УДК 004.942 http://doi.org/10.15407/pp2024.02-03.367

G.Ch. Nabibayova

APPLICATION OF NEURAL NETWORKS IN OLAP SYSTEMS

The article highlights the main characteristics of OLAP systems that perform online analytical data processing. These systems, based on OLAP technology, are widely used both in government agencies and in private ones. The main characteristics, features and structure of OLAP systems are mentioned. The article emphasizes that OLAP is a data warehousing tool. OLAP allows analysts to explore and navigate a multidimensional structure of indicators called a data cube or OLAP cube. Indicators (measures) of OLAP cubes play an important role in the decision-making process. To solve some problems, these measures often need to be classified or clustered. Moreover, empty measures are common in OLAP cubes. Empty measures can present due to non-existing facts in data warehouse or due to empty cells which are unfilled in by mistake. The presence of empty measures negatively impacts strategic decision making. Unfortunately, OLAP itself is poorly adapted for forecasting empty measures of data cubes. Over the years, researchers and analysts have tried to improve the decision-making process in OLAP systems and add forecasting and other options to OLAP applications. Today, in the era of Industry 4.0, with the availability of big data, there is a need to apply new technologies to solve such problems. These technologies include neural networks. The article examines the problem of integrating OLAP and a neural network. In this regard, the article provides information about neural networks: information about their properties, types, as well as their capabilities. The article shows the possibility and advantages of integrating OLAP and neural network. It mentions that in the case of big data, the integration of OLAP and neural networks is very effective for solving problems of classification, clustering and prediction of empty measures of OLAP cubes. An architectural and technological model for integrating OLAP and neural networks is presented. It is noted what types of neural networks can be used to solve the problems of classification, clustering and forecasting specified in the model.

Keywords – data warehouse, OLAP, artificial intelligence, machine learning, neural network, forecasting, clustering, classification.

Г.Ч. Набібаєва

ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ В OLAP-СИСТЕМАХ

У статті висвітлено основні характеристики OLAP-систем, що здійснюють оперативну аналітичну обробку даних. Ці системи, засновані на технології OLAP, широко використовуються як в державних установах, так і в приватних. Наведено основні характеристики, особливості та структуру OLAP-систем. У статті підкреслюється, що OLAP є інструментом сховища даних. OLAP дозволяє аналітикам досліджувати та орієнтуватися в багатовимірній структурі індикаторів, яка називається кубом даних або OLAP-кубом. Індикатори (заходи) кубів OLAP відіграють важливу роль у процесі прийняття рішень. Щоб вирішити деякі проблеми, ці заходи часто потрібно класифікувати або кластеризувати. Крім того, порожні міри часто зустрічаються в кубах OLAP. Порожні показники можуть бути через неіснуючі факти в сховищі даних або через порожні клітинки, які помилково не заповнені. Наявність порожніх заходів негативно впливає на прийняття стратегічних рішень. На жаль, сам OLAP погано підходить для прогнозування порожніх розмірів кубів даних. Протягом багатьох років дослідники та аналітики намагалися вдосконалити процес прийняття рішень у системах OLAP і додати прогнозування та інші параметри до програм OLAP. Сьогодні, в епоху Індустрії 4.0, з доступністю великих даних виникає потреба застосовувати нові технології для вирішення таких проблем. Ці технології включають нейронні мережі. У статті розглядається проблема інтеграції OLAP і нейронної мережі. У зв'язку з цим наводяться відомості про нейронні мережі: їхні властивості, види, а також їхні можливості. У статті розглядаються можливості та переваги інтеграції OLAP та нейронної мережі. Зроблено висновок, що у випадку великих даних, інтеграція OLAP і нейронних мереж дуже ефективна для вирішення проблем класифікації, кластеризації та прогнозування порожніх мір кубів OLAP. Представлено архітектурно-технологічну модель інтеграції OLAP і нейронних мереж. Відзначено, які типи нейронних мереж можна використовувати для вирішення заданих у моделі задач класифікації, кластеризації та прогнозування.

Ключові слова – сховище даних, OLAP, штучний інтелект, машинне навчання, нейронна мережа, прогнозування, кластеризація, класифікація.

