ENSEMBLE
Journal of Applied Mathematics, Statistics, and Data Science
eISSN: xxxx-xxxx
Vol. 1, No. 1, June 2026
Page: 29-34

Fuzzy Time Series K-Medoids for Inflation Forecasting in West Sulawesi Province

Ardi[1], Muh. Hijrah[2*], Fardinah[3]

[1,2,3]Study Program of Statistics, Faculty of Mathematics and Natural Science, Universitas Sulawesi Barat
Jl. Prof. Baharuddin Lopa, Majene, 91413, Indonesia

**E-mail Correspondence Author: ***muhhijrah@unsulbar.ac.id

doi: xxxxxx

ARTICLE INFO

Article history

Received : February 3, 2026
Revised : March 16, 2026
Accepted : April 21, 2026

Keywords
Forecasting;
Fuzzy Time Series;
Inflation;
K-Medoids;
K-Means;
West Sulawesi

Abstract

Inflation is a key indicator of macroeconomic that reflects the price stability of a region and directly influences purchasing power, business planning, and public policy. West Sulawesi, an agriculture- and fishery-dependent province, exhibits a highly fluctuating monthly inflation pattern that is difficult to capture with conventional linear forecasting models. This study proposes a Fuzzy Time Series (FTS) model combined with K-Medoids clustering to forecast monthly inflation in West Sulawesi Province using Statistics Indonesia (BPS) data from January 2019 to December 2024 (72 observations). K-Medoids clustering, initialized using the optimal number of clusters selected via the Pseudo-F statistic and validated with the Silhouette Coefficient, was used to partition the historical data into representative intervals prior to fuzzification. Fuzzy Logical Relationships (FLR) and Fuzzy Logical Relationship Groups (FLRG) were then constructed to model the transition pattern between fuzzy states, and defuzzification was performed to generate numerical forecasts. Model accuracy was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The optimal number of clusters was found to be k = 7, supported by the highest Pseudo-F value (325.28) and Silhouette Coefficient (0.574). The FTS K-Medoids model achieved an MAE of 0.33, an MSE of 0.18, and a MAPE of 1.45%, corresponding to a forecasting accuracy of 98.55%, which falls into the "very good" category. The out-of-sample forecast for January 2025 was 0.253%, indicating a stable and controlled inflation outlook. These results demonstrate that the FTS K-Medoids approach is an effective, distribution-free alternative for forecasting nonlinear and volatile regional inflation data.

1. INTRODUCTION

Inflation is one of the principal indicators used to evaluate the macroeconomic stability of a region. Sustained and generalized increases in the price of goods and services erode household purchasing power, distort business investment decisions, and complicate monetary policy planning. As a developing economy, Indonesia experiences considerable regional variation in inflation, driven by differences in economic structure, dominant sectors, and supply-chain dependency. West Sulawesi, a relatively young province whose economy is dominated by agriculture and fisheries, is particularly exposed to price shocks in food commodities such as rice, chili, and fish, compounded by limited transportation infrastructure and dependence on inter-regional supply.

Forecasting inflation accurately is essential for local governments to anticipate price shocks and design timely economic interventions. Classical time series models often assume linearity and a well-behaved probability distribution, assumptions that are frequently violated by inflation data exhibiting sharp, irregular fluctuations. The Fuzzy Time Series (FTS) approach, introduced by Song and Chissom, addresses this limitation by representing historical values as fuzzy linguistic terms rather than raw numerical values, making it well suited to nonlinear and uncertain data without requiring a specific distributional assumption. Previous applications of FTS to Indonesian inflation data have reported good to very good forecasting accuracy, with MAPE values ranging from approximately 9.5% to 11.7% depending on the interval-partitioning strategy used.

A key methodological choice in FTS is how the universe of discourse is partitioned into intervals, since this step strongly influences forecasting accuracy. This study incorporates K-Medoids clustering into the FTS framework to obtain a more representative, data-driven partitioning. Unlike K-Means, K-Medoids uses actual data points (medoids) as cluster centers, which makes it more robust to outliers and noise, both of which are common in monthly inflation series. Despite the growing body of FTS research in Indonesia, studies focusing specifically on West Sulawesi remain scarce, even though understanding the province's inflation dynamics is crucial given its structural dependence on volatile food-producing sectors.

