ISSN 2712-8008
Volume / Том 23 №3 2023
Journal Homepage: http://vestnik.pstu.ru/geo/

Perm Journal of Petroleum and Mining Engineering

UDC 622.276
Article / Статья

Assessment of Rock Geochemical Properties of the Bazhenov Formation According to Well Logging Data Using Machine Learning Methods

Arsenii O. Shadrin[1], Aleksandr B. Gulin[1], Sergey G. Ashikhmin[2], Aleksandr A. Melekhin[2]

[1]KogalymNIPIneft branch of LUKOIL-Engineering LLC in Tyumen (19 Tsentral'naya st., Kogalym, 628481, Russian Federation)
[2]Perm National Research Polytechnic University (29 Komsomolskiy av., Perm, 614990, Russian Federation)

Оценка геохимических свойств пород Баженовской свиты по данным ГИС методами машинного обучения

А.О. Шадрин[1]. А.Б. Гулин[1]. С.Г. Ашихмин[2]. А.А. Мелехин[2]

[1]Филиал ООО «ЛУКОЙЛ-Инжиниринг» «КогалымНИПИнефть» в г. Тюмени (Россия, 628481, г. Когалым, ул. Центральная, 19)
[2]Пермский национальный исследовательский политехнический университет (Россия, 614990, г. Пермь, Комсомольский проспект, 29)

Received / Получена: 04.04.2023. Accepted / Принята: 28.08.2023. Published / Опубликована: 25.03.2024

Keywords:
geochemical parameters. Bazhenov formation. database. geochemical studies. regression models. machine learning. pyrolysis. Rock-Eval. Boosted Trees.

The Bazhenov formation in the West Siberian oil and gas province is of interest to researchers not only as a potential industrial development object. but also in connection with forecasting oil content and assessing the risks of exploratory drilling. The results of a study aimed at developing machine learning models to assess the relationships between the geochemical parameters of the Bazhenov formation and data obtained from well logging are presented.
Modern machine learning methods provide powerful tools for data analysis and forecasting. Geological data is often characterized by a large number of parameters and complex relationships that can be difficult for humans to understand. In this context. the application of machine learning methods provides researchers with new tools.
This work focuses on the Boosted Trees model. which demonstrated better results compared to other regression methods. having the lowest error (MAE and MSE) and the highest coefficient of determination (R^2). Boosted Trees models provided accurate and stable prediction results. which were confirmed by visual analysis.
As a result of the work. using machine learning methods. geochemical parameters were calculated. including S_1. S_2 and T_{max}. which was previously difficult. This increased the accuracy of the geochemical parameters assessment in wells. which contributed to the improvement of models for the rock properties distribution along the well section. and in the future will lead to an increase in the detail of the parameters distribution over the area.
It also emphasizes the importance of careful data preprocessing and selecting the best models and learning methods. Despite its technical challenges. machine learning provides researchers with a powerful tool to more accurately analyze and interpret geodata and make operational decisions based on this data.

Ключевые слова:
геохимические параметры. Баженовская свита. база данных. геохимические исследования. регрессионные модели. машинное обучение. пиролиз. Rock-Eval. Boosted Trees.

Баженовская свита в Западно-Сибирской нефтегазоносной провинции представляет интерес для исследователей не только как потенциально промышленный объект разработки. но и в связи с прогнозированием нефтеносности и оценкой рисков поискового бурения. Представлены результаты исследования. нацеленного на разработку моделей машинного обучения для оценки зависимостей между геохимическими параметрами Баженовской свиты и данными. полученными в результате геофизических исследований скважин.
Современные методы машинного обучения предоставляют собой мощные инструменты для анализа данных и прогнозирования. Геологические данные часто характеризуются большим числом параметров и сложными взаимосвязями. которые могут быть труднопонимаемыми для человека. В этом контексте применение методов машинного обучения обеспечивает исследователям новые инструменты.
В данной работе уделено внимание модели Boosted Trees. которая продемонстрировала лучшие результаты по сравнению с другими методами регрессии. обладая наименьшей ошибкой (MAE и MSE) и наивысшим коэффициентом детерминации (R^2). Модели Boosted Trees обеспечивают точные и стабильные результаты прогнозирования. что было подтверждено визуальным анализом.
В результате проведенной работы благодаря применению методов машинного обучения был проведен расчет геохимических параметров. включая S_1. S_2 и T_{max}. что ранее было затруднительно. Это повысило точность оценки геохимических параметров в скважинах. что способствовало улучшению моделей распределения свойств пород по разрезу скважин. а в дальнейшем приведет к повышению детальности картины распределения параметров по площади.
Также подчеркивается важность тщательной предобработки данных. выбора наилучших моделей и методов обучения. Несмотря на технические сложности. машинное обучение предоставляет исследователям мощный инструмент для более точного анализа и интерпретации геоданных. а также для принятия производственных решений на основе этих данных.

© **Arsenii O. Shadrin** (Author ID in Scopus: 57222469593. ORCID: 0000-0003-2736-0252) – 1st Category Geologist (tel.: +007 (995) 093 94 53. e-mail: Arseni.Shadrin@lukoil.com).
© **Aleksandr B. Gulin** – Lead Geologist (tel.: +007 (918) 742 83 43. e-mail: Alexandr.Gulin@lukoil.com).
© **Sergey G. Ashikhmin** (Author ID in Scopus: 6603057955. ORCID: 0000-0001-7850-3415) – Doctor in Engineering. Professor at the Department of Mine Surveying. Geodesy and Geoinformation Systems (tel.: +007 (342) 219 84 22. e-mail: A_s_g_perm@mail.ru).
© **Aleksandr A. Melekhin** (Author ID in Scopus: 55531747500. ORCID: 0000-0002-0737-1360) – PhD in Engineering. Associate Professor at the Department of Oil and Gas Technologies (tel.: +007 (342) 219 82 07. e-mail: melehin.sasha@mail.ru).

© **Шадрин Арсений Олегович** (ORCID: 0000-0003-2736-0252) – геолог 1-й категории (тел.: +007 (995) 093 94 53. e-mail: Arseni.Shadrin@lukoil.com).
© **Гулин Александр Борисович** – ведущий геолог (тел.: +007 (918) 742 83 43. e-mail: Alexandr.Gulin@lukoil.com).
© **Ашихмин Сергей Геннадьевич** (ORCID: 0000-0001-7850-3415) – доктор технических наук. профессор кафедры «Маркшейдерское дело. геодезия и геоинформационные системы» (тел.: +007 (342) 219 84 22. e-mail: A_s_g_perm@mail.ru).
© **Мелехин Александр Александрович** (ORCID: 0000-0002-0737-1360) – кандидат технических наук. доцент кафедры «Нефтегазовые технологии» (тел.: +007 (342) 219 82 07. e-mail: melehin.sasha@mail.ru).

Please cite this article in English as:
Shadrin A.O.. Gulin A.B.. Ashikhmin S.G.. Melekhin A.A. Assessment of Rock Geochemical Properties of the Bazhenov Formation According to Well Logging Data Using Machine Learning Methods. *Perm Journal of Petroleum and Mining Engineering.* 2023. vol.23. no.3. pp.141-150. DOI: 10.15593/2712-8008/2023.3.5

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:
Шадрин А.О.. Гулин А.Б.. Ашихмин С.Г.. Мелехин А.А. Оценка геохимических свойств пород Баженовской свиты по данным ГИС методами машинного обучения // Недропользование. – 2023. – Т.23. №3. – С.141–150. DOI: 10.15593/2712-8008/2023.3.5

Introduction

This work is part of an integrated study of Jurassic deposits in the northern part of the Surgut arch aimed at examining and evaluating the geochemical and source rock properties of the Bazhenov Formation regarding its influence on the oil-bearing potential of the Vasyugan Formation [1–5].

