Received: June 9, 2021. Revised: September 17, 2021.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Absolute Differences K-Means Clustering Approach on Determining Intervals to Optimize Fuzzy Time Series Markov Chain Model

Ahmad Alwarid[1] **Agus Sihabuddin**[1]*

[1]*Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia*
* Corresponding author's Email: a_sihabudin@ugm.ac.id

Abstract: Fuzzy Time Series models have been developed in various ways, one of which is determining the intervals. Several methods were applied to determine the intervals, but the performances are still not optimal. This paper proposes a new approach that uses a combination of Absolute Differences and K-means Clustering in the Fuzzy Time Series Markov Chain model. K-means Clustering made the interval more flexible and compact based on the data it clustered. In addition, Absolute Differences was used as the based method to define how many intervals to be made. This study used Taiwan Capitalization Weighted Stock Index (TAIEX) as benchmark data to evaluate the proposed method, which produced an average Mean Absolute Percentage Error (MAPE) value of 0.42, and an average Root Mean Squared Error (RMSE) value of 51.09 for the test data. The proposed method outperformed other compared researches at the end of this paper in terms of prediction accuracy.

Keywords: Forecasting, Fuzzy time series, Markov chain, Absolute difference, K-means clustering.

1. Introduction

Nowadays, there are many cases where forecasting can be implemented. Forecasting is very useful for many people, such as stock or forex traders and the government. For example, predicting the currency exchange rates and stocks can be useful for traders to determine when to sell and buy currencies and stocks. Moreover, forecasting can also be used by the government to prepare economic plans based on forecasting results.

Currency exchange rates have several characteristics, such as that their future value is not easy to predict and fluctuates against other currencies [1]. The change of currency exchange rates will impact a country's economy, one of which is the price of goods. Many analytical methods in statistics, such as linear regression, moving average, autoregressive integrated moving average, can be forecast but have weaknesses in analyzing uncertainty from data [2]. One method to overcome uncertainty in time series data is using Fuzzy Time Series (FTS).

Fuzzy sets theory was introduced by Zadeh [3] to handle uncertainty data and compute linguistic variables. Song and Chissom [4] proposed the FTS model for the first time to forecast enrollment of the University of Alabama. Since then, various developments of FTS models have been proposed to get better results. Chen [5] then refined the model by grouping the fuzzy relations based on the antecedent state. Huarng [6] found that intervals used in FTS models can affect the prediction accuracy and then proposed average-based length and distribution-based length to determine the interval length. Huarng's [6] study started the research focus on the process of determining the interval in FTS model. Cheng et al. [7] developed the FTS model using weights for each relation and grouped them chronologically. The model [7] assigned weight for each fuzzy logic relationship to affect differently based on the weight in the calculation of forecast results. Since then, the FTS model has been developed in various ways.

Tsaur [8] stated that various unknown factors influence the trend of currency exchange rate developments, and it is impossible to form a forecasting model that can consider all unknown factors. One of the factors is investor speculation on currency exchange rates. Then, Tsaur [8] proposed combining FTS with the Markov Chain method to become Fuzzy Time Series Markov Chain (FTSMC) because the Markov process has a good performance in forecasting using change of states probability based on fuzzy relations. This combination produced a better result and higher accuracy than the previous FTS models. Rukhansah et al. [9] combined Tsaur's [8] FTS model with Huarng's average-based length [6] method for determining the interval length and got a better result than the Tsaur's [8] model.

Apart from the combination with Markov Chain, FTS models have been developed in other ways. Chen and Phuong [10] used Particle Swarm Optimization (PSO) method in the FTS model for interval and vector weight determination. On the other hand, Dong and Ma [11] used the Fuzzy Silhouette criterion to get interval length and applied an error learning mechanism. Based on those papers, only a few studies develop the process of determining the interval of the FTS model. However, Huarng [6] previously stated that it could affect the overall process and possibly produce better results. This study focuses on developing the process of determining the interval of FTS model so it can be used to improve the model performance and be reused to improve the model in future studies. In the FTS model, the methods to determine the interval have been done using various methods. Apart from Huarng's [6] method, Zhang [12] implemented K-means clustering, that originally introduced by MacQueen [13], for determining interval in FTS model in forecasting University of Alabama enrollment and get better performance compared to [4, 5] model.

