Markov average-based weighted fuzzy time series model to predict PT Kimia farma Tbk stock price

Rediva Azzahra*, Firdaniza, Nurul Gusriani
Universitas Padjajaran, Indonesia

ARTICLE INFO

Article History
Received : 10-08-2021
Revised : 17-10-2021
Accepted : 01-11-2021
Published : 30-11-2021

Keywords:
Average-Based; COVID-19; Stock Price; Markov Chain; Weighted Fuzzy Time Series.

*Correspondence: E-mail: rediva17001@mail.unpad.ac.id
Doi: 10.24042/djm.v4i3.9675

ABSTRACT

The COVID-19 pandemic impacted various activities in Indonesia, including the stock market. Despite the declining economic condition, people are increasingly interested in investing. Among other companies available on the Indonesia Stock Exchange, companies in the health sector have a particular appeal to potential investors, one of which is pharmaceutical companies. This research used a Markov Average-Based Weighted Fuzzy Time Series model applied to PT Kimia Farma Tbk stock price data. This model develops the previous Markov chain–Fuzzy Time Series model, which has not calculated the weights for recurring events and used the Sturgess rule to determine the interval length. In this research, each recurring event has given a different weight that provides different probability values for transitions from one state to another. The Average-Based method is used to determine the interval length that can reflect the fluctuation of the data used. The stock price prediction of PT Kimia Farma Tbk using this model is categorized as very accurate with a MAPE of 2.632%.

INTRODUCTION

At the end of 2019, the world was shocked by the discovering of a new type of virus, SARS-CoV-2 (known as COVID-19), which was later designated as a pandemic by the World Health Organization (WHO). The spread of COVID-19 is still ongoing in Indonesia, impacting various activities, including the stock market. In their research, Liu et al. (2020) concluded that an increase in COVID-19 cases does diminish the country’s stock performance. However, public interest in investment products has increased despite the declining economic condition. This is encouraged by public awareness to start investing and looking for additional income due to uncertain economic conditions (Otoritas Jasa Keuangan, 2021). During this pandemic, companies in the health sector have more attracted investors compared to other companies available on the Indonesia Stock Exchange (Utami & Aliyansah, 2020). This is also proven by Yunpeng et al. (2021) which expressed there is a trusting investor attitude...
toward the health sector industry, which plays a crucial part in preventing this unusual infectious disease. As a state-owned pharmaceutical company in Indonesia, PT Kimia Farma Tbk is currently responsible for handling this pandemic. The company's stock price tends to rise significantly after the news of the vaccine began to emerge so that it is considered interesting to analyze.

Stock price data is time-series data. Many methods can be used to analyze this type of data, but some methods require the satisfaction of basic assumptions. One method that can be used to analyze time-series data without requiring basic assumptions is the Fuzzy Time Series (FTS) (Aladag, 2012). FTS can handle uncertain and incomplete data well, even though it is done by ignoring basic assumptions.

FTS has been used in several previous researches. Susilowati and Sulistijanti (2018) predicted the number of inpatients using FTS, and Jatipaningrum et al. (2019) predicted the Rupiah exchange rate against the Dollar with FTS combined with the Markov chains method. The accuracy of the results obtained is satisfying enough. However, in the previous FTS, the determination of the partition interval length is only based on the number of data used and not considered to reflect the variation in the data. In their research, Sun and Li (2008) stated that the calculation still cannot effectively reflect fluctuations in the data. Hence, this research uses the Average-Based method, which calculates the average difference between two consecutive data at times \( t \) and \( t + 1 \) and is considered to reflect the data's actual fluctuation.

Ignoring repeated transitions in FTS could result in data information loss. The event that keeps repeating is assumed to have the same opportunities to happen again in the future as events that only occur once. Therefore, the weighting procedure gives different probability values for each event (Yu, 2005). By considering these two things, the Markov Average-Based Weighted FTS Model is studied to complete the weaknesses that exist in the previous model.

