Variance: Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X
Volume 7 Issue 1 | April 2025
10.30598/variancevol7iss1page83-92

FORECASTING THE COMBINED STOCK PRICE INDEX (IHSG) USING THE RADIAL BASIS FUNCTION NEURAL NETWORK METHOD

Della Fitriawan[1], Neva Satyahadewi[2]*, Wirda Andani[3]
[1, 2,3]Statistics Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University
Jl. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, West Kalimantan, Indonesia

Corresponding Author's Email: *neva.satya@math.untan.ac.id*

Abstract: *The capital market is one of the most critical factors in national economic development in Indonesia, as many industries and companies have previously used the capital market as a medium to absorb investment so that their financial position can be strengthened. The main indicator that can reflect the performance of the capital market is the Composite Stock Price Index (IHSG). The IHSG can be used to assess the general situation occurring in the market. Data IHSG is data obtained from the past and used to predict the future, also called time series data. Predictions on IHSG data need to be made so that investors can easily see capital market movements and know the policies that will be taken in the future. The Radial Basis Function Neural Network (RBFNN) method is used. RBFNN aims to get more efficient results because this method does not need to make the data stationary. The analysis results were carried out on a secondary data sample size of 1114 data, which obtained the highest forecasting price of Rp6157,619 on August 2, 2023. Meanwhile, the lowest forecast price on August 5, 2023, is IDR 5564,828 from August 1, 2023, to August 5, 2023.*

Keywords: *IHSG, Stock Index, Radial Basis Function Neural Network.*

I. INTRODUCTION

Investment is a capital investment for many existing activities, usually long-term, aiming to gain future profit [1]. Investments can be in buildings, gold, land, machinery, or financial assets such as stocks, deposits, and bonds [2]. The Composite Stock Price Index (IHSG) is an economic index that shows the movement of stock prices and describes the combined performance of all stocks listed on the Indonesia Stock Exchange (IDX). IHSG on the IDX includes moving common and preferred stock prices [3]. IHSG data has huge fluctuations. The IHSG price is challenging to predict the increase or decrease caused by rapid movements [4].

Therefore, an analysis is needed, namely stock price forecasting, which is aimed at considering and anticipating risks that may occur in the future by looking at the picture and information [5]. The forecast is also defined as an activity to estimate the possibility of events that occur in the future based on previous events [6]. Problem forecasting can be calculated using the Artificial Neural Networks (ANN) method.

JST is defined as nonlinear modeling with modeling functions with a complex relationship between input and output in discovering data patterns used in research [7]. JST can solve many very complex problems [8]. The JST has excellent capabilities in pattern recognition techniques [9]. JST has two components, namely neurons and weights [10]. The JST model has been continuously developed from 1940 to 1988; one of the forecasting or prediction methods included in the artificial neural network method group is the Radial Basis Function Neural Network (RBFNN) [11].

The RBFNN model consists of an input layer, a hidden layer, and an output layer. RBFNN only has weights on the network connected from the hidden layer to the output layer. There is an activation function in the hidden layer that outputs a value in the form of a nonlinear equation, while in the output layer or the result of the RBFNN process, it is a value in the form of a linear equation.

Autocorrelation Function (ACF) and Partial Autocorrelation Function PACF are tools used in RBFNN. ACF shows the correlation between data at a certain lag. PACF eliminates the effects of intermediate correlations. This method is expected to provide a good picture or forecast for the future so that investors will know the policies that will be adopted later.

Submitted: 2024-03-20 *Accepted: 2025-06-08*

2. METHODOLOGY

2.1. Data Source

The data used in this research is secondary data, namely daily data from the Composite Stock Price Index (IHSG) for the last 5 years, from January 2, 2019, to July 31, 2023. Data obtained from http://www.finance.yahoo.com/. The amount of data used was 1114 data.

2.2. Data Analysis

The analytical method used in this research is the Radial Basis Function Neural Network (RBFNN) method to predict the Composite Stock Price Index (IHSG) value for the next few days. This research was analyzed using RStudio software. The testing stages using the RBFNN method can be explained as follows:

2.2.1. Divide Data Based on Training and Testing Data

Some frequently used training data and testing data compositions are 80% and 20%, 75% and 25%, 60% and 40%, and so on. This research uses 80% training data and 20% testing data. So, the training data is formed with a proportion of 80%, namely 891 data and testing data with a proportion of 20%, namely 223 data. The training data is used to determine the next stage.