In this study, the aime is to develop models to assess the dependencies between the geochemical parameters of the Bazhenov Formation rocks and the data obtained from logging. For this purpose, various methods of machine learning (ML) were used in the research [6–15].

Currently, modern machine learning methods provide powerful tools for data analysis and forecast, and their application is also relevant in solving geological problems. In this field, standard practice involves working with large data volumes, complex interrelationships between geological parameters and the heterogeneity of geological formations. In this context, the use of machine learning methods is a new tool that significantly improves the analysis processes and model development based on geodata. The major advantage of machine learning application in geological research is its ability to automate and optimize data analysis. Geological data is often characterized by a multitude of parameters, and their interrelationships can be complex for humans to understand. Machine learning models, through more advanced algorithms for identifying correlation relationships and discriminant features, can increase the reliability of forecasts and data assessment models. These models can be used to forecast various parameters, such as the quality and composition of rocks, the distribution of mineral resources, hydrogeological characteristics, and for other purposes.

In this study, geochemical properties of rocks were taken as the major parameters for establishing dependencies. These data were obtained using the Rock-Eval pyrolysis method, which determines the content and composition of organic matter in rocks. The high cost does not allow core sampling in a large number of wells, which increases the value of the collected core material. In contrast, the relatively low cost and high information content of geophysical surveys have allowed a large share of the well stock to be characterized by a relatively unified set of studies, facilitating comparison of the geophysical rock properties from different wells. Establishing dependencies between geochemical and geophysical properties will enable a qualitative and quantitative assessment of the studied rock properties, as well as expand the number of wells where geochemical properties can be assessed.

Object of Study. Research methods

The Bazhenov Formation is widely regarded as the main hydrocarbon-generating formation of the West Siberian petroleum province. The deposits were predominantly formed in deep-water marine sedimentation conditions and are widespread across the territory of Western Siberia. In this study, the main characteristics of the rocks obtained through the Rock-Eval pyrolysis method were studied. The pyrolysis method is based on the thermal decomposition of core samples, and the recorded chromatogram analysis allows for obtaining data such as the content of free hydrocarbons (S_1), residual generation potential (S_2), total organic carbon (TOC), and the temperature of maximum hydrocarbon yield during cracking (T_{max}). The obtained characteristics enable conclusions to be drawn about the oil generation productivity of the rock, its maturity and the amount of dispersed organic matter. Also, assessment of rock's hydrocarbon potential has been realized [16–19].

The issues of interpreting well logging data in unconventional reservoirs, including the challenges of assessing geochemical properties based on geophysical research data, have been covered in a large number of studies, particularly in works [19–35].

The dependencies were identified with parameters obtained from geophysical well surveys. The standard set of well logging measurements used as predictor variables in most wells included: GK (gamma logging), NKT (thermal neutron logging), BK (resistivity logging), IK (induction logging), as well as derived parameters from these measurements (lnBK, lnIK, GK/NKT, ln[GK/NKT]).

Obtaining "core-well logging" dependencies will allow for more accurate averaging of geochemical data in wells with geochemical studies, assessment of rock properties in wells lacking core data, and model construction of property distribution across the area based on well data.

The analysis was carried out using a database of 1,217 studied samples taken from wells in the territory of the Surgut and Vartovsk arches, which were linked to intervals of the Bazhenov Formation or stratigraphically similar deposits.

The methods used to search for dependencies in the work included the multiple linear regression method, as well as three machine learning methods: the support vector machine (SVM), neural networks, and the boosted trees method. Detailed characteristics and distinctive features of different methods are presented in numerous papers, particularly in works [36–44].

A distinctive feature of "core-well logging" dependencies is the different vertical resolution of the methods used. Pyrolytic studies of core samples are conducted with a diameter of about 1 cm, which increases the resolution of the method but can also lead to anomalously low or high values that are not representative of the entire interval of sediments. In contrast, well-logging has a greater discretization step, where there is only one cut per 10 cm measurement. To level out this difference, the geochemical research data were smoothed using a sliding window method within a ± 1 m interval. This approach allowed for identifing a relatively high-frequency trend in the changes of geochemical properties and reducing the number of anomalous values.

Next, the geophysical data was standardized, for which the normalization method described in the work of D.E. Shira [45] was used.

According to this method, the normalization process can be performed using the following equation:

$$V_{norm} = R_{min} + \left(R_{max} - R_{min}\right) \cdot \left(\frac{V_{log} - W_{min}}{W_{max} - W_{min}}\right).$$

Table 1

Statistical characteristics of parameters

Variable	Arithmetic mean	Minimum	Maximum	Standard deviation	Asymmetry	Excess
S_1r, mg HC/g.	3.2968	0.0000	12.45	2.6387	0.50229	-0.5573
S_2, mg HC/g.	28.5440	0.0100	148.57	26.1681	0.68483	-0.1650
T_{max}, °C	436.8596	403.0000	458.00	6.8586	-1.96788	5.1094
TOC, %	5.9896	0.0900	24.23	4.9310	0.67065	-0.2007
BK, Ohm	265.6223	0.7220	10600.70	630.8205	8.44372	104.7258
IK, mSm	76.1296	0.1000	2018.85	126.3915	7.08556	77.8917
GK, mkrRG/h	23.0309	2.2610	111.80	16.6028	1.04645	0.5909
NKT, eV	4.2373	1.1670	16.39	2.2635	1.49880	2.2558
lnBK	4.1817	-0.3257	9.27	1.7960	0.08127	-1.0821
lnIK	3.5168	-2.3026	7.61	1.3844	-0.38762	0.5261
GK/NKT, mcRG/h·eV	8.0811	0.1836	53.44	8.2495	1.79523	3.8280
lnGK/NKT	1.5474	-1.6947	3.98	1.1257	-0.27131	-0.6440

In the equation above, the normalized value of the curve (V_{norm}) is calculated from the unnormalized curve value (V_{log}), the minimum and maximum values from the reference curve (R_{min} and R_{max}), as well as the minimum and maximum values from the curve that needs to be normalized (W_{min} and W_{max}). Typically, the 5th and 95th percentiles are used instead of the minimum and maximum values. This reduces the influence of data outliers (anomalies) that could lead to incorrect results. However, careless application of such methods may exclude some important data from the analysis that reflects the actual properties of the rocks.

After performing the procedures of averaging and normalizing the data, a training sample of 2109 values was obtained. The statistical characteristics of the parameters are presented in Table 1.

Choosing the optimal ratio between training and testing samples in machine learning issues is an important aspect of model-building methodology that requires logical justification. This choice lies at the center of balancing two key aspects: maximizing the generalization ability of models trained on data and ensuring a reliable evaluation of their performance on independent data.

In this research work, a ratio of 70 % training data to 30 % testing data was decided to be applied in the context of training machine learning models. Firstly, having 70 % of the training data provides the models with a substantial amount of information to learn the major patterns in the data. At the same time, 30 % of the data is reserved for testing purposes and subsequent performance evaluation of the models.

This approach helps to balance the training process and model assessment, which helps to avoid overfitting, as models adapt too closely to the training data and fail to generalize adequately to new data.

Considering the accepted ratio between the training and testing samples, experimental training of various machine learning models was conducted, followed by their performance assessment on the test data. The results obtained indicate that the chosen 70/30 ratio was optimal for this task and provided the best practical results.

The optimal ratio between training and testing samples may vary depending on the data characteristics and the specific task. Nonetheless, in this work, the choice of this ratio is justified by achieving high performance of machine learning models while ensuring an optimal assessment of their performance.

