

PAPER • OPEN ACCESS

Comparison multi-layer perceptron and linear regression for time series prediction of novel coronavirus covid-19 data in West Java

To cite this article: I N Yulita *et al* 2021 *J. Phys.: Conf. Ser.* **1722** 012021

View the [article online](#) for updates and enhancements.

You may also like

- [Causal impact of PSBB and new normal in West Java](#)
R S Pontoh, S Zahroh and N Sunengsih
- [The effect of COVID-19 outbreak on the competitive and comparative advantages of rice production in West Java, Indonesia](#)
J F Sinuraya and A Setiyanto
- [Population density and rainfall increase the incidence of Dengue Haemorrhagic Fever in West Java, Indonesia: a secondary data analysis from 2017 to 2021](#)
Juariah, Samiran and H Fuadzy



ECS The Electrochemical Society
Advancing solid state & electrochemical science & technology

ECS UNITED

247th ECS Meeting
Montréal, Canada
May 18-22, 2025
Palais des Congrès de Montréal

Early registration deadline: April 21, 2025

Unite with the ECS Community

Comparison multi-layer perceptron and linear regression for time series prediction of novel coronavirus covid-19 data in West Java

I N Yulita *, A S Abdullah, A Helen, S Hadi, A Sholahuddin, and J Rejito

Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjadjaran, Jl. Raya Bandung – Sumedang KM 21, Sumedang, Jawa
Barat, Indonesia

*intan.nurma@unpad.ac.id

Abstract. Until now, the pandemic conditions of Covid-19 are still ravaging the world, even in Indonesia and West Java. Various attempts have been made to stop it. West Java implements Large Scale Social Restrictions, is known as Pembatasan Sosial Skala Besar (PSBB). However, over time, a discourse emerged to loosen PSBB. One of the World Health Organization's (WHO) requirements to loosen is the effective reproduction rate of Corona Virus cases below 1. Therefore, this study focuses on predicting the number of cases in West Java. The methods based on multi-layer perception (MLP) and linear regression (LR). The data were obtained from the C Covid -19 positive case from March to mid-August 2020 in West Java. The experiments show that MLP reaches optimal if it used 13 hidden layers with learning rate and momentum = 0.1. The MLP had a smaller error than LR. Both of them predict the number of cases in the next 30 days from August 14, 2020. The results show that West Java will still have an increase in the number of new cases of Covid -19.

1. Introduction

China reported many cases of acute respiratory syndrome in China at the end of December 2019 [1]. Chinese scientists immediately identified the new Coronavirus as the leading causative agent. This disease is now referred to as corona virus disease 2019 or Covid-19 [2]. The initial outbreak in Wuhan spread rapidly, affecting other regions in China. Similar cases were immediately detected in several other countries. Almost every continent has countries infected with Coronavirus such as Asia, Europe, Australia, Africa, and America. On March 2nd, 2020, the Indonesian government announced two positive cases of Covid-19 patients [3]. The cases came from West Java. However, the University of Indonesia (UI) Epidemiology expert said that the SARS-CoV-2 Coronavirus as a cause of Covid-19 had entered Indonesia since early January [4]. Indonesia is less responsive so that it allowed the door to Indonesia to be a Covid-19 entry. The Indonesian government did not immediately close direct flight access to and from Wuhan. The cumulative positive case report of Covid-19 shows that from March to June graph data has increased significantly in several regions in Indonesia.

The government handled this virus spreading through PSBB [5]. It is an implementation of restrictions on activities in public places and quarantines themselves in the house. The purpose of the PSBB is to prevent the widespread of the Covid-19 virus that is currently happening [6]. The restrictions on activities carried out include the consolation of schools and workplaces, religious, and restrictions on other activities in public places. However, not everything can go well with this policy. Its side effect causes many industries and livelihoods to be hampered [7]. The sectors most affected



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

are tourism and public transport services. It has stopped the economy [8]. As a result, the finance minister in Indonesia predicts economic growth this year is expected to decline. The condition aroused the discourse to loosen PSBB. However, this easing has several prerequisites by the direction of the World Health Organization (WHO). One of the main requirements is the effective reproduction rate of cases below 1 [9]. Other requirements are that the health service system must have a maximum capacity and adequate supervision and swab test capacity.

