



Forecasting the Number of Tuberculosis Patients Using Automatic Clustering and Fuzzy Logical Relationship Method

Rahmawati*, Sarbaini, Ade Novia Rahma, Tri Uci Lestari, Fitri Aryani

Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Email: rahmawati@uin-suska.ac.id

ABSTRACT

The lungs are respiratory organs that have an important role in the human body, so if they are infected they will have a very dangerous impact. Tuberculosis is an infectious disease caused by the bacterium *Mycobacterium tuberculosis bacillus*, which infects the lungs and can potentially cause death. The goal of this study is to predict how many people in Kampar Regency will have tuberculosis in 2022. The Automatic Clustering and Fuzzy Logical Relationship (ACFLR) technique is employed. ACFLR is one approach for modeling time series data that incorporates the idea of fuzzy logic. According to a number of previous studies, this method is accurate. Secondary data from the Kampar District Health Office from 2017 to 2021 comprise the analyzed data. The result of research it was gotten to estimate the quantity of tuberculosis patients in 2022 upwards of 944 individuals with MAPE of 0.0882, the precision of gauging aftereffects of 99.9118, and an expansion in the quantity of tuberculosis victims from 2021 to 2022 upwards of 4 individuals. The forecasting result can be used by the government and medical staff so that they can take the right initial steps to reduce the increasing rate of tuberculosis.

Keywords: Automatic Clustering; Forecasting; Fuzzy Logical Relationship; MAPE; Tuberculosis.

Copyright © 2023 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0/>)

INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by the bacterium *Mycobacterium Tuberculosis Bacillus* which infects the lungs [1] and has the potential to cause death. Infectious diseases are very dangerous if not handled properly, therefore it is necessary to know the initial cause [2]. TB transmission occurs through the air when TB sufferers splash mucus or phlegm when coughing or sneezing. TB bacteria will come through the mucus and be carried into the air. Then TB bacteria will enter other people's bodies through the air they breathe [3]. Indonesia is ranked third with the highest number of TB sufferers in the world after India and China, the number of TB sufferers in Indonesia is around 5.8% of the number of TB sufferers in the world, and it is estimated that every year there are 528,000 TB sufferers with deaths of around 91,000 people [4].

Based on the results of the report of the Kampar District Health Office, Riau Province. In 2017, there were 1,069 TB sufferers. In 2018 there were 1,079 patients, this has an increase from the previous year, in 2019 TB sufferers there were 1,006, and from 2020 to 2021 TB sufferers there were 938 and 940. In 2021 TB sufferers experienced an increase from the previous year, although not as many as in 2018. However, the

significant and slight increase in TB sufferers requires all parties to be able to commit and cooperate in TB management because this has a considerable impact not only from health aspects but also social and economic aspect [5]. Therefore, to see how TB disease progresses in the next year requires forecasting or prediction. Forecasting is defined as a way of systematically and pragmatically predicting what will happen in the future based on past relevant data, so forecasting results are expected to provide greater objectivity [6].

Solving the problem of forecasting the number of tuberculosis sufferers can be optimized using the application of fuzzy science, namely using the Automatic Clustering and Fuzzy Logical Relationship (ACFLR) method. This method is a modified method of the fuzzy time series, the advantage of fuzzy time series modeling is being able to formulate a problem based on expert knowledge or empirical data [26] where the fuzzy time series is a method proposed by Song and Chissom to solve problems in the form of forecasting if the historical data is in the form of linguistic values [7][8]. Fuzzy time series is a data forecasting method that uses fuzzy principles as its basis [9][10]. An automatic clustering algorithm is used to determine the interval length [11], every the length of the interval is determined first at the beginning of the calculation process due to the interval process very influential in the formation of Fuzzyrelationship and the final result [12], and an algorithm is used to determine to forecast [13][14]. The advantage of the ACFLR method is that it can determine forecasts using short data and based on previous research the ACFLR method has a high level of accuracy compared to other time series methods. The disadvantages of the ACFLR method is in determining the length of the interval and interpretation of the output of the forecasting model, different interpretation will produce different forecasting results. Fuzzy logic too related to artificial intelligence [15], besides that fuzzy can also be used in grouping to diagnose a disease [16][17].

Fuzzy Clustering is a technique for determining optimal clusters in a vector space based on the normal Euclidian form for the distance between vectors [18]. There are many fuzzy clustering methods such as k-means, c-means and others [19]. Besides that there are many more fuzzy theories such as in [20][21] regarding FIS Sugeno, then [22][23] regarding fuzzy mamdani, then [24] regarding FIS tsukomoto, and the last [25] discuss about fuzzy linear programming that is one model that is often used to solve optimization problems.