The comparison and analysis of the developed models were conducted using standard methods, including visual analysis of cross-plots and graphs comparing actual and calculated data, correlation coefficient assessment, as well as a quantitative assessment of metrics used for analyzing the reliability of machine learning regression models. Specifically, the metrics considered were Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and R^2.

The MAE metric (Mean Absolute Error) is a way to measure how much a model or forecast deviates from the actual data. It measures the average absolute difference between the predictions and the actual values. In other words, MAE shows on average how much the model is wrong, simply by summing all the absolute differences and averaging them, its formula is:

$$MAE\left(y^{true}, y^{pred}\right) = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - f\left(x_i\right)\right|.$$

MAPE (Mean Absolute Percentage Error) is a metric used to assess the accuracy of forecasts or models. It measures the percentage error of the predictions compared to the actual values of the data, its formula is:

$$MAPE\left(y^{true}, y^{pred}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_i - f\left(x_i\right)\right|}{\left|y_i\right|}.$$

Table 2

Best independent variable for continuous dependence

Parameter	S_1		S_2		TOC		T_{max}	
	F-value	*p*-value	*F*-value	*p*-value	*F*-value	*p*-value	*F*-value	*p*-value
IK	128.565	< 10–6	93.6135	< 10–6	70.0153	< 10–6	87.0745	< 10–6
BK	94.7325	< 10–6	81.4088	< 10–6	64.8119	< 10–6	79.9992	< 10–6
GK	88.6548	< 10–6	150.454	< 10–6	115.153	< 10–6	68.8253	< 10–6
GK/NKT	90.0042	< 10–6	184.877	< 10–6	141.035	< 10–6	58.5804	< 10–6
lnGK/NKT	61.9647	< 10–6	124.936	< 10–6	102.278	< 10–6	45.8759	< 10–6
NKT	49.1096	< 10–6	96.9236	< 10–6	88.2218	< 10–6	39.3733	< 10–6
lnIK	1.51223	0.15865	1.73305	0.09718	0.75432	0.62586	0.63281	0.72908
lnBK	1.76144	0.10343	1.63941	0.13251	1.54953	0.15834	0.18478	0.98111

MSE, or Mean Squared Error, is a metric used to measure how close predicted values are to actual values in regression tasks. In other words, it measures the average of the squared differences between predicted and actual values, its formula is:

$$MSE\left(y^{true}, y^{pred}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f\left(x_i\right)\right)^2,$$

R^2 metric (or coefficient of determination) is a statistical measure that helps assess how well a model (such as linear regression) fits the data. It measures the proportion of variance (spread) in the dependent variable that the model explains. In other words, R^2 shows how close the model's predictions are to the actual data. Its formula, is:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - f\left(x_i\right))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}.$$

Research results

To perform the task of foreastig the values of dependent variables (S_1, S_2, TOC, T_{max}) based on independent variables (BK, IK, GK, NKT, lnBK, lnIK, GK/NKT, ln(GK/NKT)), the software product from StatSoft was chosen."

During the first iteration, after loading the training sample, feature selection was performed using the Feature Selection method. In this iteration, the criteria such as the *F*-value and *p*-value were calculated for each independent variable and each dependent variable in various learning methods. These criteria served for identifying the most significant variables, while variables with low statistical significance were excluded from further analysis.

Results of the analysis are presented in Table 2, which lists the *F*-values and *p*-values for each variable. The analysis of the table, made it possible to draw the following conclusions:

• significant variables: from the *p*-value analysis, the variables IK, BK, GK/NKT, GK, GK/NKT and NKT are identified, which have *p*-values significantly below 0.05 This indicates a statistically significant relationship between these variables and the dependent variables;

• insignificant variables: variables lnBK and lnIK have higher *p*-values, indicating a lack of statistically significant association with the dependent variable, making their inclusion in the models unjustified;

• *F*-test values: The *F*-test values for IK and BK are higher compared to other variables, indicating a stronger relationship between them and the dependent variable.

To achieve optimal performance and results in the data analysis task, the following initial model settings were chosen:

Boosted Trees:

– minimum number of elements in a node to stop splitting: 79;

– minimum size of child node to stop splitting: 1;

– maximum tree depth: 32 levels;

– maximum number of nodes in a tree: 15.

These parameters help control the tree structure and prevent overfitting.

Neural Network (MLP):

– minimum number of neurons in the hidden layer: 3;

– the maximum number of neurons in the hidden layer: 64;

– Number of trained neural networks: 500.

These settings help to experiment with different network architectures and save the best result.

SVM: kernel type: RBF (radial basis function).

Using the RBF kernel allows working with nonlinear data and improves the ability of SVM to separate complex data.

These initial parameters are selected considering the specific characteristics of the task and will allow our models to analyze the data effectively."

Next, dependency models for the parameters S_1, S_2, T_{max} и TOC were built using the selected machine learning methods, as well as multiple linear regression. Table 3 presents the correlation coefficient values for pairs of actual and calculated values.

Table 3

Correlation coefficients of ML models

Method	S_1	S_2	S_1+S_2	T_{max}	TOC
Boosted trees	0.819315	0.815852	0.823565	0.802799	0.825891
Neural Network	0.735394	0.774090	0.785101	0.542445	0.752758
SVM	0.681148	0.699781	0.719783	0.528209	0,666530
Multiple regression	0.730231	0.769166	0.781709	0.490886	0.742394

Table 4

Comparison of predictive models' correlation, using linear regression and Boosted Trees methods

Linear Regression / Machine learning	ML_S_1	ML_S_2	ML_S_1+S_2	ML_T_{max}	ML_TOC
RegrssS_1	**0.926029**	0.919872	0.928922	0.562976	0.863218
RegrssS_2	0.902885	**0.955211**	0.957815	0.524011	0.905242
RegrssS_1+S_2	0.905664	0.955484	**0.958337**	0.528347	0.905805
RegrssT_{max}	0.826240	0.776076	0.788178	**0.634128**	0.733300
RegrssTOC	0.881811	0.950323	0.950109	0.516769	**0.914215**

The analysis of all three models for predicting the variables allows us to draw the following general conclusions:

1. Boosted Trees Model:

• for predicting target variables, this model demonstrated excellent results with low mean square error and high correlation on the training sample;

• the importance of independent variables is diferrent for S_2 and T_{max}, but in both cases the model showed their statistical significance;

• The Boosted Trees model is the best of those considered for both target variables.

2. Neural network (MLP):

• The neural network demonstrated good correlation with the target variables, although the mean square error and correlation were lower, compared to the Boosted Trees model;

• after additional hyperparameter tuning an increase in the model performance was achieved.

3. SVM (Support Vector Machine) model:

• The SVM model showed low correlation with target variables, mean square error was also higher than the Boosted Trees model and neural network;

• This model has the lowest correlation coefficient and the highest standard error among all the models considered.

For all target variables, the Boosted Trees model is the best because it has the optimal combination of low error and high correlation. The neural network can be used if an alternative model is required, but it will require additional tuning. The SVM model may be less preferable due to higher mean square error and low correlation. To make a final choice of model, it is also necessary to evaluate their performance on a test sample.

The results' analysis showed that among the regression models obtained by machine learning methods, the Boosted Trees model stands out with the highest correlation coefficient values. The lowest correlation coefficient was achieved by the Support Vector Machine (SVM), which is a natural result, since this method operates best in classification tasks, not regression. The multiple linear regression and neural network methods showed approximately the same average result, which is also a natural result, given the general similarity in the methodology for developing dependency models.