This study therefore aims to: (1) forecast monthly inflation in West Sulawesi Province using an FTS K-Medoids model, and (2) evaluate the resulting forecast accuracy using MAE, MSE, and MAPE. The findings are intended to support local policymakers, researchers, and business actors in anticipating inflationary dynamics and formulating more adaptive strategies.

2. METHOD

2.1 Data and Study Period

The data used in this study are the monthly inflation rates (%) of West Sulawesi Province from January 2019 to December 2024, obtained from the official website of Statistics Indonesia (Badan Pusat Statistik/BPS). The dataset consists of 72 monthly observations, with the highest recorded value of 1.43% and the lowest of –1.44%.

2.2 Fuzzy Time Series (FTS)

FTS transforms a numerical time series into linguistic (fuzzy) values and models the transition pattern between them. The procedure applied in this study consists of five main stages, following Florencia and Lilis (2024):

(1) Definition of the universe of discourse $U = [D_{min} - D_1, D_{max} + D_2]$, where D_{min} and D_{max} are the minimum and maximum observed values and D_1, D_2 are small positive constants used to extend the range slightly beyond the observed extremes.

(2) Interval partitioning, where the number of intervals is initially estimated using Sturges' rule (1 + 3.3 log n) and interval length is computed as $(D_{max} - D_{min})$ / number of intervals; in this study, the interval structure was refined through K-Medoids clustering (Section 2.3) rather than equal-width partitioning alone.

(3) Fuzzification of the actual inflation data into linguistic values (fuzzy sets A_1 through A_7) based on cluster membership.

(4) Construction of Fuzzy Logical Relationships (FLR), of the form $A_i \rightarrow A_j$, describing the transition from the fuzzy state at month n to the fuzzy state at month n+1, followed by grouping of these relationships into Fuzzy Logical Relationship Groups (FLRG).

(5) Defuzzification, in which the forecast value F(t) is computed as the average of the linguistic values associated with the FLRG of the current state:

$$F(t) = (m_1 + m_2 + \cdots + m_n)/n$$

where $m_1, m_2, \ldots, m_n$ are the medoid values of the fuzzy states associated with a given FLRG and n is the number of relations within that group.

2.3 K-Medoids Clustering

K-Medoids partitions the dataset into k clusters represented by actual data points (medoids) rather than computed means, which increases robustness to outliers relative to K-Means. The clustering algorithm applied is as follows: (1) determine the number of clusters k; (2) randomly initialize k medoids; (3) assign each observation to its nearest medoid using Euclidean distance; (4) select new candidate medoids at random within each cluster; (5) recompute the total distance and calculate the swap cost S as the new total distance minus the old total distance; and (6) if S < 0, accept the new medoids and repeat steps 3–5, otherwise stop. Euclidean distance was computed as:

$$d = \sqrt{\sum(y_i - \hat{y}_i)^2}, i = 1, \dots, n$$

The optimal number of clusters k was determined by evaluating the Pseudo-F statistic across k = 2 to 10, computed as $Pseudo - F = \left(R^2/(k-1)\right)/\left((1-R^2)/(n-k)\right)$, where $R^2 = (SST - SSW)/SST$, SST is the total sum of squares, and SSW is the within-cluster sum of squares. The choice of k was cross-validated using the Silhouette Coefficient, $s(i) = \left(b(i) - a(i)\right)/max\left(a(i), b(i)\right)$, where $a(i)$is the average within-cluster distance and $b(i)$ is the average distance to the nearest neighboring cluster; values closer to 1 indicate stronger cluster structure.

2.4 Model Evaluation

Forecast accuracy was assessed using three standard error metrics:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

where y_i and $\hat{y}_i$ denote the actual and forecast values, respectively, and n is the number of observations. MAPE was interpreted using the classification: < 10% = very good, 10–20% = good, 20–50% = adequate, and > 50% = poor. All computations for cluster-number selection (Pseudo-F and Silhouette Coefficient) were performed in R using the pam() function from the cluster package, looping over k = 2 to 10.

