Intelligent Forecasting in Multi-Criteria Decision-Making

Vladyslav Borysenko[0000-0002-9381-3080], Galyna Kondratenko[0000-0002-8446-5096], Ievgen Sidenko[0000-0001-6496-2469], Yuriy Kondratenko[0000-0001-7736-883X]

Intelligent Information Systems Department, Petro Mohyla Black Sea National University, 68th Desantnykiv Str., 10, Mykolaiv, 54003, Ukraine,
kalambyr11@gmail.com, halyna.kondratenko@chmnu.edu.ua,
ievgen.sidenko@chmnu.edu.ua, yuriy.kondratenko@chmnu.edu.ua

Abstract. In this paper, the subject of the research is the intelligent forecasting in multi-criteria decision-making, in particular, methods of prediction of time series: traditional models of autoregression, smoothing techniques and machine learning methods with the use of artificial neural networks and deep learning. Time series of Amazon share prices from the official sources over the past years serve as a base for exploration. The purpose of the work is to find out the parameters that influence the efficiency and accuracy of models for analysis and forecast the share prices. Such assessment is complicated and vital because various types of criteria, in particular, sales and profitability, market indexes, exchange rates and general trends, etc., usually influence the decision-making process. The methods of research include the consideration and analysis of prediction methods using specific metrics. The relevance of the topic assumes that the accurate prediction at the financial market may contribute to the financial benefits for companies, government and other players on stock exchanges. A comparative analysis of the considered forecasting methods was conducted. It allows choosing the most appropriate intellectual method for increasing the efficiency of a specific share price prediction. The further development of the subject of research includes ensemble-learning methods for neural networks, feature engineering, collecting a more extensive data set for forecasting.

Keywords: time series, autoregression, smoothing, artificial neural networks, convolutional neural networks, recurrent neural networks, decision-making, multi-criteria approach, stock market, share price.

1 Introduction

The analysis of the behavior of stock prices is characterized by the ambiguous behavior of the process, which is usually affected by many factors (trend, seasonality, the geopolitical situation, etc.). Forecasting is a key point when making investment decisions. The ability to predict the behavior of stock for making final decisions allows you to make the best choice which otherwise might be unsuccessful [1].

However, not all investors successfully profit from their investment. This is because the stock price is constantly fluctuating, and at any moment the price may fall below the price at which it was purchased. Therefore, to predict how the financial market will behave is one of the most difficult tasks in the economy. In this prediction, it is necessary to consider various factors, such as physical, psychological, rational and irrational behavior, and the like. All these aspects lead to the conclusion that stock prices are very volatile and it's difficult to predict them with a high degree of accuracy. However, this task is urgent for the world and the whole international economy, since the possibility of accurate prediction of the value of the stock is closely linked to the financial gain of the companies, the government or personal capital and the formation of the more rational financial behavior [2-3].

Accurate predicting the value of assets on the exchange will also help to reduce investment risk and protect the investment income from the market volatility.

2 Related Works and Problem Statement

A stock market (hereinafter the SM) is an organized market where the securities owners make the agreements of purchase and sale with the members of the SM as intermediaries. The prices of these securities are determined by supply and demand, and the process of sale is governed by the rules and regulations [4].

There are three different approaches to research and predict the asset prices on the stock market: technical, technological and fundamental analysis. In this paper, the combination of the first two approaches was used.

Technical analysis is the study of the dynamics of the main indicators of the market with the help of graphical methods to predict the future direction of their movement. A significant number of participants in stock and OTC markets use technical analysis [5]. A successful trader O. Elder has spoken figuratively on this subject: "Technical analysis is related to public opinion polls. It is a combination of science and art. The scientific part consists in using statistical methods and computers; the creative part is the interpretation of the data."

Technological analysis originated in the era of the application of computer technology in business and other fields. The success of its application in solving complex tasks is largely determined by the possibilities of modern information technologies. The result of technological research is usually a selection of well-defined alternatives. Therefore, the origins of this analysis, its methodological concepts lie in those disciplines that deal with decision-making: Theory of Operations and General Management Theory.

