

УДК 622.32(004.85)  
DOI: 10.18799/24131830/2024/5/4553  
Шифр специальности ВАК: 1.6.11; 2.3.1

## Прогнозирование дебита горизонтальных скважин с применением модели машинного обучения

С.А. Пискунов✉, Ш. Давуди

Национальный исследовательский Томский политехнический университет, Россия, г. Томск

✉piskynovsa@hw.tpu.ru

**Аннотация.** *Актуальность* исследования связана с необходимостью точно и быстро прогнозировать дебиты горизонтальных скважин. Это позволит оптимизировать составление графиков бурения, программ увеличения нефтеотдачи, стратегии разработки месторождения, а также сделать экономическую модель более точной и предсказуемой. В настоящее время для прогнозирования дебитов скважин используются аналитические расчеты и численные методы моделирования. Данные методы имеют ограничения как по точности, так и по времени. Для решений данной задачи предлагается использовать машинное обучение, которое за счет своей точности, адаптивности и скорости позволяет исключить недостатки ранее перечисленных методов. **Цель:** создание модели машинного обучения для количественной оценки дебита газовых скважин на основе геологических свойств на различных временных шагах. **Объект:** фонд горизонтальных скважин газоконденсатного месторождения Западной Сибири. **Методы:** математическое моделирование, машинное обучение и статистические методы. **Результаты.** Проведено 300 итераций гидродинамического моделирования в симуляторе. Собран исходный набор данных со следующими параметрами: временной шаг, пористость, проницаемость, исходная водонасыщенность, толщина пласта, давление в зоне забоя на различных расстояниях от ствола скважины, дебит газа. Созданы модели машинного обучения на основе алгоритмов случайного леса и градиентного бустинга с различными соотношениями тестовой выборки к тренировочной. Модели машинного обучения позволяли точно прогнозировать дебит газа горизонтальной скважины. Градиентный бустинг показал лучшие результаты прогнозирования по сравнению со случайным лесом: квадрат средней квадратичной ошибки равен 8440 ст. м<sup>3</sup>/сут., средняя абсолютная процентная ошибка равна 3,95 %, коэффициент детерминации ( $R^2$ )=0,991.

**Ключевые слова:** машинное обучение, градиентный бустинг, случайный лес, прогнозирования дебита горизонтальной скважины, закон Дарси, гидродинамическое моделирование пласта

**Для цитирования:** Пискунов С.А., Давуди Ш. Прогнозирование дебита горизонтальных скважин с применением модели машинного обучения // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2024. – Т. 335. – № 5. – С. 118–130. DOI: 10.18799/24131830/2024/5/4553

---

UDC 622.32(004.85)  
DOI: 10.18799/24131830/2024/5/4553

## Horizontal well flow rate prediction applying machine-learning model

S.A. Piskunov✉, Sh. Davoodi

National Research Tomsk Polytechnic University, Tomsk, Russian Federation

✉piskynovsa@hw.tpu.ru

**Abstract. Relevance.** The need to accurately and quickly predict flow rates of horizontal wells. This allows optimizing drilling schedules, enhanced oil recovery programs, and field development strategy, as well as making the economic model more accurate and predictable. Currently, analytical calculations and numerical modeling methods are used to predict well production rates. These methods have limitations in both accuracy and time. To solve this problem, it is proposed to use machine learning, which due to its accuracy, adaptability, and speed, allows excluding the disadvantages of the above-mentioned methods. **Aim.** To create a machine-learning model to quantify gas well flow rates based on geological properties at different

time steps. **Object.** Stock of horizontal wells in a gas condensate field in Western Siberia. **Methods.** Mathematical modelling, machine learning and statistical methods. **Results.** The authors have carried out 300 iterations of hydrodynamic modeling in a simulator. They collected an initial data set with the following parameters: time step, porosity, permeability, initial water saturation, reservoir thickness, bottom hole pressure at different distances from the wellbore, and gas flow rate. Machine learning models based on random forest and gradient boosting algorithms were created with different ratios of testing/training samples. The machine learning models were able to accurately predict the gas flow rate of a horizontal well. Gradient boosting showed better prediction results compared to random forest: root mean square error is equal to 8440 std. m<sup>3</sup>/day, mean absolute percentage error is equal to 3.95%, and coefficient of determination ( $R^2$ )=0.991.

**Keywords:** machine learning, gradient boosting, random forest, horizontal well flow rate prediction, Darcy's law, reservoir simulation

**For citation:** Piskunov S.A., Davoodi Sh. Horizontal well flow rate prediction applying machine-learning model. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2024, vol. 335, no. 5, pp. 118–130. DOI: 10.18799/24131830/2024/5/4553

## Введение

Важным условием успешной разработки месторождений является прогнозирование параметров работы скважин. Точный прогноз добычи газа необходим для составления планов бурения, программ повышения нефтегазоотдачи и стратегий разработки пластов. Знание этих параметров значительно повышает шансы на успешную и эффективную добычу. Кроме того, это помогает сделать экономическую модель более точной и предсказуемой.

