Methodology for Predicting Time Processes using Neural Technologies on the example of the Spread of Covid-19 in the Russian Federation and Moscow

Aduard Dadyan*

Senior Lecturer, Department of Data Analysis and Machine Learning Financial University under the Government of the Russian Federation, Moscow, Russia.

*Corresponding Author: Aduard Dadyan

Email: dadyan60@yandex.ru

Abstract

This paper solves the problem of predicting Covid-19 diseases in Moscow and the Russian Federation using neural networks. This approach is useful in cases where it is necessary to overcome difficulties related to non-stationarity, incompleteness, unknown distribution of data, or when statistical methods are not completely satisfactory. The problem of forecasting is solved using the analytical platform Deductor Studio, developed by specialists of Intersoft Lab of the Russian Federation. When solving the problem, we used mechanisms for clearing data from noise and anomalies, which ensured the quality of building a forecast model and obtaining forecast values for tens of days ahead. The principle of time series forecasting was also demonstrated: import, seasonal detection, cleaning, smoothing, building a predictive model, and predicting diseases with Covid-19 in Moscow and the Russian Federation using neural technologies for twenty days ahead.

Keywords: Time series; Forecasting; Neural network; Data preprocessing; Training and control samples; Pandemic; Coronavirus infection; Russia; Moscow; Deductor studio; Data clearing; Partial processing; Spectral processing; Autocorrelation; Sliding window.

Introduction

The task of predicting time-dependent processes (TDP) has been and remains relevant, especially in recent years, when powerful tools for collecting and processing information have appeared. Forecasting TDP is an important scientific and technical task, as it allows you to predict the behavior of various factors in environmental, economic, social, and other systems. The development of forecasting as a science in recent decades has led to the creation of many models and methods, procedures and methods of forecasting that have different values. According to estimates of foreign and domestic experts in the field of forecasting, there are already more than a hundred forecasting methods, which raises the problem of choosing methods that would give adequate forecasts for the processes or systems under study. Strict statistical assumptions about the properties of TDP often limit the capabilities of classical forecasting methods. The use of neural networks (NN) in this task is due to the presence of complex patterns in most TDP that are not detected by known linear methods. Neural network methods of information processing began to be used several decades ago. Over time, interest in neural network technologies faded, then revived again. This variability is directly related to practical research results. Today, the capabilities of neural network technologies are used in many branches of science, ranging from medicine and astronomy, to computer science and Economics. The ability of a neural network to process information in various ways stems from its ability to generalize and identify hidden dependencies between input and output data. The great advantage of neural networks is that they are capable of learning and generalizing the accumulated knowledge.

The goal of any forecast is to create a model that allows you to investigate the future and assess trends in a factor. The quality of the forecast in this case depends on the presence of a background variable factor, the measurement error of the value in question, and other factors. Formally, the prediction problem is formulated as follows: find a function f that allows you to estimate the value of the variable x at time (t + d) from its N previous values, so that $x(t + d) = f(x(t), x(t-1)..., x(t-N+1))$.

Usually d is assumed to be equal to one, i.e. the function f predicts the next value of x. TDP is a sequence of observed attribute values ordered at non - random time points. It is already clear that the coronavirus pandemic has affected the economies of all countries of the world. In this situation, Russia is at the very epicenter of the crisis. On the one hand, there is an urgent need to address the problems caused by the reduction in consumption of almost all resources that form the basis of the country's export potential. On the other hand, it is necessary to solve the problems of stimulating the production and consumption of goods and services within the country. In this situation, it is important to obtain forecast values for the process of infection with the Covid-19 coronavirus in Russia in General and in Moscow by date. From the point of view of data analysis technologies, forecasting can be considered as determining some unknown quantity from a set of associated values. Therefore, forecasting is performed using data mining tasks such as regression, classification, and clustering. Coronavirus can be viewed as a time-distributed process. The data collected and used to develop forecasts are most often time series, i.e. they describe the development of a process over time. Therefore, forecasting in the field of coronavirus is usually associated with time series analysis.

Due to a number of advantages, the analytical neural network platform Deductor Studio, developed by BaseGroup Labs, was chosen as a tool for predicting coronavirus in the Russian Federation and Moscow in modern conditions (www.basegroup.ru, Ryazan, Russian Federation). A few words about this software product. Deductor Studio provides the development of deep data analysis systems that cover data collection, consolidation, data cleaning, modeling, forecasting, and visualization. The Deductor Studio platform is designed to solve a wide range of tasks related to processing structured data presented in the form of tables. These tables form a sample intended for training a neural network, forming an expert system for the subject area under study, and predicting processes that depend on various factors. At the same time, the scope of the system can be almost any - the mechanisms implemented in the system are successfully used in financial markets, insurance, trade, telecommunications, industry, medicine, logistics and marketing tasks, and many others. Today, the whole world is working on the problem of creating mechanisms for detecting the spread of Covid-19 to eliminate it. Forecasting would help solve this very serious problem

Objective: To propose a reasonable method for predicting the number of covid-19 coronavirus infections by date using neural technologies on the example of Moscow and the Russian Federation.