The technological analysis includes the ability to analyze, to predict, and to design decision-making in complex systems of various nature, which are based on the time series.

A time series can be represented by four components: *trend, seasonal variations, cyclical variations and irregular factors* [6]. The classical approach to the construction of the time series model is to schedule it into several components, each of which examines specific methods. A trend is a general systematic linear or non-linear component that can change over time. The seasonal and cyclical components are periodically repeating components. These components are not always present simultaneously in the time series. In our case, there are no seasonal and cyclical components.

3 Basic Concepts and Intellectual Forecasting Methods for Stock Prices Prediction

Since a prediction based on statistical data collected with the same interval is required, we are dealing with the time series. Therefore, the time series models will be considered, and those that suit the task of forecasting data will be selected. The most popular forecasting time series models are *autoregressive models, autoregressive models with a moving average and models derived from them*. Thus, we will focus on these models [3].

After selecting the specific methods of solving tasks, it is necessary to directly conduct a prediction of selected characteristics. Constructing the desired models will be implemented by their gradual complication, in other words, we will move from simple to complex. In this paper, models and methods are evaluated using MAE and MSE metrics [1, 7].

MAE measures the average absolute value of errors in a set of predictions for continuous variables. Assuming that $\hat{y}_t$ is the values of the time series forecast in period t, the metric is given by (1):

$$MAE = \sum_t \left| y_t - \hat{y}_t \right| . \tag{1}$$

MAE is a linear score, which means that all individual differences are resolved to the same average. MSE is the difference between the forecast and corresponding observed values squared at each iteration. Because errors rise to the square before they are averaged, MSE has a relatively high weight to large bias. This means that MSE is most useful when large errors are particularly undesirable, which is consistent with the objectives of this work:

$$MSE = \sum_t \left(y_t - \hat{y}_t \right)^2 . \tag{2}$$

Both metrics, MAE and MSE, can take values from 0 to ∞ and do not take the direction of errors into account. The smaller is the value the index takes, the more accurate is the forecast.

Smoothing is an important and widespread method of financial market forecasting. Smoothing methods are used to reduce the influence of random components (random fluctuations) in time series. They provide an opportunity to obtain more "pure" values that consist only of deterministic components. Some of the methods are aimed at recovering some of the components such as trend.

The authors present 5 smoothing methods typically found when forecasting financial data: the simple moving average; the weighted moving average; the exponential moving average; the double exponential moving average; the triple exponential moving average [8].

The basic assumption of these methods is that the fluctuations in past values represent random deviations from a smooth curve, which can be extrapolated to create a forecast.

The autoregressive model is another effective tool for understanding and predicting future values of the time series, which includes devolution of a variable on values of series in the past. The importance of ARMA models lies in their flexibility and in their ability to describe almost all features of the stationary time series. Autoregressive of these models describe how consecutive observations in time affect each other, while parts of the moving averages capture some possible unobserved upheavals, and that allows simulating various phenomena that can be observed in a variety of fields from biology to finance [1-3].

The main idea of autoregression methods is that future values of the time series cannot deviate to higher or lower than the previous values of the time series whatever the reasons that caused those deviations are. The paper has presented such models of autoregression as simple AR (Autoregression Model), ARMA (Autoregressive Moving Average Model) and ARIMA (Autoregressive Integrated Moving Average Model) [1, 8].

We've also proved that Artificial Neural Networks (ANN) have a significant advantage in time series forecasting because they are endowed with the capacity to solve complex problems of forecasting [2, 9, 10].

The output value of the neural network is defined mathematically as:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0i} + \sum_{i=1}^{p} \beta_{ij} y_i \right) + \varepsilon_t \, , \tag{3}$$

where p is the number of the input variables, q is the number of the hidden nodes, α_j and β_{ij} are the weights, ε_t is the random noise.

