

УДК 622.276

Статья / Article

© ПНИПУ / PNRPU, 2023

Оценка геохимических свойств пород Баженовской свиты по данным ГИС методами машинного обучения**А.О. Шадрин¹, А.Б. Гулин¹, С.Г. Ашихмин², А.А. Мелехин²**¹Филиал ООО «ЛУКОЙЛ-Инжиниринг» «КогалымНИПИнефть» в г. Тюмени (Россия, 628481, г. Когалым, ул. Центральная, 19)²Пермский национальный исследовательский политехнический университет (Россия, 614990, г. Пермь, Комсомольский проспект, 29)**Assessment of Rock Geochemical Properties of the Bazhenov Formation According to Well Logging Data Using Machine Learning Methods****Arsenii O. Shadrin¹, Aleksandr B. Gulin¹, Sergey G. Ashikhmin², Aleksandr A. Melekhin²**¹KogalymNIPIneft branch of LUKOIL-Engineering LLC in Tyumen (19 Tsentral'naya st., Kogalym, 628481, Russian Federation)²Perm National Research Polytechnic University (29 Komsomolskiy av., Perm, 614990, Russian Federation)

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

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.

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 training 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.

© 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).

© 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).

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

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

Введение

Данная работа выполнена в рамках комплексного исследования юрских отложений на территории северной части Сургутского свода и представляет собой часть исследования, направленную на изучение и оценку геохимических и нефтематеринских свойств пород Баженовской свиты в контексте ее влияния на нефтеносность пластов Васюганской свиты [1–5].

В настоящем исследовании поставлена задача разработки моделей оценки зависимостей между геохимическими параметрами пород Баженовской свиты и данными геофизических исследований, полученными в результате каротажа. Для этого в данной работе применялись различные методы МО [6–15].

В настоящее время современные методы машинного обучения представляют мощные инструменты для анализа данных и прогнозирования, и их применение также актуально в контексте решения геологических задач. В этой области стандартной практикой является работа с большими объемами данных, сложными взаимосвязями между геологическими параметрами пород и неоднородностью геологических формаций. В этом контексте использование методов машинного обучения является новым инструментом, который значительно улучшает процессы анализа и создания моделей, основанных на геоданных. Основное преимущество применения машинного обучения в геологических исследованиях заключается в его способности автоматизировать и оптимизировать анализ данных. Геологические данные часто характеризуются множеством параметров, и их взаимосвязи могут быть сложными и труднопонимаемыми для человека. Модели машинного обучения, благодаря более сложному алгоритму поиска корреляционных связей и дискриминантных признаков, способны повысить надежность прогнозов и моделей оценки данных. Эти модели могут использоваться для прогнозирования различных параметров, таких как качество и состав пород, распределение полезных ископаемых, гидрогеологические характеристики и многое другое.

В данном исследовании в качестве основных параметров для установления зависимостей являлись геохимические свойства пород. Эти данные были получены с использованием метода пиролиза Rock-Eval, который позволяет определить содержание и состав органических веществ в породах. Высокая стоимость отбора керн не позволяет проводить отбор образцов в большом количестве скважин, что повышает ценность отобранного кернового материала. Напротив, относительно низкая стоимость проведения геофизических исследований и их высокая информативность способствовали тому, что большая доля фонда скважин охарактеризована относительно единым комплексом исследований, что позволяет сравнивать геофизические свойства пород разных скважин друг с другом. Установление зависимостей между геохимическими и геофизическими свойствами позволит выполнить качественную и количественную оценку свойств

изучаемых пород, а также расширить количество скважин, в которых возможно оценить геохимические свойства пород.

Объект исследования. Методы исследования

Баженовская свита по праву считается главной нефтегенерирующей формацией Западно-Сибирской нефтегазоносной провинции. Отложения сформированы преимущественно в глубоководно-морских условиях осадконакопления и распространены практически повсеместно на территории Западной Сибири. В рамках данной работы исследовались основные характеристики пород, полученные методом пиролиза Rock-Eval. Метод пиролиза основан на термическом разложении образцов керн с записью хроматограммы, анализ которой позволяет получить такие данные, как содержание свободных углеводородов (S_1), остаточный генерационный потенциал (S_2), общий органический углерод (ТОС) и температура максимального выхода углеводородов в процессе крекинга (T_{max}). Полученные характеристики позволяют сделать выводы о нефтегенерационной продуктивности породы, зрелости и количестве рассеянного органического вещества, а также оценить степень реализации нефтематеринского потенциала породы [16–19].

Проблемы интерпретации данных ГИС в нетрадиционных коллекторах и в том числе проблемы оценки геохимических свойств по данным геофизических исследований освещены в большом количестве работ, в частности в работах [19–35].

Поиск зависимостей осуществлялся с параметрами, полученными при проведении геофизических исследований скважин. В качестве переменных предикторов использовался стандартный комплекс ГИС, выполненный в большинстве скважин, – GK (гамма-каротаж), NKT (каротаж по тепловым нейтронам), BK (каротаж сопротивлений), IK (индукционный каротаж), а также производные от этих параметров ($\ln BK$, $\ln IK$, GK/NKT , $\ln[GK/NKT]$).

Получение зависимостей «керн – ГИС» позволит более точно осреднить геохимические данные в скважинах с наличием геохимических исследований, оценить свойства пород в скважинах, не освещенных керновыми данными, а также построить модели распределения свойств по площади, основываясь на скважинных данных.

Для анализа использовалась база данных, состоящая из 1217 исследованных образцов, взятых в скважинах на территории Сургутского и Вартовского сводов, имеющих привязку к интервалам пород баженовской свиты или стратиграфически близких отложений.

