mathematics

MDPI

Article

Determination of Reservoir Oxidation Zone Formation in Uranium Wells Using Ensemble Machine Learning Methods

Ravil I. Mukhamediev [1,2], Yan Kuchin [1,2,*], Yelena Popova [3], Nadiya Yunicheva [2,4,*], Elena Muhamedijeva [2], Adilkhan Symagulov [1,2], Kirill Abramov [2], Viktors Gopejenko [5,6], Vitaly Levashenko [7], Elena Zaitseva [7], Natalya Litvishko [2] and Sergey Stankevich [8]

1 Institute of Automation and Information Technologies, Satbayev University (KazNRTU), Almaty 050013, Kazakhstan; r.mukhamediev@satbayev.university (R.I.M.); a.symagulov@satbayev.university (A.S.)
2 Institute of Information and Computational Technologies CS MSHE RK, 28 Shevchenko Str., Almaty 050010, Kazakhstan; muhamedijeva@gmail.com (E.M.); abramc@mail.ru (K.A.)
3 Transport and Management Faculty, Transport and Telecommunication Institute, 1 Lomonosov Str., LV-1019 Riga, Latvia; popova.j@tsi.lv
4 Institute of Automation and Information Technologies, Almaty University of Energy and Communications, Baitursynov Str., 126/1, Almaty 050013, Kazakhstan
5 International Radio Astronomy Centre, Ventspils University of Applied Sciences, LV-3601 Ventspils, Latvia; viktors.gopejenko@isma.lv
6 Department of Natural Science and Computer Technologies, ISMA University of Applied Sciences, LV-1019 Riga, Latvia
7 Faculty of Management Science and Informatics, University of Zilina, 010 26 Žilina, Slovakia; vitaly.levashenko@fri.uniza.sk (V.L.); elena.zaitseva@fri.uniza.sk (E.Z.)
8 Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Sciences of the National Academy of Sciences of Ukraine, 01054 Kyiv, Ukraine; st@casre.kiev.ua
* Correspondence: ykuchin@mail.ru (Y.K.); naduni@mail.ru (N.Y.)

check for updates

Citation: Mukhamediev, R.I.; Kuchin, Y.; Popova, Y.; Yunicheva, N.; Muhamedijeva, E.; Symagulov, A.; Abramov, K.; Gopejenko, V.; Levashenko, V.; Zaitseva, E.; et al. Determination of Reservoir Oxidation Zone Formation in Uranium Wells Using Ensemble Machine Learning Methods. *Mathematics* **2023**, *11*, 4687. https://doi.org/10.3390/math11224687

Academic Editors: Xiang Li, Ioannis G. Tsoulos, Jie Wen and Jianping Gou

Received: 12 September 2023
Revised: 3 November 2023
Accepted: 13 November 2023
Published: 17 November 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Approximately 50% of the world's uranium is mined in a closed way using underground well leaching. In the process of uranium mining at formation-infiltration deposits, an important role is played by the correct identification of the formation of reservoir oxidation zones (ROZs), within which the uranium content is extremely low and which affect the determination of ore reserves and subsequent mining processes. The currently used methodology for identifying ROZs requires the use of highly skilled labor and resource-intensive studies using neutron fission logging; therefore, it is not always performed. At the same time, the available electrical logging measurements data collected in the process of geophysical well surveys and exploration well data can be effectively used to identify ROZs using machine learning models. This study presents a solution to the problem of detecting ROZs in uranium deposits using ensemble machine learning methods. This method provides an index of weighted harmonic measure (f1_weighted) in the range from 0.72 to 0.93 (XGB classifier), and sufficient stability at different ratios of objects in the input dataset. The obtained results demonstrate the potential for practical use of this method for detecting ROZs in formation-infiltration uranium deposits using ensemble machine learning.

Keywords: uranium mining; machine learning; reservoir oxidation zone; ensemble machine learning

MSC: 68T05; 68Q32; 68-04

1. Introduction

Uranium deposits in Kazakhstan are mined using the environmentally efficient in situ leaching (ISL) method, which, however, requires a fairly accurate determination of the lithologic structure of the host rocks. Since about 41% of the world uranium production and almost the entire production volume in Kazakhstan is carried out using ISL [1], the relevance of the task of determining the characteristics of the host rocks is extremely high. To solve

this task, in some cases, machine learning methods are employed, increasing the degree of automation of this process and reducing the influence of human error. The characteristics of rocks are determined through the process of geophysical research of boreholes (GRB). The GRB process is different for exploration and production wells. However, in both, as a standard set of GRB methods in Kazakhstan fields, apparent resistance logging (AR) and spontaneous polarization (SP) potential are used for lithologic classification and the determination of the filtration properties of rocks. Moreover, gamma ray logging (GR) is used for the calculation of uranium content on the basis of the gamma radiation of radium and its decay products with the use of conversion through the radioactive equilibrium coefficient. The log data are the recorded physical parameters inside the well in 10 cm increments in depth, and are visualized for evaluation by experts as graphs (curves). At the same time, the automatic interpretation of electric logs (AR and SP) is implemented only on the basis of AR logs, without taking into account data from other logs, information on neighboring wells, etc., which leads to significant manual adjustments by the interpreting engineer. In addition, this leads to the impossibility of correct lithologic interpretation in case of AR curve distortion. In this connection, the main directions of the application of machine learning methods in the processing of logging data at uranium deposits emerge are as follows:

- Lithologic classification
- Determination of the permeability of host rocks
- Determination of reservoir oxidation zones (ROZs) (zones with disturbed radioactive equilibrium).
- Determination of technological acidification zones (zones with distorted AR curve).

The methods of applying machine learning to lithologic classification have been discussed in a number of papers (see Section 2). The problem of determining the filtration coefficient is considered in [2,3], in which machine learning methods demonstrate an almost twofold increase in accuracy compared to the currently used methodology.

This study deals with the problem of ROZ determination. A machine learning-based method for determining ROZs from the data of exploratory wells with acceptable accuracy for practice is proposed.

In the research process, we answer two questions:

1. Is it possible to identify ROZs by machine learning using a standard log dataset?
2. Which machine learning methods give the best classification result?

The novelty of this study includes two important aspects.

1. The problem of ROZ detection using machine learning techniques is considered for the first time according to the review of investigations in this domain.
2. An acceptable ROZ for practical applications is obtained, and the limitations of the proposed method are identified.

The paper consists of the following sections (Figure 1).



Figure 1. Main sections of the article.

Section 2 provides a literature review concerning the interpretation of logging data using machine learning.

Section 3 describes the physical principles of ROZ formation and the limitations in determining ROZs from log data.

Section 4 describes the data used and the data processing methods.

Section 5 describes the computational experiments and demonstrates the results.

The obtained results are discussed in Section 6.

Finally, in Section 7, the limitations of the method and directions for future research are discussed.

2. Related Works

The use of machine learning in log data interpretation has aroused the interest of researchers since the 1970s. In the 1990s, active research on automatically interpreted log data using feedforward artificial neural networks began [4–7]. A wide range of classical and modern methods of machine learning was applied. In particular, in article [8], a number of classical algorithms, such as support vector machine (SVM), decision tree (DT), random forest (RF), multi-layer perceptron (MLP), and ensemble machine learning method (XGBoost) are used for the classification of lithofacies in the Talcher coalfield, Eastern India, with an accuracy rate of more than 80% for binary classification (carbonaceous and non-coal lithofacies). In paper [9], the application of a convolutional neural network for lithofacies classification in the Eagle Ford and Austin Chalk shale oil fields is considered. The problem of lithologic classification of four geothermal wells in the Snake River Plain (SRP) in Idaho is considered in paper [10]. The authors compared k-nearest neighbor (kNN), SVM, and eXtreme Gradient Boosting (XGBoost). The latter algorithm showed the highest classification accuracy (90.67%). The combined method of kernel principal component analysis-Bayesian optimization-categorical boost (KPCBA-CatBoost) demonstrated approximately the same accuracy in the task of lithologic classification in an oil and gas field (accuracy 90%) [11]. In article [12], the task of classifying and predicting the geological facies using well log data in the Anadarko Basin oil field, Kansas, is considered. The feedforward neural network (FFNN) showed an accuracy of 88%. A similar task of classifying eight rock classes in the Vikulov Formation (western Siberia) is discussed in paper [13]. The authors compared the CatBoost, RF, and MLP algorithms. Although the achieved classification accuracy is in the order of 64%, the authors conclude that machine learning algorithms can predict lithology from a standard set of log diagrams without normalization to reference formations, which can significantly reduce the time required to prepare log curves in advance.

The issue of determining rock permeability is considered in a number of works [14–17]. For example, the problem of determining the rock permeability of oil reservoirs carbonate reservoirs (Ilam and Sarvak) in the southwest of Iran is discussed in article [18]. The authors considered the following methods: multi-layer perceptron neural network (MLP), radial basis function neural network (RBF), SVM, DT, and RF, and achieved a coefficient of determination in the rock permeability prediction problem of 0.97 (SVM), which is higher than that obtained using traditional methods.

Machine learning methods are used to interpret acoustic logging data [19], seismic data [20,21], geologic mapping [22,23], stratigraphic classification [23,24], geochemical anomaly detection [25,26], and others.