This research proposed a new method to determine the interval in the Tsaur's [8] FTSMC model. The proposed method's algorithm uses absolute differences based on Huarng's method [6] and K-means clustering [13] to determine the fuzzy sets and intervals of the FTSMC model. Absolute differences will be used to get the number of intervals, while K-means clustering will be used to group the data and set the length of each interval. Each data that has the greatest similarity will be put in the same interval. The flexibility of intervals and data clustering is expected to give better forecasting results.

This paper used TAIEX data from 2014 to 2018 to evaluate the performance and compared the results with some well-known models such as Song and Chissom's model [4], Chen's model [5], Huarng's model [6], Dong and Ma's model [11], Chen and Tanuwijaya's model [14], Ye et al. [15], and Wu et al. [16] which also used TAIEX data from 2014 to 2018 based on the comparison in Dong and Ma's [11] paper in Section 1.

Section 2 describes the reference method, proposed method, and performance measurement method used in this study. Section 3 describes the study results consisting of implementation results in Section 3.1 and performance comparison and analysis of results in Section 3.2. Lastly, the conclusion of the study is described in Section 4.

2. Methods

2.1 Fuzzy time series Markov chain model

Tsaur's [8] FTSMC model was developed, combining the conventional FTS model with the Markov chain. The steps of Tsaur's [8] FTSMC model are:

1. Define the universe of discourse and divide it into several partitions of intervals as in Eq. (1).

   $$U = [D_{min} - D_1, D_{max} + D_2] \quad (1)$$

   Universe of discourse U have lower bound as $D_{min} - D_1$ and upper bound as $D_{max} + D_2$ where D_{min} is the minimum value of the data, D_{max} is the maximum value of the data, while D_1 and D_2 are selected integers as tolerance values.
2. Partition universal of discourse into several sets and intervals.
   Universe of discourse U will be partitioned into n intervals based on the method used. In this step, the proposed method will be used to determine the sets and intervals.
3. Fuzzification of the data.
   Each data will be identified based on the fuzzy sets. If $F(t-1)$ is in the interval length of fuzzy set A_k then $F(t-1)$ is fuzzified as A_k [17].
4. Identify fuzzy logical relationships (FLR).
   Fuzzy relationships are formed by connecting each data to the next data. In this process, each relation used must be distinct without duplication. If $F(t-1)$ is fuzzified as A_k and $F(t)$ is fuzzified as A_k, then relation $A_k \rightarrow A_m$ is identified with A_k as current state and A_m as the next state [17].
5. Form fuzzy logical relationship groups (FLRG).
   FLR from the previous step is grouped based on the antecedent/first state. For example, $A_1 \rightarrow A_1$ and $A_1 \rightarrow A_2$ have the same antecedent, which is A_1 . Both relations then grouped as one relationship group as $A_1 \rightarrow A_1, A_2$ because data in fuzzy sets A_1 have ever changed to A_1 or A_2 in historical data used.
6. Create a probability matrix.

Based on FLRG from the previous step, a probability matrix R formed such as in Eq. (2).

$$R = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \cdots & P_{nn} \end{bmatrix}, P_{ij} = \frac{S_{ij}}{S_i} \quad (2)$$

where P_{ij} as change probability from state A_i to A_j , S_{ij} as change frequency from A_i to A_j, and S_i as number of data identified as A_i.

7. Defuzzification to get forecast results.
   In this step, there are three rules to fulfil the defuzzification:
   a) If $F(t-1)$ is in fuzzy set A_i and there is a one-to-one relationship in the FLRG, which is $A_i \rightarrow A_j$, then the forecast result is the median of interval in fuzzy set A_j.
   b) If $F(t-1)$ is in fuzzy set A_i and there is no relationship at all in FLRG, and then the forecast result is the median of interval in fuzzy set A_i.
   c) If $F(t-1)$ is in fuzzy set A_i and there is a one-to-many relationship in the FLRG, which is $A_i \rightarrow A_1, \ldots, A_i, \ldots, A_n$, then the forecast result is formulated as in Eq. (3).

$$F(t) = m_1 \times P_{i1} + m_2 \times P_{i2} + \cdots + Y_{i-1} \times P_{ii} + \cdots + m_n \times P_{in} \quad (3)$$

where m_i as the mid-value of fuzzy sets A_i.

