Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

DATA MINING AND ITS IMPACT ON INVESTORS' DECISIONS: A CASE STUDY IN THE IRAQI FINANCIAL MARKET USING THE RBF MODEL

Amna Shehab Ahmed Al-Hassan
Assistant lecturer at Al-Nahrain University, College of Business Economics. Baghdad, Iraq.
E-mail: amna.shehab@nahrainuniv.edu.iq, Orcid: https://orcid.org/0000-0002-8885-5390

Abstract

The research addressed the definition of data mining technology in general and the RBF model in particular and its importance in analyzing financial market data to facilitate investors' decisions. The data was collected based on the daily and weekly reports of the Iraqi stock market, where the research sought to test a main hypothesis that states that adopting artificial neural network algorithms The radial basis function (**RBF**) is one of the best data mining algorithms for predicting and analyzing financial indicators for the sample studied. The research reached a set of conclusions, the most important of which is confirming the research hypothesis. Based on this, the researcher recommended the necessity of studying other indicators within a longer period to obtain higher accuracy of the results, as the larger the data sample, the more accurate the techniques used in data mining.

Keywords: data mining, artificial neural networks, RBF model.

Introduction:

As a result of the development taking place in the financial markets and the huge amount of financial data, the need has emerged to use data mining techniques to search and extract knowledge and provide a database that helps absorb this huge amount of financial information to provide an appropriate environment that helps in the process of making the appropriate decision for investors in the stock markets.

Deciding to invest in stocks requires providing prior information about market conditions and price trends to rationalize the correct decision to reduce the incidence of error and the occurrence of financial losses. Therefore, it required the availability of accurate technology that helps provide the required information to guide investors in estimating future stock prices, especially after the financial crisis caused it. The global economy is in a state of shaken confidence among investors in financial markets in general and emerging financial markets in particular, as the economies of these countries are still affected by this. In light of this reality, there remains an urgent need to make the greatest possible effort to restore investor confidence as it is a basic pillar for achieving the fair value of security and practicing effective avoidance. For future crises, therefore, it is necessary to find a means and mechanism that helps investors determine the appropriate and best option for investing in the stock market, by analyzing the market and its fluctuations and studying the variables and their impact on its direction to predict what it will be like in the future. Based on these data, it was necessary to find a solution through predictive models

that can describe the random movement of fluctuations in the returns of financial market indicators and that take into account the linear and nonlinear characteristics of financial time series. This is what prompted us to choose the most prominent data mining techniques, which are functional artificial neural networks. The radial basis function (RBF) is used as a forecasting tool, as it has the appropriate features for forecasting a series of financial indicators and does not require a prior formulation or specific structure.

The research hypothesis also stated that adopting functional artificial neural network algorithms with a radial basis function is one of the best data mining algorithms to predict and analyze the data of the general index of the Iraqi financial market. To achieve the research hypothesis, the variables of the general market index were obtained from the daily and weekly reports of the Iraqi stock market for the period extending between (2017-2023), where the data was collected from the official website of the Iraqi Stock Exchange, then the database was built using the SPSS26 program, and then applied RBF model after handling missing data.

Data mining:

Data mining is a branch of data analysis or part of an analytics strategy used to find hidden and previously unknown patterns in data. It is also known as a computerized sorting process to analyze, process, and explore large sets of data. The use of data mining techniques enables organizations to discover hidden patterns and relationships in their data by converting data into information (knowledge) that is used to solve problems, analyze the future impact of business decisions, and increase profit margins For institutions(Nasser and Ahmed, 2017).

Artificial Neural Networks:

An interconnected group of virtual neurons created by computer programs to resemble the functioning of a biological neuron, or electronic structures (electronic chips designed to simulate the functioning of neurons) that use a mathematical model to process information based on the communication method in computing. Neural networks generally consist of simple processing elements that perform simple work, but the overall behavior of the network is determined by the connections between the various elements, which are called neurons, and the indicators of these elements (gunasekaran and ramaswami, 2011). The first suggestion for the idea of neural networks came from the mechanism of action of brain neurons, which can be likened to biological electrical networks for processing information coming into the brain. In these networks, Donald Hebb suggested that the synapse plays a fundamental role in directing the processing process, and this prompted thinking about the idea of connectivity and artificial neural networks. Artificial neural networks consist of nodes, or as we mentioned previously, neurons or processing units, connected to form a network of nodes, and each connection between these nodes has a set of values called weights that contribute to determining the values resulting from each processing element based on the values entering this element (MAJUMDER and HUSSIAN, 2011).