2.2.2. Carrying out the RBFNN training process

1. Pre-processing data

   Pre-processing is used for data that has been obtained and then processed for each anomaly [12]. Data normalization is a process or action to change data form into a more specific value in the range 0-1 [13]. The Min-Max Normalization method maintains the relationship between the distribution form and the original data values. Data normalization is formulated with the following equation:

$$x_r' = \frac{x_r - x_{min}}{x_{max} - x_{min}} \tag{1}$$

   Information:

   x_r' : Data normalization results.

   x_r : Data actual stock data value.

   x_{min} : Minimum data value.

   x_{max} : Maximum data value.

2. Determine the input variables that will be used in modeling

   The ACF plot is identified to determine the number of input layers in the RBFNN network structure. Input determination is carried out based on significant lags with the help of the ACF plot on the stock closing data.

3. Determining hidden layers

   At this stage, the k-means method is applied to divide into several groups. Then, k-means will group data with the same characteristics and form the same cluster. Likewise, data with different characteristics will form different clusters.[14]. K-means clustering is grouping previously obtained input data into several clusters so that the center value and variance of each cluster can be calculated. The cluster center is the average (mean) of the cluster. The equation that is often used to solve problems in determining the closest distance is formulated with the following equation:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

Information:

$d(x, y)$: The distance between data at points x and y.

x_i : The i-th attribute value.

y_i : Attribute value i-th comparison.

n : Number of data attributes.

Function radial basis is a function used to calculate the distance between data points and the center of the function. There are several functions that can be used, but the most common is the Gaussian function. Before that, it is necessary to calculate the value first with the following equation:

$$\sigma = \frac{Dmax}{n} \tag{3}$$

Information:

σ : Standard deviation on hidden neurons.

$Dmax$: Maximum distance between clusters.

n : Amount clusters which exists.

So, the equation of the Gaussian function can be formulated as follows:

$$\phi(x) = e^{\left(\frac{-x^2}{2\sigma^2}\right)} \tag{4}$$

Information:

$\phi(x)$: Hidden neuron activation function.

x : Euclidean norms between the input vector x and the center of the hidden neuron.

σ : Standard deviation on hidden neurons.

4. Count forecasting accuracy from RBFNN training results

   Forecasting error can be measured by several criteria. Determining the best model is done by paying attention to the smallest Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) values. From each aid method used, namely Least Square Method and Global Ridge Regression.

5. Compare ACF and PACF plots of LS and GR residuals

2.2.3. Forecasting Data from 1 August to 5 August 2023

In forecasting, data denormalization is carried out again, so that previously normalized data values can return to their actual value based on the forecast results [15]. The data denormalization is formulated as follows:

$$x_r = x'_r(x_{max} - x_{min}) + x_{min} \tag{5}$$

Information:

x_r : Data denormalization results

x'_r : Prediction results before denormalization

x_{min} : Minimum data value

x_{max} : Maximum data value

3. RESULTS AND DISCUSSION

3.2. Data Description

Jakarta Stock Exchange Composite (JKSE), also known as the Composite Stock Price Index (IHSG), is a stock index that shows the stock market's performance in Indonesia. This stock index contains many well-known Indonesian Stock Exchange (BEI) companies. BEI is one of the largest stock exchanges in Southeast Asia. Descriptive statistics from JKSE closing price data can be seen in Table 1.

Table 1. Descriptive Statistics of Closing Stock Prices JKSE

Min	1st Qu	Median	3rd Qu	Max	N
3938	6031	6339	6258	6768	7318

Table 1 shows that the amount of data used in the research was 1114, with the lowest closing stock price of 3938 recorded on March 24, 2020. Meanwhile, the highest closing stock price was recorded at 7318 on September 13, 2022. The average value of the closing price of JKSE shares over the last five years was recorded at 6258 with a standard deviation of 675.2856.