Regarding the multiple linear regression method, the results of the Boosted Trees model show an increased correlation coefficient. Consequently, further comparative analysis was carried out between the results of these models.

Correlation analysis of the models developed using the Boosted Trees and multiple linear regression methods, showed a fairly high linear relationship between the parameters. The correlation matrix is given in Table 4, cross-plots of dependencies for the calculated values of the S_1, S_2, T_{max} and TOC parameters are shown in Fig. 1.

The analysis of the correlation matrix and dependence graphs of forecast results obtained through linear regression and the Boosted Trees method showed that the relationship between the parameters is linear with a high correlation coefficient. The most significant difference in results was observed when predicting the parameter T_{max}, where the Boosted Trees method demonstrated greater forecast accuracy compared to linear regression. Additionally, the graph showing the relationship between parameters calculated by different methods does not exhibit a clear linear connection.

Next, a visual comparison of the forecast results for geochemical parameters obtained by different methods with experimental data was conducted (Fig. 2). The Boosted Trees method was chosen as the method with the highest correlation and the linear regression method as the most common for solving similar problems.

The visual analysis of the comparison between synthetic parameter curves and laboratory research

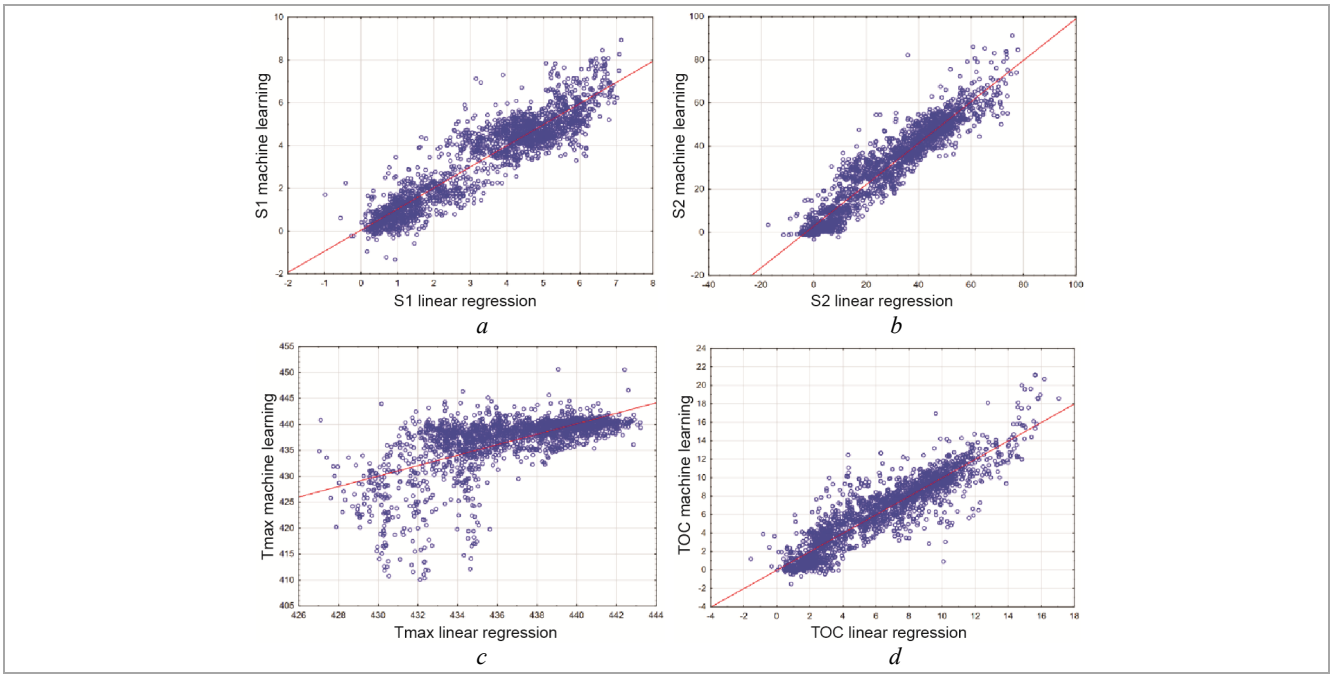

Fig. 1. Comparison Graphs of estimated values using linear regression methods and Boosted Trees for parameters: *a* – S_1; *b* – S_2; *c* – T_{max}; *d* – TOC

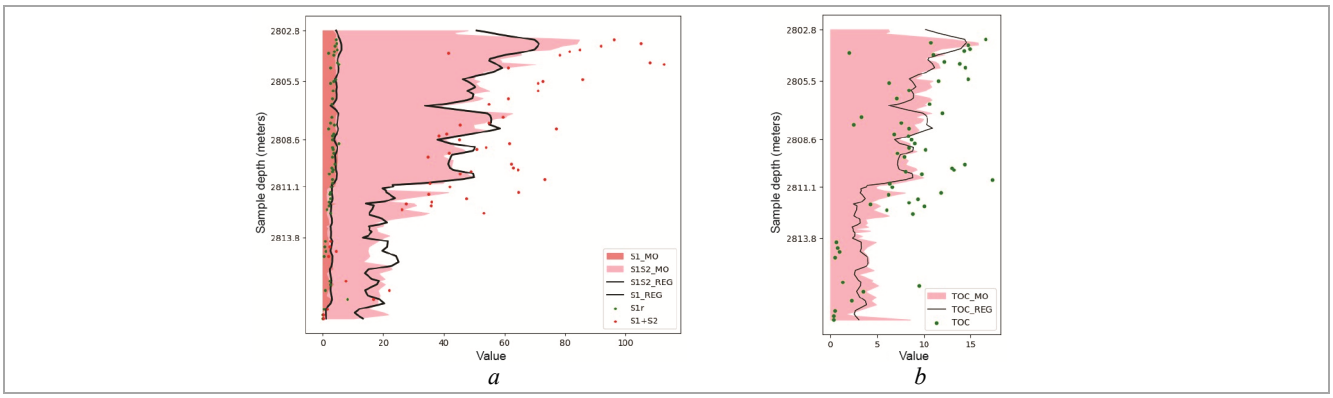

Fig. 2. Comparison of experimental data from geochemical studies with a synthetic curve of parameters: *a* – S_1 and S_2; *b* – TOC

Table 5

Quantitative metrics for assessing model reliability

Parameter	Metric/Method	MAE (less – better)	MAPE (less – more)	MSE (less – better)	R_2(bigger – better)
S_1	SVM	1.601	56.44%	2.564	0.464
	Neural network	1.342	51.98%	1.801	0.541
	Boosted trees	1.125	50.20%	1.265	0.671
	Linear regression	1.362	58.07%	1.854	0.533
S_2	SVM	21.561	54.93%	464.860	0.490
	Neural network	11.423	102.13%	130.480	0.599
	Boosted trees	10.179	38.68%	103.612	0.666
	Linear regression	11.383	61.30%	129.571	0.591
T_{max}	SVM	6.342	1.46%	40.220	0.279
	Neural network	3.967	0.91%	15.735	0.294
	Boosted trees	2.872	0.66%	8.249	0.644
	Linear regression	4.587	0.61%	21.045	0.239
TOC	SVM	3.690	50.76%	13.620	0.444
	Neural network	2.318	44.42%	5.375	0.567
	Boosted trees	1.967	54.24%	3.870	0.682
	Linear regression	2.206	50.76%	4.868	0.551

N o t e : * – cells that characterize the model as more accurate are marked in green, and cells that characterize the model as less accurate are marked in red.

