

ОРИГИНАЛЬНАЯ СТАТЬЯ

УДК 681.518+622.276.63

doi: 10.26907/2542-064X.2024.4.623-639

## ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ ВЫБОРА ОБЪЕКТОВ ВОЗДЕЙСТВИЯ И ПРОЕКТИРОВАНИЯ КИСЛОТНОЙ ОБРАБОТКИ

*И.И. Маннанов, М.А. Варфоломеев, Г.Р. Ганиева, А.Р. Гимаева,  
Р.Р. Гиниятуллин*

*Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия*

### Аннотация

Для карбонатных отложений кислотная обработка является одним из наиболее востребованных решений для интенсификации добычи нефти, которое активно применяется на практике. Анализ успешности соответствующих работ показывает высокую дифференцированность промысловых результатов. Современные подходы к решению задачи проектирования дизайна кислотной обработки основываются на применении программных продуктов, позволяющих учесть многие факторы, влияющие на процесс обработки. Применение машинного обучения может быть эффективным дополнением к подходам в части выбора объектов воздействия и обоснования исходных данных для проектирования дизайна в реализованных и эффективно функционирующих программных продуктах. В работе приводятся решения по использованию машинного обучения для выбора объектов воздействия на основе данных ранее выполненных работ по влиянию начальных режимных скважинных условий, технологических условий выполнения работ, кратности обработки, объемов, типов кислотных систем, предобработки, применения отклонителей кислотных систем, а также времени выдерживания кислоты на результативность обработки. Проектирование кислотной обработки предусматривает выполнение сложных лабораторных исследований для изучения кинетики процесса взаимодействия кислотных систем с породой, ключевыми параметрами которого являются минералогический состав горной породы и характеристики применяемых кислотных составов, в частности концентрации компонентов кислотной композиции. В работе рассмотрена задача прогнозирования кинетики реакции кислотных систем методами машинного обучения на основе обработки массива данных лабораторных исследований. В качестве базовых решений прогнозирования использованы методы линейной регрессии и случайного леса. Привлечение машинного обучения позволяет создать алгоритмы принятия решений для оптимизации кислотной обработки с учетом многофакторного влияния условий ее выполнения. Разработанные алгоритмы значительно упрощают задачи проектирования кислотной обработки.

**Ключевые слова:** колебательные спектры, молекулярная структура, электронография, квантово-химические расчеты.

### Введение

Доля карбонатных отложений в мировой структуре извлекаемых запасов нефти и газа по различным оценкам составляет более 60 % и продолжает расти [1–3].

Традиционно для разработки коллекторов, содержащих в составе пород карбонатные минералы, применяют различные модификации технологии кислотной обработки пластов [4]. Этот подход можно считать одним из основных, универсальных и действенных методов, обеспечивающих гидродинамическую связь скважины и пласта коллектора. Характерным признаком опыта применения кислотной обработки является факт различной технологической успешности в разновозрастных или техногенно измененных карбонатных отложениях. Основная причина таких результатов заключается в природном разнообразии карбонатных коллекторов по генезису, минеральному составу, структуре порового пространства, неоднородности и проницаемости. Другой, не менее важной, причиной различий в эффективности применяемого кислотного состава являются влияние накопленной техногенной нагрузки на пласт и ее рост при кислотном воздействии. Проявления этой нагрузки снижают проницаемость призабойной зоны пласта [5].

Одна из проблем, с которыми сталкиваются нефтедобывающие предприятия, состоит в снижении плановых и фактических приростов дебитов после выполнения работ по интенсификации добычи. Успешность кислотной обработки, особенно на месторождениях трудноизвлекаемых запасов, составляет по различным оценкам не более 30–40 %, при этом основными причинами низкой успешности являются сложное геологическое строение, свойства минералов, слагающих пласт, термобарические условия призабойной зоны, невозможность создания оптимального дизайна обработки скважины [6, 7].

Изменение проницаемости и пористости породы в результате кислотной обработки – сложный процесс, так как на него оказывают влияние несколько различных, порой конкурирующих, явлений в пористой среде. Проницаемость возрастает с увеличением размеров пор за счет растворения минералов скелета пород. В то же время при растворении цементирующего материала породы освобождаются твердые частицы различных размеров, не взаимодействующие с кислотой [8, 9].

Важным этапом применения кислотных композиций является лабораторное изучение процесса взаимодействия кислотных составов с породой. Зачастую это трудоемкие и сложные эксперименты, направленные на получение данных, отражающих условия протекания процесса, и ориентированные на дальнейшее их применение в симуляторах процесса кислотного воздействия.

Перенос результатов лабораторного изучения растворения породы в скважинах невозможен без применения методов численного моделирования протекающих процессов. Математическое моделирование и симуляции являются эффективными решениями при проектировании успешных кислотных обработок в масштабах скважины [10].

Главная задача планирования выполнения технологий обработки призабойной зоны (ОПЗ) пласта связана с решением по выбору оптимального темпа закачки кислотного состава в пласт, что обеспечивает максимальное снижение скин-фактора и образование червоточин в пласте, обеспечивающих, в свою очередь, эффективное сообщение пласта со скважиной [13]. Скорость закачки, которая позволяет провести кислотную обработку с максимальной эффективностью, относят к оптимальным параметрам. Для подтверждения оценки эффективности необходимо исследовать различные параметры, такие как температура, давление, минералогический состав породы и влияние химических отклонителей. Также необходимо учесть влияние

конвекции и диффузии [14]. Современные задачи планирования и проектирования выполнения кислотных обработок могут быть успешно дополнены при использовании методов машинного обучения.

## 1. Методы машинного обучения

Методы машинного обучения, получившие существенное развитие в последнее время, позволяют по-новому взглянуть на многочисленные фактические результаты, полученные в нефтегазовой отрасли за годы эксплуатации скважин, применения методов интенсификации добычи, выполнения лабораторных исследований по изучению показателей эффективности кислотных составов и моделированию процесса кислотной обработки [15, 16].

Для решения задачи оптимизации кислотной обработки необходимо разработать комплексную методику прогнозирования с использованием многомерного регрессионного анализа с учетом данных лабораторных исследований. Регрессионный анализ включает в себя статическую модель, для создания которой необходимо учесть все виды операций с привлечением геолого-физических и технологических показателей, таких как:

- проницаемость ствола скважины и удаленных зон пласта;
- обводненность;
- отношение пластового и скважинного давлений к давлению насыщения;
- удельный дебит нефти перед остановкой скважины на капитальный ремонт;
- удельный расход кислоты;
- давление впрыска;
- скорость впрыска;
- время воздействия на реакцию.