Neural networks in Covid-19 prediction: A brief overview Literature Review

Predicting the spread of coronavirus is important in developing protective measures and behavioral measures for the population. The problem with modeling such a system is that every day COVID-19 and the number of new potential cases cannot be determined in a simple mathematical equation. There are many reasons for such problems. The spread of human filaments generally depends on various features, depending on both human behavior and the biological structure of the coronavirus itself. In any case, research needs to be done to biologically describe the coronavirus in order to develop a medical treatment, as well as to model the spread that will help prevent new cases and focus on the places with the greatest potential needs. According to authors Michał Wieczorek, Jakub Siłka, and Marcin Woźniak [1], predicting the spread of coronavirus is very Important for active action planning. Unfortunately, coronaviruses are not easy to control, as the speed and reach of their spread depend on many factors, from environmental to social. The article by these authors presents the results of research on the development of a neural network model for predicting the spread of COVID-19. The prediction process itself is based on the classical approach of training a neural network with a deep architecture using the NAdam training model. For training, the authors of the article used official data from government and open repositories.

The COVID-19 pandemic has challenged global science. The international community is trying to find, apply or develop new methods for the diagnosis and treatment of patients with COVID-19 as soon as possible. The work of authors Shayan Hassantabara, MohsenA hmadib, Abbas Sharifie [2] is devoted to the use of deep learning to identify and diagnose patients with COVID-19 using x-ray images of the lungs. To diagnose the disease, they presented two algorithms: deep neural network (DNN) on the fractal feature of images and neural network (SNN) methods using lung images directly. The results of the authors ' work show that the presented methods allow detecting infected areas of the lungs with high accuracy-83.84%.

Several works are devoted to the detection of COVID-19 disease using neural network technologies, among which we can distinguish [3-5]. The authors of these papers propose a method based on a convolutional neural network (CNN) developed using the EfficientNet architecture for automated COVID-19 diagnostics. The architecture of an automated medical diagnostics system is also proposed to support healthcare professionals in the decision-making process for the diagnosis of diseases.

Several important models have been introduced in recent months. In [6], machine learning was applied to evaluate how the flash of this stream will take place. However, predicting the situation in the case of COVID-19 is not easy, since there are many factors that determine rapid changes [7]. Therefore, many approaches have been used to help.
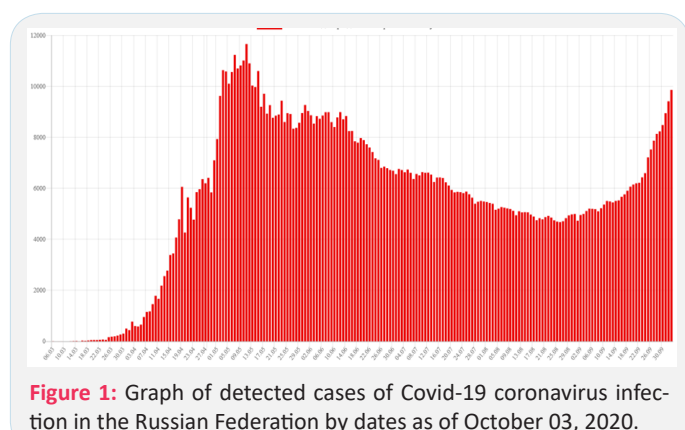
In [8], the flow prediction was performed using a mathematical model that evaluated undetected infections for the Chinese region. Sometimes even very simple techniques are used. When a solution is needed immediately, we can start predicting based on preprocessing, in which some cases are simply removed for the applied model on the Euclidean network [9]. In Japan, prognostic models also evaluated the first symptoms of the disease [10]. One of the first models presented for Italy was the use of the Gauss error function and Monte Carlo simulation on registered cases [11]. In addition, stochastic predictors provide potential help in the early days when not much data is available for machine learning approaches [12]. Such stochastic models also seem to work even for very large societies, such as India [13]. Therefore, when artificial intelligence is applied in the first days of forecast periods, the results are mostly related to a single region or country. One of the first approaches for China was