Radial Basis Function (RBF): A functional neural network with radial basis function is one of the artificial neural networks that was first formulated by two scientists (Lowe and Broomhead) in 1988(DESAI, 2011). It consists of three layers. The first layer represents the network inputs, the second layer is the hidden layer consisting of several nonlinear activation units with radial function. The last layer represents the network outputs. The neurons in the output layer contain linear activation functions, as they have a very effective role in the process of classifying data that contains a percentage of noise. The network is also used in linear and non-linear representation models. It has been used in many fields as an approximation function in predicting time series. Classification and system monitoring. It is used in functions of temporal arrangement and control of machines. It also has hybrid characteristics compared to other networks, in addition to its ability to adapt and modify through learning to find functions linking inputs and outputs to produce a specific decision according to the problem to be solved. The RBF network relies on two types of education, which are as follows:

Unsupervised learning: We find this type of learning in the hidden layer, and this method is called self-learning, meaning that the resulting output values are directly as inputs to the output layer.

Supervised learning: This type is found in the output layer, where the network is trained to reach or obtain the required outputs, and training here is according to the training algorithm (SUTHEEBANJARD and PREMCHAISWADI). Among the most widely used activation functions in applications is the Gaussian activation function. It is used in the hidden layer and takes the following mathematical form, which illustrates the structure of functional neural networks with a radial basis function.

$$\mu_i = exp\left[-\frac{\|X-C_i\|^2}{2\sigma_i^2}\right], \text{j= 1, 2, K, N}$$

Where:

μ_i: represents the output of cell i, where the output values are limited to (1, 0).

X: network input vector.

C_i: The vector of weights accompanying the inputs, which are generated random values.

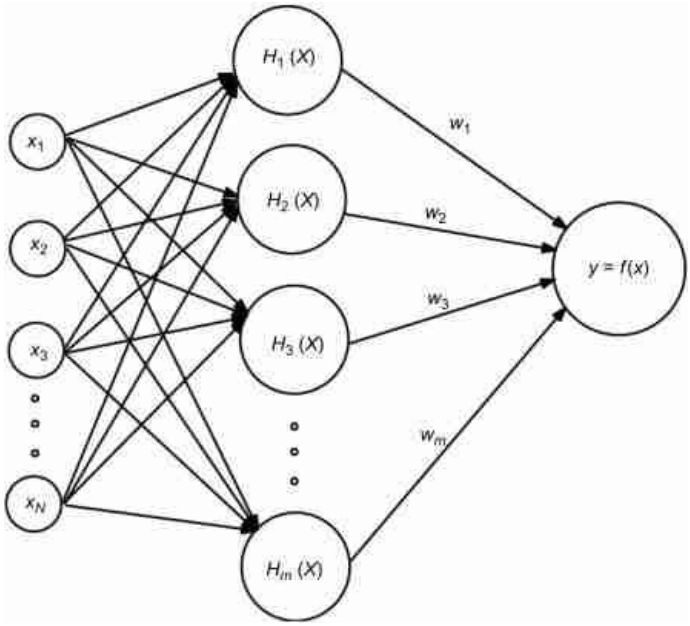


Figure (1) RBF network architecture

Results and discussion:

To carry out the forecasting process using the RBF model, data from the weekly reports of the Iraqi stock market were used for the period extending (1/1/2017 to 11/1/2023). The RBF model was applied and the data was processed using the SPSS26 program.

The method of artificial neural networks with radial basis function was applied to predict the indicators of the Iraqi financial market according to the following steps:

1. Selection of variables

The research variables are the time series of general market indicators represented by opening and closing prices and the percentage of change for the period from (1/1/2017 to 11/1/2023) with (310) observations. After identifying the study variables, we processed the missing data using SPSS according to the Replace missing values function. Table (1) shows the columns of the processed variables. Then, the RBF model for neural networks was applied. The artificial neural network consists of three layers.