As a function of g the following functions can be used [2, 11, 14].

Sigmoidal function:

$$f(x) = \frac{1}{1+e^{-x}} \, . \tag{4}$$

In some literature, this is called a logistic function. This nonlinear function is one of the most common activation functions for deep learning [12].

Hyperbolic tangent:

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \, . \tag{5}$$

This function gives the best result for multilayer neural networks compared with the sigmoidal function. However, the function does not solve the problem of the vanishing gradient. The main benefit offered by this function is that it is centered relative to zero, which helps in the process of error backpropagation [11].

Softmax function:

$$f(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}. \tag{6}$$

The Softmax function is another type of activation function used in neural computing. It is used to calculate the probability distribution of a vector of real numbers and gets the value range from 0 to 1.

ReLU function:

$$f(x) = \max(0, x). \tag{7}$$

This function is considered to be the most successful and widely used transfer function. ReLU shows the best productivity in deep learning compared to sigmoid and hyperbolic tangent. ReLU is a nearly linear function and therefore preserves the properties of linear models, which make them easily optimized by methods of gradient descent. The main advantage of using ReLU in calculus is that this function does not require computing the exponent or dividing, therefore it guarantees a more rapid execution [11, 13].

ANN has been and continues to be actively used in the financial markets; one of the main advantages of ANN that make them so popular as harbingers of the market is the natural nonlinearity that allows them to learn nonlinear mapping and correlation of data. ANN also works on data, can be trained in real-time; they are highly adaptive, easy to retrain in the event of market fluctuations and, finally, do well with data that contain a certain number of errors [14-17]. ANN mechanism implies minimum participation of the analyst in the model shaping, as far as the learning ability is characteristic of all neural network models and learning algorithms adapt (adjust) the weights according to the structure of the data presented for training.

Consider the most common optimization techniques.

Gradient descent [18]:

$$x_{t+1} = x_t - \alpha \cdot f'(x_t). \tag{8}$$

According to this method, the steps proportional to the opposite gradient value are taken for minimizing functions. The parameter to this method is a descent speed α. When reaching a large value α will possibly make large steps for finding the minimum, but there is a risk of skipping the lowest point. At a very low speed of learning, the algorithm will move confidently in the direction of the negative gradient, but the implementation, in this case, will take a long time.

The method of stochastic gradient descent [19]. This is a type of gradient descent, which handles 1 element for learning at each iteration. So, parameters are updated even after one iteration, which processed only one value of the variable. This allows optimizing the target function much faster than ordinary gradient descent. But if the number of training data is very large, the number of iterations is therefore sufficiently large.

Mini-batch gradient descent [20]. This is the type of gradient descent, which is faster than batch and stochastic gradient descent. Let there be the m number of input data, then in one iteration of this method $b<m$ elements will be processed. Consequently, even if the number of training data is large, it is processed in smaller training groups at once. Thus, the method works for a large amount of training data by optimizing an objective function with less number of iterations.

Gradient descent with momentum is the method that helps to accelerate a descent in the corresponding direction and dampens the oscillations, approaching the local minimum. This is implemented by adding a part of the update vector from the previous step to the vector of the current update.

Adam (Adaptive Moment Estimation) [21]. Adam optimizer is one of the most popular optimization algorithms of gradient descent since it is computationally efficient and has very little memory requirements. This method calculates an individual adaptive learning rate for each parameter from the evaluation of the first and second moments of the gradients.

In the simulation of isolated time series using ANN models a transformation of the original data for increasing the number of input neurons and, consequently, increasing predictive ability is allowed. Modeling of the time series using the ANN mechanism consists in the formation of ANN as a certain structure, which describes the behavior of the system under study in points in time, and forecasting is the prediction of the future behavior of the system in the background.

To sum it up, this work explains the mathematical foundations of neural networks, namely multilayer perceptron of Rumelhart, convolutional neural networks and recurrent neural networks. The correctly chosen input is significant. In this paper, only the value of the stock price in the past was taken as the initial data.

