

D.S.¹ Shybaiev
T.V.¹ Otradsкая, Ph. D.
N.O.² Shybaieva, Ph. D.
N.D.² Rudnichenko, Ph. D.
M.V.¹ Stepanchuk

*Odessa college of computer technologies «Server», Odessa, Ukraine¹
Odessa National Polytechnic University, Odessa, Ukraine²*

Predicting system for the estimated cost of real estate objects development using neural networks

Automation of user workflows is an integral part of the development of modern information and software systems. Many specialists in various subject areas perform most of their daily tasks using computer technology. This allows you to improve communication, automate part of workflows, provide unified data storage.

The purpose of the work is to develop mobile software to predict the estimated value of real estate based on the use of artificial neural networks.

The scientific and practical significance of the work is to study the algorithms of artificial neural networks in the field of forecasting and the formation of estimated value of real estate objects and the practical application of such algorithms in the concept of mobile applications.

The general work methodology includes the use of a multi-layer perceptron architecture and an error back-propagation algorithm as a method for teaching a neural network, as well as methods and skills for designing and developing mobile software solutions for automating realtor work activities.

The peculiarity of the developed software solution is the system of forecasting the optimal value of the price of a real estate object. Interaction of software components is carried out by using search criteria and real estate database. The values obtained are input data for the operation of the artificial neural network model.

The results of the study provide a high level of final accuracy of calculations in solving the problems of forecasting the estimated value of real estate using a multilayer artificial neural network.

The result of the work is the development of a mobile application for predicting the financial assessment of real estate. The average accuracy of the forecast value of the property is 95.7%, which is a high value for the use of the system in practice.

Keywords: *data mining, artificial neural networks, classification, evaluation of real estate, data analysis.*

Introduction. Automation of user workflows is an integral part of the development of modern information and software systems. Many specialists in various subject areas perform most of their daily tasks using computer technology. This allows us to improve communication, automate part of workflows and provide unified data storage. The sphere of real estate trading is no exception and allows us to use information technology to solve various applied problems. Modern real estate business involves the use of distributed information necessary for the sale or purchase of real estate, their lease and other actions carried out in accordance with the requirements of the target client. Realtor's main tools are specialized trading platforms which provides ads placement, searching and selecting processes, graphic files viewing, communication processes with the seller or buyer support using the built-in messaging tools.

One of the realtor's activities is to assist in the assessment of the property for the target client, which will correspond to the market value. Formation of the property value includes a large set of input parameters which the pricing for each specific object of evaluation depends on. An important fact is the external independent assessment by a specialist, who can make his preferences regarding the object of evaluation. However, the compilation of the actual price at the time of the sale or purchase of the property is a complex computational task with a large number of input data sets, the analysis of which requires considerable forces, resources and time [1-3]. Automating such a process will improve the dynamics of realtors, optimize the time spent on conducting real estate evaluation procedures and ensure greater accuracy in prices determining for each property.

The aim of the work is to develop and study the mobile software for predicting the estimated value of real estate based on the use of the artificial neural networks approach.

Problem description. Analysis of existing literature sources for solving problems of forecasting and regression in areas related to real estate and other related areas using ANN [1-3] revealed the absence of existing applied software solutions and ANN models that allow taking into account the realtors real estate activity specifics in optimal apartment options selection. The advantages of artificial neural networks (ANN), indicating its practical potential, is to support the capabilities of parallel processing of information, which significantly speeds up the process of computational operations. With a large number of interneuron connections, the ANN model acquires resistance to learning errors that may occur in certain parts of its structure. The use of neural networks in modern software, allows us to automate the processes of solving problems of forecasting, classification or data retrieval [4, 5].

Currently, ANN capabilities are often used in practice for solving forecasting problems. In this regard, the proposed approach for information system development, which implements the INS model, can serve as a base for selecting the optimal values when evaluating real estate objects [6-8].

Solving the problem of estimating the property value by manual methods, which include analyzing the real estate market, currency markets and taking into account many factors related to the pricing of real estate, is an extremely complex and computationally expensive process. Automating such a process will significantly improve the efficiency of realtors, increase the accuracy of pricing for a property and search for a similar group of objects according to the established list of criteria.

A sample of data for analyzing key market value indicators for a property is performed by processing free databases to search for similar offers (fig.1). This task was performed by developed automated data parser usage, which can collect information on all real estate objects, according to a predetermined range. The data samples are based on the use of multi-criteria conditions created by the system's user, ensuring the search for the best matches between the existing objects in the real estate market and the optimal search result of the user.

