IRSTI 28.23.37 https://doi.org/10.26577/jpcsit2024020103

B.A. Muratbek*, G.U. Bektemisova

International Information Technology University, Almaty, Kazakhstan
*e-mail: 36138@iitu.edu.kz

USING NEURAL NETWORKS FOR DEMOGRAPHIC PREDICTION

Abstract. Currently, the big problem or opportunity is the collection and correct processing of data. Using a large amount of reliable data, as well as several types of data processing methods, it is possible to optimize many processes and the allocation of funds. Now, many neural networks are used throughout the Internet, an example of this is the use of neural networks to predict sales in retail. This method saves resources for analysts, speeds up work and provides round-the-clock monitoring, so it is possible to perfect sales in the market. This article also discusses the way of using a neural network to predict population demographic growth using neural networks. The proposed system should help perfect the distribution of funds and reduce the cost of demographic analysis. The article also discusses several methods of training a neural network to achieve this result. As a result of this research, it was possible to create a neural network model with a prediction accuracy of 99.7 percent. It is worth noting that testing occurred with the data we collected, and the accuracy may not be as accurate when used with other data.
Key words: Machine Learning, Deep Learning, Artificial Intelligence, Data Analysis, Big Data.

1 Introduction

Nowadays, we are accustomed to solving many problems using statistics, and when doing this we do not always realize that we are using statistics, for example, in the morning when leaving home for work, we look in map applications for what period we can now get to work. This is one of the most common examples of use, applications that calculate and show you the current time use a large number of algorithms, and some of them have begun to switch to the use of neural networks, because thanks to the training of the neural network, the assumption of the given time is more accurate and cheaper.

At the moment, a large amount of analytical or statistical work is performed by workers manually. That is, the creation of reports, assumption of sales growth, analysis of population growth and many other examples are performed by trained, highly paid employees. This leads to large resource and time costs in developing a future analysis and plan. At the same time, this analysis process is extremely necessary, since without careful planning and data analysis it is hardly possible to achieve success in a competitive environment. After all, by analyzing the available data, we find patterns and using them correctly, we adjust the plan or give results about any expected growth. It turns out that this expensive and resource-intensive element is necessary in the work process but needs to be optimized and improved.

However, now you can find the use of neural networks everywhere in different industries, sometimes this solution can be found in very unexpected tasks and industries. Neural networks are a technique in artificial intelligence that trains computers to process information in an equivalent way to the human brain. They fall under the category of deep learning, which uses interconnected nodes or neurons organized in a layered structure, similar to the structure of the human brain. These neural networks form an adaptive system in which computers learn from their mistakes and continually improve their performance. Thus, artificial neural networks strive to solve complex problems such as information compression or face recognition with increased accuracy.

High accuracy of the assumption of a particular result is achieved by careful sampling of data, as well as a well-prepared training model in which dependencies important for the result are determined in the data for training and testing. Also, to obtain a more accurate result when training a neural network, it is necessary to determine the exclusion factors or cases that fall outside the main data flow and understand the reason for the data exception.

The use of neural networks has fundamentally changed the processes of application development, data analysis and in many different industries. This has allowed the emergence of new directions in the

IT industry such as "Machine Learning," and a new type of optimization of work with data. It is also noteworthy that the use of these methods seems expensive but brings great profits to companies and can help in some very unexpected industries.

There are several reasons for this, the main one of which is that this method of solving problems, that is, the use of neural networks or artificial intelligence, makes production in the future many times cheaper, because training, testing and universally using a neural network or artificial intelligence is many times cheaper than maintain a whole team of workers for this.

In conclusion, the article examines the way to use neural networks to solve problems in statistical fields; to be more precise, we will consider the use of neural models in analysis and training of neural networks to create predictions about demographic growth. We will also analyze articles related to our topic and other solutions to this problem.

1.1 Paper organization

This article will consider the use of neural networks for data analysis and forecasting demographic growth. It will also consider the use of neural networks in other areas and compare the result of this research with other existing similar analogues.

