VARIANSI: Journal of Statistics and Its Application on Teaching and Research **ISSN 2684-7590 (Online)**
Vol. 08 No. 02 (2026), 167-178

DOI: *10.35580/variansiunm463*

Implementation of the Hierarchial Clustering Method to Predict the Number of Domestic Tourist Trips in East Java Province

Dara Agustina[1]* & Muayyad[2]

[1*] *Department of Information System, Faculty of Informatics and Computer, Institute of Technology and Business STIKOM Bali, Indonesia*
[2] *Department of Business Administration, Faculty of Business Administration, Polytechnic of Bali, Indonesia*

Keywords: Tourism, Domestic Tourists, Cluster, ARIMA Model

Abstrack:

Indonesian tourism represents a wealth of natural and cultural assets that play a vital role in boosting state revenue. Data from Statistics Indonesia indicates a shift in the travel patterns of domestic tourists following the COVID-19 pandemic in 2022. East Java Province emerged as the primary destination for domestic travelers within Java. The province comprises 29 Regencies and 9 Cities, each offering its own unique tourism appeal. In this study, the travel patterns of domestic tourists across these Regencies and Cities were categorized into three clusters. Subsequently, ARIMA modeling was employed to forecast the volume of domestic tourist trips in East Java for each cluster. This information provides an estimate of future domestic travel activity in the Province, thereby assisting the government in considering improvements to accessibility and tourism management, as well as the development of new tourism offerings. The ultimate goal is to maintain the quality of the tourism sector and foster a continued desire among tourists to revisit the region.

1. Introduction

Indonesia's tourism sector represents a wealth of natural and cultural assets that play a crucial role in boosting state revenue (Cooper, 2008). It is well known that tourism activities in Indonesia were previously suspended, impacting natural resources and the maintenance of tourism infrastructure due to activity restrictions during the COVID-19 pandemic. However, as the pandemic subsided, Indonesia's tourism sector began to recover and successfully overcame the earlier setbacks. So far, several Indonesian tourist destinations have attracted significant attention from both domestic and international travelers. While Bali is one of Indonesia's most famous destinations, the Country's natural beauty and cultural richness are not limited to that island alone; there are many other stunning destinations, including those on the island of Java.

According to data from Statistics Indonesia, the travel patterns of domestic tourists in 2022 showed a structural shift compared to pre-pandemic conditions. In 2019, West Java Province was the primary destination for domestic tourists, recording 107,451,428 trips, followed by East Java Province (99,579,835 trips) and Central Java Province (59,838,282 trips). The situation was different in 2022, with East Java Province emerging as the top destination for domestic tourists, recording 198,913,339 trips, followed by West Java Province (128,667,116 trips) and Central Java Province (103,991,668 trips). This is positive news for East Java's tourism sector, positioning it to attract even more interest from travelers, not only domestic tourists but also international visitors as well.

* Corresponding author.
E-mail address: daraagustina@stikom-bali.ac.id

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

East Java Province is located in the eastern part of Java Island, Indonesia. East Java Province covers an area of 48,033 km², comprising 29 Regencies and 9 Cities. Tourism in the Province is currently experiencing significant growth. Data released by Statistics Indonesia in March 2025 recorded 16,430,577 domestic tourist visits to East Java. This dataset not only covers the province as a whole but also provides information on the number of domestic tourist trip for each Regency and City within the Province. On March 2025, Pasuruan City recorded the lowest number of domestic tourist trips, while Surabaya City recorded the highest. One key factor influencing a tourist's decision to visit a destination is its appeal whether based on natural beauty, culture, or history. Other considerations include the quality of supporting facilities and services such as accommodation, restaurants, and entertainment as well as the ease of transportation to the site. Ancillary services like banks, travel agencies, and consulting services also factor into tourist decisions. Consequently, the volume of domestic tourist trips varies across the different Regencies and Cities of East Java due to these factors. In statistics, these varying travel patterns can be categorized using grouping techniques; one such approach is the clustering method. Clustering is the process of grouping data so that items within a single cluster share maximum similarity, while items across different clusters share minimum similarity (Tan, 2006). Cluster analysis aims to map the distribution of domestic tourist travel patterns across East Java, with hierarchical clustering serving as one possible methodological approach. Beyond grouping, statistical methods can also be employed to forecast the volume of domestic tourist trips in the Province. Thus, having an overview of the volume of domestic tourist trips across East Java Province can assist the Government in planning and managing tourism development more effectively and efficiently in the future.

2. Literature Review

2.1 National Development Economy

Tourism in Indonesia has demonstrated significant growth over the last few decades, driven by unparalleled natural beauty and cultural richness (Bonaraja Purba et al., 2024). Tourism has proven to play a vital role in supporting regional economies. The tourism industry across various regions generates significant positive impacts on economic development, such as creating jobs, fostering new business opportunities, and increasing regional revenue (Hamzah and Hermawan, 2018).

