

Impact of DTW-Based K-Medoids and Fuzzy C-Means Clustering on the Forecasting Accuracy of ARIMA, TCN, and Hybrid Models in Anomalous Time Series

Meavi Cintani¹, I Made Sumertajaya², Yenni Angraini³

^{1,2,3} Statistics and Science Data, School of Data Science Mathematics and Informatics, IPB University, Indonesia

Article Info

Article history:

Received Mar 12, 2026

Revised May 22, 2026

Accepted July 12, 2026

Keywords:

Time series clustering

DTW-MDS

K-Medoid

Fuzzy C Means

Hybrid ARIMA-TCN

Export Forecasting

ABSTRACT

This study investigates the impact of Dynamic Time Warping (DTW)-based clustering on forecasting accuracy in anomalous time series data. Monthly export values from 30 provinces in Indonesia are clustered using K-Medoids and Fuzzy C-Means (FCM) based on DTW distances reduced through Multidimensional Scaling (MDS). Cluster validation results indicate that FCM demonstrates better stability under anomalous conditions, with a Silhouette Score of 0.62 and a Davies–Bouldin Index of 0.57, compared to K-Medoids with a Silhouette Score of 0.61 and a Davies–Bouldin Index of 0.58. Forecasting performance is systematically evaluated through an expanding-window scheme using standalone ARIMA and Temporal Convolutional Network (TCN) models as baselines against a hybrid ARIMA–TCN approach. The results show that the hybrid ARIMA–TCN model achieves the lowest Mean Absolute Percentage Error (MAPE) in clusters with relatively stable patterns by effectively combining linear and non-linear components. However, for clusters characterized by higher volatility and irregular patterns, the standalone TCN model yields better forecasting accuracy. Furthermore, FCM clustering produces better overall forecasting accuracy, with an average MAPE of 7.11%, compared to 8.42% for K-Medoids. The relatively small difference in evaluation results between clean and empirical data suggests that the proposed DTW–MDS–clustering–forecasting framework maintains consistent performance in the presence of anomalies. The final model is then applied to generate export forecasts for the next 12 periods.

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



Corresponding Author:

Meavi Cintani,

Statistics and Science Data,

IPB University,

Jl. Raya Dramaga Kampus IPB, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680.

Email: meavi2501cintani@apps.ipb.ac.id

1. INTRODUCTION

The monthly export value between provinces in Indonesia shows significant variation and large differences in contributions across regions. These disparities are influenced by regional economic structures, global market dynamics, exchange rates, and various external events that may trigger sharp increases or decreases over certain periods [1], [2]. These conditions cause the time series pattern of export values to be complex, heterogeneous across provinces, and potentially to contain anomalies. Accurate information on export trends, particularly based on regions with complete and continuous historical data, is essential to support regional economic planning and sustainable trade policies, including the achievement of the Sustainable Development Goals (SDGs), especially Goal 8 (Decent Work and Economic Growth), Goal 9 (Industry, Innovation, and Infrastructure), and Goal 10 (Reduced Inequality).

Time series forecasting plays an important role in supporting decision-making in the economic and international trade sectors. However, individual forecasting for each province may increase analytical complexity and lead to unstable model performance, especially when provinces share similar temporal patterns. To address this issue, this study applies time series clustering to group provinces based on export pattern similarity. The forecasting model is then developed at the cluster level to represent its members [3]. This approach reduces the number of models required, improves computational efficiency, and enhances the stability of forecasting results. Previous studies have shown that clustering before forecasting can simplify data complexity and improve prediction consistency in time series data [4]. The presence of anomalies in time series data may obscure historical patterns and reduce the effectiveness of clustering when conventional distance measures are used [5]. Dynamic Time Warping (DTW) is a more flexible similarity measure because it can align two with similar patterns occurring at different time points [6]. Several studies have shown that DTW more representative clustering results for economic data with complex time series [7].

The K-Medoids method is relatively flexible because it can operate directly on a DTW distance matrix without requiring a Euclidean representation. This method selects actual observations as cluster representatives, making it more robust for non-Euclidean data. Chen et al. [8], showed that this approach performs well for temporal data due to its ability to preserve original patterns. In contrast, Fuzzy C-Means (FCM) relies on centroid-based representations and requires data in a Euclidean space. Izakian et al. [9], explained that DTW cannot be directly applied to FCM because the averaging process becomes problematic in non-Euclidean space. To address this limitation, Multidimensional Scaling (MDS) is used to transform the DTW distance matrix into a lower-dimensional Euclidean space. Although this transformation may introduce distortion in the local structure of the original distances, it enables both clustering methods to be applied within a comparable framework and reduces the influence of extreme fluctuations [10].

The clustering results are then integrated into the forecasting stage. Instead of building separate models for each province, this study uses cluster prototypes to represent dominant regional patterns. In general, cluster-level representations tend to have lower variability because the aggregation process helps reduce noise [11]. The forecasting process employs three approaches: ARIMA, Temporal Convolutional Network (TCN), and a hybrid ARIMA–TCN model. In this framework, both ARIMA and TCN are explicitly utilized as comparative baselines to systematically evaluate the hybrid model. ARIMA serves as a traditional baseline model to capture linear structures through autoregressive and moving-average components [12]. However, linear models may not fully capture the complexity of real-world data. Serving as the stronger, deep learning baseline, Borovykh et al. [13], demonstrated that TCN can effectively model nonlinear patterns and long-term dependencies in dynamic time series. Furthermore, Chakraborty et al. [14], showed that combining ARIMA with neural network models to capture residual nonlinear components can improve forecasting accuracy by synergizing the strengths of both baseline models.

Several studies have supported the integration of clustering and time series forecasting methods. Sumertajaya et al. [15], showed that DTW-based clustering was able to produce more representative clusters than conventional approaches. Tsabitah et al. [16], also found that forecasting based on clustering results was more efficient than forecasting at the individual level. In the context of modeling, Ghiffary et al. [17], highlighted the difference in performance between classical and deep learning models, so that the selection of methods needs to be adjusted to the characteristics of the data being analyzed.

Despite these advancements, studies that integrate DTW-based clustering with hybrid forecasting models remain limited, particularly in the context of interprovincial export data. Moreover, existing research rarely evaluates model performance under different data conditions, such as anomalous (raw) and cleaned datasets, to assess model stability. This gap is important because Indonesia's export data often exhibit sudden fluctuations that can significantly influence both clustering structure and forecasting accuracy. Therefore, this study proposes an integrated framework that combines DTW-based K-Medoids and FCM clustering with ARIMA, TCN, and

hybrid ARIMA–TCN forecasting models. This study also evaluates model performance under both empirical and cleaned data conditions using an expanding-window scheme. By linking clustering and forecasting stages, this approach provides deeper insights into how anomalies affect grouping patterns and prediction results, while examining the stability and consistency of the proposed framework.

2. RESEARCH METHOD

2.1 Data Sources

The data used in this study were obtained from the official website of Badan Pusat Statistik (BPS). The dataset contains monthly export values for 30 provinces in Indonesia from January 2013 to March 2025. Out of the total 38 provinces, only 30 were included in the analysis based on data completeness and continuity. Provinces resulting from the administrative division of Papua, as well as those with a high proportion of missing values, were excluded because they did not meet the requirements for consistent time series modeling. This selection ensures the reliability of the clustering and forecasting processes, as incomplete or highly irregular time series may introduce bias and reduce model stability. The export values were originally reported in millions of US dollars and were converted into thousands of US dollars to facilitate numerical stability during modeling. This dataset enables the analysis of temporal changes in export performance, including periods of significant fluctuation, as well as cross-provincial differences in export levels.