In the second chapter, we will explore the use of neural networks for similar purposes and the different approaches to using neural networks.

The third section introduces the very concept of neural networks and examines different methods and models for training neural networks. Also, this chapter compares the used models of training neural networks.

The fourth section proposes to consider the use of neural networks in optimizing demographic analysis and predicting demographic growth. This part discusses the complete development cycle of a neural network model. Which begins with collecting the required amount of data and training a neural network on this data using different methods and models, among which you can see models such as "Random Forest."

The fifth section discusses the results of testing the work performed and comparing it with existing analogues. Not only is the prediction accuracy of the model compared, but also the relevance and necessity in a competitive environment of use.

The last chapter of the article describes the result of the study, which is based on the study itself.

Research contributions

Optimization of time and resource costs of analytical work during studies of demographic growth. We offer a web application that will make demographic analysis using a neural network accessible to everyone. Which will entail a more competent distribution of funds and timely preparation of the resources and benefits necessary for the population. Why web application? Since today most of the population has access to the global network, that is, the Internet, and by creating a web application, we will give each of them access to our application.

A new way to use population growth analysis. At the moment, demographic analysis is used only by states or large companies, but using the solution presented in this article, the results of demographic analysis can also be used by small/medium businesses, and thanks to a more detailed consideration of demographic growth and assumptions about subsequent growth, it will be possible to more accurately and better stimulate economic growth.

2 Related literature

Recently, there has been a great development in the field of machine learning in a very short period, making the widespread use of neural networks profitable and necessary for competitiveness. The purpose of the first article [1], that we will consider is to assume demographic information by collecting data from customer mobile applications, based on the fact that in our time almost every person has a mobile phone, it was assumed that data collection would be very successful. Thanks to this, they collected the necessary data from 5,000 users and trained a neural network with a CNN model on the collected data. Based on the results of the comparative experiment, the predictive effect of our model is also better than the effect of three classical machine learning algorithms.

The following article [2] examines the problem of student dropouts and proposes to predict the chance of a student dropping out by training a model that would take the necessary demographic data of students and based on this data would issue a verdict. For this work, they used a learning model called "Random Forest" and prepared three different datasets. With this solution, it is expected to predict the performance of students before they drop out.

The third article [3] that we selected for analysis of similar literature talks about modular neural networks. Their types and key features and advantages are discussed, as well as a comparative analysis between them. More in this article, we describe a number of neural networks based on self-organizing Kohonen maps and which can be successfully used to identify dynamic objects, and also describe new ones that have been developed and successfully used to identify dynamic objects.

The article [4] proposes a new method for crowd counting using background-elimination convolutional neural network in scenes with uneven population distribution. First, the image is divided into blocks, where the network detects the presence of a crowd in each block, excluding background noise. Special convolutions are then used to estimate crowd density. Experiments on two datasets confirm the superiority of this method compared to existing approaches.

The article [5] outlines a new crowd counting method based on cooling convolutional neural networks. The method uses various filters to more accurately measure the number of people in the scene, ensuring the safety of the cameras. It also applies an appropriate training strategy, including the use of 1x1 filters to speed up training and reduce model parameters. Experimental results show that this method outperforms seven popular algorithms on three public datasets, improving counting accuracy by 7% and 8%.

Our next paper [6] directly addresses our research topic. Article [6] discusses the significance and application of deep learning in crowd counting and population density estimation problems in computer vision. She describes the limitations of traditional methods, such as manual feature extraction and model complexity, and discusses the advantages of deep learning methods such as CNN and RNN on these tasks.

The paper [7] is also closely related to our research topic and presents a new method for counting population in dense scenes using a deep perception network. It proposes a framework to train a model end-to-end using multi-scale kernels to capture semantic information about the context of an image. This method adaptively encodes the scale of context information needed to accurately predict population density. The results show that the proposed algorithm outperforms existing crowd counting methods, especially when the perspective effect is strong.