2.2 Cluster

Clustering is a technique used to group data into several clusters based on similarities between data points. This analysis aims to identify groups exhibiting similar characteristics, thereby enabling the discovery of underlying relationships and structures within the data. The distribution patterns revealed by cluster analysis can indicate groups of data with uniform characteristics that are either distinct or concentrated in specific areas. Based on cluster analysis methods considering the presence of occurrence locations with or without specific data information regarding distribution patterns can be obtained, categorizing them as clustered, random, or dispersed.

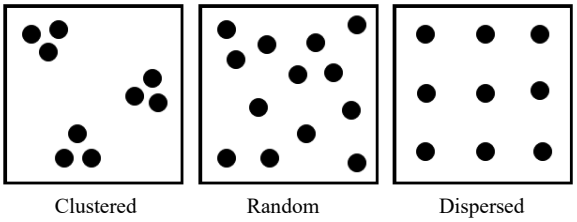


Figure 1: Cluster Distribution Patterns

Several clustering methods are available, one of which is hierarchical clustering. Hierarchical clustering begins the grouping process with two or more objects that share the greatest similarity. The process then proceeds to other objects based on the next level of proximity. This continues until the clusters form a "tree-like" structure, establishing

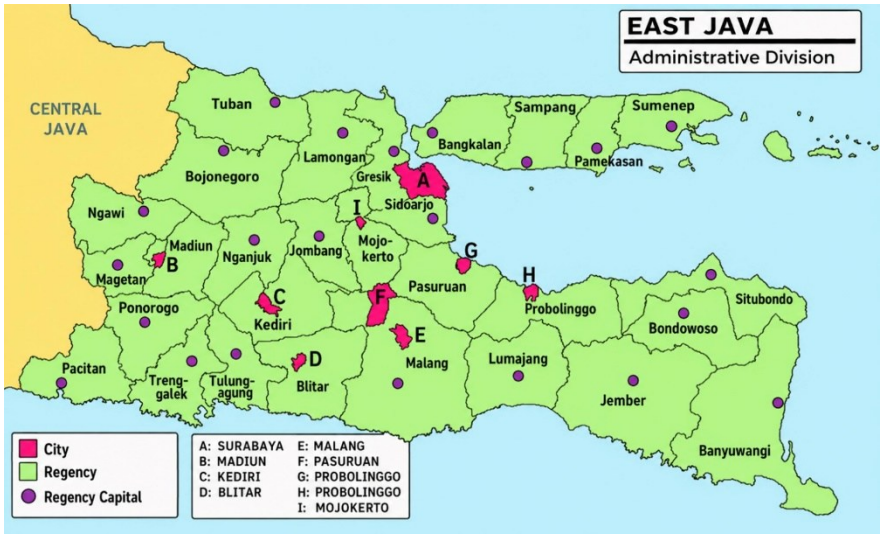
a clear hierarchy (levels) among objects, ranging from the most similar to the least similar. Hierarchical clustering offers significant advantages in data visualization and the ability to handle non-linear relationships and time-dependent patterns, making it suitable for grouping time-series data (Aghaborzogi et al., 2014).

2.3 Time Series Data Forecasting

Time series forecasting is a method used to predict future values over several periods based on a constructed time series model. The time series model employed is the Autoregressive Integrated Moving Average (ARIMA). According to Wei (2006), the ARIMA model is generally denoted as $ARIMA(p,d,q)$, where p represents the AR order, d the differencing order, and q the MA order. The accuracy of a forecasting model is crucial for obtaining precise predictions. This accuracy can be determined by the lowest value obtained from various data accuracy metrics; the lower the value, the more accurate the model's predictions. In statistics, the methods used rely on the average of a function representing the difference between the actual value and the forecasted value. This difference is known as the residual (Arsyad, 2001). One such method is the Mean Absolute Percentage Error (MAPE).

3. Methodology

3.1 Data



Source: https://jatim.bpk.go.id/peta-wilayah/

Figure 2: Map of East Java Province

This study was conducted in East Java Province, which comprises 29 Regencies and 9 Cities, as shown in Figure 2. The data used are secondary data obtained from the website of the Statistics Indonesia of East Java Province. Statistics Indonesia routinely uploads monthly data on the number of domestic tourist trips. The data utilized covers the period from January 2019 to March 2025 for the 29 Regencies and 9 Cities in East Java Province.