The steps conducted as follows:
1. Determine the universal set \( U \) from the data used using the following equation:

\[
U = [D_{\text{min}} - D_1, D_{\text{max}} + D_2]
\]  

where \( D_{\text{min}} \) is the lowest data value, \( D_{\text{max}} \) is the highest data value, \( D_1 \) and \( D_2 \) arbitrary positive numbers used for interval adjustment (Tsaur, 2012).
2. Divide $U$ into several partitions following the Average-Based method steps below:
   a. Determine the length of the initial interval ($l^1$)
   \[
   l^1 = \frac{1}{2} \left( \sum_{t=1}^{N-1} |Y(t + 1) - Y(t)| \right) / (N - 1)
   \]  
   with $N$ is the number of data used.
   b. According to $l^1$, determine the base for the interval length based on Table 1.
   Table 1. Interval Length Base
<table>
<thead>
<tr>
<th>Range</th>
<th>Base</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 - 1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>1.1 - 10</td>
<td>1</td>
</tr>
<tr>
<td>11 - 100</td>
<td>10</td>
</tr>
<tr>
<td>101 - 1000</td>
<td>100</td>
</tr>
</tbody>
</table>
   (Sun & Li, 2008)
   c. Rounding $l^1$ according to the basis obtained from Table 1 and get the length of the final interval used ($l$).
   d. Divide $U$ into several partitions with interval length $l$.

3. Determine the fuzzy set $A_i$ for the universal set $U$, with:
   \[
   A_i = \begin{cases} 
   \frac{1}{u_1} + \frac{0.5}{u_2}, & i = 1 \\
   \frac{0.5}{u_{i-1}} + \frac{1}{u_i} + \frac{0.5}{u_{i+1}}, & 2 \leq i \leq k - 1 \\
   \frac{0.5}{u_{k-1}} + \frac{1}{u_k}, & i = k 
   \end{cases}
   \]  

4. Data fuzzification
   Fuzzification is a process of mapping numerical values into fuzzy sets (Díaz-Cortés et al., 2017). If the observation data is the element of the interval $u_i$, then the data is fuzzified to the $A_i$ fuzzy set.

5. Define Weighted Fuzzy Logical Relationship (Weighted FLR)
   Let $w_{i,j}$ be the weight of the transition from state $i$ to state $j$, $w_{i,j} = \varphi$ where $\varphi$ is the $T$th repetition of transition $i$ to another state.

6. Establish Fuzzy Logical Relationship Group (FLRG)
   Assume there are Weighted FLRs, a transition from a state $i$ as follow:
   \[
   A_i \rightarrow A_1 \\
   A_i \rightarrow A_2 \\
   \vdots \\
   A_i \rightarrow A_k
   \]
   In that case, they can be grouped into an FLRG as $A_i \rightarrow A_1, A_2, ..., A_k$ (Khuat & Le, 2017)

7. Construct a Markov chain transition probability matrix with its element is calculated by:
   \[
   p_{i,j} = \frac{\sum w_{i,j}}{\sum_{j=1}^{k} \sum w_{i,j}}
   \]  
   where $\sum w_{i,j}$ is the sum of the transition weight from state $i$ to state $j$ while $\sum_{j=1}^{k} \sum w_{i,j}$ is the total weight of the transition from state $i$ to state $j$ according to the FLRG that has been formed.

8. Defuzzification by observing the following rules:
   Rule 1: If FLRG of state $A_i$ is one to one ($A_i \rightarrow A_j$), with $A_i$ is a state on $t - 1$, then:
   \[
   Y^*(t) = [m_j][p_{i,j}]
   \]  
   Rule 2: If FLRG of state $A_i$ is one to many (e.g:$A_i \rightarrow A_1, A_2, ..., A_k$, $i = 1, 2, ..., k$) and $Y(t - 1)$ is a data on the previous time $(t - 1)$, then:
where $Y^*(t)$ is the result of defuzzification or the initial prediction value and $m_j$ is the median of the interval class $j$.

9. Adjusting the predicted value by observing these following rules:

Rule 1: If state $A_i$ communicate with its own state ($A_i \leftrightarrow A_i$) and experiences an increasing transition for $i < j$ or a decreasing to state $A_j$ at time $t$, the adjustment value can be measured by:

$$D_1(t) = \pm \frac{l}{2}$$

where the value will be positive when $i < j$ and negative for $i > j$.