In order to support the main requirements, it is necessary to predict new cases. This prediction is useful for determining the preparation needed in facing the new normal. In computer science, forecasting is an activity to predict future events by considering data from the past [10]. Many methods in statistics can be used to forecast data time series. These methods are expected to identify the data used to predict conditions in the future. Holt-Winters is one of the most commonly used methods for prediction of time series [11]. This method can handle seasonal behavior in data based on past data. This method is very good at predicting data patterns with seasonal influences with trend elements that emerge simultaneously. Another prediction method is the Autoregressive Integrated Moving Average (ARIMA). This method produces predictions based on the synthesis of historical data patterns [12]. The ARIMA method will work well if the data in the time series used are dependent or statistically related to each other.

Multi-layer perceptron (MLP) is one of the methods in predicting time series [13]. This method can analyze very complex problems. The training process has goals that are used to generalize well. So during the process, the network changes the weight of the network in shaping the network architecture to get better in each iteration. This method has been successful in predicting various cases, including finance [14], electricity price [15], precipitation [16], and hourly solar radiation forecasting [17]. Therefore, this study proposes MLP to predict Covid-19 data in West Java. The evaluation in this study was based Root Mean Squared Error (RMSE). The analysis is also done by comparing it with linear regression (LR) because new case data in west Java shows an upward trend, such as forming a straight line.

2. Method

The workflow of this is described in Figure 1.

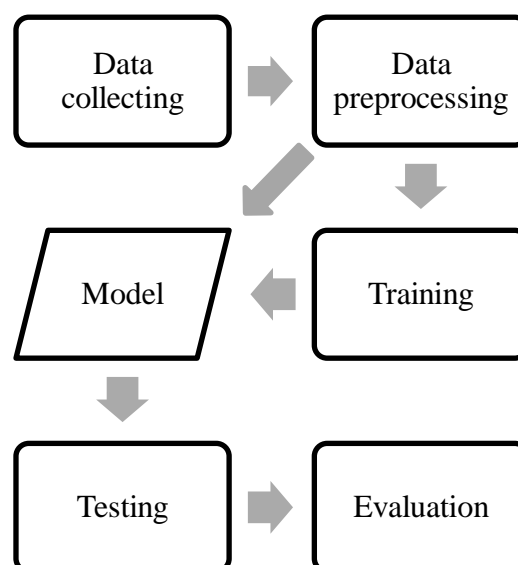


Figure 1. The methodology of Covid -19 prediction

Data comes from <https://pikobar.jabbarprov.go.id/>. This site is the center of information and coordination in West Java. The period is from March 2 to August 14, 2020. Two variables from the dataset are timestamp, which shows the date and newCase, which shows the number of new cases. These two variables are processed so that they become 25 attributes, based on:

1. The correlation with the name of the day so that there were seven attributes obtained.
2. The correlation to the weekend so that there was one attribute that is generated.
3. The linkage timestamp and remapping so that there were three additional attributes.
4. The lag with the window size was seven, so seven attributes are extracted.
5. The linkage between remapping and lag so that seven more attributes will be used.

The total number of attributes of preprocessing is 25. Data from March 2 to August 14, 2020, consisted of 166. The data was divided into two, namely 80% as training data and 20% as testing data. The training process aims to build a prediction model. The models were formed by MLP and LR. They predicted the data. The model evaluation was based on Root Mean Squared Error (RMSE). Both calculations are shown in equation 1.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x)_i - x_i)^2} \quad (1)$$

Where

i show index data based on timestamp order,

n is the number of data

xi is the i -th data

$f(xi)$ is the i -th data prediction.

2.1. Multi-layer perceptron

The artificial neural network (ANN) method has some type. One of them is MLP. It has one or more hidden layers [19], which contains neurons between the input and output layers. For each data in the training data, the input is calculated for the node based on the current input and network values. Initial weights for the input layer, hidden layer, and bias can be initialized. At random, it usually ranges from -0.1 to 1.0. The bias node consists of two, namely, the input layer connected to the nodes in the hidden layer and the hidden layer connected to the output layer. After all initial values are initialized, input, output, and error are calculated. Next, it generates output for the node using the sigmoid activation function. After obtaining the value of the activation function, calculate the error value between the predicted value and the actual value. After the error value is calculated, then reversed to the previous layer (back propagated). The error value generated from the previous step is used to update the relation weight [20].