3. RESULTS AND DISCUSSION

3.1 Descriptive Overview of Inflation Data

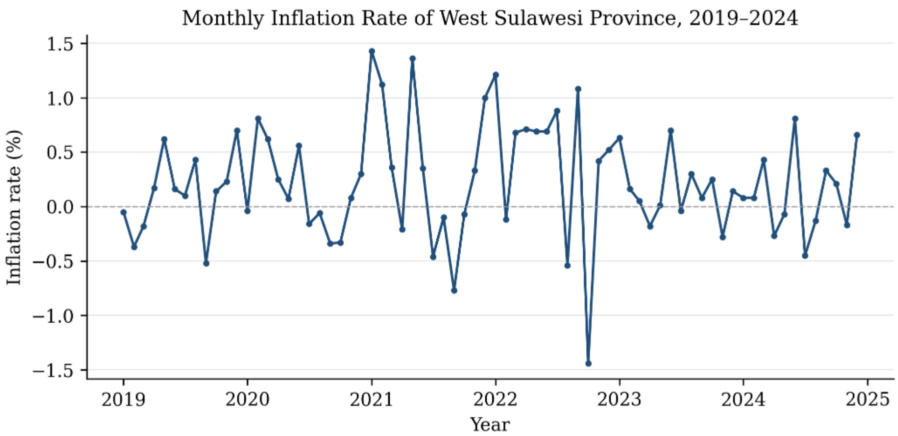

Figure 1. Monthly inflation rate of West Sulawesi Province, 2019–2024 (Source: BPS).

Figure 1 shows that monthly inflation in West Sulawesi from 2019 to 2024 fluctuated irregularly around zero, without a consistent long-term upward or downward trend. Several sharp deflationary episodes are visible, most notably in October 2022 (–1.44%), while the 2023–2024 period shows comparatively smaller swings, suggesting that inflation became somewhat more controlled toward the end of the observation window. This volatility, together with the absence of a clear linear trend, motivates the use of a fuzzy, cluster-based forecasting approach rather than a conventional linear time series model.

3.2 Determination of the Optimal Number of Clusters

The number of clusters used to partition the fuzzy intervals was selected by evaluating the Pseudo-F statistic for k = 2 through k = 10 (Table 1), and validated using the Silhouette Coefficient (Table 2).

Table 1. Pseudo-F and Silhouette Coefficient values for k = 2 to 10.

k	Pseudo-F	Silhouette Coefficient
2	115.6693	0.5619
3	122.1355	0.5095
4	121.6393	0.5208
5	163.2783	0.5390
6	157.8357	0.5485
7	325.2826	0.5738
8	313.9822	0.5527
9	294.9620	0.5495
10	318.0364	0.5734

Both criteria are maximized at k = 7 (Pseudo-F = 325.28; Silhouette Coefficient = 0.574), which is classified as a "moderate" cluster structure and clearly outperforms all other candidate values of k. Consequently, k = 7 was adopted as the number of clusters for the subsequent K-Medoids partitioning.

3.3 K-Medoids Clustering and Fuzzification

K-Medoids clustering with k = 7 was initialized using seven data points spanning the observed range of inflation values. After the first iteration, the total distance across all clusters was 5.69; a second iteration, using newly sampled candidate medoids, produced a total distance of 5.70. Since the swap cost S = 5.70 − 5.69 = 0.01 > 0, the algorithm terminated after the second iteration, and the medoids obtained at that step (−0.54, −0.17, 0.08, 0.30, 0.52, 0.70, 1.21, corresponding to fuzzy sets A_1 through A_7) were adopted as the final cluster centers.

Each of the 72 monthly observations was then fuzzified into one of the seven linguistic states (A_1–A_7) based on its nearest medoid. Fuzzy Logical Relationships (FLR) were derived from the sequential transitions between consecutive months, and these were grouped into Fuzzy Logical Relationship Groups (FLRG). Table 3 summarizes the FLRG structure, showing, for each current state, the frequency of transition to each subsequent state.