ML methods have been used at uranium deposits for lithologic classification [27]; stratigraphy [28]; the determination of the filtration properties of host rocks [2]; and the assessment of the influence of expert marking of logging data [29]. It turned out that the quality of interpretation of logging data in uranium fields, as well as in some tasks solved in oil fields (problems of classifying the connectivity quality (among six ordinal classes) and hydraulic isolation (two classes)) [30], depends crucially on expert data labeling [31]. The results of the application of machine learning methods in the processing of logging data obtained to date are briefly summarized in Table 1. Methods that showed the best results are highlighted in bold.

Table 1. Machine learning in logging data processing tasks.

Deposit Type	Task	Used Methods	Result	Ref.
Coalfield	lithologic classification	SVM, DT, RF, MLP, and **XGBoost**	Ac $\geq$ 80%	[8]
Geothermal deposit	lithologic classification	KNN, SVM, and **XGBoost**	Ac $\geq$ 90%	[10]
Oil and gas field	lithologic classification	**KPCBA-CatBoost**, and FFNN	Ac $\geq$ 88%	[11]
Oil and gas field	rock permeability	MLP, RBF, **SVM**, DT, and RF	$R^2 \geq 0.97$	[18]
Geological formation	lithologic classification	**CatBoost**, RF, and MLP	Ac $\geq$ 64%	[13]
Uranium field	lithologic classification	**FFNN**, kNN, and LSTM	Ac $\geq$ 60%	[27]
Uranium field	determination of the filtration properties of host rocks	kNN, RF, SVM, and **XGBoost**	$R^2 \geq 0.67$	[2]
Uranium field	stratigraphy	KNN, RF, linear regression, and **XGBoost**	Ac $\geq$ 97%	[28]

There are significant differences between oil well logging and uranium well logging. At oil fields, the interpretation of data is required to identify formations with a thickness of one meter or more, for one or several wells of thousands of meters deep. Meanwhile, due to the technological peculiarities of mining processes, the interpretation of logging data at uranium deposits of the infiltration-type formation requires the identification of layers with a minimum thickness of 20 cm for dozens and hundreds of wells several hundred meters deep. The wells at uranium deposits of Kazakhstan are usually drilled with a smaller diameter (118–132 mm) than oil wells, which imposes restrictions on the length and diameter of the borehole devices used, and consequently, on the geophysical research methods used. Therefore, process wells in uranium deposits are investigated by relatively inaccurate electrical logging methods with a small number of measuring electrodes. The exception is exploration wells, where core sampling is performed from different depths, but even in this case, due to the fact that the deposits are located in sedimentary rocks (sands and clays), they are often eroded, and it is often impossible to extract rock samples of the required quality.

In general, the analysis of publications shows the interest and very significant success in the interpretation of log data using ML, especially in oil and gas fields. However, we have not identified any studies devoted to ROZ identification using ML methods for uranium deposits. The authors hope that the present article will fill this gap, since the timely identification of ROZs affects the correct estimation of extractable uranium reserves and can significantly reduce costs in the mining process.

For a better understanding of the necessity of such research, we consider the causes of ROZ, the currently used methods for its identification, and their limitations.

3. Physical Principles of ROZ Formation and Applied Methods for Its Identification

The peculiarities of deposit formation predetermine their radiological situation and the element composition. In this case, typically, in the ore deposit there is a shortage of radium in comparison with the equilibrium state, and in its frame the radioactive equilibrium is disturbed in the direction of excess radium, which is a consequence of the formation of so-called “residual” and “diffusion” halos of radium. This is due to the fact that deposits of the layer-infiltration type are formed in sedimentary permeable rock strata at the boundary of the redox barrier. Since in oxidizing and reducing conditions the behavior of mobile forms of uranium and radium differ significantly, in different morphological elements of ore bodies as a result of the processes of “export intake” of “mother” uranium and “daughter” radium, geochemical zones appear, where the correlation of mass fractions of radium and uranium differs from the values corresponding to the state of radioactive equilibrium between them.

The state of radioactive equilibrium between radium and uranium is traditionally characterized by the radioactive equilibrium violation coefficient (or, merely, radioactive equilibrium coefficient) often referred to as Kpp, which is equal to the ratio of the mass fractions of radium and uranium.

Thus, the value of Kpp = 1 corresponds to the presence of radioactive equilibrium, and the difference of Kpp values from 1 indicates the presence of systems that have not reached equilibrium or have undergone violations of their closure.

In the cut, the ore body at hydrogenic uranium deposits has the form of a roll moving in the direction of formation water movement (see Figure 2), and the change in Kpp obeys the following basic laws [32]:

- The average values of Kpp for the different morphological elements of ore bodies, sites, and geochemical zones of deposits (with uranium content of more than 0.01%) vary within a fairly wide range—from 0.60 to 1.0;
- Directly behind the front of reservoir oxidation, the radioactive equilibrium is shifted towards an excess of radium (Kpp = 1.5–2.5 and more), up to almost a complete absence of uranium. These are the so-called "residual" radium halos;
- As the radioactive equilibrium gradually shifts from equilibrium ores (near the reservoir oxidation zone (ROZ)) to an excess of radium, it forms small (0.2–0.4 m) areas of radium rims (the so-called "diffusion" radium halos) at the boundary of ore bodies, with uranium content of 0.01% and higher.

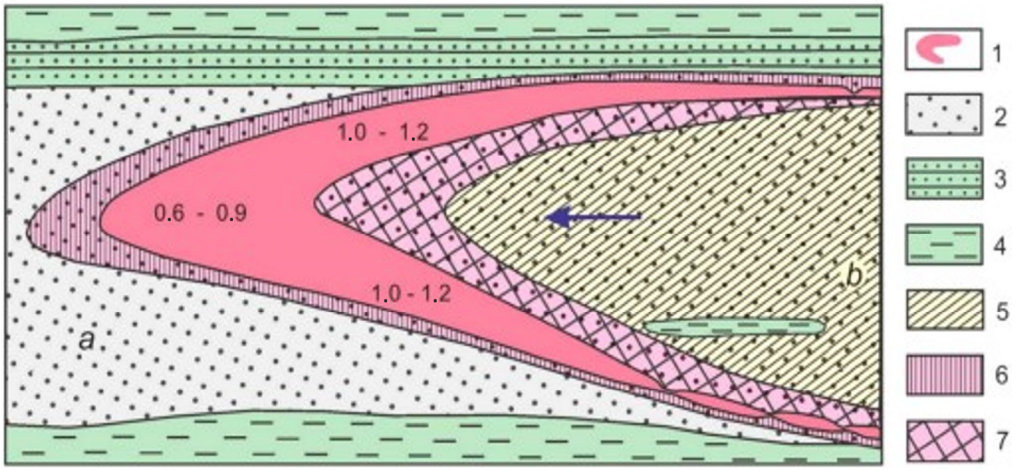


Figure 2. Scheme of radiological zonality in the section of the ore-bearing horizon of formation-infiltration uranium deposits (the arrow indicates the direction of formation water movement, numbersKpp value, arecovery zone, boxidation zone): 1—uranium ore body; 2—sands; 3—clayey sandstones; 4—clays, siltstones; 5—oxidized rocks (reservoir oxidation zone); 6—radium diffusion halo; 7—residual radium halo.

At present, the uranium content is determined by dividing the radium content obtained as a result of interpretation by Kpp [2]. Natural gamma radiation of radium and other uranium decay products is registered during GR. Since uranium is completely absent in the zones of reservoir oxidation, Kpp = ∞; in other words, in the presence of pronounced gamma anomalies, they must be taken into account in the interpretation of gamma-ray logging. ROZs can be extracted by core analysis at the exploration stage, and then geologic sections can be built and extended along geologic sections, as shown in Figure 3. High-resolution pictures are given at in the Supplementary Materials.