- The first layer (input layer): is represented by the values of the opening prices and the percentage change.
- The second layer (hidden layer): consists of 9 neurons.
- The third layer (output layer): consists of the closing prices of the Iraqi stock market. Table (2) shows the inputs of the vector-based functional neural network.

Therefore, the network architecture is in the form (1:9:2).

Table 1: Treatment of missing data

Result Variable	N of Replaced	Case Number of Non-Missing Values	N of Valid Cases	Creating Function

		Missing Values	First	Last		
1	Currentshutdown_1	17	1	301	301	SMEAN(Currentshutdown)
2	Previousclosure_1	16	1	301	301	SMEAN(Previousclosure)
3	Percentageofchange_1	16	1	301	301	SMEAN(Percentageofchange)

The source was prepared by the researcher using the spss program

Table (2) Neural network inputs

Input Layer	Factors	1	SMEAN(Previousclosure)
		2	SMEAN(Percentageofchange)
	Number of Units		362
Hidden Layer	Number of Units		9[a]
	Activation Function		Softmax
Output Layer	Dependent Variables	1	SMEAN(Currentshutdown)
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

The source was prepared by the researcher using the spss program

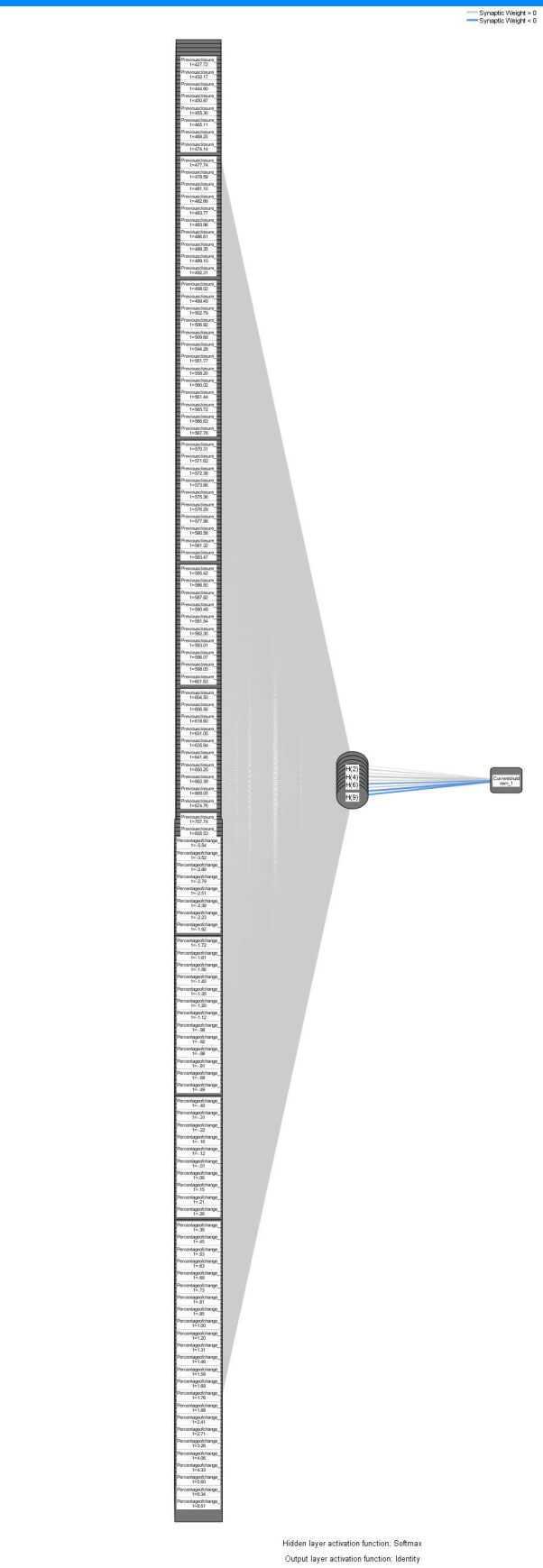

Synaptic Weight > 0
Synaptic Weight < 0
H(2)
H(4)
H(6)
H(9)
Hidden layer activation function: Softmax
Output layer activation function: Identity

Figure (2) shows the architecture of the artificial neural network

2. Data processing

After determining the input time series, the data is processed in a nonlinear autoregressive prediction form with an external input.