Table 1. Monthly Export Value Data by Province of Origin (Thousands of US\$).

Bulan	Aceh	Sumatera Utara	...	Gorontalo	Kalimantan Utara
1/1/2013	100,753.00	847,450.00	...	212	78,250.00
2/1/2013	51,438.00	809,250.00	...	369	43,510.00
...	...	...	...	...	...
2/1/2025	53,940.00	1,002,500.00	...	7,070.00	90,030.00
3/1/2025	57,030.00	990,470.00	...	7,980.00	137,190.00

2.2 Clean Data

Anomaly cleaning was conducted using the winsorization method with the three sigma (3σ) criterion to minimize the impact of extreme values on the export value time series data. This method preserves the core structure of the data while reducing the influence of outliers. The data cleaning process involved several steps. Initially, the differences in the export data were adjusted according to equation (1):

$$d' = \begin{cases} \bar{d} + 3\sigma_d, & \text{jika } d > 0, \\ \bar{d} - 3\sigma_d, & \text{jika } d < 0 \end{cases} \quad (1)$$

where d denotes the difference in export value before adjustment, $\bar{d}$ denotes the average value of all differences d , and σ_d denotes the standard deviation of differences d . Second, the export value at the time t is updated using equation (2):

$$y'_t = d' + y_{t-1} \quad (2)$$

where y'_t represents the adjusted export value at time t , y_{t-1} represents the export value at the previous time. The steps are repeated until no observations exceed the three-sigma threshold, ensuring the dataset is free from extreme values and ready for clustering and forecasting analysis.

2.3 Research Design

1. Data preprocessing includes imputation of missing data using linear interpolation, data exploration through time series plots and boxplots, and anomaly detection using the three sigma rule based on the difference in export values between months.
2. The similarity of provincial export value time series was measured using Dynamic Time Warping (DTW) and presented in the form of a distance matrix.
3. Dimension reduction was performed using Multidimensional Scaling (MDS) to map the DTW distance matrix to a low-dimensional space by minimizing the stress value.
4. Clustering analysis

- a. Apply the K-Medoids method based on MDS reduction results to minimize the total distance within clusters.
- b. Apply the Fuzzy C-Means (FCM) method, which allows each province to have a degree of membership in more than one cluster
- c. Performing defuzzification on the FCM results by assigning each province to the cluster with the highest membership degree.
- d. Evaluating cluster quality using the elbow method, silhouette score, and Davies–Bouldin Index.
- e. Identifying provincial members in each cluster based on the optimal number of clusters.
- f. Determine the cluster prototype, which is the median of the time series for K-Medoids and the average of the time series for FCM.

5. Cluster level modeling

The forecasting model is constructed using cluster prototypes derived from the cleaned data, based on the expanding window approach [18]. In this study, model development and parameter selection are performed using the cleaned dataset to ensure stable learning, and the resulting optimal models are subsequently applied to the empirical data containing anomalies for evaluation and final forecasting.

a. Data Splitting

The data partitioning scheme follows an expanding window approach to preserve the temporal structure of the time series. In this scheme, the training set is progressively extended while the testing set remains fixed for each evaluation window. The detailed partitioning configuration is presented in Table 2.

Table 2. Expanding window

Window	Training data	Testing data
Window 1	Jan 2013 – Mar 2021	Apr 2021 – Mar 2022
Window 2	Jan 2013 – Mar 2022	Apr 2022 – Mar 2023
Window 3	Jan 2013 – Mar 2023	Apr 2023 – Mar 2024

The final model validation evaluates the last 12 months (April 2024–March 2025) to ensure that the model maintains good performance in the most recent period.

a. ARIMA Model

- The ARIMA model is constructed using the Box–Jenkins approach, including order identification, parameter estimation, and residual diagnostic testing.
- This allows the model to adapt to the evolving structure of the time series data.
- This ensures that the model adapts to the evolving structure of the time series data.
- Furthermore, the consistency of selected ARIMA orders across the expanding windows is analyzed to identify stable model structures. The most frequently selected model order is then used as the final model for the last 12-month forecasting period.

b. Temporal Convolutional Network (TCN) Model

- Extracting the data and applying normalization using MinMaxScaler to ensure stable neural network training.
- Building the TCN architecture utilizing dilated causal convolutions to ensure long-term dependencies are captured without future information leakage.
- Performing a Grid Search to find the optimal combination of parameters (time steps, number of filters, dropout rate, and batch size) while implementing strict random seed controls for reproducibility.
- The performance of each parameter combination is evaluated using forecasting error metrics (MAPE and RMSE) across the expanding windows. The parameter configuration that consistently produces the lowest error is selected as the final model for the last 12-month forecasting period.

c. Hybrid ARIMA-TCN modeling

- Calculating the residuals (the difference between actual values and ARIMA fitted values) dynamically for each strictly partitioned expanding window.
 - Training the TCN model on the ARIMA residuals to capture complex nonlinear components.
 - Performing hyperparameter tuning for the TCN residual model using a grid search approach within each expanding window.
 - Evaluating hybrid model performance across expanding windows using MAPE and RMSE, and selecting the best-performing configuration for the final 12-month forecasting period.
 - Reverting the TCN residual predictions back to their original scale using the inverse transform of the MinMaxScaler.
 - Combining the linear forecast generated by ARIMA with the non-linear residual forecast generated by TCN to produce the final comprehensive predicted value.
6. Model comparison
- Compare the performance of the ARIMA, TCN, and hybrid ARIMA–TCN models in each cluster in clean data and empirical data.
 - Determine the best combination of clustering and forecasting methods based on the lowest MAPE value.
- $$MAPE = \frac{1}{n} \sum \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (3)$$
- The comparison also examines the consistency of model performance across clusters with different characteristics, including stable and volatile patterns.
7. Final forecasting
- Apply the best clustering method and forecasting model to empirical data containing anomalies.
 - Generate forecasts of export values for the next 12 periods for each cluster.