8. Calculate the adjustment value of the results.
   The adjustment value is calculated to decrease the error of the forecast result. The calculation follows these rules:
   a) Rule 1: If fuzzy set A_i have relation with A_i, from state A_i at $t-1$ so that $F(t-1) = A_i$ and make an ascending movement to A_j at t where $(i < j)$, then the adjustment value defined as in Eq. (4).

$$D_{t1} = (l/2) \quad (4)$$

Where l is the mean interval length.

   b) Rule 2: If fuzzy set A_i have relation with A_i, from state A_i at $t-1$ so that $F(t-1) = A_i$ and make a descending movement to A_j at t where $(i > j)$, then the adjustment value defined as in Eq. (5).

$$D_{t1} = -(l/2) \quad (5)$$

Where l is the mean interval length.

   c) Rule 3: If the current state is A_i, from state A_i at $t-1$ so that $F(t-1) = A_i$ and make a forward movement to A_{i+s} at t where $(1 \leq s \leq n-i)$, then the adjustment value defined as in Eq. (6).

$$D_{t2} = (l/2)s, 1 \leq s \leq n-i \quad (6)$$

Where l is the mean interval length.

   d) Rule 4: If the current state is A_i, from state A_i at $t-1$ so that $F(t-1) = A_i$ and make a backward movement to A_{i-v} at t where $(1 \leq v \leq i)$, then the adjustment value defined as in Eq. (7).

$$D_{t2} = -(l/2)v, 1 \leq v \leq i \quad (7)$$

Where l is the mean interval length.

9. Fix previous results by using adjustment values.
   The previous forecast result is fixed using the adjustment values from the previous step as in Eq. (8).

$$F'(t) = F(t) \pm D_{t1} \pm D_{t2} \quad (8)$$

where $F'(t)$ is the adjusted forecast result.

2.2 Average-based length and distributed-based length

Huarng [6] introduced average-based length and distributed-based length as a new way to get the length of intervals used in the FTS model. Algorithm for Huarng's [6] distribution-based length:

1. Calculate all the absolute differences between data x_i and x_{i+1} for $i = 1,2,\ldots,n-1$, as the first differences and the average of the first differences.
2. According to the average, determine the base for length of intervals by following Table 1.
3. Plot the cumulative distribution of the 4rst differences. The base determined in step 2 is used as an interval.
4. According to the base determined in step 2, choose as the length of intervals the largest length that is smaller than at least half the first differences.

Table 1. Base mapping table

Range	Base
0,1 – 1.0	0,1
1.1 – 10	1
11 – 100	10
101 -1000	100

As for algorithm steps of Huarng's [6] average-based length are:

1. The same as step 1 in the algorithm for distribution-based length.
2. Take one half the average (in step 1) as the length.
3. According to the length (in step 2), determine the base for the length of intervals by following Table 1.
4. Round the length according to the determined base as the length of intervals.

2.3 K-means clustering

The term K-means was first used by MacQueen [13] for the process of partitioning N-dimensional population into k parts. The K-means put each data to nearest cluster based on mid value of the cluster. The base steps of the K-means clustering are described as follows:

1. Choose the number of clusters to be used.
2. Choose random data as the mid-value of each cluster.
3. Groups the data to the nearest cluster by calculating the distance of each data and mid-value of each cluster.
4. Calculate the new mid-value of each cluster by getting the mean value of data in each cluster.
5. Repeat steps 4 and 5 until there are no changes of mid values.

2.4 The proposed method

The proposed method was used to generates the sets and intervals of Tsaur's [8] FTSMC model. The method was a combination of absolute differences from Huang's method [6] in step 1 to generate the number of sets or intervals and K-means clustering with modification in the remaining steps to groups the data and set each interval's length with a little. This method was named *Absolute Differences K-means Clustering* (ADKC) method. The steps of the method are:

1. Calculate the average of all absolute differences ($\bar{x}$) of data x_i and x_{i+1} for $i = 1, 2, \ldots, n-1$.
2. Total intervals and fuzzy sets (p) to be used in the model are defined in Eq. (9).

$$p = round(\frac{U_b - U_a}{\bar{x}}) \quad (9)$$

where U_b denotes the upper bound of the universe of discourse (U) and U_a denotes lower bound of U.

3. Clusters are formed with the same amount as intervals. The initial mid-value of each cluster except the last cluster is defined in Eq. (10).

$$m_i = \frac{(U_a + \bar{x} \times (i-1)) + (U_a + \bar{x} \times i)}{2} \quad (10)$$

where m_i denotes mid-value of cluster i for $i = 1, 2, \ldots, p-1$. For the last cluster, the initial mid-value is defined in Eq. (11).

$$m_p = \frac{(U_a + \bar{x} \times (p-1)) + U_b}{2} \quad (11)$$

and m_p denotes the mid-value of cluster p or the last cluster. This step will generate a relatively balanced initial cluster.