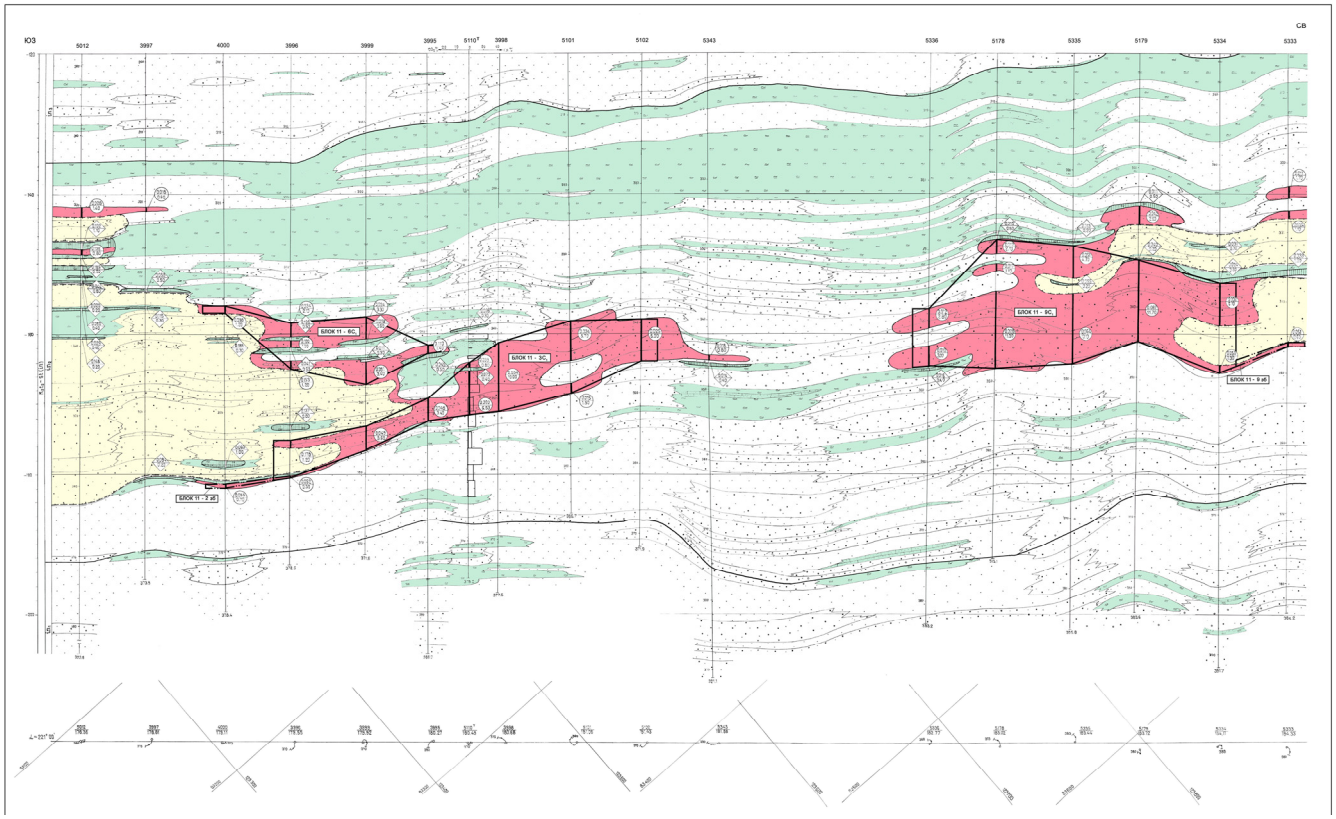


Figure 3. Geologic cut, with reservoir oxidation zones (highlighted in yellow) on the right and left.

However, core sampling and laboratory testing are a long and expensive process. Moreover, the construction of sections and the extrapolation of selected ROZs are not always carried out in a timely and correct manner, as they require high qualification and considerable manual labor. As a result, the initial interpretation is often carried out without taking ROZs into account, and then it is necessary to recalculate it considering ROZs.

Another method of ROZ extraction is fission neutron logging (FN), as it allows the direct determination of uranium content, bypassing the stage of conversion of radium content to uranium content via Kpp. In such cases, ROZs are identified where gamma log ore intervals do not correspond to FN ore intervals. An example of GR reinterpretation after ROZ delineation based on FN results is shown in Figure 4.

Figure 4 shows that the actual ore intervals for FN were much smaller than those calculated for radium. However, the FN procedure is expensive and the logging speed is not more than 50 m/h, but the main limitation is that the resource of the neutron generator tube used is extremely limited. Therefore, FN is not ordered in all fields, and it only concerns small volumes (5–10%) of the total number of wells. Failure to account for ROZ is one of the main reasons for the overestimation of available ore reserves, and often leads to significant material losses when entire geological blocks turn out to be empty. At the moment, there is no fast and reliable way of ROZ identification in the process of GR interpretation. The development of a formal way of such identification is also very problematic. Therefore, ML is one of the likely ways to solve this problem. In the next section, the data and methods that were used to solve the ROZ identification problem using machine learning are discussed.

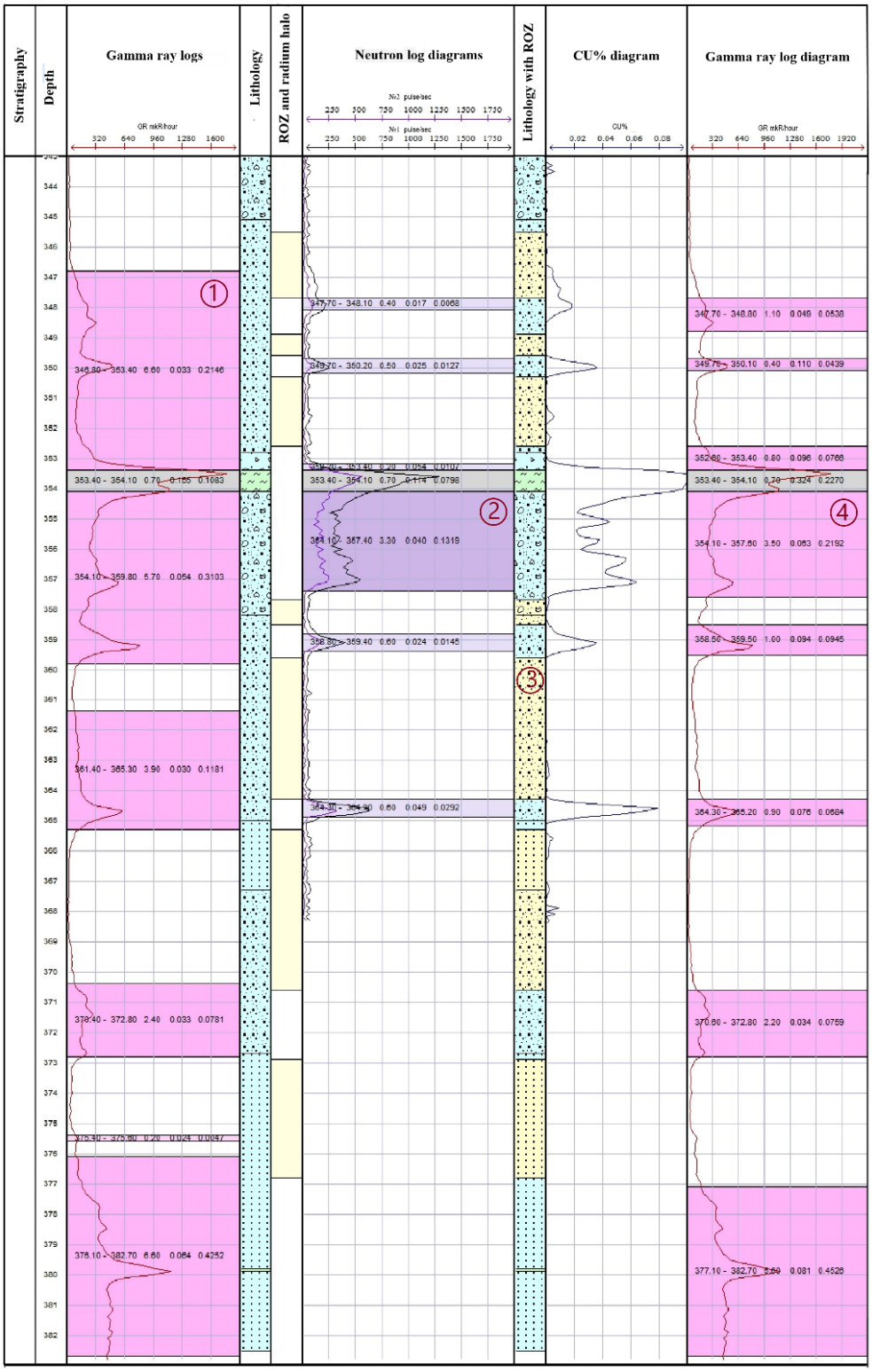


Figure 4. Example of GC reinterpretation after ROZ allocation based on FN results: 1—initial interpretation without ROZ, 2—actual reserves based on FN data, 3—ROZ allocated based on FN results, 4—final interpretation with ROZ taken into account.

4. Data and Methods

4.1. Methodological Research Design

The method of solving the problem of ROZ identification using machine learning methods seems quite obvious. The input of the machine learning model is a set of n

available log data at a certain depth, which form a vector of input values or features $x^{(i)} = (x_0, x_1, \ldots, x_n)$. A total of m training examples $\left\{x^{(1)}, x^{(2)}, \ldots x^{(i)}, \ldots, x^{(m)}\right\} \in X$ are used to train the model. Target value (y) is the rock code, which in this case takes three values: 1—permeable rocks, 2—impermeable rocks, and 8—ROZ. The ML model can afterwards be trained using the pairs $\left(x^{(i)}, y^{(i)}\right)$, and the resulting output is evaluated using the chosen metric. However, this straightforward approach yields rather poor classification performance. Significant improvement can be achieved by pre-training the data and applying feature engineering techniques, which include generating a balanced data set and windows, and using geographically close wells as a source of lithologic information. The structure of the proposed method is illustrated in Figure 5.