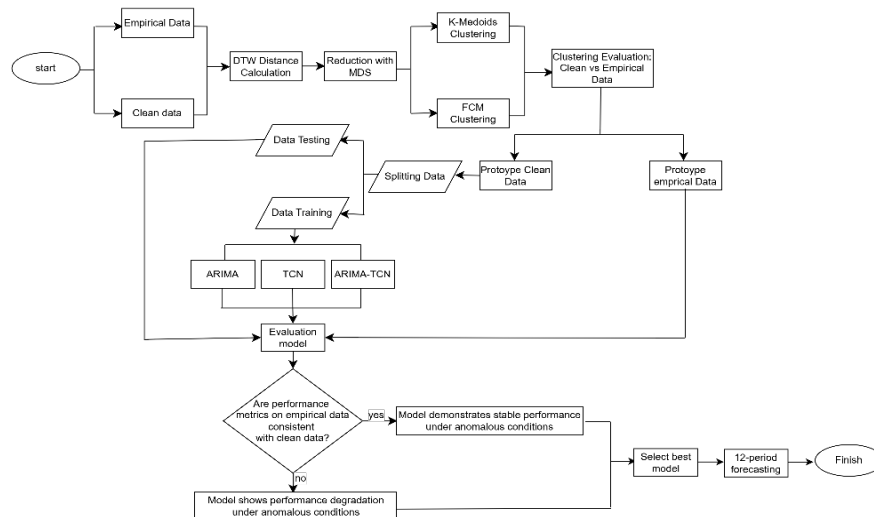


Figure 1. Research design

2.4 Dynamic Time Warping Based Multidimensional Scaling (DTW–MDS)

Dynamic Time Warping (DTW) is a time series similarity measurement method that can compare patterns based on shape (shape-based similarity) and is flexible to time shifts. Unlike Euclidean distance, which requires alignment of the same time points, DTW allows for warping paths so that two time series with different phases can still be compared optimally. The DTW distance between two time series S and T is formulated as:

$$DTW(S, T) = \min_W \left[\sum_{k=1}^p \delta(w_k) \right] \quad (4)$$

with W optimal warping trajectory and $\delta(w_k)$ local distance between data point pairs [19]. The DTW distance matrix is non-Euclidean, so it cannot be directly used in centroid or medoid based clustering algorithms in a stable manner. Therefore, dimensional reduction is performed using Multidimensional Scaling (MDS). MDS maps the DTW distance matrix into a low-dimensional space while maintaining the proximity structure between objects by minimizing the stress function, which is formulated as:

$$Stress = \sqrt{\frac{\sum_{i < j} (d_{ij} - \delta_{ij})^2}{\sum_{i < j} d_{ij}^2}} \quad (5)$$

d_{ij} interprovincial DTW distance and δ_{ij} mapping distance in low-dimensional space [20]. The integration of DTW and MDS produces a spatial representation that reflects the similarity of time series patterns more stably and facilitates the clustering process and further analysis.

2.5 Autoregressive Integrated Moving Average

ARIMA (Autoregressive Integrated Moving Average) modeling is widely used in time series forecasting with the Box–Jenkins approach. The modeling process includes identifying the order p, d, q based on data stationarity, ACF and PACF patterns, and residual diagnostic tests to ensure that the error (ε_t) is white noise, it does not contain autocorrelation and has constant variance. The ARIMA model is considered feasible if it meets these criteria, and if there is more than one candidate model, the model with the smallest forecasting error in the test data is selected. The general form of the ARIMA model is written as:

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \quad (6)$$

$\phi(B)$ and $\theta(B)$ are the autoregressive and moving average operators, respectively, ε_t is random error [12].

2.6 Temporal Convolutional Network

Temporal Convolutional Network (TCN) modeling is a dilated causal convolution-based architecture that is effective in capturing long-term dependencies in time series data and is able to overcome the vanishing gradient problem that often occurs in deep learning models. At each convolutional layer, TCN applies the principle of causality, where each neuron only receives information from time s and the previous period so that there is no information leakage from the future to the past. The convolution operation in TCN is formulated as:

$$F(s) = \sum_{k=0}^{K-1} W(k) x(s - d \cdot k) \quad (7)$$

$F(s)$ is the convolution output at time s , $W(k)$ is the weight of kernel k , $x(s - d \cdot k)$ is the input value at the previous time, d is the dilation rate, K is the size of the convolution kernel. The TCN training process is carried out through hyperparameter tuning using a grid search approach in accordance with the specified parameter combinations. The adjusted parameters include the number of filters, dropout rate, and batch size. The best model is selected based on the smallest error value in the validation data [21]. The TCN architecture can be seen in Figure 2.

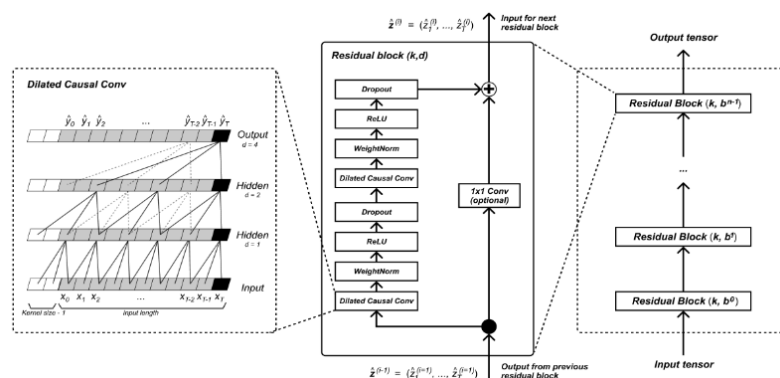


Figure 2. TCN architecture

3. RESULTS AND DISCUSSION

3.1 Data Preprocessing and Anomaly Detection

The study uses monthly provincial export value data in Indonesia from January 2013 to March 2025. Out of a total of 38 provinces, 30 were selected based on data completeness and continuity. Provinces resulting from the administrative division of Papua, as well as those with a high proportion of missing values, were excluded because they did not meet the requirements for consistent time series modeling. The exploratory analysis presented in Figure 3 shows that national export values declined during 2015–2016, followed by a sharp increase in 2021 and a peak in mid-2022, in line with post-pandemic economic recovery and rising global commodity prices. Cross-provincial analysis also reveals significant disparities in export contributions. These findings confirm the heterogeneity of regional economic structures, which forms the basis for a more targeted and systematic application of clustering and time series forecasting methods. However, this selection may limit the generalizability of the findings, particularly for regions with distinct export dynamics that are not represented in the dataset.

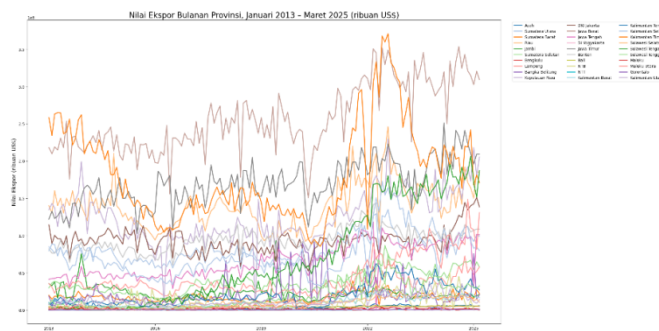


Figure 3. Monthly Export Values of Indonesian Provinces

Anomaly detection was performed because the export value data for the 30 provinces did not include predefined anomaly labels. Therefore, a statistical approach was applied to identify sudden spikes or declines that deviate from the general time series pattern, including both long-term trends and seasonal variations. The anomaly threshold was determined using the three-sigma rule, where observations falling outside the normal range of the data distribution were classified as anomalies. The results indicate that the number of detected anomalies varies across provinces, ranging from 0 to 7 occurrences. Bangka Belitung recorded the highest number of anomalies, while Bengkulu, DKI Jakarta, East Java, and Maluku showed no anomalies, as all observations remained within the defined threshold. In total, 72 anomaly points were identified and subsequently corrected using the three-sigma winsorization method by capping extreme values at the specified threshold. Most provinces returned to a stable pattern after a single iteration of the cleaning process.