Also, for a better understanding of the construction of different learning models, we also looked for scientific articles in which models are used in scientific practice. For a better understanding of how Linear Regression works, we read the articles:

Article [8] compares the accuracy of innovative linear and logistic regression algorithms for groundwater level detection. The new linear regression algorithm achieved an accuracy of 93.27%, surpassing the accuracy of the logistic regression algorithm (86.5%).

Article [9] shows that sparse linear regression with elastic net regularization successfully decodes signals from neural populations. In a study on real-world electrocorticography (ECoG) data, the elastic net achieved high accuracy in detecting hand grip and cursor movement, outperforming other regression methods. This indicates the potential of using this method in brain-computer interfaces.

Article [10] uses the multiple linear regression method on data on CPI, money supply and retail trade volume for the period from December 2019 to September 2020. The results indicate high accuracy in predicting CPI when using money supply and retail trade volume as independent variables. This confirms the influence of these factors on consumer prices in China.

Article [11] presents a forecast of demand for bicycle rentals in the Capital Bike share program in Washington DC using random forests and multiple linear regression. First, a multiple linear regression model was built, but it turned out to be less accurate. Then, a random forest model was proposed, which greatly improved the demand forecasting accuracy.

Also, for a deeper understanding of the use of Random Forest model in scientific practice:

Article [12] proposes a method for classifying and identifying network traffic based on random forest algorithm in software-defined networks (SDN). In the context of the development of machine learning, its application to the classification of network traffic is becoming a new direction of research. The paper presents a method based on random forest, which lets you effectively classify traffic with high accuracy and ensures the efficiency and stability of processing large volumes of data.

Article [13] proposes the use of genetic algorithm to optimize concrete frame structures using improved random forest. The quality of structures is assessed based on the relative dynamic modulus of elasticity. The results show that the proposed random forest model achieves high prediction accuracy and better fit, making it an effective tool for predicting concrete frame structures.

Article [14] proposes an improved random forest-based classification method using hybrid cluster

selection. The proposed approach considers the random forest model diversification to improve the classification quality. As a result of the development of a clustering method for selecting base classifiers using the Dunn index, as well as a hybrid method for selecting clusters based on hierarchical and k-medoid clustering, the effectiveness of the proposed approach was confirmed.

Article [15] evaluates the application of a random forest model for forest fire forecasting on the island of Borneo based on climate data. The study shows that the model can successfully represent wildfire events on this island.

Analyzing the above-mentioned studies, we can understand that the demand for the use of neural networks in various areas is extremely high, and the use of demographic data in this area is profitable and convenient, since this type of data is not personal and is not under any legal protection, that is they are convenient for use in research. Moreover, we see that neural networks are used in various fields, even in education, and it can bring very good results. Also, during the analysis of the literature, it was possible to find out what types of models are recommended to be used in certain data conditions.

3 Neural networks

Neural networks are a powerful tool in the field of artificial intelligence, developed by analogy with the workings of the human brain. They are used to process information, identify patterns in data, and make decisions, making them a sought-after tool in various fields such as machine learning and data analytics.