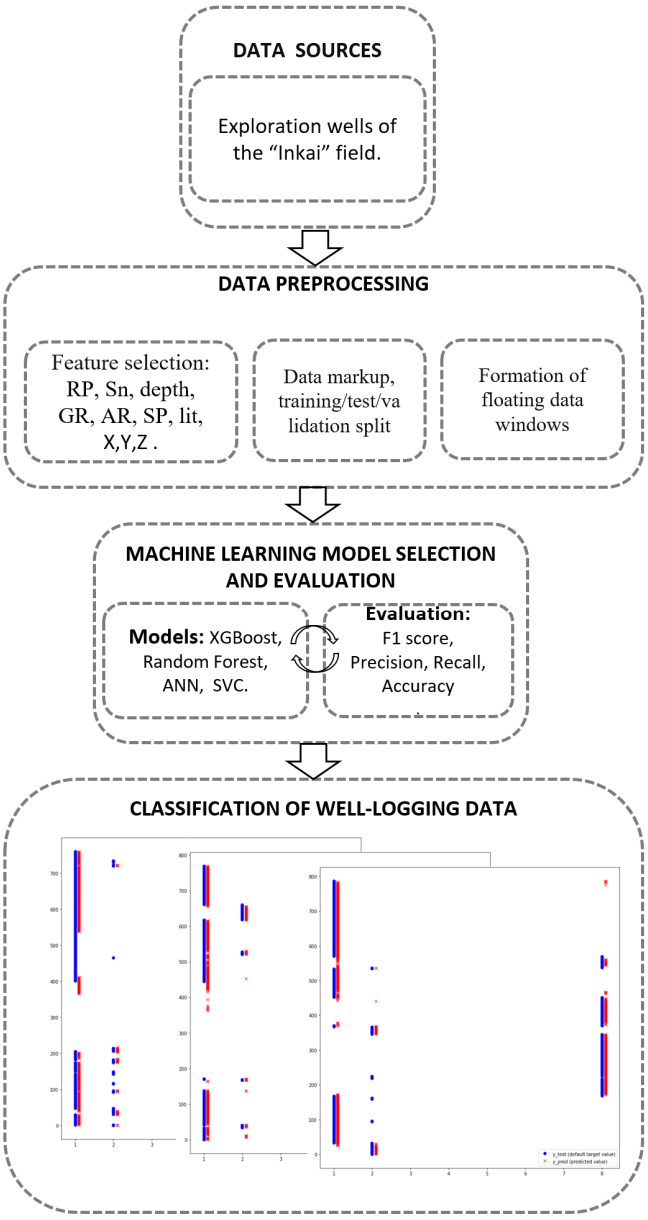


Figure 5. The structure of the proposed method.

A set of 1000 exploration wells with core sampling from the Inkai field was used to solve log data processing tasks. From this set of wells, a special set was manually generated for ROZ studies, input features were identified, and so-called floating data windows were generated for each well. The set of wells was divided into training, test, and validation

wells, and was then used in a cyclic process of machine learning model selection and evaluation. The best performing model was used to interpret the validation dataset and visualize the results.

4.2. Data Collection and Preprocessing

For the task of ROZ identification from the Inkai deposit wells, a special data set was created with the division of wells into three classes: 42 wells without ROZ (LOW_ZPO), 84 wells with a ROZ of 5–50% of the ore-bearing horizon (MEDIUM ZPO), and 42 wells with a ROZ of more than 50% of the ore-bearing horizon (HI_ZPO).

For each well, for the analysis of the data of AR (Ohm × m), SP (mV), and GR (μR × h) logs recorded in 10 cm increments, lithologic intervals (upper boundary, rock code, permeability code, filtration coefficient, and lower boundary) and wellhead coordinates (X, Y, and Z) were used. Since the physical properties of the rocks (in particular, the AR recording level) vary considerably in different geologic horizons, logging data were used only within one horizon (the Inkuduk horizon). Since we used exploratory wells, the ZPO zones were identified in the course of laboratory studies and marked with a geochemical code of 8. Well logging data are recorded in several tables, of which the following indicators are used for further processing (Figure 6):

- RP—ore intersection (value 2 is used for the Inkuduk horizon);
- Sn—well number (used only at the stage of data input and search for the nearest wells);
- depth—depth of 10 cm measurement zone (m);
- GR—gamma ray log (μr/h);
- AR—apparent resistance;
- SP—spontaneous polarization potential;
- lit—rock type by permeability (target value) was assigned based on the laboratory analyses of core samples and their visual inspection: 1—permeable rocks, 2—impermeable rocks, and 8—oxidized rocks (ROZ). We used the classification of rocks by color (geochemistry) adopted in the NAC Kazatomprom standard, where oxidized (yellow) rocks are designated by code 8;
- LIT1—lithologic difference code;
- LIT2—geochemical type of rock;
- X, Y, Z—well coordinates.

The occurrence of ROZ is attributed to the extended groundwater movement, which justifies the use of well coordinates, LIT1 and LIT2.

In the process of initial data generation, the following transformations were performed to increase the accuracy of calculations:

- Formation of the so-called floating data window [27,29] (Figure 6);
- Searching for the lithologic data of the nearest wells and using them as additional input parameters.

The use of floating data windows allowed us to increase the size of the input data vector by a multiple of the window size. For example, if the input vector size is n = 11, then using a window of size h = 3 gives the vector n′ = h × n (n′ = 33) at the model input.

In practice, the size of the floating window is chosen to be either equal to the length of the logging probe (110 cm) or larger. Since each line of the source data describes logging values for a 10 cm layer of rock, the height of the floating data window is equal to the length of the probe h = 11. If each window corresponds to a target value at a depth in the middle of the window, the window is symmetrical. For example, when h = 11, the top (tw) and bottom (bw) of the window may be equal to 5. In general, as computational experiments have shown, the use of symmetric windows is not necessary. About 800 data windows were generated from each well in this way.

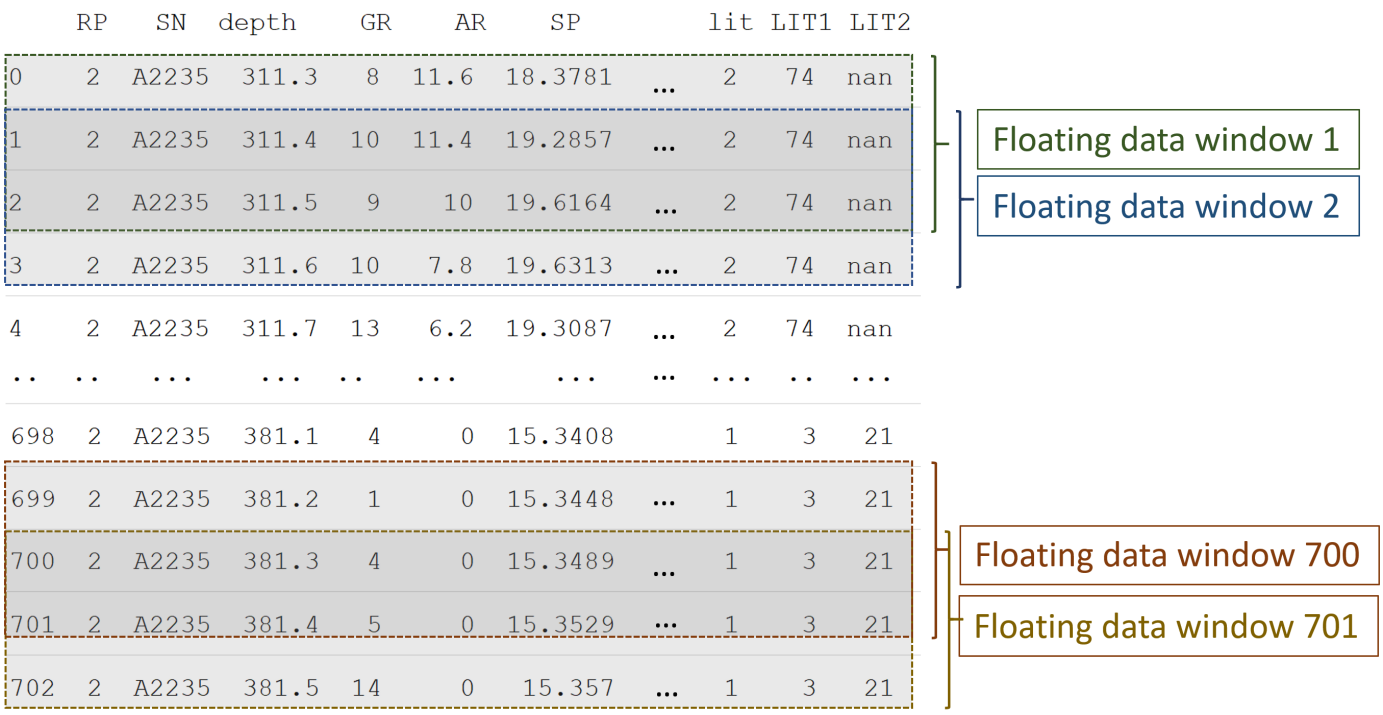


Figure 6. Floating data window with size h = 3.

We suppose that using the data from two nearest wells can improve the quality of the classification. A special program has been developed to test this assumption. The program calculates the distance to the two closest wells of the current well in a given data set and includes in the input vector the lithologic code and the square of the distance to the first (1) and second (2) closest wells: nearest_LIT1_1, nearest_LIT1_2, min_dist_sq1, min_dist_sq. Figure 7 shows a fragment of the data table for one of the wells in the training dataset.