Introduction

People have always been eager to create all the necessary conveniences for everyday life for themselves. As a result of their efforts, proper technologies and machines appeared to assist them. One of the areas of such technologies is Artificial Intelligence (AI) [1]. The term "Artificial Intelligence" was first coined in history in 1956 by John McCarthy. The term refers to any system capable of performing creative functions and solving problems that typically require human intelligence, that is, which are traditionally performed by humans. In effect, AI imitates human intelligence in machines that are programmed to think and act like humans. AI plays a critical role in today's world by enabling automation, improving decision making, increasing efficiency and productivity. It opens up new opportunities for innovation and growth in a variety of industries, including healthcare, finance, manufacturing, transportation, e-commerce, education, and more.

AI is frequently discussed and explored jointly with machine learning (ML). ML is a branch of AI. The idea of ML is that machines should be able to learn and adapt through experience, making predictions based on statistical data collected by computers. Thus, the concept of AI is a broader concept compared to the concept of ML.

One of the areas of artificial intelligence is neural networks (NN). NNs are used to recognize hidden patterns in raw data, for clustering [2] and classification [3], as well as for solving tasks in the field of AI.

Recently, both government agencies and private ones have widely used OLAP systems based on OLAP (Online Analytical Processing) technology [4]. They are applied, for example, in banking, medicine, industry, telecommunications, trade, etc. [5] describes the use of OLAP in the terminology environment, namely in the terminology system to expand its capabilities and for more efficient functioning.

NN and OLAP are essential tools for quickly and efficiently discovering valuable, non-obvious information from a large collection of data.

The goal of this paper is to study the possibility of integrating OLAP and NN, and to identify the benefits of this integration.

The first section describes the main characteristics of OLAP systems. The second section provides information about NN: historical background, types of NN and their functions. The third section reviews related work. The fourth section is devoted to the problem of integrating NN and OLAP, and also presents an architectural and technological model for integrating OLAP and NN. Finally, the fifth section presents the final conclusions of this article.

Main characteristics of OLAP systems

An OLAP system is an information and analytical data processing system developed on OLAP technologies. The popularity of OLAP is explained by the fact that it is possible to solve many problems with its help, namely: to implement operational processing of information, including issuing information in various sections and dynamic report generation and its analysis based on the data obtained, to perform monitoring and forecasting [6]. The OLAP system is designed for generating reports, constructing predictive scenarios and performing statistical calculations based on large collection of data with a complex structure [7]. The key components of the OLAP system are a data warehouse (DW), an OLAP server and applications.

DW is a source of processed information accumulated from already existing systems of geographically distributed units. DW is a domain-specific, non-volatile, integrated, time-varying set of data for decision support [8].

OLAP is an element of the DW and takes advantage of its information.

The OLAP server is the core of the system, with the help of which multidimensional data structures are processed and communication between the DW and system users is ensured.

Applications are used for user work. They formulate queries and visualize the responses received. OLAP applications are used to store DW analysis contexts in multidimensional data structures, i.e., in OLAP cubes. OLAP cubes enable analysts to explore information and report through interactive, easy-to-use dashboards. It is OLAP cubes that contain indicators (measures) used for analysis and management decision-making.

One of the important goals of OLAP is to make decisions based on historical data.

Note that OLAP provides any analytical report within a few seconds due to its wide visualization functionality.

Neural network as a method in artificial intelligence

The basic principles of NN operation were described back in 1943 by Warren McCulloch and Walter Pitts [9]. In 1957, neuroscientist Frank Rosenblatt developed the first NN. He was the author of the first paper on perceptrons [10]. In 2010, large amounts of training data opened up the possibility of using NN for machine learning.

Each NN includes a first layer of neurons called the input layer. This layer does not perform any transformations or calculations; it receives and distributes input signals to other neurons. This layer is the only one that is common to all NN types.

The main types of NN are as follows:

Perceptron. Perceptrons are single-layer or multilayer feed-forward artificial NNs with binary or analog outputs that are supervised learning.

Single layer neural network. It is a structure for the interaction of neurons, in which signals from the input layer are immediately sent to the output layer. The output layer converts the signal and immediately produces a response.

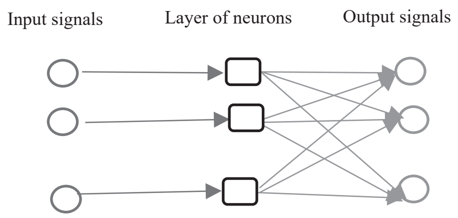


Fig. 1. Example of a single layer neural network

Multilayer neural network. This NN, in addition to the output and input layers, has several hidden intermediate layers. The number of these layers depends on the complexity of the NN.