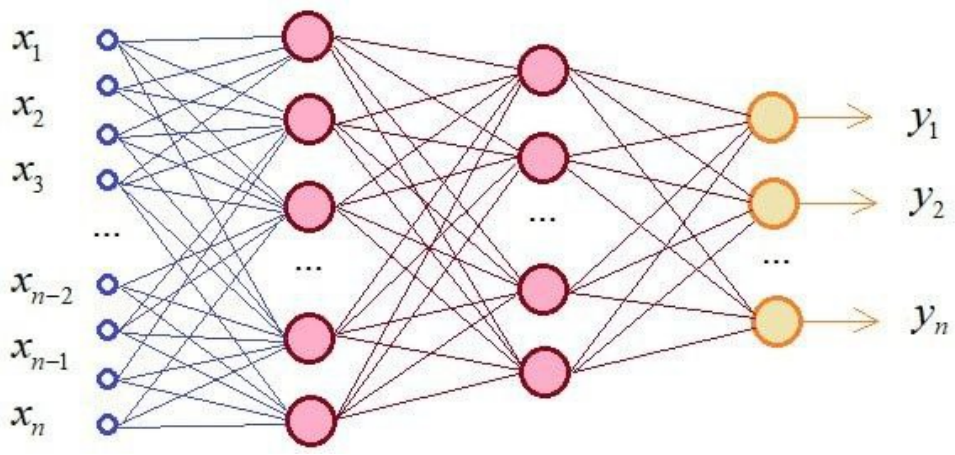

Figure 1 – Visualization of the structure and operating principle of neural networks

Neural networks are built on the principle of interconnected and interacting neurons combined into layers. The basic components of a neuron include inputs, weights, adder, activation function, and output. Neuron layers are usually divided into input, hidden and output.

- Input Layer: Receives input data.
- Hidden layers: Perform calculations and discover patterns in data.
- Output layer: Provides results or makes decisions.

Neural networks are trained through a process called backpropagation. In this process, the network predicts the outcome and then compares it with the desired outcome. The error is propagated back through the network and the neuron weights are updated to reduce the error. This process is repeated until the network reaches acceptable accuracy.

Usage of Neural Networks:

- Pattern Recognition: Neural networks have been successfully used for pattern recognition in images, audio, and video. This may include recognizing faces, objects, handwritten text and more.
- Natural Language Processing: In natural language processing, neural networks are used for automatic translation, sentiment detection, chatbots, and other applications.
- Forecasting: Neural networks are effectively used in forecasting time series, financial markets, weather, and other variables where it is important to predict future values.

- Medical Diagnostics: In the medical field, neural networks help in diagnosing diseases, analyzing medical images, and predicting response to treatment.

- Automatic control and robotics: Neural networks are used for automatic control in various fields, including autonomous cars and robotics.

Training neural networks involves various models, each designed to solve specific problems. Some examples of basic neural network training models:

- Feedforward Neural Networks: This is the basic and simplest model of neural networks. Data is transferred from the input layer through hidden layers to the output layer without loops or feedback. Widely used in classification, regression, and pattern recognition problems.

- Recurrent Neural Networks (RNN): These networks have feedback loops that allow them to retain information about previous inputs. This allows sequences of data to be processed. Effective for processing sequential data such as time series, natural languages, and speech.

- Convolutional Neural Networks (CNN): Specialized in processing multidimensional data using the convolution operation. Effective for highlighting spatial patterns. Main application is processing images, videos, and other multi-dimensional data.

- Deep Autoencoders: These are neural networks used to learn a compact representation of data through compression and reconstruction. They consist of an encoder and a decoder. They are used in data dimensionality reduction tasks and in data generation and recovery.

- Long Short-Term Memory (LSTM) networks: This is a type of recurrent neural network designed to combat the decaying gradient problem. They are better able to retain and use information for the long term. Widely used in tasks where processing and understanding long-term dependencies is important, such as machine translation and text analysis.

Neural networks play a key role in the development of artificial intelligence and machine learning. Their ability to process complex data and extract important patterns makes them a useful tool in solving a variety of problems in science, technology, and business. With the constant development of technology, the role of neural networks in modern society is expected to only expand.

4 Predicting demographic growth

Let me remind you that the goal of the study is to build a neural network that should predict demographic growth. To train the neural network and the research itself, we chose the demographics of the country of Kazakhstan.

4.1 Preparing data

As previously stated, training neural networks begins with collecting the necessary data. In our case, collecting data was very labor-intensive. The reason for this is the lack of data for all time in our country, which is why we had to manually collect data from diverse sources and average them when merging.

We used the Kazakhstan's government website [16] with statistics to take the data, that was the first part of data for training. In this dataset, we only have collected data on the population of Kazakhstan since 2000. That is, we only have data for 23 years, which is extraordinarily little for training a neural network. But on the other hand, the data is official and accurate, moreover, it is not poorly distributed; in the table we have a division into types, namely, divided into the population of cities, villages, and the total population. They are also divided by gender.