The next step is a statistical test for the data obtained effectiveness, which is ensured through continuous training of the ANN model. Based on the fact that in most situations, the problem of forecasting is a special case of the regression problem, it is advisable to carry out its solution on the basis of such architectures: multilayer perceptron, radial-basic network, Volterra network, Elman network.

In connection with the frequent and effective use of the ANN multilayer perceptron architecture in the practice of solving problems of classification and regression [9], it is possible to construct a model based on it.

When solving the problem of real estate appraisal, a multilayer perceptron architecture was chosen. The criteria that were formed for the implementation of training processes and the use of the ANN model are: area, infrastructure development, age of the house, layout, number of floors, condition of the room, floor of living space, type of house, number of rooms, area of real estate, house condition.

Each criterion is used to determine the optimal variant of the user's search query and to build the final sample for all the available options that are suitable for the condition. In order to create a neural network model, each criterion corresponds to an input neuron that is part of the ANN.

The formed ANN model is unable to process such descriptions properly, resulting in the need to improve the quality of the final data selection, which is a time consuming process.

There are 11 neurons on the first hidden layer, the second and the third layers have 15 and 25 neurons accordingly. 5800 objects from the learning sample are assigned for training. The values of the layers number and hidden neurons are determined empirically to ensure the ANN correct learning process.

To evaluate the performance of the ANN model, the least square method is used to find an approximate solution of an excess-level system. To evaluate the performance of the ANN model, the approximate solution method is used.

This practical method is applicable when solving systems of linear equations. In particular, the expediency of its use in our case lies in the possibility of estimating parameters in a linear regression.

The neuron activation function must be delimited to use the back propagation method.

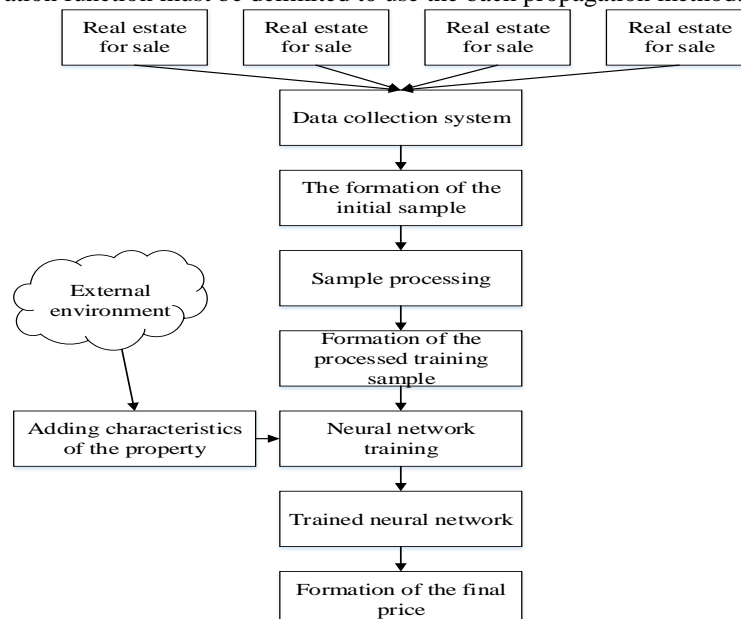


Fig. 1. The scheme of the neural network sampling

To implement developed ANN model we'd used the backpropagation algorithm. The network has a set of inputs $x_{\{1\}}, \dots, x_{\{n\}}$ and target output price variable and many internal nodes. Let $w_{\{i, j\}}$ be the weight of the connection connecting the i -th and j -th nodes, and $o_{\{i\}}$ - the output of the i -th node. According to the selected test data set, it is possible to form the initial error function obtained by the least squares method, which has the form:

$$E(\{w_{i,j}\}) = \frac{1}{2} * \sum_{k \in \text{Outputs}} (t_k - o_k)^2 \quad (1)$$

As design methods, use-case means are applied. In order to formalize the logic of developing a future software system, a diagram of use cases has been formed (fig. 2).

