The comparison between recurrent neural network and grey model to predict Indonesia tuberculosis morbidity rate

D. Harlianto, *S. Mardiyati, and D. Lestari
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok 16424, Indonesia.

*sri_math@sci.ui.ac.id

Abstract. Many insurance companies offer health insurance products that cover the risk of tuberculosis (TB). The risk of disease is generally stated as the morbidity rate that is the ratio of the number of residents suffers from an illness to the total population. In 2018, the World Health Organization (WHO) reported that Indonesia was at rank two for the case numbers of TB, so Indonesia has a high risk of this disease. The recurrent neural network (RNN) and grey model are two models that can be employed to predict TB morbidity rates. In this research, the accuracy of these two models was compared. The results of this research may give the insurance company an assist to choose an appropriate mathematical model to provide a competitive and profitable premium. TB morbidity rate in a certain year had been predicted based on the past several year’s morbidity rates as model input. The size of past years data used as model input was made varied to observe how information availability influences the model accuracy measured by mean squared error (MSE) and mean absolute percentage error (MAPE). The results show that the grey model has better accuracy when the small data used as input. On the other hand, the accuracy of RNN is not affected significantly by the setting of the data input size used in this research.

1. Introduction
Tuberculosis (TB) is a disease caused by *Mycobacterium tuberculosis*. In 2018, the World Health Organization (WHO) reported that Indonesia was the second-highest of the total cases of TB in the world [1]. On the other hand, insurance companies offer health insurance products that cover the risk of TB. The morbidity rate is the indicator used to represent the risk of disease in a certain area. It is defined as a proportion between the number of persons suffering from particular health problems and the total number of persons living in a certain area [2]. Therefore, insurance companies need to forecast TB morbidity rates accurately to provide a premium that is not only competitive but also profitable.

The TB morbidity rate data available in sequential form, so time series models like ARIMA will appropriate to forecast the TB morbidity rate in the future. In 2015, the ARIMA-ARCH model was used to predict the TB morbidity rate in Xinjiang province, China [3]. The other model currently developed to deal with sequential data is the recurrent neural network (RNN). RNN has been used in many fields such as stock prediction [4], vehicle trajectory projection [5], and extraction of chemical protein interaction [6]. It has also been implemented to forecast the TB morbidity rate in Xinjiang Province, China, and the result shows that RNN outperforms ARIMA-ARCH in terms of accuracy measured by residual mean squared error (RMSE)[7].
The grey model is the other model that can be implemented to predict the sequential data. It is usually implemented to the small size data with a minimum size of 4 [8]. The model has been used widely in various fields such as natural gas consumption forecasting [9], Taiwan’s major stock indices forecasting [10], and short-term manufacturing demand forecasting [8]. Furthermore, the recording of the morbidity rate of TB in Indonesia just be done annually, so it forms small size data. Therefore, implementation of the grey model on Indonesia TB morbidity rate data may be appropriate.

As previously mentioned, RNN has good performance to predict the TB morbidity rate in China, so this result can be a motivation to employ this model for Indonesia TB morbidity rate data. Besides that, the grey model is also an appropriate model to predict TB morbidity rate in Indonesia because it has good performance to predict small size sequential data. This research compares RNN and grey model accuracy in terms of mean squared error (MSE) and mean absolute percentage error (MAPE) to predict TB morbidity rate in Indonesia. The availability of information may influence the model accuracy, so the size of the model input was made varied. Especially for RNN, we also evaluate the model on a varying number of hidden units to know the influence of that aspect on the model accuracy.

2. Method
In this section, we discussed the RNN and grey model. We also give the steps needed to implement these models to the data.

