Forecasting model of Covid-19 cases using fuzzy time series using percentage change

A A Husain¹, B Surarso², Farikhin², B Irawanto⁴
¹Undergraduate Program of Mathematics, Departemen of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
²Departemen of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Corresponding author: aisyahauliahusain@students.undip.ac.id

Abstract. The pandemic caused by the novel corona virus (covid-19) has affected various aspects of life throughout the world. Indonesia is one of the country with a daily high number of Covid-19 spread cases. This study aims to obtain a forecasting model of Covid-19 cases that can be used to predict Covid-19 cases daily and it can increase the readiness of Covid-19 health protocols system. In this study, we get a very good model for Covid-19 forecasting in Indonesia obtained by the fuzzy time series method using frequency density-based partitioning. The universe of this method is the percentage of case changes from day to day. The percentage change as a universe in fuzzy time series forecasting method makes the results of comparison of actual data and predictions increasingly similar. We use data of the Covid 19 cases taken from the Nasional Kompas website during June 2020. Forecast results show very good with MSE value of 457.83 and small AFER value of 0.0425138%.

1. Introduction
Over the last few months of 2020, the Covid-19 pandemic has shocked the whole world. Affect all aspects of life and many take lives. The novel plague originating from Wuhan China, are spreading so fast in various parts of the world and one of them is Indonesia, which has a fairly high increase in daily cases. Forecasting the case of Covid-19 can be a reference for increase preparedness and making decisions for the right solutions in the future.

Many forecasting methods have been introduced, but if the historical data is available in the form of linguistic values, the classical time series method has not been able to solve it, so that a fuzzy time series method appears to fill the shortcomings of the classical time series method. Fuzzy Time Series (FTS) is a method introduced by Song and Chissom (1993) which is a concept used to predict problems in which actual data is formed in linguistic values. Many FTS methods were developed, including the FTS Chen method, FTS using percentage change, weighted FTS, FTS Sah and Degtiarev, FTS Cheng [1][2][3].

Jilani, Burney, and Ardil introduced a method FTS forecasting based frequency density based partitioning the historical enrollment data with case study of enrollments at the University of Alabama. The proposed method belongs to the class k-step first-order univariate time variant method. This method can obtain a better level of forecast accuracy than the existing methods [6]. The method proposed by Jilani, Burney, and Ardil was modified again by Meredith Stevenson, and John E. [4] using the percentage change from year to year as the universe of discourse and shows better results of
accuracy than previous existing methods, including the methods proposed by Jilani, Burney, and Ardil [4].

In this paper, we will discuss the Covid-19 case forecasting model using fuzzy time series with the percentage change of daily case covid-19 as the universe of discourse. The result of the forecast method will be evaluated by examine value of MSE and AFER. The rest of this paper is organized as follows. In Section 1 introduction, Section 2 brief definition of fuzzy time series, Section 3 the method of forecasting and result, Section 4 conclusion.

2. Fuzzy time series definition

There are several definitions of fuzzy time series.

Definition 1 [5],[1]
Assume \( Y(t) \subset R \) (real line), \( t=...,0,1,2,.... \), to be a universe of discourse defined by the fuzzy set \( f_i(t) \). \( F(t) \) consisting of \( f_i(t),i=1,2,.... \) is defined as a fuzzy time series on \( Y(t) \). At that, \( F(t) \) can be understood as a linguistic variable, whereas \( f_i(t),i=1,2,.... \) are a possible linguistic value of \( F(t) \). Using fuzzy time series, we define \( F(t) = F(t - 1) \circ R(t - 1) \) where \( R(t,t - 1) \) is a fuzzy relation and “o” is the max-min composition operator, then \( F(t) \) is caused by \( F(t - 1) \), denote by the fuzzy logical relationship or FLR “\( F(t - 1) \rightarrow F(t) \)” where \( F(t) \) and \( F(t - 1) \) are fuzzy sets. If \( F(t - 1) = A_1 \) and \( F(t) = A_j \), where \( A_1 \) and \( A_j \) is fuzzy sets, then the fuzzy logical relationship between \( F(t - 1) \) and \( F(t) \) is represented by “\( A_1 \rightarrow A_j \)”.

Definition 2 [6]
A fuzzy set \( A \) of the data \( Y \), \( Y = \{y_1,y_2,...,y_n\} \) defined as \( A = \{f_A(y_i)\},i=1,2,....,n \) where \( f_A \) is the member’s function of the fuzzy set \( A \), \( f_A: Y \rightarrow [0,1], f_A(y_i) \) denotes the grade of membership of \( y_i \) in the fuzzy set \( A \), and \( 1 \leq i \leq n \).