2.2. Linear regression

Simple linear regression is a linear relationship between the independent (X) and the dependent variable (Y) [21]. It also determines the direction of the relationship between the independent variable with the dependent variable, whether positive or negative. The relationship builds a model which is used to predict the future data. If the value of the independent variable is known that we can calculate value of the dependent variable. The equation is shown in equation 2.

$$f(x) = ax + b \quad (2)$$

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(f(x)_i - f(\bar{x}))}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

$$b = f(\bar{x}) - a\bar{x} \quad (4)$$

Where

a is the regression coefficient

b is a constant value

3. Result and Discussion

The main objective of various experiments was to find the optimal parameters of MLP, namely the number of hidden layers, the value of learning rate and momentum. The MLP ran for 500 iterations. After the optimal parameters were found, MLP was compared to LR based on RMSE.

3.1. Analysis of hidden layers in MLP

This experiment aim to find the number of hidden layers from MLP that produced the smallest RMSE. Some values were tested as based on:

1. The number of predictive outputs was one
2. Half of the sum of attributes and outputs was 13
3. The number of attributes was 25
4. The number of attributes and classes was 26

The learning rate in this experiment was 0.1, and the momentum was 0.1. The results are shown in Table 1.

Table 1. Analysis of hidden layers

Hidden layers	RMSE
1	81.01
13	80.52
25	81.28
26	81.37

Table 1 shows that MLP has the highest error if the hidden layer used was 26. Decreasing the number of hidden layers showed a decrease in error. On the other hand, the number of hidden layers as much as 13 had a smaller error than 1. It showed that the hidden layer was a generated feature extraction in MLP. Each layer represented a feature. If the number of hidden layers was above the number of features, it caused the feature not to be extracted properly. It became a factor that increased errors.

3.2. Analysis of learning rate in MLP

After the hidden layer parameters were obtained in the previous experiment, the next target was to find the optimal learning rate. The tested values were 0.1, 0.2, and 0.3. The experiments used 13 hidden layers and momentum = 0.1. The learning rate aims to correct the weight during the training process. The values ranged between 0 and 1. The greater the value of the learning rate, the training process would run faster. However, it caused the process to jump over optimal conditions. But if the value was too small, the process runs slowly. Table 2 shows that the most optimal conditions when the learning rate = 0.1. The error got bigger if the value was enlarged.

Table 2. Analysis of learning rate

Learning rate	RMSE
0.1	80.52
0.2	81.63
0.3	103.59

3.3. Analysis of momentum in MLP

The momentum played a role in changing weights. The purpose of momentum in the backpropagation was to launch training and prevent the weight from stopping at a value that was not optimal. Thus, momentum could improve MLP performance when compared to gradient descent. The values ranged from 0 to 1. The experiments of momentum from this study only used three values: 0.1, 0.2, and 0.3, while the number of hidden layers = 13 and learning rate = 0.1. Table 3 shows that the best momentum value was 0.1. The higher the momentum value, the higher the error.

Table 3. Analysis of momentum

Momentum	RMSE
----------	------

0.1	80.52
0.2	107.68
0.3	228.29

3.4. Comparison of MLP and LR

Tables 1, 2, and 3 show that the optimal condition of MLP when there were 13 hidden layers and the learning rate and momentum is 0.1. The optimal conditions of the MLP were compared to the LR, as shown in Table 4. The difference in error between the two ranged from 2% to RMSE. It shows that MLP was better than LR in this study. The validation process of the test data was visualized by Figure 2, and Figure 3. On the dates when the peak cases were added, the two methods got the most prediction failures. But it seems that LR had failed more than MLP. The data on the peak of cases in July were better modeled by MLP than LR.

Table 4. Analysis of methods

Method	RMSE
MLP	80.52
LR	82.37

The data test provided an overview of how the model can be trusted in providing predictions. The purpose of modeling was to predict future data. Through this research, the data from the next 14 days were predicted using both methods. The data prediction from August 15 to 28, 2020, is presented in Table 5. Also, its visualization is presented in Figure 4 and 5. It show that in the next 14 days, West Java has a definite case increase. But, the number of cases predicted by MLP and LR is different. The difference in their predictions ranges from 54 to 163.