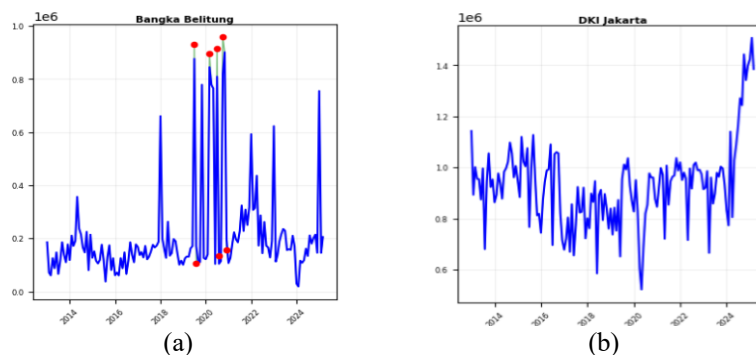


Figure 4. Visualization of anomaly detection: (a) Bangka Belitung (b) DKI Jakarta

Figure 4 illustrates the differences in anomaly characteristics across provinces. In Bangka Belitung, a significant spike in export values can be observed, deviating markedly from the surrounding trend, and thus identified as an anomaly. In contrast, DKI Jakarta exhibits a relatively

stable fluctuation pattern, with all observations remaining within the detection threshold, indicating the absence of anomalies.

3.2 Time series Clustering

Time series clustering was applied to group provinces exhibiting similar export patterns. Dynamic Time Warping (DTW) distance metric was used to measure temporal similarity, enabling the identification of comparable dynamics despite potential misalignment in timing between regions. This procedure was applied to both empirical data and data that had been cleaned of anomalies, so that comparisons of clustering structures could be made consistently in both data conditions. The results of the DTW distance matrix calculations for each condition are shown in Figure 5.

$$\begin{array}{cc} \begin{bmatrix} 0 & 9446915 & 1641719 & \dots & 1219401 \\ 9446915 & 0 & 7454712 & \dots & 8088850 \\ 1641719 & 7454712 & 0 & \dots & 683099 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1219401 & 8088850 & 683099 & \dots & 0 \end{bmatrix} & \begin{bmatrix} 0 & 9425141 & 1609609 & \dots & 1233669 \\ 9425141 & 0 & 7445106 & \dots & 8012347 \\ 1609609 & 7445106 & 0 & \dots & 695297 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1233669 & 8012347 & 695297 & \dots & 0 \end{bmatrix} \\ \text{(a)} & \text{(b)} \end{array}$$

Figure 5. 30×30 DTW distance matrix: (a) clean data (b) empirical data

Based on the results of the DTW distance matrix, it can be seen that anomaly cleaning using the three sigma rule makes the similarity pattern between provinces more stable. The provinces with the smallest DTW distance are dominated by provinces with small export scales, namely Gorontalo–East Nusa Tenggara (19,943.5). The largest pattern occurs between West Java and East Nusa Tenggara (3,1597,888.3). DTW analysis of empirical data shows a more diverse level of similarity in export patterns between provinces. The most similar pattern is between East Nusa Tenggara and Gorontalo (21,178.4), while the largest pattern difference occurs between West Java and East Nusa Tenggara (3,160,353.1). These findings indicate that anomaly cleaning does not fundamentally change the main similarity patterns between provinces, but only reduces the influence of extreme values that previously exaggerated the distance.

The non-Euclidean DTW distance matrix is projected using MDS into a low-dimensional space so that it can be represented geometrically and evaluated fairly by FCM and K-Medoids. In empirical data, the stress values are 0.0282 (two dimensions) and 0.0255 (three dimensions), while in clean data they are 0.0273 (two dimensions) and 0.0253 (three dimensions). All values are below the threshold of 0.05, indicating that the global dissimilarity structure is well preserved in the projection space [20]. However, it should be noted that dimensionality reduction techniques such as MDS may introduce some distortion in local distance relationships. Despite this limitation, the relatively consistent stress values before and after anomaly cleaning suggest that anomalies primarily affect local structures without substantially altering the overall relationship patterns between objects, in line with the principle of robust statistics [22].

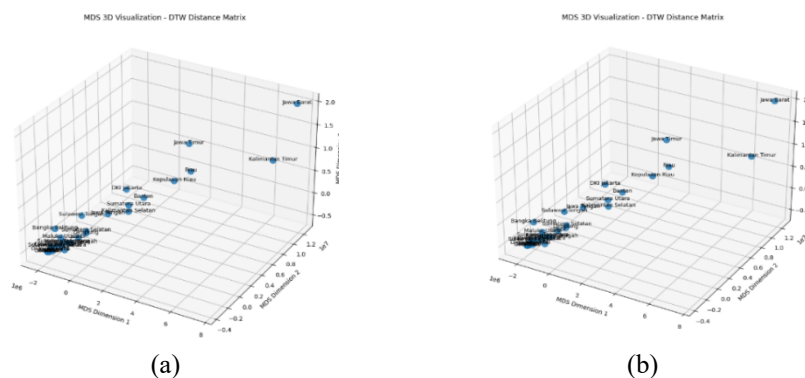


Figure 6. 3-dimensional MDS visualization: (a) clean data (b) empirical data

The optimal number of clusters for both FCM and K-Medoids was determined using the Elbow method, Silhouette Score (SS), and Davies–Bouldin Index (DBI). The evaluation was conducted separately on empirical and cleaned data to ensure the stability of the clustering structure. Based on these criteria, the optimal number of clusters was consistently identified as $k = 3$ for both methods, indicating that the clustering configuration is stable and appropriate for subsequent analysis.

To assess clustering stability, the performance of FCM and K-Medoids was compared across clean and empirical data. As shown in Table 3, the FCM method on clean data achieves a Silhouette score of 0.63 and a DBI of 0.56, indicating good cluster compactness. Under the same condition, K-Medoids produces comparable results, with a Silhouette score of 0.62 and a DBI of 0.57. These results suggest that both methods are able to capture similar underlying patterns, although FCM shows a slight advantage in cluster compactness.

In empirical data, the evaluation metrics for both methods show a slight decrease. FCM yields a Silhouette score of 0.62 and a DBI of 0.57, while K-Medoids achieves a Silhouette score of 0.61 and a DBI of 0.58. The relatively small differences between clean and empirical data indicate that the clustering structure remains stable in the presence of anomalies. This stability suggests that the overall pattern grouping is not significantly affected by extreme values, indicating that anomalies primarily introduce local distortions without substantially altering the global structure of the data, in line with the principle of robust statistics [22]. This finding is also consistent with the use of DTW, which flexibly captures similarity in time series patterns, and the application of MDS, which maps DTW distances into a Euclidean space while preserving the global structure of inter-provincial relationships [23].

Table 3. Evaluation of clustering results

Method	Cluster	Province	Clean data		Empirical data	
			SS	DBI	SS	DBI
FCM	1	Jawa Barat, Jawa Timur, Kalimantan Timur	0.63	0.56	0.62	0.57
	2	Sumatera Utara, Riau, Kepulauan Riau, DKI Jakarta, Jawa Tengah, Banten, Kalimantan Selatan, Sulawesi Tengah				
	3	Aceh, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, DI Yogyakarta, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Gorontalo, Kalimantan Utara				
K-Medoids	1	Riau, Kepulauan Riau, Jawa Barat, Jawa Timur, Kalimantan Timur	0.62	0.57	0.61	0.58
	2	Sumatera Utara, DKI Jakarta, Jawa Tengah, Banten, Kalimantan Selatan, Sulawesi Tengah				
	3	Aceh, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, DI Yogyakarta, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Gorontalo, Kalimantan Utara				

Table 3 shows differences in provincial composition between the clustering results of the FCM and K-Medoids methods, even though both methods use three clusters. The FCM method places West Java, East Java, and East Kalimantan in a single cluster, reflecting provinces with relatively high export values. The K-Medoids method includes Riau and the Riau Islands in the same group as those provinces. This difference arises because the two methods employ distinct clustering principles. The FCM method forms clusters based on the overall degree of pattern similarity, so provinces with similar characteristics but not dominant ones tend to fall into intermediate clusters. The K-Medoids method determines clusters based on proximity to a central object (medoid), so provinces with large export contributions, such as Riau and the Riau Islands, are closer to the group of high-contribution provinces. These differences in mechanism result in the composition of cluster members not being entirely the same, but both methods still depict a similar structure of interprovincial export patterns.