Definition 3 [7]
The member function of triangle is presented by 3 parameters fuzzy \( \tilde{\alpha} = (\alpha_1,\alpha_2,\alpha_3) \), with \( \alpha_1, \alpha_2, \alpha_3 \in R \). This curve is essentially a combination of function as this following equations:

\[
\mu_{\tilde{\alpha}}(x) = \begin{cases} 
\frac{x - \alpha_1}{\alpha_3 - \alpha_1}, & \alpha_1 \leq x \leq \alpha_2 ; \\
\frac{\alpha_3 - x}{\alpha_3 - \alpha_2}, & \alpha_2 \leq x \leq \alpha_3 ; \\
0, & \text{else} 
\end{cases}
\]

3. Forecasting method using percentage change and result

The method for forecasting Covid-19 cases in Indonesia using fuzzy time series with percentage change [1] of daily cases as universe discourse will be explained below.

Step 1 : Define the universe of discourse \( U \) and partition it into intervals \( u_1,u_2,..,u_n \) of equal length.

The percentage change of daily cases Covid-19 is given in Table I and ranges from 1.91% to 3.75%.

Next we will look for the universe \( U \) as follows then partition \( U \) into seven equal intervals.

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

\[
U = [1.91 - 0.11, 3.75 + 0.25]
\]

\[
U = [1.8,3.8]
\]
| Date | Daily case | Change | Percentage |
|------|------------|--------|------------|
| 1    | 26940      |        |            |
| 2    | 27549      | 609    | 2.26       |
| 3    | 28233      | 684    | 2.48       |
| 4    | 28818      | 585    | 2.07       |
| 5    | 29521      | 703    | 2.44       |
| 6    | 30514      | 993    | 3.36       |
| 7    | 31186      | 672    | 2.2        |
| 8    | 32033      | 847    | 2.72       |
| 9    | 33076      | 1043   | 3.26       |
| 10   | 34316      | 1240   | 3.75       |
| 11   | 35295      | 979    | 2.85       |
| 12   | 36406      | 1111   | 3.15       |
| 13   | 37420      | 1014   | 2.79       |
| 14   | 38277      | 857    | 2.29       |
| 15   | 39294      | 1017   | 2.66       |
| 16   | 40400      | 1106   | 2.81       |
| 17   | 41431      | 1031   | 2.55       |
| 18   | 42762      | 1331   | 3.21       |
| 19   | 43803      | 1041   | 2.43       |
| 20   | 45029      | 1226   | 2.8        |
| 21   | 45891      | 862    | 1.91       |
| 22   | 46845      | 954    | 2.08       |
| 23   | 47896      | 1051   | 2.24       |
| 24   | 49009      | 1113   | 2.32       |
| 25   | 50187      | 1178   | 2.4        |
| 26   | 51427      | 1240   | 2.47       |
| 27   | 52812      | 1385   | 2.69       |
| 28   | 54010      | 1198   | 2.27       |
| 29   | 55092      | 1082   | 2          |
| 30   | 56385      | 1293   | 2.35       |

**Step 2:** Find the density based distribution of the daily percentage change by sorting the data into the corresponding intervals shown in Table 2. Then determine the number of percentage data that falls into each interval. Table 2 contains the density based distribution of the percentage data displayed in Table 1 with seven intervals. Find the interval that has the largest number of data frequencies, the second largest, and the third largest then divide each interval into 4, 3, and 2 sub-intervals with equal length, respectively and sequentially. Let all subsequent intervals remain unchanged in length. After completing this step, the universe of discourse is divided into 16 intervals shown in Table 3.
**Table 2.** Frequency density based distribution of the daily percentage change of data

| Name | Interval  | Frequency |
|------|-----------|-----------|
| $u_1$ | [1.8, 2.09] | 4         |
| $u_2$ | [2.09, 2.38] | 7         |
| $u_3$ | [2.38, 2.67] | 7         |
| $u_4$ | [2.67, 2.96] | 6         |
| $u_5$ | [2.96, 3.25] | 2         |
| $u_6$ | [3.25, 3.54] | 2         |
| $u_7$ | [3.54, 3.80] | 1         |

**Table 3.** Fuzzy sub-interval using frequency density based partitioning

| Linguistic Sub-interval | Sub-interval |
|-------------------------|--------------|
| $X_1$                   | 1.80, 1.95   |
| $X_2$                   | 1.95, 2.09   |
| $X_3$                   | 2.09, 2.16   |
| $X_4$                   | 2.16, 2.24   |
| $X_5$                   | 2.24, 2.31   |
| $X_6$                   | 2.31, 2.38   |
| $X_7$                   | 2.38, 2.45   |
| $X_8$                   | 2.45, 2.53   |
| $X_9$                   | 2.53, 2.60   |
| $X_{10}$                | 2.60, 2.67   |
| $X_{11}$                | 2.67, 2.77   |
| $X_{12}$                | 2.77, 2.86   |
| $X_{13}$                | 2.86, 2.96   |
| $X_{14}$                | 2.96, 3.25   |
| $X_{15}$                | 3.25, 3.54   |
| $X_{16}$                | 3.54, 3.80   |