В качестве методов поиска зависимостей в работе использовался метод множественной линейной регрессии, а также три метода машинного обучения: метод опорных векторов (SVM), нейронные сети (Neural Networks) и метод усиленных деревьев (Boosted trees). Подробная характеристика и отличительные особенности разных методов представлены в большом числе публикаций, в частности, в работах [36–44].

Таблица 1

Статистические характеристики параметров

Переменная	Среднее арифметическое	Минимум	Максимум	Станд. отклон	Асимметрия	Экссесс
S_1r , мгУВ/г.обр	3,2968	0,0000	12,45	2,6387	0,50229	-0,5573
S_2 , мгУВ/г.обр	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
ТОС, %	5,9896	0,0900	24,23	4,9310	0,67065	-0,2007
ВК, Омм	265,6223	0,7220	10600,70	630,8205	8,44372	104,7258
ИК, мСм	76,1296	0,1000	2018,85	126,3915	7,08556	77,8917
GK, мкрРГ/ч	23,0309	2,2610	111,80	16,6028	1,04645	0,5909
NKT, эВ	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, мкрРГ/чэВ	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

Особенностью поиска зависимостей «кern – ГИС» является разная вертикальная разрешающая способность методов. Пиролитические исследования керна выполняются на образцах, имеющих диаметр около 1 см, что повышает разрешающую способность метода, однако способствует появлению аномально низких или высоких значений, не являющимся характерными для всего интервала отложений. Каротаж ГИС имеет больший шаг дискретизации, где на 10 см разреза приходится всего одно измерение. Для нивелирования этой разницы данные геохимических исследований были сглажены методом скользящего окна в интервале ± 1 м, что позволило выделить относительно высокочастотный тренд изменения геохимических свойств и снизить количество аномальных значений.

Далее выполнялась нормировка геофизических данных, для чего применялся метод нормализации, описанный в работе Д.Э. Шира [45].

Согласно этому методу, процесс нормализации может быть выполнен с использованием следующего уравнения:

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

В вышеприведенном уравнении нормализованное значение кривой (V_{norm}) вычисляется из ненормализованного значения кривой (V_{log}), минимальных и максимальных значений из эталонной кривой (R_{min} и R_{max}), а также минимальных и максимальных значений из кривой, которую нужно нормализовать (W_{min} и W_{max}). Обычно вместо минимальных и максимальных значений используются 5-й и 95-й перцентили. Это уменьшает влияние выбросов данных (аномалий), которые могли бы привести к неверным результатам, однако неосторожное применение подобных методов может исключить часть важных данных из анализа, которые отражают действительные свойства пород.

После проведения процедур осреднения и нормализации данных была получена обучающая выборка, состоящая из 2109 значений, статистические характеристики параметров приведены в табл. 1.

Выбор оптимального соотношения между обучающей и тестовой выборками в задачах

машинного обучения представляет собой важный аспект методологии построения модели, требующий логического обоснования. Этот выбор находится в центре баланса между двумя ключевыми аспектами: максимизацией обобщающей способности моделей, обучаемых на данных, и обеспечением надежной оценки их производительности на независимых данных.

В данной исследовательской работе было решено применить соотношение 70 % обучающих данных и 30 % тестовых данных в контексте обучения моделей машинного обучения. В первую очередь, наличие 70 % обучающих данных обеспечивает моделям значительное количество информации для изучения основных закономерностей в данных. Одновременно с этим 30 % данных отводятся для целей тестирования и последующей оценки производительности моделей.

Данный подход способствует балансировке между процессом обучения и оценкой моделей, что помогает избежать явления переобучения, при котором модели адаптируются слишком тесно к обучающим данным, недостаточно обобщая полученные знания на новые данные.

С учетом принятого соотношения между обучающей и тестовой выборками было проведено экспериментальное обучение различных моделей машинного обучения и последующая оценка их производительности на тестовых данных. Полученные результаты свидетельствуют, что выбранное соотношение 70/30 оказалось оптимальным для данной задачи и обеспечило наилучшие практические результаты.

Оптимальное соотношение между обучающей и тестовой выборками может различаться в зависимости от особенностей данных и конкретной задачи. Тем не менее в данной работе выбор данного соотношения обоснован в контексте достижения высокой производительности моделей машинного обучения при оптимальной оценке их работы.

Сравнение и анализ разрабатываемых моделей проводились стандартными методами, выполнялся визуальный анализ кросс-плотов и графиков сопоставлений фактических и расчетных данных, оценка коэффициентов корреляции, а также количественная оценка метрик, используемых для анализа надежности регрессионных моделей машинного обучения, в частности: MAE, MAPE, MSE и R^2 .

Наилучшие предикторы для непрерывной зависимости

Параметр	S_1		S_2		ТОС		T_{max}	
	F-value	p-value	F-value	p-value	F-value	p-value	F-value	p-value
ИК	128,565	< 10 ⁻⁶	93,6135	< 10 ⁻⁶	70,0153	< 10 ⁻⁶	87,0745	< 10 ⁻⁶
БК	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

Метрика MAE (средняя абсолютная ошибка) – это способ измерения того, насколько сильно модель или прогноз отклоняются от реальных данных. Она измеряет среднее абсолютное значение разницы между прогнозами и фактическими значениями. Иными словами, MAE показывает в среднем, насколько ошибается модель, просто складывая все абсолютные разницы и делая усреднение, его формула имеет вид:

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

MAPE (Mean Absolute Percentage Error) – это метрика, которая используется для оценки точности прогнозов или моделей. Она измеряет процентную ошибку прогнозирования по сравнению с фактическими значениями данных, ее формула имеет вид:

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

MSE, или среднеквадратичная ошибка (Mean Squared Error) – это метрика, которая используется для измерения того, насколько близки предсказанные значения к истинным значениям в задачах регрессии. Иными словами, она измеряет среднее значение квадратов разницы между предсказанными и фактическими значениями, ее формула имеет вид:

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

Метрика R² (или коэффициент детерминации) – это статистическая мера, которая помогает оценить, насколько хорошо модель (например, линейная регрессия) соответствует данным. Она измеряет долю дисперсии (разброса) зависимой переменной, которую модель объясняет. Иначе говоря, R² показывает, насколько близки прогнозы модели к реальным данным, его формула имеет вид:

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

Результаты исследования

Для выполнения поставленной задачи, заключающейся в предсказании значений зависимых переменных (S_1 , S_2 , ТОС, T_{max}) на основе независимых переменных (БК, ИК, GK, NKT, lnBK, lnIK, GK/NKT, ln(GK/NKT)), был выбран программный продукт от компании StatSoft.