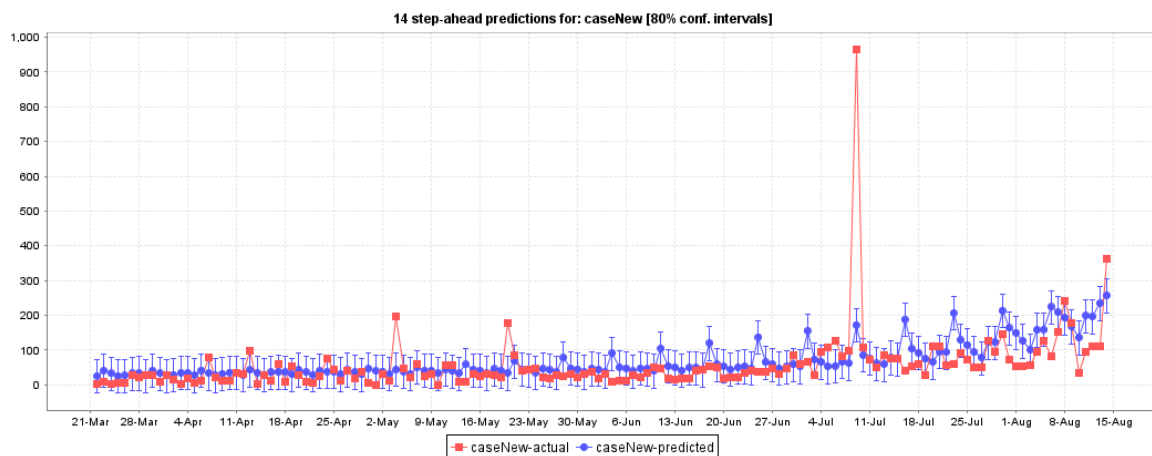


Figure 2. Validation data of MLP

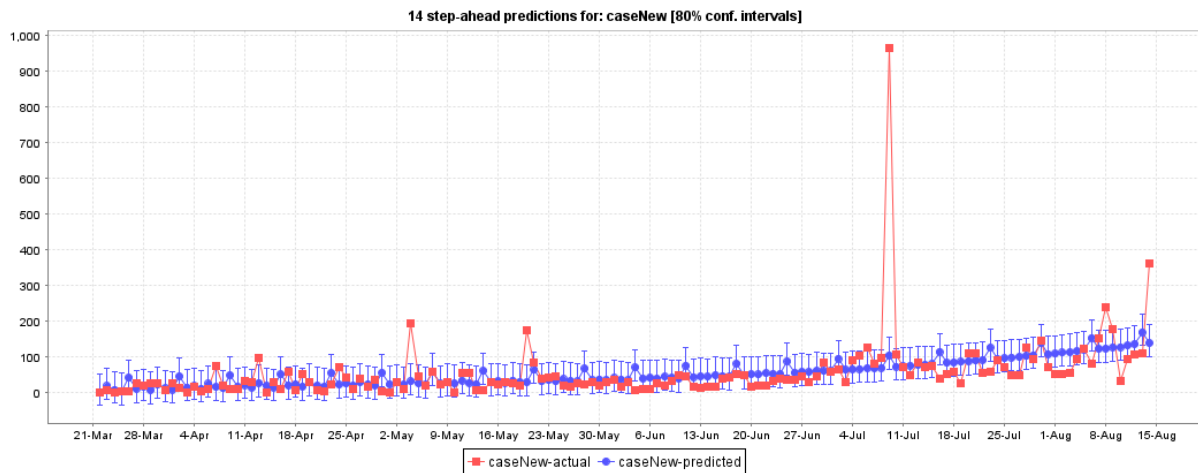


Figure 3. Validation data of LR

Table 5. Data prediction of MLP and LR for the next 14 days

Date	Prediction data	
	MLP	LR
08-15-2020*	219	149
08-16-2020*	183	128
08-17-2020*	165	146
08-18-2020*	229	133
08-19-2020*	226	150
08-20-2020*	254	207
08-21-2020*	294	159
08-22-2020*	269	157
08-23-2020*	245	163
08-24-2020*	217	160
08-25-2020*	263	169
08-26-2020*	248	177
08-27-2020*	262	208
08-28-2020*	340	177

