FOLIA GEOGRAPHICA • ISSN 1336-6157 (hard copy) • ISSN 2454-1001 (online)

SOLVING OF CLASSIFICATION PROBLEM IN SPATIAL ANALYSIS APPLYING THE TECHNOLOGY OF GRADIENT BOOSTING CATBOOST

Ruslan Z. SAFAROV [A], **Zhanat K. SHOMANOVA** [B*], **Yuriy G. NOSSENKO** [C], **Zharas G. BERDENOV** [D], **Zhuldyz B. BEXEITOVA** [E], **Adai S. SHOMANOV** [F], **Madina MANSUROVA** [G]

Received: January 19, 2019 | Revised: May 1, 2020 | Accepted: May 7, 2020
Paper No. 20-62/1-558

Abstract
In the paper two models of spatial analysis are considered. The models are dedicated for spatial analysis of ecological factors distribution, such as distribution of contaminant concentration on researched territory. The models are created using the method of machine learning – gradient boosting. In order to build the models we have used open source effective library CatBoost. Functions AUC and Accuracy were calculated for each model. MultiClass – integrated function of CatBoost library was used for loss minimization. For solving the problem, it was necessary to define affiliation of searched point from test dataset to one of four classes. This problem belongs to the type of classification, or rather multiclassification. As a result of the studies, an effective model was obtained that allows one to perform with sufficient accuracy the spatial forecast of the factor distribution at points and regions of the studied field with an unknown gradient value of this factor. This model works adequately with a training dataset of 0.5% of all analyzed information about the object.

Key words
Spatial analysis, gradient boosting, CatBoost, machine learning, neural networks, computer modeling, geoecological maps.

A L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
ruslanbox@yandex.ru
B* Pavlodar State Pedagogical University, Pavlodar, Kazakhstan
zshoman@yandex.ru (corresponding author)
C Innovative University of Eurasia, Pavlodar, Kazakhstan
nosenko1980@yandex.ru
D L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan
berdenov-z@mail.ru
E Eurasian Center of Innovative Development DARA, Nur-Sultan, Kazakhstan
femida.pvl@gmail.com
F Nazarbayev University, Nur-Sultan, Kazakhstan
adai.shomanov@nu.edu.kz
G Al-Farabi Kazakh National University, Almaty, Kazakhstan
mansurova.madina@gmail.com

INTRODUCTION

Geographical approaches to understanding origins of environmental problems are quite often used (Ilieş et al., 2017). Currently, an urgent problem in ecology is the study of the spatial distribution of various factors (Beketova et al., 2019; Gallego-álvarez et al., 2014; Glass et al., 1995; Jerrett et al., 2003; Ramazanova et al., 2019). These factors can be divided into groups:

1) natural factors (temperature, pressure, soil salinity, biodiversity, etc.);
2) technogenic factors (concentration of pollutants, soil degradation, drying out of water bodies, etc.);
3) factors associated with human life (the spread of diseases, mortality and fertility, population, etc.).

A huge amount of work has been devoted to studying the influence and distribution of these factors. (Azhayev et al., 2020; Berdenov et al., 2017; T. Z. Li et al., 2009; Matlovič and Matlovičová, 2012, 2020; Ribeiro et al., 2013; Safarov et al., 2019; Sexton et al., 2002; Vicente et al., 2014). The knowledge of the spatial distribution of such factors allows us to draw conclusions about the degree of influence, dynamics and directions of distribution, helps to identify the main trends in the change in the situation associated with the influence of the factor on the environment. In addition, these data allow us to make predictions of the situation over time. (Berrocal et al., 2010; Ilieş et al., 2019; Mihincău et al., 2019; Paci et al., 2013; Sahu et al., 2009).

When implementing spatial analysis, the problem of lack of data, or uneven distribution of data, often arises. For example, Meyer et al. (Meyer et al., 2019) presented a map of the location of weather stations in Antarctica (Figure 1). The figure shows that the weather stations are located unevenly on the mainland and there are not many data sources. When the question arises of obtaining extended data at other points located not near the stations, it is necessary to consider various approaches of spatial forecasting. The most common computer modeling methods are statistical methods of interpolation and extrapolation. Currently, computer models obtained through the application of machine learning approaches, in particular neural networks, are gaining more and more popularity (R. Li et al., 2020; Murtagh et al., 2000; Shoji and Kawakami, 2006).

The problem of spatial prediction of the distribution of factors is the significant non-linearity of the existed dependencies (Casati et al., 2004; Madadgar and Moradkhani, 2014). Non-linearity is due to the specificity of the task, the influence of the landscape, weather conditions, the mixed sources of the factor, etc. In such analytical problems as forecasting, including spatial forecasting, decisions are based on known available data and the more known data we have, the more detailed the forecast can be. But, often, as in the example of Antarctica, there is not much data. This is due to the technical complexity or cost of analysis, the inaccessibility of locations for sampling, lack of human resources to perform monitoring analytical work.