Удельный прирост добычи нефти после операций ОПЗ принимали за показатель эффективности проведенных работ. Для оценки достоверности регрессионных моделей на первом этапе сравнивали результаты расчетных и фактических значений, а на втором – оценивали эффективность обработок, включенных в адресную программу геолого-технических мероприятий, с использованием статистических зависимостей, а также проводили лабораторную оценку исследований. Способ прогнозирования эффективности кислотной обработки позволил разработать, обосновать и апробировать подход на основе математического и физического моделирования. Реальные геологические полевые материалы дали возможность получить уравнения многомерной регрессии для прогнозирования эффективности кислотных обработок с целью точной оценки результата обработки призабойных зон скважин различными химическими реагентами [11].

Оценку эффективности кислотной обработки проводили с помощью пошагового регрессионного анализа и вычисления значения коэффициента корреляции. Расчет регрессионного коэффициента в разрабатываемой модели выполняли при помощи метода наименьших квадратов. Для расчета рассматривали два варианта: в первом случае использовали геолого-физические и технологические параметры, а во втором – скорость закачки применяемой для обработки кислоты. Все данные необходимо получать с учетом многомерных статистических моделей [12].

Для достоверного прогнозирования проводят оценку эффективности кислотной обработки с помощью регрессии опорных векторов, нейронных сетей, методов дерева решений, случайного леса (метод, который объединяет несколько посредственных базовых оценок, чтобы обеспечить лучшую общую производительность) и  $k$ -ближайших соседей (интуитивная идентификация аналогового и целевого объектов на основе близости их параметров). Показано, что модель, основанная на нейронных сетях, дает надежные и более точные результаты по оценке давления насыщения, объемного коэффициента нефти и плотности газа и нефти [17].

Рассмотрим применение методов машинного обучения на примере решения двух задач:

- выбор объектов воздействия для достижения максимального эффекта на основе машинного обучения по данным статистики применения методов ОПЗ;
- прогноз результатов кинетики и констант растворения образцов, состоящих из различных минералов, по результатам комплекса лабораторных исследований.

## **2. Прогноз выбора предпочтительных объектов для ОПЗ с применением метода машинного обучения на основе результатов ранее выполненных работ**

В качестве исходных данных использованы результаты применения технологий ОПЗ на скважинах Бавлинского месторождения.

Для построения различных моделей из набора данных были сформированы следующие подвыборки объектов:

- 1) все объекты, в которых применялись химические методы ОПЗ (230 объектов);
- 2) результаты первичной обработки призабойной зоны кислотными составами (146 объектов);
- 3) результаты повторной обработки призабойной зоны кислотными составами (84 объекта);
- 4) результаты обработки вертикальных скважин (после первичной и повторной кислотной обработки) (132 объекта);
- 5) результаты обработки горизонтальных скважин (после первичной и повторной кислотной обработки) (98 объектов);
- 6) результаты обработки вертикальных скважин после первичной кислотной обработки (83 объекта);
- 7) результаты обработки вертикальных скважин после повторной кислотной обработки (49 объектов);
- 8) результаты обработки горизонтальных скважин после первичной кислотной обработки (63 объекта);
- 9) результаты обработки горизонтальных скважин после повторной кислотной обработки (35 объектов).

При построении прогнозных моделей и проведении анализа учитывали следующие показатели:

- геолого-эксплуатационные характеристики – коэффициент продуктивности, пористость, плотность перфорации, дебит жидкости, дебит нефти,

обводненность, пластовое и забойное давление, перфорированная мощность, сумма перфорационных отверстий и наличие открытого ствола;

- технологические характеристики выполнения ОПЗ, в частности значения ожидания реакции кислоты (ОРК), давление при продавке, количество обработок скважины, объем кислотных составов;

- химические характеристики – концентрация активных компонентов в кислотных составах, виды кислотных составов, наличие и объем отклонителя.

Следует отметить, что категориальные признаки были преобразованы в числовые с использованием метода One-Hot Encoding.

Для каждой подвыборки данных строили модели, используя:

- 1) Все перечисленные факторы (дескрипторы);
- 2) Параметры, характеризующие эксплуатационные характеристики (дебит жидкости, дебит нефти, обводненность и т. д.);
- 3) Параметры, характеризующие эксплуатационные характеристики (дебит жидкости, дебит нефти, обводненность) и объемы и виды технологических жидкостей;
- 4) Параметры, характеризующие технологические жидкости, типы кислотных составов и объемы, без эксплуатационных характеристик (дебит жидкости, дебит нефти, обводненность и т. д.).

При этом анализировали данные, отражающие фактическое исполнение дизайна обработки с учетом давления закачки.

В результате было построено 36 моделей (9 подвыборок по четырем типам перебора дескрипторов). Проведенный сравнительный анализ всех моделей показал, что среди лучших моделей оказалась библиотека градиентного бустинга на деревьях решений с открытым исходным кодом (CatBoost).

В результате проведенного исследования установлено, что высокие значения предсказательной способности классификационных моделей наблюдаются для системы первичной обработки вертикальных скважин и системы первичной обработки вертикальных и горизонтальных скважин. Значения сбалансированной точности (BA), полноты (TPR) и точности (PPV) разработанной модели на основе CatBoost для системы первичной обработки вертикальных скважин составили 0.84, 0.82 и 0.61 соответственно, а для системы первичной обработки вертикальных и горизонтальных скважин – 0.79, 0.56 и 0.72 соответственно (табл. 1).

На основе анализа зависимостей дескрипторов с целевой переменной с помощью библиотеки SHAP (она позволяет оценить, насколько рассматриваемый признак изменил предсказание модели по сравнению с моделью при некоем базовом значении этого признака) выявлено, что в среднем для всех систем обработка скважины успешнее при высоких значениях дебита жидкости до обработки, пористости, пластового и забойного давления до обработки, а также содержания нефти в продукции скважины. Таким образом, отсутствует явная зависимость эффективности обработки от типа использованной кислотной композиции. Эффективность обработки определяется почти исключительно начальными условиями на скважине, также заметен эффект от учета объема используемой кислотной композиции. Поскольку в выборке присутствовали кислотные композиции, оптимально подобранные для скважины, то они работали почти одинаково хорошо, и при моделировании влияние типа композиции на успешность ОПЗ не наблюдается.