In general, Fuzzy sets defined as a class of numbers with fuzzy boundaries [26]. Research conducted by [11] in 2018 regarding the analysis of gold price forecasting using the ACFLR method resulted in a gold price forecast with a MAPE value of 5.3%, which means that the accuracy value of the forecast results with the ACFLR method is outstanding. Further research by [13] Makassar City resulted in a forecasting of the number of residents from 2017 to 2019 decreased. In 2020 to 2021, it increased, with a MAPE value of <10% which means that the level of accuracy of forecasting results using the ACFLR method is excellent. Further research by [27] 2021 regarding forecasting the number of foreign tourist visits using the fuzzy time series ACFLR resulted in a forecasting error value with a MAPE calculation of 6.22% which means that the error value of the forecasting results using the ACFLR method is excellent. Based on the events above, fuzzy theory has a big role in the health sector [28]. Tuberculosis is an infectious disease where the most significant impact is death, so to reduce the mortality rate due to TB, it is necessary to forecast the number of TB sufferers so that the health department can anticipate or provide counseling to the community to maintain cleanliness and health, so the purpose of this study is the results of forecasting the number of TB

sufferers in 2022 in Kampar Regency using the ACFLR. This research is expected to be used to anticipate an event for a program and as a benchmark for the number of TB sufferers in Kampar Regency Riau Province. So that the government and medical personnel can take appropriate initial steps to reduce the increase in TB disease.

METHODS

The ACFLR method is a method used to form intervals and forecasting arranged into a single unit [29]. The ACFLR algorithm is divided into three main stages. The first stage carried out is to form intervals using an automatic clustering algorithm, the interval referred to is the distance for each clustering. The steps used are:

1. Sort the data from smallest to largest, and assume that no data is the same. Shown as follows:

$$d_1, d_2, d_3, \dots, d_i, \dots, d_n$$

Based on the row above, calculate the value of "Average_diff" using the following equation:

$$\text{Average_diff} = \frac{\sum_{i=1}^{n-1} (d_{i+1} - d_i)}{n-1} \quad (1)$$

The Average_diff value indicates the average difference between each adjacent data in successive data.

2. Put the first numeric data (the smallest data in the data sequence) on an existing cluster (current cluster). Based on the value of average_diff it will be determined whether numerical data can then be entered into the current cluster or needs to be entered into a new cluster based on the following principles:
 - i. Principle 1: It is assumed that the current cluster is the first cluster and it is assumed that numeric data d_2 is the closest numerical data from d_1 . Displayed as follows:

$$\{d_1\}, d_2, d_3, \dots, d_n$$

If $d_2 - d_1 \leq \text{average_diff}$, then d_2 it is put into the current cluster that contains d_1 . If not, then a new cluster is created for and assigned d_2 a new cluster that loads into the current d_2 cluster.

- ii. Principle 2: Assume that the current cluster is not the first cluster and assume that d_i it is the largest in the current cluster, and assume d_j that it is d_i the closest to. Displayed as follows:

$$\{d_1, \dots\}, \dots, \{\dots\}, \{\dots, d_i\}, d_j, \dots, d_n$$

If $d_j - d_i \leq \text{average_diff}$ and $d_j - d_i \leq \text{cluster_diff}$, then it is put into d_j the current cluster that contains d_i . If not, then a new cluster forms for d_j and assigns a new cluster that loads d_j into the current cluster. Cluster_diff is the average difference between every two adjacent data in a cluster that has been formed. The calculation of cluster_diff can be determined by the formula:

$$\text{cluster_diff} = \frac{\sum_{i=1}^{n-1} (c_{i+1} - c_i)}{n-1} \quad (2)$$

where c_1, c_2, \dots, c_n is the data in an existing cluster.

- iii. Principle 3: Assume that the current cluster is not the first cluster and d_j is the only one in the current cluster. Assumed that d_k it was data adjacent to data d_j and d_i was the largest in the previous cluster, and was shown as follows:

$$\{d_1, \dots\}, \dots, \{\dots\}, \{\dots, d_i\}, \{d_j\}, d_k, \dots, d_n$$

If $d_k - d_j \leq \text{average_diff}$ and $d_k - d_j \leq d_j - d_i$ then d_k is put into the current cluster that contains d_j . If not, then a new cluster forms for d_k and assigns a new cluster that loads d_k into the current cluster.

3. Based on the clustering in Step (2), readjust the members of the cluster using the following principle:
 - i. Principle 1: If a cluster has more than two data, then the smallest numerical data and the largest numerical data are retained and delete the other data.
 - ii. Principle 2: If a cluster has only two numerical data, then both data are preserved.
 - iii. Principle 3: If on a cluster only there is one numerical data d_q , then the values " $d_q - \text{average_diff}$ " and " $d_q + \text{average_diff}$ " are placed in the cluster and d_q removed from cluster. If the following situations occur, the cluster needs to be adjusted again:
 - a. Situation 1: If principle 3 occurs on the first cluster, then delete the value from " $d_q - \text{average_diff}$ " and set d_q it as its replacement.
 - b. Situation 2: If principle 3 occurs on the last cluster, then delete the value from " $d_q + \text{average_diff}$ " and set d_q the value as its replacement value.
 - c. Situation 3: If the value of " $d_q - \text{average_diff}$ " is smaller than the smallest data in the previous cluster then principle 3 does not apply.
4. Transform the formed clusters into adjacent intervals. It is assumed that the result of clustering in Step (3) is as follows:

$$\{d_1, d_2\}, \{d_3, d_4\}, \{d_5, d_6\}, \dots, \{d_r\}, \{d_s, d_t\}, \dots, \{d_{n-1}, d_n\}$$

