

Distribution-Based Fuzzy Time Series Markov Chain Models for forecasting Inflation in Bandung

Salsabila Ayu Pratiwi^{1, a)}, Dewi Rachmatin^{1, b)} and Rini Marwati^{1, c)}

¹Mathematics Study Program, Department of Mathematics Education, FPMIPA, Indonesia University of Education, Indonesia

^{a)} email: ayusalsabila958@gmail.com

^{b)} email: dewirachmatin@upi.edu

^{c)} email: rinimarwati@upi.edu

Abstract

This study discusses the application of the Fuzzy Time Series Markov Chain method which was developed by determining the length of the interval using the distribution method. In the fuzzy forecasting method, the determination of the length of the interval is an important thing that will affect the accuracy of the forecasting results. The development of this forecasting model aims to get better forecasting accuracy results. In this study, general inflation data for the city of Bandung is used for the period January 2016 – June 2021. The data is divided into two groups, namely in sample data and out sample data with a ratio of 90: 10. In the data processing process, the Python programming language is used. Based on the accuracy test using the MAPE method, it can be concluded that this method provides better forecasting results with a MAPE value of 1.16%.

Keywords: Fuzzy Time Series Markov Chain, Distribution Based Interval, MAPE, Inflation, Python.

Introduction

Time series forecasting according to Makridakis, et.al. [1] is a forecasting process based on historical data observations. The time series forecasting method is divided into two, forecasting techniques based on statistics and mathematical models and forecasting techniques based on artificial intelligence[2]. Conventional methods such as ARIMA require many conditions that must be met so that this method can be applied properly, while not all data can meet these conditions, therefore fuzzy forecasting techniques are a solution for this problem[3].

Fuzzy Time Series (FTS) is a forecasting concept that was first proposed by Song and Chissom in 1993 [4]. The development of fuzzy forecasting models generally focuses on achieving a high level of forecasting accuracy by increasing the three main stages, namely, fuzzification, defuzzification, and fuzzy inference [5]. The development of the fuzzy method was carried out by several researchers including the development of methods by adding fuzzy logic relation tables [6], development by adding weight to the fuzzy relation process and repetition of relations [7][8], development by induction of Markov chains [9]. Determination of the length of the interval in the fuzzy method is

crucial in the forecasting process because it will affect the accuracy of forecasting results. Among the methods for determining the length of the interval are intervals based on the mean and distribution, ratio, and granular computational approaches and entropy methods [1][10][11][12][13]. In several studies evaluating fuzzy forecasting models by comparing the Markov Chain method with the Chen and Weighted Markov Chain methods, from these studies the Markov Chain method gives the best results [14][15].

Fuzzy time series forecasting is widely used for solving forecasting problems. However, the fuzzy method specifically in the Markov Chain method has deficiency in determining arbitrary intervals. Some improve it with the struges formula method [15][16], the average method [17] distribution method. Based on this explanation, this study focuses on the application of the Distribution Based Fuzzy Time Series Markov Chain method in forecasting inflation.

Method

1. Fuzzy Time Series (FTS)

FTS is defined as follows : Let U be universe of discourse, where $U = \{u_1, u_2, \dots, u_n\}$, then fuzzy set A_i of u_i is defined as $A_i = \left\{ \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n} \right\}$, where $f_A(u_i)$ is the membership function u_i of A with $f_A(u_i) \in [0,1]$ and $1 \leq i \leq n$.

Definition 1. Let the universe of discourse $Y(t)$ with $(t = \dots, 0, 1, 2, \dots, n, \dots)$ is a subset of real number defined by fuzzy set A_i ($i = 1, 2, \dots$). If $F(t)$ consisted of A_i , $F(t)$ is defined as a fuzzy time series on $Y(t)$ ($t = \dots, 0, 1, 2, \dots, n, \dots$).

Definition 2. Suppose $F(t)$ is caused by $F(t-1)$, then the relation between $F(t)$ and $F(t-1)$ can be expressed as $F(t) = F(t-1) \circ R(t-1, t)$ where $R(t, t-1)$ is the relation matrix to describe the fuzzy relationship between $F(t-1)$ and $F(t)$, where ' \circ ' is the operator.

