Fuzzy time series based on frequency density-based partitioning and k-means clustering for forecasting exchange rate

U S Mukminin\textsuperscript{1}, B Irawanto\textsuperscript{2}, B Surarso\textsuperscript{2}, Farikhin\textsuperscript{2}

\textsuperscript{1}Undergraduate Program of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
\textsuperscript{2}Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding author : uskarsabililmukminin@students.undip.ac.id, b_irawanto@yahoo.co.id

Abstract. Fuzzy Time Series (FTS) has been growing rapidly in recent years. There are many models that were developed. In this paper, we propose a new method to forecast exchange rate data by combining some models. Firstly, we use the average-based interval to make optimal interval numbers. Secondly, we use frequency density-based partitioning for optimal partitioning. In this part, we divide the three highest frequency of intervals into four, three, and two sub-intervals, respectively, and discarding intervals if there is no data distributed. And thirdly, we use k-means clustering to construct the Fuzzy Logical Relationship Group (FLRG). We divide Fuzzy Logical Relationship (FLR) into 16 initial clusters. Then we evaluate model by calculating the error value using MSE (Mean Squared Error) and AFER (Average Forecasting Error Rates). The study case of this paper is daily exchange rate data (USD to IDR) started from January until May 2020 that we got from www.bi.go.id. We also evaluate the forecasting data by calculating MSE (Mean Squared Error).

1. Introduction
Fuzzy time series method was firstly introduced by Song and Chissom [1] in 1993. They introduced a method for forecasting time series data in the future. The data that is used should be collected periodically based on days, months, years.

This method is growing rapidly in recent years. Some researches that were using fuzzy time series for forecasting TAIEX are [2] that using particle swarm optimization algorithm for optimal partition, and k-means clustering algorithm to build fuzzy logical relationship group. Another method for forecasting TAIEX is [3] Another developing application of fuzzy time series researched for forecasting stock index [4].

In this paper, we modify fuzzy time series for the forecasting exchange rate (USD to IDR) using frequency density-based partitioning for optimal partitions [5], using average-based interval for making an optimal interval [6], and modify the model by constructing the Fuzzy Logical Relationship Group (FLRG) using k-means clustering method [2].

Study case of this paper is the daily exchange rate of USD to IDR from January-May 2020 that we got from www.bi.go.id. We also evaluate the forecasting data by calculating MSE (Mean Squared Error)
value and AFER (Average Forecasting Error Rate) value. It aims to obtain a forecasting model of exchange rate data as preparation and evaluation for future conditions.

2. Fuzzy Time Series

The roughly fuzzy set can be interpreted as a number class with vague boundaries. Let \( U \) be a universe of discourse, where \( U = \{u_1, u_2, \ldots, u_n\} \), then a fuzzy set \( A_i \) in \( U \) with the membership function generally represented as follows:

\[
A_i = \mu_A(u_i) / (u_1) + \mu_A(u_2) / (u_2) + \ldots + \mu_A(u_p) / u_p
\]

(1)

Where \( \mu_A \) is a membership function of fuzzy set \( A_i \), \( \mu_A(u_i) \in [0,1] \) and \( 1 \leq i \leq p \).

Based on FTS that proposed by Song and Chissom [1], fuzzy time series is following concepts:

**Definition 2.1 [7]:**

A fuzzy set \( A \) of exchange rate \( X \), where \( X = \{x_1, x_2, \ldots, x_n\} \) defined as \( A = \{f_A(x_i)\}, i = 1, 2, \ldots, n \) where \( f_A \) is a membership function of fuzzy set \( A \), \( f_A : X \rightarrow [0,1] \). \( f_A(x_i) \) denotes membership degree of \( x_i \) in fuzzy set \( A \),\( 1 \leq i \leq n \).

**Definition 2.2 [5]:**

Let \( F(t) \) be a forecasted data of \( t \) time that is caused by \( F(t-1) \) which is noted by \( F(t-1) \rightarrow F(t) \) so, there is fuzzy relationship that is denoted \( F(t) = F(t-1) \odot R(t,t-1) \). Where \( R(t,t-1) \) is a fuzzy relationship and \( \odot \) is the max min composition operator.