3.3 Forecasting Model

The forecasting phase is centered on the cluster prototypes derived from the clustering analysis. For the Fuzzy C-Means (FCM) method, the prototype is represented by the average export

values of all provinces within the cluster, whereas for K-Medoids, the prototype is defined by the medoid—the actual province exhibiting the minimum aggregate distance to other cluster members. This study utilizes monthly export data from 34 provinces spanning 147 observations. To preserve the temporal integrity of the series, the data are partitioned into a training set (approximately 92% or 135 observations) and a testing set (8% or 12 observations) using a chronological split. Crucially, the initial forecasting models comprising ARIMA, TCN, and a hybrid ARIMA-TCN configuration are constructed using cleaned data to establish a robust baseline. Hyperparameters are optimized via a systematic grid search and evaluated through expanding window cross-validation with a 12-month forecasting horizon. This approach ensures that the models capture the underlying structural patterns before being subjected to empirical data containing anomalies.

The ARIMA modeling procedure, which serves as the primary conventional baseline for this study, followed the Box–Jenkins approach. First, the training time series data were plotted to observe the overall pattern of the series. Stationarity was then examined using ACF and PACF plots along with the Augmented Dickey–Fuller (ADF) test. After applying differencing, the ADF test resulted in p -values below 0.05, indicating that the series had become stationary. Based on the patterns observed in the ACF and PACF plots, several candidate ARIMA models were identified. The parameters of these models were estimated using maximum likelihood estimation (MLE). Model adequacy was subsequently evaluated using residual diagnostics, including the Ljung–Box test. The test results yielded p -values greater than 0.05, indicating that the residuals showed no significant autocorrelation. Furthermore, the Jarque–Bera test indicated that most residuals were approximately normally distributed. In cases where the residuals slightly deviated from normality, the models were still considered acceptable based on the Central Limit Theorem. Crucially, across the three expanding windows, the identified (p, d, q) orders remained remarkably consistent, demonstrating the structural stability of the cluster prototypes. The final optimal ARIMA configurations, which were re-estimated independently within each window to adapt to the evolving temporal structure, are summarized in Table 4.

Table 4. ARIMA model identification

Method	Cluster	Model	Parameter	coefficient	p -value	AIC
K-Medoids	1	ARIMA (1,1,0)	AR(1)	-0.4324	0,000	-233.986
	2	ARIMA (0,1,1)	MA(1)	-0,5468	0,000	3372.169
	3	ARIMA (0,1,1)	MA(1)	-0.5469	0,000	-14.858
FCM	1	ARIMA (2,1,0)	AR(1)	-0.6372	0,000	-293.977
			AR(2)	-0.2591	0,003	
	2	ARIMA (2,1,0)	AR(1)	-0.6120	0,000	-263.858
			AR(2)	-0.2045	0,027	
			AR(1)	-0.5174	0,000	
	3	ARIMA (3,1,0)	AR(2)	-0.3041	0,001	-133.996
			AR(3)	-0.2781	0,000	

The TCN model is developed by tuning several hyperparameters using a grid search approach. The parameter combinations include the input sequence length (n_steps_in) {3, 6, 12}, the number of convolution filters in two consecutive layers {16, 32, 64}, dropout values {0.1, 0.2, 0.3}, and batch sizes {4, 8, 16}. The model is trained using the Adam optimizer with the mean squared error (MSE) loss function. The optimal parameter configuration is selected based on performance across expanding windows, ensuring that the model generalizes well to unseen data. The final hyperparameter configurations for each cluster are summarized in Table 5.

Table 5. Hyperparameter optimization TCN

Method	Cluster	Input length	Filter 1	Filter 2	Drop out	Batch size
K-Medoids	1	3	64	64	0,2	16
	2	3	64	64	0,1	4
	3	3	16	32	0,1	8
FCM	1	3	32	8	0,1	4
	2	3	32	32	0,1	4
	3	3	32	64	0,2	16

The hybrid ARIMA–TCN model is constructed using a residual-based integration strategy to address the multifaceted nature of export data. In this framework, the ARIMA component is first utilized to extract the underlying linear trends from the cleaned dataset. As illustrated in Figure 7, the resulting residuals from Cluster 3 (FCM) represent the stochastic and nonlinear components that the linear model fails to capture. Although these residuals oscillate around a zero mean, they exhibit significant high-frequency volatility and complex patterns, which serve as the essential input for the TCN model.

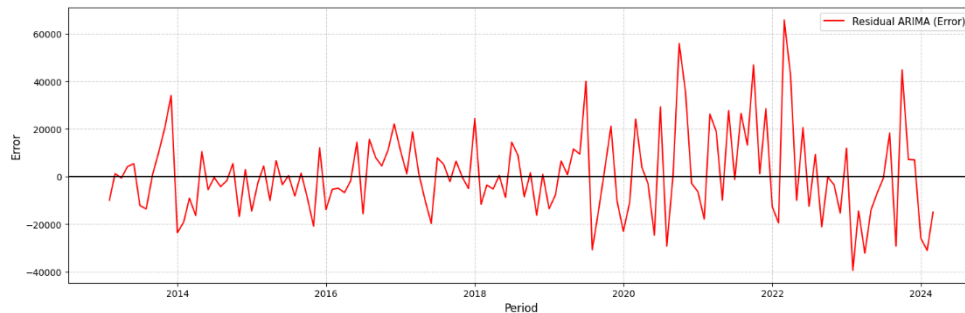


Figure 7. ARIMA residual patterns cluster 3 based on FCM clustering

Table 6 presents the hyperparameter configuration used in this study, which follows the same structure as the baseline TCN model. The optimal values are selected based on the lowest RMSE and MAPE obtained from the ARIMA residual data. In the expanding window framework, residuals are generated separately within each window to preserve the temporal structure. These residual based predictions are then combined with ARIMA forecasts to produce the final hybrid output.