Табл. 1

## Результаты предсказательной способности классификационных моделей

Наборы дескрипторов	Подвыборка объектов набора данных	Кол-во данных	Значения, отражающие сбалансированность и точность модели													
			TN	FP	FN	TP	Acc	TPR	TNR	BA	PPV	PNV	AUC	F_1	F_0.5	F_2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
С учетом эксплуатационных характеристик и химических параметров	Все обработки ПЗ	230	230	106	42	33	49	0.674	0.598	0.716	0.539	0.763	0.725	0.566	0.549	0.585
	Первичная обработка ПЗ	146	146	71	29	16	30	0.692	0.652	0.710	0.509	0.816	0.764	0.571	0.532	0.617
	Повторные обработки ПЗ	84	84	30	18	19	17	0.560	0.472	0.625	0.486	0.612	0.530	0.479	0.483	0.475
	ГС	98	98	35	19	21	23	0.592	0.523	0.648	0.548	0.625	0.631	0.535	0.542	0.528
	ВС	132	132	75	19	15	23	0.742	0.605	0.798	0.548	0.833	0.745	0.575	0.558	0.593
	Первичная обработка ГС	63	63	22	12	13	16	0.603	0.552	0.647	0.571	0.629	0.590	0.561	0.567	0.556
	Первичная обработка ВС	83	83	57	9	3	14	0.855	0.824	0.864	0.609	0.950	0.909	0.700	0.642	0.769
	Повторные обработки ГС	35	35	11	9	8	7	0.514	0.467	0.550	0.438	0.579	0.523	0.452	0.443	0.461
	Повторные обработки ГС	49	49	20	8	12	9	0.592	0.429	0.714	0.529	0.625	0.648	0.474	0.506	0.446
	Все обработки ПЗ	230	230	93	55	44	38	0.570	0.463	0.628	0.409	0.679	0.589	0.434	0.419	0.451
Без учета эксплуатационных характеристик, но с учетом химических параметров	Первичная обработка ПЗ	146	146	66	34	18	28	0.644	0.609	0.660	0.452	0.786	0.722	0.519	0.476	0.569
	Повторные обработки ПЗ	84	84	26	22	23	13	0.464	0.361	0.542	0.371	0.531	0.487	0.366	0.369	0.363
	ГС	98	98	35	19	19	25	0.612	0.568	0.648	0.568	0.648	0.596	0.568	0.568	0.568
	ВС	132	132	68	26	19	19	0.659	0.500	0.723	0.422	0.782	0.681	0.458	0.436	0.482
	Первичная обработка ГС	63	63	18	16	17	12	0.476	0.414	0.529	0.429	0.514	0.507	0.421	0.426	0.417
	Первичная обработка ВС	83	83	59	7	5	12	0.855	0.706	0.894	0.632	0.922	0.824	0.667	0.645	0.690
	Повторные обработки ГС	35	35	11	9	7	8	0.543	0.533	0.550	0.471	0.611	0.523	0.500	0.482	0.519
	Повторные обработки ГС	49	49	19	9	12	9	0.571	0.429	0.679	0.500	0.613	0.478	0.462	0.484	0.441

Продолжение табл. 1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
С учетом эксплуатационных характеристик, без учета химии и объемов химии	Все обработки ПЗ	230	106	42	33	49	0.674	0.598	0.716	0.657	0.539	0.763	0.719	0.566	0.549	0.585
		146	70	30	13	33	0.706	0.717	0.700	0.709	0.524	0.843	0.786	0.606	0.554	0.668
	Первичная обработка ПЗ	84	30	18	15	21	0.607	0.583	0.625	0.604	0.539	0.667	0.586	0.560	0.547	0.574
		98	35	19	22	22	0.582	0.500	0.648	0.574	0.537	0.614	0.615	0.518	0.529	0.507
	ВС	132	69	25	13	25	0.712	0.658	0.734	0.696	0.500	0.842	0.758	0.568	0.525	0.619
		63	20	14	10	19	0.619	0.655	0.588	0.622	0.576	0.667	0.616	0.613	0.590	0.638
	Первичная обработка ГС	83	55	11	2	15	0.843	0.882	0.833	0.858	0.577	0.965	0.898	0.698	0.620	0.798
		35	15	5	6	9	0.686	0.600	0.750	0.675	0.643	0.714	0.673	0.621	0.634	0.608
	Повторные обработки ГС	49	23	5	10	11	0.694	0.524	0.821	0.673	0.688	0.697	0.672	0.595	0.647	0.550
		230	104	44	30	52	0.678	0.634	0.703	0.668	0.542	0.776	0.730	0.584	0.558	0.613
С учетом эксплуатационных характеристик и объемов химии, без учета химии	Первичная обработка ПЗ	146	73	27	13	33	0.726	0.717	0.730	0.724	0.550	0.849	0.794	0.623	0.577	0.676
		84	31	17	15	21	0.619	0.583	0.646	0.615	0.553	0.674	0.600	0.568	0.559	0.577
	Повторные обработки ПЗ	98	35	19	21	23	0.592	0.523	0.648	0.585	0.548	0.625	0.619	0.535	0.542	0.528
		132	72	22	13	25	0.735	0.658	0.766	0.712	0.532	0.847	0.784	0.588	0.553	0.628
	ВС	63	19	15	11	18	0.587	0.621	0.559	0.590	0.546	0.633	0.621	0.581	0.559	0.604
		83	55	11	3	14	0.831	0.824	0.833	0.828	0.560	0.948	0.897	0.667	0.598	0.753
	Первичная обработка ГС	35	16	4	6	9	0.714	0.600	0.800	0.700	0.692	0.727	0.633	0.643	0.672	0.616
		49	22	6	12	9	0.633	0.429	0.786	0.607	0.600	0.647	0.662	0.500	0.556	0.455

Примечания: ПЗ – призабойная зона; ГС – горизонтальные скважины; ВС – вертикальные скважины; TN – истинно отрицательные случаи прогноза; FR – ложноположительные случаи прогноза; FN – ложноотрицательные случаи прогноза; TP – истинно положительные случаи прогноза; Асс – площадь под кривой; TPR – истинно положительный показатель; TNR – ложноположительный показатель; BA – сбалансированная точность; PPV – прогностическая значимость положительной классификации; PNV – доля ложных пропусков; AUC – качество работы классификатора; F\_1 – среднее гармоническое точности и полноты; F\_0.5 – F-мера с приоритетом точности; F\_2 – F-мера с приоритетом полноты. Изменение цвета от зеленого к красному отражает снижение сбалансированной точности предсказания.

Также проанализировано влияние на качество модели учета коэффициента продуктивности и пористости. Для линейных моделей учет таких характеристик, как коэффициент продуктивности и плотность перфорации, не позволил выявить зависимости. Как видно из табл. 2, для нелинейных моделей лучше исключить комплексный коэффициент продуктивности, не учитывающий абсолютные значения входящих в него показателей. Лучшие результаты получены в модели, учитывающей абсолютные значения дебита жидкости и пластового и забойного давления, т. е. комплексный коэффициент продуктивности является менее информативным критерием для модели.

