Forecasting the mortality rates of Indonesian population by using neural network

Lutfiani Safitri¹, Sri Mardiyati¹, and Hendrisman Rahim²

¹ Department of Mathematics, FMIPA Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia
² PT. Asuransi Jiwasraya, Indonesia

Abstract. A model that can represent a problem is required in conducting a forecasting. One of the models that has been acknowledged by the actuary community in forecasting mortality rate is the Lee-Carter model. Lee Carter model supported by Neural Network will be used to calculate mortality forecasting in Indonesia. The type of Neural Network used is feedforward neural network aligned with backpropagation algorithm in python programming language. And the final result of this study is mortality rate in forecasting Indonesia for the next few years

1. Introduction

Mortality is the number of death sizes of a population. Mortality is one component that influences the change of population structure [3]. In general, the mortality of an area is expressed in a mortality rate, which describes how large the rate of death in an area per year and at a certain age. Mortality rate is beneficial in several areas, such as economics, health and life insurance [13]. For example in life insurance, mortality rates are typically used to determine the amount of a premium a person pays to a life insurance company [7].

Due to the importance of this mortality rate, it is necessary to forecast mortality rates for the foreseeable future. In order to predict future mortality rates, a model can be used to describe the mortality rate of a region. There are two approaches in a modeling, namely a deterministic approach and a stochastic approach. In the deterministic model, it will produce a fixed output value, and the output is fully influenced by the value of the parameter and the initial value. While the output value of the stochastic model is probability, and the output value is not fully dependent on the value of the parameter and the initial value [6]. One model of mortality rate using a deterministic model is the model proposed by Helligman and Pollard in 1980, they used a deterministic approach based on their age [5]. one model of mortality rate with a stochastic approach is the Lee-Carter model. In 1992, Lee and Carter introduced a stochastic model to predict mortality rates in the United States of America. The Lee-carter model has been successfully applied for mortality rates forecasting in several countries and at different time periods, including the United States [8], Chile (1994) [9], Singapore (2004) [10], Hungary (2007) [1], and Malaysia (2016) [11].

From Lee Carter model, The parameters will be estimate using least square method and Singular Value Decomposition (SVD) [11]. After the parameter value obtained then forecasting from stochastic process will be conducted. This paper will present forecasting mortality rate by using Neural Network. Neural network (NN) is a tool that used to complete a calculation process, where the NN work system is following the learning process in the human brain that is implemented in computing systems. NN
itself is divided into 2 types, namely feedforward NN and recurrent NN, This Research will use Feedforward Neural Network type [2].

2. Methodology

2.1 Feedforward Neural Network
Before going to deeper understanding of Feedforward Neural Network, we will explain Lee Carter model. Lee and Carter introduced a stochastic model that was used to predict a region's mortality rate, in 1992. Its first study predicted mortality rates in the United States. The study was quite successful because it adequately describes the mortality rate in the American population. Then several years after that, many countries use the model to forecast mortality rates in their countries and get pretty good results. Here is a model from Lee Carter

\[ \ln(m_{x,t}) = a_x + b_x \cdot k_t + \varepsilon_{x,t} \]  

(1)

with constraints:

\[ \sum_{t=1}^{T} k_t = 0 \] \[ \sum_{x=1}^{N} b_x^2 = 1 \]

where \( m_{x,t} \) is the central death rate at age \( x \) in year \( t \), with \( x = 1, 2, ..., N \), and \( t = 1, 2, ..., T \). Then \( a_x \) and \( b_x \) are age-dependent parameters, and \( k_t \) is a stochastic process that depends on the observed year and \( \varepsilon_{x,t} \) is independent error, with mean 0 and its variance \( \sigma^2 \) [11].