4 Practical Implementation of the Described Methods and Techniques

This section describes the architecture of the developed program and the final comparative results of all models. Compared to the classical mathematical methods, this work has allowed confirming the viability and the feasibility of using artificial neural networks for further study of their application in the financial markets [1].

For this study, AMAZON stock statistics were taken from the official NASDAQ (National Association of Securities Dealers Automated Quotation) website for the past 5 years. Dataset (Fig. 1) is daily information on the price of shares at the beginning and at the end of the day, the minimum and maximum value of a share during the day, and the number of shares sold. The above parameters can be considered as criteria since they have an optimization direction (maximum or minimum). Moreover, the considered methods and approaches are also used to solve multi-criteria decision-making problems with a more complex structure and the number of criteria. Therefore, this task of forecasting the stock price with the considered methods and approaches can be considered in the context of multi-criteria decision-making.

The moving average method was implemented for various values of the N parameter of the previous time points number that had to be taken into account when building a forecast (Fig. 2). It was checked that decreasing the N window length model shows the more accurate result on the test sample, indicating that the latest data is the most influential to the prediction, i.e. prediction considering the past 5 days is more accurate than the forecast, which takes the last 20 days into account.

	date	close	volume	open	high	low
0	5/10/19	1164.27	1314546	1162.38	1172.600	1142.50
1	5/9/19	1162.38	1185973	1166.27	1169.660	1150.85
2	5/8/19	1166.27	1309514	1174.10	1180.424	1165.74
3	5/7/19	1174.10	1551368	1189.39	1190.440	1161.04
4	5/6/19	1189.39	1563943	1185.40	1190.850	1166.26

Fig. 1. Dataset as input data

Fig. 2. The schedule of the simple moving average forecast

For the exponential moving average (Fig. 3) parameter is the level of the α smoothing coefficient, which represents the degree of reduction of weighting from 0 to 1. The lower the level of smoothing is, the more accurate is the forecast, as every previous reading weighs more [2-3].

This part of the paper also contains a comparison of the results of forecasting (by terms of MAE and MSE) stock prices by various types of neural networks (MLP, CNN, LSTM) with an additional comparison of their models, which have different structures and parameters. Therefore, a detailed review of the architecture of neural networks, justification for choosing the number of hidden layers, the number of neurons in them, is not necessary.

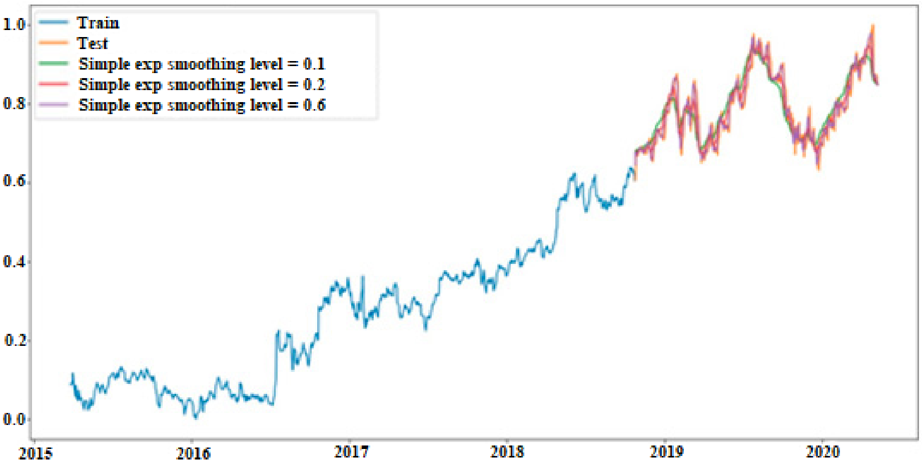


Fig. 3. The schedule of exponential moving average method forecast

For the data with a distinct trend, which corresponds to the input operation, the method of double exponential smoothing that is the recursive application of exponential filter twice, gives the best result. This method has provided an additional parameter β, which is responsible for the smoothing of the trend (Fig. 4). The combination of the α and β pair has adjusted the accuracy and the quality of the forecast [7].