3. Data analysis stage

At this stage, the program analyzes the data in a file and describes it in columns. The program also divides the data into two groups based on training and testing as follows:

- 70% of the data was used for training, equivalent to (210) views.
- 30% of the data was used as an independent test on the network, equivalent to (90) observations, to determine the efficiency and accuracy of the model used.

Table (3) Breakdown of the data sample

		N	Percent
Sample	Training	205	95.8%
	Testing	9	4.2%
Valid		214	100.0%
Excluded		87	
Total		301	

The source was prepared by the researcher using the spss program

The number of hidden units is determined by the test data criterion: the "best" number of hidden units is the one that produces the smallest error in the test data.

The final step in building the model is the implementation stage, in which it is ensured that the model is ready for prediction after conducting the training and verification process. This is done by measuring the level of model accuracy performance.

After completing the model training process on the studied data sample, the model's performance is tested, as the model's accuracy indicates the extent of the neural network model's ability to estimate the test sample. Based on this, the model's performance is evaluated according to the measure of the mean square error, the relative error, and the correlation coefficient to measure the relationship between the real values and the predicted values. As shown in Table (4) below:

Table 4: Model performance evaluation

Training	Sum of Squares Error	.33871
	Relative Error	.0332
	Training Time	0:00:03.45
Testing	Sum of Squares Error	.0324[a]
	Relative Error	.0469

The source was prepared by the researcher using the spss program

The relative error is used to calculate significant digits, as the smaller the resulting error value, the better.

Discuss the results

The results of the analysis after implementing the network model showed that the network excluded (87) views that it considered invalid, and the remaining views (214) were divided into (96%) of the sample that was trained to build the model, and approximately (4%) of the sample was used in testing. The model. As for the verification stage, the entire sample was used to estimate the performance of the model, as shown in Table (4) above. To evaluate the ability of the RBF model to predict the values of the financial market index, the researcher calculated the rate of fit or correspondence between the actual values and the predicted values, as this indicator measures the extent to which the predicted values tend to have the same actual index values. Figure (5) shows the difference between the true values and the values. Predicted. The calculated match rate reached (75%), which is a high percentage that indicates the accuracy of the model. The Pearson correlation coefficient was also calculated between the actual data series and the predicted series, and the correlation value reached (99%), which indicates the existence of a strong relationship between the predicted time series and the actual value series, and this confirms the accuracy of the model used in analyzing the studied market data. Based on what was presented previously, this confirms the stated research hypothesis (adopting functional artificial neural network algorithms with a radial basis function is one of the best data mining algorithms to predict and analyze the data of the general index of the Iraqi financial market). Artificial neural networks have the attractive ability to reduce the error bias factor in visual time series. Analyzes of the first neuron network showed that the error coefficient decreased to the level of (0.01), meaning that the accuracy of the predictions reached approximately (99%). The independent variables were also analyzed by measuring their importance, as shown in Table (5) below.

The importance of the independent variable measures the extent to which the expected value of the network model changes for different values of the independent variable. Normalized importance is simply the importance values divided by the largest importance values and is expressed as percentages. Figure (3) the importance of independent variables.

Table (5) The importance of independent variables

	Importance	Normalized Importance
SMEAN(Previousclosure)	.480	92.2%
SMEAN(Percentageofchange)	.520	100.0%