2.1. Recurrent Neural Network
Recurrent neural network (RNN) is the variant of neural network used to deal with sequential data [11]. Observations in the sequential data linked each other in an order and RNN has capability to accumulate the information contained in those observations. RNN consist of three layer. They are input layer, hidden layer, and output layer as shown in left side of Figure 1. There are reverse connections at hidden layer that make capability of RNN to accumulate information. The recurrent form of hidden unit of RNN can be unfolded as shown in right side of Figure 1 [12]. The unfolding form of RNN can show the process that will be experienced by each observation of the data more clearly.

Let $X = \{x_1, x_2, ..., x_N\}$ be a sequential data used to train the model. Form subsequent of the data as follows

\[
X_1 = \{x_1, x_2, ..., x_T\} \\
X_2 = \{x_2, x_3, ..., x_{T+1}\} \\
\vdots \\
X_{N-T-1} = \{x_{N-T-1}, x_{N-T}, ..., x_{N-2}\} \\
X_{N-T} = \{x_{N-T}, x_{N-T+1}, ..., x_{N-1}\}
\]

where $T \in \{t \in \mathbb{N} | 0 < t < N\}$ is called step size. The RNN model is trained to predict $x_{k+T}$ according to input $x_k, x_{k+1}, ..., x_{k+T-2}, x_{k+T-1}$, where $k = 1, 2, ..., N - T$. For instance, if step size be 7 then $x_8$ is predicted according to $x_1, x_2, ..., x_7$, then $x_9$ is predicted according to $x_2, x_3, ..., x_8$, and so on. In

![Figure 1. Recurrent neural network diagram](image)
Figure 2. Example of recurrent neural network

the sequence \(x_k, x_{k+1}, \ldots, x_{k+T-2}, x_{k+T-1}\), we say \(x_k\) as the first step, \(x_{k+1}\) as the second step, and so on, until the end \(x_{k+T-1}\) is said as the \(T\)-th step. The hidden unit \(j\) resulted from input \(x_{k+1}\) denoted by \(a_{j,k+1}^{(1)}\) and defined as follows

\[
a_{j,k+1}^{(1)} = w_j^{(1)} x_{k+1} + \sum_{r=1}^{M} h \left( a_{r,k}^{(1)} \right) w_{jr}^{(3)} + w_{j0}^{(1)}
\]  (1)

where \(h\) is the hidden unit activation function, \(w_j^{(1)}\) is the weight from input to hidden unit \(j\), \(w_{jr}^{(3)}\) is the weight from hidden unit \(r\) to hidden unit \(j\) at the next time step, and \(w_{j0}^{(1)}\) is the bias for hidden unit \(j\). The complete form of RNN model consists of \(M\) hidden unit to process \(x_k, x_{k+1}, \ldots, x_{k+T-2}, x_{k+T-1}\) as input is given by Equation 2.

\[
\hat{x}_{k+T} = \sum_{j=1}^{M} w_{outj}^{(2)} h \left( w_j^{(1)} x_{k+T-1} + \sum_{r=1}^{M} h \left( a_{r,k+T-2}^{(1)} \right) w_{jr}^{(3)} + w_{j0}^{(1)} \right) + w_{out0}^{(2)}
\]  (2)

where \(\hat{x}_{k+T}\) denote predicted value for \(x_{k+T}\). To be clearer with the notation which has been introduced before, Figure 2 shows RNN diagram consists of one input unit, two hidden unit, and three output unit There are three main processes to train RNN. They are forward propagation, backpropagation, and weight updating. Equation 2 is used for forward propagation processes with the error function would be minimized as follows

\[
e_k = \frac{1}{2} (\hat{x}_{k+T} - x_{k+T})^2
\]  (3)

where \(w\) is weight vector.