Definition 3. Suppose $F(t) = A_i$ is caused by $F(t-1) = A_j$ and the relation between $F(t)$ and $F(t-1)$ can be denoted $F(t-1) \rightarrow F(t)$ ($t = \dots, 0, 1, 2, \dots, n, \dots$), then Fuzzy Logical Relationship (FLR) is defined as $A_i \rightarrow A_j$. Where A_i is called the current stage and A_j is the next stage.

2. Distribution Based

Distribution based is an algorithm in FTS which is used to determine the length of intervals. Distribution based length can be determined by the following algorithm [10]:

- 1) Calculate all the absolute differences between A_{i+1} and A_i ($i = 1, \dots, n-1$) as the first differences and the average of the first differences.
- 2) According to the average, determine the base for length of intervals based by following Table 1.
- 3) Plot the cumulative distribution of the first differences. The base in step 2 is used as interval on the plot.
- 4) Choose the largest interval length that has a data value less than at least half the amount of the differences data.

Table 1. Base Mapping Table

Range	Base
0,1-1,0	0,1
1,1-10	1
11-100	10
101-1000	100

3. Distribution Based Fuzzy Time Series Markov Chain (DFTSMC)

1) Determine the universe of discourse U:

$$U = [D_{min} - B_1, D_{max} + B_2]$$

Where B_1, B_2 : positive real number, D_{min} : minimum data, D_{max} : maximum data

2) Partition interval U with the length according to the distribution algorithm.

3) Define the fuzzy set for each partition interval U with the following equation:

$$A_i = \left\{ \sum_{j=1}^n \frac{\mu_{ij}}{u_i} \right\} \quad (1)$$

$$\mu_{ij} = \{1, 0, 5, 0, i = j \quad j = i - 1 \text{ or } i = j - 1 \text{ other}$$

Where μ_{ij} : membership function

4) Fuzzified the data.

5) Determine Fuzzy Logical Relationship (FLR) and Fuzzy Logical Relationship Group (FLRG).

6) Determine the Markov transition probability matrix based on FLRG. The Markov transition probability matrix has dimension of $n \times n$, where n is the number of fuzzy sets. The state transition probability is formulated as follows:

$$P_{ij} = \frac{M_{ij}}{M_i}, i, j = 1, 2, \dots, n \quad (2)$$

Where P_{ij} : Probability of transition from state A_i to A_j , M_{ij} : Number of transitions from state A_i to A_j , M_i : Number of data included in state A_i .The transition probability matrix \mathbf{R} can be written as [9]:

$$R = [P_{11} \ P_{12} \ \dots \ P_{1n} \ P_{21} \ P_{22} \ \dots \ P_{2n} \ ; \ P_{n1} \ ; \ P_{n2} \ \ddots \ \dots \ ; \ P_{nn} \]$$

Definition 1. If $P_{ij} \geq 0$ then state A_j is accessible from state A_i .Definition 2. If state A_i and A_j are accessible to each other, then A_i communicates to A_j .

7) Calculate the initial forecast value with the following rules:

- If FLRG A_i is the empty set ($A_i \rightarrow \emptyset$), then $F(t) = m_i$, where m_i is midpoint of interval U_i .
- If FLRG A_i is one to one relation, then $F(t) = m_k P_{ik}$, where m_k is midpoint of interval U_k .
- If FLRG A_i is one to many relation, where data retrieved by $Y(t-1)$ at time $(t-1)$ is in state A_j , then $F(t) = m_1 P_{j1} + m_2 P_{j2} + \dots + m_{j-1} P_{j(j-1)} + Y(t-1) P_{jj} + m_{j+1} P_{j(j+1)} + \dots + m_n P_{jn}$, with $m_1, m_2, \dots, m_{j-1}, m_{j+1}, \dots, m_n$ are midpoint of interval $U_1, U_2, \dots, U_{j-1}, U_{j+1}, \dots, U_n$.