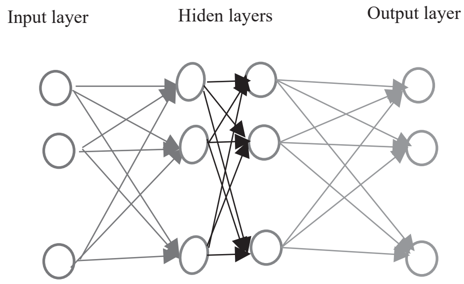
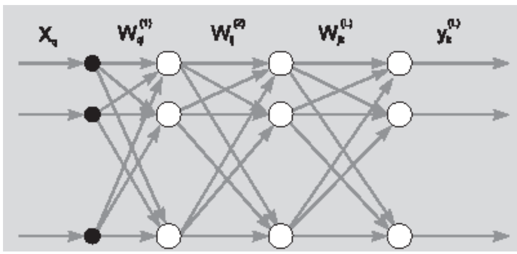


Fig. 2. Example of a multilayer neural network

NN can be classified not only for the number of layers, but also according to the direction of information distribution along synapses (connections) between neurons:

Feed-forward neural network (unidirectional). In this structure (Fig. 3.) the signal moves strictly in the direction from the input layer to the output layer.



Input layer Hidden layers Output layer

Fig. 3. Architecture of a three-layer feed-forward neural network.

Recurrent neural networks (with feedback) (RNN). Here the signal moves both forward and backward. As a result, the output result can be returned to the input. Types of RNN:

- One to one
- One to many
- Many to one
- many to many

Self-organizing maps. They include self-organizing Kohonen maps. They are a powerful, self-learning clustering engine: the results are displayed in compact and easy-to-interpret two-dimensional maps. The Kohonen map is used for exploratory data analysis. It is able to recognize clusters in data and also establish class proximity. In addition, the Kohonen card is able to predict client behavior. If it is built a Kohonen map containing clusters for each group of clients according to their degree of loyalty, then with its help the expected behavior of the client can be predicted and applied appropriate marketing policies to them. The Kohonen map is also capable of detecting anomalies. It distinguishes clusters in the training data and assigns all data to one cluster or another. If after this the map encounters a data set that is unlike any of the known samples, then it will not be able to classify such a set and thereby reveal its anomaly [11].

There are also other criteria for NN classification:

- depending on the types of neurons: homogeneous and hybrid;
- depending on the NN learning method: supervised learning, unsupervised learning, reinforcement learning;
- according to the type of input information, NNs are: analogous (use information in the form of real numbers), binary (operate with information presented in binary form); figurative (operate with information presented in the form of images, signs, hieroglyphs, symbols);
- according to the nature of synapse setup: with fixed connections (NN weight coefficients are selected immediately based on the conditions of the problem, with $dW/dt=0$, where W denotes NN weight coefficients); with dynamic connections (when the learning process is in progress in the settings of synaptic connections, that is, $dW/dt\neq0$, where W denotes NN weight coefficients).

Related works

In recent years, there has been a need to add new capabilities to OLAP. First of all, this is due to the fact that when solving some problems, issues arise due to the sharp increase in the flow of data entering systems from various sources. Therefore, there are big problems in solving some tasks. On the other hand, it should be noted that today the requirement of the time is the solution of many intellectual tasks. Intellectual tasks based on big data are most effectively solved using NN. Thus, it can be argued that the integration of two technologies such as OLAP and NN is very useful and important, as it enriches each of them: OLAP is the ability to navigate the multidimensional structure of indicators, and NN is the ability to intelligently solve tasks on large amounts of data .

If a large amount of data in intelligent systems uses OLAP, NNs are effective for solving clustering and classification tasks. Note that clustering, unlike classification, does not have predefined categories into which all data should be grouped. In this case, the NN itself generates clusters based on common features of the data.

Clustering is one of the most important methods of data analysis. The article [12] provides a comprehensive overview of clustering methods such as the self-organizing Kohonen map, as well as clustering algorithms such as k-means, fuzzy means algorithms, etc.

One of the classes of NNs primarily used to solve clustering tasks is the Kohonen neural network [13]. In this case, learning occurs without a teacher, that is, only input data sets are used, and no output values are required.

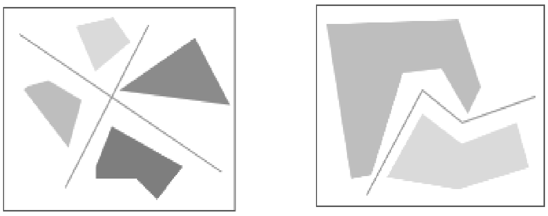
k-means is the most popular and simplest clustering method. Its main disadvantages are: you need to know the number of clusters in advance; very sensitive to the choice of initial cluster centers.