Rule 2: If state $A_i$ is a state at time $t-1$ and experience the increasing transition to the state of $A_{i \pm v}$ at time $t$, then:

$$D_2(t) = \pm \frac{l}{2}v$$

10. Calculate the final prediction result $\hat{Y}(t)$ considering these points below:

- If FLRG of state $A_i$ is one to many and $A_{i+1}$ can be reached from state $A_i$ with $A_i \leftrightarrow A_i$, then:

$$\hat{Y}(t) = Y^*(t) + D_1(t) + D_2(t)$$

- If FLRG of state $A_i$ is one to many and $A_{i \pm j}$ can be reached from state $A_i$ but $A_i$ is not communicated with its own state, then:

$$\hat{Y}(t) = Y^*(t) \pm D_2(t)$$

11. Measure the prediction accuracy using MAPE

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{Y(t) - \hat{Y}(t)}{Y(t)} \right| \times 100\%$$

The accuracy level of the prediction results based on the MAPE value can be seen in Table 2.

**Table 2. MAPE Accuracy Level**

<table>
<thead>
<tr>
<th>MAPE</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10%</td>
<td>Very Accurate</td>
</tr>
<tr>
<td>11% - 20%</td>
<td>Good</td>
</tr>
<tr>
<td>21% - 50%</td>
<td>Make Sense</td>
</tr>
<tr>
<td>&gt; 50%</td>
<td>Not Accurate</td>
</tr>
</tbody>
</table>

(Klimberg et al., 2010).

**RESULTS AND DISCUSSION**

The data used in this research is the stock price of PT Kimia Farma Tbk from March 09, 2020, to May 31, 2021, collected from the Indonesia Stock Exchange official site (https://www.idx.co.id).

**Table 3. PT Kimia Farma Tbk Stock Price**

<table>
<thead>
<tr>
<th>No.</th>
<th>Date</th>
<th>Stock Price (Rupiah)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>09/03/2020</td>
<td>740</td>
</tr>
<tr>
<td>2</td>
<td>10/03/2020</td>
<td>830</td>
</tr>
<tr>
<td>3</td>
<td>11/03/2020</td>
<td>770</td>
</tr>
<tr>
<td>4</td>
<td>12/03/2020</td>
<td>695</td>
</tr>
<tr>
<td>5</td>
<td>13/03/2020</td>
<td>720</td>
</tr>
<tr>
<td>6</td>
<td>16/03/2020</td>
<td>675</td>
</tr>
<tr>
<td>7</td>
<td>17/03/2020</td>
<td>630</td>
</tr>
<tr>
<td>8</td>
<td>18/03/2020</td>
<td>630</td>
</tr>
<tr>
<td>9</td>
<td>19/03/2020</td>
<td>600</td>
</tr>
<tr>
<td>10</td>
<td>20/03/2020</td>
<td>670</td>
</tr>
</tbody>
</table>
Thus, there are 128 partition classes as the interval length of each class is 50.

The next step is determining the fuzzy sets using Equation (3). The number of fuzzy sets formed is 128, as much as the number of partitions formed in the previous step. The fuzzy set formed is as follows:

\[
\begin{align*}
A_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} \\
A_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} \\
A_3 &= \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} \\
A_4 &= \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5} \\
A_5 &= \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6} \\
A_6 &= \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7} \\
A_7 &= \frac{0.5}{u_6} + \frac{1}{u_7} + \frac{0.5}{u_8} \\
A_8 &= \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9} \\
&\quad \vdots \\
A_{124} &= \frac{0.5}{u_{124}} + \frac{1}{u_{125}} + \frac{0.5}{u_{126}} \\
A_{125} &= \frac{0.5}{u_{124}} + \frac{1}{u_{125}} + \frac{0.5}{u_{126}} \\
A_{126} &= \frac{0.5}{u_{125}} + \frac{1}{u_{126}} + \frac{0.5}{u_{127}} \\
A_{127} &= \frac{0.5}{u_{126}} + \frac{1}{u_{127}} + \frac{0.5}{u_{128}} \\
A_{128} &= \frac{0.5}{u_{127}} + \frac{1}{u_{128}} 
\end{align*}
\]

After the fuzzy set is formed, the stock price data is fuzzified and the weighting process is carried out for each repetition that occurs in each fuzzy set. Table 4 shows the results of fuzzification and weighting for each transition.