3.2 Analysis Steps

There are 38 regions requiring modeling to predict the number of domestic tourist trips in East Java Province; therefore, grouping the regions can enhance research efficiency. Previous studies have demonstrated the benefits of using clustering techniques. Magriaty et al. (2022) identified waste management zones in Tapin Regency, South Kalimantan Province, using k-means cluster analysis. Wardani et al. (2013) sought solutions for tuberculosis control in Indonesia using spatial-based cluster analysis. Deliar et al. (2023) employed a spatial clustering method to analyze land cover change patterns in West Java Province.

This study adopts a feature-based approach to clustering time-series data. This feature-based approach aims to represent the dynamic structure of each series using lower dimensional features, thereby enabling the reduction of inherently high dimensional time series data. Serial features are extracted using the autocorrelation function (ACF), and clustering is performed based on ACF values within each group to ensure an efficient clustering procedure (Montero and Vilar, 2015).

The process involves calculating the distance or dissimilarity measure between data points; the specific dissimilarity measure employed is correlation distance, calculated using the following equation (Golay et al., 1998):

$$COR(x_t, v_t) = \frac{\sum_{t=i}^{T}(X_t - \bar{X}_T)(V_t - \bar{V}_T)}{\sqrt{\sum_{t=1}^{T}(X_t - \bar{X}_T)^2}\sqrt{\sum_{t=1}^{T}(V_t - \bar{V}_T)^2}} \quad (1)$$

where $\bar{X}_T$ and $\bar{V}_T$ are the mean of X_t and V_t, with the following distance matrix

$$d_{COR(x_t,v_t)} = \sqrt{2\big(1 - COR(x_t, v_t)\big)} \quad (2)$$

Next, hierarchical clustering is performed using the ward linkage method. The ward linkage method is defined as follows

$$d_{(A,B)} = \frac{n_A n_B}{n_A + n_B}\left(\bar{y}_i - \bar{y}_j\right)^2 \quad (3)$$

with $d_{(A,B)}$ is the distance between objects y_i in cluster A and object y_j in cluster B. While n_A and n_B represents the number of objects in cluster A and cluster B. The dissimilarity measure is evaluated by calculating the cophenetic correlation coefficient; a higher cophenetic correlation coefficient indicates a better dissimilarity measure (Sokal and Rohlf, 1962).

Determining the number of clusters or the k optimal value can be achieved by calculating the silhouette coefficient and the dispersion ratio. A higher silhouette value indicates good cluster quality (Kaufman and Rousseeuw, 2009). In addition to the silhouette coefficient, determining the optimal number of clusters can also involve comparing within cluster standard deviation with between cluster standard deviation as measures of within cluster and between cluster dispersion (Kalkstein et al., 1987). This allows for the identification of cluster members for each k selected.

For each selected k cluster, the cluster representative value is calculated using a technical approach to calculate the average for each time series point in that cluster. So that the value of the number of domestic tourist trips which is representative of the cluster (prototype) is obtained, which will later be continued in the time series analysis modeling process for each cluster.

ARIMA modeling is used to predict prototypes for the number of domestic tourist trips in each cluster. Forecasting is carried out using the ARIMA method (Box and Jenkins, 1970). ARIMA is a statistical method that utilizes historical data patterns to produce forecasts. ARIMA combines an autoregressive (AR) model which explains the relationship between the current value and the next value, an integrated model which involves differencing to reach stationary and a moving average (MA) model which explains the current residual value with past residuals. By utilizing the properties of AR and MA, ARIMA can capture patterns and trends in groupings. The general equation for ARIMA modeling is as follows.

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)e_t \quad (4)$$

where $\phi_p(B)$ is the coefficient of parameter p of the AR model, $\theta_q(B)$ is the coefficient of parameter q of the MA model, B is backshift operator, $(1 - B)^d$ is differencing process, Y_t is the actual data at the t period, and e_t is residual at the t period.

The accuracy of the ARIMA model intended for forecasting is assessed using MAPE. The MAPE value is obtained by calculating the percentage difference between the actual values and the model's predicted values. MAPE provides insight into the model's ability to generate accurate predictions. MAPE is formulated as follows (Montgomery et al., 2008).

$$MAPE = \frac{\sum_{t=1}^{T}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|}{T} x100\% \quad (5)$$

where Yt represents the actual value at the t period, $\hat{Y}_t$ represents the predicted value at the t period, and T represents the number of observations in the time series data. A model is considered accurate if it achieves a MAPE value below 10% (Lewis, 1982).

4. Results

4.1 Exploration of Domestic Tourist Travel Data

Data exploration was conducted by generating time-series graphs to observe patterns in the volume of domestic tourist trips within East Java Province over time. Generally, the travel patterns across the regencies and cities of East Java Province from January 2019 to March 2025 show a similar trend, characterized by an annual increase. Certain Regencies and Cities notably the City of Surabaya experienced a sharper rise in travel volume compared to others. Conversely, some areas saw only modest growth in tourist travel; the City of Pasuruan is one such example.