В ходе первой итерации после загрузки обучающей выборки был проведен отбор признаков с применением метода Feature selection. В рамках этой итерации для каждой независимой переменной и каждой зависимой переменной были рассчитаны критерии, такие как F-значение и p-значение, в контексте различных методов обучения. Эти критерии послужили основой для выявления наиболее значимых переменных, в то время как переменные с низкой статистической значимостью были исключены из дальнейшего анализа.

Результаты анализа приведены в табл. 2, в которой указаны F-значения и p-значения для каждой переменной. В процессе анализа таблицы были сделаны следующие выводы:

- значимые переменные: из анализа p-значений выделяются переменные ИК, БК, GK/NKT, GK, GK/NKT и NKT, у которых p-значения значительно ниже 0,05. Это свидетельствует о статистически значимой связи этих переменных с зависимыми переменными;
- незначимые переменные: переменные lnBK и lnIK имеют более высокие p-значения, указывая на отсутствие статистически значимой связи с зависимой переменной и их включение в модели не обосновано;
- значения F-критерия: значения F-критерия для ИК и БК выше, по сравнению с другими переменными, указывая на более сильную связь между ними и зависимой переменной.

Для достижения оптимальной производительности и результатов в задаче, связанной с анализом данных, были выбраны следующие первичные настройки моделей:

Boosted Trees:

- минимальное количество элементов в узле для остановки разделения: 79;
- минимальный размер дочернего узла для остановки разделения: 1;
- максимальная глубина дерева: 32 уровня;
- максимальное количество узлов в дереве: 15.

Таблица 3

Коэффициенты корреляции моделей МО

Метод	S_1	S_2	$S_1 + S_2$	T_{\max}	ТОС
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
Множественная регрессия	0,730231	0,769166	0,781709	0,490886	0,742394

Таблица 4

Сравнение корреляции прогнозных моделей методами линейной регрессии и Boosted Trees

Линейная регрессия / Машинное обучение	ML_ S_1	ML_ S_2	ML_ $S_1 + S_2$	ML_ T_{\max}	ML ТОС
Regrss_ S_1	0,926029	0,919872	0,928922	0,562976	0,863218
Regrss_ S_2	0,902885	0,955211	0,957815	0,524011	0,905242
Regrss_ $S_1 + S_2$	0,905664	0,955484	0,958337	0,528347	0,905805
Regrss_ T_{\max}	0,826240	0,776076	0,788178	0,634128	0,733300
RegrssТОС	0,881811	0,950323	0,950109	0,516769	0,914215

Эти параметры позволяют контролировать структуру дерева и предотвращать переобучение.

Neural Network (MLP):

– минимальное количество нейронов на скрытом слое: 3;

– максимальное количество нейронов на скрытом слое: 64;

– количество обучаемых нейронных сетей: 500.

Эти настройки позволяют экспериментировать с различными архитектурами сетей и сохранить наилучший результат.

SVM: тип ядра (kernel): RBF (радиально-базисная функция).

Использование RBF-ядра позволяет работать с нелинейными данными и повышает способность SVM разделять сложные данные.

Эти начальные параметры подобраны с учетом специфики задачи и позволяют нашим моделям эффективно анализировать данные.

Далее выбранными методами машинного обучения, а также методом множественной линейной регрессии были построены модели зависимостей для параметров S_1 , S_2 , T_{\max} и ТОС. В табл. 3 приведены значения коэффициента корреляции для пар фактических и расчетных значений.

Анализ всех трех моделей для прогнозирования переменных позволяет сделать следующие общие выводы:

1. Модель Boosted Trees:

- для прогнозирования целевых переменных данная модель продемонстрировала отличные результаты с низкой среднеквадратичной ошибкой и высокой корреляцией на обучающей выборке;

- важность независимых переменных различается для S_2 и T_{\max} , но в обоих случаях модель показала их статистическую значимость;

- модель Boosted Trees является наилучшей из рассмотренных для обеих целевых переменных.

2. Нейронная сеть (MLP):

- нейронная сеть продемонстрировала неплохую корреляцию с целевыми переменными, хотя среднеквадратичная ошибка и корреляция были ниже, по сравнению с моделью Boosted Trees;

- после дополнительной настройки гиперпараметров достигнуто увеличение результативности модели.

3. Модель SVM (Support Vector Machine):

- модель SVM показала низкую корреляцию с целевыми переменными, среднеквадратичная ошибка также оказалась выше, чем у модели Boosted Trees и нейронной сети;

- эта модель имеет наименьший коэффициент корреляции и наибольшую среднеквадратичную ошибку среди всех рассмотренных моделей.

Для всех целевых переменных модель Boosted Trees является наилучшей, так как она обладает лучшей комбинацией низкой ошибки и высокой корреляции. Нейронная сеть может быть использована в случае, если требуется альтернативная модель, но она потребует дополнительной настройки. Модель SVM может быть менее предпочтительной из-за более высокой среднеквадратичной ошибки и низкой корреляции. Для окончательного выбора модели необходимо также провести оценку их производительности на тестовой выборке.