С целью получения правил при принятии решений по выбору объектов для выполнения ОПЗ с учетом рассмотренных моделей использован алгоритм «дерево принятия решений» – это метод обучения с учителем (с англ. supervised machine learning algorithm), когда необходимо найти функциональную зависимость результатов (моделируемых свойств) от параметров объектов и построить алгоритм, принимающий на входе описание объекта и дающий на выходе ответ (свойство). В качестве входных данных выбраны параметры объектов, а результаты предсказанных значений успешности ОПЗ из моделей CatBoost приняты за значения успешности для построения дерева принятия решений. Использование предсказанных значений успешности позволяет сгладить выбросы, которые могут наблюдаться в исходных значениях. Для построения дерева принятия решений применяли алгоритм обучения CART, имплементированный из библиотеки scikit-learn [18, 19].

Для системы первичной ОПЗ вертикальных и горизонтальных скважин в качестве критерия выбора узла для расщепления в процессе роста классификационного дерева взят дебит жидкости, а в качестве числа признаков для выбора расщепления – все рассматриваемые параметры (рис. 1). Для данной системы потребовалось построение дерева принятия решений до глубины 5. При построении модели учитывали все параметры, т. е. правила по критерию успешности.

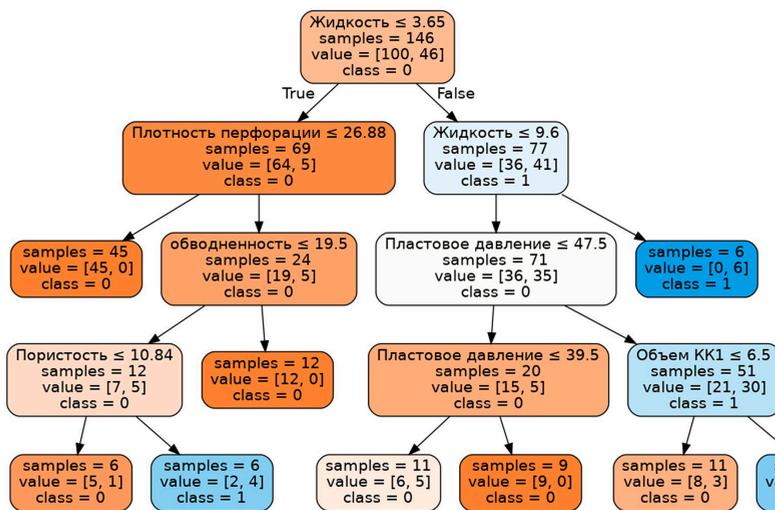


Рис. 1. Иерархическая древовидная структура, состоящая из решающих правил для системы первичной обработки вертикальных и горизонтальных скважин

Табл. 2

Результаты предсказательной способности классификационных моделей

Наборы общих дескрипторов	Параметры		Значения, отражающие сбалансированность и точность модели													
			TN	FP	FN	TP	Acc	TPR	TNR	BA	PPV	PNV	AUC	F_1	F_0.5	F_2
1	2		3	4	5	6	7	8	9	10	11	12	13	14	15	16
Первичная обработка вертикальных скважин																
С учетом эксплуатационных характеристик и химических параметров	Все 27 (факторов) дескрипторов		58	8	4	13	0.855	0.765	0.879	0.822	0.619	0.936	0.868	0.684	0.644	0.730
	Без коэффициента продуктивности		57	9	3	14	0.855	0.824	0.864	0.844	0.609	0.950	0.909	0.700	0.642	0.769
	Без дебита нефти забойного и пластового давления		57	9	7	10	0.807	0.588	0.864	0.726	0.526	0.891	0.823	0.556	0.538	0.575
Без учета эксплуатационных характеристик, но с учетом химических параметров	Все 27 (факторов) дескрипторов		61	5	5	12	0.880	0.706	0.924	0.815	0.706	0.924	0.854	0.706	0.706	0.706
	Без коэффициента продуктивности		59	7	5	12	0.855	0.706	0.894	0.800	0.632	0.922	0.824	0.667	0.645	0.690
	Без дебита нефти забойного и пластового давления		53	13	6	11	0.771	0.647	0.803	0.725	0.458	0.898	0.777	0.537	0.487	0.598
С учетом эксплуатационных характеристик, без учета химии и объемов химии	Все 27 (факторов) дескрипторов		52	14	4	13	0.783	0.765	0.788	0.776	0.482	0.929	0.876	0.591	0.520	0.684
	Без коэффициента продуктивности		55	11	2	15	0.843	0.882	0.833	0.858	0.577	0.965	0.898	0.698	0.620	0.798
	Без дебита нефти забойного и пластового давления		50	16	6	11	0.735	0.647	0.758	0.702	0.407	0.893	0.801	0.500	0.440	0.579
С учетом эксплуатационных характеристик и объемов химии, без учета химии	Все 27 (факторов) дескрипторов		53	13	3	14	0.807	0.824	0.803	0.813	0.519	0.946	0.896	0.636	0.560	0.737
	Без коэффициента продуктивности		55	11	3	14	0.831	0.824	0.833	0.828	0.560	0.948	0.897	0.667	0.598	0.753
	Без дебита нефти забойного и пластового давления		51	15	5	12	0.759	0.706	0.773	0.739	0.444	0.911	0.811	0.545	0.480	0.632

Продолжение табл. 2

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Первичная обработка вертикальных и горизонтальных скважин														
С учетом эксплуатационных характеристик и химических параметров	Все 27 (факторов) дескрипторов	74	26	13	33	0.733	0.717	0.740	0.729	0.559	0.851	0.769	0.629	0.585	0.679
	Без коэффициента продуктивности	71	29	16	30	0.692	0.652	0.710	0.681	0.509	0.816	0.764	0.571	0.532	0.617
	Без дебита нефти забойного и пластового давления	75	25	17	29	0.712	0.630	0.750	0.690	0.537	0.815	0.732	0.580	0.553	0.609
Без учета эксплуатационных характеристик, но с учетом химических параметров	Все 27 (факторов) дескрипторов	68	32	18	28	0.658	0.609	0.680	0.644	0.467	0.791	0.740	0.528	0.490	0.574
	Без коэффициента продуктивности	66	34	18	28	0.644	0.609	0.660	0.634	0.452	0.786	0.722	0.519	0.476	0.569
	Без дебита нефти забойного и пластового давления	71	29	14	32	0.706	0.696	0.710	0.703	0.525	0.835	0.733	0.598	0.552	0.653
С учетом эксплуатационных характеристик, без учета химии и объемов химии	Все 27 (факторов) дескрипторов	73	27	14	32	0.719	0.696	0.730	0.713	0.542	0.839	0.782	0.610	0.567	0.658
	Без коэффициента продуктивности	70	30	13	33	0.706	0.717	0.700	0.709	0.524	0.843	0.786	0.606	0.554	0.668
	Без дебита нефти забойного и пластового давления	74	26	13	33	0.733	0.717	0.740	0.729	0.559	0.851	0.749	0.629	0.585	0.679
С учетом эксплуатационных характеристик и объемов химии, без учета химии	Все 27 (факторов) дескрипторов	71	29	13	33	0.712	0.717	0.710	0.714	0.532	0.845	0.787	0.611	0.561	0.671
	Без коэффициента продуктивности	73	27	13	33	0.726	0.717	0.730	0.724	0.550	0.849	0.794	0.623	0.577	0.676
	Без дебита нефти забойного и пластового давления	75	25	12	34	0.747	0.739	0.750	0.745	0.576	0.862	0.764	0.648	0.603	0.700