8) Calculate the value of the forecast adjustment with the following rules:

- If state A_i communicates to A_j , with $F(t-1) = A_i$ and makes an increasing transition to state A_j at time where $(i < j)$, then $D_{t1} = \frac{l}{2}$, where l : the length of interval.
- If state A_i communicates to A_j , with $F(t-1) = A_i$ and makes a decreasing transition to state A_j at time t where $(i > j)$, then $D_{t1} = -\frac{l}{2}$.
- If transition of state A_i to $F(t-1) = A_i$ and make a jump forward transition to state A_{i+s} at time t where $(1 \leq s \leq n - i)$ then $D_{t2} = \left(\frac{l}{2}\right)s, (1 \leq s \leq n - i)$, s the number of forward jumps.
- If transition of state A_i to $F(t-1) = A_i$ and make a jump backward transition to state A_{i-v} at time t where $(1 \leq v \leq i)$, then $D_{t2} = -\left(\frac{l}{2}\right)v, (1 \leq v \leq i)$, v the number of backward jumps.

9) Calculate the final forecast value with the following equation:

$$F'(t) = F(t) \pm D_{t1} \pm D_{t2} \tag{3}$$

4. Forecast accuracy

Determine forecasting accuracy using MAPE with the following formula:

$$MAPE = \frac{\sum_{t=1}^n \left| \left(\frac{Y_t - F_t}{Y_t} \right) \times 100\% \right|}{n} \tag{4}$$

Where F_t : forecast value, Y_t : actual data at time t , and n : number of data.

5. Research Design

In this study, secondary data for general inflation in Bandung is used from January 2016 – June 2021. The data is divided into two categories, in sample data or data to be used in the model and out sample data to be used as forecasting accuracy calculations. In the data processing, the Python programming language is used. The flow in the study is described in Figure 1.

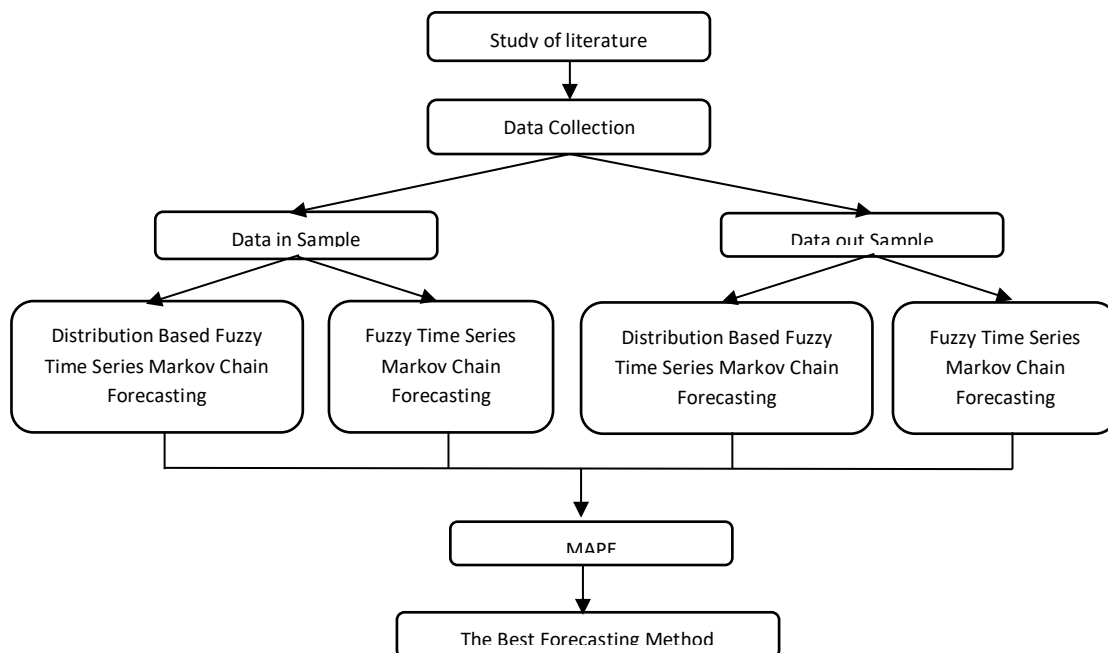


Figure 1. Research Design

Result and discussion

In this study, general inflation data in Bandung from January 2016 to December 2020 was applied as the in-sample data. The data are presented in Table 2.