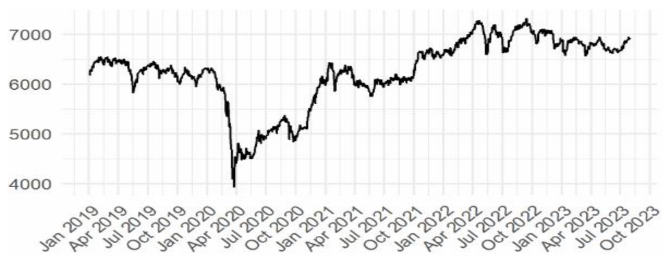

Figure 1. Plot of Closing Prices of JKSE Shares

Based on the JKSE closing share price plot, Figure 1 shows a graph of JKSE closing share prices for the last five years from January 2, 2019, to July 31, 2023. It can be seen that the graph shows a random, non-seasonal, and non-linear pattern. The RBFNN method does not explain the exact assumptions of the data patterns. However, the data characteristics generally used in applying the RBFNN method are non-linear patterns with a large amount of data, so in this research, the RBFNN method can be applied to predict the closing price of JKSE shares.

3.1. Data Sharing

A total of 1114 data were used in this research, which was divided into two, namely training and testing data. Training data is used to form the model, and testing data is used to see the model's accuracy and make forecasts for the future. The proportion of JKSE closing price data that has been divided can be seen in Table 2.

Table 2. Proportion of Training Data and Testing Data

Data	Proportion	Data Amount
Training	80%	891
Testing	20%	223
Total	100%	1114

Table 2 shows the total data, the composition of training data, and the testing data used to forecast the closing price of JKSE shares. The training data used is 80% of the total data, namely 891 data. Meanwhile, testing data is 20% of the total data, namely 223.

3.2. RBFNN Training Process

3.2.1. Pre-processing Data

Data normalization is the first step that must be taken in the forecasting process using the RBFNN method. Data normalization in this study was carried out using the Min-Max Normalization method using Equation (1) to produce numbers 0-1. The calculations for the first 3 data in the data normalization are as follows.

$$z_i = z_1 = \frac{6181.17 - 3937.63}{7318.02 - 3937.63} = 0.66369265$$

$$z_2 = \frac{6221.01 - 3937.63}{7318.02 - 3937.63} = 0.67547827$$

$$z_3 = \frac{6274.57 - 3937.63}{7318.02 - 3937.63} = 0.69131372$$

The normalization results of the training data used can be seen in Table 3.

Table 3. Data Normalization Results

Date	JKSE Price
01-02-19	0.66369265
01-03-19	0.67547827
01-04-19	0.69131372
⋮	⋮
07-27-23	0.8753517
07-28-23	0.8764078
07-31-23	0.8856167

Table 3 shows the results of the normalization of training data with a composition of 1114 JKSE stock closing data, where the total data is 1114.

3.2.2.Identify Models with PACF Plots

Model identification is carried out to obtain a determination of neuron input based on significant lag with the help of the ACF plot. The ACF plot of the closing price of JKSE shares can be seen in Figure 2.

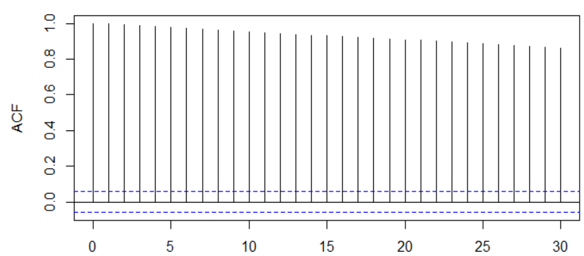

Figure 2. ACF Plot of Closing Prices of JKSE Shares

Based on the ACF plot of the closing price of JKSE shares in Figure 2, the ACF plot does not show lags that have a specific pattern, so 3 input neurons can be taken for the RBFNN model, namely lag-1, lag-2, and lag-3. After that, data preparation was carried out by making the first three columns as representations of the three input neurons and one column as the target data. Data preparation is obtained by filling in the first three columns successively with values $X_{3:n-1}$, $X_{2:n-2}$, $X_{1:n-3}$, and the fourth column (target) with the value. In this context, successively stating the data values from the 3rd to the 3rd order, the 2nd to the 2nd order, and the 1st to the 1st order will represent lag-1, lag-2, and lag-3 data for the data target. This process can be seen in Table 4.