4. For each data will be calculated, the distance to each cluster mid-value. If m_i is the closest to the data, then the data will be put to cluster A_i.
5. The value of m_i will be redefined by calculating the rounded average value of the data in the cluster A_i.
6. Steps 4 and 5 are looped until there are no changes of value in each mid-value.
7. The lower bound of the first interval and upper bound of the last interval are defined in Eq. (12).

$$a_1 = U_a \, , b_p = U_b \quad (12)$$

and the other intervals, lower bound and upper bound, are defined in Eq. (13).

$$b_i = \frac{value\ in\ A_i + value\ in\ A_{i+1}}{2}$$
$$a_i = b_{i-1} \quad (13)$$

where a_i denotes the lower bound of cluster i and b_i denotes the upper bound of cluster i.

The intervals have various lengths, so the average length of the intervals was used to calculate the adjustment value in the FTSMC model.

2.5 Performance evaluation

The performance of the ADKC method was evaluated using the forecast accuracy method Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The formula of MAPE is described in Eq. (14).

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|Y(t) - F'(t)|}{Y(t)} \times 100\% \quad (14)$$

while the formula of RMSE is described in Eq. (15).

Table 2. The properties of datasets

No	Data	Total Size	Training Size	Test Size
1	TAIEX-2014	247	204	43
2	TAIEX-2015	244	200	44
3	TAIEX-2016	241	197	44
4	TAIEX-2017	243	200	43
5	TAIEX-2018	245	203	42

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y(t) - F'(t))^2} \qquad (15)$$

where $Y(t)$ denotes the forecast results, and $F'(t)$ denotes the actual results.

3. Results

TAIEX data from 2014 to 2018 are used to test the proposed method in the FTSMC model. The data are used separately per year in the model to evaluate the performance and accuracy of the method in the FTSMC model. The first ten months of each data (January to October) are used as training data, while the last two (November and December) are used as test data. The property details of the datasets that are used in this study are shown in Table 2.

3.1 Implementation

This section utilizes sub-dataset TAIEX-2014 to illustrate the process of the proposed method.

Step 1. Define the universe of discourse.

The minimum and maximum values of the data are used to get the lower bound and upper bound of the universe of discourse. The minimum value is rounded down to the nearest hundred as a lower bound, while the maximum value is rounded up to the nearest hundred as the upper bound. Based on that, the universe of discourse is defined as follows:

$$U = [8200, 9600]$$

Step 2. Define fuzzy sets and intervals.

Fuzzy sets and intervals are defined using the proposed method. The mean absolute difference generates many sets and intervals that are not too many and not too few. The formula then calculates the mid-value of each interval with a balanced distance from each other. The K-means clustering puts the data in the same cluster with other data that are quite similar, or in this case, the data are close to each other. Based on that clusters, several sets and intervals can be made. The method produces unequal interval lengths in each set to be more flexible based on the data used. In this implementation, twenty-eight sets and intervals are generated and shown in Table 3.

Some of the previous methods that were used to define interval length in fuzzy sets are the average-based length and distribution-based length [6]. Those methods sometimes may produce many fuzzy sets without data in the interval because the methods generate too many intervals and not flexible. Large amounts of data are needed to be evenly distributed and prevent empty fuzzy sets. The proposed method used the absolute difference to get the number of sets and intervals. On the other hand, K-means clustering grouped the data and set the interval length based on that, which is more flexible.

Step 3. Fuzzification of data.

Each data is fuzzified based on the fuzzy sets from the previous step. The fuzzified data are shown in Table 4. Each data is identified below the upper bound and above or equal to the lower bound of a set interval. If the condition is fulfilled, then the data is a part of that fuzzy set.

Steps 4-5. Identify FLR and FLRG.

Using the fuzzified data from the previous step, FLR is identified. Then, the relationships are grouped in FLRG based on the transition's antecedent and then count the number of times each relation occurs from the FLR. The FLRG can be seen in Table 5.

