to outliers. In addition, comparisons of deep learning models also need to be carried out to evaluate more complex approaches in capturing post-pandemic recovery patterns. The addition of exogenous variables such as the national holiday calendar also has the potential to increase forecasting accuracy in peak months which are still difficult to predict.

ACKNOWLEDGEMENT

The author would like to thank the Department Statistics, Universitas Brawijaya, for the academic support during this study. The author also thanks Statistics Indonesia (BPS) for providing the railway passenger data used in this research, and all those who have helped and supported this work.

REFERENCES

- Badan Pusat Statistik. (2025). *Statistik Transportasi Darat 2024*.
- Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>
- Chan, F., & Pauwels, L. L. (2018). Some Theoretical Results on Forecast Combinations. *International Journal of Forecasting*, 34(1), 64–74. <https://doi.org/https://doi.org/10.1016/j.ijforecast.2017.08.005>
- Chen, P., Vivian, A., & Ye, C. (2022). Forecasting carbon futures price: a hybrid method incorporating fuzzy entropy and extreme learning machine. *Annals of Operations Research*, 313(1), 559–601. <https://doi.org/10.1007/s10479-021-04406-4>
- Claeskens, G., Magnus, J. R., Vasnev, A. L., & Wang, W. (2016). The forecast combination puzzle: A simple theoretical explanation. *International Journal of Forecasting*, 32(3), 754–762. <https://doi.org/https://doi.org/10.1016/j.ijforecast.2015.12.005>
- Clements, M. P. (2017). Sir Clive W.J. Granger's Contributions to Forecasting. *European Journal of Pure and Applied Mathematics*, 10(1), 30–57. <https://doi.org/https://www.ejpam.com/ejpam/article/view/2949>
- de Mattos Neto, P. S. G., de Oliveira, J. F. L., Bassetto, P., Siqueira, H. V., Barbosa, L., Alves, E. P., Marinho, M. H. N., Rissi, G. F., & Li, F. (2021). Energy consumption forecasting for smart meters using extreme learning machine ensemble. *Sensors*, 21(23), 1–18. <https://doi.org/10.3390/s21238096>
- Diebold, F. X., & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, 35(4), 1679–1691. <https://doi.org/https://doi.org/10.1016/j.ijforecast.2018.09.006>
- Hightower, A., Ziedan, A., Guo, J., Zhu, X., & Brakewood, C. (2024). A comparison of time series methods for post-COVID transit ridership forecasting. *Journal of Public Transportation*, 26, 1–15. <https://doi.org/10.1016/j.jpubtr.2024.100097>
- Houssein, E. H., Mohamed, M., Younis, E. M. G., & Mohamed, W. M. (2025). Artificial intelligence and classical statistical models for time series forecasting: a comprehensive review. *Journal of Big Data*, 12(1), 1–38. <https://doi.org/10.1186/s40537-025-01318-z>
- Huang, G. Bin, Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>

- Kourentzes, N., Barrow, D., & Petropoulos, F. (2019). Another look at forecast selection and combination: Evidence from forecast pooling. *International Journal of Production Economics*, 209, 226–235. <https://doi.org/10.1016/j.ijpe.2018.05.019>
- Lawson, C. L., & Hanson, R. J. (1995). *Solving Least Squares Problems* (Vol. 15). Society for Industrial and Applied Mathematics (SIAM). <https://doi.org/10.1137/1.9781611971217>
- Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200209. <https://doi.org/10.1098/rsta.2020.0209>
- Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527. <https://doi.org/10.1198/jasa.2011.tm09771>
- Mahsup, Febriani, P. A., Syaharuddin, Mandailina, V., Abdillah, & Ibrahim. (2024). Accuracy Rate of Least Square Support Vector Machine Method and Its Various Modifications: A Forecasting Evaluation on Multi-Type Data. *Ingenierie Des Systemes d'Information*, 29(3), 1209–1218. <https://doi.org/10.18280/isi.290337>
- Munim, Z. H. (2022). State-space TBATS model for container freight rate forecasting with improved accuracy. *Maritime Transport Research*, 3, 1–14. <https://doi.org/10.1016/j.martra.2022.100057>
- Nar, M., & Arslankaya, S. (2022). Passenger demand forecasting for railway systems. *Open Chemistry*, 20(1), 105–119. <https://doi.org/10.1515/chem-2022-0124>
- Perone, G. (2022). Comparison of ARIMA, ETS, NNAR, TBATS and Hybrid Models to Forecast the Second Wave of COVID-19 Hospitalizations in Italy. *European Journal of Health Economics*, 23(6), 917–940. <https://doi.org/10.1007/s10198-021-01347-4>
- Roccazzella, F., Gambetti, P., & Vrans, F. (2022). Optimal and robust combination of forecasts via constrained optimization and shrinkage. *International Journal of Forecasting*, 38(1), 97–116. <https://doi.org/10.1016/j.ijforecast.2021.04.002>
- Sattar, A. M. A., Ertugrul, Ö. F., Gharabaghi, B., McBean, E. A., & Cao, J. (2019). Extreme learning machine model for water network management. *Neural Computing and Applications*, 31(1), 157–169. <https://doi.org/10.1007/s00521-017-2987-7>
- Segura-Heras, J. V., Bermúdez, J. D., Corberán-Vallet, A., & Vercher, E. (2022). Analysis of Weighting Strategies for Improving the Accuracy of Combined Forecasts. *Mathematics*, 10(5), 1–12. <https://doi.org/10.3390/math10050725>
- Svetunkov, I., & Boylan, J. E. (2020). State-space ARIMA for supply-chain forecasting. *International Journal of Production Research*, 58(3), 818–827. <https://doi.org/10.1080/00207543.2019.1600764>
- Zhao, D., & Zhang, H. (2022). The research on TBATS and ELM models for prediction of human brucellosis cases in mainland China: a time series study. *BMC Infectious Diseases*, 22(1), 1–12. <https://doi.org/10.1186/s12879-022-07919-w>
- Zhao, S., & Mi, X. (2019). A Novel Hybrid Model for Short-Term High-Speed Railway Passenger Demand Forecasting. *IEEE Access*, 7, 175681–175692. <https://doi.org/10.1109/ACCESS.2019.2957612>
- Zhu, K., Xun, P., Li, W., Li, Z., & Zhou, R. (2019). Prediction of Passenger Flow in Urban Rail Transit Based on Big Data Analysis and Deep Learning. *IEEE Access*, 7, 142272–142279. <https://doi.org/10.1109/ACCESS.2019.2944744>