The clusters that have been formed are converted into intervals with the following sub-steps:

- i. The first cluster $\{d_1, d_2\}$ is converted into the interval $[d_1, d_2]$.
 - ii. If the current interval $[d_i, d_j]$ and the current cluster $\{d_k, d_l\}$, then
 - a. If $d_j \geq d_k$, then an interval $[d_j, d_l]$ is formed which is set to the current interval and the next cluster $\{d_m, d_n\}$, set to the current cluster.
 - b. If $d_j < d_k$, then $\{d_k, d_l\}$ it is converted into the interval $[d_k, d_l]$ and a new interval $[d_j, d_k]$ is formed between $[d_i, d_j]$ and $\{d_k, d_l\}$. $[d_k, d_l]$ is set to the current interval and next cluster $\{d_m, d_n\}$ is further set to the current cluster. If the current interval $[d_i, d_j]$ and the current cluster $\{d_k\}$, then the current interval $[d_i, d_j]$ is converted into $[d_i, d_k]$ and set to the current interval and the next cluster is current cluster.
 - c. The last interval is the closed-hose interval $[d_m, d_n]$.
 - iii. Double-check all established intervals until the entire cluster is transformed into intervals.
5. Each interval obtained in Step (4) is divided into p sub-intervals, where the values p are adjusted according to the amount of data used.

After forming the interval using the automatic clustering algorithm, the second stage that is done is to calculate the forecast value with the ACFLR algorithm using the following steps:

1. Set of Universes

The set of U universes is determined according to existing historical data by finding the minimum value and the maximum value. Next calculate the midpoint value of each interval obtained from Step (5).

2. Defining Fuzzy Sets

It is assumed n that there is an interval obtained from Step (5) is u_1, u_2, \dots, u_n further every fuzzy set A_i with the following conditions: $1 \leq i \leq n$ [30].

$$\mu_j(u_i) = \begin{cases} 1 & ; j = i \\ 0,5 & ; j = i - 1 \text{ or } j = i + 1 \\ 0 & ; \text{others} \end{cases} \quad (3)$$

Rule 1 : If there is with then $\mu_j(u_i)$ with $j = i$ then $\mu_j(u_i) = 1$

Rule 2 : If there is $\mu_j(u_i)$ with $j = i - 1$ or $j = i + 1$ then $\mu_j(u_i) = 0,5$

Rule 3: If it does not include the above 2 conditions then $\mu_j(u_i) = 0$

Using Equation (3) the following results were obtained:

$$\begin{aligned} A_1 &= \frac{1}{u_1} + \frac{0,5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \dots + \frac{0}{u_{n-1}} + \frac{0}{u_n} \\ A_2 &= \frac{0,5}{u_1} + \frac{1}{u_2} + \frac{0,5}{u_3} + \frac{0}{u_4} + \dots + \frac{0}{u_{n-1}} + \frac{0}{u_n} \\ A_3 &= \frac{0}{u_1} + \frac{0,5}{u_2} + \frac{1}{u_3} + \frac{0,5}{u_4} + \dots + \frac{0}{u_{n-1}} + \frac{0}{u_n} \\ &\vdots \\ A_n &= \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \dots + \frac{0,5}{u_{n-1}} + \frac{1}{u_n} \end{aligned}$$

3. Fuzzification Process

Based on the definition of fuzzy sets in Step (2), the next step is to fuzzify each data into fuzzy sets. If the data belongs to u_i where $1 \leq i \leq n$, then the data is fuzzified in A_i , and so on until all data is fuzzified.

4. Fuzzy Logical Relationship

In this step, form a fuzzy relationship based on the data obtained in Step (3) i.e. if the data has been fuzzified from the year t and $t + 1$ is $A_j \rightarrow A_k$, then a fuzzy logical relationship is formed " $A_j \rightarrow A_k$ ", which A_j is referred to as the current condition while referred to A_k as the next condition. Based on the current state of the fuzzy logical relationship, fuzzy logical relationship groups are formed, where fuzzy logical relationships that have the same current linguistic value are incorporated into the same fuzzy logical relationship.