The early step is to estimate parameter value of \( a_x \) by using Least Square method, so the equation is obtained as follows:

\[ \hat{a}_x = \frac{\sum_{t=1}^{T} \ln(m_{x,t})}{T} \]  

(2)

then parameter value of \( b_x \) and \( k_t \) will be estimated by using Singular Value Decomposition, so the equation is obtained as follows:

\[ \hat{b}_x = \frac{1}{\sum_x u_{x,1}^2} (u_{1,1}, u_{2,1}, ..., u_{x,1})^T \]  

(3)

\[ \hat{k}_t = \sum_x u_{x,1} \times \sigma_1 \times (v_{1,1}, v_{2,1}, ..., v_{t,1}) \]  

(4)

then the value of \( k_t \) for \( t = T + 1, T + 2 \) will be forecasted and so on by using the neural network, since the main discussion of this paper is forecasting using Neural Network [11].

According to the encyclopedia of Machine learning and data mining, Neural network is a learning algorithm whose work system follows the brain's working system in humans. The learning is obtained by adjusting the value of weights between nodes, analogous to synapses and neurons [12]. In building a neural network framework that will be used, it requires 3 things, namely:

1. Pattern of relationship between neurons, or its architecture.
   
   Based on the architecture, there are 2 types of single-layer net and multi-layer net in the feedforward neural network. In single-layer net there is only one layer with the connected weights. This network only accepts input and is directly processed into output value through its weights, without going through the hidden layer. While in a multi-layer net network there is one layer or more located between the input layer and the output layer, called the hidden layer.

2. The method used to determine the weight of the correction (learning algorithm)
   
   there are many of Algorithms that can be used in the learning process such as, Backpropagation, Delta Learning Rule, Forward Propagation, Hebb Learning Rule, and Simulated Annealing.

3. The activation function used
   
   Some commonly used activation functions are the Identity function, Binary step function, Binary sigmoid, and Sigmoid Bipolar [4].
Feedforward neural network with multi-layer net will be used in this paper, which there is one layer in hidden layer. Then the training algorithm used is backpropagation algorithm, and the activation function used is sigmoid function. Here is an overview of the neural network structure to be applied in this paper.

![Multilayer Structure of Feedforward Neural Network](image)

**Figure 1.** Multilayer Structure of Feedforward Neural Network

The steps in the backpropagation algorithm can be explained as follow:

1. The number of input nodes will be determined by looking at the relationship proximity between variables, by calculating the correlation value between variables, or by describing the correlogram by using the device R
2. Data will be divided into two parts, namely training data and testing data. This training data will be used for learning while the testing data will be used to validate the model obtained, whether the model obtained is sufficient to describe the existing data or not. After that, data normalization will be conducted. Then, early initialization will be conducted on all weights and bias values, with random number range (0,1) and determine the activation function to be used.
3. Then, enter the input value and the input will be forwarded to the next layer that is the hidden layer, by multiplying each input by its respective weight and adding it, like the following formula

$$Z_{netj} = v_{j0} + \sum_{i=1}^{n} x_i v_{ji}$$

(5)

once the $Z_{netj}$ value obtained on each node in the hidden layer, calculate the output signal from the hidden layer as follows $Z_j = f(Z_{netj})$ where f is the predefined activation function
4. Perform step 3 as much as the number of hidden layers used.
5. Then the output value from the hidden layer will be obtained. The output of the hidden layer will be passed to the output layer, through this following formula

$$Y_{netk} = w_{k0} + \sum_{j=1}^{p} z_j w_{kj}$$

(6)

the value of the activation function of $Y_k = f(Y_{netk})$, will be determined and the output value will be obtained.
6. The output obtained and the expected output then will be checked. If not appropriate or the error is significant, weight correction on each layer will be done. The weight correction can be conducted in two ways. It can starts from the weight between hidden layer with output layer, or
done from behind. Here is a formula for updating the weights and biases between the hidden layer and the output layer
\[
w_{kj}(\text{new}) = w_{kj}(\text{existing}) + \Delta w_{kj}
\]
\[
\Delta w_{kj} = \alpha \delta_k z_j, j \neq 0
\]
is correction formula for weight value and \( \Delta w_{k0} = \alpha \delta_k 
\]
is correction formula for bias value, where \( \delta_k \) was obtained from formula presented below:
\[
\delta_k = (t_k - Y_k)f'(Y_\text{net}_k)
\]
7. Then value of weight and bias between input layer and hidden layer will be renewed with formula as follows:
\[
v_{ji}(\text{new}) = v_{ji}(\text{existing}) + \Delta v_{ji}
\]
\[
\Delta v_{ji} = \alpha \delta_j x_i, i \neq 0
\]
is correction formula for weight value, and \( \Delta v_{j0} = \alpha \delta_j 
\]
is correction formula for bias value, where \( \delta_j \) was obtained from formula presented below:
\[
\delta_j = \delta_\text{net}_j f'(Z_\text{net}_j)
\]
\[
\delta_\text{net}_j = \sum_{k=1}^m \delta_k w_{kj}
\]
is this process is repeated until the smallest error value is obtained. Then we will get the model with the best estimation weight, which will then be done for future forecasting [4].