Table 6. Hyperparameter optimization of TCN for ARIMA residuals

Method	Cluster	Input length	Filter 1	Filter 2	Drop out	Batch size
K-Medoids	1	6	16	64	0,2	16
	2	12	16	32	0,1	8
	3	16	16	32	0,2	16
	1	6	16	16	0,1	8
FCM	2	12	16	16	0,1	4
	3	12	16	16	0,3	16

Table 7 compares the performance of three forecasting models: ARIMA, TCN, and the ARIMA–TCN hybrid. The results indicate that MAPE values tend to be higher for the empirical data than for the cleaned data. This is understandable, since real-world time series often contain noise, volatility, and unexpected shocks that increase modeling difficulty [24]. After applying clustering using K-Medoids and FCM, the forecasting models are estimated separately for each cluster. In this framework, provinces with relatively similar export patterns are grouped together

Table 7. Forecast evaluation

Method	Cluster	MAPE Clean Data			MAPE Empirical data		
		ARIMA	TCN	ARIMA-TCN	ARIMA	TCN	ARIMA-TCN
K-Medoid	1	8.56	8.23	7.53	8.63	8.57	7.68
	2	6.97	5.77	5,13	7.25	6.81	5,34
	3	12,90	11.66	11.21	13.98	12.23	13.12
FCM	1	6,08	5.96	5,91	6,24	6.07	6,02
	2	7,97	7.71	7,32	8,04	7.92	7.40
	3	8,03	7.51	7,96	8,83	7.90	7.99

Based on the K-Medoids clustering results, the hybrid ARIMA–TCN model gives the lowest MAPE values in Cluster 1 and Cluster 2 for both the cleaned data and the empirical data. In Cluster 2, the hybrid model records the smallest error among all clusters, with a MAPE of about 5.13% for the cleaned data and 5.34% for the empirical data. This result suggests that the export

pattern in this cluster is relatively stable, so the model can predict future values more easily. In contrast, Cluster 3 shows the highest forecasting error compared with the other clusters. The MAPE in this cluster is above 11% for the cleaned data and increases to around 13% in the empirical data, indicating stronger fluctuations in the data and a more difficult forecasting task.

The results from the FCM method show a slightly different and more stable pattern. The MAPE values across clusters are closer to each other compared with the results from the K-Medoids method. In the cleaned data, the MAPE of the hybrid model ranges from about 5.91% to 7.96%, while in the empirical data it ranges from about 6.02% to 7.99%. The relatively small difference between the cleaned data and the empirical data suggests that the clusters produced by the FCM method remain more stable when anomalies are present in the data. This also indicates that FCM forms more homogeneous clusters, which helps the forecasting model produce more consistent results.

The comparison results show that the hybrid ARIMA–TCN model generally performs better than the standalone baseline models (ARIMA and TCN), particularly for clusters with more stable patterns, as it combines the linear and non-linear strengths of each method. However, for clusters with more irregular patterns such as Cluster 3 in K-Medoids and FCM the standalone TCN model actually yields better results because it is not constrained by linear assumptions and is more flexible in adapting to sudden changes in the data.

The results are consistent with the study by Tsabitah, et al. [25] which integrates DTW-based time series clustering with a hybrid Vector Error Correction Model with Exogenous Variables–Long Short-Term Memory model, demonstrating that combining clustering with hybrid forecasting approaches can more effectively capture the dynamics of commodity price patterns. Similarly, Bai et al. [21] show that TCN models are capable of capturing complex patterns and long-term dependencies, making them particularly suitable for highly volatile time series data. Therefore, the selection of the most appropriate model ultimately depends on the characteristics of the data being analyzed. Furthermore, the relatively small differences in MAPE between clean and empirical data across clusters indicate that the proposed clustering–forecasting framework maintains consistent performance under different data conditions, suggesting a good level of stability in the presence of anomalies.

A comparison of the average MAPE for the hybrid model between clean data and empirical data can be seen more clearly in Figure 8. Based on the systematic expanding-window evaluation, in the K-Medoids method, the average MAPE increased from around 7.96% in clean data to 8.42% in empirical data. This means that there was an increase in error of around 0.46%. Meanwhile, in FCM, the increase is much smaller, from 6.91% to 7.11%, or around 0.20%. This empirical evidence clearly demonstrates that the increase in error in K-Medoids is greater than in FCM, definitively confirming that FCM possesses superior stability.

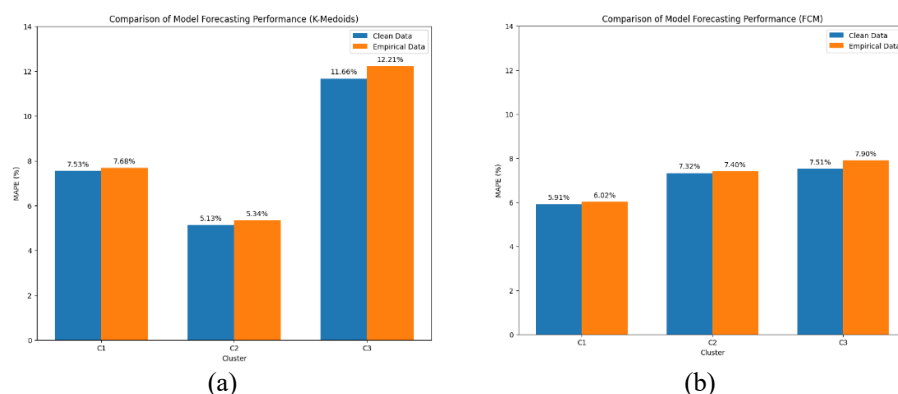


Figure 8. Comparison of MAPE hybrid ARIMA-TCN performance: (a) K-medoids (b) FCM

This difference is fundamentally related to how group prototypes are determined. In K-Medoids, prototypes are determined from group time medoids, which are actual single time series that are considered to be the most representative of the group. The time series selected as the

medoid can carry extreme patterns if the empirical data contains anomalies, resulting in a less stable representation of group patterns [26]. In contrast, in FCM, group representation is calculated using the average of the time series, so that the general pattern of the group remains more stable even if there are anomalies in the data [27]. In this study, the FCM clustering results are represented using the average time series within each cluster, resulting in smoother pattern representation. When combined with DTW, this approach enables more effective grouping of time series based on temporal similarity, which in turn supports more consistent and reliable forecasting performance.