presented in [14]. An interesting discussion of the principles of using mathematical modeling was presented in [15]. Some methodologies not only predict the number of new cases, but also make some assumptions about the growth dynamics [16]. There are many sources of information for predicting the situation. As reported in [17], social networks can bring valuable information not only about confirmed cases of the disease, but also about further spread. The relationship between new cases and the rate or coverage of growth can be transformed into a prediction elsewhere, as shown in [18]. this transfer of knowledge to model another region was carried out between Italy and Hunan province in China. The case of the ship "Diamond Princess" was discussed in [19]. There are also models that assess the situation in larger regions or in more than one country. In [20] and [21], an applied forecasting model was defined for working with data from China, Italy, and France. Some models only consider the total number of cases worldwide as a whole [22].

The model proposed in [23] is a complex solution. The proposed neural network architecture was developed for flexible forecasting of new cases in various countries and regions of the world. The architecture consists of seven layers, and the output predicts the number of new cases. Today, models for predicting inflation based on artificial neural networks in most studies are evaluated both in terms of the accuracy of the forecasts themselves, and in comparison, with regression models. Unfortunately, it is almost impossible to achieve high accuracy in predicting monthly inflation based on neural networks, but their superiority over regression models, which also did not succeed in this task, is demonstrated in the short and especially in the long term.
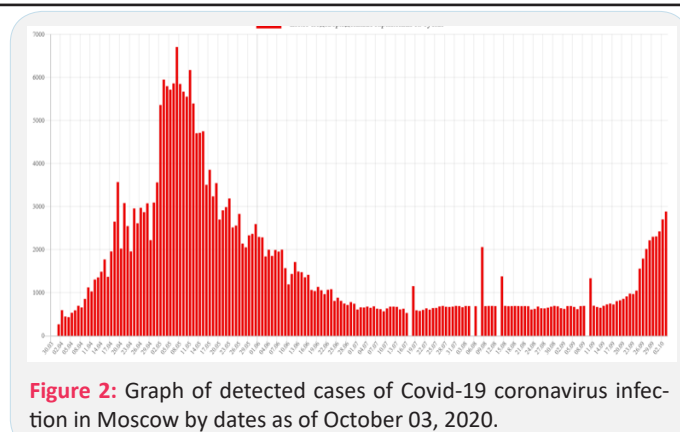
General scheme for building an analytical solution for forecasting covid-19

The solution of the problem of forecasting with the help of a trained neural network assumes, first of all, the availability of statistical data on the spread of this disease by day, provided by Rospotrebnadzor of the Russian Federation (https://coronavirus-monitor.info/country/russia/moskva/) for the Russian Federation (Figure 1) and Moscow (Figure 2).

Figure 1: Graph of detected cases of Covid-19 coronavirus infection in the Russian Federation by dates as of October 03, 2020.

The statistical data obtained in the form of a time series require significant processing in order to form a training sample of the neural network and obtain the necessary data for the operation of the neural network dataset. This process usually includes the following steps:

• Time series adjustment – smoothing and removing anomalies.

Figure 2: Graph of detected cases of Covid-19 coronavirus infection in Moscow by dates as of October 03, 2020.

• Study of the time series, highlighting its components (trend, seasonality,

cyclicity, noise) – auto correlation analysis.

• Data processing using the sliding window method.

• Data processing using a multi-layer neural network, neural network training

• Selecting the appropriate forecasting method.

• Assessment of the accuracy of forecasting and the adequacy of the chosen forecasting method.

The analysis of the above points and numerous experiments allowed us to propose a General scheme for analytical processing of statistical source data in order to obtain a dataset for a neural network with subsequent training of the neural network and forecasting. The block diagram of the dataset generation algorithm for the neural network and predicting cases of Covid-19 coronavirus infection is shown in figure 3.