These varying travel patterns across Regencies and Cities serve as the basis for grouping the data within East Java Province. Figure 3 displays the graphs showing the volume of domestic tourist trips for the 28 Regencies and 9 Cities across East Java Province from January 2019 to March 2025.

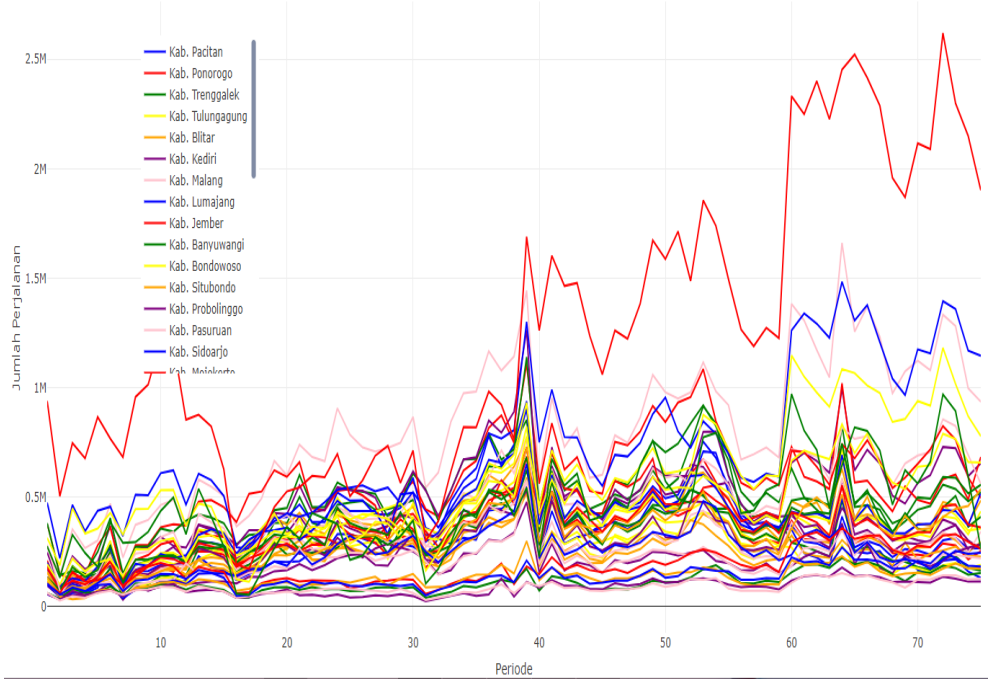


Figure 3. Plot of Domestic Tourist Travel in East Java Province

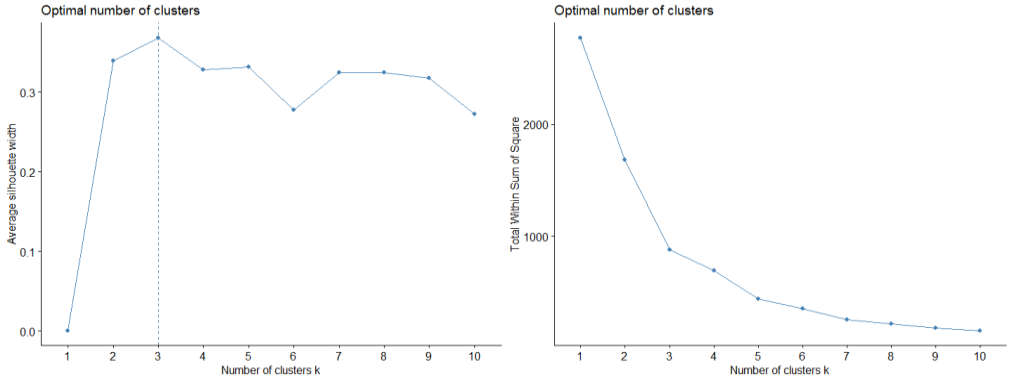
4.2 Clustering of Regencies/Cities in East Java Province

In this study, the clustering method employed is hierarchical clustering, testing all linkage methods with correlation distance used as the measure of dissimilarity. Based on the analysis results, the comparative linkage values presented in Table 1 were obtained.

Table 1. Cophenetic Correlation Coefficient Value

Linkage	Cophenetic Correlation
Single Linkage	0.8462
Average Linkage	0.8913
Complete Linkage	0.9065
Ward Linkage	0.9212

One measure of the quality of time-series data clustering is the cophenetic correlation coefficient, which represents the correlation between the cophenetic distances derived from the dendrogram and the original object distances used to construct it. The correlation value ranges from $-1 \leq r \leq 1$, cophenetic correlation value approaching 1 indicates a satisfactory clustering process. The highest cophenetic correlation coefficient, achieved using the ward linkage method, was 0.92 signifying that the clustering of domestic tourist trip volumes in East Java Province based on ACF values for each regency and city was effective. Furthermore, the optimal number of clusters can be determined based on the maximum silhouette coefficient and the minimum variance ratio.