Table 2. Fuzzy Logical Relationship Groups (FLRG).

Current state	Next-state transitions	Total
A1	3(A1), 3(A2), 1(A3), 2(A4), 2(A5), 1(A7)	12
A2	2(A1), 4(A2), 4(A3), 1(A4), 3(A6), 1(A7)	15
A3	2(A1), 1(A2), 2(A3), 2(A4), 2(A5), 3(A6)	12
A4	1(A1), 3(A2), 1(A3), 1(A4), 1(A6), 2(A7)	9
A5	1(A1), 2(A2), 1(A5), 1(A6)	5
A6	2(A1), 4(A3), 1(A4), 5(A6)	12
A7	1(A1), 1(A2), 2(A4), 2(A7)	6

Defuzzification of each FLRG — computed as the average of the medoid values associated with the transitions in that group — produced the representative forecast value for each fuzzy state: A_1 = 0.067, A_2 = 0.144, A_3 = 0.220, A_4 = 0.272, A_5 = 0.068, A_6 = 0.253, and A_7 = 0.385. These seven values constitute the full set of possible point forecasts generated by the model.

3.4 Forecasting Results

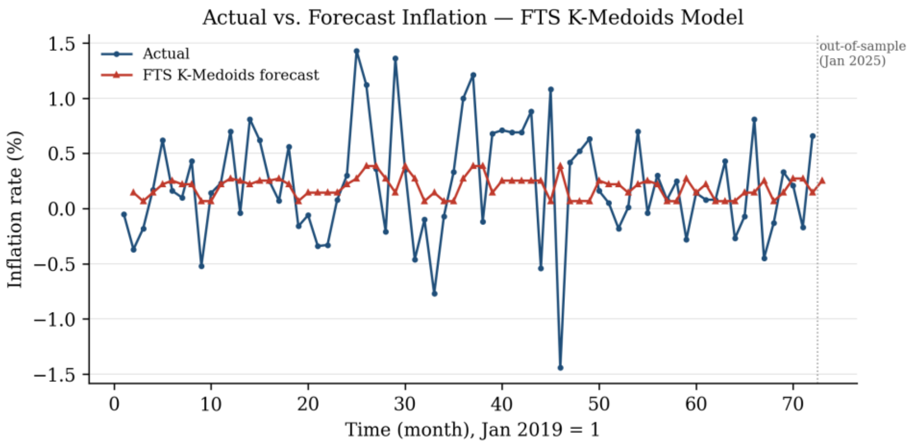

Figure 2. Actual vs. forecast inflation rate, FTS K-Medoids model (t = 1, Jan 2019; t = 73, Jan 2025 out-of-sample)

Figure 2 compares the actual monthly inflation values with the forecasts produced by the FTS K-Medoids model. The model tracks the general direction of the series reasonably well and captures its recurrent oscillation around zero, although, as expected for a model with only seven discrete output levels, it underestimates the magnitude of extreme spikes and troughs (e.g., the October 2022 deflation of –1.44%). The forecast for January 2025, generated from the FLRG associated with the final observed state (A_7, January 2022 pattern group), is 0.253%, indicating a low and stable inflation outlook for the province going into the new year.

3.5 Model Accuracy Evaluation

Table 4. Forecast accuracy of the FTS K-Medoids model.

MAE	MSE	MAPE
0.33	0.18	1.45%

As shown in Table 4, the FTS K-Medoids model achieved an MAE of 0.33, an MSE of 0.18, and a MAPE of 1.45%, corresponding to a forecasting accuracy of 98.55%. Following the MAPE classification of Rais et al. (2020), a value below 10% falls into the "very good" category, indicating that the combination of K-Medoids-based interval partitioning with the FTS framework is highly effective for modeling the nonlinear and volatile inflation pattern of West Sulawesi Province. This result is also favorable relative to prior FTS-based inflation forecasting studies in Indonesia, which have reported MAPE values in the range of 9.5%–11.7%, suggesting that data-driven, medoid-based interval construction offers a meaningful accuracy improvement over conventional equal-width partitioning.