Анализ результатов показал, что среди регрессионных моделей, полученных методами машинного обучения по величине коэффициента корреляции, выделяется модель Boosted Trees как имеющая наиболее высокие значения. Наименьший коэффициент корреляции достигнут методом опорных векторов (SVM), что является закономерным результатом, так как этот метод наилучшим образом работает в задачах классификации, а не регрессии. Методы множественной линейной регрессии и нейронных сетей показали примерно одинаково средний результат, что также является закономерным результатом, учитывая общую схожесть методики разработки моделей зависимостей.

Относительно метода множественной линейной регрессии результаты модели Boosted Trees отличаются повышенным коэффициентом корреляции, в связи с этим дальнейший сравнительный анализ проводился между результатами именно этих моделей.

Корреляционный анализ моделей, разработанных методами Boosted Trees и множественной линейной регрессии, показал достаточно высокую линейную связь параметров друг с другом, корреляционная матрица приведена в табл. 4, кросс-плоты зависимостей для расчетных величин параметров S_1 , S_2 , T_{\max} и ТОС приведены на рис. 1.

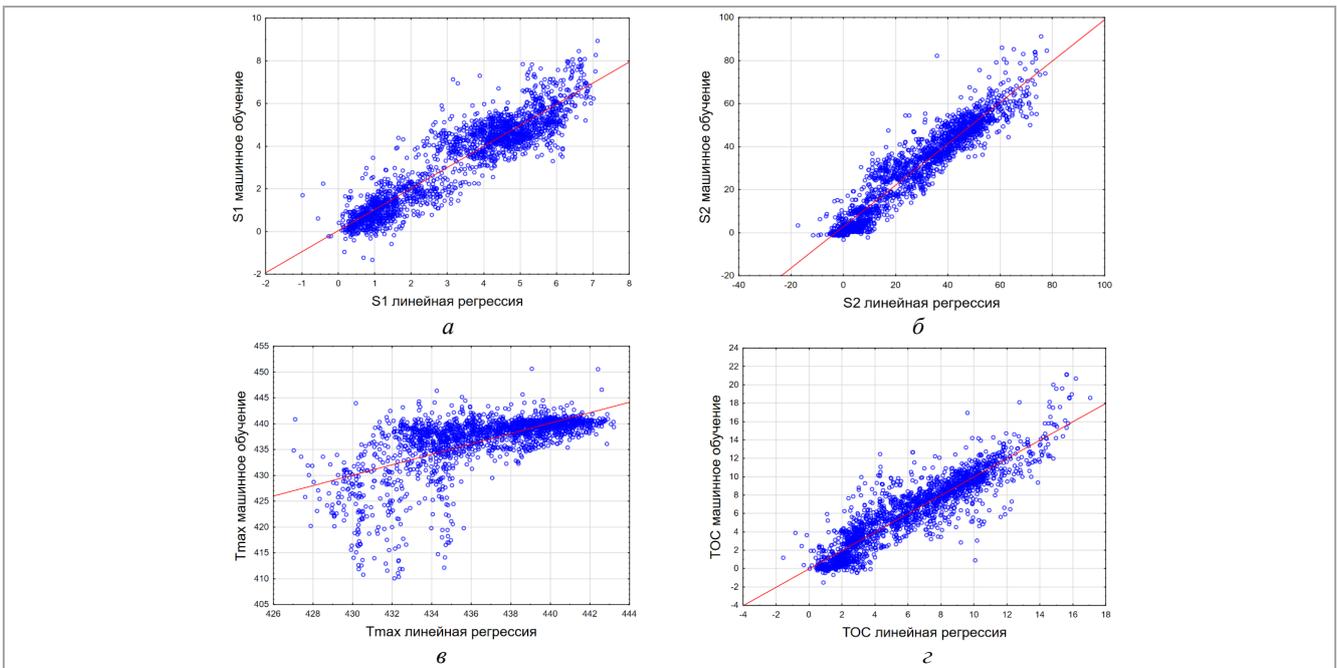


Рис. 1. Графики сопоставления оцененных значений методами линейной регрессии и Boosted Trees для параметров: а – S_1 ; б – S_2 ; в – T_{\max} ; г – TOC

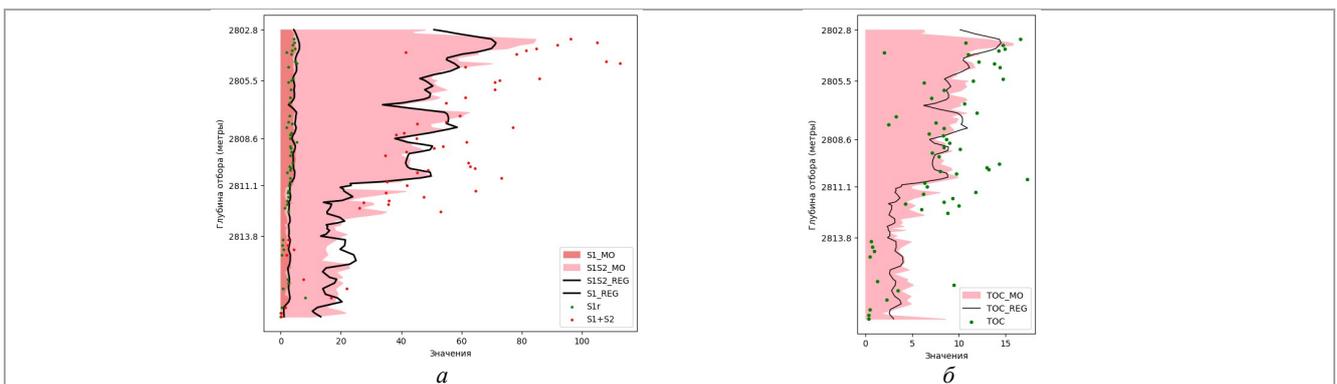


Рис. 2. Сопоставление экспериментальных данных геохимических исследований с синтетической кривой параметров: а – S_1 и S_2 ; б – TOC