The source was prepared by the researcher using the spss program

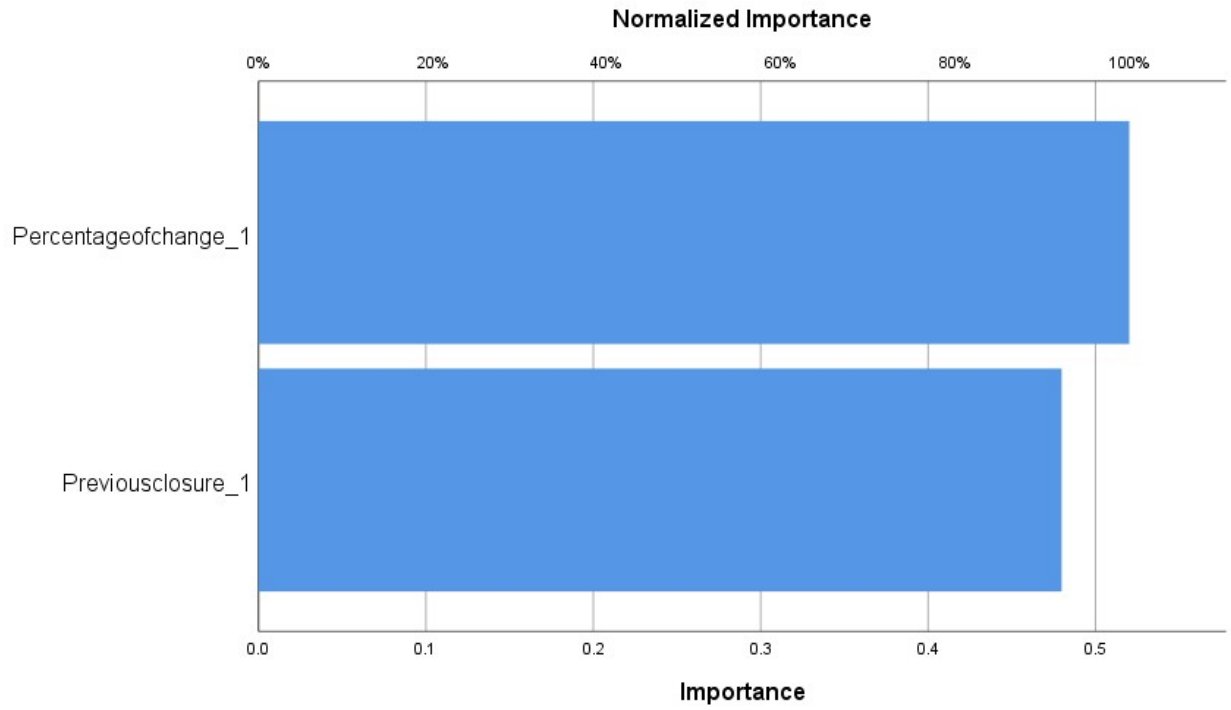