In the decision-making process, you can encounter many fuzzy tasks. Therefore, the queries to the DW that the analyst is trying to formulate may often contain uncertainties. Clustering applied to the dimensions of an OLAP cube using NN produces linguistic variables. This will make it possible to solve fuzzy tasks as well [14].

For data mining, the task of data classification plays an important role. Currently, a large number of different types of classifiers have been developed, including those built on machine learning. These include NN. Although the classification task for NNs is not the main one, their use has a number of advantages:

- NNs are self-learning models, the operation of which requires almost no user intervention;

- NNs are universal approximators that allow you to approximate any continuous function with suitable accuracy;
- NNs are nonlinear models. It allows to solve effectively classification tasks even in the absence of linear separability of classes (Fig. 4) [15].



Linearly separable classes   Linearly inseparable classes

Fig. 4. Options for linear separability of classes

By means of NN, forecasting problems that are of great importance in the production, economic and financial spheres are also solved. Forecasting in OLAP is important because when looking at the contents of a cube, it can often be sparse, meaning it is missing some measures, and may also be missing dimensions. This happens due to missing information or input errors. The absence of any measures and measurements is undesirable and can lead to incorrect analysis when making strategic decisions.

In OLAP systems, NN can be used in parallel with OLAP, i.e. OLAP cubes are created on historical DW data, and NN forecasting work is based on the same historical data [16]. The disadvantage of this approach is that there are no training data sets.

The approach proposed in [17] includes two stages. First, principal components are analyzed to reduce the dimensionality of the data cube and special training sets are created. Then, in the second stage, a new OLAP-oriented multi-layer perceptron network (MLP) architecture is proposed whereby training is implemented on each training set and predicted dimensions are generated.

In [18], the possibilities proposed in [17] are expanded. First, the authors introduce a generalized framework, i.e., Multi-perspectives Cube Exploration Framework (MCEF), for applying the classical data mining algorithm to OLAP cubes. Secondly, the authors refer to modular NNs that apply a neural approach to predicting multidimensional cubes (NAP-NN). Modular NNs are a collection of several different networks that operate independently and contribute to the result. Each NN has its own set of input data. These networks do not interact with each other during task execution. The main advantage of modular NN is that the huge computational process can be divided into several subprocesses. This reduces computational complexity and increases computational speed. But ultimately, the processing time will depend on the number of neurons and their participation in calculating the results. Note that NAP-NN includes a preprocessing step. In this step, principal component analysis (PCA) is performed to reduce the size of the OLAP cube of the proposed method. Modular neural networks work effectively in cases where several directions of the system are simultaneously processed.

Note that the article presents experimental results showing the effectiveness of NN.

OLAP and neural networks integration model

Big data includes huge volumes of heterogeneous and rapidly arriving digital information that cannot be processed with traditional tools. Very effective analysis of big data is carried out using machine learning methods, in particular NN. It is very important that the advantage of NN, such as the detection of hidden patterns that are invisible to humans, also works well on big data. The integration of OLAP and NN also provides these benefits.

Figure 5 illustrates the architectural and technological model for integrating OLAP and NN.

According to the Figure 5, data from various sources, before entering the DW or data mart (DM), first goes through ETL technology. During the ETL process, data is cleared of duplication, contradictions and typos and brought into a common format. OLAP cubes are built based on DW (or DM) data. The figure shows the integration of NN with an OLAP cube to classify and cluster OLAP cube data and predict empty measures.