Table 4. Results of Normalization of Training Data

$X_{3:n-1}$	$X_{2:n-2}$	$X_{1:n-3}$	$Y_{4:n}$
0.8121903	0.8146131	0.80269436	0.8082766
0.6332405	0.6476442	0.64765012	0.6369975
0.9330669	0.9225001	0.93517612	0.9189206
0.6194404	0.6371691	0.62937708	0.5945557
0.9282213	0.9285881	0.91384426	0.9272983
0.6824183	0.6785726	0.67109712	0.6919320
0.7470114	0.7425918	0.72037842	0.7327882
...	...	...	...

3.2.3.Determination of Hidden Layers

Below is the calculation of K-means clustering using Equation (2). The centroid and data distance calculations at this stage are carried out starting from 2 clusters to 3 clusters, with the initial center initialization determined randomly. All data used in this research was calculated until the end using centroid values that had been randomly initialized previously, and then iteration continued to obtain new centroid values. The calculation of K-means clustering with 3 clusters is as follows.

1. Clusters first

   The following is a K-means calculation for the first and last data with an initial initialization of .0,8830174

$$d(x_1, \mu_1) = \sqrt{(0{,}8082766 - 0{,}8830174)^2}$$
$$d(x_1, \mu_1) = \sqrt{0{,}00558618718464}$$
$$d(x_1, \mu_1) = 0{,}0747408$$

   The final k-means data calculation is as follows.

$$d(x_{1111}, \mu_1) = \sqrt{(0{,}6636927 - 0{,}8830174)^2}$$
$$d(x_{1111}, \mu_1) = \sqrt{0{,}04810332403009}$$
$$d(x_{1111}, \mu_1) = 0{,}2193247$$

2. Clusters second

   The following is a K-means calculation of the first and last data with an initial centroid initialization of 0.9404673.

$$d(x_1, \mu_2) = \sqrt{(0{,}8082766 - 0{,}9404673)^2}$$
$$d(x_1, \mu_2) = \sqrt{0{,}01747438116649}$$
$$d(x_1, \mu_2) = 0{,}1321907$$

   The final k-means data calculation is as follows.

$$d(x_{1111}, \mu_2) = \sqrt{(0{,}6636927 - 0{,}9404673)^2}$$
$$d(x_{1111}, \mu_2) = \sqrt{0{,}07660417920516}$$
$$d(x_{1111}, \mu_2) = 0{,}2767746$$

3. Clusters third

   The following is a K-means calculation from the first 3 data with an initial centroid initialization of 0.0455565.

$$d(x_1, \mu_3) = \sqrt{(0{,}8082766 - 0{,}0455565)^2}$$
$$d(x_1, \mu_3) = \sqrt{0{,}58174195094401}$$
$$d(x_1, \mu_3) = 0{,}7627201$$

   The final k-means data calculation is as follows.

$$d(x_{1111}, \mu_3) = \sqrt{(0{,}6636927 - 0{,}0455565)^2}$$
$$d(x_{1111}, \mu_3) = \sqrt{0{,}38209236175044}$$
$$d(x_{1111}, \mu_3) = 0{,}6181362$$

   Next, the distance values for each cluster are averaged, then the center value for each cluster will be obtained. This calculation was also carried out on the variables so the results can be seen in Table 5.

Table 5. Center of Each Cluster on Each Variable

Clusters	X_1	X_2	X_3	Y
1	0.6640345	0.6641593	0.6641389	0.6637206
2	0.3015841	0.3017037	0.3021433	0.3020311
3	0.8625079	0.8627757	0.8630651	0.8622047

The radial basis function is a function used to calculate the distance between data points and the center of the function. There are several functions that can be used, but the most common is the Gaussian function with Equation (4). Before that, it is necessary to calculate the standard deviation value for each layer first using Equation (3), which is as follows.

$$\sigma_1 = \frac{6.368475}{495} = 0.01286561$$
$$\sigma_2 = \frac{6.184740}{174} = 0.03554448$$
$$\sigma_3 = \frac{6.059059}{442} = 0.01370828$$

After calculating the standard deviation in each layer, then calculate each hidden layer using the previous Equation (4).