4. CONCLUSION

This study applied a Fuzzy Time Series model combined with K-Medoids clustering to forecast monthly inflation in West Sulawesi Province using 72 months of BPS data (2019–2024). The optimal number of clusters, selected jointly via the Pseudo-F statistic and the Silhouette Coefficient, was k = 7. The resulting FTS K-Medoids model produced a forecast of 0.253% for January 2025 and achieved a MAPE of 1.45% (MAE = 0.33; MSE = 0.18), corresponding to a "very good" forecasting accuracy of 98.55%. These findings confirm that combining medoid-based clustering with fuzzy time series modeling is an effective, distribution-free strategy for forecasting nonlinear and volatile regional inflation data. Future research is encouraged to incorporate exogenous variables (e.g., staple food prices, exchange rates, interest rates), to benchmark the model against alternatives such as ARIMA, LSTM, or SVR through cross-validation, and to extend the observation period to further test the model's robustness to long-run structural change.

REFERENCES

[1] S. Amir, Fadilla, and P. Anggun, "Pengaruh inflasi terhadap pertumbuhan ekonomi Indonesia," Ekonomica Sharia, vol. 7, no. 1, pp. 17–28, 2021.

[2] F. F. Aulia and A. Defri, "Implementasi fuzzy time series logika Lee untuk peramalan inflasi di Indonesia," J. Ilm. Pendidik. Mat. Stat., vol. 4, no. 2, pp. 1083–1092, 2023.

[3] BPS, "Inflasi Sulawesi Barat Juni 2024 capai 2.32 persen," Badan Pusat Statistik Provinsi Sulawesi Barat, 2024.

[4] H. Dedy, Wanayumini, and S. D. Irfan, "Pengelompokan algoritma K-Means dan K-Medoid berdasarkan lokasi daerah rawan bencana di Indonesia dengan optimasi Elbow, DBI, dan Silhouette," Build. Informatics Technol. Sci., vol. 6, no. 2, pp. 1151–1158, 2024.

[5] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokkan data obat dengan Silhouette Coefficient," J. Appl. Informatics Comput., vol. 5, no. 2, pp. 109–116, 2021.

[6] A. D. Florencia and N. Lilis, "Implementasi metode fuzzy time series dalam peramalan penjualan produk unggulan perusahaan," JUTIN, vol. 7, no. 1, pp. 176–185, 2024.

[7] H. Gabriella and Matdoan, "Algoritma K-Medoids clustering untuk mengelompokkan tingkat kemiskinan pada kabupaten dan kota di Kepulauan Maluku dan Papua," J. Stat. Its Appl., vol. 4, no. 2, pp. 81–87, 2022.

[8] R. A. Mella and A. Ulil, "Optimasi portofolio saham menggunakan model mean-variance dan mean absolute deviation berdasarkan K-Medoids clustering," J. Mat. Stat. Komputasi, pp. 164–183, 2023.

[9] I. A. Putri, N. El Maidah, and M. A. Furqon, "Penerapan metode fuzzy time series Cheng pada peramalan inflasi di Indonesia," Komputika J. Sist. Komput., vol. 13, no. 2, pp. 183–191, 2024.

[10] A. N. Rais et al., "Evaluasi metode forecasting pada data kunjungan wisatawan mancanegara ke Indonesia," EVOLUSI J. Sains Manaj., vol. 8, no. 2, pp. 104–115, 2020.

[11] N. A. Siagian, A. Rikki, and P. B. N. Simangunsong, "Clustering menggunakan metode K-Medoids dengan pendekatan Manhattan distance," KAKIFIKOM, pp. 169–175, 2023.

[12] Q. Song and B. S. Chissom, "Fuzzy time series and its models," Fuzzy Sets Syst., vol. 54, no. 3, pp. 269–277, 1993.

[13] V. Vita, M. Shantika, and I. Nurfitri, "Penerapan fuzzy time series Chen average based," Bul. Ilm. Mat. Stat. Terapannya, vol. 10, no. 4, pp. 485–494, 2021.

[14] W. Wardani, "Analisis fluktuasi inflasi pertanian di Indonesia," J. Agribus. Sci., vol. 3, no. 2, pp. 100–104, 2020.

.