Table 3. Fuzzy sets and intervals

Fuzzy set	Interval	Fuzzy set	Interval	Fuzzy set	Interval
A1	(8200, 8287.74)	**A10**	(8706.86, 8758.61)	**A19**	(9120.83, 9169.11)
A2	(8287.74, 8349.18)	**A11**	(8758.61, 8800.03)	**A20**	(9169.11, 9210.44)
A3	(8349.18, 8411.26)	**A12**	(8800.03, 8838.82)	**A21**	(9210.44, 9259.94)
A4	(8411.26, 8483.86)	**A13**	(8838.82, 8890.36)	**A22**	(9259.94, 9297.44)
A5	(8483.86, 8522.08)	**A14**	(8890.36, 8937.53)	**A23**	(9297.44, 9343.9)
A6	(8522.08, 8570.91)	**A15**	(8937.53, 8981.78)	**A24**	(9343.9, 9396.84)
A7	(8570.91, 8617.05)	**A16**	(8981.78, 9029.74)	**A25**	(9396.84, 9444.47)
A8	(8617.05, 8659.33)	**A17**	(9029.74, 9065.6)	**A26**	(9444.47, 9487.78)
A9	(8659.33, 8706.86)	**A18**	(9065.6, 9120.83)	**A27**	(9487.78, 9548.05)

Table 4. Fuzzified data

Date	Price	Fuzzy set
1/2/2014	8612.54	A7
1/3/2014	8546.54	A6
1/6/2014	8500.01	A5
1/7/2014	8512.3	A5
1/8/2014	8556.01	A6
...	...	...
10/27/2014	8627.78	A8
10/28/2014	8773.55	A11
10/29/2014	8903.68	A14
10/30/2014	8888.07	A13
10/31/2014	8974.76	A15

Table 5. Fuzzy Logical Relationship Groups (FLRG)

State	Next State
A1	{'A2': 1}
A2	{'A3': 1}
A3	{'A4': 1, 'A3': 1}
A4	{'A5': 2, 'A1': 1}
A5	{'A6': 3, 'A5': 2, 'A4': 1, 'A9': 1}
A6	{'A7': 4, 'A5': 2, 'A6': 2, 'A8': 1}
A7	{'A7': 6, 'A6': 4, 'A8': 3, 'A4': 1, 'A9': 1}
A8	{'A7': 3, 'A10': 2, 'A8': 2, 'A11': 1, 'A5': 1}
A9	{'A10': 3, 'A9': 3, 'A8': 1, 'A7': 1}
A10	{'A9': 3, 'A10': 2, 'A11': 2, 'A8': 1}
A11	{'A13': 2, 'A8': 1, 'A11': 1, 'A14': 1, 'A12': 1}
A12	{'A13': 2, 'A12': 1}
A13	{'A13': 7, 'A14': 5, 'A15': 2, 'A11': 1, 'A12': 1}
A14	{'A13': 5, 'A14': 3, 'A15': 2}
A15	{'A15': 7, 'A16': 2, 'A10': 1, 'A11': 1, 'A14': 1, 'A18': 1}
A16	{'A15': 2, 'A16': 1, 'A17': 1}
A17	{'A19': 1, 'A17': 1, 'A15': 1}
A18	{'A18': 3, 'A19': 2, 'A17': 1, 'A20': 1, 'A16': 1}
A19	{'A19': 4, 'A18': 4, 'A21': 3, 'A20': 1}
A20	{'A20': 2, 'A19': 2, 'A21': 2}
A21	{'A21': 6, 'A22': 2, 'A19': 2, 'A20': 2, 'A23': 1, 'A24': 1}
A22	{'A21': 2, 'A23': 2}
A23	{'A22': 2, 'A19': 1, 'A23': 1, 'A21': 1, 'A24': 1}
A24	{'A26': 2, 'A24': 2, 'A23': 1, 'A25': 1}
A25	{'A25': 5, 'A26': 2, 'A27': 2, 'A24': 2}
A26	{'A25': 3, 'A26': 1, 'A27': 1, 'A23': 1}
A27	{'A27': 6, 'A28': 2, 'A25': 2}
A28	{'A26': 1, 'A27': 1}

Step 6. Create probability matrix

The probability matrix is made by using the FLRG based on Eq. (2). Then, the probability is used to calculate the forecast results using Eq. (3). This matrix generates different probabilities for each transition in the prediction results. Table 6 gives the probability matrix from this step.