Figure 3: Block diagram of the dataset generation algorithm for the neural network and prediction of cases of Covid-19 coronavirus infection.
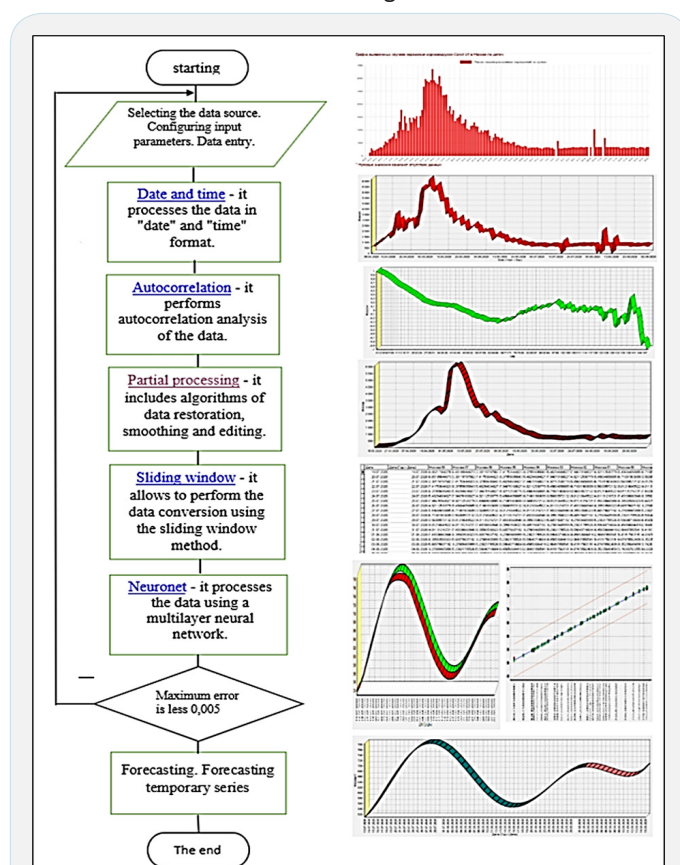
Formation of dataset for neural network

Adjustment of time series

Graphs of detected cases of Covid-19 coronavirus infection in the Russian Federation and Moscow as of September 30, 2020 were shown in figures 1 and 2. to get a forecast in the required scale, you need to change the time scale of the data series in order to optimize it for further processing procedures. If you submit data by day to the input of the predictive model (neural network, linear model), then the forecast will be by day. If you previously convert the data to weekly intervals, then the forecast will be based on weeks. In addition, the date can be converted to a number or string, if necessary, for further processing.

In our case, we proceeded from the need to get a forecast by day, so by performing the necessary transformations of the source data to the " date "scale: Year+Day" we get the corresponding two graphs of the source data for the Russian Federation and Moscow in the specified scale (Figure 4 & 5):
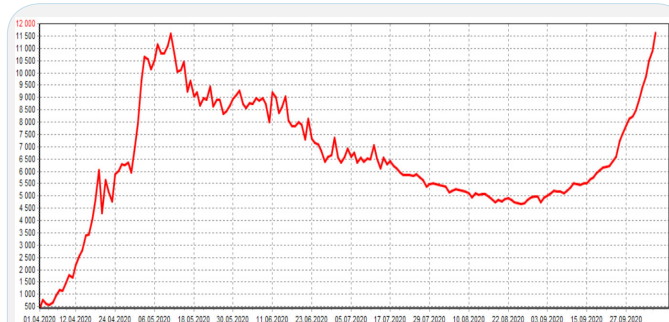

Figure 4: Graph of detected cases of Covid-19 coronavirus infection in Russia by dates as of September 30, 2020 on the "date: Year+Day" scale».

Figure 5: Graph of detected cases of Covid-19 coronavirus infection in Moscow by dates as of September 30, 2020 on the "date: Year+Day" scale».

Smoothing and removal of anomalies: Spectral data processing

The purpose of spectral processing is to smooth ordered data sets using a wavelet or Fourier transform. The principle of such processing is to decompose the original time series function into basic functions. It is most often used for preliminary data preparation in forecasting tasks.

At the "Spectral processing" step of the processing wizard, the "Wavelet transform" method was selected, and the decomposition depth and order of the wavelet were set. The depth of decomposition determines the "scale" of the parts to be filtered out: the larger this value, the «larger» parts in the source data will be discarded. If the parameter values are large enough (about 7-9), the data is not only cleared of noise, but also smoothed (sharp outliers are "cut off"). Using too large values of the decomposition depth can lead to loss of useful information due to too much "coarsening" of the data. The order of the

wavelet determines the smoothness of the reconstructed data series: the lower the parameter value, the more pronounced the "outliers" will be, and, conversely, if the parameter values are large, the "outliers" will be smoothed.

Figures 6 and 7 show graphs of the results of smoothing and removing anomalies using spectral processing using the "Wavelet transform" method and setting the average values of the parameters of this method.
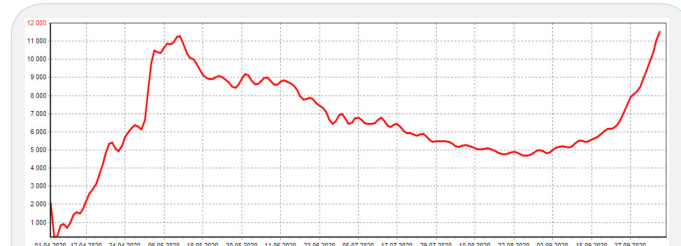

Figure 6: Graph of detected cases of Covid-19 coronavirus infection in Russia smoothed using spectral processing using the "Wavelet transform" method.