	RP	SN	depth	GR	AR	SP	lit	LIT1	LIT2	X	Y	Z	nearest_LIT1	min_dist_sq	nearest_LIT1_2	min_dist_sq2
0	2	A2235	311.3	8	11.6	18.3781	2	74	nan	337,509.55	105,714.6	213.6	7	3225.1141	7	4145.1781
1	2	A2235	311.4	10	11.4	19.2857	2	74	nan	337,509.55	105,714.6	213.6	7	3225.1141	7	4145.1781
2	2	A2235	311.5	9	10	19.6164	2	74	nan	337,509.55	105,714.6	213.6	7	3225.1141	7	4145.1781
3	2	A2235	311.6	10	7.8	19.6313	2	74	nan	337,509.55	105,714.6	213.6	7	3225.1141	7	4145.1781
4	2	A2235	311.7	13	6.2	19.3087	2	74	nan	337,509.55	105,714.6	213.6	7	3225.1141	7	4145.1781
..	..	...	...	..	...	...	...	..	...	...	...	...	...	...	...	...
698	2	A2235	381.1	4	0	15.3408	1	3	21	337,509.55	105,714.6	213.6	0	3225.1141	1	4145.1781
699	2	A2235	381.2	1	0	15.3448	1	3	21	337,509.55	105,714.6	213.6	0	3225.1141	1	4145.1781
700	2	A2235	381.3	4	0	15.3489	1	3	21	337,509.55	105,714.6	213.6	0	3225.1141	1	4145.1781
701	2	A2235	381.4	5	0	15.3529	1	3	21	337,509.55	105,714.6	213.6	0	3225.1141	1	4145.1781
702	2	A2235	381.5	14	0	15.357	1	3	21	337,509.55	105,714.6	213.6	0	3225.1141	1	4145.1781

Figure 7. Dataset fragment from well A2235.

The formation of training, test, and validation datasets deserves a special discussion. In order to avoid data leakage and, as a consequence, to obtain an overestimated classification quality, data from different wells should be used in the training and test sets. The division of the dataset into test and training was performed as follows. From the N wells available in a particular experiment, 0.1*N co-comprised the test set and 0.9*N the training set. However, since in some experiments N was not large (less than 40), we applied the k-fold validation approach. The division was performed 9 times (k = 9), and the obtained estimates were

averaged. The final model evaluation was performed on the validation set, which was specially formed depending on the experiment objectives (see below) and whose wells were not included in either test or training sets.

4.3. Machine Learning Model Selection and Evaluation

In the process of preliminary experiments, the following list of machine learning algorithms was used: support vector machines [33], artificial neural network [34,35], random forest [36], and eXtreme Gradient Boosting (XGBoost) [37], which are traditionally used to solve lithologic and stratigraphic classification problems [2,31,38,39], as well as LightGBM [40]. It turned out, however, that the most stable results are demonstrated by the ensemble learning methods. Therefore, in the final experiments, the ensemble learning method based on the gradient boosted trees algorithm (XGBClassifier, LightGBM) and the ensemble learning method based on bagging technique (random forest classifier) were used as the machine learning models.

XGBClassifier uses a boosting technique, where the next ensemble algorithm (t) is trained by taking into account the error gradient of the previous algorithm (t − 1). In other words, the subsequent algorithm is tuned so that the target value is not the target value ($\left(y^{(i)}\right)_{i=1}^{m}$ is a target value for the i-th example out of m training examples), but the antigradient of the error function of the previous algorithm.

$$\Delta_{t-1} = -L_{t-1}^{(i)}\left(y^{(i)}, \hat{y}_{t-1}\left(x^{(i)}\right)_{i=1}^{m}\right),$$

where $\hat{y}_{t-1}\left(x^{(i)}\right) = h_{t-1}\left(\theta_{t-1}, x^{(i)}\right)$, $h_{t-1}\left(\theta_{t-1}, x^{(i)}\right)$ is the hypothesis function of the previous algorithm, and θ are the parameters of hypothesis function (the weight of the leaves of the decision tree).

This means that the next training algorithm uses pairs $\left(x^{(i)}, -L_{t-1}^{(i)}\left(y^{(i)}, \hat{y}_{t-1}\left(x^{(i)}\right)\right)\right)$ instead of the traditional pairs $\left(x^{(i)}, y^{(i)}\right)$). The optimal parameters at step t are found by minimizing the cost function of the following type:

$$J_t = \min\left(\sum_{i=1}^{\mathrm{m}} L_t^{(i)}\left(x^{(i)}, \Delta_{t-1}\right) + \Omega_t\right),$$

$\Omega_t = \sum_{i=1}^{t} \Omega_i\left(h_i(\theta_i, x)\right)$ is the sum of regularizers of the type

$\Omega_i = \gamma \mathrm{L}_i + \frac{1}{2}\lambda \sum_{j=1}^{L} \theta_j^2$, where L_i − number of leaves of the tree i; γ—the parameter regulating the division of the leaf into subtrees; and λ—the regularization parameter for the sum of tree weights.

The division of a leaf into subtrees is performed using the function of prediction of the ensemble of T algorithms (trees), which is found as

$$\hat{y} = \sum_{t=1}^{\mathrm{T}} h_t(\theta, x)$$

The RF model uses the bootstrap aggregation technique, where a separate decision tree is constructed for each random subsample of the training dataset based on only a portion of the features. The final result of the classification task is generated by voting between the constructed trees.

$$\hat{y} = \max(T_1, T_2, \ldots T_{n_c}),$$

where T_i is the number of trees that "voted" for class i, and n_c is the number of classes.

To perform the computation experiments, a Python language programming system was developed with the application of numpy, sklearn, mat-plotlib, cv2, alive_progress, pickle, and tensorflow libraries, which solves the tasks of reading and preparation of the initial data and includes 21 functions on data-frame formation, the selection and use of nearby well data, the resizing of the floating data window, the formation of the training and test set of wells, the use of underlying horizon data, the training and evaluation of machine learning models, and others.

Computational experiments were performed on a computer equipped with 32 GB RAM, Intel(R) Core(TM) i7-10750H processor, and discrete video card Nvidia GeForce GTX 1650 Ti. Program example and data sets are given at in the Supplementary Materials.

It should be noted that when using SVC from the Sklearn library, it was not possible to obtain the results during the experiments with a floating data window.

The performance of machine learning models in this case was evaluated using a confusion matrix and measures of accuracy (*Ac*), precision (Precision), completeness (Recall), and harmonic measure (*f1*-Score):

$$Ac = \frac{N_t}{N},$$

where N_t is the number of correct answers, and N is the total number of possible model answers.

$$\text{Precision} : P = \frac{TP}{(TP + FP)}.$$

$$\text{Recall} : R = \frac{TP}{(TP + FN)}.$$

$$F1score = \frac{2 \times P \times \text{R}}{(P + R)}.$$

where true positive (*TP*) and true negative (*TN*) are cases of correct classifier performance, meaning cases where the predicted class matched the expected class. Correspondingly, false negative (*FN*) and false positive (*FP*) are cases of misclassification.

5. Results

The preliminary experiments showed that the application of an ensemble learning method allows to detect ROZs with an accuracy of up to 90% in individual experiments. Nevertheless, for a productive evaluation of machine learning methods, it is necessary to take into account the fact that, depending on the deposit, the amount of data with and without acidification may significantly differ. Therefore, a group of experiments was designed to evaluate the robustness of the methods depending on the content of the training dataset.

First, we evaluated the effect of the size of the floating data window on classification quality (Table 1).

It can be seen that the f1-score increases by 3–4% when the floating window size increases. The most stable results are demonstrated by XGB (f1_macro > 0.7). The AdaBoost and Naive Bayes classifiers demonstrated unsatisfactory results. It is worth noting that the results shown in Tables 2–4 and Appendix A are the average of a k-fold validation at k = 9. In other words, during the computational experiments, the wells were divided into training (90%) and test (10%) well sets nine times. Each time, machine learning models were trained and the results were evaluated. The resulting machine learning model score is the average of these evaluations. The detailed results of the computational experiments are provided in Appendix A (Table A1).

Table 2. Classification quality (f1_macro) when changing the floating window size.

Classifier	tw = 5, bw = 5	tw = 5, bw = 50	tw = 5, bw = 150	tw = 5, bw = 250
LGBM	0.699898	0.728492	0.742686	0.747335
RFC	0.678747	0.679514	0.697662	0.70538
XGB	0.710678	0.732387	0.742319	0.74074
MLP	0.690518	0.700193	0.686201	0.675613
Naive Bayes	0.512381	0.401539	0.376814	0.375069

Note: tw and bw are top and bottom of the floating window.

Table 3. Classification quality (f1_macro) when varying the floating window size. Two additional input parameters are used—the lithologic codes of the two nearest wells.

Classifier	tw = 5, bw = 50	tw = 5, bw = 150
LGBM	0.731947	0.742312
RFC	0.6716	0.699612
XGB	0.73279	0.740219
MLP	0.707169	0.683285

Table 4. Performance of machine learning models with different combinations of training and test data.