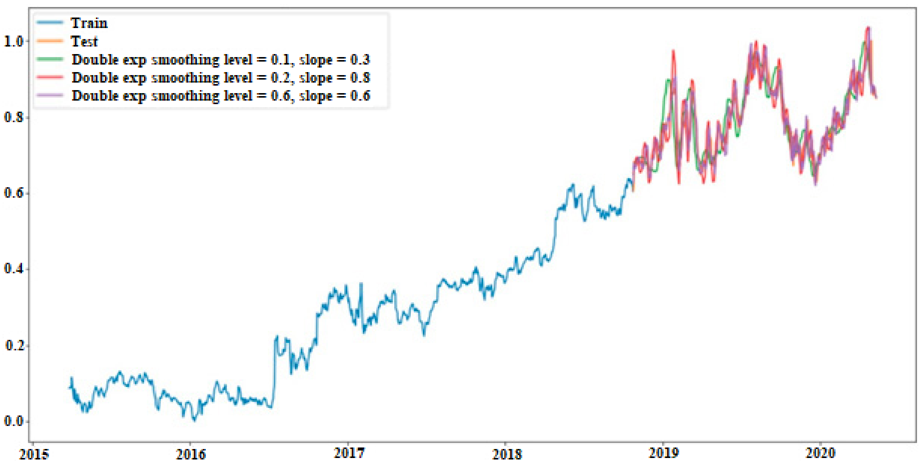


Fig. 4. The schedule of the double exponential moving average method forecast

The results of implementation of all smoothing methods are summarized in Table 1.

The autoregressive model [22] is an effective tool for understanding and predicting future values of the time series, which includes the devolution of a variable on values of series in the past. Autoregressive of these models describe how consecutive observations in time affect each other, while the parts of the moving averages capture some possible unobserved upheavals. For the considered models, the autoregressive Akaike coefficient allowed the software to choose the model that gives the best forecast. Therefore, in the case of the dataset, it was the model ARIMA (2,1,1).

Table 1. The smoothing methods implementation results

Method	Specific parameters	Test MSE	Test MAE
Simple moving average	Window $N = 30$	0.0029	0.0455
	Window $N = 10$	0.0012	0.0281
	Window $N = 5$	0.0005	0.0180
Exponential moving average	The level of smoothing $\alpha = 0.1$	0.0018	0.0340
	The level of smoothing $\alpha = 0.2$	0.0011	0.0259
	The level of smoothing $\alpha = 0.6$	0.00049	0.0165
Double exponential moving average	$\alpha = 0.1$, $\beta = 0.3$	0.00244	0.0388
	$\alpha = 0.2$, $\beta = 0.8$	0.00163	0.0322
	$\alpha = 0.6$, $\beta = 0.6$	0.00054	0.0177

The execution results of all three models were summarized in a single comparative Table 2.

Table 2. The autoregressive models implementation results

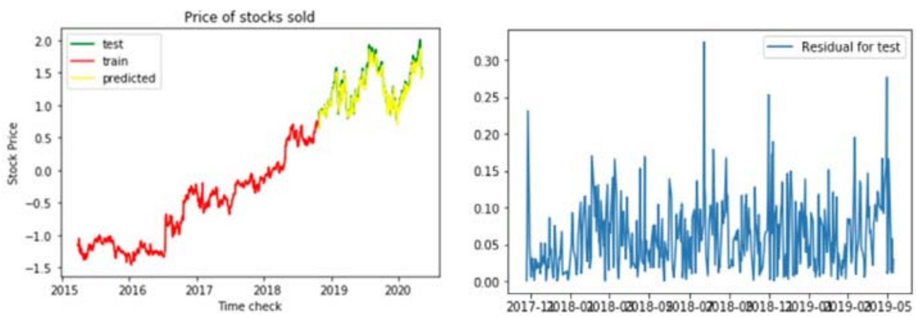
Model	Specific parameters	Test MSE	Test MAE
Simple autoregressive model	$p = 2$	0.000347	0.013322
Autoregressive moving average model	$p = 2$, $q = 1$	0.000348	0.013346
Autoregressive integrated moving medium model	$p = 2$, $q = 1$, $d = 1$	0.000345	0.013335