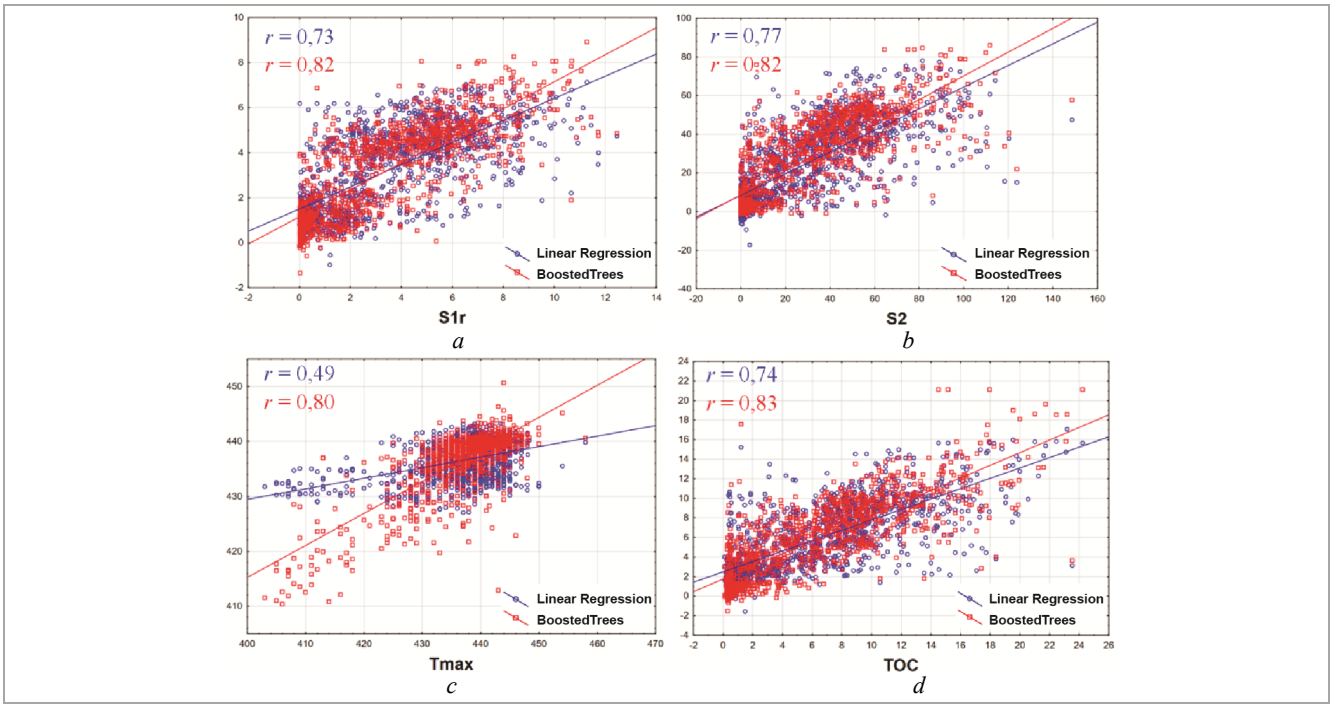

Fig. 3. Comparison of crossplots of estimated pairs and experimental values
for linear regression models and Boosted Trees parameters: *a* – S_1; *b* – S_2; *c* – T_{max}; *d* – TOC

results showed that both models adequately capture the trend in the changes of geochemical properties. However, the visual differences between the predicting models are insufficient to draw conclusions about the superiority of one over the other. Due to the lack of visual differentiation between the models, standard metrics were calculated and compared. The results of the metric comparison are presented in Table 5.

Considering the presented metrics (MAE, MAPE, MSE, and R^2) for all four parameters (S_1, S_2, T_{max} and TOC), a general conclusion can be drawn that Boosted Trees is the best machine learning model among all the methods considered.

The Boosted Trees model is defined in the following ways:

1. Prediction Accuracy: Boosted Trees have the lowest MAE and MSE values, which means they predict values closer to the actual data and have smaller errors in absolute and squared values.

2. Percentage Errors: Boosted Trees method also shows the lowest MAPE values, which indicates the lowest percentage errors in predictions. This means that the predictions generated by Boosted Trees are the least distorted in relative terms.

3. Ability to explain variability: Boosted Trees show the highest R^2 values for all parameters, which indicates their superior ability to explain variability in the data.

Consequently, Boosted Trees method is the best choice for this task, providing the most accurate and reliable predictions with the lowest errors in both absolute values and percentage terms, as well as having the highest ability to explain variability in the data.

A comparative visual analysis of the correlation fields, presented in Fig. 3, also showed an increase in the prediction accuracy of the model developed using the Boosted Trees method, which also confirms the conclusions obtained from the comparison of numerical metrics for assessing the prediction models.

Conclusion

The study conducted an analysis of various regression methods, among which the Boosted Trees model stands out as the leader. This model demonstrates the best results among all the tested methods, with the lowest errors (MAE and MSE) and the highest coefficient of determination (R^2). Boosted Trees provide accurate and stable forecasting results, supported by visual analysis. The research shows that the Boosted Trees model is more effective and accurate than the linear regression method for establishing the "core – well-logging" relationship for geochemical data.

By using alternative regression methods based on machine learning algorithms, geochemical parameters were calculated, including S_1, S_2 and T_{max}, which could not be performed earlier. The use of machine learning methods has improved the accuracy of assesing the geochemical parameters across well section, which will lead to increased accuracy of models for the distribution of rock properties over an area.

Machine learning methods in geology have great potential for applications in regression tasks, discrimination, and making various predictions based on the identified patterns. However, this requires careful data preprocessing, selection of the best models and training methods, and careful verification of the results. Despite these technical challenges, the use of machine learning in geology opens up new prospects for scientists and researchers, allowing for more accurate data analysis and interpretation, as well as for production decisions based on this data.

The geochemical properties assessment of rocks in this formation significantly contributes to forecasting hydrocarbon potential, as well as assessing the prospects of underexplored areas and deep horizons.

The results of the studies will be used for further reservoir identification and their classification into two types according to methodological recommendations for calculating oil reserves in the deposits of the Bazhenov horizon in the Western Siberian oil and gas province.