RandomForestClassifier		**Validation_Set**		
Train_set		HI_ZPO_val	LOW_ZPO_val	MEDIUM_ZPO_val
HI_ZPO_Train	f1_weighted	0.8876	0.4862	0.5600
	f1_macro	0.6812	0.3740	0.5700
LOW_ZPO_train	f1_weighted	0.0866	0.9296	0.5074
	f1_macro	0.2586	0.6802	0.4662
MEDIUM_ZPO_train	f1_weighted	0.3801	0.8319	0.6943
	f1_macro	0.3586	0.4897	0.6556
HI_LOW_MED_ZPO_train	f1_weighted	0.8901	0.9250	0.7792
	f1_macro	0.6869	0.5801	0.7091
XGBClassifier		**Validation_set**		
Train_set		HI_ZPO_val	LOW_ZPO_val	MEDIUM_ZPO_val
HI_ZPO_Train	f1_weighted	0.8946	0.6070	0.6073
	f1_macro	0.7096	0.4360	0.6285
LOW_ZPO_train	f1_weighted	0.0884	0.9343	0.5017
	f1_macro	0.2613	0.7020	0.4620
MEDIUM_ZPO_train	f1_weighted	0.4463	0.8437	0.7523
	f1_macro	0.3926	0.5164	0.7089
HI_LOW_MED_ZPO_train	f1_weighted	0.8968	0.9373	0.7264
	f1_macro	0.7024	0.7146	0.6701

Second, we have assessed the impact of the nearest well lithology data (Table 3). To save time, the AdaBoost and Naive Bayes classifiers were excluded from further experiments.

The results of this experiment demonstrate that the use of the information about the lithologic composition of the nearest wells practically does not affect the results. This is probably due to the principle of its formation, in which wells were selected based on the percentage

of ZPO content, and not proximity. A slight increase in the results was observed when using non-normalized input data (Appendix A. Table A2). The LGBM classifier was used with default settings. The performance of decision-tree-based classifiers depends on the depth of the tree (max_depth) and the number of trees (n_estimators) [41]. For RFC, the best results are obtained with max_depth = 16, n_estimators = 150. For XGB max_depth = 12. The RFC results are significantly affected by the number of features for splitting selection (max_features), which is chosen as $n_x/3$, where n_x is the number of the input model parameters. The results of the experiments on tuning hyperparameters are given in the Supplementary Materials.

To evaluate the influence of the training dataset, experiments with different ratios of training and test data were performed (see Table 4). In Table 4, HI stands for high ROZ ratio in the dataset, LOW stands for low, and MEDIUM stands for medium. A total of four training datasets with different well ratios were generated:

- HI_ZPO_Train—only wells with high ZPO share (32 wells);
- LOW_ZPO_train—only wells with low ZPO share (32 wells);
- MEDIUM_ZPO_train—only wells with medium ZPO share (32 wells);
- HI_LOW_MED_ZPO_train—mixed set formed from the previous three (96 wells).

These datasets were used for training, and were evaluated on validation sets:

- HI_ZPO_val (10 wells);
- MEDIUM_ZPO_val (10 wells);
- LOW_ZPO_val (10 wells).

When evaluating the results of ML models, it should be taken into consideration that when classifying unbalanced datasets, three variants of estimating the main indicators are possible. In the first case, the quality metric is calculated within the objects of each class and then averaged across all classes (macro average). In the second case, objects of all classes contribute equally to the quality metrics (micro average). In the third case, the contribution of a class to the overall score depends on its size (weighted). In our case, since it is important for us to consider the influence of all classes, macro average and weighted average will give a more objective evaluation. The results of machine learning models are shown in Table 2. The best f1_macro and f1_weighted are highlighted in bold.

6. Discussion

The obtained results allow us to draw important conclusions about the influence of the size and balance of the training dataset on the results. It can be seen that when using a mixed balanced dataset, which includes data from wells with high, medium, and low ROZ, the results are on average no worse than when the type of training dataset coincides with the type of test dataset. If the training set type matches the test set type in terms of ROZ content, the XGBClassifier provides f1_macro estimates of 0.7096, 0.7020, and 0.7089 for HI_ZPO_val, LOW_ZPO_val, and MEDIUM_ZPO_val, respectively. Moreover, the estimates of the validation datasets in which the ROZ ratio is significantly different from the training set range from 0.2613 to 0.6285. If the balanced set HI_LOW_MED_ZPO_train is used for training, the f1_macro estimates for the sets HI_ZPO_val, LOW_ZPO_val, and MEDIUM_ZPO_val are 0.7024, 0.7146, and 0.6701, respectively. In other words, this training set provides a high stability of the model. The XGBClassifier and LGBM show slightly better results compared to the RandomForestClassifier. Certainly, the classification result significantly depends on the classifier settings. The meta parameters of the classifiers were fine-tuned to maximize the model score on one of the test sets. The meta-parameters were then fixed and the series of experiments mentioned above were conducted. Expanding the dataset may provide an opportunity to apply deep learning models and improve classification accuracy. However, this expansion of the dataset is not always possible because the number of exploration wells is limited.

Figure 8 shows examples of classification for six test wells, where expected classes are indicated by numbers: 1—permeable rocks, 2—impermeable rocks, and 8—oxidized rocks (ROZ). Blue shows the actual data, and red shows the predicted values at depths

between 300 and 400 m within the Inkuduk stratigraphic horizon. Each well is divided into approximately 800 sections (interbeds) by depth. Each section has a thickness (thickness) of 10 cm. The minimum depth of the analyzed section of wells is about 300 m, while the maximum depth is about 400 m.

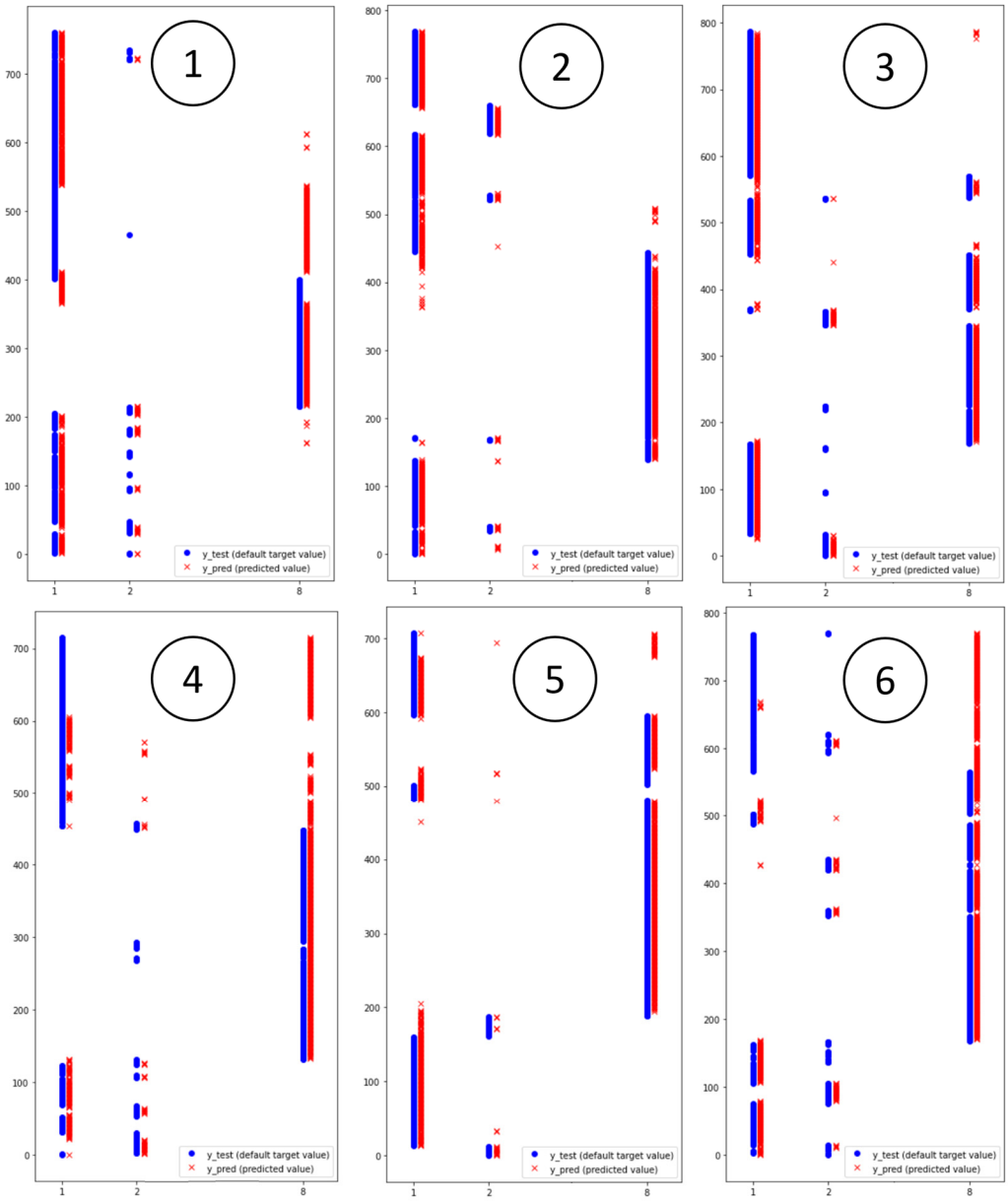


Figure 8. ROZ detection based on log data for six test wells marked in the picture with numbers ①–⑥. Blue are the actual values, while red are the predicted values. The vertical line represents the number of the 10 cm layer, while the horizontal axis represents the rock codes, of which only three are used: 1—permeable rocks, 2—impermeable rocks, and 8—oxidized rocks (ROZ).