The simplest type of the neural network is a single-layer perceptron network (usually a multilayer perceptron of Rumelhart), which consists of a single layer of output nodes, and the inputs provided directly to the outputs via a series of weights. Universal approximation theorem for neural networks states that each continuous function, which maps intervals of real numbers to a certain initial interval of real numbers, can be arbitrarily closely approximated by a multilayer perceptron with only one hidden layer [2, 12]. The main characteristic of the multilayer perceptron is its architecture, namely the number of hidden layers, nodes and activation functions. Besides, since the study outcome partially depends on the initialization of variables, each model was trained separately 5 times, and all the characteristics for a comparative table are average values. For a more in-depth study of the model 5 architectures were implemented in the paper, the accuracy of each of them is represented in Table 3.

According to the comparative results (Table 3), the dependence of the quality of the forecast on the number of nodes in the hidden layer, the optimization method and the activation function is completely traced. The best result was the MLP 20-100-1 model with the Adam optimization algorithm and hyperbolic tangent as an activation function. Here is a graph of the price of stocks sold predicted by this model (Fig. 5, a), as well as a graph of the forecast error (Fig. 5, b).

Table 3. MLP learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
MLP 20-10-1	221	0.0645	0.0074	0.090	0.014	adam	relu
MLP 20-20-1	441	0.0462	0.0038	0.052	0.005	adam	relu
MLP 20-100-1	2201	0.0438	0.0038	0.037	0.002	adam	relu
MLP 20-100-1	2201	0.0035	0.0428	0.032	0.002	adam	tanh
MLP 20-100-1	2201	0.0668	0.0078	0.093	0.014	sgd	relu



a) b)

Fig. 5. MLP 20-100-1: a) price of stocks sold; b) forecast error

Convolutional neural network [23] is a type of artificial neural network, in which the pattern of connectivity between neurons is inspired by the organization of the visual cortex of animals, individual neurons that are arranged in such a way that they respond to overlapping regions of the field. The main advantage of convolutional neural networks is that we use the convolutional layers to identify the signs of a network that allows training of a neural network without complex pre-processing since useful features are learned during training [13]. Unlike multilayer perceptrons, convolutional neural networks have a complex structure, since they need to gauge the number and type of layers, the number of nodes in each layer, the optimization method, the activation function and the number of filters. So CNN-20-200-3 will designate 20 knots at the entrance as the first layer, 200 filters sized 3x3, MaxPooling layer and 2 fully-connected layers. In this work, 5 different CNN architectures were tested, the results are shown in Table 4.

The CNN-10-256-3 model with the Adam optimization algorithm and ReLU activation function showed the best result.

A recurrent neural network (RNN) is an artificial neural network, the neurons of which transmit feedback signals to each other [24]. The idea of RNN is to use consistent information. In traditional neural networks, we assume that all the inputs (and the outputs) are independent of each other. But for many tasks, this is not the best idea. This type of neural network is well suited to predict the value of the stock, as further steps may depend on the past.