Table 6. Probability matrix

Next Present	A1	A2	A3	A4	...	A27	A28
A1	0	1	0	0	...	0	0
A2	0	0	1	0	...	0	0
A3	0	0	0,5	0,5	...	0	0
A4	0,33	0	0	0	...	0	0
...	...	...	...	...	...	...	...
A25	0	0	0	0	...	0,18	0
A26	0	0	0	0	...	0,16	0
A27	0	0	0	0	...	0,6	0,2
A28	0	0	0	0	...	0,5	0

Steps 7-9. Get forecast results and adjustment value.

The forecast results are calculated based on the step in the FTSMC model using the probability from the previous step. Then, the adjustment values are also calculated to fix the results as adjusted forecast. The test data results are shown in Table 4, and the comparison of the adjusted result and actual data can be seen in Fig.1.

The actual price, forecast result, adjustment value, and adjusted result are presented in Table 6. Each forecast has a different adjustment value, be it negative or positive, according to the rules that make the final adjusted forecast result different. As for Fig.1, the prediction error with the actual price can be concluded that is quite good for this method.

3.2 Forecast comparison and analysis

Several FTS models are used as performance comparison in MAPE and RMSE value based on the comparison in Dong and Ma’s [11] paper. The models that will be used are some well-known models such as Song and Chissom’s model [4], Chen’s model [5], Huarng’s model [6], Chen and Tanuwijaya’s model [14], Ye et al. [15], Wu et al. [16], and Dong and Ma’s model [11].

The performance comparison used in this study was the comparison from Dong and Ma’s [11] paper that used the TAIEX data from 2008 to 2018. The other methods used in Table 8 and Table 9, such: [4-6, 14-16], originally were not TAIEX data, but the models were regenerated and tested with TAIEX data in [11]. In this study, we assume that models generated by [11] using TAIEX data is comparable.

Each year the data was treated as different instances and implemented to the model, and then the performance is calculated using MAPE and RMSE algorithm shown in Table 8 and 9.

Based on the MAPE results comparison in Table 8, the ADKC method got the best result from 2014 until 2016.

Received: June 9, 2021. Revised: September 17, 2021.

Table 7. Forecast results of test data

Date	Actual price	Forecast result	Adjustment value	Adjusted forecast
11/3/2014	9004.86	8950.311923	50	9000.311923
11/4/2014	8989.18	8992.96	0	8992.96
11/5/2014	8962.6	8989.04	-50	8939.04
11/6/2014	8891.02	8943.764231	-50	8893.764231
11/7/2014	8912.62	8891.532	0	8891.532
11/10/2014	9049.98	8898.012	100	8998.012
11/11/2014	9034.14	9051.535	0	9051.535
11/12/2014	8918.95	9046.255	-100	8946.255
11/13/2014	8980.67	8899.911	50	8949.911
11/14/2014	8982.88	8953.494231	50	9003.494231
...	...	...	...	...
12/25/2014	9158.7	9188.78	-50	9138.78
12/26/2014	9214.07	9158.58375	75	9233.58375
12/29/2014	9286.28	9228.737857	50	9278.737857
12/30/2014	9268.43	9277.93	0	9277.93
12/31/2014	9307.26	9277.93	25	9302.93