<table>
<thead>
<tr>
<th>No.</th>
<th>Date</th>
<th>Stock Price (Rupiah)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>23/03/2020</td>
<td>835</td>
</tr>
<tr>
<td>201</td>
<td>12/01/2021</td>
<td>6.975</td>
</tr>
<tr>
<td>202</td>
<td>13/01/2021</td>
<td>6.500</td>
</tr>
<tr>
<td>203</td>
<td>14/01/2021</td>
<td>6.050</td>
</tr>
<tr>
<td>204</td>
<td>15/01/2021</td>
<td>5.650</td>
</tr>
<tr>
<td>205</td>
<td>18/01/2021</td>
<td>5.275</td>
</tr>
<tr>
<td>283</td>
<td>11/05/2021</td>
<td>2.640</td>
</tr>
<tr>
<td>284</td>
<td>17/05/2021</td>
<td>2.580</td>
</tr>
<tr>
<td>285</td>
<td>18/05/2021</td>
<td>2.610</td>
</tr>
<tr>
<td>286</td>
<td>19/05/2021</td>
<td>2.550</td>
</tr>
<tr>
<td>287</td>
<td>20/05/2021</td>
<td>2.530</td>
</tr>
<tr>
<td>288</td>
<td>21/05/2021</td>
<td>2.420</td>
</tr>
<tr>
<td>289</td>
<td>24/05/2021</td>
<td>2.370</td>
</tr>
<tr>
<td>290</td>
<td>25/05/2021</td>
<td>2.420</td>
</tr>
<tr>
<td>291</td>
<td>27/05/2021</td>
<td>2.420</td>
</tr>
<tr>
<td>292</td>
<td>28/05/2021</td>
<td>2.430</td>
</tr>
<tr>
<td>293</td>
<td>31/05/2021</td>
<td>2.600</td>
</tr>
</tbody>
</table>
To construct a Markov chain transition probability matrix \( P \), calculate the probability values for each state \( i \) transitions to state \( j \) using Equation (4). Below is the calculation of \( p_{i,j} \):

- Since \( A_1 \) transitions to \( A_1 \) and \( A_2 \), thus:
  \[
p_{1,1} = \frac{\sum w_{1,1}}{\sum_{j=1}^{5} w_{1,j}} = \frac{3}{6} = \frac{1}{2}
  \]
  \[
p_{1,2} = \frac{\sum w_{1,2}}{\sum_{j=1}^{5} w_{1,j}} = \frac{3}{6} = \frac{1}{2}
  \]

- Since \( A_2 \) transitions to \( A_1, A_3 \) and \( A_5 \), thus:
  \[
p_{2,1} = \frac{\sum w_{2,1}}{\sum_{j=1}^{5} w_{2,j}} = \frac{2}{6} = \frac{1}{3}
  \]
  \[
p_{2,3} = \frac{\sum w_{2,3}}{\sum_{j=1}^{5} w_{2,j}} = \frac{1}{6}
  \]
  \[
p_{2,5} = \frac{\sum w_{2,5}}{\sum_{j=1}^{5} w_{2,j}} = \frac{3}{6} = \frac{1}{2}
  \]

- Since \( A_3 \) transitions to \( A_2 \) and \( A_5 \), thus:
  \[
p_{3,2} = \frac{\sum w_{3,2}}{\sum_{j=1}^{5} w_{3,j}} = \frac{2}{3}
  \]
  \[
p_{3,5} = \frac{\sum w_{3,5}}{\sum_{j=1}^{5} w_{3,j}} = \frac{1}{3}
  \]
  
and so on, until we obtained all elements values for the 128x128 matrix \( P \).

\[
P = \begin{bmatrix}
  1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
  1/3 & 1/6 & 0 & 1/2 & 0 & 0 & 0 \\
  0 & 2/3 & 0 & 1/3 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 1/3 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots 
\end{bmatrix}
\]

From the matrix \( P \) above, it can be seen that there are some \( p_{i,j} \) with a value of 0. It means that no transition occurs from state \( i \) to state \( j \).
The defuzzification process calculates the initial prediction value using Equations (5) and (6). After calculating the prediction adjustment value by Equations (7) and (8), then obtained the final prediction results calculate following Equation (9) or (10). The results of defuzzification and the final prediction result can be seen in Table 5.