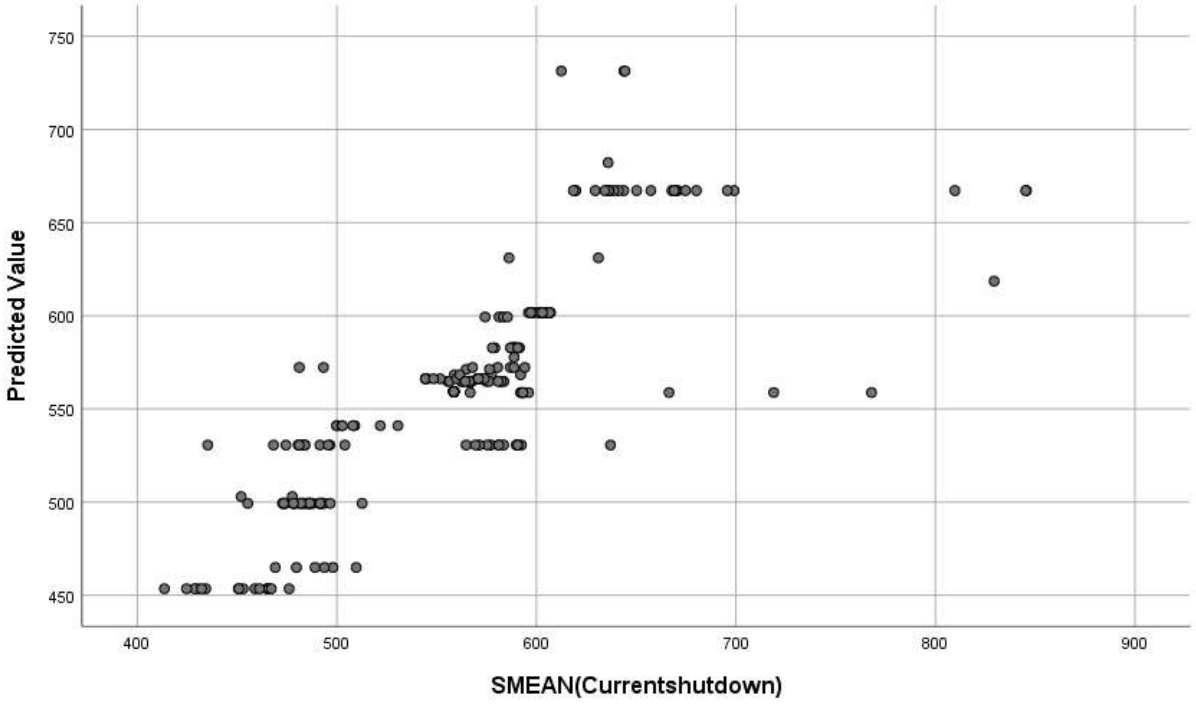
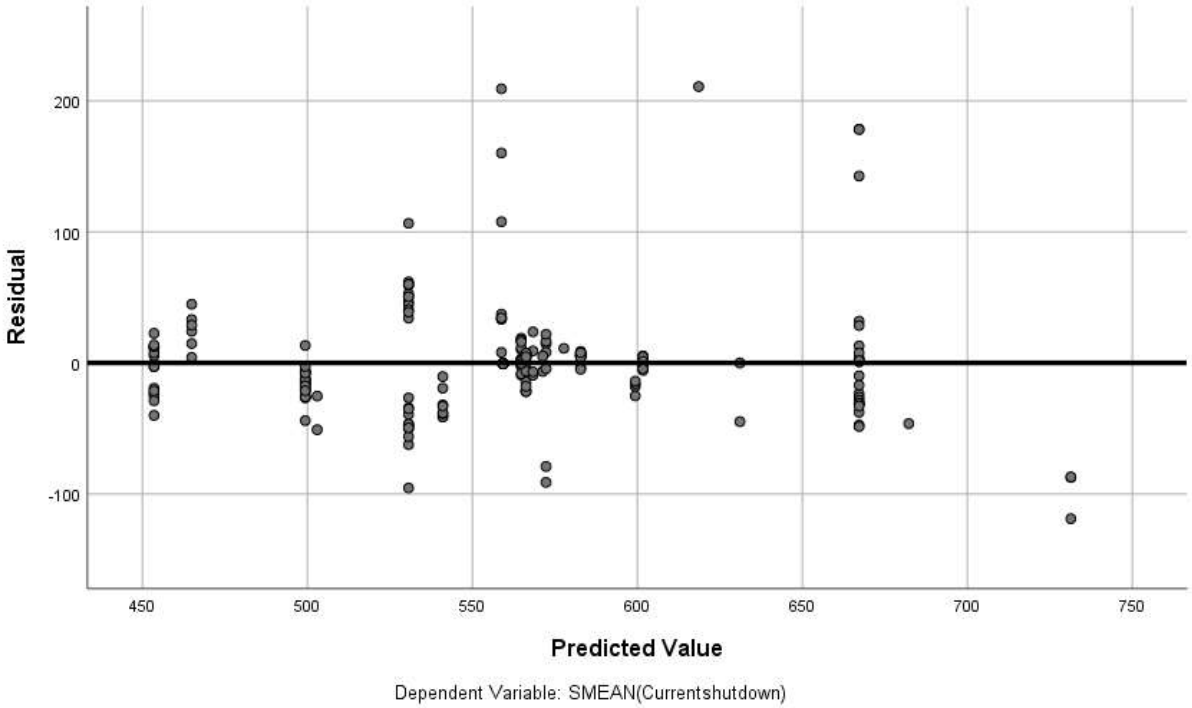