After we tried to train a neural network using this data, we still accepted that there was very little data, and that working with the data would be labor-intensive and time-consuming, after which we decided to look towards other sites with statistical data, this was the World Bank website [17] as well as the "macrotrends" website [18], unfortunately the first site did not include much useful data about the demography of Kazakhstan. However, the second one had data from 1950 to 2023. As it turned out, this site also links to the World Bank website [17] to obtain data.

After much work, we were able to collect approximate data on the population of Kazakhstan from 1950 to 2022 and were even able to break it down into regions of the country. However, due to the averaging of data, and obtaining data not only from government sources, we say that this is useful data, but there is still some inaccuracy.

4.2 Training a neural network model

Now that we have collected all the necessary data to train the neural network, it is time to look at the final data and start training the neural network.

As you can see from "Figure 2", the data does not have several unique features, that is, our data has only one feature, this is the population, so the first attempt at training can be carried out in a linear regression model, since it is excellent for such tasks. Also, we cannot forget about dividing the data by 2, the first for training and the second for testing.

	City	1950	1951	1952	1953	1954	1955	1956	1957	1958	...
2	Абай облысы	206236.493201	213434.516870	220750.506386	228132.063469	235682.094495	243515.125802	251740.261078	260462.513785	269544.334571	...
3	Семей қ.ә.	110697.530589	114561.073034	118487.933701	122449.987789	126502.470342	130706.853423	135121.698486	139803.371554	144678.042950	...
4	Семей қаласы	103708.548256	107328.162678	111007.097671	114719.004119	118515.629755	122454.565550	126590.675631	130976.767306	135543.672195	...
5	Шаған к.	158.517880	164.050438	169.673668	175.347294	181.150413	187.171052	193.493071	200.197185	207.177671	...
6	Шүлбі к.	969.358803	1003.191165	1037.577993	1072.273003	1107.759880	1144.576923	1183.236945	1224.233532	1266.920172	...
...	...	...	...	...	...	...	...	...	...	...	...
481	Шымкент қаласы	400219.790904	414188.179734	428385.491513	442710.042840	457361.532456	472562.208646	488523.796575	505450.083968	523074.144382	...
482	Абай ауданы	116118.909668	120171.667972	124290.845487	128446.940011	132697.891708	137108.183217	141739.244022	146650.200655	151763.607652	...
483	Әл фараби ауданы	68147.140620	70525.598107	72943.035285	75382.138088	77876.910075	80465.194416	83183.042464	86065.154027	89066.078395	...
484	Еңбекші ауданы	97732.863585	101143.915890	104610.870715	108108.897181	111686.761316	115398.735702	119296.523195	123429.888347	127733.648259	...
485	Қаратау ауданы	118220.877031	122346.997767	126540.740026	130772.067560	135099.969357	139590.095311	144304.986895	149304.840940	154510.810076	...

484 rows × 74 columns

Figure 2 – Final data columns and examples.

Based on our "Literature review", we decided to start training the model using Linear Regression, more precisely from the article [8], article [9], article [10] and article [11]. Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (attributes). The model is based on the assumption that these relationships are linear, that is, they can be approximated by a straight line. Least squares method (LSM):

- The goal is to minimize the sum of squared differences between the actual and predicted values.

- The method finds optimal values of model parameters that minimize the sum of squared errors.

Using this learning model, we achieved a prediction accuracy of 88 percent. These are incredibly good results, but if we remember that our initial data also has its own errors, then we need to show the result as accurately as possible, that is, strive for 100 percent accuracy. At the same time, we spend a lot of time on the process of training the neural network itself, that is, due to the fact that training a neural network requires a lot of computing power, we have to wait until the training process is completed. Therefore, to speed up this moment, we decided that we would train a neural network and experiment with data cut into areas. Which, not much, but still speeded up the learning process.