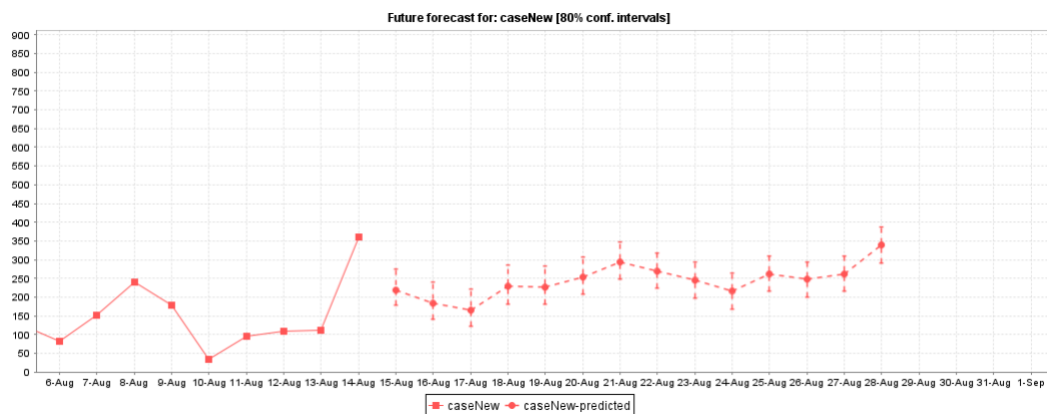


Figure 4. Prediction data of MLP

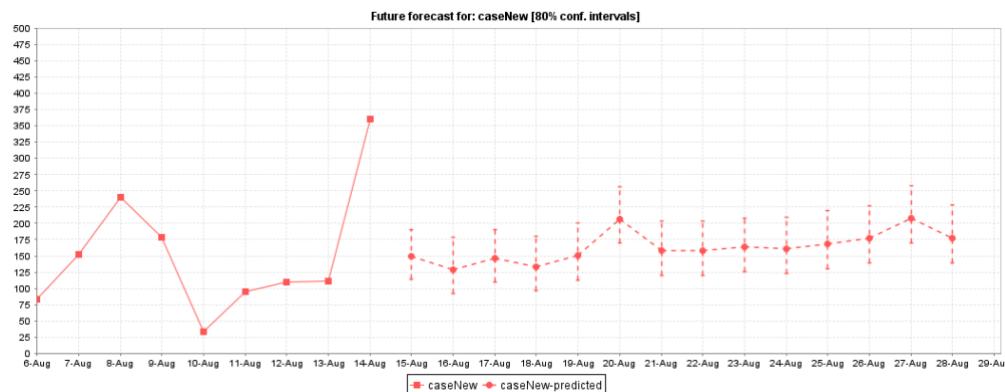


Figure 5. Prediction data of LR

4. Conclusion

Based on Tables 1, 2, and 3, the optimal conditions were reached if MLP used 13 hidden layers for learning rate and momentum was 0.1, respectively. The hidden layer has a function as a generated feature extraction. The pre-processing produced 25 features so that the optimal conditions were reached if the number of hidden layers is half of the number of attributes. The optimal learning rate in this study was the smallest value in the experiment. If the learning rate was higher, then the optimal conditions were skipped over. The momentum with a value = 0.1 also proved useful in improving the weights in the network.

Table 4 shows that the optimal MLP had a smaller error than the LR. Both methods also indicate that West Java will experience an increase in the number of new cases in the next 14 days from August 15, 2020. It indicates that West Java must be strict in implementing health protocols, especially in the new normal. However, the results cannot be used as the primary reference for Covid-19 handling in West Java because many factors influence the spread of Covid-19. It is not enough to just based on time series data.

Acknowledgments

We thanks to Rector Universitas Padjadjaran. Financial support was received from Academic Leadership Grant UNPAD contract number: 1427/UN6.3.1/LT/2020.