Примечания: TN – истинно отрицательные случаи прогноза; FR – ложноположительные случаи прогноза; FN – ложноотрицательные случаи прогноза; TP – истинно положительные случаи прогноза; Ass – площадь под кривой; TPR – истинно положительный показатель; TNR – ложноположительный показатель; BA – сбалансированная точность; PPV – прогностическая значимость положительной классификации; PNV – доля ложных пропусков; AUC – качество работы классификатора; F\_1 – среднее гармоническое точности и полноты; F\_0.5 – F-мера с приоритетом точности; F\_2 – F-мера с приоритетом полноты. Изменение цвета от зеленого к красному отражает снижение сбалансированной точности предсказания.

### 3. Прогноз результатов лабораторных исследований с использованием методов машинного обучения

В качестве исходных данных использованы результаты лабораторного изучения константы скорости реакции на основе 114 уникальных параметров, включающих результаты рентгенофазного анализа пород, SARA-анализа насыщающих флюидов (асфальтены, смолы и парафины) и содержание серы.

Задача по изучению предсказательной способности моделей для прогнозирования скорости кинетических реакций при известных составах слагающих породу минералов и содержаниях асфальтенов, смол, парафинов, серы и кислотных составов в насыщающей нефти рассматривается как задача регрессии. Для этого использованы линейные и нелинейные методы машинного обучения, такие как линейная регрессия и метод случайного леса.

На первом этапе из общего набора дескрипторов были отобраны дескрипторы для построения модели. Затем проводили усреднение дескрипторов, масштабирование и удаление дубликатов. На третьем этапе осуществляли построение и обучение модели с использованием пятикратной перекрестной валидации с десятью повторениями. На заключительном этапе собирали и анализировали статистические данные по всем моделям (всего было построено более 7000 моделей).

Для нахождения линейной зависимости использовали основанный на методе наименьших квадратов метод LinearModel, имплементированный в библиотеке scikit-learn [19]. Были построены модели на малом наборе дескрипторов, и лучшие модели затем использованы для предсказания константы скорости реакции.

Результаты предсказательной способности наилучшей регрессионной модели представлены на рис. 2, а. Значения коэффициента детерминации ( $R^2$ ) и среднеквадратичной ошибки RMSE равны 0.6 и 0.18 соответственно.

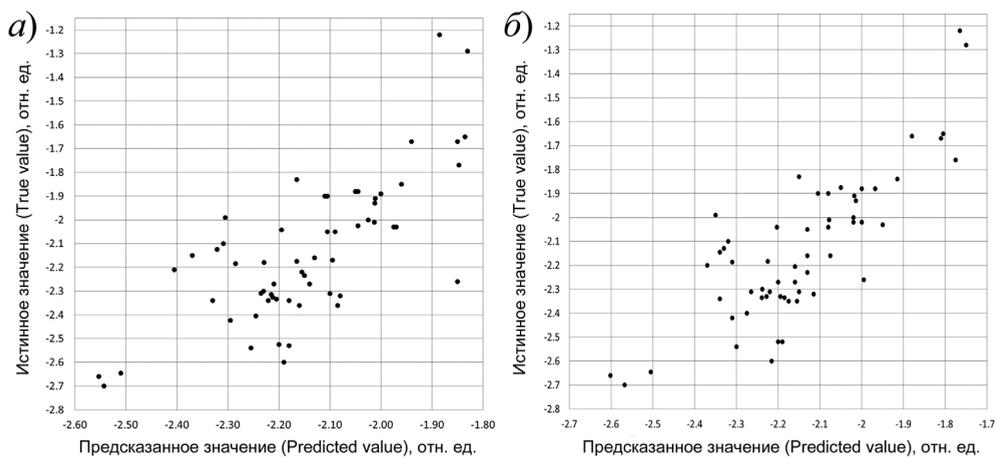


Рис. 2. Качество регрессионной модели, построенной на малом наборе дескрипторов: а) метод линейной регрессии; б) метод Random Forest

Для нахождения нелинейной зависимости использовали метод случайного леса (Random Forest, RF), имплементированный в библиотеку scikit-learn [19]. Количество деревьев установлено на уровне 100, настраиваемый гиперпараметр

модели –  $\max\_features(\log 2)$ . В целях получения надежной оценки предсказательной способности модели и во избежание переобучения использована процедура вложенной перекрестной валидации. Выбор гиперпараметра  $\max\_features$  модели RF осуществляли таким образом, чтобы обеспечить максимальную предсказательную способность моделей во внутреннем скользящем перекрестном контроле на основе RMSE. Лучшие модели, построенные на малом наборе дескрипторов, использованы для предсказания константы скорости реакции (рис. 2, б). Значения коэффициента детерминации ( $R^2$ ) и среднеквадратичной ошибки (RMSE) равны 0.59 и 0.19 соответственно.

Исходя из этих результатов, можно сделать вывод, что полученные регрессионные модели обладают средней предсказательной способностью.

Для определения вклада каждого признака, описывающего объект, в предсказанное значение свойства этого объекта был проведен SHAP-анализ. Ввиду более дифференцированного влияния факторов в модели случайного леса для дальнейшего предсказания применяли этот метод.

Анализ значений SHAP позволяет констатировать, что наибольшее влияние на константу скорости реакции оказывают следующие факторы (по степени убывания):

- типы коммерческих кислотных составов, использованных для ОПЗ;
- содержание в составе карбонатных пород примесей кварца и полевого шпата;
- содержание в насыщающих нефтях серы, смол, асфальтенов и парафинов, причем влияние уменьшается в представленном ряду от серы к парафинам.

Таким образом, установлено, что успешная предсказательная модель основывается на методе случайного леса и использует такие признаки, как состав нефти, наличие кварца и полевого шпата, а также принадлежность нефти к эксплуатационному горизонту.

Для интерпретации лучшего набора признаков, выбранного в методе случайного леса, был использован метод дерева принятия решений. В данном случае масштабирование данных не проводили. Визуализация сложных деревьев принятия решений в виде решающих правил вместо иерархической структуры из узлов и листьев может оказаться более удобной для восприятия (рис. 3). Следует отметить, что чем больше значение константы скорости реакции, тем ближе цвет вершины к темно-оранжевому в представленной визуализации, и наоборот, чем меньше, тем цвет ближе к белому.