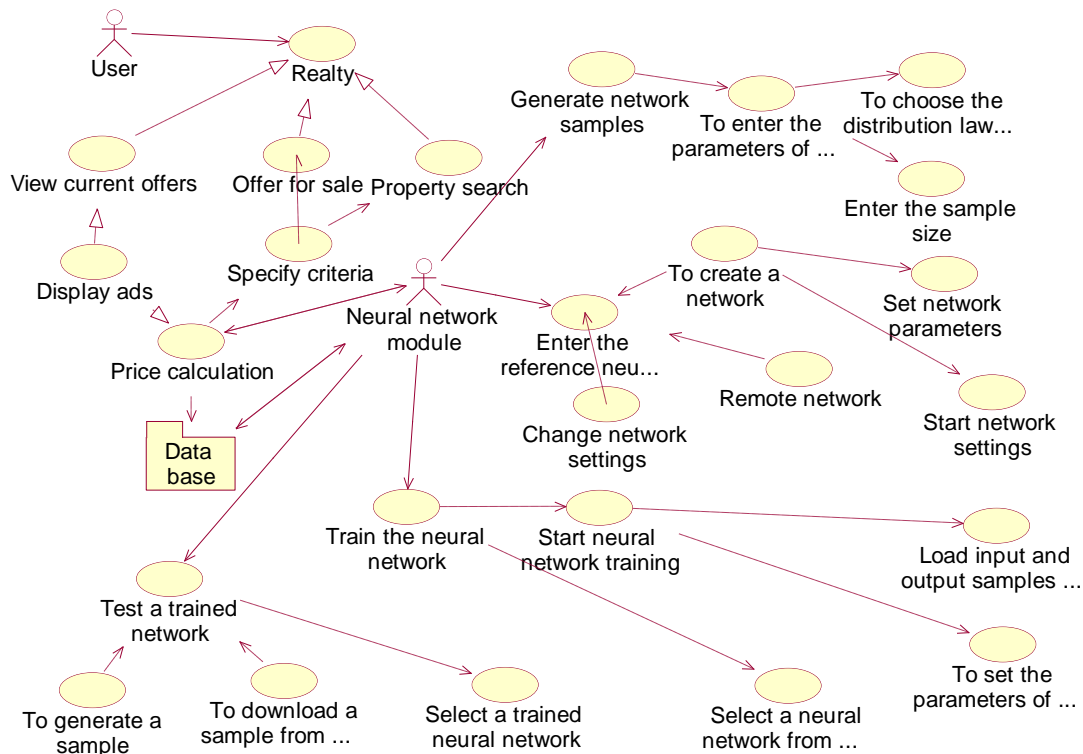


Fig. 2. Software Solution Use Case

A key feature of the developed software solution is a component for predicting the optimal value of the property price. The interaction of software components is performed by using search criteria and a real estate database. The values obtained are the input data for the operation of the ANN model.

The developed software solution has a client-server architecture and allows processing a large number of user requests on a centralized server. The target platform is the Android operating system, the software implementation was performed using Java, Android SDK, the Google TensorFlow library to implement machine learning functions, and the Google Firebase database was selected to store data. The client side is designed to set the values of the input criteria, generate and send a set of input data for ANN learning at the server part and display the results obtained. The server part responsible for supporting the construction of the ANN is implemented on the basis of the python language using Google TensorFlow API. The server's backend based on serverless approach using Firebase. In order to form the logic of user behavior when working with a software solution, a diagram of a sequence of actions was drawn up, which is shown in Fig. 3.

The developed mobile application is highly loaded and allows processing a large number of user transactions passing through the created INS model.

After completing the design stage of a software application, its development has been completed with subsequent correction of errors that have occurred, and functional components have been reworked to increase the efficiency of the application on different versions of mobile devices and running different Android operating system versions.

The software interface is designed, two main activities are shown in Fig. 4.

Each user has a personal account that allows him to save the selected results, as well as view the statistics of search queries for a period of time. This improves the quality of work of realtors, provides a convenient system for viewing data and the ability to predict the real estate value.