5. Defuzzification Process

In this step, the results of forecasting from historical data will be determined by taking into account the following principles:

- i. If the result of fuzzification in the year t is A_j and there is only one fuzzy logical relationship in fuzzy logical relationship groups where the current condition is $A_j \rightarrow A_k$, then the forecasted data in the year $t + 1$ is m_k where m_k is the middle value of the interval u_k .
- ii. If the fuzzified data from the year t is A_j and there is a fuzzy logical relationship in fuzzy logical relationship groups where the current state is A_j , then the forecasted data in the year $t + 1$ is calculated using the following formula:

$$\frac{(x_1 \times m_{k1}) + (x_2 \times m_{k2}) + \dots + (x_p \times m_{kp})}{x_1 + x_2 + \dots + x_p} \quad (4)$$

$x_1, x_2, x_3, \dots, x_p$ is the number of fuzzy logical relationships in fuzzy logical relationship groups, and $m_{k1}, m_{k2}, \dots, m_{kp}$ is the middle value of the interval $u_{k1}, u_{k2}, \dots, u_{kp}$.

- iii. If the result of fuzzification in the year t is A_j and there is a fuzzy logical relationship in fuzzy logical relationship groups where the current state is $A_j \rightarrow \emptyset$, the symbol \emptyset is expressed as an unknown value, then the forecasting data for the year $t + 1$ is m_i , where m_i is the middle value of the interval u_i .

Evaluation of forecasting results is the last stage of the ACFLR method which is a comparison and measurement process in order to determine the accuracy of forecasting results from actual data, there are several methods to calculate the level of accuracy in a forecast, one of which is the Mean Absolute Percentage Error (MAPE) [31]. MAPE is a more ideal way to calculate error, because it states that the percentage of error of forecast results against actual conditions over a certain period shows a low percentage. MAPE calculation is done by calculating the difference between the original data and the forecasting data divided by the amount of original data that is absolutized and then calculated into a percentage. The MAPE value is obtained from the calculation using the following formula:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{(y_t - \hat{Y}_t)}{y_t} \right| \times 100\% \tag{5}$$

Information:

- n : Amount of data
- \hat{Y}_t : Forecasting value in year to t
- y_t : Original value in year to t

According to [31] If the resulting MAPE is <10%, the forecasting result is perfect. The forecasting result is good if the resulting MAPE is between 10%-20%. If the resulting MAPE is between 20%-50%, then the forecasting result is said to be sufficient. Furthermore, if the resulting MAPE is >50%, the forecasting result is said to be inaccurate, and it is advisable not to use it.

RESULTS AND DISCUSSION

The data used by the authors in this study is secondary data on the number of tuberculosis sufferers from 2017 to 2021 sourced from the Kampar District Health Office Riau Province. The population of this study was Kampar District, and the sample used by the author was 21 health centers in Kampar Regency. Data on the number of tuberculosis patients in Kampar Regency Riau Province from 2017 to 2021 experienced fluctuations. Here is the data on the number of tuberculosis sufferers from 2017 to 2021.

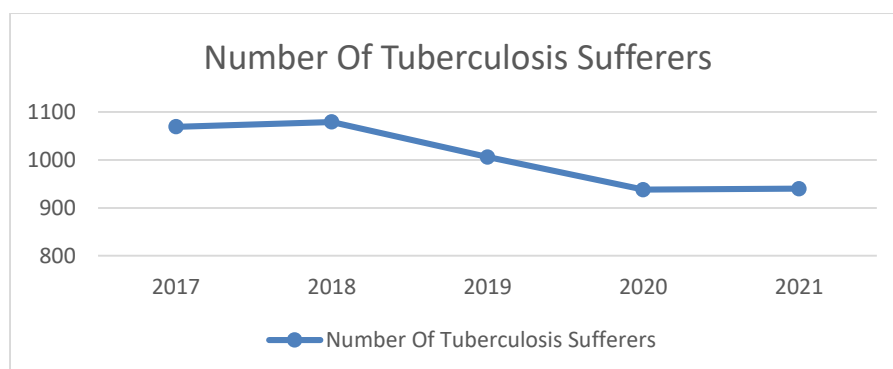


Figure 2. Data on the Number of TB Patients in 2017-2021 in Kampar Regency

Forecasting the number of TB patients using the ACFLR method consists of three main stages. The following are the first stages, namely:

1. Applying an automatic clustering algorithm to cluster historical data on the number of TB sufferers from 2017-2021 into intervals. The steps performed are:
 - i. Sort data from smallest to largest.
Data order: 938, 940, 1,006, 1,069, 1,079.
Based on the data, the average_diff value is obtained by using the formula in the Equation (1), so obtained

$$\text{average_diff} = \frac{\sum_{i=1}^{5-1} (940-938) + (1.006-940) + (1.069-1.006) + (1.079-1.069)}{5-1} = 35.25$$
 - ii. Based on the average_diff value, a cluster will be formed from data that has been sorted using the three principles and with Equation (2) obtained

$$\text{cluster_diff} = \frac{\sum_{i=1}^{2-1} (940-938)}{2-1} = 2$$
, then the clustering results are obtained as follows:
 - a. First Cluster = {938 ; 940}
 - b. Second Cluster = {1.006}
 - c. Third Cluster = {1.069 ; 1.079}
 - iii. Based on the clustering results in step (ii), the clustering members are adjusted again and obtained as follows:
 - a. First Cluster = {938 ; 940}
 - b. Second Cluster = {970,75 ; 1.041,25}
 - c. Third Cluster = {1.069 ; 1.079}
 - iv. Based on the cluster results in Step (iii) the following intervals are obtained:
 - a. First Interval = [938; 940]
 - b. Second Interval = [940; 970.75]
 - c. Third Interval = [970.75; 1,041.25]
 - d. Fourth Interval = [1,041.25 ; 1,069]
 - e. Fifth Interval = [1,069; 1,079]
 - v. Divide the interval into sub-intervals. In this study, the author used the value $p=5$ because the data used was 5 years. Based on the results of the division of the above intervals, the following sub-intervals are obtained
 - a. First Interval = $\frac{940-938}{5} = 0,4$
 - b. Second Interval = $\frac{970,75-940}{5} = 6,15$
 - c. Third Interval = $\frac{970,75-1.041,25}{5} = 14,1$
 - d. Fourth Interval = $\frac{1.041,25-1.069}{5} = 5,55$
 - e. Fifth Interval = $\frac{1.069-1.079}{5} = 2$

Table 1. Results of the universe set and middle value

Universe Set	Middle Value
$u_1 = [938,0 ; 938,4)$	$m_1 = 938,2$
$u_2 = [938,4 ; 938,8)$	$m_2 = 938,6$
$u_3 = [938,8 ; 939,2)$	$m_3 = 939,0$
$u_4 = [939,2 ; 939,6)$	$m_4 = 939,4$
$u_5 = [939,6 ; 940,0)$	$m_5 = 939,8$
$u_6 = [940 ; 946,15)$	$m_6 = 943,075$
$u_7 = [946,15 ; 952,30)$	$m_7 = 949,225$
$u_8 = [952,30 ; 958,45)$	$m_8 = 955,375$

Universe Set	Middle Value
$u_{10} = [964,60 ; 970,75)$	$m_{10} = 967,675$
$u_{11} = [970,75 ; 984,85)$	$m_{11} = 977,8$
$u_{12} = [984,85 ; 998,95)$	$m_{12} = 991,9$
$u_{13} = [998,95 ; 1.013,05)$	$m_{13} = 1.006$
$u_{14} = [1.013,05 ; 1.027,15)$	$m_{14} = 1.020,1$
$u_{15} = [1.027,15 ; 1.041,25)$	$m_{15} = 1.034,2$
$u_{16} = [1.041,25 ; 1.046,80)$	$m_{16} = 1.044,025$
$u_{17} = [1.046,80 ; 1.052,35)$	$m_{17} = 1.049,575$
$u_{18} = [1.052,35 ; 1.057,90)$	$m_{18} = 1.055,125$
$u_{19} = [1.057,90 ; 1.063,45)$	$m_{19} = 1.060,675$
$u_{20} = [1.063,45 ; 1.069,00)$	$m_{20} = 1.066,225$
$u_{21} = [1.069 ; 1.071)$	$m_{21} = 1.070$
$u_{22} = [1.071 ; 1.073)$	$m_{22} = 1.072$
$u_{23} = [1.073 ; 1.075)$	$m_{23} = 1.074$
$u_{24} = [1.075 ; 1.077)$	$m_{24} = 1.076$
$u_{25} = [1.077 ; 1.079)$	$m_{25} = 1.078$

2. The second stage is to calculate the forecasting value with the following steps:
 - i. Determine the set of universes, i.e., $U = [938 ; 1.079]$. Furthermore, calculate the middle value in a way so that based on step (v) in the first stage, the result of the middle value on the Table 1 for each interval $m_i u_i$. With is as follows: $1 \leq i \leq 25$
 - ii. It is assumed that there are intervals. Next, define each u_1, u_2, \dots, u_{25} fuzzy set A_i where $1 \leq i \leq n$ and $n = 25$ by using the formula in equation (3)

Based on research [15], So that the following results are obtained:

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \dots + \frac{0}{u_{24}} + \frac{0}{u_{25}}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \dots + \frac{0}{u_{24}} + \frac{0}{u_{25}}$$

$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \dots + \frac{0}{u_{24}} + \frac{0}{u_{25}}$$

⋮

$$A_{25} = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \dots + \frac{0.5}{u_{24}} + \frac{1}{u_{25}}$$

- iii. Based on the definition of the fuzzy set in step (ii), the number of TB patients is fuzzified by looking at the data on the number of TB sufferers based on the year and then fuzzification based on the results of the sub-interval in step (v) of the first stage so that results are obtained as shown in the table below:

Table 2. Fuzzification Results of the Number of TB Patients in 2017-2021

Year	Number of TB sufferers	Fuzzification
2017	1.069	A_{21}
2018	1.079	A_{25}
2019	1.006	A_{13}
2020	938	A_1
2021	940	A_6

- iv. The next step is to determine the Fuzzy Logical Relationship of the fuzzification results in Table 2 by connecting the fuzzification results year t to year $t + 1$ so that results are obtained, as shown in the table below:

Table 3. FLR Results on the Number of TB Patients in 2017-2021

Year	Fuzzification	Fuzzy Logical Relationship
2017	A_{21}	$A_{21} \rightarrow A_{25}$
2018	A_{25}	$A_{25} \rightarrow A_{13}$
2019	A_{13}	$A_{13} \rightarrow A_1$
2020	A_1	$A_1 \rightarrow A_6$
2021	A_6	$A_6 \rightarrow \emptyset$

After the fuzzy logical relationship is formed, the next step is to group the fuzzy logical relationship into a fuzzy logical relationship group based on the results of the smallest to largest fuzzy logical relationship so that the following results are obtained:

$$\begin{array}{ll}
 \text{Group 1 : } A_1 \rightarrow A_6 & \text{Group 4: } \rightarrow A_{21}A_{25} \\
 \text{Group 2 : } A_6 \rightarrow \emptyset & \text{Group 5 : } \rightarrow A_{25}A_{13} \\
 \text{Group 3 : } A_{13} \rightarrow A_1 &
 \end{array}$$

- v. Based on the results of step (iv) in the second stage, the results of forecasting the number of TB patients were obtained by looking at the results of the fuzzy logical relationship year t to year $t + 1$ then the forecasting value was seen from the midpoint of the year $t + 1$ fuzzification results so that the results were obtained in the table below:

Table 4. Results of Forecasting the Number of TB Patients in 2022

Year	Number of Tuberculosis Sufferers	Forecasting
2017	1.069	-
2018	1.079	1.078
2019	1.006	1.006
2020	938	938,2
2021	940	943,075
2022	-	943,075

Based on Table 4, it can be seen that the results of forecasting the number of TB sufferers in 2022 are 944 people. A comparison between the actual data and the forecasting result data can be seen in the chart below:

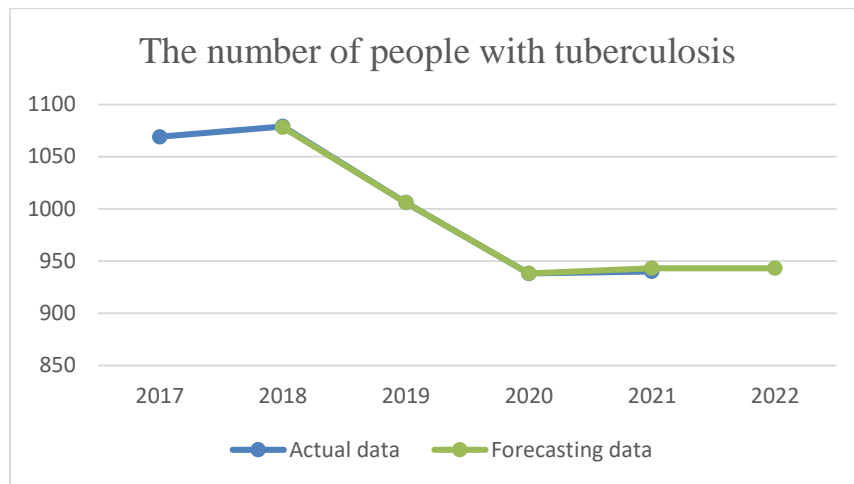


Figure 3. Graph of Actual Data Pattern Difference in Actual Data with Forecasting Results Data

In Figure 3, the graph shows that the forecasting data pattern is almost the same value as the actual data pattern.

- The last stage of the ACFLR method is to determine the forecasting error value using the formula on Equation (5) So, forecasting errors are obtained from each year as follows:

Table 5. Results of Calculating the Forecasting Error Value for 2018-2021

Year	Number of TB Patients	Forecasting	Error
2017	1.069	-	-
2018	1.079	1.078	0,093%
2019	1.006	1.006	0%
2020	938	938,2	0,021%
2021	940	943,075	0,327%
2022	-	943,075	

Based on the results of Table 5 using the calculation of the forecasting average MAPE error, this shows that the forecast results are close to the actual data. Based on the calculation criteria, $MAPE = \sum_{t=1}^5 \frac{0,00441}{5} \times 100\% = 0,0882$, the resulting MAPE error is excellent because the error obtained is <10%. Based on the analysis, the forecast result obtained in 2022 amounted to 944 tuberculosis sufferers. The results of this forecasting use to anticipate an event against a program and as a benchmark for the number of tuberculosis sufferers in Kampar Regency Riau Province. So that appropriate initial steps can be taken by the government and medical professionals to stop the rise in tuberculosis.