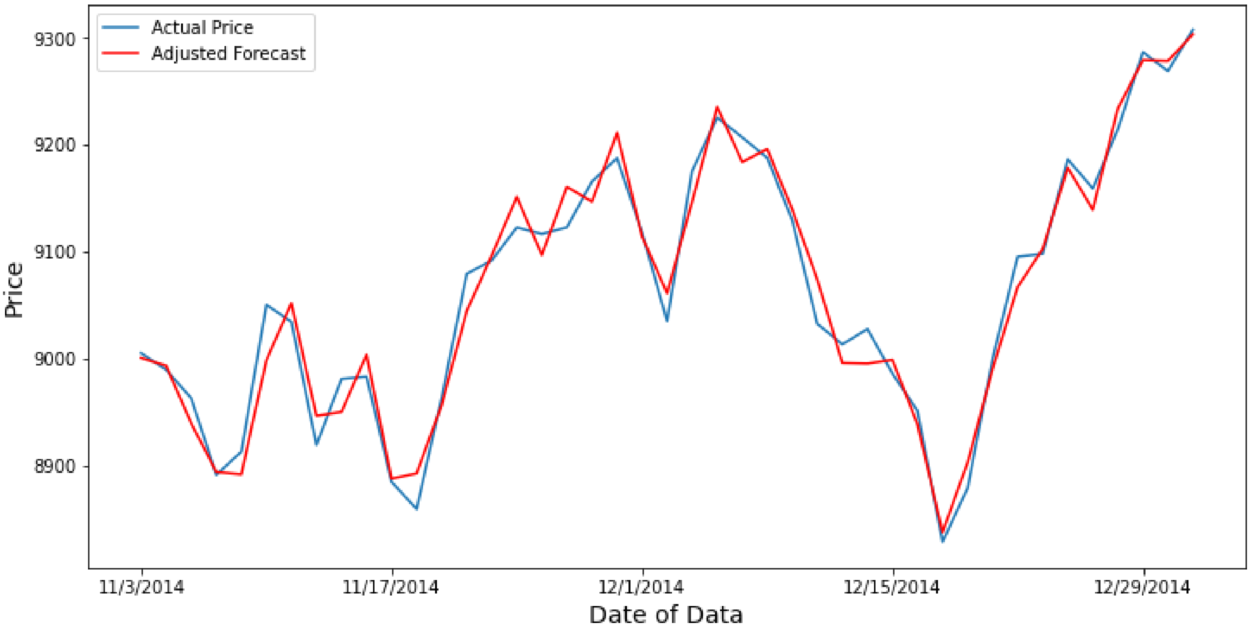

Figure. 1 Graphic of adjusted forecast and actual price of test data

Table 8. Performance evaluation (MAPE) of various model for TAIEX test data

Model	2014	2015	2016	2017	2018	Average
Song and Chissom [4]	0,86	1,78	1,1	0,71	1,29	1,15
Chen [5]	3,31	0,97	3,2	8,45	4,73	4,13
Huarng [6]	0,79	0,86	0,63	0,47	1,71	0,89
Chen & Tanuwijaya [14]	0,66	0,98	0,7	0,67	1,42	0,89
Ye et al. [15]	0,58	0,8	0,62	0,45	0,76	0,64
Wu et al. [16]	0,59	0,77	0,62	0,47	0,82	0,65
Dong and Ma [11]	0,44	0,59	0,43	0,27	0,53	0,45
Proposed: FTSMC with ADKC	**0,2**	**0,35**	**0,35**	**0,27**	**0,93**	**0,42**

Table 9. Performance evaluation (RMSE) of various model for TAIEX test data

Model	2014	2015	2016	2017	2018	Average
Song and Chissom [4]	98,64	176,6	122,71	91,69	173,36	132,6
Chen [5]	308,72	100,75	316,57	902,18	492,89	424,22
Huarng [6]	87,52	90,72	80,98	63,2	186,21	101,73
Chen & Tanuwijaya [14]	74,85	94,33	86,59	88,93	161,78	101,3
Ye et al. [15]	66,26	82,78	80,14	61,98	99,03	78,04
Wu et al. [16]	68,28	78,65	82,43	62,49	104,49	79,27
Dong and Ma [11]	44,23	58,79	53,15	36,99	70,67	52,76
Proposed: FTSMC with ADKC	**21,79**	**37,59**	**43,15**	**38,71**	**114,22**	**51,09**

In 2017, this method also got the best performance even though it is the same as Dong and Ma's [11] model. On the other hand, the performance was not good in 2018. However, on average, the performance of ADKC is better.

The RMSE comparison performance of ADKC with some other methods is almost the same as MAPE, as shown in Table 9. From 2014 to 2017, ADKC got the best performance compared to other models. Even though, just like the MAPE result, the ADKC's performance was pretty bad in 2018. This kind of performance may happen due to critical economic conditions or other factors, both internal or external.

The used of K-means Clustering made the cluster and interval more flexible because, in step 4-6 of the proposed method, each data affect the creation of clusters which in step 7 the length of each interval will be decided based on those clusters. Slight differences in training data will result in different clusters and intervals. This method produced clusters with similar data in each cluster so that they are more compact as a unit and are more suitable to be used as intervals that are used as data identities in the fuzzification process.

Based on the comparison of MAPE and RMSE results, the FTSMC model using the ADKC method presents better performance, although, in 2018, the performance is not good. This study proves that using a different method to determine the interval in the FTS model can produce different performances. Previous methods that are still not optimal can be improved, changed, or combined to generate a better method. That way, the forecast result and prediction accuracy will be better and can be used for many useful purposes.