Figure 7: Graph of detected cases of Covid-19 coronavirus infection in Moscow smoothed Using spectral processing using the "Wavelet transform" method

Autocorrelation analysis of data

The purpose of autocorrelation analysis is to find out the degree of statistical dependence between different values (counts) of a random sequence formed by the data sample field. In the process of autocorrelation analysis, correlation coefficients (a measure of mutual dependence) are calculated for two sample values that are separated by a certain number of samples, also called lag. The set of correlation coefficients for all lags is an autocorrelation function of the series (ACF):

R(t) = corr(X(t), X(t+k)), where k > 0 is an integer (lag).

The behavior of the ACF can be used to judge the nature of the analyzed sequence, i.e. the degree of its smoothness, and the presence of periodicity (for example, seasonal) or a trend.

For k = 0, the autocorrelation function will be maximal and equal to 1.as the number of lags increases, i.e. the distance between two values for which the correlation coefficient is calculated increases, the ACF value will decrease due to a decrease in the statistical interdependence between these values (the probability of occurrence of one of them less affects the probability of occurrence of the other). At the same time, the faster the ACF decreases, the faster the analyzed sequence changes. Conversely, if the ACF decreases slowly, then the corresponding process is relatively smooth. If there is a trend in the original sample (a smooth increase or decrease in the values of the series), then a smooth change in the ACF will also occur. If there are seasonal fluctuations in the original data set, the ACF will also have periodic spikes.

Figures 8 and 9 show graphs of autocorrelation functions of

detected cases of Covid-19 coronavirus infection in Russia and Moscow, respectively. Using these graphs, you can visually determine the presence of trends in the first and second curves with lags of 65 and 51, respectively.
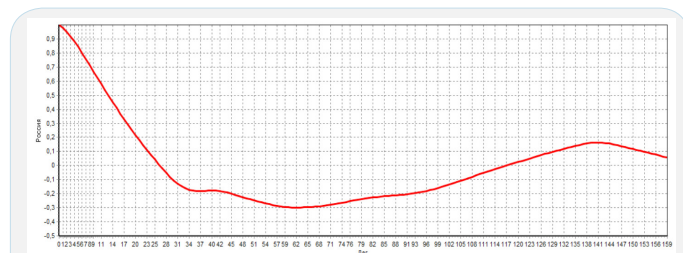

Figure 8: Graph of autocorrelation functions of detected cases of Covid-19 coronavirus infection in Russia.

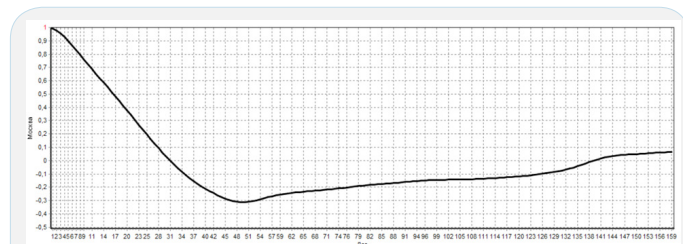**Data processing by a sliding window**

Figure 9: Graph of autocorrelation functions of detected cases of Covid-19 coronavirus infection in Moscow.

Data processing using the sliding window method is used for preprocessing data in forecasting tasks, when the input of the neural network requires feeding the values of several adjacent samples of the original data set. The term "sliding window" reflects the essence of processing – a certain continuous piece of data is allocated, called a window, and the window, in turn, moves, "slides", over the entire source data set. This operation results in a sample where each record contains a field corresponding to the current sample (it will have the same name as in the original sample), and to the left and right of it are fields containing samples shifted from the current sample to the past and to the future, respectively.

Processing by the sliding window method has two parameters: the depth of immersion – the number of counts in the "past" and the forecast horizon – the number of counts in the "future".

The paper used the sliding window method to smooth out the graphs of detected cases of Covid-19 coronavirus infection in Russia and Moscow with diving depths of 65 and 51, respectively, using spectral processing. The forecast horizon in both cases was taken equal to one. As a result, two datasets were obtained for training the neural network.

The neural network training, Data processing using a multilayer neural network

In this mode, the processing wizard allows you to set the structure of the neural network, determine its parameters, and train it is using one of the algorithms available in the system.

Configuring and training a neural network consists of the following steps:

1. Configure field assignments. Here you need to determine how the fields of the source data set will be used when training the neural network and working with it in practice.