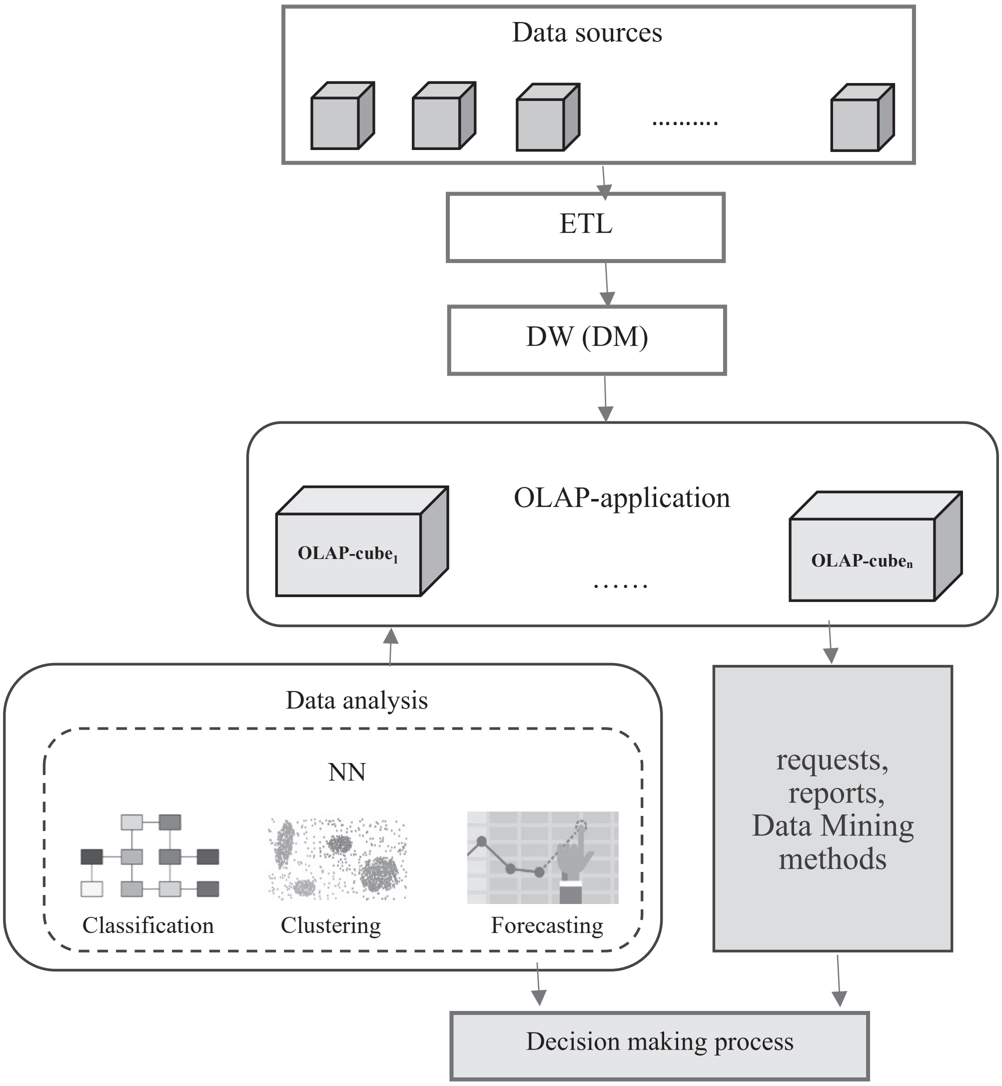


Fig. 5. Architectural and technological model of integration OLAP and NN

The integration of NN and OLAP is that the NN environment is built into OLAP applications that operate on a multidimensional structure and a large volume of data cubes.

Note that at this stage, the execution of processes characteristic of traditional OLAP is also ensured, namely: analytical queries are implemented on data for their rapid viewing and analysis, reports are issued based on the OLAP cube data, which can be with either intermediate or final results. It is also possible to view the same data from different angles.

Depending on the task set, the most appropriate NN is selected from the above types. For example,

- single-layer and multi-layer perceptrons are used for classification;
- single-layer and multi-layer perceptrons are used for classification;
- single-layer or multilayer perceptrons and Kohonen map are used for forecasting [11, 19].

Finally the results obtained will serve to make management decisions.

Conclusion

Currently, in the era of Industry 4.0, there is a dramatic increase in the flow of data.

This creates great complications when solving some problems related, for example, to classification, data clustering, forecasting, and even in some cases makes solving these problems impossible. AI can provide quick and effective solutions to such problems. AI refers to any system capable of performing creative functions and solving problems that would typically require human intelligence. AI contributes to development and innovation in various industries, such as healthcare, finance, manufacturing, transport, e-commerce, education, etc. NNs are one of the areas of AI. NNs, being implemented into systems, can solve important tasks. Such systems include OLAP systems based on OLAP technology. Recently, these OLAP systems have been widely used both in government agencies and in private ones. OLAP enables analysts to explore and navigate a multidimensional structure of metrics called an OLAP cube. The purpose of the article is to study the possibility of integrating OLAP and NN, as well as to demonstrate the benefits of such integration.

This article presented an architectural and technological model, according to which data analysis is performed using NN. The integration of NN and OLAP is achieved by embedding the NN framework into OLAP applications that operate on a multidimensional structure and a large volume of data cubes.

Further studies will develop methods for embedding NN environment into OLAP applications to integrate NN and OLAP.