It can be seen that, in some cases, the prediction accuracy is very high (wells 2 and 3). However, there are also errors (wells 6 and 4), and overcoming them may be the subject of future research. It can be preliminary noted that since the field is located in sedimentary rocks (sands and clays), core losses during extraction amount to up to 20%. In addition, the error in tying core samples to depth can be 1–2 m. This leads to errors in expert assessments on which models are trained.

The developed model can be used for ROZ identification when interpreting the data of the technological wells in real time. First, the algorithm is trained on exploration wells;

then, when logging data from a new well are obtained, the trained model interprets them in order to obtain rock codes, for example, as in one of the authors' works [27,29]. In practice, when data from the nearby wells become available, the model can be adjusted to use them to improve the classification quality.

7. Conclusions

This article describes a method developed by the authors to identify the formation of the oxidation zones from well logs. To the authors' knowledge, this problem of log data interpretation has not previously been considered in the literature. In contrast to the currently used methods of manual ROZ estimation, the proposed method based on the use of machine learning algorithms is low-cost and fast, and does not require additional logging operations.

To identify the most accurate algorithms, computational experiments were performed. During the experiments, the influence of the size of the floating data window, the use of lithologic data from the nearby wells, and the normalization of input parameters were determined. The results were assessed using k-fold validation. The best results for all sets of input parameters and data preprocessing methods were demonstrated by ensemble machine learning algorithms.

The method has a fairly high accuracy (the value of the harmonic measure f1-score is about 0.7), which allows for a significantly reduction in errors in the calculation of ore reserves and, consequently, improves the economic performance of mining processes.

At the same time, some limitations of the proposed approach can be noted.

7.1. Limitations

1. In some cases, as illustrated above, machine learning algorithms produce incorrect ROZ results.
2. A specific set of wells already marked in terms of ROZ is required to train the machine learning models.

In addition, it remains unclear whether it is possible to apply machine learning models trained on data from one field to ROZ detection in another field.

Therefore, future research can address the following challenges.

7.2. Future Research

1. Improving the accuracy of the method, for example, by applying a deep learning and stacking technique.
2. Evaluation of the accuracy of uranium reserve determination with the application of the developed method.
3. Assessment of the possibilities and limitations of applying the algorithms trained on data from one deposit to identify ROZs in another deposit.

Supplementary Materials: The following high-resolution pictures can be downloaded at: https://www.dropbox.com/sh/jsnwz84orbamwyz/AADCbatPHxAqC-utUN2lEy32a?dl=0, accessed on 10 October 2023. The following data and program example can be downloaded at: https://www.dropbox.com/sh/twon4ukskdqbmgd/AAB4a4Ad2v1_cPtCk7hFUMFpa?dl=0, accessed on 10 October 2023. The results of experiments on tuning hyperparameters can be downloaded at: https://www.dropbox.com/scl/fi/r6tbqdc3vrj08qdlx1h0q/ROZ_tun_4_folders_1.xlsx?rlkey=osc4gej07yxakf6686bqotcdp&dl=0, accessed on 22 October 2023.

Author Contributions: Conceptualization, R.I.M. and Y.K.; methodology, R.I.M.; software, R.I.M. and Y.K.; validation, A.S., Y.P., S.S. and N.Y.; formal analysis, Y.K., K.A. and V.G.; investigation, R.I.M., E.M., A.S. and V.L.; resources, N.L., K.A. and V.G.; data curation, K.A. and N.L.; writing—original draft preparation, R.I.M., Y.K. and E.Z.; writing—review and editing, Y.P., E.M. and S.S.; visualization, K.A., Y.K. and E.M.; supervision, R.I.M. and Y.K.; project administration, N.Y.; funding acquisition, S.S., E.Z. and V.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP14869110), "Improving the accuracy of solving problems of interpretation of geophysical well research data on uranium deposits using machine learning methods", BR21881908 «Complex of urban ecological support», and BR18574144 «Development of a Data Mining System for Monitoring Dams and Other Engineering Structures under the Conditions of Man-Made and Natural Impacts". This research was supported by the Ministry of Education, Science, Research, and Sport of the Slovak Republic, "New approaches of reliability analysis of non-coherent systems" (VEGA 1/0165/21), and by the Slovak Research and Development Agency, "Risk assessment of environmental disturbance using Earth observation data" (reg.no. SK-UA-21-0037); and develops ideas of the project "Earth Observation for Early Warning of Land Degradation at European Frontier (EWALD)" under the European Union's Framework Programme for Research and Innovation Horizon Europe (Grant Agreement No. ID 101086250).

Data Availability Statement: The data presented in this study are openly available in https://www.dropbox.com/sh/twon4ukskdqbmgd/AAB4a4Ad2v1_cPtCk7hFUMFpa?dl=0, accessed on 10 October 2023.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Ac	Accuracy
AdaBoost	Adaptive boosting
AR	Apparent resistance logging
CatBoost	Categorical boosting
DT	Decision tree
FFNN	Feedforward neural network
FN	Fission neutron logging
GR	Gamma ray logging
GRB	Geophysical research of boreholes
IL	Induction log
ISL	In situ leaching (ISL)
K-NN	K-nearest neighbors
LGBM	Light gradient boosting machine
ML	Machine learning
MLP	Multilayer perceptron
RBF	Radial basis function neural network
RF	Random forest
ROZ	Reservoir oxidation zone
SP	Spontaneous polarization potential
SVM	Support vector machine
XGB	eXtreme Gradient Boosting
Kpp	Radioactive equilibrium coefficient

Appendix A. Results of Calculation Experiments

Table A1. Results of computational experiments (tw = 5, bw = 5).

Floating Data Window	Classifier	Accuracy	f1_Score_Macro	f1_Score_Micro	Duration
tw = 5, bw = 5	LGBM	0.81633	0.699898	0.81633	2.206632
	RFC	0.807601	0.678747	0.807601	20.00478
	XGB	0.818902	0.710678	0.818902	50.16421
	Neural Net	0.81206	0.690518	0.81206	43.8792

Table A1. *Cont.*

Floating Data Window	Classifier	Accuracy	f1_Score_Macro	f1_Score_Micro	Duration
	AdaBoost	0.783008	0.632727	0.783008	56.24385
	Naive Bayes	0.564419	0.512381	0.564419	0.587876
tw = 5. bw = 50	LGBM	0.826549	0.728492	0.826549	27.71743
	RFC	0.808657	0.679514	0.808657	99.984
	XGB	0.828718	0.732387	0.828718	178.4619
	Neural Net	0.809933	0.700193	0.809933	58.01762
	AdaBoost	0.763817	0.614482	0.763817	88.14933
	Naive Bayes	0.438764	0.401539	0.438764	2.138282
tw = 5, bw = 150	LGBM	0.83064	0.742686	0.83064	27.74648
	RFC	0.816573	0.697662	0.816573	271.1427
	XGB	0.831109	0.742319	0.831109	393.3063
	Neural Net	0.78915	0.686201	0.78915	142.7333
	AdaBoost	0.78848	0.644886	0.78848	1202.09
	Naive Bayes	0.424534	0.376814	0.424534	11.7524
tw = 5, bw = 250	LGBM	0.83453	0.747335	0.83453	40.39162
	RFC	0.824367	0.70538	0.824367	395.0657
	XGB	0.831407	0.74074	0.831407	651.8487
	Neural Net	0.765472	0.675613	0.765472	101.8281
	AdaBoost	0.79016	0.630911	0.79016	936.2833
	Naive Bayes	0.45133	0.375069	0.45133	8.438352

Table A2. Results of computational experiments for non-normalized input data.