**Definition 2.3 [7]:**

The membership function of fuzzy triangular is represented by 3 fuzzy parameters \( a = (a_1, a_2, a_3) \) where \( a_1, a_2, a_3 \in \mathbb{R} \). It curve is a combination of functions of the equations:

\[
\mu_a(x) = \begin{cases} 
\frac{x-a_1}{a_3-a_1}, & a_1 \leq x \leq a_2 \\
\frac{a_3-x}{a_3-a_2}, & a_2 \leq x \leq a_3 \\
0 & \text{else}
\end{cases}
\]

(2)

3. K-Means Clustering Algorithm

K-means clustering that firstly proposed by MacQueen [8] is one of the non-hierarchical data grouping method that can partition data into two or more groups. This method will partition the data into a group where the data with the same characteristics will be included in the same group, while the data that has different characteristics will be grouped into other groups. The K-Means algorithm is one of the most widely used algorithms in grouping because of simplicity and efficiency [9] and is recognized as one of the top 10 data mining algorithms by IEEE International Conference on Data Mining (ICDM) [10].

K-Means Clustering firstly was proposed by [8] with the following concepts:
Definition 2.4 [2] :

Let \( X = \{ x_1, x_2, x_3, \ldots, x_n \} \) be a set of data points and let \( C = \{ c_1, c_2, c_3, \ldots, c_k \} \) be a set of cluster centers, where \( k \) is a number of cluster. Then, it aims to minimizing an objective function \( J \) (Squared Error function) given by equation:

\[
J(C) = \sum_{j=1}^{n} \sum_{i=1}^{k} \left( \| x_i - c_j \| \right)^2
\]

where \( \| x_i - c_j \| \) is Euclidean distance, \( n_j \) is the number of data points of cluster \( j \), and \( k \) is the number of clusters.

Here the steps for forecasting data using frequency density based partitioning [7] and k-means clustering [2] :

Step 1 : Determine universe discourse \( U \). Then calculate the interval data using average-based interval [9]. Let \( X_{\min} \) and \( X_{\max} \) be the minimum and maximum datum of historical data, respectively. Then we define universe of discourse \( U \) as \( [X_{\min} - D_1, X_{\max} - D_2] \) where \( D_1 \) and \( D_2 \) are two proppers possitive numbers. Then partition the universe of discourse \( U \) into several intervals \( u_1, u_2, \ldots, u_n \) using average-based interval [6]. It follows several steps as follows :

1. Calculate the absolute mean of data

\[
|X| = \sum_{i=1}^{n} |\Delta_i|
\]

2. Calculate range value

\[
Range(r) = X_{\max} - X_{\min}
\]

3. Calculate basis by following form :

\[
Basis = \frac{|X|}{2}
\]

4. Calculate the number of intervals

\[
I = \frac{Range}{basis}
\]

Step 2 : Divide into sub-interval based on frequency density. Interval with the three highest frequency are divided into four, three, and two sub-intervals equal length, respectively. If there’s no data distributed in the intervals, it will be discarded [5].

Step 3 : Based on step 2, define fuzzy sets \( A_1, A_2, \ldots, A_6 \) and define the membership function of triangular fuzzy as shown as eq. 1.

Step 4 : Construct FLR and determine FLRG by using k-means algorithm [8]. Each group cluster has some fuzzy set members. We assume that FLR are shown as follows :
In FLR above we know that $A_{m_1}, A_{m_2}, A_{m_3}, \ldots, A_{m_p}$ are called current state and $A_{n_1}, A_{n_2}, A_{n_3}, \ldots, A_{n_p}$ are called next state. Then we cluster the subscripts of current state fuzzy sets that shown as $m_1, m_2, \ldots, m_p$ into several groups/ clusters $G_1, G_2, G_3, \ldots, G_k$.