### 3. Main Result

The data used in this study is the data of Indonesia’s state mortality rate from 1950-2015, with the data range of five years. The death rate data is taken from https://www.un.org. Available data from age 0 to 90 years, with age range is 0, 1-5, 6-10, and so on until age 86-90. Here is an overview of the data used in this study

#### Table 1. Indonesia Mortality rate

| Age Group | 1950-1955 | 1955-1960 | 1960-1965 | 1965-1970 | 1970-1975 | 1975-1980 | 1980-1985 | 1985-1990 | 1990-1995 | 1995-2000 | 2000-2005 | 2005-2010 | 2010-2015 |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 0       | 0.220     | 0.187     | 0.159     | 0.135     | 0.112     | 0.097     | 0.084     | 0.072     | 0.058     | 0.047     | 0.037     | 0.030     | 0.025     |
| 2 1-5     | 0.036     | 0.028     | 0.021     | 0.016     | 0.012     | 0.010     | 0.008     | 0.006     | 0.004     | 0.003     | 0.002     | 0.001     | 0.001     |
| 3 6-10    | 0.005     | 0.004     | 0.004     | 0.003     | 0.002     | 0.002     | 0.001     | 0.001     | 0.001     | 0.000     | 0.000     | 0.000     | 0.000     |
| ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       | ...       |
| 8 8-12    | 0.203     | 0.197     | 0.191     | 0.185     | 0.179     | 0.175     | 0.171     | 0.168     | 0.163     | 0.157     | 0.151     | 0.146     | 0.142     |
| 9 9-12    | 0.305     | 0.303     | 0.298     | 0.293     | 0.287     | 0.284     | 0.281     | 0.278     | 0.273     | 0.267     | 0.261     | 0.254     | 0.250     |

From the data of Indonesia mortality rate, the estimation of parameters value of \( a_x, b_x \) and \( k_x \) can be defined by using Least Square (2) and Singular Value Decomposition method (3), (4), and the estimation result which will be used in the calculation of mortality forecasting in Indonesia are presented in Table 2 and Table 3.

Later, the value of \( k_x \) will be forecasted by using the neural network that is implemented in python programming language, with the steps described as follows:

1. Determine the number of input nodes, in this paper 2 nodes are required.
2. Data shall be divided into training data and testing data. Since the data obtained is in small amount, then all data will be used for training data. The testing data will use random data picked from training data. Then, data normalization and initialization of weight and value of bias will be conducted.
3. Calculate the output from hidden layer according to formula (5)
4. Calculate the output value of output layer according to formula (6)
5. Update the weight of each edge and bias by using formulas (7) and (8).
6. After weight and bias value being updated, do step 3 repeatedly until minimum error value is achieved.
7. Then will be obtained neural network model for forecasting.