2. Setting the normalization field. The purpose of normaliz-

ing field values is to transform data to the form that is most suitable for processing by means of a neural network.

3. Setting the training sample. Here you need to split the training sample for building a model based on a neural network into two sets – training and test. Training set – includes entries that will be used as input data, as well as the corresponding desired output values.

Test set – also includes records that contain input and desired output values, but are used not for training the model, but for testing its results.

4. Set up the structure of the neural network. At this step, parameters are set that determine the structure of the neural network, such as the number of hidden layers and neurons in them, as well as the activation function of neurons. In the "Neurons in layers" section, you can set the number of hidden layers, i.e. layers of the neural network located between the input and output layers.

5. The choice of algorithm and parameters training. At this step, we select the neural network training algorithm and set its parameters.

6. Setting the stop conditions of training. At this step, we set the conditions under which training will be terminated: the condition that the mismatch between the reference and real network output becomes less than the specified value, and we set the number of epochs (training cycles) after which training stops, regardless of the error value.

7. Start the learning process. At this step, we start the actual process of training the neural network.

8. Select the data display method. At this step, select the form in which the imported data will be presented. In our case, the following specialized Visualizers are interesting:

8.1. Contingency table or a scatterplot. The choice of the appropriate forecasting method consists in determining whether this method gives satisfactory forecast errors. In addition to calculating errors, their comparison is carried out in a special Visualizer – "scatter diagram". The scatter plot shows the output values for two sets of training samples (dataset) for Russia (figure 10) and Moscow (figure 11).
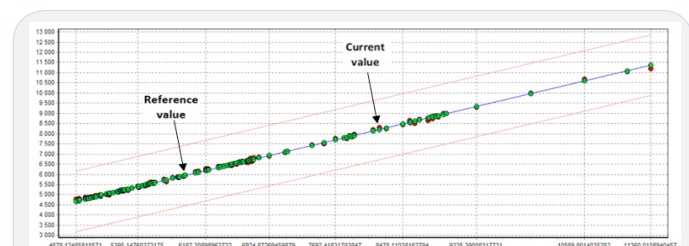

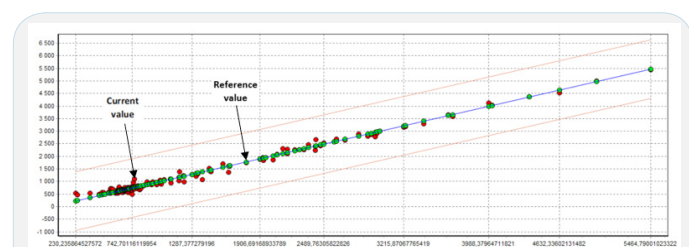**Figure 10:** Scatter diagram of a trained neural network for dataset Russia.

Figure 11: Scatter diagram of a trained neural network for dataset Moscow.

Coordinate the X — axis is the output value on the training sample (reference), and the Y - axis is the Value output calculated by the trained model using the same example. Straight diagonal line it is a reference point (a line of ideal values). The closer the point is to this line, the less model error.

The scatter plot allowed us to compare several models to determine which model provides the best accuracy on the training set.

8.2. Diagram. The diagram visually shows the dependence of the values of one field on another. The most used type of chart is a two - dimensional graph. The values of the independent column are plotted along the horizontal axis, and the corresponding values of the dependent column are plotted along the vertical axis.

After building a model for evaluating the quality of training, we present the data obtained in the form of diagrams for the current and reference values of the dataset for Russia (Figure 12) and Moscow (Figure 13).
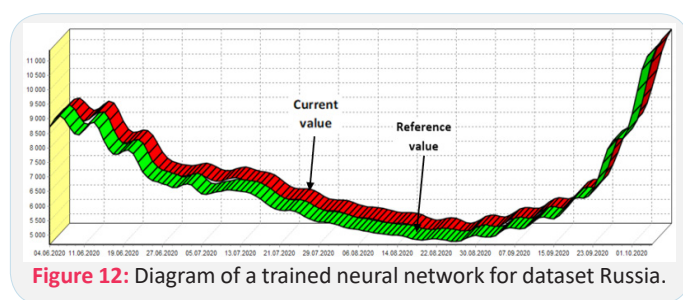

Figure 12: Diagram of a trained neural network for dataset Russia.
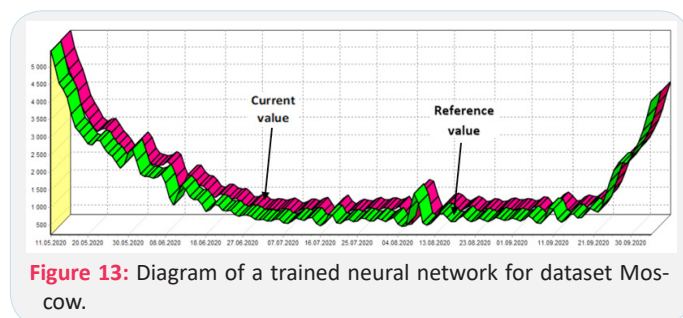

Figure 13: Diagram of a trained neural network for dataset Moscow.