$$\phi_1(x_1) = e^{\left(-\frac{1}{2}\times\left(\frac{||0.8082766-0.6640345||}{(0.01286561)^2}+\frac{||0.8082766-0.3015841||}{(0.01286561)^2}+\frac{||0.8082766-0.8625079||}{(0.01286561)^2}\right)\right)}$$
$$= e^{-1096.0625}$$
$$= 0$$

This calculation is carried out down to the last element in X_1, $,X_2X_3$ and here is an example of calculating the last in variable Y.

$$\phi_3(y_{1111}) = e^{\left(-\frac{1}{2}\times\left(\frac{||0.6636927-0.6637206||}{(0.01286561)^2}+\frac{||0.6636927-0.3020311||}{(0.01286561)^2}+\frac{||0.6636927-0.8622047||}{(0.01286561)^2}\right)\right)}$$
$$= e^{-21.78010075}$$
$$= 3.475541 \times 10^{-10}$$

3.2.4.RBFNN Training Architecture

After determining the input variables that will be used in modeling, the next step is to determine the number of hidden layers. The RBFNN method has no clear limit regarding the maximum number of hidden layers that can be used. Apart from that, the number of hidden layers acts as a link between the input layer and the output layer, affecting the accuracy level of the resulting forecast. Therefore, the more hidden layers used can impact the RBFNN method's learning process. In the case of forecasting the closing price of JKSE shares, 3 hidden layers are used. Further information regarding the variables used in the research can be seen in Table 6.

Table 6. RBFNN Method Input, Output, and Hidden Nodes Variables

Variable	Category
X_1	
X_2	Inputs
X_3	
Hidden layers 3	Hidden layers
Y	Outputs

Table 6 shows an input architecture with 3 hidden layers can be formed. RBFNN training process with architecture $input-$ hidden layers $-output$. That is 3-3-1, as shown in Figure 3.

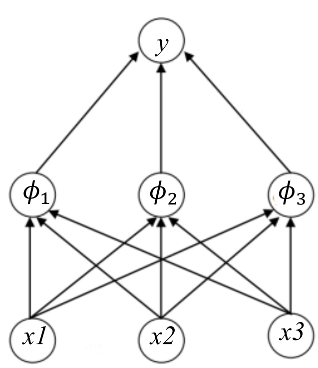

Figure 3. 3-3-1 Architecture

3.2.5.RBFNN Method Data Processing

The architecture formed through input and hidden layers is processed using the RBFNN learning method via the RSNNS and RBF packages in R-studio software. The composition of the data used in the training process is 80% of the JKSE stock closing price data, which has been pre-processed. By applying the architecture and parameters that have been formed using the Least Square Method and Global Ridge Regression, the training results obtained can be seen in Table 7.

Table 7. RBFNN Training Results with LS and GR

Least Square Method		Global Ridge Regression	
MAPE	MSE	MAPE	MSE
30.42788%	0.0005701245	28.7914%	0.001031517

Based on Table 7, it can be seen that the results of RBFNN training with the LS and GR methods produce the smallest MAPE value for the GR method of 28.7914%, and the smallest MSE value for the LS method is 0.0005701245. Because the smallest MAPE and MSE values are not found in the same process, further decision-making will be carried out to see which model fits the original data better using ACF and PACF plots of the residuals of each technique. The suitability of the model can be seen from the lack of significant lag in the residuals of a method.

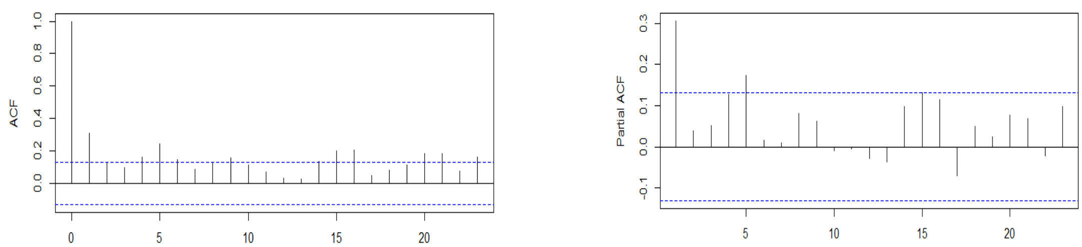

Figure 4. ACF and PACF Plots of LS Model Residuals

Based on the ACF and PACF plots of the LS model residuals in Figure 4, you can see a comparison of the significant lags found in the ACF and PACF plots in the LS method. In the ACF and PACF plots, it is clear that the residuals from the LS method have several significant lags.