References

1. Shadrin AO, Krivoshchekov SN Studying the Structural and Thickness Characteristics of the Sedimentary Mantle of the Northern Part of the Surgut Arch// IOP Conference Series: Earth and Environmental Science. – IOP Publishing, 2021. – Vol. 666, no. 5. – P. 052056. DOI: 10.1088/1755-1315/666/5/052056
2. Shadrin A.O., Krivoshchekov S.N. Development of probabilistic-statistical models for forecasting oil content based on structural parameters of the YUS1 formation in the northern part of the Surgut dome // Geology of oil and gas. - 2022. - No. 2. - P. 53-65. DOI: 10.31087/0016-7894-2022-2-53-65. - EDN TUHENY.
3. Shadrin A.O., Akhmetova L.V. Analysis of the influence of structural parameters of the sedimentary cover on the oil and gas potential of the YUS1 formation in the northern part of the Surgut arch // Oil and Gas - 2021: proceedings of the 75th International Youth Scientific Conference, Moscow, April 26-30, 2021. - M .: Gubkin Russian State University of Oil and Gas (National Research University), 2021. - P. 349-359. - EDN WDVQPP.
4. Shadrin A.O. Forecast of oil and gas potential of the YUS1 formation in the northern part of the Surgut arch based on trend analysis // Problems of development of hydrocarbon and ore mineral deposits. - 2020. - Vol. 1. - P. 154-160.
5. Shadrin A.O. Development of a zonal model for forecasting the oil and gas potential of the YUS1 formation based on geochemical parameters // Subsoil Use. - 2023. - Vol. 23, No. 1. - Pp. 2–10. DOI: 10.15593/2712-8008/2023.1.1. - EDN LLSEXO.
6. On the age of deposits of anomalous sections of the boundary layers of the Jurassic and Cretaceous along wells of the Severo-Konitlorskoye field (preliminary report) / Yu.V. Braduchan, N.K. Glushko, V.K. Komissarenko [et al.]. // Bulletin of the subsoil user of the Khanty-Mansiysk Autonomous Okrug. - 2005. - No. 16. - P. 20-24.
7. Experience of petrographic analysis of the mechanism of formation of the anomalous section of the Bazhenov suite in the Imilorskaya area of Western Siberia/ V.F. Grishkevich, N.N. Gatina, A.O. Sidorenko, E.V. Karpova // Lithosphere. – 2019. – No. 2 (19). – pp. 209–227. DOI: 10.24930/1681-9004-2019-19-2-209-227
8. Shaikhutdinova G.Kh. Petrographic study of oil migration using the example of the Imilorskoye field (Kogalym oil and gas region, Western Siberia) // Lithosphere. - 2020. - No. 4 (20). - P. 592-600. DOI: 10.24930/1681-9004-2020-20-4-592-600
9. Skachek K.G., Osyka A.V., Garifullin I.I. Prospects for oil-bearing capacity of the Bazhenov formation of the Kogalym region // Ways to realize the oil and gas potential of the Khanty-Mansiysk Autonomous Okrug: the seventh scientific and practical conference / edited by Karasev V.I., Akhpatelov E.A., Volkov V.A. - Khanty-Mansiysk, 2004. - V. 1. - P. 162-170.
10. Nemova V.D. Lithology and reservoir properties of the Bazhenov horizon deposits in the west of the Latitudinal Ob region: diss. Cand. Geol.-Min. Sciences: 25.00.06; 25.00.12. – M., 2012.
11. Nemova V.D., Bedretdinov R.Yu., Kirsanov A. Interpretation of well logging data in the interval of the Bazhenov formation under conditions of a limited well logging complex in the territory of the Sredne-Nazymskoye field // Petrophysics of complex reservoirs: problems and prospects 2015. - M .: OOO "EAGE Geomodel", 2015. - P.174-159.
12. Nemova V.D., Gavrilov S.S. Studies of core samples of the Bazhenov horizon deposits as a basis for interpreting seismic exploration data // Petrophysics of complex reservoirs: problems and prospects 2014. - M .: OOO "EAGE Geomodel", 2014. - P. 212-230.
13. Grishkevich V.F. Anomalous sections of the Bazhenov formation: model of formation and problems of geometrization // Tyumen 2015 – Deep Subsoil and Science Horizons, Mar, 2015. – Vol. 2015. – P. 1–5. DOI: 10.3997/2214-4609.201412065
14. Grishkevich V.F. Bazhenov horizon of Western Siberia: search for new harmony: monograph – Tyumen: TIU, 2022. – 279 p.
15. Nesterov I.I. New type of oil and gas reservoirs // Geology of oil and gas. – 1979. – No. 10. – P. 26–29.
16. On the possibility of predicting the oil and gas potential of Famennian deposits using the construction of probabilistic-statistical models / V.I. Galkin, I.A. Kozlova, S.N. Krivoshchekov [et al.] // Geology, geophysics and development of oil and gas fields. – 2007. – No. 10. – P. 22–27. – EDN IBLHOX.
17. Solution of regional problems of oil-bearing forecasting based on geological and geochemical analysis of dispersed organic matter of Domanik-type rocks / V.I. Galkin, I.A. Kozlova, M.A. Nosov, S.N. Krivoshchekov // Oil industry. – 2015. – No. 1. – P. 21–23. – EDN TIJTZT.
18. On the scale of hydrocarbon migration within the Solikamsk depression of the Pre-Ural trough and the possibilities of its use for forecasting oil and gas potential / V.L. Voevodkin, V.I. Galkin, I.A. Kozlova [et al.] // Geology, geophysics and development of oil and gas fields. – 2010. – No. 12. – P. 6–11. – EDN NCQIEL.
19. Krivoshchekov S.N. Development of regional-zonal criteria for forecasting oil and gas potential of the Perm Kama region using probabilistic-statistical methods // Oil Industry. – 2011. – No. 10. – P. 10–14. – EDN OHKPOJ.
20. The Petrophysics of Shale Gas Reservoirs: Technical Challenges and Pragmatic Solutions / VK Bust, AA Majid, JU Oletu, PF Worthington // International Petroleum Technology Conference, Bangkok, Thailand, November. – 2011. DOI: 10.2523/IPTC-14631-MS
21. Thaimar R., Ramirez JD, Klein RJ Comparative Study of Formation Evaluation Methods for Unconventional Shale-Gas Reservoirs: Application to the Haynesville Shale (Texas) // North American Unconventional Gas Conference and Exhibition, The Woodlands, Texas, USA, June 2011. – 2011. DOI: 10.2118/144062-MS
22. Evaluation of shale oil reserves using geochemical parameters / M.V. Dakhnova, S.V. Mozhegova, E.S. Nazarova, I.L. Payzanskaya // Geology of oil and gas. - 2015. - No. 4. - P. 55-61. - EDN UEKTBP.
23. Jarvie DM, Shale resource systems for oil and gas: Part 2: Shale-oil resource systems / ed.: JABreger // Shale reservoirs - Giant resources for the 21st century: AAPG Memoir 97. DOI: 10.1306/13321447M973489
24. Kirsanov A.M., Skvortsov M.B. Determination of calculation parameters based on well logging data for assessing the resources of the Bazhenov Formation // Oil Capital: Proceedings of the 4th International Youth Scientific and Practical Forum, Khanty-Mansiysk, March 24–25, 2021. – Khanty-Mansiysk: Center for Scientific and Technical Solutions, 2021. – P. 100–103. – EDN CBBBCW.
25. Automatic Determination Of Lithology From Well Logs / Pierre Delfiner [et al.] // SPE Formation Evaluation, Sep. – 1987. – Vol. 2, iss. 03. – P. 303–310. DOI: 10.2118/13290-PA
26. Toumani A., Schmitz D., Schepers R. Automatic determination of lithology from well logs using fuzzy classification // 56th EAEG Meeting. – EAGE Publications BV, 1994. – P. 47-00144. DOI: 10.3997/2214-4609.201409899
27. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: A case study from Sichuan Basin, China / D. Zheng [et al.] // Journal of Petroleum Science and Engineering. – 2022. – Vol. 215. – P. 110610. DOI: 10.1016/j.petrol.2022.110610
28. Probabilistic logging lithology characterization with random forest probability estimation / Y. Ao [et al.] // Computers & Geosciences. – 2020. – Vol. 144. – P. 104556. DOI: 10.1016/j.cageo.2020.104556
29. Alfred D., Vernik L. A new petrophysical model for organic shells // Petrophysics. – 2013. – Vol. 54, No. 03. – P. 240–247.
30. A new approach to calculate gas saturation in shale reservoirs / S. Chen [et al.] // Energy & Fuels. – 2022. – Vol. 36, No. 4. – P. 1904–1915. DOI: 10.1021/acs.energyfuels.1c04067
31. The key parameter of shale oil resource evaluation: Oil content / M. Wang [et al.] // Petroleum Science. – 2022. – Vol. 19, No. 4. – P. 1443–1459. DOI: 10.1016/j.petsci.2022.03.006
32. Schmoker JW Use of formation-density logs to determine organic carbon content in Devonian shale of the western Appalacian basin // Bull, of US Geol. Surv. – 1993. – 1909. – P. 71–74.
33. Belokon T.V., Naborshchikova I.I., Ersulova I.S. Determination of DOM in rocks by natural gamma activity // Collection. Evolution of oil and gas formation in the history of the Earth: abstracts of reports of the IV All-Union seminar. - M.: Moscow State University, 1984. - P. 236.
34. Meyer BL, Nederlof MH Identification of sours-rocks on wireline logs by density / resistivity and sonic transit time / resistivity crossplots // Bulletin of the American Association of Petroleum Geologists. – 1984. – Vol. 68, no. 2. – P. 121–129. DOI: 10.1306/AD4609E0-16F7-11D7-8645000102C1865D
35. US Patent No. 14/977.336 (22 Jun. 2022) / L. Mosse, E. Rylander, P. Craddock.
36. Rashka S., Mirjalili V. Python and Machine Learning: Machine and Deep Learning Using Python, scikit-learn, and TensorFlow-2. – St. Petersburg: Dialectika, 2020. – Vol. 848.