Таблица 5

Количественные метрики оценки надежности моделей

Параметр	Метрика/Метод	MAE (меньше – лучше)	MAPE (меньше – лучше)	MSE (меньше – лучше)	R_2 (больше – лучше)
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
	Лин.регрессия	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
	Лин.регрессия	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
	Лин.регрессия	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
	Лин.регрессия	2,206	50,76 %	4,868	0,551

Примечание: * – зеленым цветом отмечены ячейки, характеризующие модель как более точную, красным – менее точную.

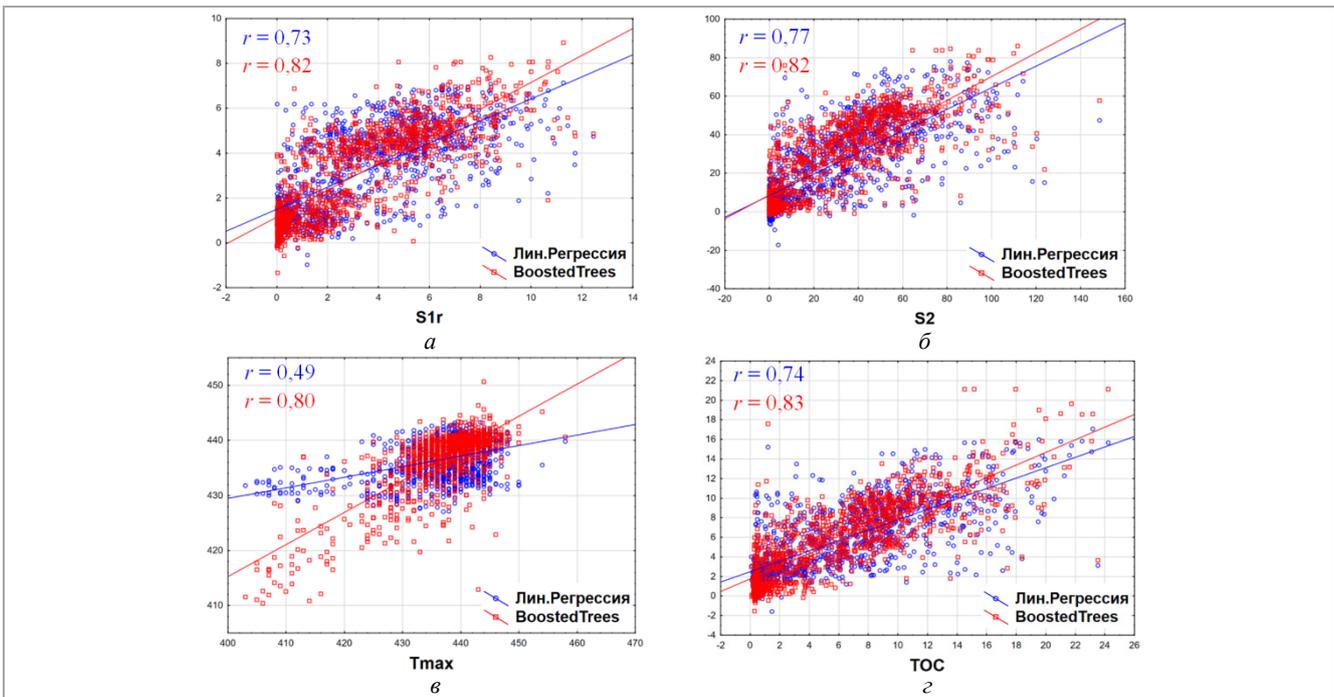


Рис. 3. Сравнение кросс-плотов оцененных пар и экспериментальных значений для моделей линейной регрессии и Boosted Trees параметров: $a - S_1$; $б - S_2$; $в - T_{max}$; $г - ТОС$

Анализ корреляционной матрицы и графиков зависимости результатов прогноза методами линейной регрессии и методом Boosted Trees показал, что связь между параметрами является линейной с высоким коэффициентом корреляции. Наибольшее различие между результатами получено при прогнозировании параметра T_{max} , где метод Boosted Trees показал большую точность прогноза, относительно линейной регрессии, а на графике зависимости между параметрами, рассчитанными разными методами, не наблюдается четкой линейной связи.

Далее было проведено визуальное сопоставление результатов прогноза геохимических параметров разными методами с экспериментальными данными (рис. 2). В качестве методов, между которыми производилось сравнение, выбраны метод Boosted Trees как имеющий наибольшую корреляцию и метод линейной регрессии как наиболее распространенный для решения аналогичных задач.

Визуальный анализ сопоставления синтетических кривых параметров с результатами лабораторных исследований показал, что обе модели достаточно хорошо характеризуют тренд изменения геохимических свойств, однако визуальных отличий между моделями прогноза недостаточно для формирования выводов о приоритете одной из них. В связи с отсутствием визуальной разницы между моделями был произведен расчет стандартных метрик и проведено сопоставление. Результаты сопоставления метрик приведены в табл. 5.

С учетом представленных метрик (MAE, MAPE, MSE и R^2) для всех четырех параметров (S_1 , S_2 , T_{max} и $ТОС$) можно сделать общий вывод, что Boosted Trees являются наилучшей моделью машинного обучения среди всех рассмотренных методов.