To get more accurate results, we looked towards another, more complex learning model, namely Random Forest, also after reading the article [12], article [13], article [14] and article [15]. Random Forest is an ensemble machine learning that combines multiple decision trees into a single model. Each tree in the forest is built independently, and then their results are combined to improve the accuracy and robustness of the model. Basic Components of Random Forest:

- Decisive Trees: Each tree is built based on a random subsample of data and a random subset of features. Each node of the decision tree makes a division of data based on a certain attribute, with the goal of maximizing the information gain or reducing the uncertainty criterion.

- Bootstrap Sampling: For each tree, a bootstrap sample is constructed, which is created by randomly selecting and returning from the original data set. This allows each tree to see different subsets of the data, improving generalization ability.

- Randomness of Signs: When constructing each node of the tree, a subset of features is randomly selected for decision making. This prevents the dominance of individual traits and promotes the creation of diverse trees.

- Voting and Averaging: The predictions of the individual trees are combined to make a decision. In a classification problem this could be voting, and in a regression problem it could be averaging.

By retraining the model using the "Random Forest" learning model, we were able to greatly improve the model's prediction accuracy, ultimately achieving 99.7 percent accuracy. Which is an excellent result even if we consider the aforementioned error problem due to data averaging.

5 Testing and comparing results

During testing, it was revealed that the main trend of growth and decline in demographics persists for all regions, although there is a difference in the pace, the general trend is still maintained. At the same time, when obtaining data prediction results, linear regression does not behave as well as Random Forest.

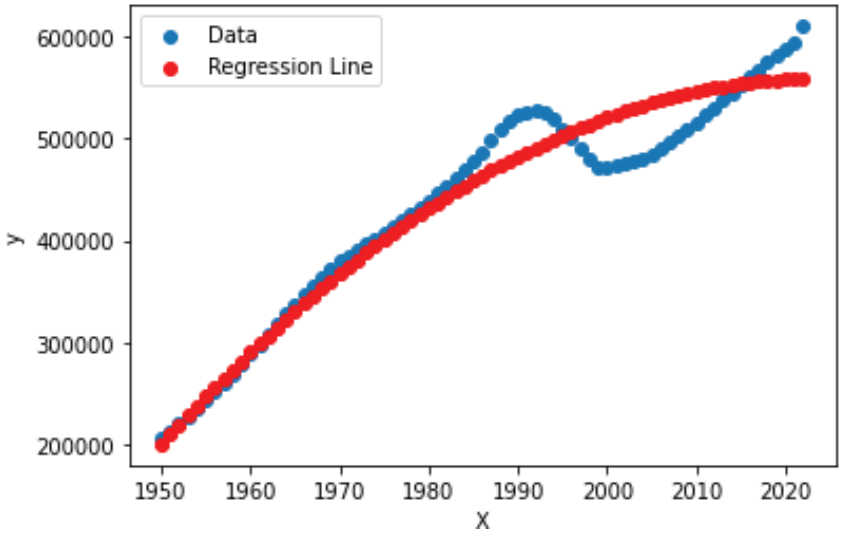

Figure 3 – Visualization of data prediction (Linear Regression)

```
0.9979305857383898
Trained slope: 0.49394455552101135
Trained intercept: 0.322654664516449
```