CONCLUSION

Based on the analysis and discussion results using the ACFLR method, 944 people were obtained from forecasting the number of TB patients in Kampar Regency in 2022. Forecasting error value of 0.0882 the accuracy rate of forecasting results was 99.9118, this means that the forecasting results using the ACFLR method have an excellent level of validity and the increase in the number of TB sufferers in Kampar Regency from 2021 to 2022 was 4 people.

REFERENCES

- [1] F. Luis and G. Moncayo, "Global Tuberculosis Report", WHO, 2021.
- [2] A. Gailan Qasem, S. S. Lam, and F. Aqlan, "A Fuzzy-Based Approach for Cholera Risk Assessment and Vaccine Allocation," *International Journal of Fuzzy Systems*, vol. 24, no. 8, pp. 3366–3383, 2022, doi: 10.1007/s40815-022-01388-8.
- [3] dr. F. R. Makarim, "Ways of Transmission of Tuberculosis that are Often Ignored," *09 Oktober*, 2020.
- [4] R. U. Fadillah, D. Wulan, S. Rengganis, F. Kedokteran, and U. Lampung, "New Case of Tuberculosis Disease Management in 30 Years Old Woman," *Journal Medula Unila*, vol. 5, pp. 69–74, 2016.
- [5] H. D. Sambudi *et al.*, "Steering Team, Kampar District Health Office Profile," 2018.
- [6] M. Muhammad, "Distribution and Forecasting of New Students in Mathematics Education at Muhammadiyah University Purwokerto with the Time Invariant Fuzzy Time Series Method," *MAJU: Scientific Journal of Mathematics Education*, vol. 3, no. 2, pp. 48–58, 2018.
- [7] P. Budianto, H. Prabaningtyas, S. E. Putra, diah K. Mirawati, F. Muhammad, and M. Hafizan, "Acute Ischemic Stroke: Basic and Clinical," *Univesrsitas Sebelas Maret*, no. January, p. 84, 2021.
- [8] Bagus Handoko, "Forecasting Short Term Electricity Loads in East Java and Bali Electrical Systems Using Fuzzy Time Series," pp. 1–6, 2009.
- [9] Rahmawati, E. P. Cynthia, and K. Susilowati, "Cheng's Fuzzy Time Series Method for Predicting the Number of Tourists in West Sumatra Province," *Journal of Education Informatic Technology and ...*, vol. 1, no. 1, pp. 11–23, 2019, [Online]. Available: <http://ejurnal.umri.ac.id/index.php/JeITS/article/view/1222>
- [10] A. Novia Rahma, Rahmawati, and W. Septia, "Prediction of Rupiah Exchange Rate Against Australian Dollarusing the Chen Fuzzy Time Series Method," *Sintechcom: Science, Technology, and Communication Journal*, vol. 1, no. 2, p. 74, 2021.
- [11] H. T. Ikshsanto, *Comparison of Accuracy Levels of Automatic Clustering, Average Based, and Markov Chain Fuzzy Time Series Methods at Rupiah Exchange Rates. Thesis. Faculty of Mathematics and Natural Sciences, Semarang State University.*, vol. 7, no. 1. 2016.
- [12] M. R. R. Putra, Rahmawati, and F. Muttakin, "Prediction of the Number of Visitors to the Regional Library of Batang Regency Using the Fuzzy Time Series Chen-Hsu Method vol. 8, no. 1, pp. 110–119, 2023.
- [13] I. L. Khoiriyah, I. Cholissodin, and B. D. Setiawan, "Forecasting the Number of Tourist Visitors Using Fuzzy Logical Relationship and Genetic Algorithm (Case Study of Tourists in Banyuwangi Regency)," *Journal of Information Technology Development and Computer Science (J-PTIIK) Universitas Brawijaya*, vol. 2, no. 3, pp. 1109–1118, 2018.
- [14] M. Abdy, R. Syam, and E. Haryanensi, "M Automatic clustering-fuzzy logical relationships method on Population Forecasting in Makassar City," *Journal of Mathematics, Computations, and Statistics*, vol. 1, no. 2, p. 193, 2018, doi: 10.35580/jmathcos.v1i2.9242.
- [15] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *Journal of Ambient Intelligence and Humanized Computing*, 2022, doi: 10.1007/s12652-021-03612-z.
- [16] Y. P. Huang *et al.*, "Wavelet K-Means Clustering and Fuzzy-Based Method for Segmenting MRI Images Depicting Parkinson's Disease," *International Journal of*