(a) Silhouette Coefficient (b) Diversity Ratio

Figure 4. Plot of Silhouette Coefficient and Diversity Ratio

One method for determining the optimal number of clusters involves the subjective interpretation of silhouette coefficients across various possible cluster counts. The silhouette coefficients shown in Figure 4(a) indicate a maximum value at $k = 3$. Additionally, Figure 4(b) displays the variance ratio for cluster counts ranging from $k = 3$ to $k = 10$. Consequently, based on both the silhouette and variance coefficients, the optimal number of clusters is determined to be $k = 3$. Data regarding the volume of domestic tourist trips across regencies and cities in East Java Province will be grouped into these three clusters. The members of each cluster are presented in Figure 5.

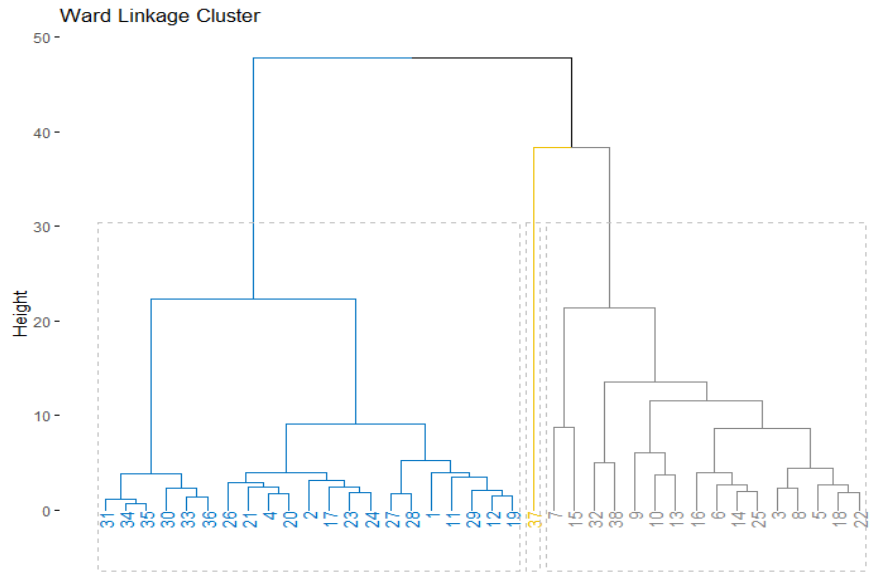


Figure 5. Dendrogram of Clustering Results

The clustering results align with field data. Three clusters were identified; one includes the City of Surabaya, which indeed records a high volume of domestic tourist travel compared to other Regencies or Cities. This outcome stems from the fact that the field data on domestic tourist travel was derived from tourist arrival figures, and Surabaya City possesses an actively operating airport. The formation of the other two clusters was influenced by the types of tourism and the accommodation facilities available in the vicinity of those tourist sites. Details regarding the members of each cluster are presented in Table 2.

Table 2. Details of Cluster Members

Cluster	Number of Member	Members
1	21	Blitar City, Pasuruan City, Mojokerto City, Kediri City, Probolinggo City, Madiun City, Bangkalan Regency, Ngawi Regency, Tulungagung Regency, Magetan Regency, Ponorogo Regency, Jombang Regency, Tuban Regency, Lamongan Regency, Sampang Regency, Pamekasan Regency, Pacitan Regency, Bondowoso Regency, Sumenep Regency, Situbondo Regency, Madiun Regency
2	1	Surabaya City
3	16	Malang Regency, Sidoarjo Regency, Malang City, Batu City, Jember Regency, Banyuwangi Regency, Probolinggo Regency, Mojokerto Regency, Kediri Regency, Pasuruan Regency, Gresik Regency, Trenggalek Regency, Lumajang Regency, Blitar Regency, Nganjuk Regency, Bojonegoro Regency

4.3 Time Series Data Forecasting Based on Formed Clusters

This study utilizes data on the number of domestic tourist trips in East Java Province from January 2019 to March 2025, covering 29 Regencies and 9 Cities. To validate the consistency of the ARIMA model, modeling was conducted using three different training datasets. The first training dataset spanned from January 2019 to December 2023; the second from January 2019 to May 2024; and the third from January 2019 to October 2024. Subsequently, modeling was performed on the domestic tourist trip data covering the period from January 2019 to March 2025 using the best selected model, followed by a forecast for the subsequent six months.