Table 4. CNN learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
CNN 10-256-3	198.657	0.0238	0.0011	0.096	0.013	adam	relu
CNN 20-256-5	201.367	0.0182	0.0074	0.174	0.038	adam	relu
CNN 20-256-3	199.937	0.0513	0.0050	0.290	0.110	sgd	relu
CNN 20-500-3	756.501	0.0515	0.0050	0.259	0.089	sgd	relu
CNN 20-300-3	273.901	0.0414	0.0030	0.255	0.079	sgd	tanh

For the implementation of the forecasting recurrent neural networks, the LSTM model (Long Short-Term Memory Units) was chosen [25]. LSTMs help to save the bug, which can be spread through time and layers. Maintaining a more constant error, they allow the networks to continue exploring for many time steps [16]. To analyze the effect of parameters on the prediction value of the shares 5 variants of the LSTM architecture summarized in the following Table 5 were implemented. LSTM-20-30-30-30-1 was used to identify neural LSTM networks with initial data of the previous 20 days, the number of nodes 30 on the first, the second and the third layers of LSTM and one fully connected layer with the resulting output of one element.

Table 5. LSTM learning results

Model	Parameters number	Training MAE	Training MSE	Test MAE	Test MSE	Algorithm	Activation function
LSTM 20-30-30-30-1	18,631	0.0074	1.0564e-04	0.0198	0.0005	adam	relu
LSTM 20-30-30-40-1	22,681	0.0079	1.8716e-04	0.0188	0.0005	adam	relu
LSTM 30-30-30-30-1	18,631	0.0106	1.9467e-04	0.0159	0.0004	adam	relu
LSTM 10-30-30-30-1	18,631	0.0099	1.7086e-04	0.0150	0.0003	adam	relu
LSTM 10-30-30-30-1	18,631	0.0082	1.2895e-04	0.0152	0.0003	sgd	relu

The best result was shown by the LSTM model 10-30-30-30-1 with the Adam optimization algorithm and ReLU activation function, and with the least lag window.

The performance of classical algorithms and machine-learning algorithms in section 4 can be summarized in a comparative Table 6.

Table 6. The comparative results of the implemented methods

Model/Method	Advantages	Disadvantages
Smoothing	Ability to process trends of varying levels and seasonality components	Vulnerability to extreme values
Autoregression	Easy to automation	Limitations in the assumptions
ANN	Ability to handle complex nonlinear patterns. High forecast accuracy	Require a large amount of data

5 Conclusions

For investors, capital management is becoming increasingly important today. Professional investment managers and individual investors tend to have effective tools for understanding the trends in the stock market, to minimize investment risk and increase their profits. Some people believe that it is very difficult to predict stock prices. However, in the real business world, successful traders carry out thousands of transactions every year.

Since the very beginning of the financial operations on the stock exchange, people developed methods to predict the value of assets in the future. With advances in computing technology and artificial intelligence, the accuracy of the methods is growing every day.

In this paper, the authors compared the indicators of forecasting between the neural network and the classic time series forecast method, namely, the value of the company's shares on the stock exchange. The analysis showed that the neural network models described in this study showed a very much better able to accurately predict, and therefore confirmed the viability and the feasibility of using artificial neural networks for further study of their application in the financial markets. The predictive power of the model was influenced by different factors, depending on the architecture and input data.

References

1. Kingdon, J.: Intelligent Systems and Financial Forecasting. Springer, London (1997).
2. Palit, A., Popovic, D.: Computational Intelligence in Time Series Forecasting. Theory and Engineering Applications. Springer, London (2005). doi: 10.1007/1-84628-184-9
3. Guerard, J.: Introduction to Financial Forecasting in Investment Analysis. Springer, New York (2013). doi: 10.1007/978-1-4614-5239-3
4. Pavlov, V., Pilipenko, I., Krivov'yazyuk, I.: Securities in Ukraine: Students Book. Condor, Kyiv (2004). (in Ukrainian)
5. Sokhatska, O.: Stock Exchange. Carte Blanche, Ternopil (2003). (in Ukrainian)

6. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis: Forecasting & Control. John Wiley & Sons, New Jersey (2008)
7. Wei, W.: Time Series Analysis: Univariate and Multivariate Methods. Pearson, New York (2005)
8. Sitte, R., Sitte, J.: Neural Network Systems Technology in the Analysis if Financial Time Series. In: Leondes, C.T. (eds.) Intelligent Knowledge-Based Systems, pp. 1564–1615. Springer, Boston, MA (2005). doi: 10.1007/978-1-4020-7829-3_43
9. Kondratenko, Y., Gordienko, E.: Neural Networks for Adaptive Control System of Caterpillar Turn. In: Annals of DAAAM for 2011 & Proceeding of the 22th Int. DAAAM Symp. "Intelligent Manufacturing and Automation", pp. 0305–0306 (2011)
10. Gomolka, Z., Dudek-Dyduch, E., Kondratenko, Y.P.: From homogeneous network to neural nets with fractional derivative mechanism. In: Rutkowski, L. et al. (eds.) International Conference on Artificial Intelligence and Soft Computing (ICAISC-2017), Part I, pp. 52–63, Zakopane, Poland (2017)
11. Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S.: Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. arXiv:1811.03378 [cs.LG] (2018)
12. Robert, J.: The Application of Neural Networks in the Forecasting of Share Prices. Finance and Technology Publishing, New York (1996)
13. Di Persio, L.: Artificial Neural Networks architectures for stock price prediction: comparisons and applications. International journal of circuits, systems and signal processing **31(10)**, 404-405 (2016)
14. Kushneryk, P., Kondratenko, Y., Sidenko, I.: Intelligent dialogue system based on deep learning technology. In: 15th International Conference on ICT in Education, Research, and Industrial Applications: PhD Symposium (ICTERI 2019: PhD Symposium), vol. 2403, pp. 53–62, Kherson, Ukraine (2019)
15. Sidenko, I., Filina, K., Kondratenko, G., Chabanovskyi, D., Kondratenko, Y.: Eye-tracking technology for the analysis of dynamic data. In: IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), pp. 479–484, Kiev, Ukraine (2018). doi: 10.1109/DESSERT.2018.8409181
16. Cryer, J., Chan K.-S.: Time Series Analysis. With Applications in R. Springer, New York (2008). doi: 10.1007/978-0-387-75959-3
17. Beran, J.: Mathematical Foundations of Time Series Analysis. A Concise Introduction. Springer, Cham (2017). doi: 10.1007/978-3-319-74380-6
18. Paper, D.: Gradient Descent. In: Data Science Fundamentals for Python and MongoDB, pp. 97–128. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3597-3_4
19. Ketkar, N.: Stochastic Gradient Descent. In: Deep Learning with Python, pp. 113–132. Apress, Berkeley, CA (2017). doi: 10.1007/978-1-4842-2766-4_8
20. Danner, G., Jelasity, M.: Fully Distributed Privacy Preserving Mini-batch Gradient Descent Learning. In: Bessani, A., Bouchenak, S. (eds.) Distributed Applications and Interoperable Systems. DAIS 2015. Lecture Notes in Computer Science, vol. 9038, pp. 30–44. Springer, Cham (2015). doi: 10.1007/978-3-319-19129-4_3
21. Jiang, Y., Han, F.: A Hybrid Algorithm of Adaptive Particle Swarm Optimization Based on Adaptive Moment Estimation Method. In: Huang, D., Bevilacqua, V., Premaratne, P., Gupta, P. (eds.) Intelligent Computing Theories and Application. ICIC 2017. Lecture Notes in Computer Science, vol. 10361, pp. 658–667. Springer, Cham (2017)
22. Olson, D., Wu, D.: Autoregressive Models. In: Predictive Data Mining Models. Computational Risk Management, pp. 55–69. Springer, Singapore (2017)
23. Ayyadevara, V.: Convolutional Neural Network. In: Pro Machine Learning Algorithms, pp. 179–215. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3564-5_9

24. Hammer, B.: Learning with recurrent neural networks. Springer, London (2000)
25. Manaswi, N.: RNN and LSTM. In: Deep Learning with Applications Using Python, pp. 115–126. Apress, Berkeley, CA (2018). doi: 10.1007/978-1-4842-3516-4_9