Analysis of the scattering diagrams (Figures 10 and 11) and diagrams of the trained neural network for the dataset of Russia and Moscow (Figures 12 and 13) allows us to assert that the neural network was successfully trained with both the dataset of Russia and the dataset of Moscow.

Forecasting itself, research result. Points for discussion

Forecasting allows you to get a prediction of the values of a time series for the number of samples corresponding to the specified forecast horizon.

What is the maximum forecast horizon? The following rule is recommended: the amount of statistical data should be 10-15 times greater than the forecast horizon [24]. This means that in our case, the maximum forecast horizon can be 20-25 days.

When performing the actual forecast, we pre-configure several fields: forecast horizon (set 20 days), request the "forecast step" and "source data" fields, and set color and scale parameters. Adding the" forecast step "field (check the box) allows you to add an additional" forecast Step " field to the resulting selection, which will indicate the number of the forecast step that resulted in it for each record.

"Source data" – selecting this check box allows you to include in the resulting selection not only those records that contain the predicted values, but also all those that contain the source data. In this case, the records containing the forecast will be located at the end of the resulting selection.

The final graphs for predicting the number of COVID-19 infections by date using neural technologies are shown in figures 14 (Russia) and 15 (Moscow).
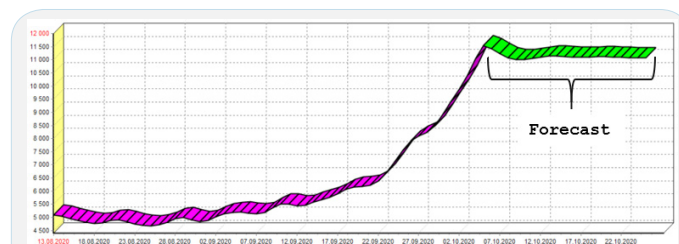

Figure 14: Graph for predicting the number of COVID-19 infections by date using neural technologies (Russia).
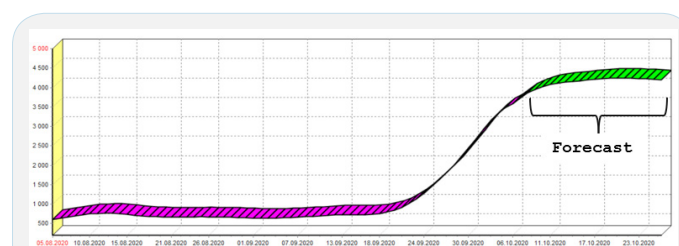

Figure 15: Graph for predicting the number of COVID-19 infections by date using neural technologies (Moscow).

The proposed model for predicting the number of COVID-19 infections by date using neural technologies, built once, cannot "work" indefinitely. There are new data on the number of infections in Russia and Moscow. Therefore, the model should be periodically reviewed and retrained.

The forecast error

A forecast error is the difference between the actual value of yt and its forecast yt* at time t. Deviations and errors of the forecast and fait accompli calculated using standard expressions are used to evaluate accuracy. The most common are the following: Mean Absolute Deviation (MAD):

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|\,y_t - y_t^*\,|.$$

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_t - y_t^*\right)^2.$$

Mean absolute percentage error (Mean Absolute Percentage Error, MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\,y_t - y_t^*\,|}{y_t},$$

where n is the number of training examples.

In our case, using the data values shown in figures 1, 2, 14 and 15 in the calculations, the average absolute error was 2.32%. The choice of the appropriate forecasting method consists in determining whether this method gives satisfactory forecast errors. In addition to calculating errors, their comparison is carried out in a special Visualizer – "scatter diagram" (Figures 10

and 11). The scatterplot displays the output values for each of the training sample examples, whose X - axis coordinate is the output value in the training sample (reference), and Y – axis coordinate is the output value calculated by the trained model in the same example. A straight diagonal line is a reference point (a line of ideal values). The closer the point is to this line, the smaller the model error. The scatter plot is useful when comparing multiple models. It is often enough to look at the spread of points from the diagonal line to determine which model provides the best accuracy on the training set. The work performed by the author on estimating the scatter diagrams of various forecasting models showed that the neural network provides the smallest spread within the boundaries of a given error.