Модель Boosted Trees выделяется в следующих аспектах:

1. Точность прогнозирования: Boosted Trees имеют наименьшие значения MAE и MSE, это означает, что они предсказывают значения ближе к фактическим данным и имеют меньшие ошибки в абсолютных и квадратичных значениях.

2. Процентные ошибки: метод Boosted Trees также показывает наименьшие значения MAPE, что свидетельствует о наименьших процентных ошибках в предсказаниях. Это означает, что прогнозы Boosted Trees наименее искажены в процентном соотношении.

3. Способность объяснения изменчивости: Boosted Trees демонстрирует наивысшие значения R^2 для всех параметров, что указывает на их способность наилучшим образом объяснять изменчивость в данных.

Следовательно, Boosted Trees является наилучшим выбором в данной задаче и обеспечивают наиболее точные и надежные прогнозы с наименьшими ошибками как в абсолютных значениях, так и в процентном соотношении, а также обладает наивысшей способностью объяснения изменчивости в данных.

Сравнительный визуальный анализ корреляционных полей, представленный на рис. 3, также показал увеличение точности прогноза модели, разработанной методом BoostedTrees, что также подтверждает выводы, полученные по результатам сравнения численных метрик оценки моделей прогноза.

Закключение

В работе проведен анализ различных методов регрессии, среди которых лидирует модель Boosted Trees. Эта модель демонстрирует лучшие результаты среди всех рассмотренных методов, имея наименьшую ошибку (MAE и MSE) и наивысший коэффициент детерминации (R^2). Boosted Trees обеспечивают точные и стабильные результаты прогнозирования, подтвержденные

визуальным анализом. Исследование показывает, что модель Boosted Trees относительно метода линейной регрессии более эффективна и точна для решения задачи установления зависимости керн-ГИС для геохимических данных.

Благодаря применению альтернативных методов регрессии на основе алгоритмов машинного обучения был проведен расчет геохимических параметров, в том числе S_1 , S_2 и T_{\max} , которые не удавалось рассчитать ранее. Применение методов МО повысило точность оценки геохимических параметров в разрезе скважин, что приведет к увеличению точности моделей распределения свойств пород по площади.

Методы машинного обучения в геологии имеет большой потенциал применения в задачах регрессии, дискриминации и создания различных прогнозов на основании выявленных закономерностей. Однако это требует тщательной предобработки данных, выбора

наилучших моделей и методов обучения, а также внимательной проверки результатов. Несмотря на эти технические сложности, использование машинного обучения в геологии открывает новые перспективы для ученых и исследователей, позволяя им более точно анализировать и интерпретировать данные, а также принимать производственные решения на основе этих данных.

Оценка геохимических свойств пород этой формации вносит большой вклад в прогнозирование нефтеносности, а также в оценку перспектив малоизученных областей и глубокозалегающих горизонтов. Результаты исследований будут использованы для дальнейшего определения коллекторов и разделения их на два типа в соответствии с методическими рекомендациями по подсчету запасов нефти в отложениях баженовского горизонта Западно-Сибирской нефтегазоносной провинции.

Библиографический список

1. Shadrin A.O., Krivoshechekov 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., Кривошеков 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. Брэдучан, 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. Гагина, 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 Широкого 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 свиты 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. Кривошеков [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. Кривошеков // Нефтяное хозяйство. – 2015. – № 1. – С. 21–23. – EDN T1JZT.
18. О масштабах миграции углеводородов в пределах Соликамской депрессии Предуральяского 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. Кривошеков 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 UEKTBV.
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 Appalachian basin // *Bull. of US Geol. Surv.* – 1993. – 1909. – P. 71–74.
33. Белоконов Т.В., Наборщикова И.И., Ерсулова И.С. Определение ПОВ 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 sour-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., Yunusov 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 son 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.