**Step 5**: Calculate the distances $d_1, d_2, d_3, \ldots, d_k$ between $y$ and center clusters by the following equation:

$$d_i = |y - C_i|$$

Where $1 \leq i \leq k$, $y$ is a subscript of fuzzy set $A_y$, and $C_i$ is cluster center. Minimum distances from each $y$ to the center cluster show that $y$ is the member of the cluster. Then re-calculate the center of new clusters with new members until there is no data is reassigned.

**Step 6**: Calculate the forecasted fuzzy set $A_y$ using cluster centers as a subscript of the final forecasted fuzzy set $A_y$. Then calculate the result forecasted data using centroid method by the following forms [7]:

$$t_j = \begin{cases} 
\frac{\alpha_j + 0.5}{\alpha_j + \alpha_{j+1}}, & \text{if } j = 1 \\
\frac{0.5 + 1 + 0.5}{0.5 + \frac{1}{\alpha_j + \alpha_{j+1}}}, & \text{if } 2 \leq j \leq n - 1 \\
\frac{0.5 + 1}{\frac{0.5}{\alpha_j} + \frac{1}{\alpha_{j+1}}}, & \text{if } j = n 
\end{cases}$$

Where $t_j$ is forecasted data, and $\alpha_{j-1}, \alpha_j, \alpha_{j+1}$ are the middle value of interval forecasted fuzzy based on k-means clustering $A_{j-1}, A_j, A_{j+1}$, respectively.

**Step 7**: Calculate error value using MSE (Mean Square Error) and AFER (Average Forecasting Error Rates) by the following forms [7]:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (X_i - F_i)^2$$

$$AFER = \frac{\|X - F\|}{\bar{X}} \times 100\%$$

Where $X$ is an actual data, $F$ is a forecasted data, and $1 \leq i \leq 99$ with criteria of AFER is shown as table 1.
4. Numerical Simulation

In this session, we demonstrate the proposed method for exchange rate data of IDR and USD from January 2020 until May 2020 which can be accessed in www.bi.go.id. Here the step-by-step to forecast data.

Step 1: Define the universe of discourse $U$. From the data, we can see that $X_{\min} = 13680.06$ and $X_{\max} = 16824.71$ with $D_1$ and $D_2$ are 680.06 and 175.29 respectively. So, we get $U = (13000, 17000)$. From the calculation to find the optimal interval based on average-based interval [6], we get the universe of discourse $U$ partition into 71 intervals.

Step 2: We divide intervals into 46 sub-interval based on frequency density-based partitioning [5]. Table 2 shows sub-intervals of data.

### Table 2. Sub-interval

| Criteria | AFER Value | Criteria |
|----------|------------|----------|
| Poor     | $>50\%$   |         |
| Good     | $20\% - 50\%$ |         |
| Very good| $10\% - 20\%$ |         |
| Good enough | $<10\%$ |         |

Step 3: Define fuzzy sets $A_1, A_2, \ldots, A_{46}$ and define membership function of triangular fuzzy. For Example, membership function of $u_1$ can be made from interval [13672;13686] shown as figure 1.

$$
\mu_{A_1}(x) = \begin{cases} 
\frac{x - 13672}{13679 - 13672}, & \text{if } \quad 13672 \leq x \leq 13679 \\
\frac{13686 - x}{13686 - 13679}, & \text{if } \quad 13679 \leq x \leq 13686 \\
0, & \text{else}
\end{cases}
$$

(7)
**Step 4**: Construct FLR and determine FLRG by k-means clustering algorithm [8]. The FLR are divided into 16 initial clusters. For example, we have FLRG of Group $G_1$, which has membership FLR as follows:

$$
G_1 : A_1 \rightarrow A_{11} \\
A_1 \rightarrow A_{13} \\
A_3 \rightarrow A_{12} \\
A_{12} \rightarrow A_{12} \\
A_2 \rightarrow A_{10} \\
A_{10} \rightarrow A_{19}
$$

By the same way, we can determine FLRG based on k-means clustering of 16 initial clusters.