Figure (3) the importance of independent variables

It is clear from Figure (4) the extent of convergence between the estimated values of the previous closing price and the actual values, which explains the strength of the degree of correlation in the testing phase. This confirms the accuracy of the model used in the testing and the possibility of using it to predict future price values and also supports the validity of the research hypothesis in that adopting neural network algorithms Functional synthetic with radial basis function is one of the best data mining algorithms for predicting and analyzing data for the general index of the Iraqi financial market.





Conclusions:

The RBF model can be used to predict the values of financial indicators with high accuracy. The results of predicting the value of the market index and its direction for the next trading day can also be relied upon to make important investment decisions and achieve profitable returns with

a high degree of confidence. The accuracy of the model is also related to the length of the period of the series studied. The longer the time series, the more accurate the model used is. Increasing the size of the data for the study sample leads to improving the model's performance.

Recommendations:

The researcher recommended the necessity of studying other indicators within a longer period to obtain higher accuracy of results, as the larger the data sample, the more accurate the techniques used in data mining become.

Sources:

1. Al-Hassan, A,S,A. Using the Induced Response Function to Measure and Analyze the Impact of Fiscal Policy Shocks on Some Macroeconomic Variables, International Journal of Professional Business Review, Vol. 8, n. 5, e01128, 2023, p. 01-19.
2. Al-Rubaie, Q, L, A, Ahmed, A, S. Measuring and analyzing the repercussions of public debt in financing the general budget deficit for the iraqi economy after 2003 using the (Eviews) program, Materials Today: Proceedings, Vol. 80, Part 3, 2023, Pages 3144-3154.
3. DASE, R. K.; PAWAR, D. D. Application of Artificial Neural Network for Stock Market Prediction: A Review of Literature, International Journal of Machine Intelligence, Vol. 2, Issue 2, 2010, 14-17.
4. DESAI, J. et al. Forecasting of Indian Stock Market Index S&P CNX Nifty 50 Using Artificial Intelligence, Behavioral & Experimental Finance E-journal, Vol. 3, No. 79, 2011.
5. GUNASEKARAN, M.; RAMASWAMI, K. S. Evaluation of Artificial Immune System with Artificial Neural Network for Predicting Bombay Stock Exchange Trends, Journal of Computer Science, Volume 7, Issue 7, 2011, 967-972.
6. HAMID, S. A.; IQBAL, Z. Using Neural Networks for Forecasting Volatility of S&P500 Index Futures Prices, Journal of Business Research 57, 2004, 1116 – 1125.
7. KARA, Y. et al. Predicting the direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, Expert Systems with Applications 38, 2011, 5311–5319.
8. KAZEM, B. I.; MUTLAG A. K. Optimal Brain Surgeon Pruning of Neural Network Models of Manufacturing Processes, Journal of Engineering, N. 3, Vol. 11, 2005, 495-508.
9. MEHRARA, M. et al. Using Technical Analysis with Neural Network for Forecasting Stock Price Index in Tehran Stock Exchange, Middle Eastern Finance and Economics, Volume 6, Issue 6, 2010, 50-61.
10. NAEINI, M. P. et al. Stock Market Value Prediction Using Neural Networks, International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010, 132-136.
11. Nasser, S, A, Ahmed, A, S. Detecting financial fraud using some statistical methods: An applied study in the Iraqi stock market, Al-Mustansiriya Science Journal, Vol. 28, n.1, 2017, Pages 176-186.

12. SUTHEEBANJARD, P.; PREMCHAISWADI, W. Stock Exchange of Thailand Index Prediction Using Back Propagation Neural Networks, Second International Conference on Computer and Network Technology, 2010, 377-380.
13. AAMODT, R. Using Artificial Neural Networks To Forecast Financial Time Series, Master thesis, Norwegian University of Science and Technology, Department of Computer and Information Science, 2010, 95 Pages.
14. LARSEN, J. I. Predicting Stock Prices Using Technical Analysis and Machine Learning, Master thesis, Norwegian University of Science and Technology, Department of Computer and Information Science, 2010.
15. SENOL, D. Prediction of Stock Price Direction by Artificial Neural Network Approach, Master thesis, Bogazici University, Institute for Graduate Studies in Social Sciences, 2008.
16. DESAI, J. et al. Forecasting of Stock Market Indices Using Artificial Neural Networks. Working Paper No. CPI/MBA/2013/0003, Shri Chimanbhai Patel Institutes, Ahmadabad, 2013.
17. MAJUMDER, M.; HUSSIAN, M. A. Forecasting of Indian Stock Market Index Using Artificial Neural Network, Working Paper, 2010.