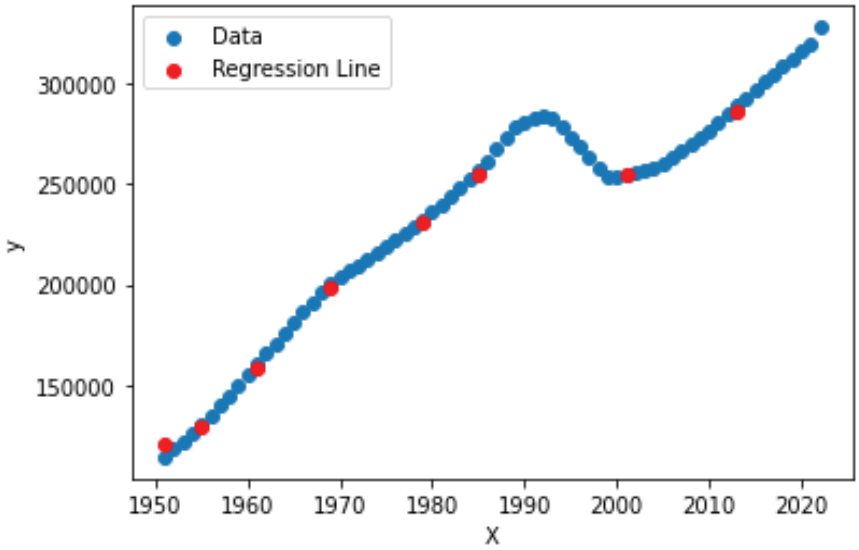

Figure 4 – Visualization of data prediction (Random Forest)

As can be seen from Figure 3 and Figure 4, there is a clear difference between the results of the two models. While Linear Regression stops working with non-linear growth, Random Forest predicts demographic growth or decline in a more accurate way regardless of the type of growth or decline. As a result of all the training and testing, the best result of prediction accuracy was:

- Linear Regression: 88%
- Random Forest: 99.7%

During the research, we also searched for similar solutions and found a couple of similar solutions. Namely [19] "worldometers" and [18] "macrotrends". However, these web applications have a slightly different purpose; instead of providing an extremely accurate guess at the demographics of one country or region, they provide a general overview and provide basic statistics of countries in different categories. Whereas our goal is more accurate in predictions and a more detailed solution. Therefore, direct competition in use in the market should not arise.

6 Conclusion

To summarize, we can say that the use of neural networks in working with data analysis, moreover, in creating predictions from big data is more than relevant, and every day there are more and more works on similar topics.

The article substantiates the advantages of using neural networks in various fields, provides and discusses other analogues of the use of neural networks in different fields, and proposes a solution to the problem of manual data analysis. It is obvious that the optimization of such expensive processes will bring benefits both in terms of resource and time costs, which in turn will improve the availability of market analysis for various businesses and governments. By improving the availability of market analysis, we receive more thoughtful business plans and social actions on the part of the government, which in turn will improve the quality of life of citizens.

The article also discusses the types of neural network models, their features and examples of use. Moreover, more successful cases of using different models depending on the data are discussed. For example, when training a neural network model to predict demographic growth, two different models were considered, namely Linear regression and Random Forest.

In the future, this study could be improved several times and the result of the study could be made more reasonable; for this we need to collect a new database that includes data on demographic development, economic development, social development and other parameters of the country over a sufficiently long period of time and based on train a new neural network with this data. Which is our goal soon.

References

1. Gong, N. Jia, Q. Zhang and L. Yu, "A Deep Learning Framework Using App Usage Record to Predict Demographic Information," *2021 International Conference on Computer Engineering and Application (ICCEA)*, Kunming, China, 2021, pp. 390-394, doi: 10.1109/ICCEA53728.2021.00083.

2. S. Batool, J. Rashid, M. W. Nisar, J. Kim, T. Mahmood and A. Hussain, "A Random Forest Students' Performance Prediction (RFSPP) Model Based on Students' Demographic Features," *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, Karachi, Pakistan, 2021, pp. 1-4, doi: 10.1109/MAJICC53071.2021.9526239.

3. S. Yarushev and A. Averkin, "Modular Neural Networks in Time Series Modelling," *2016 Fifteenth Mexican International Conference on Artificial Intelligence (MICAI)*, Cancun, Mexico, 2016, pp. 63-68, doi: 10.1109/MICAI-2016.2016.00018.

4. X. Zhang, L. Zuo, B. Yang, S. Chen and B. Li, "A Convolutional Neural Network with Background Exclusion for Crowd Counting in Non-uniform Population Distribution Scenes," *2021 IEEE 13th International Conference on Computer Research and Development (ICCRD)*, Beijing, China, 2021, pp. 16-20, doi: 10.1109/ICCRD51685.2021.9386369.