### Заключение

Анализ результатов применения методов машинного обучения показывает, что на сегодняшний день резерв повышения эффективности технологий ОПЗ не исчерпан. Прогноз результатов выполнения обработки и прогноз исходных данных, получаемых в лабораторных условиях на основе сформированной по результатам экспериментов базы данных, позволяет через систему правил значительно упростить задачи, решаемые при планировании и проектировании работ по ОПЗ.

Наполнение базы данных, используемой при создании прогнозных моделей, позволяет значительно повысить адекватность полученных моделей. По результатам выполненных работ получены два решения, позволяющие использовать полученные модели прогноза показателей.

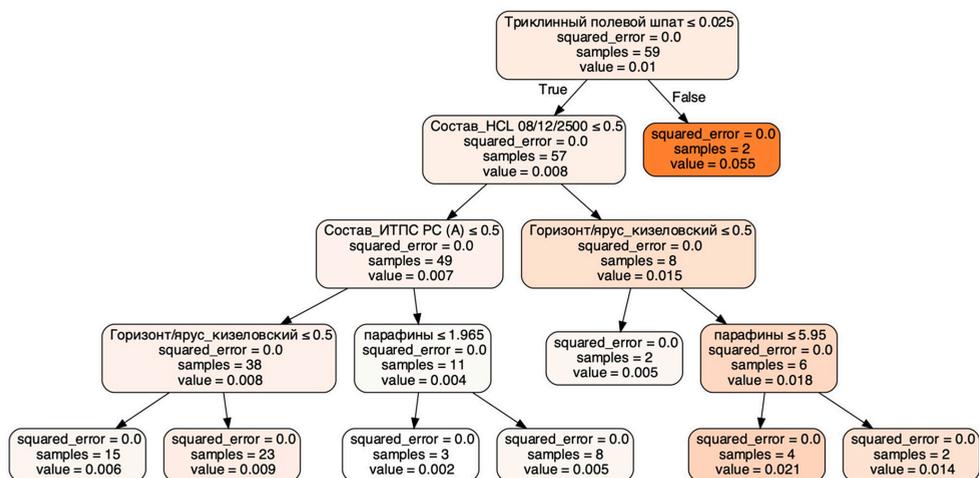


Рис. 3. Иерархическая древовидная структура, состоящая из решающих правил для предсказания константы скорости реакции

Первая модель – это система подбора скважин для интенсификации добычи нефти, встроенная в технологическую базу мониторинга работы скважин АРМИТС (автоматизированные рабочие места инженерно-технологической службы) и позволяющая в режиме реального времени отслеживать потенциальный фонд для выполнения ОПЗ.

Вторая модель – система прогнозирования кинетики реакции кислотных композиций с породами дифференцированного минералогического состава, реализованная с помощью статистически значимой модели на основе случайного леса в виде интернет-ресурса [tatneft2023.cimm.site](http://tatneft2023.cimm.site). Прогнозная константа кинетики реакции используется в дизайне и проектировании ОПЗ.

В дополнение к существующим подходам к решению задачи по оптимизации ОПЗ, положения, изложенные в данной работе, позволяют значительно упростить систему прогнозирования и подготовки данных для проектирования и выполнения ОПЗ.

**Благодарности.** Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации по договору № 075-15-2022-299 в рамках программы развития Научного центра мирового уровня «Рациональное освоение запасов жидких углеводородов планеты».

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

### Литература

1. Akbar M., Vissapragada B., Alghamdi A.H., Allen D., Herron M., Carnegie A., Dutta D., Olesen J.R., Chourasiya R.D., Logan D., Stief D., Netherwood R., Russell S.D., Saxena, K. A snapshot of carbonate reservoir evaluation // *Oilfield Rev.* 2000. V. 12, No 4. P. 20–41.
2. Burchette T.P. Carbonate rocks and petroleum reservoirs: A geological perspective from the industry // Garland J., Neilson J.E., Laubach S.E., Whidden K.J. (Eds.) *Advances in Carbonate Exploration and Reservoir Analysis*. Ser.: GSL Special Publications. V. 370. Geol. Soc. London, 2012. P. 17–37. <https://doi.org/10.1144/SP370.14>.
3. Janjuhah H.T., Alansari A. Offshore carbonate facies characterization and reservoir quality of Miocene rocks in the southern margin of South China Sea // *Acta Geol. Sin.* 2020. V. 94, No 5. P. 1547–1561. <https://doi.org/10.1111/1755-6724.13880>.
4. Hall-Thompson B., Ernesto A.R., Abdulrahman N., Alsuhaime A. Acid stimulation-best practices for design, selection and testing of acid recipes in low permeability carbonate reservoirs // *Proc. Int. Pet. Technol. Conf. (January 13–15, 2020)*. Dhahran, 2020. Art. IPTC-19690-MS. <https://doi.org/10.2523/IPTC-19690-MS>.
5. Насибулин И.М. Обоснование технологии кислотной стимуляции карбонатных коллекторов на основе многофункциональных композиций с регулируемой реакционной способностью: автореф. дис. ... канд. техн. наук. СПб., 2013. 20 с.
6. Копытов А.Г., Левкович С.В., Левитина Е.Е., Ковалев И.А. Эффективность кислотной обработки призабойной зоны скважины для отложений с трудноизвлекаемыми запасами // *Наука. Инновации. Технологии*. 2021. № 4. С. 27–40. <https://doi.org/10.37493/2308-4758.2021.4.2>.
7. Sengul M., Remisio L.H.A. Applied carbonate stimulation – an engineering approach // *Proc. Abu Dhabi Int. Pet. Exhib. Conf. (October 13–16, 2002)*. Abu Dhabi, UAE, 2002. Art. SPE-78560-MS. <https://doi.org/10.2118/78560-MS>.
8. Бурнашев В.Ф., Хужаеров Б.Х. Моделирование кислотной обработки доломитового коллектора призабойной зоны нефтяного пласта с учетом кольматации породы // *Изв. РАН. МЖГ*. 2015. № 1. С. 78–85.
9. Economides M.J., Nolte K.G. (Eds.) *Reservoir Stimulation*. New York, NY: Prentice Hall, 1989. 440 p.
10. Liu X., Ormond A., Bartko K., Li Y., Ortoleva P. A geochemical reaction-transport simulator for matrix acidizing analysis and design // *J. Pet. Sci. Eng.* 1997. V. 17, Nos 1–2. P. 181–196. [https://doi.org/10.1016/S0920-4105\(96\)00064-2](https://doi.org/10.1016/S0920-4105(96)00064-2).
11. Новиков В.А. Методика прогнозирования эффективности матричных кислотных обработок карбонатов // *Недропользование*. 2021. Т. 21, № 3. С. 137–143. <https://doi.org/10.15593/2712-8008/2021.3.6>.
12. Галкин В.И., Хижняк Г.П., Амиров А.М., Гладких Е.А. Оценка эффективности воздействия кислотных составов на керны с использованием регрессионного анализа // *Вестник ПНИПУ. Геология. Нефтегазовое и горное дело*. 2014. № 13. С. 38–48. <https://doi.org/10.15593/2224-9923/2014.13.4>.
13. Fredd C.N. Dynamic model of wormhole formation demonstrates conditions for effective skin reduction during carbonate matrix acidizing // *Proc. SPE Permian Basin Oil Gas Recovery Conf. (March 21–23, 2000)*. Midland, TX, 2000. Art. SPE-59537-MS. <https://doi.org/10.2118/59537-MS>.
14. Safari A., Dowlatabad M.M., Hassani A., Rashidi F. Numerical simulation and X-ray imaging validation of wormhole propagation during acid core-flood experiments in a carbonate gas reservoir // *J. Nat. Gas Sci. Eng.* 2016. V. 30. P. 539–547. <https://doi.org/10.1016/j.jngse.2016.02.036>.