References

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, no. 5, 052056 p. DOI: 10.1088/1755-1315/666/5/052056
2. Shadrin A.O., Krivoshchekov S.N. Razrabotka veroiatnostno-statisticheskikh modelei prognoza neftenosnosti po strukturnym parametram plasta IuS1 v severnoi chasti Surgut'skogo svoda [Prediction of oil occurrence using structural parameters of YUS1 reservoir in the northern part of Surgut Arch: development of probabilistic-statistical models]. *Geologiya nefi i gaza*, 2022, no. 2, pp. 53–65. DOI: 10.31087/0016-7894-2022-2-53-65. EDN TUHENY.
3. Shadrin A.O., Akhmetova L.V. Analiz vliianiia strukturnykh parametrov osadochnogo chekhla na neftegazonosnost' plasta IuS1 severnoi chasti Surgut'skogo svoda [Analysis of the influence of structural parameters of the sedimentary cover on the oil and gas content of the YUS1 formation in the northern part of the Surgut arch]. *Nef' i gaz - 2021. Sbornik trudov 75-i Mezhdunarodnoi molodezhnoi nauchnoi konferentsii, Moskva, 26-30 April 2021. Moscow: Rossiiskii gosudarstvennyi universitet nefi i gaza (natsional'nyi issledovatel'skii universitet) imeni I.M. Gubkina*, 2021, pp. 349–359. EDN WDVQPP.
4. Shadrin A.O. Prognoz neftegazonosnosti plasta IuS1 severnoi chasti Surgut'skogo svoda na osnove provedeniia trend-analiza [Forecast of oil and gas bearing field JS1 of the northern part of Surgut Arch based on trend-analysis]. *Problemy razrabotki mestorozhdenii uglevodorodnykh i rudnykh poleznykh iskopaemykh*, 2020, vol. 1, pp. 154–160.
5. Shadrin A.O. Razrabotka zonal'noi modelei prognoza neftegazonosnosti plasta IuS1 po geokhimicheskim parametram [Development of a Zonal Model for Forecasting the Oil and Gas Potential of the YUS1 Formation by Geochemical Parameters]. *Nedropol'zovanie*, 2023, vol. 23, no. 1, pp. 2–10. DOI: 10.15593/2712-8008/2023.1.1. EDN LLSEXO.
6. Braduchan Iu.V., Glushko N.K., Komissarenko V.K. et al. O vozraste otlozhenii anomal'nykh razrezov pogranichnykh sloev iury i mela po skvazhinam Severo-Konitlorskogo mestorozhdeniia (predvaritel'noe soobshchenie) [On the age of deposits of anomalous sections of the Jurassic and Cretaceous boundary layers in the wells of the North Konitlorskoye field (preliminary report)]. *Vestnik nedropol'zovatel'ia Khanty-Mansiiskogo avtonomnogo okruga*, 2005, no. 16, pp. 20–24.
7. Grishkevich V.F., Gatina N.N., Sidorenko A.O., Karpova E.V. Opyt petrograficheskogo analiza mekhanizma formirovaniia anomal'nogo razreza bazhenovskoi svity na Imilorskoi ploskhadi Zapadnoi Sibiri [A petrographic study of the genesis of anomalous sections in the Bazhenov Formation, the Imilor deposit, Western Siberia]. *Litosfera*, 2019, no. 2 (19), pp. 209–227. DOI: 10.24930/1681-9004-2019-19-2-209-227
8. Shaikhutdinova G.Kh. Petrograficheskoe izuchenie migratsii nefi na primere Imilorskogo mestorozhdeniia (Kogalymskii neftegazonosnyi raion, Zapadnaia Sibir') [Petrographic study of oil migration on the example of Imilorskoye field (Kogalymsky petroleum region, Western Siberia)]. *Litosfera*, 2020, no. 4 (20), pp. 592–600. DOI: 10.24930/1681-9004-2020-20-4-592-600
9. Skachek K.G., Osyka A.V., Garifullin I.I. Perspektivy neftenosnosti bazhenovskoi svity Kogalym'skogo regiona [Oil-bearing prospects of the Bazhenov Formation in the Kogalym region]. *Puti realizatsii neftegazovogo potentsiala Khanty-Mansiiskogo avtonomnogo okruga: sed'maia nauchno-prakticheskaiia konferentsiia*. Ed. V.I. Karasev, E.A. Akhpatelov, V.A. Volkov. Khanty-Mansiisk, 2004, vol. 1, pp. 162–170.
10. Nemova V.D. Litologiya i kollektorskie svoistva otlozhenii bazhenovskogo gorizonta na zapade Shirotnogo Priob'ia [Lithology and reservoir properties of deposits of the Bazhenov horizon in the west of the Latitudinal Ob region]. Ph. D. thesis. Moscow, 2012.
11. Nemova V.D., Bedretdinov R.Iu., Kirsanov A. Interpretatsiia dannykh GIS V intervale bazhenovskoi svite v usloviakh ogranichenno kompleksa GIS na territorii Sredne-Nazym'skogo mestorozhdeniia [Interpretation of well logging data in the Bazhenov formation interval under the conditions of a limited well logging complex in the territory of the Sredne-Nazym'skoye field]. *Petrofizika slozhnykh kollektorov: problemy i perspektivy 2015*. Moscow: OOO "EAGE Geomodel", 2015, pp. 174–159.
12. Nemova V.D., Gavrilov S.S. Issledovaniia kerna otlozhenii bazhenovskogo gorizonta, kak osnova dlia interpretatsii dannykh seismorazvedki [Studies of sediment cores from the Bazhenov horizon as a basis for interpreting seismic data]. *Petrofizika slozhnykh kollektorov: problemy i perspektivy 2014*. Moscow: OOO "EAGE Geomodel", 2014, pp. 212–230.
13. Grishkevich V.F. Anomal'nye razrezy bazhenovskoi svity: model' obrazovaniia i problemy geometrizatsii [Enormous Sequences of Bajenovskaya Svita: Genetic Model and Geometrical Approach], *Tyumen 2015 - Deep Subsoil and Science Horizons, Mar 2015, Vol. 2015, pp.1-5*. DOI: 10.3997/2214-4609.201412065
14. Grishkevich V.F. Bazhenovskii gorizont Zapadnoi Sibiri: poiski novoi harmonii [Bazhenov horizon of Western Siberia: search for a new harmony]. *Tiumen: Tiument'skii industrial'nyi universitet*, 2022, 279 p.
15. Nesterov I.I. Novyi tip kollektorov nefi i gaza [New type of oil and gas reservoirs]. *Geologiya nefi i gaza*, 1979, no. 10, pp. 26–29.
16. Galkin V.I., Kozlova I.A., Krivoshchekov S.N. et al. O vozmozhnosti prognozirovaniia neftegazonosnosti famenskikh otlozhenii s pomoshch'iu postroeniia veroiatnostno-statisticheskikh modelei [On the possibility of predicting the oil and gas content of Famenskiy deposits using the construction of probabilistic-statistical models]. *Geologiya, geofizika i razrabotka nefiannykh i gazovyykh mestorozhdenii*, 2007, no. 10, pp. 22–27. EDN IBLHOX.
17. Galkin V.I., Kozlova I.A., Nosov M.A., Krivoshchekov S.N. Reshenie regional'nykh zadach prognozirovaniia neftenosnosti po dannykh geologo-geokhimicheskogo analiza rasseianiya organicheskogo veshchestva porod domanikovogo tipa [Solutions to regional problems of forecasting oil bearing according to geological and geochemical analysis of dispersed organic matter of Domani type rocks]. *Nef'tianoe khoziaistvo*, 2015, no. 1, pp. 21–23. EDN TLJTZT.