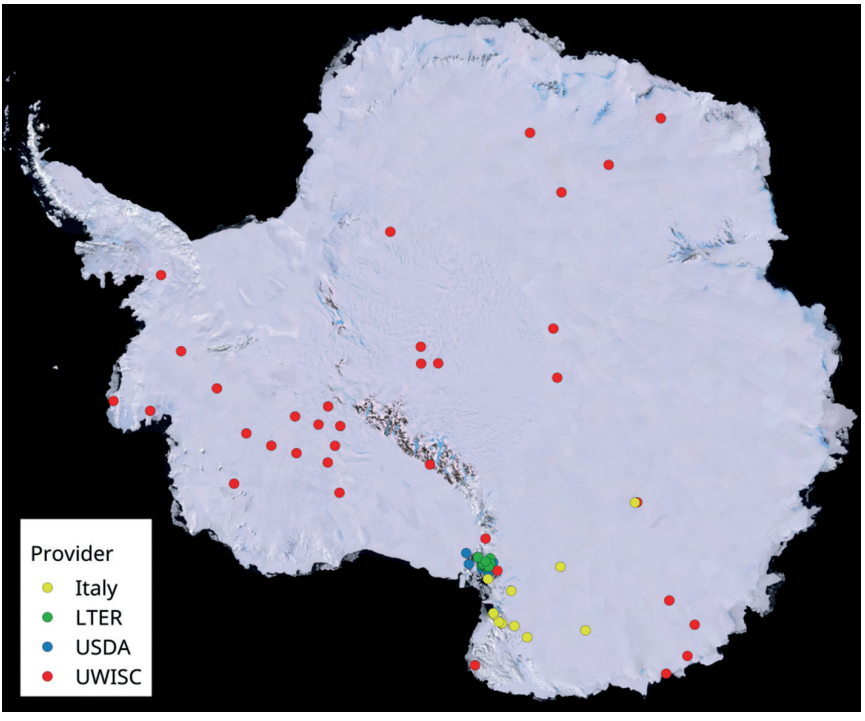

Figure 1
Location of weather stations in Antarctica (according to the Antarctic Meteo-Climatological Observatory of Italy, the Long Term Ecological Research (LTER) programme, the United States Department of Agriculture (USDA), the Antarctic Meteorological Research Center at the University of Wisconsin (UWISC))
Source: *Meyer et al., 2019*

Classical interpolation methods can be used to build good enough forecasts only if the set of known data is evenly distributed in space, in compliance with a certain mesh structure. In cases where this is not possible, it is recommended to use more complex computer models, for example, using neural network technologies or gradient boosting (GB) (Abadi et al., 2016; Yuan, 2015).

Along with neural networks, gradient boosting technology has confidently take the place in solving complex nonlinear problems. Gradient boosting is a class of algorithms that works best on heterogeneous data, that is, in cases where an object is described by a set of different data: temperature, height, weight, age, other information (Körner et al., 2018; Xiong et al., 2018). Therefore, gradient boosting is widely used to solve various problems in medicine, in the field of economics, in marketing, in industry, when creating artificial intelligence, when performing search queries, etc. Also, unlike neural networks, the use of gradient boosting is much simpler. Since for the successful use of a neural network it is necessary to be able to design

a network architecture that can effectively solve the problem. In the case of using gradient boosting, the task is simplified, since we feed the available data, and at the output we get a working model. In other words, gradient boosting works like a "black box", the model is automatically generated based on the algorithms embedded in the library (Natekin and Knoll, 2013; Prokhorenkova et al., 2018).

Gradient boost models work on the basis of the so-called "decision trees" (Figure 2). The model is built in parts. First, one decision tree is built, which may not produce very good quality. Then another decision tree is built, which improves the quality. This happens hundreds, thousands or tens of thousands of times, after which a large model is obtained that can already find complex patterns within the data (Natekin and Knoll, 2013).

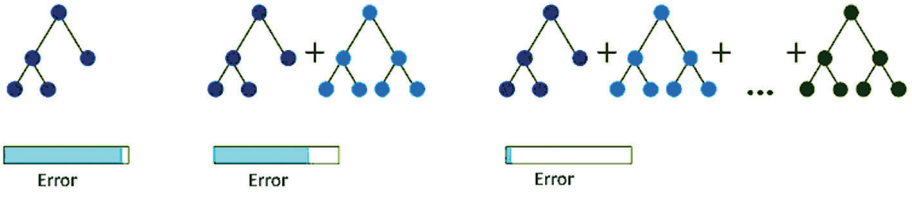

Figure 2
The principle of constructing a gradient boosting model based on decision trees

In this article, we present the results of a study of solving the problem of spatial forecasting of the distribution of a factor classified into several groups. For solving the problem, the machine learning method - gradient boosting was used. The proposed model can be applied, for example, to the study of the spatial distribution of the pollutant (for example, heavy metals in the soil), divided into hazard classes (I-IV hazard class) or into classes by relation the concentration of pollutant to MPC (class 1 - pollutant concentration below MPC, Grade 2 - at the MPC level, Grade 3 - above the MPC). There can be many approaches to classification; the type of classification is selected when solving an applied problem.