Table 2. In Sample data of General Inflation in Bandung

Month	2016	2017	2018	2019	2020
January	0,53	0,49	0,83	0,09	0,38
February	-0,15	0,38	0,22	-0,08	0,35
March	0,2	-0,02	0,21	0,03	0,25
April	-0,17	0,1	0,27	0,43	0,16

Month	2016	2017	2018	2019	2020
May	0,24	0,47	0,22	0,84	-0,25
June	0,6	0,99	0,48	0,1	0,41
July	0,71	-0,27	0,17	0,55	-0,14
August	-0,49	0,06	-0,02	0,61	-0,1
September	0,14	0,11	-0,24	-0,28	-0,05
October	0,14	-0,03	0,5	-0,13	0,08
November	0,52	0,39	0,36	0,14	0,26
December	0,63	0,73	0,71	0,45	0,39

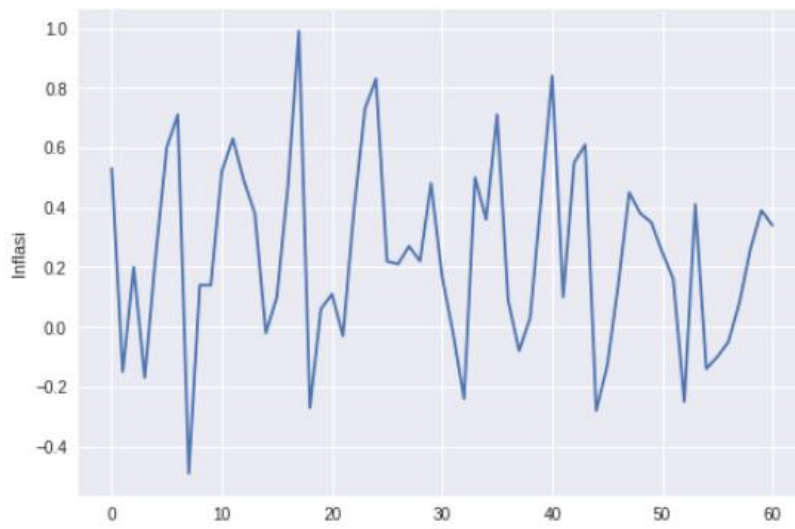


Figure 2. Training Data Plot of General Inflation in Bandung

Based on data in Table 2, the universe of discourse is $U = [D_{min} - B_1; D_{max} + B_2] = [-0,5 ; 1,1]$. Next, determine the length of interval according to distribution algorithm.

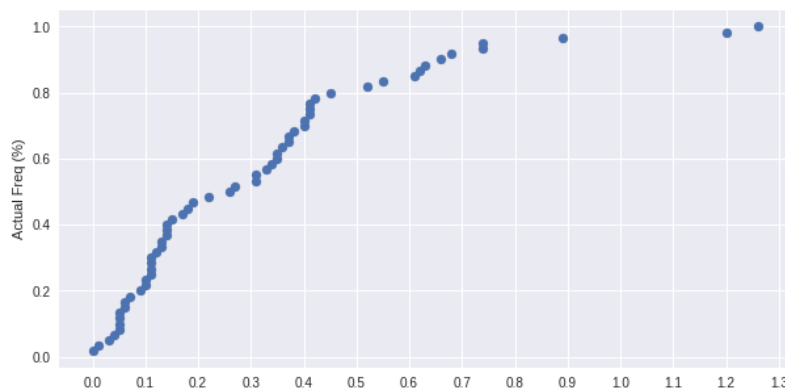


Figure 3. Cumulative Distribution Plot

Based on Figure 3, the largest interval length is 0.2. By using equation 1, the fuzzy set is obtained, then continue the process of fuzzification, FLR, and FLRG. After obtaining FLRG, a Markov

transition probability matrix is made where each element is a probability value obtained from equation 2. So that a Markov transition probability matrix of order 8 x 8 is obtained as follows:

$$R = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0,22 & 0,22 & 0,33 & 0,22 & 0 & 0 & 0 & 0 \\ 0,09 & 0,36 & 0,18 & 0,27 & 0,09 & 0 & 0 & 0 \\ 0,14 & 0,14 & 0,36 & 0,21 & 0,14 & 0 & 0 & 0 \\ 0 & 0,08 & 0,83 & 0,16 & 0,33 & 0 & 0,25 & 0,08 \\ 0 & 0,33 & 0 & 0,16 & 0,33 & 0,16 & 0 & 0,2 \\ 0 & 0,4 & 0,2 & 0 & 0,2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The process of calculating the initial forecast, the forecast adjustment value, and the final forecast value as described in the rules in the DSMC method. The results of the in-sample data forecasting are presented in Table 3.