18. Voevodkin V.L., Galkin V.I., Kozlova I.A. et al. O mashtabakh migratsii uglevodorodov v predelakh Solikamskoi depressii Predural'skogo progiba i vozmozhnostiakh ee ispol'zovaniia dlia prognoza neftegazonosnosti [Hydrocarbons migration volumes within the limits of Solikamsk depression (Pre-Ural deflection) and possibilities of its use for the oil and gas content forecast]. *Geologiya, geofizika i razrabotka neftiannykh i gazovykh mestorozhdenii*, 2010, no. 12, pp. 6-11. EDN NCQIEL.
19. Krivoshchekov S.N. Razrabotka regional'no-zonal'nykh kriteriev prognoza neftegazonosnosti territorii Permskogo Prikam'ia veroiatnostno-statisticheskimi metodami [Establishment of regional and zonal criteria of Perm region's oil forecast probabilistic and statistical methods]. *Neftianoe khoziaistvo*, 2011, no. 10, pp. 10-14. EDN OHKPOJ.
20. Bust V.K., Majid A.A., Oletu J.U., Worthington P.F. The Petrophysics of Shale Gas Reservoirs: Technical Challenges and Pragmatic Solutions. *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*. DOI 10.2118/144062-MS
22. Dakhnova M.V., Mozhegova S.V., Nazarova E.S., Paizanskaia I.L. Otsenka zapasov "slantsevoi nefti" s ispol'zovaniem geokhimicheskikh parametrov [Evaluation of reserves of shale oil using geochemical parameters]. *Geologiya nefi i gaza*, 2015, no. 4, pp. 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. Kirsanov A.M., Skvortsov M.B. Opreделение podschetykh parametrov po dannym GIS dlia otsenki resursov bazhenovskoi svity [Determination of calculation parameters for the problem of resource estimation based on well logs data in the Bazhenov Formation interval]. *Neftianaia stolitsa. Materialy 4-i Mezhdunarodnogo molodezhnogo nauchno-prakticheskogo foruma, Khanty-Mansiisk, 24-25 March 2021*. Khanty-Mansiisk: Tsentr nauchno-tekhnicheskikh reshenii, 2021, pp. 100-103. EDN CBBBCW.
25. Delfiner Pierre et al. Automatic Determination of Lithology From Well Logs. *SPE Formation Evaluation, 1987, vol. 2, iss. 03*, pp. 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, 47-00144 p. DOI: 10.3997/2214-4609.201409899
27. Zheng D. et al. Application of machine learning in the identification of fluvial-lacustrine lithofacies from well logs: A case study from Sichuan Basin, China. *Journal of Petroleum Science and Engineering*, 2022, vol. 215, 110610 p. DOI: 10.1016/j.petrol.2022.110610
28. Ao Y. et al. Probabilistic logging lithology characterization with random forest probability estimation. *Computers & Geosciences*, 2020, vol. 144, 104556 p. DOI: 10.1016/j.cageo.2020.104556
29. Alfred D., Vernik L. A new petrophysical model for organic shales. *Petrophysics*, 2013, vol. 54, no. 03, pp. 240-247.
30. Chen S. et al. A new approach to calculate gas saturation in shale reservoirs. *Energy & Fuels*, 2022, vol. 36, no. 4, pp. 1904-1915. DOI: 10.1021/acs.energyfuels.1c04067
31. Wang M. et al. The key parameter of shale oil resource evaluation: Oil content. *Petroleum Science*, 2022, vol. 19, no. 4, pp. 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, pp. 71-74.
33. Belokon' T.V., Naborschchikova I.L., Ersulova I.S. Opreделение ROV v porodakh po estestvennoi gamma-aktivnosti [Determination of DOM in rocks by natural gamma activity]. *Evolutsiia neftegazooobrazovaniia i istorii Zemli. Tezisy dokladov IV Vsesotuznogo seminara*. Moscow: Moskovskii gosudarstvennyi universitet, 1984, 236 p.
34. Meyer B.L., Nederlof M.H. Identification of sour-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, pp. 121-129. DOI: 10.1306/AD4609E0-16F7-11D7-8645000102C1865D
35. Mosse L., Rylander E., Craddock P. US Patent No 14/977.336 (22 Jun. 2022).
36. Rashka S., Mirdzhalili V. Python i mashinnoe obuchenie: mashinnoe i glubokoe obuchenie s ispol'zovaniem Python, scikit-learn i TensorFlow-2 [Python and Machine Learning: Machine and deep learning using Python, scikit-learn and TensorFlow-2]. Saint Petersburg: Dialektika, 2020, vol. 848.
37. Breiman L. Random forests. *Machine learning*, 2001, vol. 45, pp. 5-32. DOI: 10.1023/A:1010933404324
38. Friedman J.H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 2001, pp. 1189-1232. DOI: 10.1214/aos/1013203451
39. Doveton J.H. Compositional analysis of mineralogy. *Principles of Mathematical Petrophysics*. Oxford University Press, 2014, pp. 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, pp. 206-212. DOI: 10.1109/ICECCO.2018.8634775
41. Mohamed I.M. et al. Formation lithology classification: Insights into machine learning methods. *SPE Annual Technical Conference and Exhibition, OnePetro*, 2019. DOI: 10.2118/196096-MS
42. Sapatelli L. et al. Deriving permeability and reservoir rock typing supported with self-organized maps som and artificial neural networks ann-optimal workflow for enabling core-log integration. *SPE Reservoir Characterisation and Simulation Conference and Exhibition? SPE*, 2019, D021S015R001 p. 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. Karpatne A. et al. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 2018, vol. 31, no. 8, pp. 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, no. 03, pp. 268-280.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов равноценен.