In the backpropagation process, derivative of error respect to each weights and biases are given by
chain rule. The equation employed in backpropagation as follows

\[
\frac{\partial e_k}{\partial w_j^{(1)}} = \sum_{i=k}^{k+T-1} \frac{\partial e_k}{\partial a_{ji}^{(1)}} x_i
\]

(4)

\[
\frac{\partial e_k}{\partial w_j^{(2)}} = (\hat{x}_{k+T} - x_{k+T}) h(a_{j,k+T-1}^{(1)})
\]

(5)

\[
\frac{\partial e_k}{\partial w_{jr}^{(3)}} = \sum_{i=k}^{k+T-2} \frac{\partial e_k}{\partial a_{ji+1}^{(1)}} h(a_{ri}^{(1)})
\]

(6)

\[
\frac{\partial e_k}{\partial w_j^{(1)}} = \sum_{i=k}^{k+T-1} \frac{\partial e_k}{\partial a_{ji}^{(1)}}
\]

(7)

\[
\frac{\partial e_k}{\partial w_{out}^{(2)}} = \hat{x}_{k+T} - x_{k+T}
\]

(8)

The derivative obtained in backpropagation process is used to compute weight and bias updating as follows

\[
w' = w - \eta \nabla e(w)
\]

(9)

where \( w' \) is the updated weight vector. The Equation 9 move the weight value to the direction with smaller gradient magnitude, so the training process for RNN that have been explained here is known as gradient descent algorithm. The steps of training process of RNN can be summarized as follows:

1. Do forward propagation using Equation 2 to obtain \( \hat{x}_{k+T} \) with initialization of weight and bias selected randomly.
2. According to \( \hat{x}_{k+T} \) obtained from poin 1 and \( x_{k+T} \) from the data, compute error function using Equation 3.
3. Do backpropagation using Equation 4 - 8.
4. The weight and biases are updated using Equation 9.
5. Repeats poin 1 - 4 for each subsequence \( X_1, X_2, ..., X_{N-T} \). When every subsequence has been involved in training process then we said that training has been done for 1 epoch.

2.2. Grey Model

Grey model is a sequential data predictive model usually implemented on small size data, with minimum size of 4 [13]. Let \( X = \{x_1, x_2, ..., x_N\} \) be sequential data satisfy \( x_i > 0 \) for \( i = 1, 2, ..., N \). The objective of implementation of the model is to predict \((N + 1)\)-th observation value according to given \( X = \{x_1, x_2, ..., x_N\} \). Define

\[
x_j^{(0)} = x_j \text{ for } j = 1, 2, ..., N
\]

(10)

\[
X^{(1)} = \{x_j^{(1)}\} \text{ dengan } x_j^{(1)} = \sum_{i=1}^{j} x_j^{(0)} \text{ for } j = 1, 2, ..., N
\]

(11)

\[
z_k^{(1)} = \alpha_k x_k^{(1)} + (1 - \alpha_k)x_{k-1}^{(1)} \text{ for } 0 < \alpha_k < 1 \text{ dan } k = 2, 3, ..., N
\]

(12)

then grey model can be defined as follows

\[
x_k^{(0)} + a_k^{(1)} = b
\]

(13)
where \( a \) and \( b \) are the model parameters [14]. These parameters are estimated using least square method as shown in Equation 14.

\[
\hat{w} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{x} \quad (14)
\]

where \( \mathbf{x}, \mathbf{B}, \) and \( \mathbf{w} \) defined as follows

\[
\mathbf{x} = \begin{bmatrix} x_{2}^{(0)} \\ x_{3}^{(0)} \\ \vdots \\ x_{N}^{(0)} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -z_{2}^{(1)} & 1 \\ -z_{3}^{(1)} & 1 \\ \vdots & \vdots \\ -z_{N}^{(1)} & 1 \end{bmatrix}, \quad \text{dan} \ \mathbf{w} = \begin{bmatrix} a \\ b \end{bmatrix} \quad (15)
\]

The parameter \( \alpha_k \) in the Equation 13 represent the importance level of \( x_k^{(0)} \) and can be estimated using some steps below [8] :

1. Compute central location of the data

\[
\text{CL} = \frac{x_{\text{min}} + x_{\text{max}}}{2} \quad (16)
\]

where \( x_{\text{min}} \) and \( x_{\text{max}} \) are observations in the data with minimal and maximal values.