**Step 3:** Define each fuzzy set $X_i$ based on the re-divided intervals and fuzzify the historical enrollments shown in table I, where fuzzy set $X_i$ denotes a linguistic value of the year to year percentage change represented by a fuzzy set. As in [9] we use a triangular membership function to define the fuzzy sets $X_i$ [10], the result shown in table 4.

**Table 4.** Fuzzy membership function

| Date | Daily Case | Percentage Change | Linguistic |
|------|------------|-------------------|------------|
| 1    | 26940      | 2.26              | $X_5$      |
| 2    | 27549      | 2.48              | $X_8$      |
| 3    | 28233      | 2.07              | $X_2$      |
| 4    | 28818      | 2.44              | $X_7$      |
| 5    | 29521      |                   |            |
Step 4: Built FLR (Fuzzy Logical Relationship) and FLRG (Fuzzy Logical Relationship Group) that will be used to change the value of the fuzzy rule to become the value of the crisp value. FLR “$X_p, X_q, X_r \rightarrow X_s$” denotes that “if the fuzzified of day p, q, and r are $X_p$, $X_q$, and $X_r$ respectively, then the fuzzified data of day (s) is $X_s$” [5].

Step 5: Defuzzify the fuzzy data using the forecasting formula (see [9])

$$t_j = \begin{cases} 
\frac{1.5}{\frac{1}{a_1} + \frac{0.5}{a_2}}, & j = 1 \\
\frac{2}{\frac{0.5}{a_{j-1}} + \frac{0.5}{a_j} + \frac{1}{a_{j+1}}}, & 2 \leq j \leq n-1 \\
\frac{1.5}{\frac{0.5}{a_{n-1}} + \frac{1}{a_n}}, & j = n 
\end{cases}$$

Where $a_{j-1}, a_j, a_{j+1}$ are the midpoints of the fuzzy intervals $X_{j-1}, X_j, X_{j+1}$ respectively. $t_j$ yields the predicted day to day percentage change of covid-19 case. Use the predicted percentage on the previous day’s case to determine the forecasted case. The forecasted case is provided in Table 6.

|   |   |   |   |
|---|---|---|---|
| 6 | 30514 | 3.36 | $X_{15}$ |
| 7 | 31186 | 2.2 | $X_4$ |
| 8 | 32033 | 2.72 | $X_{11}$ |
| 9 | 33076 | 3.26 | $X_{15}$ |
| 10 | 34316 | 3.75 | $X_{16}$ |
| 11 | 35295 | 2.85 | $X_{12}$ |
| 12 | 36406 | 3.15 | $X_{14}$ |
| 13 | 37420 | 2.79 | $X_{12}$ |
| 14 | 38277 | 2.29 | $X_5$ |
| 15 | 39294 | 2.66 | $X_{10}$ |
| 16 | 40400 | 2.81 | $X_{12}$ |
| 17 | 41431 | 2.55 | $X_9$ |
| 18 | 42762 | 3.21 | $X_{14}$ |
| 19 | 43803 | 2.43 | $X_7$ |
| 20 | 45029 | 2.8 | $X_{12}$ |
| 21 | 45891 | 1.91 | $X_1$ |
| 22 | 46845 | 2.08 | $X_2$ |
| 23 | 47896 | 2.24 | $X_5$ |
| 24 | 49009 | 2.32 | $X_6$ |
| 25 | 50187 | 2.4 | $X_7$ |
| 26 | 51427 | 2.47 | $X_8$ |
| 27 | 52812 | 2.69 | $X_{11}$ |
| 28 | 54010 | 2.27 | $X_5$ |
| 29 | 55092 | 2. | $X_2$ |
| 30 | 56385 | 2.35 | $X_6$ |
Step 6: Examine and calculate the accuracy of fuzzy time series method to forecast daily Covid-19 case using percentage change as universe of discourse by MSE and AFER. Here the equation:

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

$$AFER = \frac{A_i - F_i}{n} \times 100\%$$

Where $A_i$ denotes the actual case and $F_i$ denotes the forecasting case. AFER value can be used to know the criteria of the forecasting method.

| AFER Value | Criteria of Forecasting Result |
|------------|--------------------------------|
| <10%       | Very good                     |
| 10%-20%    | Good                           |
| 20%-50%    | Good enough                   |
| >50%       | Bad                            |