37. Breiman L. Random forests // Machine learning. – 2001. – Vol. 45. – P. 5–32. DOI: 10.1023/A:1010933404324
38. Friedman JH Greedy function approximation: a gradient boosting machine // Annals of statistics. – 2001. – P. 1189–1232. DOI: 10.1214/aos/1013203451
39. Doveton JH Compositional analysis of mineralogy // Principles of Mathematical Petrophysics. – Oxford University Press, 2014. – P. 94–117. DOI: 10.1093/oso/9780199978045.003.0009
40. Merembayev T., Yunussov R., Yedilkhan A. Machine learning algorithms for classification geology data from well logging // 2018 14th International Conference on Electronics Computer and Computation (ICECCO). – IEEE, 2018. – P. 206–212. DOI: 10.1109/ICECCO.2018.8634775
41. Formation lithology classification: Insights into machine learning methods / IM Mohamed [et al.] // SPE Annual Technical Conference and Exhibition. – OnePetro, 2019. DOI: 10.2118/196096-MS
42. Deriving permeability and reservoir rock typing supported with self-organized maps som and artificial neural networks ann-optimal workflow for enabling core-log integration / L. Saputelli [et al.] // SPE Reservoir Characterization and Simulation Conference and Exhibition? – SPE, 2019. – P.D021S015R001. DOI: 10.2118/196704-MS
43. Oyewande A., Stacey L., Jon R. Synthetic well log generation using machine learning techniques // Unconventional Resources Technology Conference(URTeC). – 2018. DOI: 10.15530/URTEC-2018-2877021
44. Machine learning for the geosciences: Challenges and opportunities / A. Karpatne [et al.] // IEEE Transactions on Knowledge and Data Engineering. – 2018. – Vol. 31, No. 8. – P. 1544–1554. DOI: 10.1109/TKDE.2018.2861006
45. Shier DE Well log normalization: Methods and guidelines // Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description. – 2004. – Vol. 45, No. 03. – P. 268–280.

Библиографический список

1. Shadrin A.O., Krivoshchekov S.N. Studying the Structural and Thickness Characteristics of the Sedimentary Mantle of the Northern Part of the Surgut Arch // IOP Conference Series: Earth and Environmental Science. – IOP Publishing, 2021. – Vol. 666, №. 5. – P. 052056. DOI: 10.1088/1755-1315/666/5/052056
2. Шадрин А.О., Кривощеков С.Н. Разработка вероятностно-статистических моделей прогноза нефтеносности по структурным параметрам пласта ЮС1 в северной части Сургутского свода // Геология нефти и газа. – 2022. – № 2. – С. 53–65. DOI: 10.31087/0016-7894-2022-2-53-65. – EDN TUHENY.
3. Шадрин А.О., Ахметова Л.В. Анализ влияния структурных параметров осадочного чехла на нефтегазоносность пласта ЮС1 северной части Сургутского свода // Нефть и газ – 2021: сборник трудов 75-й Международной молодежной научной конференции, Москва, 26–30 апреля 2021 года. – М.: Российский государственный университет нефти и газа (национальный исследовательский университет) имени И.М. Губкина, 2021. – С. 349–359. – EDN WDVQPP.
4. Шадрин А.О. Прогноз нефтегазоносности пласта ЮС1 северной части Сургутского свода на основе проведения тренд-анализа // Проблемы разработки месторождений углеводородных и рудных полезных ископаемых. – 2020. – Т. 1. – С. 154–160.
5. Шадрин А.О. Разработка зональной модели прогноза нефтегазоносности пласта ЮС1 по геохимическим параметрам // Недропользование. – 2023. – Т. 23, № 1. – С. 2–10. DOI: 10.15593/2712-8008/2023.1.1. – EDN LLSEXO.
6. О возрасте отложений аномальных разрезов пограничных слоев юры и мела по скважинам Северо-Конитлорского месторождения (предварительное сообщение) / Ю.В. Брадучан, Н.К. Глушко, В.К. Комиссаренко [и др.]. // Вестник недропользователя ХМАО. – 2005. – № 16. – С. 20–24.
7. Опыт петрографического анализа механизма формирования аномального разреза баженовской свиты на Имилорской площади Западной Сибири / В.Ф. Гришкевич, Н.Н. Гатина, А.О. Сидоренко, Е.В. Карпова // Литосфера. – 2019. – № 2 (19). – С. 209–227. DOI: 10.24930/1681-9004-2019-19-2-209-227
8. Шайхутдинова Г.Х. Петрографическое изучение миграции нефти на примере Имилорского месторождения (Когалымский нефтегазоносный район, Западная Сибирь) // Литосфера. – 2020. – № 4 (20). – С. 592–600. DOI: 10.24930/1681-9004-2020-20-4-592-600
9. Скачек К.Г., Осыка А.В., Гарифуллин И.И. Перспективы нефтеносности баженовской свиты Когалымского региона // Пути реализации нефтегазового потенциала ХМАО: седьмая научно-практическая конференция / под ред. Карасева В.И., Ахпателова Э.А., Волкова В.А. – Ханты-Мансийск, 2004. – Т. 1. – С. 162–170.
10. Немова В.Д. Литология и коллекторские свойства отложений баженовского горизонта на западе Широтного Приобья: дис. канд.геол.-мин. наук: 25.00.06; 25.00.12. – М., 2012.
11. Немова В.Д., Бедретдинов Р.Ю., Кирсанов А. Интерпретация данных ГИС В интервале баженовской свите в условиях ограниченного комплекса ГИС на территории Средне-Назымского месторождения // Петрофизика сложных коллекторов: проблемы и перспективы 2015. – М.: ООО «ЕАГЕ Геомодель», 2015. – С.174–159.
12. Немова В.Д., Гаврилов С.С. Исследования керна отложений баженовского горизонта, как основа для интерпретации данных сейсморазведки // Петрофизика сложных коллекторов: проблемы и перспективы 2014. – М.: ООО «ЕАГЕ Геомодель», 2014. – С. 212–230.
13. Гришкевич В.Ф. Аномальные разрезы баженовской свиты: модель образования и проблемы геометризации // Tyumen 2015 – Deep Subsoil and Science Horizons, Mar, 2015. – Vol. 2015. – P. 1–5. DOI: 10.3997/2214-4609.201412065
14. Гришкевич В.Ф. Баженовский горизонт Западной Сибири: поиски новой гармонии: монография – Тюмень: ТИУ, 2022. – 279 с.
15. Нестеров И.И. Новый тип коллекторов нефти и газа // Геология нефти и газа. – 1979. – № 10. – С. 26–29.
16. О возможности прогнозирования нефтегазоносности фаменских отложений с помощью построения вероятностно-статистических моделей / В.И. Галкин, И.А. Козлова, С.Н. Кривощеков [и др.] // Геология, геофизика и разработка нефтяных и газовых месторождений. – 2007. – № 10. – С. 22–27. – EDN IBLHOX.
17. Решение региональных задач прогнозирования нефтеносности по данным геолого-геохимического анализа рассеянного органического вещества пород доманикового типа / В.И. Галкин, И.А. Козлова, М.А. Носов, С.Н. Кривощеков // Нефтяное хозяйство. – 2015. – № 1. – С. 21–23. – EDN TIJTZT.
18. О масштабах миграции углеводородов в пределах Соликамской депрессии Предуральского прогиба и возможностях ее использования для прогноза нефтегазоносности / В.Л. Воеводкин, В.И. Галкин, И.А. Козлова [и др.] // Геология, геофизика и разработка нефтяных и газовых месторождений. – 2010. – № 12. – С. 6–11. – EDN NCQIEL.
19. Кривощеков С.Н. Разработка регионально-зональных критериев прогноза нефтегазоносности территории Пермского Прикамья вероятностно-статистическими методами // Нефтяное хозяйство. – 2011. – № 10. – С. 10–14. – EDN OHKPOJ.
20. The Petrophysics of Shale Gas Reservoirs: Technical Challenges and Pragmatic Solutions / V.K. Bust, A.A. Majid, J.U. Oletu, P.F. Worthington // International Petroleum Technology Conference, Bangkok, Thailand, November. – 2011. DOI: 10.2523/IPTC-14631-MS
21. Thaimar R., Ramirez J.D., Klein R.J. Comparative Study of Formation Evaluation Methods for Unconventional Shale-Gas Reservoirs: Application to the Haynesville Shale (Texas) // North American Unconventional Gas Conference and Exhibition, The Woodlands, Texas, USA, June 2011. – 2011. DOI: 10.2118/144062-MS
22. Оценка запасов «сланцевой нефти» с использованием геохимических параметров / М.В. Дахнова, С.В. Можегова, Е.С. Назарова, И.Л. Пайзанская // Геология нефти и газа. – 2015. – № 4. – С. 55–61. – EDN UEKTBP.
23. Jarvie D.M., Shale resource systems for oil and gas: Part 2: Shale-oil resource systems / ed.: J.A.Breger // Shale reservoirs - Giant resources for the 21st century: AAPG Memoir 97. DOI: 10.1306/13321447M973489
24. Кирсанов А.М., Скворцов М.Б. Определение подсчетных параметров по данным ГИС для оценки ресурсов баженовской свиты // Нефтяная столица: материалы 4-й Международного молодежного научно-практического форума, Ханты-Мансийск, 24–25 марта 2021 года. – Ханты-Мансийск: Центр научно-технических решений, 2021. – С. 100–103. – EDN CBBBCW.
25. Automatic Determination Of Lithology From Well Logs / Pierre Delfiner [et al.] // SPE Formation Evaluation, Sep. – 1987. – Vol. 2, iss. 03. – P. 303–310. DOI: 10.2118/13290-PA
26. Toumani A., Schmitz D., Schepers R. Automatic determination of lithology from well logs using fuzzy classification // 56th EAEG Meeting. – EAGE Publications BV, 1994. – P. 47-00144. DOI: 10.3997/2214-4609.201409899
27. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: A case study from Sichuan Basin, China / D. Zheng [et al.] // Journal of Petroleum Science and Engineering. – 2022. – Vol. 215. – P. 110610. DOI: 10.1016/j.petrol.2022.110610