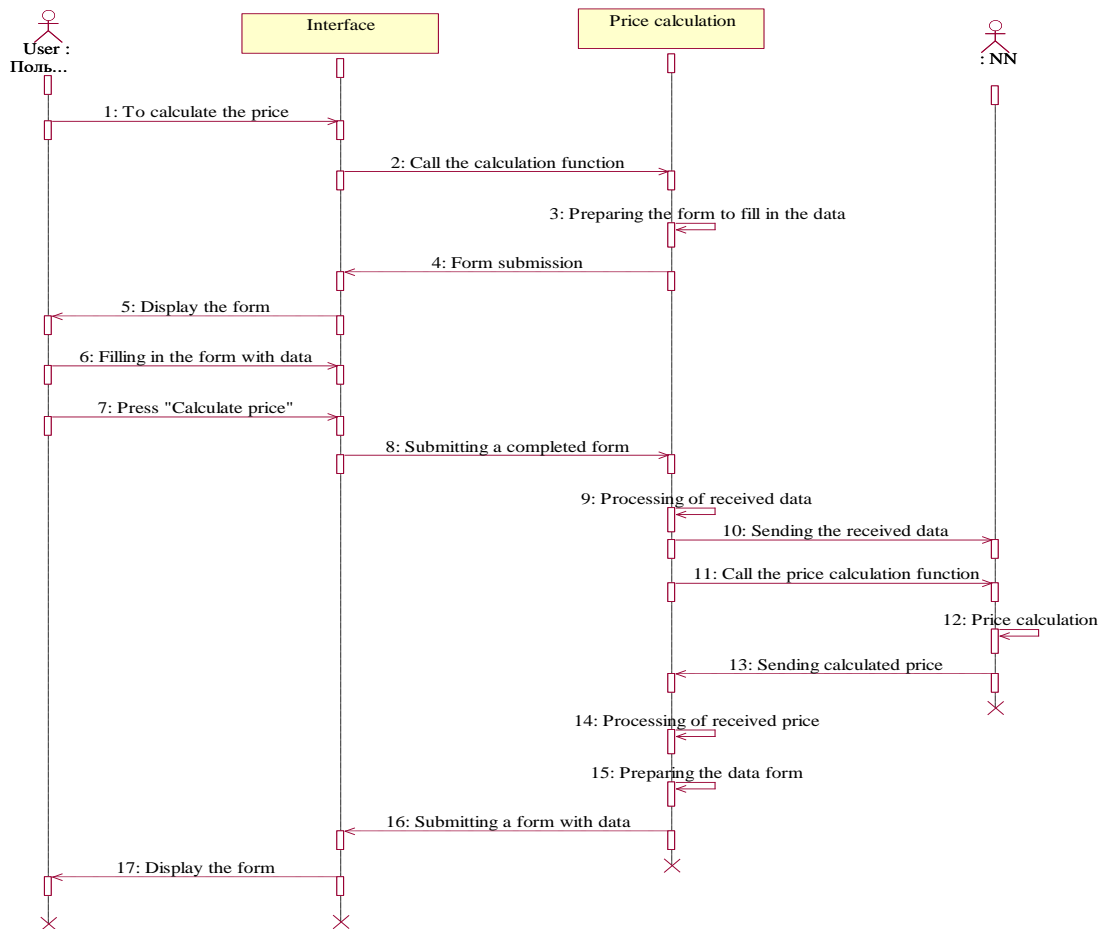


Fig. 3. User actions sequence diagram in real estate valuation

ANN work analysis. Studing the ANN work as part of a mobile software solution is a complex task that requires the use of modern solutions that will improve the accuracy of the final forecasting, improve the data sampling. One of the input data normalization methods used in the system is batch normalization (packet normalization).

Batch normalization provides the ability to normalize the input data in order to obtain a zero expectation and unit variance. Normalization is performed before entering each layer. This method allows solving the problem of distortion due to expectation and dispersion of the ANN signal, even after the normalization of such a signal at the model inputs. This can lead to inconsistencies between gradients at different levels of the ANN.

For the selected batches, the expectation and variance were calculated according to:

$$\mu_{\beta} = \frac{1}{m} \sum_{i=1}^m x_i \tag{2}$$

$$\sigma_{\beta}^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\beta})^2 \tag{3}$$

Using the obtained statistical characteristics, the activation function was transformed to bring it to zero expectation and zero dispersion in the entire batch.

$$\hat{x}_l = \frac{x_l - \mu_{\beta}}{\sqrt{\sigma_{\beta}^2 + \epsilon}} \tag{4}$$

where $\epsilon > 0$ is a parameter performing the division test by 0. To obtain the final value of the activation function y , it is necessary to test for the ability of generalization, thereby obtaining the final activation function:

$$y_i = \gamma \hat{x}_l + \beta \tag{5}$$

where β and γ are parameters of the batch normalization with which the system is trained.