Table 3. In Sample Data Forecasting Result

t	Actual Data	Initial Forecast Value	Adjustment Value	Final Value
1	0,53	-	-	-
2	-0,15	0,31	-0,4	-0,09
3	0,2	0,122222222	0,2	0,322222222
...
5	0,26	0,210909091	0,1	0,310909091
6	0,39	0,235714286	0,1	0,335714286

After doing the forecasting process on the in-sample data with t = 60, a looping process will be carried out by substituting the last forecast result as data to t = 61 as much as the number of out sample data. Forecasting results are presented in Table 4.

Table 4. Out Sample Data Forecasting Result

t	Actual Data	Initial Forecast Value	Adjustment Value	Final Value
60	0,39	0,235714286	0,1	0,335714286
61	0,34	0,372777778	0	0,372777778
62	0,43	0,347777778	0	0,347777778
63	0,4	0,392777778	0	0,392777778
64	0,44	0,377777778	0	0,377777778
65	0,43	0,397777778	0	0,397777778
66	-0,45	0,392777778	-0,4	-0,007222222

The result of the test is the MAPE value calculated based on equation 4 of the DFTSMC method compared to the MAPE value of the FTS Markov Chain method. The MAPE value of all data carried out by the DFTSMC method is 1.16%, the MAPE value of the out-sample data from the DFTSMC method is 2.2%. The MAPE value of all data carried out by the FTS Markov Chain method is 1.24% while the MAPE value of the out-sample data from the FTS Markov Chain method is 1.84%. The comparison of the two methods is presented in Figure 4.

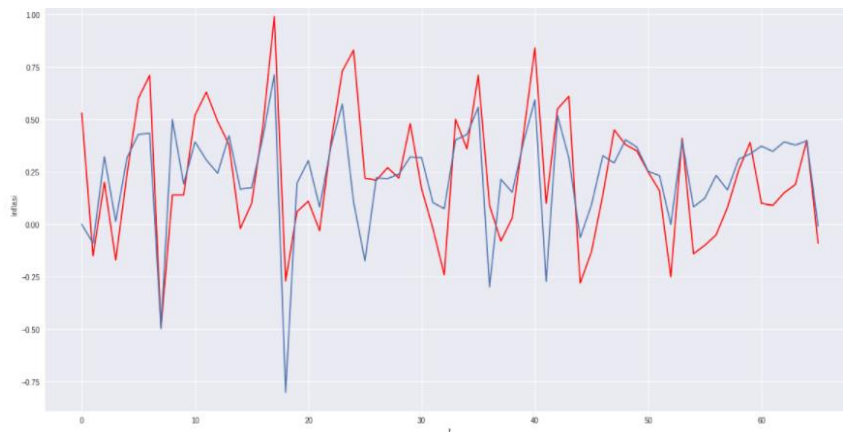


Figure 4. MAPE Plot

Based on Figure 4, forecasting using the DFTSMC method results in a data pattern following the actual data. However, the DFTSMC method produces forecasting values that tend to be constant while the original data has a fluctuating pattern.

Conclusions

Based on the results and discussion of the conclusions obtained from this study are the application of the DFTSMC method to forecasting inflation in the city of Bandung produces a data pattern that is close to the actual data pattern. The DFTSMC method has a better accuracy rate than the FTS Markov Chain with a MAPE value of 1.16%. The results of inflation forecasting in Bandung using the DFTSMC method provide the same data pattern as the previous data pattern.