DATA AND METHODS

1. Object of study

As the object of study, we take an image (Figure 3), which rather well imitates the spatial distribution of the concentration factor, for example, the distribution of the concentration of the pollutant. The image is made in four gradients of gray. So, it is known that in computer graphics the gradation of gray (from white to black) lies in the range from 0 to 255. Four gradients are included in the given image: 0 - black, 85 - dark gray, 170 - light gray, 255 - white. Image dimensions 197x256 pixels, total pixels - 50432.

Figure 3
Model image taken as a data source for solving an analytical problem

2. Gradient boosting library

To solve this problem, we used the CatBoost library, an open source product from Yandex. The CatBoost library is distinguished by the fact that it has a wide functionality for studying the learning process, a wide range of functions for assessing the effectiveness of model training, and good tools for visualizing learning results (Prokhorenkova et al., 2018). In addition, since gradient boosting models can work with heterogeneous data, there is no need to implement procedures of normalization - renormalization data. The model works with its own input parameters "as is", showing adequate results.

3. Gradient boosting models

Models were developed in Python (ver. 3.7) in the Jupyter Notebook environment.

During the study, two models were built. In the first model (SpatialCB1), the coordinates X and Y were used as input parameters, and the gray gradient at X.Y point was obtained as the output parameter (Figure 4).

In the second model (SpatialCB2), in addition to the X, Y coordinate parameters, it was also supplied the data about the nearest point with known data from the training dataset. So, additional parameters were chosen: the distance to the nearest point, a gray gradient at the nearest point (Figure 5).

The prototype for this approach was the well-known interpolation method called the "k-nearest neighbor method". K-nearest neighbor method (k-nearest neighbors algorithm, k-NN) is a metric algorithm for automatic classification of objects or regression. In the case of using the method for classification, the object is assigned to the class that is the most common among k neighbors of this element,

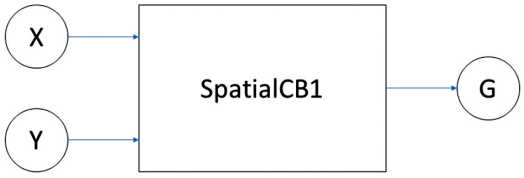

Figure 4
Scheme of the gradient boost model SpatialCB1.
X, Y – coordinates, G – the gradient of the desired point.

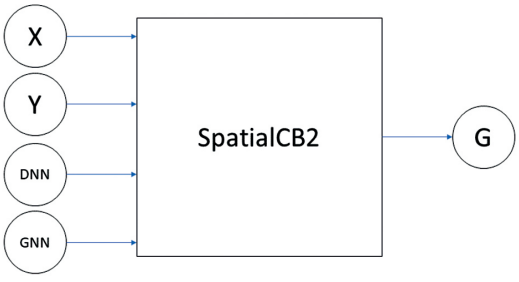

Figure 5
Scheme of the gradient boost model SpatialCB2.
X, Y – coordinates, DNN – distance to nearest neighbor point, GNN – gradient of nearest neighbor point, G – gradient of the desired point.

the classes of which are already known (Kung et al., 2012; Lee, 2017). In the case of using the method for regression, the object is assigned an average value over the k nearest objects to it, the values of which are already known. In a weighted way of implementing this method, not only the number of certain classes that fall into the region, but also their remoteness from the new value are considered (Biau et al., 2011). Figure 6 shows an example of k-nearest-neighbor classification. The test

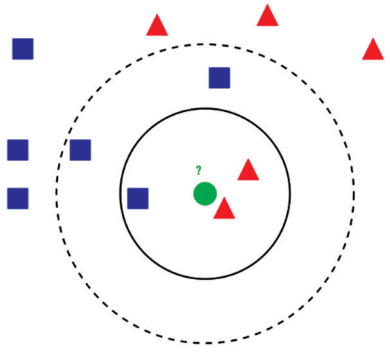

Figure 6
Example of classification by the method of k-nearest neighbors
Author: Antti Ajanki ("File:KnnClassification.svg - Wikimedia Commons," n.d.)

sample (green circle) must be classified as a blue square (class 1) or as a red triangle (class 2). If k = 3, then it is classified as 2nd class, because inside the smaller circle there are 2 triangles and only 1 square. If k = 5, then it will be classified as 1st class (3 squares against 2 triangles inside the larger circle).

For building the models CatBoostClassifier the class of the CatBoost module was used. Model parameters were set the same and have the following values: eval_metric = 'AUC' – evaluation metric, the metric by which the model's performance on a test set is calculated. AUC – one of the most popular quality assessment functions in classification problems – the area under the error curve or ROC curve (Cortes and Mohri, 2004). The error curve or ROC-curve is a graphical characteristic of the quality of the binary classifier, the dependence of the rate of true positive classifications on the rate of false positive classifications when varying the threshold of the decision rule (Hernández-Orallo, 2013). Figures 7 and 8 show examples of ROC curves. The graphs show that for the most accurately working model, the area under the error curve (AUC) will be 1.