28. Probabilistic logging lithology characterization with random forest probability estimation / Y. Ao [et al.] // Computers & Geosciences. – 2020. – Vol. 144. – P. 104556. DOI: 10.1016/j.cageo.2020.104556
29. Alfred D., Vernik L. A new petrophysical model for organic shales // Petrophysics. – 2013. – Vol. 54, № 03. – P. 240–247.
30. A new approach to calculate gas saturation in shale reservoirs / S. Chen [et al.] // Energy & Fuels. – 2022. – Vol. 36, № 4. – P. 1904–1915. DOI: 10.1021/acs.energyfuels.1c04067
31. The key parameter of shale oil resource evaluation: Oil content / M. Wang [et al.] // Petroleum Science. – 2022. – Vol. 19, № 4. – P. 1443–1459. DOI: 10.1016/j.petsci.2022.03.006
32. Schmoker J.W. Use of formation-density logs to determine organic carbon content in devonian shales of the western Appalacian basin // Bull, of US Geol. Surv. – 1993. – 1909. – P. 71–74.
33. Белоконь Т.В., Наборщикова И.И., Ерсулова И.С. Определение РОВ в породах по естественной гамма-активности // Сб. Эволюция нефтегазообразования в истории Земли: тезисы докладов IV Всесоюзного семинара. – М.: МГУ, 1984. – С. 236.
34. Meyer B.L., Nederlof M.H. Identification of sours-rocks on wireline logs by density / resistivity and sonic transit time / resistivity crossplots // Bulletin of the American Association of Petroleum Geologists. – 1984. – Vol. 68, no. 2. – P. 121–129. DOI: 10.1306/AD4609E0-16F7-11D7-8645000102C1865D
35. US Patent No 14/977.336 (22 Jun. 2022) / L. Mosse, E. Rylander, P. Craddock.
36. Рашка С., Мирджалили В. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow-2. – СПб.: Диалектика, 2020. – Т. 848.
37. Breiman L. Random forests // Machine learning. – 2001. – Vol. 45. – P. 5–32. DOI: 10.1023/A:1010933404324
38. Friedman J.H. Greedy function approximation: a gradient boosting machine // Annals of statistics. – 2001. – P. 1189–1232. DOI: 10.1214/aos/1013203451
39. Doveton J.H. Compositional analysis of mineralogy // Principles of Mathematical Petrophysics. – Oxford University Press, 2014. – P. 94–117. DOI: 10.1093/oso/9780199978045.003.0009
40. Merembayev T., Yunussov R., Yedilkhan A. Machine learning algorithms for classification geology data from well logging // 2018 14th International Conference on Electronics Computer and Computation (ICECCO). – IEEE, 2018. – P. 206–212. DOI: 10.1109/ICECCO.2018.8634775
41. Formation lithology classification: Insights into machine learning methods / I.M. Mohamed [et al.] // SPE Annual Technical Conference and Exhibition. – OnePetro, 2019. DOI: 10.2118/196096-MS
42. Deriving permeability and reservoir rock typing supported with self-organized maps som and artificial neural networks ann-optimal workflow for enabling core-log integration / L. Saputelli [et al.] // SPE Reservoir Characterisation and Simulation Conference and Exhibition? – SPE, 2019. – P. D021S015R001. DOI: 10.2118/196704-MS
43. Oyewande A., Stacey L., Jon R. Synthetic well log generation using machine learning techniques // Unconventional Resources Technology Conference (URTeC). – 2018. DOI: 10.15530/URTEC-2018-2877021
44. Machine learning for the geosciences: Challenges and opportunities / A. Karpatne [et al.] // IEEE Transactions on Knowledge and Data Engineering. – 2018. – Vol. 31, № 8. – P. 1544–1554. DOI: 10.1109/TKDE.2018.2861006
45. Shier D.E. Well log normalization: Methods and guidelines // Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description. – 2004. – Vol. 45, № 03. – P. 268–280.

Funding. The study had no sponsorship support.
Conflict of interest. The authors declare no conflict of interest.
The authors' contribution is equal.