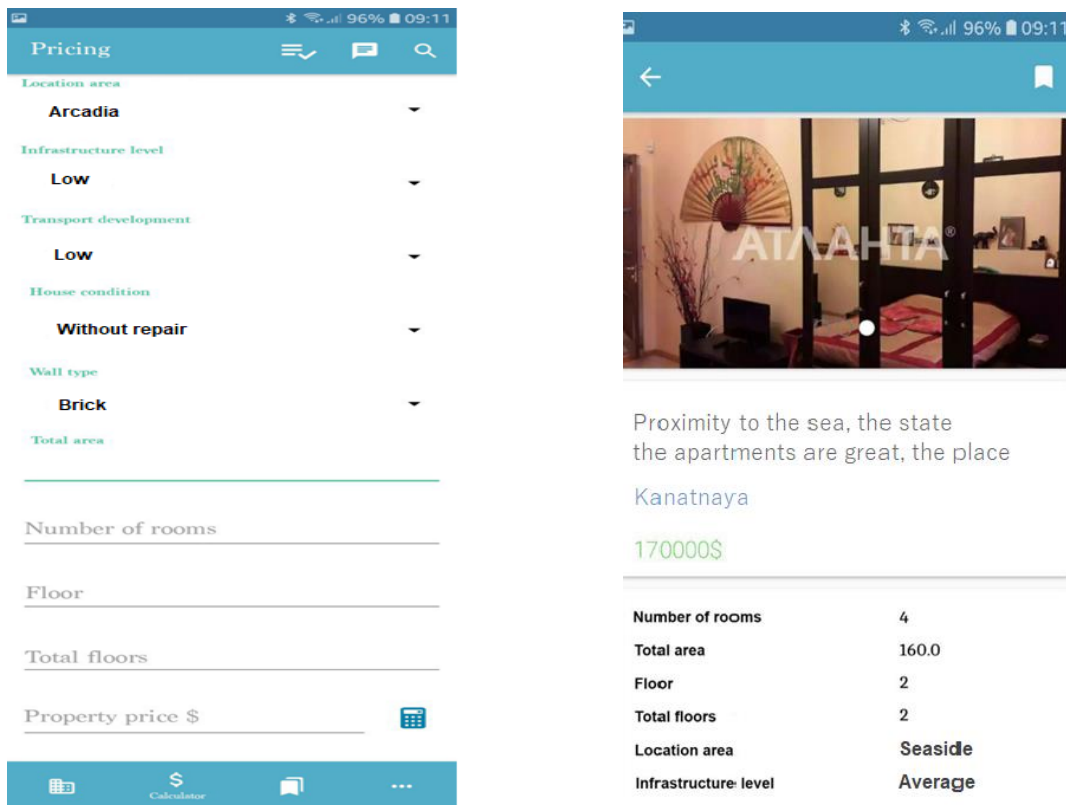


Fig. 4. Interface windows of a developed mobile application

The minimization of the neural network training mean square error is carried out on the basis of the least squares method application. The splitting is performed according to the batch, as the main type of ANN training periods formation. Dependency between batch means square error value and batch number is shown in Fig. 5.