2. Determine skewness of the data, that are skew\(_U\) and skew\(_L\) as follows

\[
\text{skew}_U = \frac{|X^+|}{|X^+| + |X^-|} \quad (17)
\]

\[
\text{skew}_L = \frac{|X^-|}{|X^+| + |X^-|} \quad (18)
\]

where \( |X^+| \) and \( |X^-| \) are the the number of observation values in the data that smaller and higher, respectively, than \( \text{CL} \).

3. Estimates upper bound (UB) and lower bound (LB) of the data using Equation 19.

\[
UB = \begin{cases} 
U = \text{CL} + \text{skew}_U \sqrt{-2 \hat{s}_x^2/|X^+|ln(10^{-20})} & \text{for } U \geq x_{\text{max}} \\
U = x_{\text{max}} & \text{for } U < x_{\text{max}} 
\end{cases} \quad (19)
\]

\[
LB = \begin{cases} 
L = \text{CL} - \text{skew}_L \sqrt{-2 \hat{s}_x^2/|X^-|ln(10^{-20})} & \text{for } L \leq x_{\text{min}} \\
L = x_{\text{min}} & \text{for } L > x_{\text{min}} 
\end{cases} \quad (20)
\]

where \( \hat{s}_x^2 \) is the variance of the data.

4. Compute membership function (MF) for each observation

\[
MF_i = \begin{cases} 
\frac{x_i^{(0)} - LB}{UB - LB} & \text{for } x_i^{(0)} \leq \text{CL} \\
\frac{CL - x_i^{(0)}}{UB - CL} & \text{for } x_i^{(0)} > \text{CL} 
\end{cases} \quad (21)
\]

5. Finally, parameter \( \alpha_k \) can be determined using Equation 22.

\[
\alpha_k = \frac{\sum_{i=1}^{k} i \cdot MF_i}{\sum_{i=1}^{k} i} \quad \text{for } k \geq 2 \quad (22)
\]
Grey model defined by Equation 13 is a delicate approximation of first order differential equation

$$\frac{dy}{dt} + ay = b \quad (23)$$

that has solution

$$y = \left( y_0 - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \quad (24)$$

where \( y_0 \) is the value of \( y \) when \( t = 0 \) [15]. Therefore, the predicted value for \( x_k^{(1)} \), where \( k = 1, 2, \ldots, N \), given by Equation 25.

$$\hat{x}_{k+1}^{(1)} = \left( x_1^{(0)} - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (25)$$

Then, using Equation 11 and 25, we obtain

$$\hat{x}_{k+1}^{(0)} = \hat{x}_{k+1}^{(1)} - \hat{x}_k^{(1)} \quad (26)$$

Substitute \( k = N \) into Equation 26 give the predicted value of \( x_{N+1} \) which is we looking for

$$\hat{x}_{N+1}^{(0)} = \hat{x}_{N+1}^{(1)} - \hat{x}_N^{(1)} \quad (27)$$

### 3. Result and Discussion

The morbidity rate data were obtained from Departemen Kesehatan Republik Indonesia [16]. The data was displayed in Table 1, where JPKK is the number of cases of TB in Indonesia, JP is the Indonesia total population, and AM is the morbidity rate. Morbidity rate (AM) was computed based on the definition given by Badan Pusat Statistik (BPS)[2] as follows

$$AM = \frac{JPKK}{JP} \times 100 \quad (28)$$

The following first two subsection explain the results of implementation of the models on the data, then those results was compared in the third subsection.