15. Курганов Д.В. Оценка эффективности обработок призабойных зон нефтяных скважин с применением методов машинного обучения // Автоматизация процессов управления. 2020. № 1 (59). С. 47–57. <https://doi.org/10.35752/1991-2927-2020-1-5-47-54>.
16. Zangl G., Hannerer J. Data Mining: Applications in the Petroleum Industry. East Nynehead: Round Oak Publ., 2003. 222 p.
17. Хадавимогоддам Ф., Мищенко И.Т., Мостаджеран М. Применение методов искусственного интеллекта в прогнозировании основных свойств нефти // Газовая промышленность. 2019. № 12 (794). С. 28–32.
18. Scikit-Learn User Guide. URL: [https://scikit-learn.org/stable/\\_downloads/scikit-learn-docs.pdf](https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf).
19. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python // J. Mach. Learn. Res. 2011. V. 12. P. 2825–2830.

Поступила в редакцию 27.04.2024

Принята к публикации 29.07.2024

---

**Маннанов Ильдар Илгизович**, кандидат технических наук, доцент кафедры разработки и эксплуатации месторождений трудноизвлекаемых углеводородов Института геологии и нефтегазовых технологий

Казанский (Приволжский) федеральный университет  
ул. Кремлевская, д. 18, г. Казань, 420008, Россия  
E-mail: [ildarmannanov@mail.ru](mailto:ildarmannanov@mail.ru)

**Варфоломеев Михаил Алексеевич**, кандидат химических наук, доцент, заведующий кафедрой разработки и эксплуатации месторождений трудноизвлекаемых углеводородов Института геологии и нефтегазовых технологий

Казанский (Приволжский) федеральный университет  
ул. Кремлевская, д. 18, г. Казань, 420008, Россия  
E-mail: [vma.ksu@gmail.com](mailto:vma.ksu@gmail.com)

**Ганиева Гузель Рафиковна**, кандидат технических наук, доцент кафедры разработки и эксплуатации месторождений трудноизвлекаемых углеводородов Института геологии и нефтегазовых технологий

Казанский (Приволжский) федеральный университет  
ул. Кремлевская, д. 18, г. Казань, 420008, Россия  
E-mail: [gguzelrafikovna@mail.ru](mailto:gguzelrafikovna@mail.ru)

**Гимаева Алина Рашитовна**, кандидат технических наук, доцент кафедры разработки и эксплуатации месторождений трудноизвлекаемых углеводородов Института геологии и нефтегазовых технологий

Казанский (Приволжский) федеральный университет  
ул. Кремлевская, д. 18, г. Казань, 420008, Россия  
E-mail: [argimaeva@yandex.ru](mailto:argimaeva@yandex.ru)

**Гиниятуллин Радель Рамилевич**, аспирант кафедры разработки и эксплуатации месторождений трудноизвлекаемых углеводородов Института геологии и нефтегазовых технологий

Казанский (Приволжский) федеральный университет  
ул. Кремлевская, д. 18, г. Казань, 420008, Россия  
E-mail: [radel0099@gmail.com](mailto:radel0099@gmail.com)

## ORIGINAL ARTICLE

doi: 10.26907/2542-064X.2024.4.623-639

**Application of Machine Learning for Target Selection and Acid Treatment Design***I.I. Mannanov*<sup>\*</sup>, *M.A. Varfolomeev*<sup>\*\*</sup>, *G.R. Ganieva*<sup>\*\*\*</sup>, *A.R. Gimaeva*<sup>\*\*\*\*</sup>,  
*R.R. Giniyatullin*<sup>\*\*\*\*\*</sup>*Kazan Federal University, Kazan, 420008 Russia*E-mail: *\*ildarmannanov@mail.ru*, *\*\*vma.ksu@gmail.com*, *\*\*\*gguzelrafikovna@mail.ru*,  
*\*\*\*\*argimaeva@yandex.ru*, *\*\*\*\*\*radel0099@gmail.com*

Received April 27, 2024; Accepted July 29, 2024

**Abstract**

Acid treatment is commonly used to enhance the production capacity of wells drilled in carbonate deposits. However, field outcomes of this procedure may vary significantly. Current approaches to acid treatment design rely on advanced software tools that evaluate major acidizing factors. Machine learning is a valuable complement to the existing techniques: it facilitates the selection of target wells and aids in defining initial parameters for design engineering on reliable and effective software platforms. This study examines potential applications of machine learning in target selection based on the history of treatment outcomes influenced by the initial well conditions, operational conditions, treatment frequency, acid volumes, acid system types, pretreatment strategies, acid system diverters, and acid residence time. Acid treatment design requires complex laboratory work to investigate the kinetics of acid-rock interactions determined by the mineral composition of the rock formation and the chemical properties of the acid system, including the concentrations of its components. The problem of predicting the reaction kinetics of acid systems by processing an array of laboratory data using machine learning methods, specifically linear regression and random forest methods, was discussed. It was demonstrated that the incorporation of machine learning enables the development of robust decision-making algorithms that optimize acid treatment by considering its multifactorial effects. These algorithms significantly simplify the tasks of acid treatment design.

**Keywords:** vibrational spectra, molecular structure, electronography, quantum chemical calculations

**Acknowledgments.** This study was supported by the Ministry of Science and Higher Education of the Russian Federation under agreement no. 075-15-2022-299 within the framework of the development program for a world-class research center “Efficient development of the global liquid hydrocarbon reserves.”

**Conflicts of Interest.** The authors declare no conflicts of interest.

**Figure Captions**

Fig. 1. Hierarchical tree structure of decision rules for the primary processing system of vertical and horizontal wells.