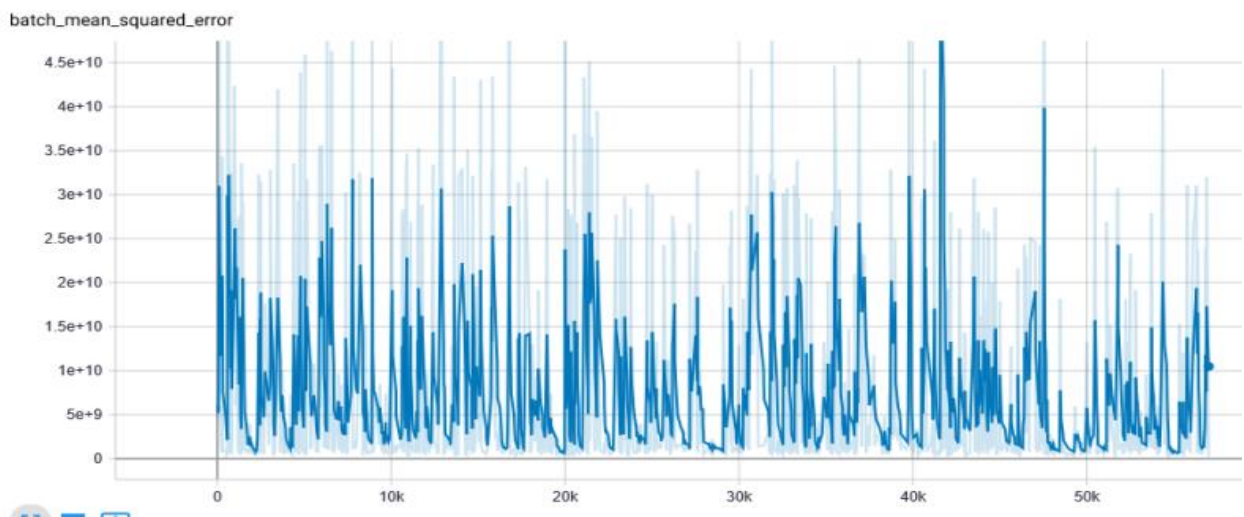


Fig. 5. Batch mean square error

After identifying the errors for each batch separately, it becomes possible to form a chart of errors for all epochs (Fig. 6), which went through the constructed model in both directions during the entire ANN cycle.

The result of the obtained results analysis is to identify the total error of the system for predicting the value of real estate. The prediction accuracy is 95.7%, which is a high accuracy parameter given the complex criteria that affect pricing.

In order to provide experimental research estimates of the various apartments cost were carried out, results are shown in table 1.

Table 1

The results of experiments on the use of the developed software application

Sample size	Flat Type	Average cost, dollars	NN estimated cost, dollars
150	One room	24 000	20 550 – 21 520
250	Two rooms	32 000	31 920 – 32 750
200	Three rooms	40500	40 250 – 43 125
100	Four rooms	52 000	50 000 – 53 500

The results of the actual data samples studying show the variation in the predicted values of apartment prices are not significant, within 6-8 percent, which testify to the adequacy of the created model work.

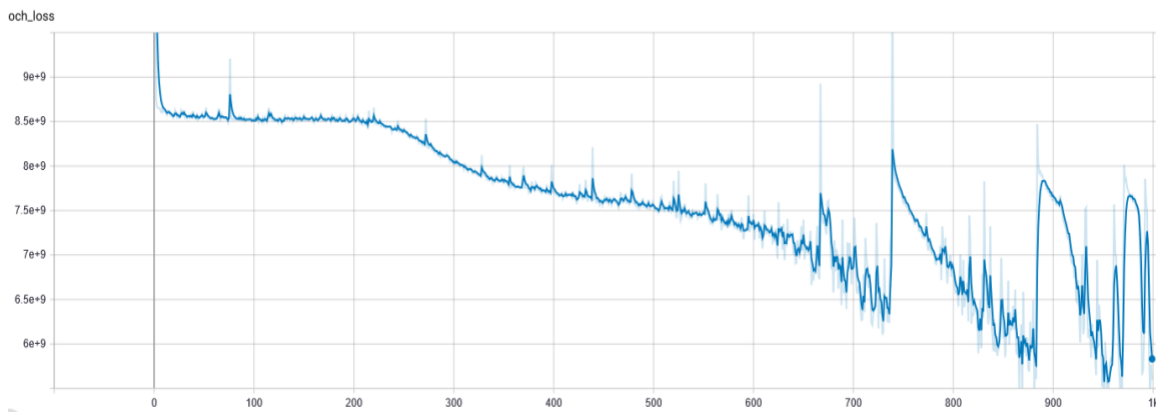


Fig. 6. The schedule of error by epoch

Application of the developed ANN model as the main algorithmic tool capable of analyzing large amounts of information and selecting the optimal search results will further improve the accuracy of the resulting estimates.

Conclusion. The result of the work is the development of a mobile application for predicting the financial assessment of real estate using ANN. The mobile application is designed for use by realtors for the selection and assessment of real estate and for market ordinary consumers. As a data processing and decision-making system, the created ANN model is used, which functions on the basis of the Rumelhart multilayer perceptron algorithm application, the method of training a neural network based on the back propagation error algorithm. In the future, it is possible to add a number of regularization functions to the model in order to improve its accuracy.