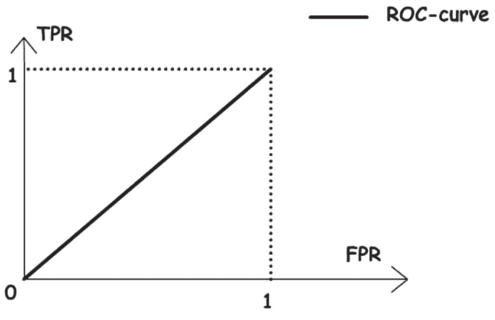

Figure 7
The shape of the ROC-curve in the case when the right or wrong decision is made as a result of „random fortune-telling"

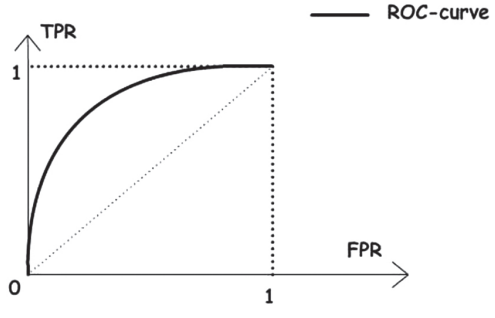

Figure 8
The shape of the ROC curve when using a fairly good model

iterations = 40 – iterations number. The number of repetitions of the model training procedure.
random_seed = 63 – the number used to generate random numbers when working with the built-in randomizer of the CatBoostClassifier class.
learning_rate = 0.01 – learning rate.
loss_function = 'MultiClass' – loss function. This function is used to reduce errors during model training. The 'MultiClass' function allows you to find the probability of each given class to be the right answer, while the total probability of all classes is 1.0 ("Multiclassification: objectives and metrics - CatBoost. Documentation," n.d.). Since in our problem the number of classes in the output is four, the type of the problem is multiclassification.
custom_loss = ['AUC', 'Accuracy'] – functions calculated during model training. Observation of these functions allows you to better understand how well the model is trained and how adequate results it can show on the test sample. 'Accuracy' function – proportion of correct algorithm responses (Jiang et al., 2018):

$$Accuracy = \frac{number\ of\ correct\ answers}{number\ of\ all\ answers}$$

early_stopping_rounds = 20. This parameter is responsible for the number of training iterations during which a decrease or increase in the loss function is observed. If the loss function does not decrease, the learning algorithm stops.

RESULTS AND DISCUSSION

Two gradient boosting models were tested with varying training dataset sizes from 50 to 10000 elements. The results of the models are shown in table 1.

Table 1 Spatial prediction results using gradient boosting models

Training dataset size	% of the training dataset from the total data	Model					
		SpatialCB1			SpatialCB2		
		AUC	Accuracy	Learning time	AUC	Accuracy	Learning time
50	0.099	0.28	0.2	132 ms	0.16	0.90	256 ms
100	0.198	0.19	0.6	94 ms	0.84	0.70	215 ms
250	0.496	0.62	0.52	141 ms	0.91	0.86	446 ms
500	0.991	0.62	0.62	185 ms	0.96	0.91	371 ms
1000	1.983	0.83	0.65	269 ms	0.98	0.95	548 ms
5000	9.914	0.95	0.77	7 s 437 ms	0.98	0.97	547 ms
10000	19.827	0.96	0.76	9 s 678 ms	0.99	0.98	847 ms

The study revealed that already with a training sample of 250 points (~ 0.5% of the total data), the SpatialCB2 model shows an efficiency similar to that achieved with the SpatialCB1 model with a training sample of 5000 points, which is almost 10% of the total data. In other words, in order to get adequate results on the SpatialCB1 model, we need to have 10% of the total amount of information, while in the case of the SpatialCB2 model, 0.5% of the known amount of data will be enough. Figures 9 and 10 show the results of the two models under consideration with different sizes of the training dataset. Figure 11 shows an approximate distribution of points of a training dataset of different volumes in order to visually see the required amount of information about the object under study.

Analyzing the visualized results of the models, it is clear that with a small sample size of 50 points, the SpatialCB1 model detects the most common gradient and gives a "dark gray" answer to any coordinate. In this case, all the points that fall into the area with the given color gradient are automatically determined as "correct",