#### 3.1. RNN Implementation

The data was split into two parts that are the first nine observations (2004-2012) be the training data and the remainder be validation data. The training data was employed to establish the model. In other words, the parameter of RNN is estimated based on that data. Then, the accuracy of the model is measured in term of mean squared error (MSE) and mean absolute percentage error (MAPE) on the validation data.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \quad (29)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (30)$$

where \( x_i, \hat{x}_i, \) and \( n \) denote actual value, predicted value, and the size of data, respectively. Data standardization was performed using min-max scaler defined by Equation 31 before the training process is started

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (31)$$
Table 1. TB morbidity rate from 2004 - 2018

| Years | JPKK   | JP      | AM(%) |
|-------|--------|---------|-------|
| 2004  | 214.658| 217.072.346 | 0.0989 |
| 2005  | 259.969| 218.868.791 | 0.1188 |
| 2006  | 277.589| 222.192.000 | 0.1249 |
| 2007  | 268.042| 225.642.124 | 0.1188 |
| 2008  | 298.329| 228.523.342 | 0.1305 |
| 2009  | 294.731| 231.369.592 | 0.1274 |
| 2010  | 302.861| 237.641.326 | 0.1274 |
| 2011  | 321.300| 241.182.182 | 0.1332 |
| 2012  | 324.086| 244.775.797 | 0.1324 |
| 2013  | 327.094| 248.422.956 | 0.1317 |
| 2014  | 285.254| 252.124.458 | 0.1131 |
| 2015  | 330.910| 255.461.686 | 0.1295 |
| 2016  | 351.893| 258.704.986 | 0.1360 |
| 2017  | 425.089| 261.890.872 | 0.1623 |
| 2018  | 511.873| 265.015.313 | 0.1931 |

Table 2. Standardized data

| i | Years | $x_i$ | $x'_i$ |
|---|-------|-------|--------|
| 1 | 2004  | 0.0989| 0      |
| 2 | 2005  | 0.1188| 0.57935976 |
| 3 | 2006  | 0.1249| 0.75861466 |
| 4 | 2007  | 0.1188| 0.57973843 |
| 5 | 2008  | 0.1305| 0.92214034 |
| 6 | 2009  | 0.1274| 0.83006612 |
| 7 | 2010  | 0.1274| 0.83181381 |
| 8 | 2011  | 0.1332| 1      |
| 9 | 2012  | 0.1324| 0.97617314 |
| 10| 2013  | 0.1317| 0.95482217 |

where $x_i$ and $x'_i$ denote the $i$-th observation before and after standardization, $x_{max}$ and $x_{min}$ denotes the observation maximum and minimum values of the data, respectively. Standardization is performed to achieve better performance of the model [17]. The standardized training data was displayed in Table 2.

Python programming language with Keras module was used to implement the model on the data. We selected $tanh$ as a hidden unit activation function [11]. The number of hidden units and step size were made variedly, which is from $M=1$ to $M=15$ for hidden units and from $T=4$ to $T=8$ for the step size. Combining those settings gave $15 \times 5 = 75$ condition where the training process would be run with 100 epoch in each. To get more general results, the 50 times repetition was applied to all 75 conditions. The results were displayed in the appendix.

The smallest MSE value resulted from the RNN model was 0.00075014 achieved when RNN consists of 10 hidden units ($M=10$) and step size of 6 ($T=6$). Furthermore, the smallest MAPE value, that was 0.11682883, was achieved when $M=2$ and $T=6$. Figure 3 shows the MSE and MAPE values for all combinations the number of hidden units and step size. Based on Figure 3, the MSE values as well
as MAPE values has a darker color when step size values around 6 that indicates the model has lower values of MSE and MAPE there. But those two graphs have a different pattern for the use of the hidden unit. When the RNN model consists of more hidden units then MSE values tend to smaller but MAPE values tend to greater. This different pattern arises because when the RNN model has more hidden units, the accuracy of the model be better generally but still poor to predict observation with small values such that the ratio in Equation 30 becomes larger.