Список використаної літератури:

1. Оптимизация отбора и анализа информации в разноструктурных хранилищах данных / Д.С. Шибает, В.В. Вычужанин, Н.О. Шибаета, Н.Д. Рудниченко // Информатика и математические методы в моделировании. – 2017. – № 3. – С. 318–324.
2. Оптимизация методов прогнозирования, обработки и анализа информации в разноструктурных хранилищах данных / Д.С. Шибает, В.В. Вычужанин, Н.О. Шибаета, Н.Д. Рудниченко // Информатика и математические методы в моделировании. – 2018. – № 1. – С. 78–85.
3. Рудниченко Н.Д. Разработка модели нейросети для прогнозирования риска отказов компонентов сложных технических систем / Н.Д. Рудниченко, В.В. Вычужанин // Информатика и математические методы в моделировании. – 2016. – № 4. – С. 333–338.
4. Michael S. Gashler Ashmore Training Deep Fourier Neural Networks To Fit Time-Series Data / S.Gashler Michael, C.Stephen // International Conference on Intelligent Computing. – Cham : Springer, 2014. – С. 48–55.
5. Горбань А.Н. Обучение нейронных сетей / А.Н. Горбань. – М. : СП ПараГраф, 1990.
6. Осовский С. Нейронные сети для обработки информации / С.Осовский ; пер. с польского И.Д. Рудинского. – М. : Финансы и статистика, 2002. – 344 с.
7. Горбань А.Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей / А.Н. Горбань // Сибирский журнал вычислительной математики. – 1998. –Т. 1, № 1. – С. 12–24.
8. Розенблатт Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга / Ф.Розенблатт. – М. : Мир, 1965. – 480 с.
9. Шадрина В.В. Применение методов прогнозирования в технических системах / В.В. Шадрина, Е.Ю. Косенко // Известия Южного федерального университета. Технические науки. – 2011. – № 2. – Т. 115. – С. 141–145.

References:

1. Shibaev, D.S., Vychuzhanin, V.V., Shibaeva, N.O. and Rudnichenko, N.D. (2017), «Optimizatsiya otbora i analiza informatsii v raznostrukturnykh khranilishchakh dannykh», *Informatika i matematicheskie metody v modelirovanii*, No. 3, pp. 318–324.
2. Shibaev, D.S., Vychuzhanin, V.V., Shibaeva, N.O. and Rudnichenko, N.D. (2018), «Optimizatsiya metodov prognozirovaniya, obrabotki i analiza informatsii v raznostrukturnykh khranilishchakh dannykh», *Informatika i matematicheskie metody v modelirovanii*, No. 1, pp. 78–85.
3. Rudnichenko, N.D. and Vychuzhanin, V.V. (2016), «Razrabotka modeli neuroseti dlya prognozirovaniya riska otkazov komponentov slozhnykh tekhnicheskikh sistem», *Informatika i matematicheskie metody v modelirovanii*, No. 4, pp. 333–338.
4. Michael S. Gashler Ashmore and Stephen, C. (2014), «Training Deep Fourier Neural Networks To Fit Time-Series Data», *International Conference on Intelligent Computing*, Springer, Cham, pp. 48–55.
5. Gorban', A.N. (1990), *Obuchenie neironnykh setei*, SP ParaGraf, M.
6. Osovskii, S. (2002), *Neironnye seti dlya obrabotki informatsii*, per. s pol'skogo Rudinskogo I.D., Finansy i statistika, M., 344 p.
7. Gorban', A.N. (1998), «Obobshchennaya approksimatsionnaya teorema i vychislitel'nye vozmozhnosti neironnykh seteĭ», *Sibirskii zhurnal vychislitel'noi matematiki*, Vol. 1, No. 1, pp. 12–24.
8. Rozenblatt, F. (1965), *Printsipy neirodinamiki: Pertseptrony i teoriya mekhanizmov mozga*, Mir, M., 480 p.
9. Shadrina, V.V. and Kosenko, E.Yu. (2011), «Primenenie metodov prognozirovaniya v tekhnicheskikh sistemakh», *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki*, No. 2, Vol. 115, pp. 141–145.

Шибасв Денис Сергійович – викладач Одеського коледжу комп'ютерних технологій «Севрер».

Отрадська Тетяна Василівна – кандидат технічних наук, директор Одеського коледжу комп'ютерних технологій «Севрер».

Шибасва Наталя Олегівна – кандидат технічних наук, доцент кафедри Інформаційні технології Одеського національного політехнічного університету.

Рудніченко Микола Дмитрович – кандидат технічних наук, доцент кафедри Інформаційні технології Одеського національного політехнічного університету.

Степанчук Михайло Васильович – студент Одеського коледжу комп'ютерних технологій «Севрер».

Стаття надійшла до редакції 18.04.2019.