The result of calculate MSE and AFER value shown in Table 6.

| Date | Daily case | Percentage change | Forecast percentage | Forecast case | $A_i - F_i$ | $(A_i - F_i)^2$ | $|A_i - F_i|/A$ |
|------|------------|-------------------|---------------------|---------------|-------------|----------------|----------------|
| 1    | 26940      | 2.26              | 2.28                | 27554.2       | -5.2        | 27.37          | 0.000189916    |
| 2    | 27549      | 2.26              | 2.28                | 27554.2       | -5.2        | 27.37          | 0.000189916    |
| 3    | 28233      | 2.48              | 2.48                | 28237.6       | -4.6        | 20.95          | 0.000162114    |
| 4    | 28818      | 2.07              | 2.03                | 28810.8       | 7.2         | 51.84          | 0.000249852    |
| 5    | 29521      | 2.44              | 2.43                | 29510.9       | 10.1        | 101.97         | 0.000342055    |
| 6    | 30514      | 3.36              | 3.37                | 30505.4       | 8.6         | 73.62          | 0.000281195    |
| 7    | 31186      | 2.2               | 2.2                 | 31176.5       | 9.5         | 89.51          | 0.000303379    |
| 8    | 32033      | 2.72              | 2.72                | 32024.5       | 8.5         | 71.56          | 0.000264081    |
| 9    | 33076      | 3.26              | 3.37                | 33053.8       | 8.6         | 73.62          | 0.000281195    |
| 10   | 34316      | 3.75              | 3.57                | 34285.6       | 30.4        | 925.85         | 0.000886694    |
| 11   | 35295      | 2.85              | 2.83                | 35255.9       | 39.1        | 1532.42        | 0.001109111    |
| 12   | 36406      | 3.15              | 3.18                | 36377         | 29          | 841.58         | 0.000796845    |
| 13   | 37420      | 2.79              | 2.83                | 37406.5       | 13.5        | 183.36         | 0.000361868    |
| 14   | 38277      | 2.29              | 2.28                | 38259.3       | 17.7        | 312.36         | 0.000461735    |
| 15   | 39294      | 2.66              | 2.66                | 39277         | 17          | 288.18         | 0.000432019    |
| 16   | 40400      | 2.81              | 2.83                | 40388.6       | 11.4        | 130.78         | 0.000283069    |
| 17   | 41431      | 2.55              | 2.57                | 41426.6       | 4.4         | 19.8           | 0.000107405    |
| 18   | 42762      | 3.21              | 3.18                | 42743.9       | 18.1        | 327.09         | 0.000422936    |
| 19   | 43803      | 2.43              | 2.43                | 43782.6       | 20.4        | 416.51         | 0.000465915    |
4. Conclusion

Based on the result of forecasting daily covid-19 cases during the month of Juni 2020 in Indonesia using Fuzzy time series with percentage change as a universe of discourse shows the forecast value that is very close to the actual data so that it produces an excellent AFER value of 0.042514 % and this means that the forecasting model of covid-19 can be categorized as very good. This excellent forecasting model using FTS can be used as an alternative model for predicting the number of Covid-19 cases in the future. Prediction with this model can be done using prediction methods such as Markov Chain so that it can be useful for making a decision and right policies during the pandemic.

![Figure 1. Comparison between actual data and forecast](image)

References

[1] Chen S M 1996 *Fuzzy sets Syst.* 81 311-318
[2] Chen S M and Hsu C C 2004 *Inter. J. Appl. Sci. Eng.* 2 234
[3] Cheng S H, Chen S M and Jian W S 2015 *Elsevier* 13:39 2-15
[4] Hwang J R, Chen S M and Lee C H 1998 *Fuzzy sets Syst.* 100 217-228
[5] Hao W, Haiming L and Jiancheng J 2019 *Appl. Soft Comput. J.* 78 109-118
[6] Irawanto B, Ningrum R W, Wulandari R, B Surarso and Farikhin 2019 *J. Phys. Conf. Ser.* 5 1321
[7] Jilani T A, Burney S M A, and Ardil C 2008 *Int. J. Comput. Intel.* 4 112-114
[8] Jilani T A, Burney S M A, and Ardil C 2010 *Int. J. Comput. Inf. Eng.* 4 1194
[9] Sah M and Degtiarev K Y 2005 *World Acad. Sci. Eng. Technol.* 1 375-378
[10] Song Q and Chissom B S 1994 *Fuzzy sets Syst.* 54 269-277
[11] Song Q and Chissom B S 1994 *Fuzzy sets Syst.* 62 1-8
[12] Stevenson M and Porter J E 2009 *World Acad. Sci. Eng. Technol.* 55 154-155