References

1. Sh. Iskanderova, The impact of artificial intelligence on the modern world, in: Science and Education Scientific Journal (2023), Impact Factor 3.848. Vol. 4, issue 4, pp. 564-570. [in Russian]
2. L. Yunusova, A. Magsumova, Clustering using neural networks and searching for dependencies, 2019. Accessed: 22.01.2024. https://cyberleninka.ru/article/n/klasterizatsiya-s-pomoschyu-neyronnyh-setey-i-poisk-zavisimostey-1. [in Russian].
3. M. Artikova, Sh. Rasulova, Data classification using neural networks, in: Scientific Collection «InterConf»: Scientific Goals and Purposes in XXI Century, 2021, (78), pp. 403-409. Accessed: 15.02.2024. https://ojs.ukrlogos.in.ua/index.php/interconf/article/view/15124, [in Russian].
4. What is Online Analytical Processing (OLAP)? Accessed: 18.12.2023. https://aws.amazon.com/ru/what-is/olap/. [in Russian].
5. R. M. Alguliyev, G. Ch. Nabibayova, A. M. Gurbanova, Development of a Decision Support System with the use of OLAP-Technologies in the National Terminological Information Environment, in: International Journal of Modern Education and Computer Science (IJMECS) (2019). Vol.11, no. 6, pp. 43-52.
6. E.F Codd, S.B. Codd, C.T. Salley, Providing OLAP (on-line Analytical Processing) to User-analysts: An IT Mandate, 1993. Accessed: 16.10.2024. http://www.estgv.ipv.pt/paginaspessoais/jloureiro/esi_aid2007_2008/fichas/codd.pdf.
7. Main characteristics of OLAP systems, 2020. Accessed 18.12.2023. https://hsbi.hse.ru/articles/osnovnye-kharakteristiki-olap-sistem/. [in Russian].
8. W.H. Inmon, Building the Data Warehouse, John Wiley & Sons, 2005, p. 596.
9. W. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, in: Bulletin of Mathematical Biology (1943), no 5, pp. 115–133.
10. F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychological Review (1958), **65**(6), pp. 386–408.
11. Kohonen map. Accessed: 18.11.2023. https://basegroup.ru/deductor/function/algorithm/kohonen [in Russian].
12. K.-L. Du, Clustering: A neural network approach, in: Neural Networks (2010), vol. 23, issue 1, pp. 89-107.
13. Kohonen neural network, self-organizing maps, learning, 2022. Accessed 20.03.2024. https://microtechnics.ru/nejronnaya-set-kohonena-samoorganizuyushhiesya-karty-obuchenie/ [in Russian].
14. Kumar K., Krishna R, Kumar De S. Fuzzy OLAP Cube for Qualitative Analysis / Proceedings of the 3rd International Conference on Intelligent Sensing and Information Processing (ICISIP), 2005, pp. 290-295, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1529464&tag=1
15. V. Oreshko, Data classification using neural networks, 2021. Accessed 15.03.2024 https://loginom.ru/blog/neural-classification [in Russian].
16. M. Gulesian. Using Neural Networks and OLAP Tools to Make Business Decisions, 2008. Accessed 11.02.2024. https://www.developer.com/database/using-neural-networks-and-olap-tools-to-make-business-decisions/
17. W. Abdelbaki, R. Messaoud, S. Yahia, Neural-Based Approach for Extending OLAP to Prediction, in: Proceedings of science conference

Data Warehousing and Knowledge Discovery (DaWaK 2012), Springer-Verlag Berlin Heidelberg, 2012, pp. 117–129.
18. W. Abdelbaki, S. Yahia, R. Messaoud, Modular Neural Networks for Extending OLAP to Prediction. Book Transactions on Large-Scale Data- and Knowledge-Centered Systems XXI, Springer Nature, 2015, pp. 73-93.
19. Neural networks, perceptron. Accessed: 10.12.2023. https://neerc.ifmo.ru/wiki/index.php?title=Нейронные_сети,_перцептрон#cite_note-3 [in Russian].

Received: 04.04.2024
Internal review received: 18.04.2024
External review received: 23.04.2024

About the author:

Nabibayova Gulnara Chingiz
PhD in technical sciences
http://orcid.org/0000-0001-8743-7579

Author's place of work:

Ministry of Science and Education
of the Republic of Azerbaijan
Institute of Information Technologies
Tel.: (+994 12) 539 01 67
E-mail: gnabibayova@gmail.com
www.ict.az