3.2. Grey Model Implementation

Grey model was implemented with step size varies from 4 to 8 in order to make the same condition as the preceding subsection. There is no number generated randomly involved in establishing grey model, so we do not repeat the implementation of the model for each condition. The model accuracy was evaluated on the last 6 observations and the result was shown in the Table 3. The smallest MSE and MAPE achieved when the model implemented with step size of 4 ($T = 4$). Furthermore, MSE and MAPE of the model increase for larger step size values. The results conform with some study that revealed grey model has good performance on small size data [8][13]. The graph of grey model along with the actual data were shown in Figure 4. It can be seen that the model follows the data trend well although there is an extremely increasing trend in the last two observations.

3.3. The Accuracy Comparison Between Two Models

Based on the results explained in the two preceding sections, the grey model has better accuracy than RNN. The smallest MSE resulted from the grey model was 63% smaller than the smallest MSE resulted from RNN according to the comparison between the table contained in the appendix file and Table 3. The MSE between the two models differs significantly but the MAPE differs slightly. The smallest MAPE resulted from the grey model only 7% smaller than the smallest MAPE of RNN. Figure 5 displays the graph at the conditions where each models attained its best accuracy, that is $T = 6$, $M = 10$ for RNN.

**Table 3. MSE and MAPE for grey model**

| No. | Step Size | MSE       | MAPE       |
|-----|-----------|-----------|------------|
| 1   | 4         | 0.00027492| 0.10774796 |
| 2   | 5         | 0.00034706| 0.11904685 |
| 3   | 6         | 0.00046777| 0.12434019 |
| 4   | 7         | 0.00055881| 0.12808526 |
| 5   | 8         | 0.00063422| 0.13166893 |

**Figure 3.** MSE and MAPE graphs for RNN model
According to Figure 5, RNN gives a better approximation to the data relatively than the grey model, but according to calculated MSE, the grey model has better accuracy considerably. It can happen because RNN can’t approximate the data when a considerable increasing trend of the data arise, which is at the 2017 and 2018 morbidity rate observations. Therefore, this greater MSE value of RNN is contributed largely from the difference between prediction values and actual values at those two last observations.

4. Conclusions

We can conclude several things about RNN and grey model accuracy under the conditions have been defined in this research as follows:

1. Grey model has better accuracy than RNN consists of hidden units range between 1 to 15 units. As has been stated in subsection 4.3 that the smallest MSE and MAPE resulted from the grey model are smaller 63% and 7%, respectively, than resulted from RNN. Moreover, the grey model doesn’t need large size data, so the model is appropriate to be implemented on the Indonesia TB morbidity rate data. That data is recorded annually, so it only contains a small number of observations.

2. The MSE and MAPE of RNN are relatively stagnant for various number hidden units and step size that is employed, which is about 0.0007 for MSE and 0.11 for MAPE. This indicates the potential of the model is not maximized properly. The small number of observations imply limited patterns that can be learned by the model when the training process was performed. These limited patterns
cause the model can’t reach better performance for the various sets of the number of the hidden unit. Moreover, the validation data has a considerable change of trend at the two last observations. The model hard to follow the trend at this part, so it contributes to increasing the MSE and MAPE values significantly.

Acknowledgements
We would like to thank DRPM UI that has fully funded this research. We would also like to express our gratitude to all actuarial research group members in Departemen Matematika, FMIPA UI, that has given much useful suggestion for this paper.