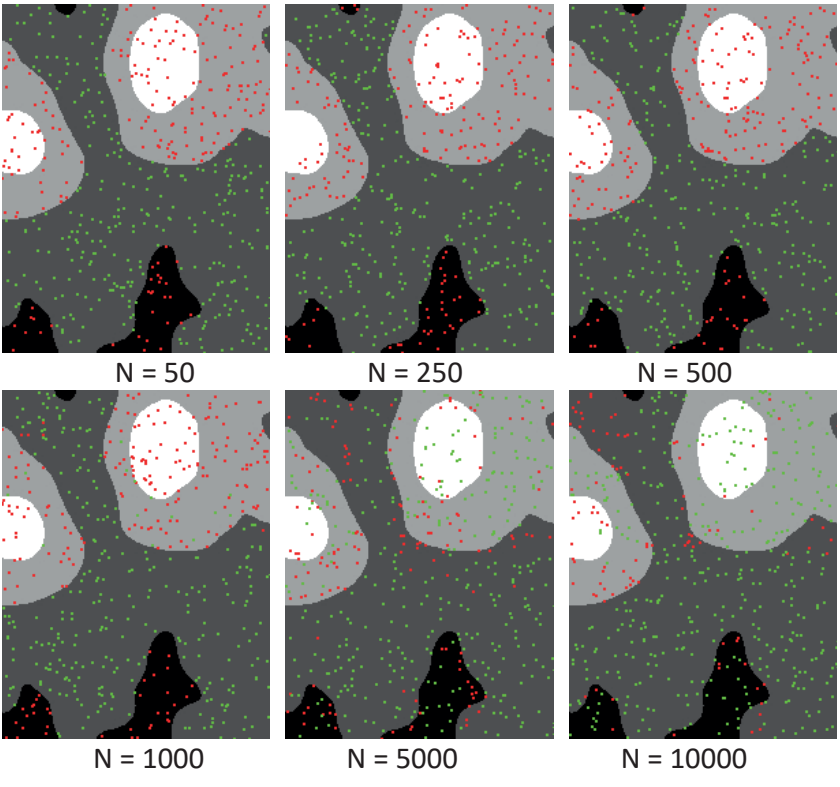

Figure 9
Visualization of the results of spatial forecasting using the SpatialCB1 model
(N - volume of the train dataset, green dots - correctly defined values,
red dots - incorrectly defined values)

and the points lying in the areas with other gradients are "incorrect". On the SpatialCB1 model, this picture is repeated until the training sample reaches 1000 points, i.e., almost 2% of all points of the model image. In this case, several points lying on a light gray gradient are correctly determined. The smallest areas - white and black on this model begin to be determined only when using the training sample size of 5000 points (~ 10%). On this model, the efficiency does not exceed 76%.

On the SpatialCB2 model, with a small training sample of 50 points, points lying in areas with dark gray and light gray gradients are quite well defined. Plots with all four gradients begin to be determined at N = 250. This model is characterized by the presence of errors at the boundaries of the transition from one gradient to another. Thus, this model shows an efficiency 20 times higher.

Based on the data obtained, it is possible to visualize the sampling grid necessary to perform a spatial forecast using the developed models, provided that we have information from 50 locations on the study area (for example, sampling sites).

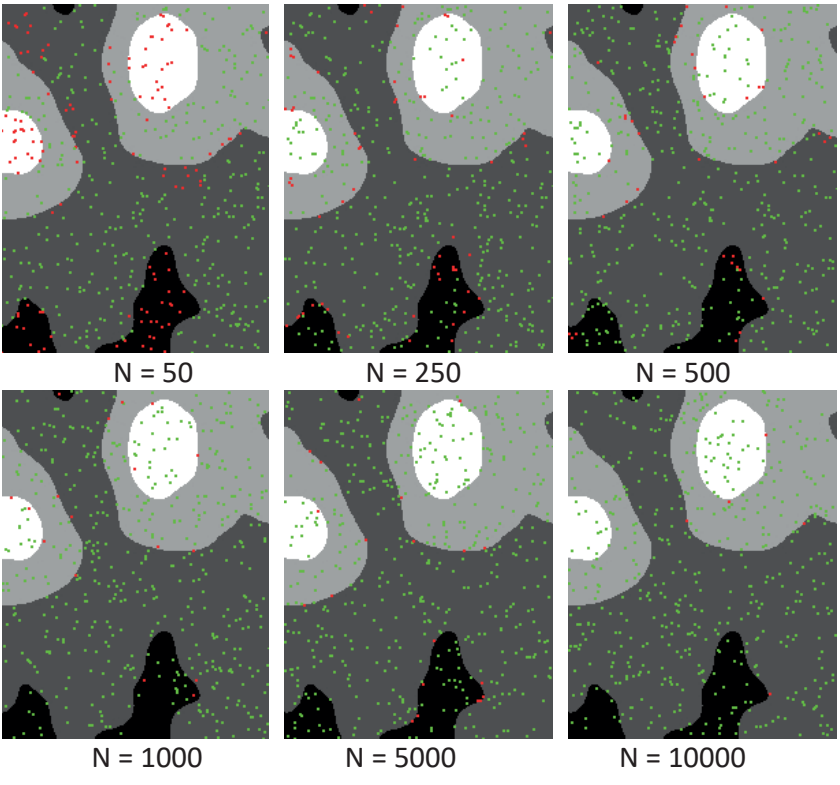

Figure 10
Visualization of the results of spatial forecasting using the SpatialCB2 model
(N - volume of the train dataset, green dots - correctly defined values,
red dots - incorrectly defined values)