**Step 5**: Calculate the distances of each subscript of the current state of fuzzy sets. Re-calculate the center of new clusters with new members of each cluster until there is no data is re-assigned. Table 3 shows the final cluster centers.

| Clusters | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 | C13 | C14 | C15 | C16 |
|----------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|
| Centers  | 11 | 4  | 2  | 5  | 7  | 7  | 14 | 17 | 24 | 41  | 44  | 38  | 33  | 29  | 26  | 21  |
| Number of members | 8  | 9  | 4  | 7  | 4  | 5  | 4  | 5  | 5  | 7   | 5   | 5   | 7   | 9   | 5   | 6   | 8   |

**Step 6**: Calculate the forecasted fuzzy set $A_j$ using cluster centers as a subscript of the final forecasted fuzzy set $A_j$ using form 6.

**Step 7**: Calculate error value using MSE (Mean Squared Error) and AFER (Average Forecasted Error Rate). From the error calculation, we get MSE value 9197.613 and AFER value 0.414755% that represented by table 4. Fig. 2 shows a comparison graphic between actual data and forecasted data.
Figure 2. Graphic of forecasting result

Table 4. Forecasting result and error value

| Day | Actual Data \((X_i)\) | Forecasted Data \((F_i)\) | \((X_i - F_i)^2\) | \(\frac{|X_i - F_i|}{X_i}\) |
|-----|-----------------|-----------------|-----------------|-----------------|
| 1   | 13964.48        | 13962.46        | 36,4816         | 0.000432        |
| 2   | 13968.50        | 13962.46        | 7865,916        | 0.006321        |
| 3   | 14030.81        | 14119.5         | 683,2996        | 0.001869        |
| 4   | 13988.60        | 13962.46        | 1698,264        | 0.002943        |
| 5   | 14003.67        | 13962.46        | 1099,586        | 0.002381        |
| 6   | 13929.30        | 13962.46        | 6625,96         | 0.005864        |
| 7   | 13881.06        | 13962.46        | 510,76          | 0.00164         |
| 8   | 13776.54        | 13799.14        | 24025           | 0.000113        |
| 9   | 13722.27        | 13723.82        | 605,6521        | 0.001787        |
| 10  | 13774.53        | 13799.14        | 61009           | 0.00018         |
| 11  | 13726.29        | 13723.82        | 574564          | 0.000553        |
| 12  | 13716.24        | 13723.82        | 24025           | 0.000113        |
| 13  | 13722.27        | 13723.82        | 61009           | 0.00018         |
| 14  | 13726.29        | 13723.82        | 574564          | 0.000553        |
| 15  | 13746.39        | 13747.06        | 4489            | 4.87E-05        |
| 16  | 13694.13        | 13692.99        | 12996           | 8.32E-05        |
| 17  | 13700.16        | 13692.99        | 51409           | 0.000523        |
| 18  | 13680.06        | 13692.99        | 1671849         | 0.000945        |
| 19  | 13715.24        | 13723.82        | 73164           | 0.000626        |
| 20  | 13702.17        | 13692.99        | 842724          | 0.00067         |
| 21  | 13720.26        | 13723.82        | 126736          | 0.000259        |
| 22  | 13730.31        | 13747.06        | 2805625         | 0.00122         |
| 23  | 13794.63        | 13799.14        | 203401          | 0.000327        |
| 24  | 13828.80        | 13799.14        | 8797156         | 0.002145        |
| Day | Actual Data ($X_i$) | Forecasted Data ($F_i$) | $(X_i - F_i)^2$ | $\frac{|X_i - F_i|}{X_i}$ |
|-----|---------------------|--------------------------|-----------------|-------------------------|
| 25  | 13785,59            | 13799,14                 | 183,6025        | 0,000983                |
| 26  | 13730,31            | 13747,06                 | 280,5625        | 0,00122                 |
| 27  | 13715,24            | 13723,82                 | 73,6164         | 0,000626                |
| 28  | 13776,54            | 13799,14                 | 510,76          | 0,00164                 |
| 29  | 13754,43            | 13747,06                 | 54,3169         | 0,000536                |
| 30  | 13727,30            | 13723,82                 | 12,1104         | 0,000254                |
| 31  | 13747,40            | 13747,06                 | 0,1156          | 2,47E-05                |
| 32  | 13775,54            | 13799,14                 | 556,96          | 0,001713                |
| 33  | 13761,47            | 13747,06                 | 207,6481        | 0,001047                |
| 34  | 13744,38            | 13747,06                 | 7,1824          | 0,000195                |
| 35  | 13785,59            | 13799,14                 | 183,6025        | 0,000983                |
| 36  | 13803,68            | 13799,14                 | 20,6116         | 0,000329                |
| 37  | 13845,89            | 13962,46                 | 13588,56        | 0,008419                |
| 38  | 13932,32            | 13962,46                 | 908,4196        | 0,002163                |
| 39  | 13962,47            | 13962,46                 | 0,0001          | 7,16E-07                |
| 40  | 14035,83            | 14119,5                  | 700,669         | 0,005961                |
| 41  | 14088,09            | 14119,5                  | 986,5881        | 0,00223                 |
| 42  | 14305,17            | 14371,89                 | 4451,558        | 0,004664                |
| 43  | 14485,07            | 14371,89                 | 12809,71        | 0,007814                |