- Fuzzy Systems*, vol. 23, no. 6, pp. 1600–1612, 2021, doi: 10.1007/s40815-021-01053-6.
- [17] M. R. Hashmi, M. Riaz, and F. Smarandache, “m-Polar Neutrosophic Topology with Applications to Multi-criteria Decision-Making in Medical Diagnosis and Clustering Analysis,” *International Journal of Fuzzy Systems*, vol. 22, no. 1, pp. 273–292, 2020, doi: 10.1007/s40815-019-00763-2.
- [18] Rahmawati, Wartono, and A. Yunika Putri, “Clustering Employee Education Levels Using the Fuzzy Subtractive Clustering Method (Case Study: Riau Province Regional Employment and Training Agency),” *Journal of Mathematics Science and Statistics*, vol. 5, no. 1, pp. 79–89, 2019.
- [19] Rahmawati, S. Saputri, A. Novia Rahma, and Aprijon, “P Clustering Temperatures in Pekanbaru City Using the Fuzzy K-Means Method,” vol. 12, no. November, pp. 2579–5406, 2019.
- [20] R. Bakri, A. N. Rahma, I. Suryani, Rahmawati, and Y. Sari, “Application of Fuzzy Logic in Determining the Number of Health BPJS Participants Using Sugeno's Fuzzy Inference System,” *Lebesgue Journal: Scientific Journal of Mathematics Education, Mathematics and Statistics*, vol. 1, no. 3, pp. 182–192, 2020, doi: 10.46306/lb.v1i3.38.
- [21] R. Rahmawati, A. N. Rahma, and H. Hernita, “Application of Sugeno's Fuzzy Method in Determining the Amount of Rahn's Financing Based on the Number of Customers and Gold Prices,” *MAP (Mathematics and Applications) Journal*, vol. 2, no. 2, pp. 40–49, 2020, doi: 10.15548/map.v2i2.2263.
- [22] Rahmawati, Rahima Dina, and Wartono, “Application of the Mamdani Fuzzy Method to Determine Pawnshop Income Based on the Total Turnover of Gold Prices and Exchange Rates,” *Buana Mathematics: Scientific Journal of Mathematics and Mathematics Education*, vol. 10, no. 1, pp. 89–104, 2020, doi: 10.36456/buanamatematika.v10i1.2420.
- [23] S. L. M. Sitio, “Application of Sugeno's Fuzzy Inference System to Determine the Amount of Drug Purchases (Case Study: Garuda Sentra Medika),” *Journal of Informatics, Pamulang University*, vol. 3, no. 2, p. 104, 2018, doi: 10.32493/informatika.v3i2.1522.
- [24] R. A. Rahmawati, “Determining the Amount of Food Commodity Supply in Riau Province Based on the Fuzzy Inference System with the Fuzzy Tsukamoto Method,” *Journal of Science, Technology and Industry*, vol. 15, no. 2, p. 105, 2018, doi: 10.24014/sitekin.v15i2.5400.
- [25] Rahmawati, “Application of Fuzzy Linear Programming in Optimizing the Development of Flats (Flats) in the PondokChina Area of West Java Province,” *Journal of Mathematics and Statistics Science*, vol. 4, no. 1, pp. 79–89, 2018.
- [26] R. Bakri, S. Inayati, Y. Yuliana, and A. Hanafiah, “Prediction of the Number of Bpjs Participants Recipients of State Budget Assistance (Pbi) Using the Fuzzy Time Series Cheng Method,” *BAREKENG: Journal of Applied Mathematics and Science*, vol. 15, no. 2, pp. 373–384, 2021, doi: 10.30598/barekengvol15iss2pp373-384.
- [27] S. Sitohang and A. Marubah Siringo, “Gold Price Forecasting Analysis With Automatic Clustering And Fuzzy Logic Relationship Methodp,” <https://Ejournal.Medan.Uph.Edu/Index.Php/Isd/Article/View/209>, vol. 3, no. 2, pp. 1–14, 2018.
- [28] K. B. Kim, G. H. Kim, D. H. Song, H. J. Park, and C. W. Kim, “Automatic segmentation of liver/kidney area with double-layered fuzzy C-means and the utility of hepatorenal index for fatty liver severity classification,” *Journal of Intelligent and*

- Fuzzy Systems*, vol. 39, no. 1, pp. 925–936, 2020, doi: 10.3233/JIFS-191850.
- [29] S. M. Chen, N. Y. Wang, and J. S. Pan, “Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships,” *Expert Systems with Applications*, vol. 36, no. 8, pp. 11070–11076, 2009, doi: 10.1016/j.eswa.2009.02.085.
- [30] Rahmawati, Yuniza, A. N. Rahma, and Zukrianto, “Prediction of the Number of Tourists in Pekanbaru City in 2019-2023 Using the Fuzzy Time Series Chen Method,” *Journal of Mathematics Education*, vol. 2, no. 1, pp. 36–44, 2020.
- [31] S. Rusydi, “Prediction of the Number of Tourists in Pekanbaru City in 2019-2023 Using the Fuzzy Time Series Chen Method,” *Journal of Mathematics Education* 2012.