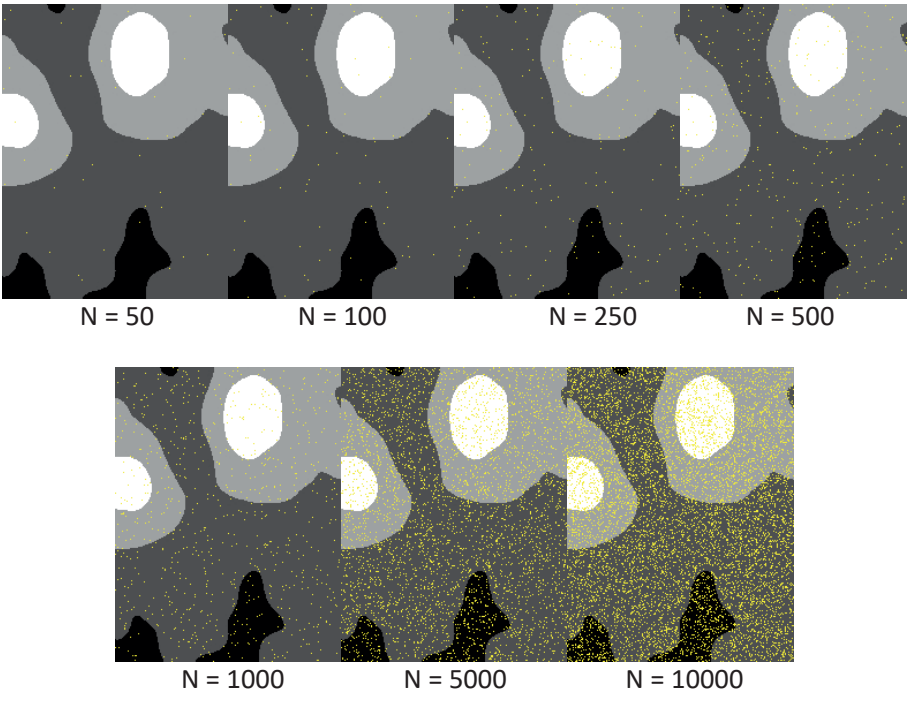

Figure 11
Visualization of train dataset (N - volume of the train dataset, yellow dots – place of certain point from dataset)

For the SpatialCB1 model, the minimum information for an adequate forecast is 10%. We take each sampling location as a square with a certain area. Then 50 squares is 10%, and 500 squares will be the entire investigated area. That is, for an adequate analysis, a grid of 500 cells can be superimposed on the study area. For the SpatialCB2 model, the minimum information for an adequate forecast is 0.5%. Then, for this model, 50 squares is 0.5%, and 10 000 squares will be the entire investigated area. That is, for an adequate analysis at 50 points studied, a grid of 10 000 cells can be superimposed on the study area. A visualized representation of such grids is shown in Figure 12. Figure 12 shows how a more detailed picture can be obtained using the SpatialCB2 model.

For an even more cleare understanding of the results, let us consider Pavlodar city (Kazakhstan) as an example. The area of the city is – 352.3 km^2. Then, in the case of the first model, each square under consideration will have an area equal to 0.7046 km^2 or 704 600 m^2. That is, if we take 50 samples to the entire territory of the city, then each sample will provide conditional information about an area of 704 600 m^2 (i.e., a square of 839 meters in length and 839 meters in width). In

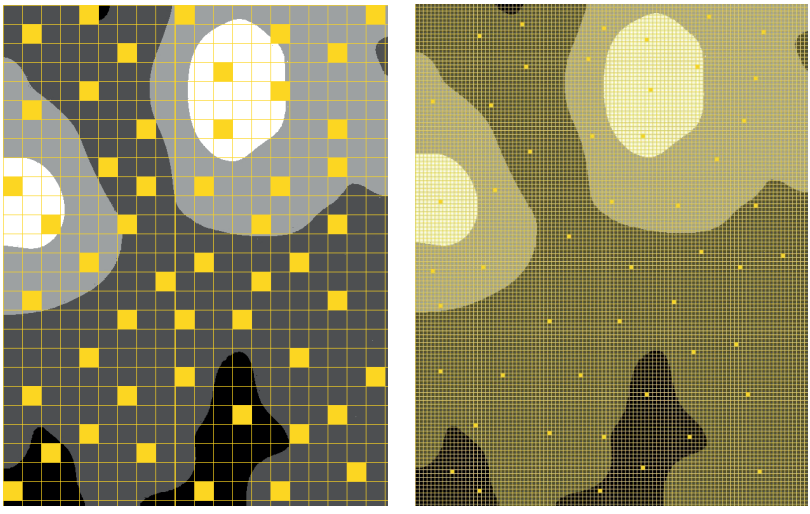

Figure 12
Visualization of spatial forecast detailing using SpatialCB1 (left) and SpatialCB2 (right) models

the case of the second model, the square area will be – 35 230 m^2 (i.e., a square of 188 m in length and 188 m in width), which is almost 20 times smaller than in the case of the first model. If, for example, we take 100 points to the city, then, using the SpatialCB2 model, one point will be a territory with an area of 17 615 m^2 (i.e., a square of 133 m in length and 133 m in width).