Modeling at the cluster level requires representative data for each cluster. These representatives were determined by calculating the average value of the members within a given cluster (creating a prototype). Using the `auto.arima` function in R software, it was found that for clusters 1, 2, and 3, the model consistently identified across all training datasets was ARIMA(1,1,0). The resulting model demonstrated good forecasting accuracy based on MAPE values. Detailed MAPE values for the modeling of each cluster across the three training datasets are presented in Table 3.

Table 3. MAPE Value from The Training Data of The Three Cluster

Training Data	Cluster 1	Cluster 2	Cluster 3
January 2019 – December 2023	19.1164	24.3085	18.7376
January 2019 – May 2024	18.9299	22.5708	18.4886
January 2019 – October 2024	18.4602	21.5430	18.0130

According to Jatmiko et al. (2017), the interpretation of MAPE calculation results indicates that a value in the 10%–20% range signifies "good forecasting," meaning the model is suitable for forecasting purposes. A value in the 20%–50% range signifies "reasonable forecasting," meaning the model is reasonably suitable for forecasting. Cluster 2, which includes the city of Surabaya exhibits a relatively high MAPE value compared to the other clusters due to highly volatile fluctuations in airport arrival data. Table 3 shows that the resulting model is suitable for forecasting, therefore all three clusters will be modeled using ARIMA(1,1,0).

The data for domestic tourist travel volume in East Java Province across clusters 1, 2, and 3 follows an ARIMA(1,1,0) model. The model's adequacy was tested by examining the residuals. It was found that for all models across the three clusters, the residuals were mutually independent as indicated by a Ljung-Box test that the p-value greater than the alpha level of 0.05. The residuals for all three clusters were not normally distributed, as shown by Kolmogorov-Smirnov test that the p-values of less than 0.05. The normality test is a secondary assumption check that affects the accuracy of confidence intervals rather than the fundamental quality of parameter estimation. Since the primary objective of this study is forecasting which can be adequately assessed through the white noise assumption test and the prediction error rate (MAPE) the failure to meet the normality assumption does not necessarily render the point prediction values inaccurate. Forecasting for the next six months will be conducted using this model based on domestic tourist travel data in East Java Province from January 2019 to March 2025. The forecasting accuracy (MAPE) using the ARIMA(1,1,0) model is 17.79 for cluster 1, 20.84 for cluster 2, and 17.73 for cluster 3, these accuracy results are considered good for all three clusters.

Domestic tourist travel patterns in East Java Province vary across the different Regencies and Cities. Field data reveals that Surabaya City exhibits travel patterns with significantly higher volumes compared to other Regencies and Cities. This is due to the presence of an active airport in Surabaya City, as Statistics Indonesia processes domestic tourist travel data based on domestic arrivals in each region. Consequently, the optimal grouping resulted in three clusters, with one cluster consisting of only a single city is Surabaya City.

The first cluster comprises 21 Regencies and Cities: Blitar City, Pasuruan City, Mojokerto City, Kediri City, Probolinggo City, Madiun City, Bangkalan Regency, Ngawi Regency, Tulungagung Regency, Magetan Regency, Ponorogo Regency, Jombang Regency, Tuban Regency, Lamongan Regency, Sampang Regency, Pamekasan Regency, Pacitan Regency, Bondowoso Regency, Sumenep Regency, Situbondo Regency, and Madiun Regency. Some of the famous tourist attractions in this cluster include Sarangan Lake in Magetan Regency, Baluran National Park and Ijen Crater in Bondowoso Regency, and White Sand Beach in Situbondo. This cluster encompasses both natural and cultural tourism. In reality, there are many other tourist attractions across these Regencies and Cities that have the potential to become famous. However, they suffer from a lack of promotion, difficult access, and inadequate infrastructure, meaning they remain known and visited only by local residents. This situation presents an opportunity for the government to promote these local attractions and improve accessibility. Forecasts for the number of domestic tourist trips in cluster 1 for the six months period from April to September 2025 are presented in Table 4.

Table 4. Domestic Tourist Travel Forecast Results for Cluster 1

Period	Forecast Value	Period	Forecast Value

April 2025	270,108.1	July 2025	274,248.5
May 2025	275,624.3	August 2025	273,453.7
June 2025	272,732.4	September 2025	273,870.4

The second cluster consists of a single City, Surabaya City due to the high volume of tourist travel to the area. This situation is driven by the presence of adequate infrastructure, specifically an actively operating airport. Forecasts for this cluster over the next six months project approximately 1 million travelers per month. Consequently, it is expected that the government will focus on further improving airport management. Furthermore, given the high volume of tourist arrivals in this cluster, it is essential to maintain Surabaya's facilities at an optimal level. Promoting local culinary offerings and nearby tourism attractions within Surabaya could also serve to inform tourists, thereby sparking their curiosity and interest in visiting other Regencies and Cities across East Java. Table 5 presents the travel forecasts for domestic tourists in cluster 2 for the six months period spanning April to September 2025.