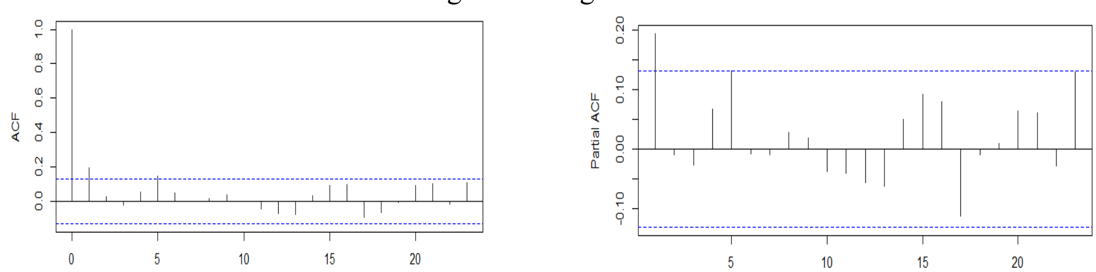

Figure 5. ACF and PACF Plots of GR Model Residuals

Based on the ACF and PACF plots of the GR model residuals in Figure 5, you can compare the significant lags found in the ACF and PACF plots in the GR method. In the ACF and PACF plots, it is clear that the residuals from the GR method have some significant lags, but they are less than those from the LS method. Therefore, it can be concluded that the GR model has a higher level of suitability to the original data than the LS method, so it can be concluded that RBFNN using the GR method is the best model that can be used to predict the JKSE closing stock price.

3.2.6.Forecasting

Forecasting JKSE stock closing price data for the next 5 days using the RBFNN method with the Global Ridge Regression method using data X_1,X_2 and normalized data compiled from random numbers can be seen in Table 8.

Table 8. JKSE Stock Closing Price Forecasting Data

Period	X_1	X_2	X_4
1	0.5	0.5	0.7
2	0.6	0.7	0.6
3	0.7	0.7	0.4
4	0.5	0.4	0.5
5	0.4	0.4	0.6

Based on Table8, the data obtained above, the further forecasting results obtained can be seen in Table 9.

Table 9. Results of Forecasting the Closing Price of JKSE Shares

Period	Forecast results
1	0.5792625
2	0.6567256
3	0.6431015
4	0.4945038
5	0.4813640

Based on Table 9, the data from the forecasting results will then be denormalized using Equation (5) to see the actual stock value obtained from forecasting using RBFNN, with the results can be seen in Table 10.

Table 10. Forecasting Results After Denormalization

Period	Forecast results
August 1, 2023	5895.763
August 2, 2023	6157.619
August 3, 2023	6111.564
August 4, 2023	5609.246
August 5, 2023	5564.828

Based on Table 10, it can be seen that the forecast value for the closing price of JKSE shares appears to be decreasing. The lowest forecast price is IDR 5564,828 on August 5 2023. It can be seen that the share price has experienced a quite significant decline. The graph of forecasting results for the next 5 days of JKSE stock closing prices using the RBFNN method is presented in Figure 6 as follows.