References

- [1] Qin C et al 2020 Dysregulation of immune response in patients with Covid-19 in Wuhan, China *Clinical Infectious Diseases* **71(15)**: 762-768
- [2] Shi H, Han X, Jiang N, Cao Y, Awalid O, Gu J, Fan Y and Zheng C 2020 Radiological findings from 81 patients with Covid-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet Infectious Diseases* **20(1)** 425-434
- [3] Tosepu R, Gunawan J, Effendy D S, Ahmad L O A I, Lestari H, Bahar H and Asfian P 2020 Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia *Science of The Total Environment* **725**
- [4] Kurniawati K, Santosa F H and Bahri S 2020 Sosialisasi Hidup Sehat di Tengah Wabah Virus Corona *JPMB: Jurnal Pemberdayaan Masyarakat Berkarakter* **3(1)** 58-65
- [5] Nasruddin R and Haq I 2020 Pembatasan Sosial Berskala Besar (PSBB) dan Masyarakat Berpenghasilan Rendah *SALAM: Jurnal Sosial dan Budaya Syar-i* **7(7)** 639-648
- [6] Muhyiddin M 2020 Covid-19, New Normal, dan Perencanaan Pembangunan di Indonesia, *The Indonesian Journal of Development Planning* **4(2)** 240-252
- [7] Hadiwardoyo W 2020 Kerugian Ekonomi Nasional Akibat Pandemi Covid-19 *BASKARA: Journal of Business & Entrepreneurship* **2(2)** 83-92

- [8] Fathoni A 2019 Dampak Covic 19 dan Kebijakan Psbb Pemerintah Terhadap UMKM di Wiyung Surabaya Dinar: *Jurnal Prodi Ekonomi Syariah* **3(1)** 30-69.
- [9] Rohmi N and Nur R 2020 Sebelum Longgarkan PSBB, Simak Syarat WHO dan Bappenas Berikut Ini” Wired May 29 2020 [Online] Available: <https://www.kompas.com/tren/read/2020/05/29/142500065/sebelum-longgarkan-psbb-simak-syarat-who-dan-bappenas-berikut-ini?page=all> [Accesed: May 30 2020]
- [10] Guijo-Rubio D, Gutiérrez P A, Casanova C, Sanz-Justo J, Salcedo-Sanz S and Martínez C H 2018 Prediction of low-visibility events due to fog using ordinal classification *Atmospheric Research* **214** 64-73
- [11] Tratar L F and Strmčnik E 2016 The comparison of Holt–Winters method and Multiple regression method: A case study *Energy* **109** 266-276
- [12] Guha B and Bandyopadhyay G 2016 Gold price forecasting using ARIMA model *Journal of Advanced Management Science* **4(2)**
- [13] Yulita I N, Purwani S, Rosadi R and Awangga R M 2017 A Quantization of Deep Belief Networks for Long Short-Term Memory in Sleep Stage Detection 2017 *IEEE International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)* pp. 1-5
- [14] Ravi V, Pradeepkumar D, and Deb K 2017 Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms *Swarm and Evolutionary Computation* **36** 136-149
- [15] Dudek, G 2016 Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting *International Journal of Forecasting* **32(3)** 1057-1060
- [16] Choubin B, Khalighi-Sigaroodi S, Malekian A and Kişi Ö 2016 Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals *Hydrological Sciences Journal* **61(6)** 1001-1009
- [17] Azimi R, Ghayekhloo M and Ghofrani M 2016 A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting *Energy Conversion and Management* **118** 331-344
- [18] Dudek G 2016 Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research* **130** 139-147
- [19] Yulita I N, Rosadi R, Purwani S and Suryani M 2018 Multi-layer perceptron for sleep stage classification *Journal of Physics: Conference Series* Vol. **1028**, No. **1**, p. 012212
- [20] Fadhlullah M U, Resahya A, Nugraha D F, and Yulita I N 2018 Sleep stages identification in patients with sleep disorder using k-means clustering *Journal of Physics: Conference Series* **1013(1)** 1-5
- [21] Farkas C, Fenyvesi L and Petróczki K 2019 Multiple linear regression model of Golden apple's failure characteristics under repeated compressive load *Potravinarstvo Slovak Journal of Food Sciences* **13(1)** 793-799