References

- [1] S. Makridakis, S. C. Wheelwright, and V. E. McGee, "Metode dan aplikasi peramalan," *Jakarta: Erlangga*, 1999.
- [2] D. Ruhiat and C. Suwanda, "Peramalan data deret waktu berpola musiman menggunakan metode regresi spektral (studi kasus: debit sungai citarum-nanjung)," *Teorema Teor. dan Ris. Mat.*, vol. 4, no. 1, pp. 1–12, 2019.
- [3] M. Bose and K. Mali, "Designing fuzzy time series forecasting models: A survey," *Int. J. Approx. Reason.*, vol. 111, pp. 78–99, 2019.
- [4] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—Part II," *Fuzzy sets Syst.*, vol. 62, no. 1, pp. 1–8, 1994.
- [5] Y. Alyousifi, M. Othman, I. Faye, R. Sokkalingam, and P. C. L. Silva, "Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution," *Int. J. Fuzzy Syst.*, vol. 22, no. 5, pp. 1468–1486, 2020.
- [6] S. S. Gautam and S. R. Singh, "A refined method of forecasting based on high-order intuitionistic fuzzy time series data," *Prog. Artif. Intell.*, vol. 7, no. 4, pp. 339–350, 2018.
- [7] X. Sang, Q. Zhao, H. Lu, and J. Lu, "Weighted fuzzy time series forecasting based on improved fuzzy C-means clustering algorithm," in *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*, 2018, pp. 80–84.
- [8] O. C. Yolcu and F. Alpaslan, "Prediction of TAIEX based on hybrid fuzzy time series model with single optimization process," *Appl. Soft Comput.*, vol. 66, pp. 18–33, 2018.
- [9] A. T. Salawudeen, M. B. Mu'azu, E. A. Adedokun, and B. A. Baba, "Optimal Determination of Hidden Markov Model Parameters for Fuzzy Time Series Forecasting," *Sci. African*, p. e01174, 2022.

- [10] M. Pant, A. K. Shukla, and S. Kumar, "A novel method to optimize interval length for intuitionistic fuzzy time series," in *Soft Computing for Problem Solving*, Springer, 2021, pp. 55–62.
- [11] R. M. Pattanayak, H. S. Behera, and S. Panigrahi, "A novel probabilistic intuitionistic fuzzy set based model for high order fuzzy time series forecasting," *Eng. Appl. Artif. Intell.*, vol. 99, p. 104136, 2021.
- [12] D. Febrian, D. Kartika, and D. A. J. Nainggolan, "Peramalan Jumlah Wisatawan Mancanegara Yang Datang Ke Sumatera Utara Dengan Fuzzy Time Series," *KUBIK J. Publ. Ilm. Mat.*, vol. 6, no. 1, pp. 13–23, 2021.
- [13] F. D. D. Hakimi, M. N. A. Anshori, and A. Hanif, "Peramalan Kecepatan Angin yang Direkam oleh Sistem AWS dengan Analisis Fuzzy Time Series," *J. Kubik*, vol. 2, no. 2, 2017.
- [14] L. M. M. E. Un and M. T. Jatipaningrum, "Perbandingan Metode Fts-Chen Dan Fts-Markov Chain Untuk Memprediksi Curah Hujan Di Nusa Tenggara Timur," *J. Stat. Ind. dan Komputasi*, vol. 4, no. 2, pp. 1–10, 2019.
- [15] R. A. Kafi, Y. R. Safitri, Y. Widyaningsih, and B. D. Handari, "Comparison of weighted Markov chain and fuzzy time series Markov chain in forecasting stock closing price of company X," in *AIP Conference Proceedings*, 2019, vol. 2168, no. 1, p. 20033.
- [16] A. Aziz, K. N. I. Khomah, and S. P. Yohanes, "Prediction The Price of National Groceries Using Average Based Fuzzy Time Series With Song-Chissom and Markov Chain Approach," in *2018 International Conference on Applied Information Technology and Innovation (ICAITI)*, 2018, pp. 129–134.
- [17] S. Hariyanto, Z. Zaenurrohman, and T. U. SRRM, "Average Based-FTS Markov Chain Based on a Modified Frequency Density Partitioning to Predict COVID-19 in Central Java," *CAUCHY*, vol. 7, no. 2, pp. 231–239, 2022.