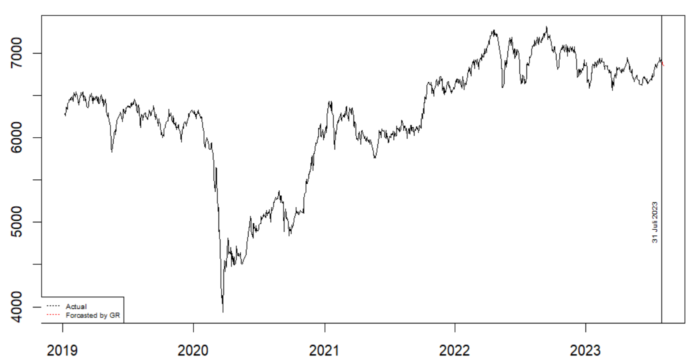

Figure 3. 3-3-1 Architecture

4. CONCLUSION

Based on research conducted by applying the Radial Basis Function Network (RBFNN) method, the results of the analysis of the composite stock price index (IHSG) data for the period from August 1, 2023, to August 5, 2023, showed that stock prices experienced fluctuating developments. The highest forecast price was IDR 6157.619. Meanwhile, the lowest forecast price was IDR 5564.828. This RBFNN method can handle non-linear relationships between input and output variables based on complex and non-linear IHSG data characteristics. Suggestions that can be given include analyzing with other neural network methods. In addition, the Global Ridge Regression can be carried out using different methods, such as Local Ridge Regression.

REFERENCES

[1] TSJ Wijaya and S. Agustin, "Factors that Influence the Value of IHSG Listed on the Indonesian Stock Exchange," J. Ilmu and Ris. Manaj., vol. 4, no. 6, 2015.

[2] RD Vivianti, Darminto, and Y. F, "Feasibility Analysis of Business Investment Based on Capital Budgeting Under Risk (Study of Banyuwangi Regency Regional Drinking Water Company)," J. Adm. Business, vol. 26, no. 1, pp. 1–7, 2015.

[3] BC Pamungkas and A. Darmawan, "The Influence of the USD Exchange Rate and the ASEAN Stock Exchange on the Composite Stock Price Index (IHSG) Study on the Indonesian Stock Exchange for the 2014-2016 Period," J. Adm. Business, vol. 60, no. 1, 2018.

[4] EMS Rochman and A. Djunaidy, "Stock Price Prediction Considering External Factors Using Artificial Neural Networks," J. Ilm. NERO, vol. 1, no. 2, pp. 5–11, 2014.

[5] A. Marjuni, "Simultaneous Stock Price Forecasting Using the Multivariate Singular Spectrum Analysis Model," J. Sist. Inf. Business, vol. 1, pp. 17–25, 2022.

[6] A. Lusiana and P. Yuliarty, "Application of Forecasting Methods to Roof Demand in PT X. Innovative Industry," J. Tek. Ind. ITN Malang, vol. 10, pp. 11–20, 2020.

[7] R. Hidayat and S. Suprapto, "Minimizing Forecasting Error Values with Extreme Learning Machine Algorithms," J. Optimization Syst. Eng., vol. 11, pp. 189–192, 2016.

[8] Sulistiyasni and E. Winarko, "Fingerprint Pattern Classification Using Backpropagation Neural Networks," J. Berk. MIPA, vol. 4, pp. 215–224, 2014.

[9] Suyanto, Artificial Intelligence: Searching, Reasoning, Planning and Learning. Bandung: Informatics, 2021.

[10] A. Sudarsono, "Artificial Neural Networks to Predict Population Growth Rates Using the Backpropagation Method (Case Study in Bengkulu City)," J. Media Infotama, vol. 12, pp. 61–69, 2016.

[11] Y. Pangaribuan and M. Sagala, "Applying Artificial Neural Networks to Recognize Letter Patterns Using the Perceptron Method," J. Tek. UNIKA St. Inframatics Thomas, vol. 2, pp. 53–59, 2017.

[12] S. Dwivedi, P. Kasliwal, and PS Soni, "Comprehensive Study of Data Analytics Tool (Rapidminer, Weka, R tool), Knime," 2016.

[13] AN Alfiyatin, WF Mahmudy, CF Ananda, and YP Anggodo, "Application of Extreme Learning Machine for Forecasting Inflation Rates in Indonesia," J. Teknol. Inf. and Computer Science., vol. 6, pp. 179–186, 2019.

[14] N. Rohmawati, S. Defiyanti, and M. Jajuli, "Implementation of the K-Means Algorithm in Clustering Student Scholarship Applicants," J. Ilm. Technol. Inf. Terap., vol. 2, no. 1, pp. 62–67, 2015.

[15] KP Dewi and A. Kudus, "Application of the Extreme Learning Machine Method in Forecasting the Number of Domestic Flight Passengers in the Riau Islands," Pros. Stat., vol. 7, pp. 589–596, 2021.