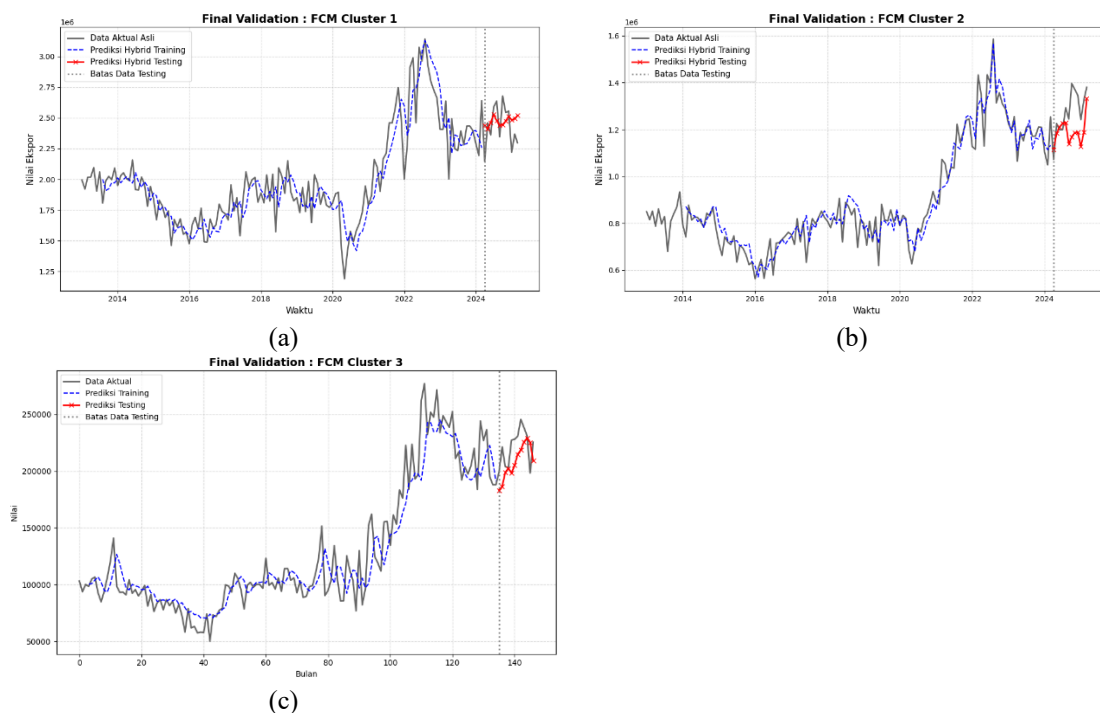


Figure 9. Final Validation: (a) FCM cluster 1 (b) FCM cluster 2 (c) FCM cluster 3

Figure 9 presents the final out-of-sample validation results for each FCM cluster, visually supporting the forecasting performance indicated by the MAPE evaluations. In Cluster 1, the predicted series closely follows the movement of the actual data during the training period, indicating that the model is able to capture the underlying trend and structural changes of the series. During the testing period (indicated by the red trajectory), the predicted values remain within a similar range as the observed data and continue to follow the upward movement at the end of the series, although small deviations are observed at several points.

In Cluster 2 and Cluster 3, the predicted values also reflect the general direction of the observed data. Cluster 2 shows a relatively smoother pattern, where the predictions follow the gradual increase after the earlier decline. In contrast, Cluster 3 exhibits higher historical fluctuations and volatility, which leads to slightly larger differences between predicted and actual values in several periods. Nevertheless, the predicted series still follows the main movement pattern of the data, indicating that the hybrid model is able to capture both linear and non-linear characteristics within each cluster.

3.4 Forecast 12 Periode

Building upon the observed stability of the FCM method, its generated cluster structure is used as the basis for compiling observation value projections for the next 12 months. Although at the empirical evaluation stage each cluster occasionally shows a different best model depending on local volatility, at the final projection stage all clusters are predicted using the hybrid ARIMA–TCN model. This decision is based on the consideration that hybrid models generally demonstrate stable and consistent performance in capturing both linear and non-linear patterns over a longer

forecasting horizon. Furthermore, utilizing a single framework ensures methodological consistency and provides a more uniform basis for handling potential future fluctuations and outliers. Thus, the forecasting approach becomes more consistent across clusters. The 12-month forward projection and its comparison with the historical pattern of the previous period are presented in Figure 10.

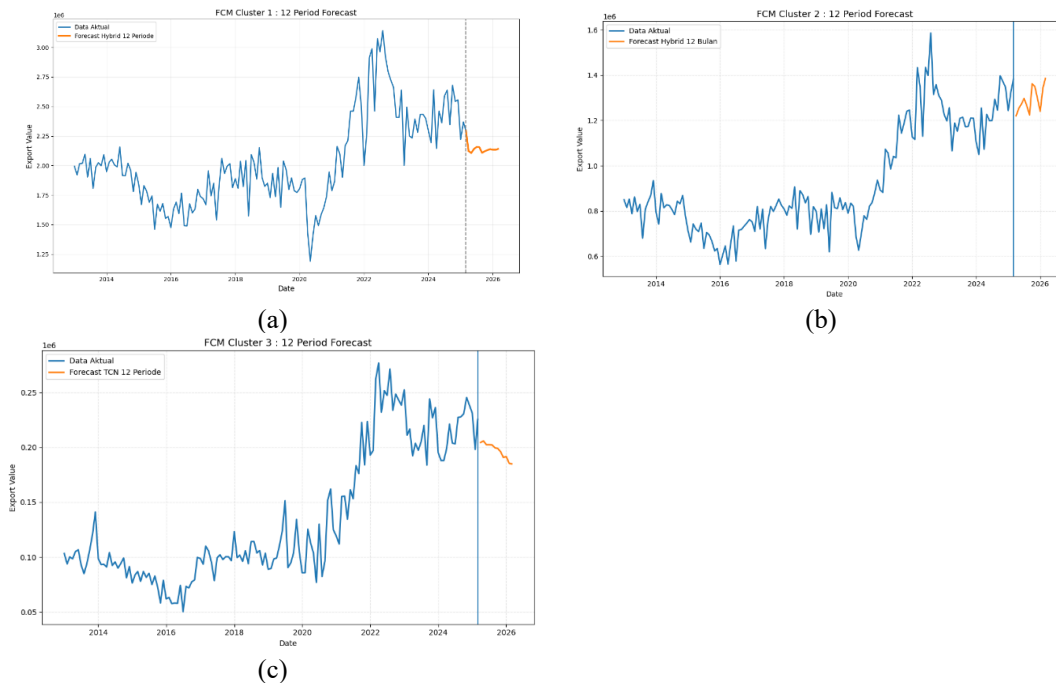


Figure 10. 12-period forecasting: (a) FCM cluster 1 (b) FCM cluster 2 (c) FCM cluster 3

Cluster 1 consists of West Java, East Java, and East Kalimantan. The graph shows that export values in this group increased sharply in the previous period. In the 12-month forecast, export movements appear more stable with a slight decline at the beginning of the forecast period. After that, export values tend to move sideways into a consolidation phase. This pattern indicates that export activity in this group is still sustained, but its growth is not as strong as the peak recovery phase observed in the previous period.

Cluster 2 includes North Sumatra, Riau, Riau Islands, DKI Jakarta, Central Java, Banten, South Kalimantan, and Central Sulawesi. The graph shows a consistent upward trend in the forecast period. Export values gradually increase until the end of the forecast horizon. This condition indicates that export activity in this group still has robust growth opportunities and economic resilience in the short term.

Cluster 3 consists of Aceh, West Sumatra, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung, Yogyakarta Special Region, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Sulawesi, Southeast Sulawesi, Maluku, North Maluku, Gorontalo, and North Kalimantan. Based on the graph, the historical movement of export values in this group is not entirely stable, showing significant fluctuations particularly after 2020. During the projection period, export values are expected to decline gradually with minor fluctuations, indicating a trend of weakening but within a relatively controlled range. This controlled projection simultaneously demonstrates the hybrid model's capability to logically smooth out extreme historical volatility to produce rational future estimates.

Based on these findings, targeted, cluster-specific policy strategies can be considered. In regions with fluctuating patterns and slowing growth projections (such as Cluster 1 and Cluster 3), policy directions may focus on developing a more diversified economic structure, strengthening the downstreaming of leading sectors, and creating new sources of growth to reduce dependence on commodity cycles. In regions with relatively stable growth trends (such as Cluster 2), maintaining this momentum may involve enhancing industrial competitiveness, improving productivity, and

supporting investment and infrastructure development. Ultimately, the application of these targeted interventions based on cluster-level forecasting can support more sustainable growth, promote more equitable economic outcomes, and contribute to the achievement of Sustainable Development Goals (SDGs) 8, 9, and 10.