References
[1] World Health Organization. (2018). Global tuberculosis report 2018, Geneva : Author.
[2] Badan Pusat Statistik. (2002). Angka kesakitan / morbiditas /persentase penduduk yang mempunyai keluhan kesehatan. Juli 16, 2019. https://sirusa.bps.go.id/sirusa/index.php/indikator/16.
[3] Zheng Y L, et al 2015 Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. PLoS ONE, 10.
[4] Berradi Z and Lazaar M 2019 Integration of principal component analysis and recurrent neural network to forecast the stock price of casablanca stock exchange. Procedia Comput. Sci. 148 55–61.
[5] Choi S, Kim J, and Yeo H 2019 Attention-based recurrent neural network for urban vehicle trajectory prediction. Procedia Comput. Sci. 151 327–334.
[6] Lu H, He X, Li L, and Liu Y 2019 Extracting chemical-protein interactions from biomedical literature via granular attention based recurrent neural networks. Comput. Methods Programs Biomed. 176 61–68.
[7] Singh J, Dhyani T, Jain A, and Agrawal P 2018 Using deep learning for forecasting tuberculosis morbidity rate. Proceedings of The International Conference on Intelligent Computing and Sustainable System, 235-239.
[8] Chang C J, Yu L, and Jin P 2016 A mega-trend-diffusion grey forecasting model for short-term manufacturing demand. Journal of the Operational Research Society 67 1439-1445.
[9] Boran F E 2015 Forecasting natural gas consumption in Turkey using grey prediction. Economics Planning and Policy 10 208-213.
[10] Chen C I, Hsin P H, and Wu C S 2010 Forecasting Taiwan’s major stock indices by the Nash nonlinear grey Bernoulli model. Expert Systems with Applications 37 7557-7562.
[11] Goodfellow I, Bengio Y, and Courville A 2016 Deep learning (MIT Press).
[12] LeCun Y, Begio Y, and Hinton G 2015 Deep Learning Nature 521 436 - 444.
[13] Yang X, Zou J, Kong D, and Jiang G 2018 The analysis of GM(1,1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China Medicine 97.
[14] Liu S and Lin Y 2006 Grey information : Theory and practical application (London : Springer-Verlag).
[15] Guo R 2006 Modeling imperfectly repaired system data via grey differential equations with unequal-gapped times Reliability Engineering and System Safety 92 378-391.
[16] Kementrian Kesehatan RI 2004 – 2018 Profil kesehatan Indonesia tahun 2004 - 2018 (Jakarta : Author).
[17] Shanker M Hu M Y and Hung M S 1996 Effect of Data Standardization on Neural Network Training Int J Mgmt Sci 24 385-397.
Table 4. MSE and MAPE for RNN model

| No. | Step Size | Hidden Units | MSE       | MAPE       |
|-----|-----------|--------------|-----------|------------|
| 1   | 4         | 1            | 0.00082314| 0.12049007 |
| 2   | 4         | 2            | 0.00083073| 0.11975093 |
| 3   | 4         | 3            | 0.00082378| 0.12019131 |
| 4   | 4         | 4            | 0.00082632| 0.12061668 |
| 5   | 4         | 5            | 0.00082106| 0.12002941 |
| 6   | 4         | 6            | 0.00082405| 0.12130354 |
| 7   | 4         | 7            | 0.00082313| 0.12148399 |
| 8   | 4         | 8            | 0.00081960| 0.12134385 |
| 9   | 4         | 9            | 0.00082241| 0.12222592 |
| 10  | 4         | 10           | 0.00082171| 0.12199121 |
| 11  | 4         | 11           | 0.00081829| 0.12184740 |
| 12  | 4         | 12           | 0.00081766| 0.12229332 |
| 13  | 4         | 13           | 0.00082206| 0.12271905 |
| 14  | 4         | 14           | 0.00081799| 0.12254723 |
| 15  | 4         | 15           | 0.00081644| 0.12272997 |
| 16  | 5         | 1            | 0.00080441| 0.11725131 |
| 17  | 5         | 2            | 0.00078019| 0.11694892 |
| 18  | 5         | 3            | 0.00077932| 0.11694710 |
| 19  | 5         | 4            | 0.00077234| 0.11685255 |
| 20  | 5         | 5            | 0.00077675| 0.11739190 |
| 21  | 5         | 6            | 0.00077294| 0.11720099 |
| 22  | 5         | 7            | 0.00076864| 0.11715454 |
| 23  | 5         | 8            | 0.00077751| 0.11768087 |
| 24  | 5         | 9            | 0.00077557| 0.11811486 |
| 25  | 5         | 10           | 0.00076787| 0.11784565 |
| 26  | 5         | 11           | 0.00077359| 0.11865324 |
| 27  | 5         | 12           | 0.00076774| 0.11868806 |
| 28  | 5         | 13           | 0.00076475| 0.11813684 |
| 29  | 5         | 14           | 0.00077497| 0.11913163 |
| 30  | 5         | 15           | 0.00076766| 0.11874250 |
| 31  | 6         | 1            | 0.00078462| 0.11711767 |
| 32  | 6         | 2            | 0.00078069| 0.11682883 |
| 33  | 6         | 3            | 0.00077722| 0.11751546 |
| 34  | 6         | 4            | 0.00076946| 0.11706437 |
| 35  | 6         | 5            | 0.00077573| 0.11783943 |
| 36  | 6         | 6            | 0.00076887| 0.11742658 |
| 37  | 6         | 7            | 0.00076350| 0.11726256 |
| 38  | 6         | 8            | 0.00076904| 0.11750124 |
| 39  | 6         | 9            | 0.00076487| 0.11780195 |
| 40  | 6         | 10           | 0.00075014| 0.11715107 |
| 41  | 6         | 11           | 0.00075423| 0.11760011 |
| 42  | 6         | 12           | 0.00076071| 0.11835229 |
| 43  | 6         | 13           | 0.00075318| 0.11770451 |
| 44  | 6         | 14           | 0.00075322| 0.11786463 |
| 45  | 6         | 15           | 0.00075331| 0.11804702 |
| 46  | 7         | 1            | 0.00076169| 0.11884111 |
### Table 4 continued