Floating Data Window	Classifier	Accuracy	f1_Score_Macro	f1_Score_Micro	Duration
tw = 5, bw = 5	LGBM	0.841322	0.735884	0.841322	7.742824
	RFC	0.820061	0.708707	0.820061	22.54099
	XGB	0.82707	0.722047	0.82707	42.69652
tw = 5, bw = 50	LGBM	0.843431	0.746168	0.843431	33.5640
	RFC	0.822069	0.705415	0.822069	113.2531
	XGB	0.829053	0.731556	0.829053	176.0672
tw = 5, bw = 150	LGBM	0.845664	0.752963	0.845664	95.00122
	RFC	0.82606	0.709584	0.82606	317.3386
	XGB	0.836221	0.745404	0.836221	407.7704

References

1. Annual Report of Kazatomprom. JSC National Atomic Company. Available online: https://www.kazatomprom.kz/storage/f4/kazatomprom_iar_2022_rus.pdf (accessed on 24 October 2023).
2. Mukhamediev, R.I.; Kuchin, Y.; Amirgaliyev, Y.; Yunicheva, N.; Muhamedijeva, E. Estimation of Filtration Properties of Host Rocks in Sandstone-Type Uranium Deposits Using Machine Learning Methods. *IEEE Access* **2022**, *10*, 18855–18872. [CrossRef]
3. Kuchin, Y.; Mukhamediev, R.; Yunicheva, N.; Symagulov, A.; Abramov, K.; Mukhamedieva, E.; Zaitseva, E.; Levashenko, V. Application of Machine Learning Methods to Assess Filtration Properties of Host Rocks of Uranium Deposits in Kazakhstan. *Appl. Sci.* **2023**, *13*, 10958. [CrossRef]

4. Baldwin, J.L.; Bateman, R.M.; Wheatley, C.L. Application of a neural network to the problem of mineral identification from well logs. *Log Anal.* **1990**, *3*, 279–293.
5. Poulton, M.M. (Ed.) *Computational Neural Networks for Geophysical Data Processing*; Elsevier: Amsterdam, The Netherlands, 2001. Available online: https://www.researchgate.net/profile/Mary-Poulton/publication/245744530_Computational_Neural_Networks_for_Geophysical_Data_Processing/links/5730b09508ae100ae55740fe/Computational-Neural-Networks-for-Geophysical-Data-Processing.pdf (accessed on 24 October 2023).
6. Benaouda, D.; Wadge, G.; Whitmarsh, R.; Rothwell, R.; MacLeod, C. Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs: An example from the Ocean Drilling Program. *Geophys. J. Int.* **1999**, *136*, 477–491. [CrossRef]
7. Saggaf, M.; Nebrija, E.L. Estimation of missing logs by regularized neural networks. *AAPG Bull.* **2003**, *87*, 1377–1389. [CrossRef]
8. Kumar, T.; Seelam, N.K.; Rao, G.S. Lithology prediction from well log data using machine learning techniques: A case study from Talcher coalfield, Eastern India. *J. Appl. Geophys.* **2022**, *199*, 104605. [CrossRef]
9. Kim, J. Lithofacies classification integrating conventional approaches and machine learning technique. *J. Nat. Gas Sci. Eng.* **2022**, *100*, 104500. [CrossRef]
10. Thongsamea, W.; Kanitpanyacharoena, W.; Chuangsuwanich, E. Lithological Classification from Well Logs using Machine Learning Algorithms. *Bull. Earth Sci. Thail.* **2018**, *10*, 31–43.
11. Liang, H.; Xiong, J.; Yang, Y.; Zou, J. Research on Intelligent Recognition Technology in Lithology Based on Multi-Parameter Fusion. 2023. Available online: https://www.researchsquare.com/article/rs-3243742/v1 (accessed on 26 October 2023).
12. Mohamed, I.M.; Mohamed, S.; Mazher, I.; Chester, P. Formation lithology classification: Insights into machine learning methods. In Proceedings of the SPE Annual Technical Conference and Exhibition, Calgary, AB, Canada, 30 September–2 October 2019.
13. Sakhnyuk, V.; Novikov, E.; Sharifullin, A.; Belokhin, V.; Antonov, A.; Karpushin, M.; Bolshakova, M.; Afonin, S.; Sautkin, R.; Suslova, A. Application of machine learning methods in processing data from geophysical studies of wells in the Vikulovsky formation. *Georesursy* **2022**, *24*, 230–238. (In Russian) [CrossRef]
14. Ahmadi, M.-A.; Ahmadi, M.R.; Hosseini, S.M.; Ebadi, M. Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: Application of artificial intelligence. *J. Pet. Sci. Eng.* **2014**, *123*, 183–200. [CrossRef]
15. Gholami, R.; Moradzadeh, A.; Maleki, S.; Amiri, S.; Hanachi, J. Applications of artificial intelligence methods in prediction of permeability in hydrocarbon reservoirs. *J. Pet. Sci. Eng.* **2014**, *122*, 643–656. [CrossRef]
16. Zhong, Z.; Carr, T.R.; Wu, X.; Wang, G. Application of a convolutional neural network in permeability prediction: A case study in the Jacksonburg-Stringtown oil field, West Virginia, USA. *Geophysics* **2019**, *84*, B363–B373. [CrossRef]
17. Khan, H.; Srivastav, A.; Kumar Mishra, A.; Anh Tran, T. Machine learning methods for estimating permeability of a reservoir. *Int. J. Syst. Assur. Eng. Manag.* **2022**, *13*, 2118–2131. [CrossRef]
18. Talebkeikhah, M.; Sadeghtabaghi, Z.; Shabani, M. A comparison of machine learning approaches for prediction of permeability using well log data in the hydrocarbon reservoirs. *J. Hum. Earth Future* **2021**, *2*, 82–99. [CrossRef]
19. Akhmetsafin, R.; Akhmetsafina, R. Applying Machine Learning Methods to Predict or Replace Missing Logging Data. *J. Instrum. Eng.* **2019**, *66*, 532–541. (In Russian) [CrossRef]
20. Priezzhev, I.; Stanislav, E. Application of machine learning algorithms using seismic data and well logs to predict reservoir properties. In Proceedings of the 80th EAGE Conference and Exhibition 2018, Copenhagen, Denmark, 11–14 June 2018; pp. 1–5.
21. Fajana, A.O.; Ayuk, M.A.; Enikanselu, P.A. Application of multilayer perceptron neural network and seismic multiattribute transforms in reservoir characterization of Pennay field, Niger Delta. *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 31–49. [CrossRef]
22. Cracknell, M.J.; Reading, A.M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* **2014**, *63*, 22–33. [CrossRef]
23. Kumar, C.; Chatterjee, S.; Oommen, T.; Guha, A. Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102006. [CrossRef]
24. Deng, C.; Pan, H.; Fang, S.; Konaté, A.A.; Qin, R. Support vector machine as an alternative method for lithology classification of crystalline rocks. *J. Geophys. Eng.* **2017**, *14*, 341–349. [CrossRef]
25. Farhadi, S.; Afzal, P.; Boveiri Konari, M.; Daneshvar Saein, L.; Sadeghi, B. Combination of Machine Learning Algorithms with Concentration-Area Fractal Method for Soil Geochemical Anomaly Detection in Sediment-Hosted Irankuh Pb-Zn Deposit, Central Iran. *Minerals* **2022**, *12*, 689. [CrossRef]
26. Afzal, P.; Farhadi, S.; Boveiri Konari, M.; Shamseddin Meigooni, M.; Daneshvar Saein, L. Geochemical anomaly detection in the Irankuh District using Hybrid Machine learning technique and fractal modeling. *Geopersia* **2022**, *12*, 191–199.
27. Kuchin, Y.I.; Mukhamediev, R.I.; Yakunin, K.O. One method of generating synthetic data to assess the upper limit of machine learning algorithms performance. *Cogent Eng.* **2020**, *7*, 1718821. [CrossRef]
28. Merembayev, T.; Yunussov, R.; Yedilkhan, A. Machine learning algorithms for stratigraphy classification on uranium deposits. *Procedia Comput. Sci.* **2019**, *150*, 46–52. [CrossRef]
29. Kuchin, Y.; Mukhamediev, R.; Yakunin, K. Quality of data classification in conditions of inconsistency of expert assessments. *Cloud Sci.* **2019**, *6*, 109–126. (In Russian)
30. Viggen, E.M.; Merciu, I.A.; Løvstakken, L.; Måsøy, S.-E. Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *J. Pet. Sci. Eng.* **2020**, *195*, 107539. [CrossRef]

31. Kuchin, Y.I.; Mukhamediev, R.I.; Yakunin, K.; Grundspenkis, J.; Symagulov, A. Assessing the Impact of Expert Labelling of Training Data on the Quality of Automatic Classification of Lithological Groups Using Artificial Neural Networks. *Appl. Comput. Syst.* **2020**, *25*, 145–152. [CrossRef]
32. Peculiarities of Formation of Uranium Deposits of Sandstone Type. Available online: https://studref.com/546203/geografiya/osobennosti_formirovaniya_mestorozhdeniy_urana_peschanikovogo_tipa (accessed on 24 October 2023).
33. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
34. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
35. Galushkin, A.I. Neural networks. Basics of the theory. *Monograph* **2012**, 497. (In Russian)
36. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
37. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
38. Chen, Y.; Wu, W. Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochem. Explor. Environ. Anal.* **2017**, *17*, 231–238. [CrossRef]
39. Caté, A.; Schetselaar, E.; Mercier-Langevin, P.; Ross, P.-S. Classification of lithostratigraphic and alteration units from drillhole lithogeochemical data using machine learning: A case study from the Lalor volcanogenic massive sulphide deposit, Snow Lake, Manitoba, Canada. *J. Geochem. Explor.* **2018**, *188*, 216–228. [CrossRef]
40. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
41. Krasnova, I.A. Analysis of the Influence of Parameters of Machine Learning Algorithms on the Results of Traffic Classification in Real Time. *T-Comm-Telecommun. Transp.* **2021**, *15*, 24–35. Available online: https://cyberleninka.ru/article/n/analiz-vliyaniya-parametrov-algoritmov-machine-learning-na-rezultaty-klassifikatsii-trafika-v-rezhime-realnogo-vremeni/viewer (accessed on 24 October 2023). (In Russian) [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.