Fig. 2. Performance of the regression model using a small set of descriptors: *a)* linear regression; *b)* random forest.

Fig. 3. Hierarchical tree structure of decision rules for predicting the reaction rate constant.

**References**

1. Akbar M., Vissapragada B., Alghamdi A.H., Allen D., Herron M., Carnegie A., Dutta D., Olesen J.R., Chourasiya R.D., Logan D., Stief D., Netherwood R., Russell S.D., Saxena, K. A snapshot of carbonate reservoir evaluation. *Oilfield Rev.*, 2000, vol. 12, no. 4, pp. 20–41.
2. Burchette T.P. Carbonate rocks and petroleum reservoirs: A geological perspective from the industry. In: Garland J., Neilson J.E., Laubach S.E., Whidden K.J. (Eds.) *Advances in Carbonate Exploration*

- and Reservoir Analysis*. Ser.: GSL Special Publications. Vol. 370. Geol. Soc. London, 2012. pp. 17–37. <https://doi.org/10.1144/SP370.14>.
3. Janjuhah H.T., Alansari A. Offshore carbonate facies characterization and reservoir quality of Miocene rocks in the southern margin of South China Sea. *Acta Geol. Sin.*, 2020, vol. 94, no. 5, pp. 1547–1561. <https://doi.org/10.1111/1755-6724.13880>.
  4. Hall-Thompson B., Ernesto A.R., Abdulrahman N., Alsuhaimeh A. Acid stimulation-best practices for design, selection and testing of acid recipes in low permeability carbonate reservoirs. *Proc. Int. Pet. Technol. Conf. (January 13–15, 2020)*. Dhahran, 2020, art. IPTC-19690-MS. <https://doi.org/10.2523/IPTC-19690-MS>.
  5. Nasibulin I.M. Justification of acid stimulation technique for carbonate reservoirs using multifunctional compositions with controlled reactivity. *Extended Abstract of Cand. Tech. Sci. St. Petersburg*, 2013. 20 p. (In Russian)
  6. Kopytov A.G., Levkovich S.V., Levitina E.E., Kovalev I.A. Efficiency of acid treatment of the bottom borehole zone of the well for sediments with hard-to-remove reserves. *Nauka. Innovatsii. Tekhnol.*, 2021, no. 4, pp. 27–40. <https://doi.org/10.37493/2308-4758.2021.4.2>. (In Russian)
  7. Sengul M., Remisio L.H.A. Applied carbonate stimulation – an engineering approach. *Proc. Abu Dhabi Int. Pet. Exhib. Conf. (October 13–16, 2002)*. Abu Dhabi, UAE, 2002, art. SPE-78560-MS. <https://doi.org/10.2118/78560-MS>.
  8. Burnashev V.F., Khuzhaerov B.K. Modeling the acid treatment of the dolomitic collector of an oil formation bottom-hole zone with account of rock colmatation. *Fluid Dyn.*, 2015, vol. 50, no. 1, pp. 71–78. <https://doi.org/10.1134/S0015462815010081>.
  9. Economides M.J., Nolte K.G. (Eds.) *Reservoir Stimulation*. New York, NY, Prentice Hall, 1989. 440 p.
  10. Liu X., Ormond A., Bartko K., Li Y., Ortoleva P. A geochemical reaction-transport simulator for matrix acidizing analysis and design. *J. Pet. Sci. Eng.*, 1997, vol. 17, nos. 1–2, pp. 181–196. [https://doi.org/10.1016/S0920-4105\(96\)00064-2](https://doi.org/10.1016/S0920-4105(96)00064-2).
  11. Novikov V.A. The method of predicting efficiency of the matrix acid treatment of carbonates. *Perm J. Pet. Min. Eng.*, 2021, vol. 21, no. 3, pp. 137–143. <https://doi.org/10.15593/2712-8008/2021.3.6>.
  12. Galkin V.I., Khizhnyak G.P., Amirov A.M., Gladkikh E.A. Assessing the efficiency of core sample acidizing using regression analysis. *Vestn. PNIPU. Geol. Neftgazov. Gorn. Delo*, 2014, no. 13, pp. 38–48. <https://doi.org/10.15593/2224-9923/2014.13.4>. (In Russian)
  13. Fredd C.N. Dynamic model of wormhole formation demonstrates conditions for effective skin reduction during carbonate matrix acidizing. *Proc. SPE Permian Basin Oil Gas Recovery Conf. (March 21–23, 2000)*. Midland, TX, 2000, art. SPE-59537-MS. <https://doi.org/10.2118/59537-MS>.
  14. Safari A., Dowlatabad M.M., Hassani A., Rashidi F. Numerical simulation and X-ray imaging validation of wormhole propagation during acid core-flood experiments in a carbonate gas reservoir. *J. Nat. Gas Sci. Eng.*, 2016, vol. 30, pp. 539–547. <https://doi.org/10.1016/j.jngse.2016.02.036>.
  15. Kurganov D.V. Assessing the effectiveness of bottomhole treatment in oil wells using machine learning techniques. *Avtom. Protsessov Upr.*, 2020, no. 1 (59), pp. 47–57. <https://doi.org/10.35752/1991-2927-2020-1-5-47-54>. (In Russian)
  16. Zangl G., Hannerer J. *Data Mining: Applications in the Petroleum Industry*. East Nynhehead, Round Oak Publ., 2003. 222 p.
  17. Hadavimoghaddam F., Mishchenko I.T., Mostajeran M. Using artificial intelligence methods in forecasting the key oil properties. *Gazov. Prom-st.*, 2019, no. 12 (794), pp. 28–32. (In Russian)
  18. Scikit-Learn User Guide. URL: [https://scikit-learn.org/stable/\\_downloads/scikit-learn-docs.pdf](https://scikit-learn.org/stable/_downloads/scikit-learn-docs.pdf).
  19. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 2011, vol. 12, no. 85, pp. 2825–2830.

Для цитирования: Маннанов И.И., Варфоломеев М.А., Ганиева Г.Р., Гимаева А.Р., Гиниятуллин Р.Р. Применение машинного обучения в задачах выбора объектов воздействия и проектирования кислотной обработки // Учен. зап. Казан. ун-та. Сер. Естеств. науки. 2024. Т. 166, кн. 4. С. 623–639. <https://doi.org/10.26907/2542-064X.2024.4.623-639>.

For citation: Mannanov I.I., Varfolomeev M.A., Ganieva G.R., Gimaeva A.R., Giniyatullin R.R. Application of machine learning for target selection and acid treatment design. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Estestvennye Nauki*, 2024, vol. 166, no. 4, pp. 623–639. <https://doi.org/10.26907/2542-064X.2024.4.623-639>. (In Russian)