CONCLUSIONS

As a result of the studies, two gradient boosting models were obtained for spatial analysis of the distribution of various factors, in particular the distribution of the concentration of the pollutant. The SpatialCB2 model showed an efficiency 20 times higher than the SpatialCB1 model. This result was achieved due to the fact that in the second model, information about the nearest neighboring point (distance to the neighboring point and the gradient of the neighboring point) was additionally introduced as parameters by analogy with the k-nearest neighbor method. This model works adequately with a training sample of 0.5% of all analyzed information about the object. The analysis showed that for the study area of 352.3 km^2 with the available data for 100 grid elements, the analytical segment of the grid will occupy an area of 17 615 m^2, which is equivalent to a square of 133x133 m.

Thus, an effective version of the model was obtained for spatial analysis of the distribution of a certain factor using the technology of machine learning - gradient boosting. The model is based on the CatBoost library.

REFERENCES

ABADI, M., AGARWAL, A., BARHAM, P., et al. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Retrieved from http://arxiv.org/abs/1603.04467

AZHAYEV, G., ESIMOVA, D., SONKO, S. M., et al. (2020). Geoecological environmental evaluation of Pavlodar region of the republic of Kazakhstan as a factor of perspectives for touristic activity. *GeoJournal of Tourism and Geosites*, *28*(1), 104–113. https://doi.org/10.30892/gtg.28108-455

BEKETOVA, A., BERDENOV, Z. G., ATAEVA, G., et al. (2019). Geochemical monitoring of industrial center for development of recreational areas (on the example of Khromtau-don industrial hub, Kazakhstan). *GeoJournal of Tourism and Geosites*, *27*(4), 1449–1463. https://doi.org/10.30892/gtg.27428-447

BERDENOV, Z., MENDIBAEV, E., SALIHOV, T., et al. (2017). Geoecological analysis of industrial cities: on the example of Aktobe agglomeration. *Chemistry*, *26*(6), 890–902.

BERROCAL, V. J., GELFAND, A. E., HOLLAND, D. M. (2010). A bivariate space-time downscaler under space and time misalignment. *Annals of Applied Statistics*, *4*(4), 1942–1975. https://doi.org/10.1214/10-AOAS351

BIAU, G., CHAZAL, F., COHEN-STEINER, D., et al. (2011). A weighted k-nearest neighbor density estimate for geometric inference. *Electronic Journal of Statistics*, *5*, 204–237. https://doi.org/10.1214/11-EJS606

CASATI, B., ROSS, G., STEPHENSON, D. B. (2004). A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, *11*(2), 141–154. https://doi.org/10.1017/S1350482704001239

File:KnnClassification.svg - Wikimedia Commons. (n.d.). Retrieved May 4, 2020, from https://commons.wikimedia.org/w/index.php?curid=2170282

GALLEGO-ÁLVAREZ, I., VICENTE-GALINDO, M. P., GALINDO-VILLARDÓN, M. P., et al. (2014). Environmental performance in countries worldwide: Determinant factors and multivariate analysis. *Sustainability (Switzerland)*, *6*(11), 7807–7832. https://doi.org/10.3390/su6117807

GLASS, G. E., SCHWARTZ, B. S., MORGAN, J. M., et al. (1995). Environmental risk factors for Lyme disease identified with geographic information systems. *American Journal of Public Health*, *85*(7), 944–948. https://doi.org/10.2105/AJPH.85. 7. 944

ILIEŞ, D. C., BAIAS, Ş., BUHAŞ, R., et al. (2017). Environmental education in protected areas. case study from Bihor County, Romania. *Geojournal of Tourism and Geosites*, *19*(1), 126–132.

ILIEŞ, D. C., ONEŢ, A., GRIGORE, H., et al. (2019). Exploring the indoor environment of heritage buildings and its role in the conservation of valuable objects. *Environmental Engineering and Management Journal*, *18*(12), 2579–2586. https://doi.org/10.30638/eemj.2019.243

JERRETT, M., BURNETT, R. T., GOLDBERG, M. S., et al. (2003, August 22). Spatial analysis for environmental health research: Concepts, methods, and examples. *Journal of Toxicology and Environmental Health - Part A*. https://doi.org/10.1080/15287390306446

KÖRNER, P., KRONENBERG, R., GENZEL, S., et al. (2018). Introducing Gradient Boosting as a universal gap filling tool for meteorological time series. *Meteorologische Zeitschrift*, *27*(5), 369–376. https://doi.org/10.1127/metz/2018/0908

KUNG, Y. H., LIN, P. S., KAO, C. H. (2012). An optimal k-nearest neighbor for density estimation. *Statistics and Probability Letters*, *82*(10), 1786–1791. https://doi.org/10.1016/j.spl.2012. 05. 017

LEE, J. M. (2017). Fast k-Nearest Neighbor Searching in Static Objects. *Wireless Personal Communications*, *93*(1), 147–160. https://doi.org/10.1007/s11277-016-3524-1

LI, R., ZHANG, X., LIU, L., et al. (2020). Application of neural network to building environmental prediction and control. *Building Services Engineering Research and Technology*, *41*(1), 25–45. https://doi.org/10.1177/0143624419838362

LI, T. Z., LIN, J. S., WU, M. T., et al. (2009). Concept and spatial analysis method of urban environmental traffic capacity. *Journal of Transportation Engineering*, *135*(11), 873–879. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000061

MADADGAR, S., MORADKHANI, H. (2014). Spatio-temporal drought forecasting within Bayesian networks. *Journal of Hydrology*, *512*, 134–146. https://doi.org/10.1016/j.jhydrol.2014. 02. 039

MATLOVIČ, R., MATLOVIČOVÁ, K. (2012). The social relevance and branding of geography [Spoločenská Relevancia a Budovanie Značky Geografie]. *Geografie-Sbornik CGS*, *117*(1), 33–51.