| 44  | 14293,11            | 14371,89                 | 6206,288        | 0,005512                |
| 45  | 14241,86            | 14119,5                  | 14971,97        | 0,008592                |
| 46  | 14238,84            | 14119,5                  | 14242,04        | 0,008381                |
| 47  | 14338,34            | 14371,89                 | 1125,602        | 0,00234                 |
| 48  | 14413,71            | 14371,89                 | 1748,912        | 0,002901                |
| 49  | 14483,06            | 14371,89                 | 12358,77        | 0,007676                |
| 50  | 14394,62            | 14371,89                 | 516,6529        | 0,001579                |
| 51  | 14562,45            | 14762,82                 | 40148,14        | 0,013759                |
| 52  | 14889,08            | 14762,82                 | 15941,59        | 0,00848                 |
| 53  | 14892,09            | 14762,82                 | 16710,73        | 0,00868                 |
| 54  | 15158,42            | 15099,9                  | 3424,59         | 0,003861                |
| 55  | 15299,12            | 15337,35                 | 1461,533        | 0,002499                |
| 56  | 15790,56            | 15659,9                  | 17072,04        | 0,008275                |
| 57  | 16354,37            | 16022,02                 | 110456,5        | 0,020322                |
| 58  | 16691,04            | 16611,91                 | 6261,557        | 0,004741                |
| 59  | 16569,43            | 16611,91                 | 1804,55         | 0,002564                |
| 60  | 16409,64            | 16443,9                  | 1173,748        | 0,002088                |
| 61  | 16311,15            | 16022,02                 | 83596,16        | 0,017726                |
| 62  | 16417,68            | 16443,9                  | 687,4884        | 0,001597                |
| 63  | 16448,84            | 16443,9                  | 24,4036         | 0,0003                  |
| 64  | 16495,07            | 16443,9                  | 2618,369        | 0,003102                |
| 65  | 16824,71            | 16611,91                 | 45283,84        | 0,012648                |
| Day | Actual Data ($X_i$) | Forecasted Data ($F_i$) | $(X_i - F_i)^2$ | $\frac{|X_i - F_i|}{X_i}$ |
|-----|-------------------|-----------------|-----------------|---------------------|
| 66  | 16546.32          | 16611.91        | 4302.048        | 0.003964            |
| 67  | 16638.78          | 16611.91        | 721.9969        | 0.001615            |
| 68  | 16492.05          | 16443.9         | 2318.422        | 0.00292             |
| 69  | 16326.23          | 16022.02        | 2953.72         | 0.018633            |
| 70  | 16322.21          | 16022.02        | 9014.04         | 0.018392            |
| 71  | 15919.20          | 16022.02        | 10571.95        | 0.006459            |
| 72  | 15800.61          | 16022.02        | 49022.39        | 0.014013            |
| 73  | 15785.54          | 15659.9         | 15785.41        | 0.007959            |
| 74  | 15865.94          | 16022.02        | 24360.97        | 0.009837            |
| 75  | 15580.52          | 15659.9         | 6301.184        | 0.005095            |
| 76  | 15620.71          | 15659.9         | 1535.856        | 0.002509            |
| 77  | 15721.22          | 15659.9         | 3760.142        | 0.0039              |
| 78  | 15644.84          | 15659.9         | 226.8036        | 0.000963            |
| 79  | 15708.15          | 15659.9         | 2328.063        | 0.003072            |
| 80  | 15630.77          | 15659.9         | 848.5569        | 0.001864            |
| 81  | 15668.96          | 15659.9         | 82.0836         | 0.000578            |
| 82  | 15565.44          | 15337.35        | 52025.05        | 0.014654            |
| 83  | 15492.08          | 15337.35        | 23941.37        | 0.009988            |
| 84  | 15232.79          | 15337.35        | 10932.79        | 0.006864            |
| 85  | 15148.37          | 15099.9         | 2349.341        | 0.0032              |
| 86  | 15179.52          | 15099.9         | 6339.344        | 0.005245            |
| 87  | 15202.64          | 15337.35        | 18146.78        | 0.008861            |
| 88  | 15084.05          | 15099.9         | 251.2225        | 0.001051            |
| 89  | 15010.68          | 14987.9         | 518.9284        | 0.001518            |
| 90  | 15052.89          | 15099.9         | 2209.94         | 0.003123            |
| 91  | 14961.44          | 14987.9         | 700.1316        | 0.001769            |
| 92  | 15020.73          | 15099.9         | 6267.889        | 0.005271            |
| 93  | 14983.55          | 14987.9         | 18.9225         | 0.00029             |
| 94  | 14959.43          | 14987.9         | 810.5409        | 0.001903            |
| 95  | 14987.12          | 14987.9         | 0.6084          | 5.2E-05             |
| 96  | 14858.93          | 14762.82        | 9237.132        | 0.006468            |
| 97  | 14847.87          | 14762.82        | 7233.503        | 0.005728            |
| 98  | 14834.81          | 14762.82        | 5182.56         | 0.004853            |
| 99  | 14842.85          | 14762.82        | 6404.801        | 0.005392            |
| 100 | 14806.67          | 14762.82        | 1922.823        | 0.002962            |