| No. | Step Size | Hidden Units | MSE      | MAPE     |
|-----|-----------|--------------|----------|----------|
| 47  | 7         | 2            | 0.00076534 | 0.12130278 |
| 48  | 7         | 3            | 0.00076462 | 0.11913926 |
| 49  | 7         | 4            | 0.00076805 | 0.11895296 |
| 50  | 7         | 5            | 0.00077427 | 0.11961136 |
| 51  | 7         | 6            | 0.00076924 | 0.11854336 |
| 52  | 7         | 7            | 0.00077267 | 0.11886378 |
| 53  | 7         | 8            | 0.00077609 | 0.11931465 |
| 54  | 7         | 9            | 0.00076898 | 0.11889780 |
| 55  | 7         | 10           | 0.00077197 | 0.11921023 |
| 56  | 7         | 11           | 0.00077533 | 0.11970290 |
| 57  | 7         | 12           | 0.00077573 | 0.11946654 |
| 58  | 7         | 13           | 0.00077810 | 0.11959656 |
| 59  | 7         | 14           | 0.00078267 | 0.12020600 |
| 60  | 7         | 15           | 0.00077495 | 0.11959854 |
| 61  | 8         | 1            | 0.00081109 | 0.12065661 |
| 62  | 8         | 2            | 0.00080565 | 0.12040508 |
| 63  | 8         | 3            | 0.00079429 | 0.12018278 |
| 64  | 8         | 4            | 0.00083069 | 0.12301454 |
| 65  | 8         | 5            | 0.00080306 | 0.12021843 |
| 66  | 8         | 6            | 0.00082366 | 0.12260300 |
| 67  | 8         | 7            | 0.00079745 | 0.11975857 |
| 68  | 8         | 8            | 0.00079486 | 0.11927986 |
| 69  | 8         | 9            | 0.00079387 | 0.11919067 |
| 70  | 8         | 10           | 0.00079367 | 0.11941345 |
| 71  | 8         | 11           | 0.00078489 | 0.11813821 |
| 72  | 8         | 12           | 0.00079883 | 0.11971756 |
| 73  | 8         | 13           | 0.00078884 | 0.11900814 |
| 74  | 8         | 14           | 0.00078837 | 0.11909340 |
| 75  | 8         | 15           | 0.00078382 | 0.11868366 |