Table 2. The estimation results of parameter $a_x, b_x$

| $X$ | Age Group | $a_x$ | $b_x$ | $t$ | Year Group | $k_t$ |
|-----|-----------|-------|-------|-----|------------|-------|
| 1   | 0         | -2.535305571 | 0.113121044 | 1   | 1950-1955   | -0.036811276 |
| 2   | 1-5       | -4.903325525  | 0.114527372 | 2   | 1955-1960   | 0.016137596 |
| 3   | 6-10      | -6.257768336  | -0.288364691 | 3   | 1960-1965   | -0.00393337 |
| 4   | 11-15     | -6.642158292  | -0.361029496 | 4   | 1965-1970   | 0.003227391 |
| 5   | 16-20     | -6.227792463  | -0.207156019 | 5   | 1970-1975   | 0.011400005 |
| 6   | 21-25     | -5.985092803  | -0.143706519 | 6   | 1975-1980   | -0.069694394 |
| 7   | 26-30     | -5.931403207  | -0.064089919 | 7   | 1980-1985   | 0.010435503 |
| 8   | 31-35     | -5.789418083  | -0.010321486 | 8   | 1985-1990   | 0.013913731 |
| 9   | 36-40     | -5.557747938  | 0.03611584  | 9   | 1990-1995   | 0.005765721 |
| 10  | 41-45     | -5.280767417  | 0.100378846 | 10  | 1995-2000   | -0.029592153 |
| 11  | 46-50     | -4.961403283  | 0.131166975 | 11  | 2000-2005   | 0.040735604 |
| 12  | 51-55     | -4.580377048  | 0.212028902 | 12  | 2005-2010   | 0.040735604 |
| 13  | 56-60     | -4.183308393  | 0.294081098 | 13  | 2010-2015   | 0.063964828 |
| 14  | 61-65     | -3.621794699  | 0.392838978 | 14  | 2015-2020   | 0.000403285 |
| 15  | 66-70     | -3.167211589  | 0.386845962 | 15  | 2020-2025   | 0.000403285 |
| 16  | 71-75     | -2.676287374  | 0.345871638 | 16  | 2025-2030   | 0.000403285 |
| 17  | 76-80     | -2.204364696  | 0.266083438 | 17  | 2030-2035   | 0.000403285 |
| 18  | 81-85     | -1.766726711  | 0.175920359 | 18  | 2035-2040   | 0.000403285 |
| 19  | 86-90     | -1.273669262  | 0.086247194 | 19  | 2040-2045   | 0.000403285 |

There are the results of testing the model that has been obtained before

Table 3. The estimation results of parameter $k_t$

| $t$ | Year Group | $k_t$ |
|-----|------------|-------|
| 1   | 1960-1965  | -0.00393337 |
| 2   | 1965-1970  | 0.003227391 |
| 3   | 1970-1975  | 0.011400005 |
| 4   | 1975-1980  | 0.004122664032152 |
| 5   | 1980-1985  | 0.0004521778477231 |
| 6   | 1985-1990  | 0.0005165814968910499 |
| 7   | 1990-1995  | 0.00048689041110894194 |
| 8   | 1995-2000  | 0.00046098903851871 |
| 9   | 2000-2005  | 0.00043739012275880324 |
| 10  | 2005-2010  | 0.00041861177683834 |
| 11  | 2010-2015  | 0.0004032820078852832 |

The forecast value obtained of $k_t$ for 5 years ahead is:

Table 5. Forecasting value of $k_t$

| $t$ | Year Group | $k_t$ |
|-----|------------|-------|
| 14  | 2015-2020  | -0.03203691 |
| 15  | 2020-2025  | 0.06385799 |
| 16  | 2025-2030  | 0.06269202 |
| 17  | 2030-2035  | -0.06117693 |
| 18  | 2035-2040  | 0.06393609 |
From Table 5 will be calculate the mortality rate with the parameters obtained in table 2 using equation (1). The result of the mortality rate forecasting will be presented in table 6.