$MSE = 9197.613 \quad AFER = 0.414755\%$
5. Conclusion
We have presented a new forecasting method for the forecasting exchange rate (USD-IDR) using frequency density-based partitioning [5], calculate the length of average-based interval [6] and modification FLRG as k-means clustering [2]. We divided fuzzy sets number of each fuzzy logical relationship into 16 clusters and then do iterations by using k-means clustering. Final cluster centers can be used to build a new Fuzzy Logical Relationship Group (FLRG). The result of forecasting data, as shown in table 4, shows that it gives an error value MSE value 9197.613 and AFER value 0.414755%.

References
[1] Song Q and Chissom B S 1993 Fuzzy Sets Syst. 54 269
[2] Cheng S H, Chen S M, Jian W S 2016 Inf. Sci. 327 272
[3] Yolcu O C and Alpaslan F 2018 Appl. Soft Comput. J. 66 18
[4] Chen S M, Zou X Y, Gunawan G C 2019 Inf. Sci. 500 127
[5] Jilani T A, Burney S M A, and Ardl C 2007 Int. J. Comp. Elec. Auto. Cont. Inf. Eng. 4 112
[6] Saxena P, Sharma K, Easo S 2012 Int. J. Comput. Technol. Appl 3 957
[7] B Irawanto et al 2019 J. Phys.: Conf. Ser. 1321 022081
[8] MacQueen J 1967 Fifth Berkeley Symp. 54 281
[9] Harding J A et al 2006 J. Manuf. Sci. Eng. Trans. 128 969
[10] Wu X et al 2008 Knowl Inf Syst 14 1