Table 5. Domestic Tourist Travel Forecast Results for Cluster 2

Period	Forecast Value	Period	Forecast Value
April 2025	1,961,271	July 2025	1,949,728
May 2025	1,947,169	August 2025	1,949,919
June 2025	1,950,529	September 2025	1,949,874

The third cluster comprises 16 Regencies and Cities: Malang Regency, Sidoarjo Regency, Malang City, Batu City, Jember Regency, Banyuwangi Regency, Probolinggo Regency, Mojokerto Regency, Kediri Regency, Pasuruan Regency, Gresik Regency, Trenggalek Regency, Lumajang Regency, Blitar Regency, Nganjuk Regency, and Bojonegoro Regency. This cluster includes Regencies and Cities known for their prominent tourism offerings, featuring natural, cultural, and educational attractions suitable for family vacations. Accessibility in these major urban areas is excellent, attracting a large number of tourists to the region.

Notable attractions in cluster 3 include Mount Bromo (spanning Malang, Probolinggo, Lumajang, and Pasuruan Regencies) and Goa Cina Beach and Balekambang Beach in Banyuwangi Regency. Batu City offers popular family destinations such as Batu Night Spectacular, Jatim Park 1, 2, and 3, and Museum Angkut. Other attractions include Taman Safari in Pasuruan Regency and Papuma in Jember Regency. The projected number of tourist arrivals for this cluster is 500,000. This projection serves as a reference for tourism operators in these major Regencies and Cities to prioritize infrastructure and maintenance. New tourism attractions could also be developed to ensure the tourism sector remains in good condition and to maintain visitor interest in returning. Forecasts for the number of domestic tourist trips in cluster 3 for the upcoming six months period (April to September 2025) are presented in Table 6.

Table 6. Forecast Results for Domestic Tourist Trips in Cluster 3

Period	Forecast Value	Period	Forecast Value
April 2025	546,030.6	July 2025	543,826.7
May 2025	543,097.2	August 2025	544,119.1
June 2025	544,457.3	September 2025	543,983.5

5. Conclusion

Multivariate time series forecasting models can be made more efficient by first grouping the data. In this study, time series data were grouped using a feature-based approach employing correlation distance. The results indicated that while correlation distance successfully captured similarities in time series patterns, it did not account for the actual

magnitude or values of the data points. Consequently, future research could investigate alternative distance metrics or combinations of metrics capable of capturing both the patterns and the values of the time series data.

ARIMA modeling performed well for each cluster. Cluster forecasting results tended to track the actual data closely, with MAPE values ranging from 10% to 25%, indicating that cluster-level modeling is effective and representative for forecasting purposes. However, achieving a MAPE below 10% would be preferable. Therefore, future research could explore alternative approaches, such as determining ARIMA orders based on ACF and PACF plots and addressing model overfitting. Additionally, analyzing seasonal effects within the data patterns could be beneficial

East Java Province has significant potential to increase both domestic and international tourist visits. By classifying regions based on visitation trends and applying precise statistical models, the local government can formulate more targeted and efficient tourism development strategies. Initiatives such as digital marketing, infrastructure improvements, inter regional destination synergy within East Java, and the implementation of predictive technologies can serve as a solid foundation for the future advancement of the tourism sector.

Reference

Aghaborzogi, S., et al. (2014). *A Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique*. The Scientific World Journal, Vol. 2014(1): 1-12. https://doi.org/10.1155/2014/562194.

Arsyad, Lincoln. (2001). Peramalan Bisnis Edisi Pertama. Yogyakarta: BPFE.

C.D., Lewis. (1982). *International and Business Forecasting Methods*. London, UK: Butterworths.

Cooper, C., Fletcher, J., Fyall, A., Gilbert, D., dan Wanhill, S. (2008). Tourism: Principles and Practice, 4th Edition. Harlow: Pearson Education.

D.C., Montgomery, C.L., Jennings, dan M., Kulachi. (2008). Introduction of Time Series Analysis and Forecasting, 3rd Edition. New Jersey, US: John Wiley & Sons Inc.

Data of domestic tourist travel to Regencies/ Cities in East Java Province from January 2019 to March 2025, from Statistics Indonesia: https://jatim.bps.go.id/.

Deliar, A., Santoso, P.B.K., Virtriana, R. (2023). Analisis Pola Perubahan Tutupan Lahan Berdasarkan Metode Spatial Cluster di Provinsi Jawa Barat. Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi, Vol. 7 No. 1: 53-62. https://doi.org/10.29408/geodika.v7i1.7105.