MATLOVIČ, R., MATLOVIČOVÁ, K. (2020). First and second order discontinuities in world geographical thought and their primary reception in Slovak geography. *Folia Geographica*, *62*(1), (online first).

MEYER, H., KATURJI, M., DETSCH, F., et al. (2019). *AntAir: satellite-derived 1 km daily Antarctic air temperatures since 2003*. https://doi.org/10.5194/essd-2019-215

MIHINCĂU, D. C., ILIES, D. C., KOROLEVA, Y., et al. (2019). The study of indoor microclimate on wooden churches to be included among Oradea's representative sights. *Geojournal of Tourism and Geosites*, *26*(3), 737–750. https://doi.org/10.30892/gtg.26305-393

MURTAGH, F., ZHENG, G., CAMPBELL, J. G., et al. (2000). Neural network modelling for environmental prediction. *Neurocomputing*, *30*(1–4), 65–70. https://doi.org/10.1016/S0925-2312(99)00144-7

NATEKIN, A., KNOLL, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC). https://doi.org/10.3389/fnbot.2013.00021

PACI, L., GELFAND, A. E., HOLLAND, D. M. (2013). Spatio-temporal modeling for real-time ozone forecasting. *Spatial Statistics*, *4*, 79–93. https://doi.org/10.1016/j.spasta.2013. 04. 003

PROKHORENKOVA, L., GUSEV, G., VOROBEV, A., et al. (2018). CatBoost: unbiased boosting with categorical features. In Bengio, S and Wallach, H and Larochelle, H and Grauman, K and CesaBianchi, N and Garnett, R (Ed.), *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 31 (NIPS 2018)* (Vol. 31). 10010 NORTH TORREY PINES RD, LA JOLLA, CALIFORNIA 92037 USA: NEURAL INFORMATION PROCESSING SYSTEMS (NIPS).

RAMAZANOVA, N., BERDENOV, Z. G., RAMAZANOV, S., et al. (2019). Landscape-geochemical analysis of steppe zone basin Zhaiyk. *News of National Academy of Sciences of the Republic of Kazakhstan*, *4*(436), 33–41. https://doi.org/10.32014/2019.2518-170x.95

RIBEIRO, M. C., PINHO, P., LLOP, E., et al. (2013). Multivariate geostatistical methods for analysis of relationships between ecological indicators and environmental factors at multiple spatial scales. *Ecological Indicators*, *29*, 339–347. https://doi.org/10.1016/j.ecolind.2013. 01. 011

SAFAROV, R. Z., SHOMANOVA, Z. K., MUKANOVA, R. Z., et al. (2019). Design of neural network for forecast analysis of elements-contaminants distribution on studied territories (on example of Pavlodar city, Kazakhstan). *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Chemistry and Technology*, *438*(6), 86–98. https://doi.org/10.32014/2019.2518-1491.78

SAHU, S. K., YIP, S., HOLLAND, D. M. (2009). Improved space-time forecasting of next day ozone concentrations in the eastern US. *Atmospheric Environment*, *43*(3), 494–501. https://doi.org/10.1016/j.atmosenv.2008. 10. 028

SEXTON, K., WALLER, L. A., MCMASTER, R. B., et al. (2002). The importance of spatial effects for environmental health policy and research. In *Human and Ecological Risk Assessment* (Vol. 8, pp. 109–125). https://doi.org/10.1080/20028091056764

SHOJI, R., KAWAKAMI, M. (2006). Prediction of genotoxicity of various environmental pollutants by artificial neural network simulation. *Molecular Diversity*, *10*(2), 101–108. https://doi.org/10.1007/s11030-005-9005-1

VICENTE, J. R., GONÇALVES, J., HONRADO, J. P., et al. (2014). A framework for assessing the scale of influence of environmental factors on ecological patterns. *Ecological Complexity*, *20*, 151–156. https://doi.org/10.1016/j.ecocom.2014. 10. 005

XIONG, D., GUI, Q., HOU, W., et al. (2018). Gradient boosting for single image super-resolution. *Information Sciences*, *454–455*, 328–343. https://doi.org/10.1016/j.ins.2018. 04. 075

YUAN, S. (2015). Random gradient boosting for predicting conditional quantiles. *Journal of Statistical Computation and Simulation*, *85*(18), 3716–3726. https://doi.org/10.1080/00949655.2014.1002099