5. J. Wu, W. Qu, H. Yu, Y. Zhou and Z. Cui, "A Novel Crowd Counting Method via Deep Convolutional Neural Network," *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Tianjin, China, 2019, pp. 162-167, doi: 10.1109/SmartIoT.2019.00033.

6. J. Fang, "Crowd Counting and Density Estimation Algorithm Based on Deep Learning," *2023 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2023, pp. 01-06, doi: 10.1109/ICDSNS58469.2023.10245292.

7. K. Yang, T. Liu, Z. Chu and L. Sun, "Population counting method based on semantic perceptual network," *2020 39th Chinese Control Conference (CCC)*, Shenyang, China, 2020, pp. 6475-6479, doi: 10.23919/CCC50068.2020.9188885.

8. C. G. Raju, V. Amudha and S. G, "Comparison of Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy," *2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)*, Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICONSTEM56934.2023.10142495.

9. J. W. Kelly, A. D. Degenhart, D. P. Siewiorek, A. Smailagic and W. Wang, "Sparse linear regression with elastic net regularization for brain-computer interfaces," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, 2012, pp. 4275-4278, doi: 10.1109/EMBC.2012.6346911.

10. J. Liang, "Multivariate linear regression method based on SPSS analysis of influencing factors of CPI during epidemic situation," *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, Chongqing, China, 2020, pp. 294-297, doi: 10.1109/ICEMME51517.2020.00062.

11. Y. Feng and S. Wang, "A forecast for bicycle rental demand based on random forests and multiple linear regression," *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, China, 2017, pp. 101-105, doi: 10.1109/ICIS.2017.7959977.

12. Y. Zhai and X. Zheng, "Random Forest based Traffic Classification Method In SDN," *2018 International Conference on Cloud Computing, Big Data and Blockchain (ICCBB)*, Fuzhou, China, 2018, pp. 1-5, doi: 10.1109/ICCBB.2018.8756496.

13. L. Wei, "Genetic Algorithm Optimization of Concrete Frame Structure Based on Improved Random Forest," *2023 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, Marseille, France, 2023, pp. 249-253, doi: 10.1109/ICEDCS60513.2023.00051.

14. D. Yuan, J. Huang, X. Yang and J. Cui, "Improved random forest classification approach based on hybrid clustering selection," *2020 Chinese Automation Congress (CAC)*, Shanghai, China, 2020, pp. 1559-1563, doi: 10.1109/CAC51589.2020.9326711.

15. A. L. Latifah, A. Shabrina, I. N. Wahyuni and R. Sadikin, "Evaluation of Random Forest model for forest fire prediction based on climatology over Borneo," *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Tangerang, Indonesia, 2019, pp. 4-8, doi: 10.1109/IC3INA48034.2019.8949588.

16. Dataset about the demography of Kazakhstan, URL: https://old.stat.gov.kz/official/industry/61/statistic/6

17. Dataset about the demography of Kazakhstan, and demography of the Central Asian counties, URL: https://data.worldbank.org/

18. Dataset about the demography of Kazakhstan and example of alternative solution, URL: https://www.macrotrends.net/countries/KAZ/kazakhstan/population

19. Example of alternative solution, URL: https://www.worldometers.info/

Information about authors:

Baglan Muratbek (corresponding author) – Master student of the International Information Technology University, Faculty of Computer Technologies and Cybersecurity (Almaty, Kazakhstan, e-mail: 36138@iitu.edu.kz).

Gulnara Bektemisova – Candidate of Technical Sciences, Associate professor of the Department of Computer Engineering, International Information Technology University (Almaty, Kazakhstan, e-mail: g.bektemisova@iitu.edu.kz).

Submission received: 27 February, 2024.
Revised: 17 March, 2024.
Accepted: 19 March, 2024.