4. CONCLUSION

This study shows that the integration of DTW, MDS, and FCM is effective in grouping interprovincial export patterns based on the similarity of their time dynamics and is more stable against anomalies than K-Medoids. Based on the evaluation of the empirical data, FCM produces a more consistent cluster structure with a Silhouette Score of 0.62 and a Davies–Bouldin Index of 0.57. This structural consistency is associated with improved forecasting performance, yielding an average MAPE of 7.11%, which is lower than that of K-Medoids at 8.42%.

The consistency of model performance between clean data and empirical data systematically evaluated through an expanding-window scheme against standalone baselines, suggests that the proposed clustering–forecasting framework maintains stable performance in the presence of noise and anomalies. However, this study also highlights an important nuance: while the hybrid approach performs well in clusters with relatively stable trajectories, the standalone TCN model proves more effective in handling highly volatile and irregular patterns. This finding emphasizes the importance of selecting forecasting models based on cluster-specific characteristics.

Overall, this study develops a comprehensive analytical framework for clustering and forecasting complex export time series data. The framework improves the identification of interregional patterns and enhances the stability of forecasting results. These findings highlight the importance of formulating targeted economic policies based on the characteristics of each regional cluster, supporting more effective regional development strategies and contributing to the achievement of the Sustainable Development Goals (SDGs), particularly Goals 8, 9, and 10.

Further research is recommended to explore alternative anomaly handling methods to identify more optimal approaches for extreme value treatment. In addition, future studies may incorporate multivariate time series models, such as ARIMAX or multivariate deep learning architectures, to improve forecasting accuracy by including additional relevant variables. Simulation-based studies with controlled anomaly scenarios are also suggested to further evaluate the robustness of the proposed framework under more rigorous conditions.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

- [1] B. P. Statistik, “Statistik Ekspor Impor Indonesia 2023,” 2023.
- [2] Adela Nursafira, Annisa Syafriani, Sabrina Natasya Matondang, Dayu Rizki Nurhamda, and Dwita Sakuntala, “Analysis of Factors Affecting Export and Import Performance,” *Indonesian Journal of Entrepreneurship and Startups*, vol. 3, no. 1, pp. 21–32, Jan. 2025, doi: 10.55927/ijes.v3i1.13355.
- [3] T. Warren Liao, “Clustering of time series data—a survey,” *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, Nov. 2005, doi: 10.1016/j.patcog.2005.01.025.
- [4] T. Fu, “A review on time series data mining,” *Eng. Appl. Artif. Intell.*, vol. 24, no. 1, pp. 164–181, Feb. 2011, doi: 10.1016/j.engappai.2010.09.007.
- [5] M. Zhang, J. Guo, X. Li, and R. Jin, “Data-Driven Anomaly Detection Approach for Time-Series Streaming Data,” *Sensors*, vol. 20, no. 19, p. 5646, Oct. 2020, doi: 10.3390/s20195646.
- [6] E. Keogh and C. A. Ratanamahatana, “Exact indexing of dynamic time warping,” *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, Mar. 2005, doi: 10.1007/s10115-004-0154-9.
- [7] A. N. Putri, N. Satyahadewi, and S. Aprizkiyandari, “Pengelompokan Provinsi di Indonesia Menggunakan Time Series Clustering pada Sektor Ekspor Nonmigas,” *Jambura Journal of Mathematics*, vol. 6, no. 1, pp. 16–22, Feb. 2024, doi: 10.37905/jjom.v6i1.21921.

- [8] Y. Chen *et al.*, “Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method,” *Landsc. Urban Plan.*, vol. 160, pp. 48–60, Apr. 2017, doi: 10.1016/j.landurbplan.2016.12.001.
- [9] H. Izakian, W. Pedrycz, and I. Jamal, “Fuzzy clustering of time series data using dynamic time warping distance,” *Eng. Appl. Artif. Intell.*, vol. 39, pp. 235–244, Mar. 2015, doi: 10.1016/j.engappai.2014.12.015.
- [10] P. J. F. Borg, I.; Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005. doi: 10.1007/0-387-28981-X.
- [11] G. Hyndman, R. J. and Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. OTexts, 2021.
- [12] G. M. Box, G. E. P. and Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.
- [13] C. W. Borovoykh, A., Bohte, S., and Oosterlee, “Conditional time series forecasting with convolutional neural networks,” *arXiv preprint*, vol. arXiv:1703, 2017.
- [14] T. Chakraborty, S. Chattopadhyay, and I. Ghosh, “Forecasting dengue epidemics using a hybrid methodology,” *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121266, Aug. 2019, doi: 10.1016/j.physa.2019.121266.
- [15] I. M. W. A. S. K. R. Embay. Sumertajaya, “Vector Autoregressive-Moving Average Imputation Algorithm for Handling Missing Data in Multivariate Time Series,” *IAENG Int. J. Comput. Sci.*, vol. 5, no. 2, pp. 727–735, May 2023.
- [16] D. Tsabitah, Y. Angraini, and I. M. Sumertajaya, “Implementation of Clustering Time Series with DTW to Clustering and Forecasting Rice Prices Each Provinces in Indonesia,” *Inferensi*, vol. 8, no. 1, p. 13, Mar. 2025, doi: 10.12962/j27213862.v8i1.21952.
- [17] G. G. Ghiffary, E. D. D. Yanuari, K. A. Notodiputro, Y. Angraini, and L. N. A. Mualifah, “COMPARATIVE PERFORMANCE OF SARIMAX AND LSTM MODEL IN PREDICTING IMPORT QUANTITIES OF MILK, BUTTER, AND EGGS,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 1, pp. 407–418, Jan. 2025, doi: 10.30598/barekengvol19iss1pp407-418.
- [18] A. Ostermann, A. Bajrami, and A. Bogensperger, “Short-term forecasting of German generation-based CO2 emission factors using parametric and non-parametric time series models,” *Energy Informatics*, vol. 7, no. 1, p. 2, Jan. 2024, doi: 10.1186/s42162-024-00303-9.
- [19] J. (1994) Berndt, D. J., & Clifford, “Using dynamic time warping to find patterns in time series. In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (AAAIWS’94)*, 1994, pp. 359–370.
- [20] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling*. New York, NY: Springer New York, 2005. doi: 10.1007/0-387-28981-X.
- [21] S. Bai, J. Z. Kolter, and V. Koltun, “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling,” 2018, [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [22] Peter J. Huber; Elvezio Ronchetti, *Robust Statistics*, 2nd edition. Hoboken, New Jersey: John Wiley & Sons, 2009.
- [23] P. D’Urso, L. De Giovanni, and R. Massari, “Robust fuzzy clustering of multivariate time trajectories,” *International Journal of Approximate Reasoning*, vol. 99, pp. 12–38, Aug. 2018, doi: 10.1016/j.ijar.2018.05.002.
- [24] G. P. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.
- [25] D. U. Tsabitah, Y. Angraini, and I. M. Sumertajaya, “Spatiotemporal Clustering of Key Food Commodity Prices Using Multivariate Time Series,” *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 638–652, Dec. 2025, doi: 10.59395/ijadis.v6i3.1422.

- [26] E. A. Maharaj, P. D'Urso, and J. Caiado, *Time Series Clustering and Classification*. Chapman and Hall/CRC, 2019. doi: 10.1201/9780429058264.
- [27] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011, doi: 10.1016/j.patcog.2010.09.013.