G.E.P., Box and G.M., Jenkins. (1970). Time Series Analysis: Forecasting and Control. San Fransisco, CA: Holden-Day Inc.

Golay X., Kollias S., Stoll G., Meier D., Valavanis A., & Boesiger P. (1998). A New Correlation Based Fuzzy Logic Clustering Algorithm for FMRI. Magnetic Resonance in Medicine, Vol. 40(2): 249-260. https://doi.org/10.1002/mrm.1910400211.

Hamzah, F., dan Hermawan, H. (2018). Evaluasi Dampak Pariwisata Terhadap Sosial Ekonomi Masyarakat Lokal. Jurnal Pariwisata, Vol. 5 No. 3. https://ejournal.bsi.ac.id/ejurnal/index.php/jp/article/view/4434.

Jatmiko, Y.A., Rahayu, R.A., dan Darmawan, G. (2017). Perbandingan Keakuratan Hasil Peramalan Produksi Bawang Merah Metode *Holt-Winters* dengan *Singular Spectrum Analysis* (SSA). Jurnal Matematika ”Mantik”, Vol. 3 No. 1: 13-23. https://doi.org/10.15642/mantik.2017.3.1.13-22.

Kalkstein, L.S., Tan, G., dan Skindlov, J.A. (1987). *An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification*. Journal of Applied Meteorology and Climatology, Vol. 26(6): 717-730. https://doi.org/10.1175/1520-0450(1987)026<0717:AEOTCP>2.0.CO;2.

Kaufman, L., and Rousseeuw, P.J. (2009). Finding Groups in Data: an Introduction to Cluster Analysis. New Jersey, US: John Wiley & Sons Inc.

Magriaty, R., Murtilaksono, K., dan Anwar, S. (2023). Analisis k-Means Cluster untuk Identifikasi Kawasan Pengelolaan Sampah di Kaputaen Tapin di Provinsi Kalimantan Selatan. Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Wilayah dan Pendesaan), Vol. 7 No. 1: 79-90. https://doi.org/10.29244/jp2wd.2023.7.1.79-90.

Montero, P., dan Vilar, J.A. (2015). TSclust: An R Package for Time Series Clustering. Journal of Statistical Software, Vol. 62(1): 1-43. https://www.jstatsoft.org/article/view/v062i01.

Peta Pembagian Administrasi Wilayah Provinsi Jawa Timur dari Badan Pemeriksaan Keuangan Perwakilan Provinsi Jawa Timur: https://jatim.bpk.go.id/peta-wilayah/.

Purba, B., Situmorang, E.A.A., Firmansyah, D., Sidebang, T.B., dan Manurung, T.A. (2024). Potensi Sumber Daya Alam Dalam Mengembangkan Sektor Pariwisata di Indonesia. Economic Reviews Journal, Vol. 3 No. 3: 2060-2065. https://doi.org/10.56709/mrj.v3i3.308.

Sokal, R.R., dan Rohlf, F.J. (1962). The Comparison of Dendrograms by Objective Methods. Taxon, Vol. 11 No. 2: 33-40. https://doi.org/10.2307/1217208.

Tan, P.N., Steinbach, M., Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Education.

Wardani, D.W., Lazuardi, L., Kusnanto, H., et al. (2013). Pentingnya Analisis Cluster Berbasis Spasial Dalam Penanggulangan Tuberkulosis di Indonesia. Jurnal Kesehatan Masyarakat Nasional Vol. 8(4): 147-151. https://doi.org/10.21109/kesmas.v0i0.391.

Wei, William W.S. (2006). Time Series Analysis Univariate and Multivariate Methods, 2nd Edition. Canada: Pearson Eddison Wesley.

Appendix

Here are the codes of Regencies and Cities in East Java Province on Table 2 for the clustering dendrogram in Figure 5.

Number	Region	Number	Region
1	Pacitan Regency	20	Magetan Regency
2	Ponorogo Regency	21	Ngawi Regency
3	Trenggalek Regency	22	Bojonegoro Regency
4	Tulungagung Regency	23	Tuban Regency
5	Blitar Regency	24	Lamongan Regency
6	Kediri Regency	25	Gresik Regency
7	Malang Regency	26	Bangkalan Regency
8	Lumajang Regency	27	Sampang Regency
9	Jember Regency	28	Pamekasan Regency
10	Banyuwangi Regency	29	Sumenep Regency
11	Bondowoso Regency	30	Kediri City
12	Situbondo Regency	31	Blitar City
13	Probolinggo Regency	32	Malang City
14	Pasuruan Regency	33	Probolinggo City
15	Sidoarjo Regency	34	Pasuruan City
16	Mojokerto Regency	35	Mojokerto City
17	Jombang Regency	36	Madiun City
18	Nganjuk Regency	37	Surabaya City