Conclusion

This paper solves the problem of predicting Covid-19 diseases in Moscow and the Russian Federation using neural networks. This approach is useful in cases where it is necessary to overcome difficulties related to non-stationarity, incompleteness, unknown distribution of data, or when statistical methods are not completely satisfactory. The forecasting problem is solved using the Deductor Studio analytical platform developed by BaseGroup Labs (www.basegroup.ru, Russian Federation, city of Ryazan). When solving the problem, we used mechanisms for clearing data from noise and anomalies, which ensured the quality of building a forecast model and obtaining forecast values for tens of days ahead. The principle of time series forecasting was also demonstrated: import, seasonal detection, cleaning, smoothing, building a predictive model and predicting Covid-19 diseases in Moscow and the Russian Federation using neural technologies for twenty days ahead.

References

1. Michał Wieczorek, Jakub Siłka, Marcin Woźniak. Neural network powered COVID-19 spread forecasting model Chaos, Solitons & Fractals. November 2020; 140: 110203.

2. Shayan Hassantabara, MohsenA hmadib, Abbas Sharific. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches Chaos. Solitons & Fractals. November 2020; 140: 110170.

3. Gonçalo Marques, Deevyankar Agarwal, Isabelde la Torre Diez. Automated medical diagnosis of COVID-19 through Efficient Net convolutional neural network. Applied Soft Computing. 2020; 96: 106691.

4. Linh T, Duong Phuong, Nguyen T, Claudio Di Sipio, Davide Di Ruscio. Automated fruit recognition using EfficientNet and Mix Net, Computers and Electronics in Agriculture. 2020; 171: 105326.

5. Nils Gessertab, Maximilian Nielsenbc, Mohsin Shaikhbc, René Wernerbc, Alexander Schlaefera. Skin lesion classification using ensembles of multi-resolution Efficient Nets with meta data. Applied Soft Computing. 2020; 7: 100864.

6. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. COVID-19 outbreak prediction with machine learning.

7. W.C. Roda, M.B. Varughese, D. Han, M.Y. Li. Why is it difficult to accurately predict the COVID-19 epidemic?. Infect Dis Model. 2020.

8. Ivorra B. Ferrández MR, Vela-Pérez M, Ramos A. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China Commun Nonlinear Sci Numer Simul. 2020; 88.

9. Biswas K, Khaleque A, Sen P. COVID-19 spread reproduction of data and prediction using a sir model on euclidean network.

10. Kuniya T. Prediction of the epidemic peak of coronavirus disease in Japan. J Clin Med. 2020; 9(3): 789.

11. Ciufolini I, Paolozzi A. Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a gauss error function and Monte Carlo simulations. Eur Phys J Plus. 2020; 135(4): 355.

12. He S, Tang S, Rong L. A discrete stochastic model of the COVID-19 outbreak: forecast and control. Math Biosci Eng. 2020; 17: 2792-2804.

13. Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and regression model based COVID-19 outbreak predictions in India.

14. Hu Z, Ge Q, Jin L, Xiong M. Artificial intelligence forecasting of COVID-19 in China.

15. Jewell NP, Lewnard JA, Jewell BL. Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. JAMA. 2020; 323(19): 1893-1894.

16. Mandal M, Jana S, Nandi SK, Khatua A, Adak S, Kar T. A model-based study on the dynamics of COVID-19: prediction and control. Chaos Solitons Fractals. 2020; 109889.

17. Qin L, Sun Q, Wang Y, Wu KF, Chen M, Shia BC, et al. Prediction of number of cases of 2019 novel coronavirus (COVID-19) using social media search index. Int J Environ Res Public Health. 2020; 17(7): 2365.

18. Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, et al. Extended sir prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. Front Med. 2020; 7: 169.

19. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the diamond princess cruise ship: a data-driven analysis. Int J Infect Dis. 2020; 93: 201-204.

20. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals. 2020; 134: 109761.

21. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020; 12(3): 165.

22. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS ONE. 2020; 15(3): e0231236.

23. Michał Wieczorek, Jakub Siłka, Marcin Woźniak, Neural network powered COVID-19 spread forecasting model, Faculty of Applied Mathematics, Silesian University of Technology, Kaszubska Gliwice, Poland, 2020; 23: 44-100.

24. Paclín N, Oreshkov O. Business intelligence: from data to knowledge, Base Group Labs, LLC. Saint-Petersburg. 2018.