Table 5. Defuzzification and Final Prediction Result

| $t$ | Date       | $Y(t)$ | Fuzzification | $Y^*(t)$ | $D(t)$ | $\bar{Y}(t)$ | $Y(t)$ | $\frac{|Y(t) - \bar{Y}(t)|}{Y(t)}$ |
|-----|------------|--------|---------------|----------|--------|--------------|--------|-------------------------------|
| 1   | 09/03/2020 | 740    | $A_3$         | 725      | 0      | 725          | 0,127  |
| 2   | 10/03/2020 | 830    | $A_5$         | 942      | 0      | 942          | 0,223  |
| 3   | 11/03/2020 | 770    | $A_4$         | 675      | 0      | 675          | 0,029  |
| 4   | 12/03/2020 | 695    | $A_2$         | 742      | 0      | 742          | 0,030  |
| 5   | 13/03/2020 | 720    | $A_3$         |          |        |              |        |
| ... |           |        | $\vdots$      | $\vdots$ | $\vdots$| $\vdots$    | $\vdots$| $\vdots$                      |
| 290 | 25/05/2021 | 2420   | $A_{37}$      | 2358     | 0      | 2358         | 0,025  |
| 291 | 27/05/2021 | 2420   | $A_{37}$      | 2498     | 0      | 2498         | 0,032  |
| 292 | 28/05/2021 | 2430   | $A_{37}$      | 2498     | 0      | 2498         | 0,028  |
| 293 | 31/05/2021 | 2600   | $A_{41}$      | 2503     | 100    | 2603         | 0,001  |

To predict the stock price for 02–07 June 2021, consider the fuzzification of the previous data ($t = 293$). Follow the defuzzification process until obtained the initial prediction value $Y^*(t)$. Do the same steps as before to gain the final prediction result $\bar{Y}(t)$. Table 6 shows the prediction stock price and the error with the actual prices. The comparison of the two can be seen in Figure 1.

Table 6. Prediction for 02 – 07 June 2021

| $t$ | Date       | $Y^*(t)$ | Fuzzification | $D(t)$ | $\bar{Y}(t)$ | $Y(t)$ | $\frac{|Y(t) - \bar{Y}(t)|}{Y(t)}$ |
|-----|------------|----------|---------------|--------|--------------|--------|-------------------------------|
| 294 | 02/06/21   | 2.599    | $A_{40}$      | -25    | 2.574        | 2.600  | 0.01                          |
| 295 | 03/06/21   | 2.575    | $A_{40}$      | 0      | 2.575        | 2.600  | 0.01                          |
| 296 | 04/06/21   | 2.575    | $A_{40}$      | 0      | 2.575        | 2.590  | 0.006                         |
| 297 | 07/06/21   | 2.575    | $A_{40}$      | 0      | 2.575        | 2.500  | 0.03                          |

Figure 1. Comparison of Prediction Results Using Markov Average-Based Weighted FTS with Actual Stock Prices

Copyright © 2021, Desimal, Print ISSN: 2613-9073, Online ISSN: 2613-9081
From Figure 1, it can be seen that the two lines almost match each other and the pattern is nearly identical. It means the prediction results obtained are close to the actual value with a relatively small error. To be more specific, the MAPE value is calculated to see the accuracy of the Markov Average-Based Weighted FTS model in predicting the stock price of PT Kimia Farma Tbk. Using the prediction data from March 09, 2020 - June 07, 2021, the MAPE obtained is 2.632% and categorized as very accurate.

CONCLUSIONS AND SUGGESTIONS

The weight calculation in the Markov Average-Based Weighted Fuzzy Time Series model is carried out by considering the repeated transitions from a state $i$ to another state. The Average-Based method is utilized to define the interval length based on the value of half the average difference between two consecutive data. Using this model, PT Kimia Farma Tbk stock price prediction obtained has a MAPE of 2.632%, which was classified as very accurate.

Further research is suggested to develop an R software script to help the calculation process, which is still done manually.

REFERENCES