4. Conclusion

In this study, ADKC as a new approach to determine interval in the FTSMC model has been applied. The key process of the method is the absolute difference and K-means clustering, which are modified and implemented in the FTSMC model. The flexibility of interval and data clustering are the main reasons that make the method generates better intervals. The cluster that is generated based on data similarity make a better interval as the identities of data in FTSMC model. Compared with many models, the performance of ADKC method is testified for the process of determining interval in FTSMC model, which reduced the MAPE and RMSE values. ADKC method was also suitable to be used in other versions of FTS model so it can improve the performance and accuracy of the results.

A future work of the method and model may improve the way to get the number of sets and intervals and find a more optimal division of the universe of discourse and forecasting formula to provide wider possibilities of forecast results. The method can also be improved by providing a way to handle unpredictable conditions so it can constantly produce better performance.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The contributions by the authors for this research are as follows: conceptualization, methodology, formal analysis, Ahmad Alwarid and Agus Sihabuddin; software, investigation, resources, writing—draft preparation, Ahmad Alwarid; writing—review and editing, validation, visualization, supervision, Agus Sihabuddin.

References

[1] A. Sihabuddin, Subanar, D. Rosadi, and E. Winarko, "A Second Correlation Method for Multivariate Exchange Rates Forecasting", *International Journal of Advanced Computer Science*, Vol. 5, No. 7, pp. 30-33, 2014.

[2] K. Bisht and S. Kumar, "Fuzzy time series forecasting method based on hesitant fuzzy sets",

Received: June 9, 2021. Revised: September 17, 2021.

Expert Systems with Applications, Vol. 64, pp. 557-568, 2016.

[3] L. A. Zadeh, “Fuzzy Sets”, *Information and Control*, Vol. 8, pp. 338-353, 1965.

[4] Q. Song and B. S. Chissom, “Forecasting enrollments with fuzzy time series”, *Fuzzy Sets and Systems*, Vol. 54, No. 1, pp. 1-9, 1993.

[5] S. M. Chen, “Fuzzy forecasting with DNA computing”, *Fuzzy Sets and Systems*, Vol. 81, pp. 311-319, 1996.

[6] K. Huarng, “Effective lengths of intervals to improve forecasting in fuzzy time series”, *Fuzzy Sets and Systems*, Vol. 123, No. 3, pp. 387-394, 2001.

[7] C. H. Cheng, T. L. Chen, H. J. Teoh, and C. H. Chiang, “Fuzzy time-series based on adaptive expectation model for TAIEX forecasting”, *Expert Systems with Applications*, Vol. 34, No. 2, pp. 1126-1132, 2008.

[8] R. C. Tsaur, “Application To Forecast the Exchange Rate”, *International Journal of Innovative Computing, Information and Control*, Vol. 8, No. 7, pp. 4931-4942, 2012.

[9] N. Rukhansah, A. Muslim, R. Arifudin, F. Matematika, D. Ipa, and U. N. Semarang, “Peramalan Harga Emas Menggunakan Fuzzy Time Series Markov Chain Model”, *Komputaki*, Vol. 1, No. 1, pp. 56-74, 2015.

[10] S. M. Chen and B. D. H. Phuong, “Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors”, *Knowledge-Based Systems*, Vol. 118, pp. 204-216, 2017.

[11] Q. Dong and X. Ma, “Enhanced fuzzy time series forecasting model based on hesitant differential fuzzy sets and error learning”, *Expert Systems with Applications*, Vol. 166, p. 114056, 2021.

[12] Z. Zhang, “Fuzzy Time Series Forecasting Based On K-Means Clustering”, *Open Journal of Applied Sciences*, Vol. 02, No. 04, pp. 100-103, 2012.

[13] J. MacQueen, “Some methods for classification and analysis of multivariate observations”, In: *Proc. of fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, No. 14, pp. 281-297, 1967.

[14] S. M. Chen and K. Tanuwijaya, “Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques”, *Expert Systems with Applications*, Vol. 38, No. 12, pp. 15425-15437, 2011.

[15] F. Ye, L. Zhang, D. Zhang, H. Fujita, and Z. Gong, “A novel forecasting method based on multi-order fuzzy time series and technical analysis”, *Information Sciences*, Vol. 367-368, pp. 41-57, 2016.

[16] H. Wu, H. Long, and J. Jiang, “Handling forecasting problems based on fuzzy time series model and model error learning”, *Applied Soft Computing Journal*, Vol. 78, pp. 109-118, 2019.

[17] J. R. Poulsen, “Fuzzy Time Series Forecasting: Developing a new forecasting model based on high order fuzzy time series”, *Aalborg University Esbjerg*, 2009.