| Year Group | 2015-2020 | 2020-2025 | 2025-2030 | 2030-2035 | 2035-2040 |
|------------|-----------|-----------|-----------|-----------|-----------|
| 0-5        | 0.078950862 | 0.079811962 | 0.079801436 | 0.078690410 | 0.079812667 |
| 6-10       | 0.001933294 | 0.001880566 | 0.001881198 | 0.001949608 | 0.001880523 |
| 11-15      | 0.001319382 | 0.001274485 | 0.001275022 | 0.001333335 | 0.001274449 |
| 16-20      | 0.001986947 | 0.001947866 | 0.001948336 | 0.001998978 | 0.001947834 |
| 21-25      | 0.002527599 | 0.002492997 | 0.002493415 | 0.002538197 | 0.002492969 |
| 26-30      | 0.002660211 | 0.002643911 | 0.002644109 | 0.002665183 | 0.002643898 |
| 31-35      | 0.003060774 | 0.003057746 | 0.003057783 | 0.003061695 | 0.003057744 |
| 36-40      | 0.003852993 | 0.003866361 | 0.003866198 | 0.00384894  | 0.003866371 |
| 41-45      | 0.005072187 | 0.005121246 | 0.005120647 | 0.005057372 | 0.005121287 |
| 46-50      | 0.006973539 | 0.007061808 | 0.007060728 | 0.006946935 | 0.00706188 |
| 51-55      | 0.010181634 | 0.010390771 | 0.010388202 | 0.01011892  | 0.010390943 |
| 56-60      | 0.015104994 | 0.015537032 | 0.015531705 | 0.014976105 | 0.015537389 |
| 61-65      | 0.026400296 | 0.027413799 | 0.027401245 | 0.026099806 | 0.02741464 |
| 66-70      | 0.041602087 | 0.043174364 | 0.043154895 | 0.041135753 | 0.043175669 |
| 71-75      | 0.068059835 | 0.070355051 | 0.070326684 | 0.067377325 | 0.070356951 |
| 76-80      | 0.10938416 | 0.112211124 | 0.112176317 | 0.1085931  | 0.112213456 |
| 81-85      | 0.169931026 | 0.172822056 | 0.172786611 | 0.169602133 | 0.172824431 |
| 86-90      | 0.27903101  | 0.281348352 | 0.281320061 | 0.278330617 | 0.281350247 |

4. Conclusion

From table 4 it can be seen that the average error for each value of $k_t$ is small enough, and the forecasting result can present the original value, so the Neural Network Method can be used in forecasting mortality rate in Indonesia.

References

[1] Baran, S., Gall, J., Ispany, M., & Pap, M. (2007). Forecasting Hungarian Mortality Rates Using the Lee-Carter Method. *Acta Oeconomica*, 57, 21-34.

[2] Bishop, C. M. (2006). Neural Network. In *Pattern Recognition and Machine Learning* (pp. 225-290). Cambridge: springer.

[3] Cox, P. R. (1976). *Demography fifth edition*. New York: Cambridge University Press.

[4] Fausett, L. (1994). Backpropagation Neural Net. In *Fundamentals of Neural Network: architectures, algorithms, and applications* (pp. 289-328). USA: rentice-Hall, Inc. Upper Saddle River, NJ.

[5] Heligman, L., & Pollard, J. H. (1980). The Age Pattern of Mortality. *Journal Institue of Actuaries*, 49-80.

[6] Hinderer, K., Rieder, U., & Stieglitz, M. (2016). *Dynamic Optimization: Deterministic and Stochastic Models*. Switzerland: Springer Nature.
[7] Khamladze, E. V. (2013). *Statistical Methods With Application to Demography and Life Insurance*. New York: Taylor & Francis Group.

[8] Lee, R. D., & Carter, L. R. (1992). Modeling and Forecasting U.S Mortality. *Journal of the American Statistical Association, 87*(419), 659-671.

[9] Lee, R. D., & Rofman, R. (1994). Modeling and Forecasting Mortality in Chile. *Notas, 22*(59), 182-213.

[10] Li, S. H., & Chan, W. S. (2004). Estimation of Complete period life tables for singaporeans. *Journal of Actuarial Practice, 11*, 129-146.

[11] Ngataman, N., Ibrahim, R. I., & Yusuf, M. M. (2016). Forecasting the Morltaity Rates of Malaysian Population Using Lee-Carter Method. *American Institute of Physics*.

[12] Sammuel, C., & Webb, G. I. (2017). *Encyclopedia of Machine Learning and Data Mining*. New York: Springer Nature.

[13] Severine, G. (2012). Forecasting mortality: when academia meets practice. *European Actuarial Journal, Volume 2 (1)*, 49-76.

[14] Zhang, F. (2010). *Matrix Theory, Basic Results and Techniques*. New York: Springer.