Наиболее распространенным в мире подходом к определению дебита скважин является закон Дарси. Инженеры нефтегазовой сферы обычно используют уравнение Дюпюи [1]. Оно представляет собой интегральную форму закона Дарси для случая плоскорадиального установившегося притока несжимаемой жидкости в вертикальную скважину. Уравнение Дюпюи связывает продуктивные характеристики скважины и фильтрационные свойства пласта. Согласно этому уравнению, основными факторами, влияющими на приток жидкости, являются: проницаемость, толщина пласта, скин-фактор, радиус скважины, забойное давление и пластовое давление. Также необходимо учитывать конструкцию скважины [2]. Оценивать все перечисленные факторы в процессе разработки месторождения становится сложной задачей, так как происходит постоянное изменение энергетики пласта (давление, аквифер), ухудшение призабойной зоны (скин-фактор), уменьшение фазовой проницаемости в связи с заводнением околоскважинной зоны, свойств пластовых флюидов (Pressure-Volume-Temperature – PVT), неоднородность коллектора в целом.

Существуют различные способы прогнозирования добычи скважин. Но они имеют разную точность и требуют разное количество времени. Аналитический подход является наиболее быстрым, но, с другой стороны, он имеет самый низкий уровень точности и ряд допущений [3]. Эмпирический подход помогает прогнозировать дальнейший дебит скважины, основываясь на стартовых значениях

дебита в скважине и анализе тенденции дальнейшего падения дебита. В частности, этот метод называется кривой падения добычи [4]. Аналитические методы прогнозируют поведение скважины при помощи установления закономерности между физическими свойствами в пласте. В основе расчетов материального баланса лежит закон сохранения массы. Материальный баланс – это учет поступления и извлечения массы из системы [5, 6]. Применение материального баланса может быть затруднительным, особенно в системах с давлением ниже давления насыщения, а также в присутствии других источников энергии (газовая шапка, аквифер) [7].

Метод использования симуляторов является более точным, чем аналитический, но в то же время требует больших временных затрат и большого количества информации о месторождении, что вносит свою неопределенность в итоговое прогнозирование. Суть метода заключается в использовании программных комплексов, которые позволяют, используя закон материального баланса и закона Дарси, итерационным путем рассчитывать показатели работы скважины. Данный подход позволяет заменить сложные аналитические формулы численными расчетами: решение на основе упрощения (аппроксимации) более явными зависимостями [8].

Симуляторы являются наиболее точным способом прогнозирования работы скважин и оценивания профиля добычи (P10, P50, P90). Они позволяют оценить неопределенность и риски дальнейшей разработки. Данные дебиты строятся на основе итерации различных вариантов геологии. Этот подход называется методом Монте Карло [9]. Но в то же время для создания данных профилей необходимо множество расчетов и большое количество времени на их осуществление. Для обеспечения точности симулятора и скорости математических расчетов предлагается использовать машинное обучение (МО) [10].

В последние годы предпринимались попытки применить алгоритмы МО и искусственный интеллект (ИИ) для прогнозирования работы скважин.

Например, в бурении [11, 12], геофизике [13, 14], инжиниринге резервуаров [15, 16]. Этот метод не только прост в реализации, но и позволяет отразить сложную взаимосвязь между входными и выходными наборами данных. В частности, он основан на анализе данных, взаимодействии и корреляции параметров системы, а не на физических процессах. В частности, популярность приобрели ИНС (искусственная нейронная сеть) и метод опорных векторов. Модель, основанная на ИНС [17], показала среднюю абсолютную ошибку в 7,9 % при прогнозировании дебита. В данной статье оценивался средний дебит после запуска горизонтальных скважин определенной длины и был получен довольно точный результат. Однако применялся только один геологический параметр – проницаемость, что не позволяет полноценно оценить возможную добычу в дальнейшем [18]. Также подобный метод применялся для прогнозирования добычи горизонтальных скважин на основе данных геологии, анизотропии, размеров скважин, и ошибка составила 7 % [10]. Но данная модель не позволяет ввести прогноз на определенный временной шаг, а дает лишь представление об ожидаемом дебите скважины. Также нейронные сети применяли на месторождении «Volve» в Норвегии [19]. В работе применяли нейронные сети прямого распространения, рекуррентную нейронную сеть, а также использовался метод роя частиц в качестве оптимизатора. Коэффициент детерминации превысил показать 94 % [19]. В данной работе в основе лежат входные параметры, основанные на забойном давлении и размере штуцера, что не позволяет вовлечь в оценку геологию.

В этой работе были разработаны две модели машинного обучения – градиентный бустинг (ГБ) и случайный лес – для прогнозирования дебита газа горизонтальных скважин. Использование петрофизической информации позволит сделать данную модель более универсальной, физичной. Также будет учитываться временной шаг, что позволит не только сделать прогноз стартового показателя дебита, но и оценить его дальнейшее изменение. Для поиска оптимальных гиперпараметров применялся половинчатый поиск по сетке (Halving Search) с использованием перекрестной проверки. Набор данных включает в себя такие геологические параметры, как: проницаемость пласта, эффективная толщина пласта, пористость, водонасыщенность вдоль ствола скважины, а также данные по падению давления на забое скважины и вблизи ствола. Часть данных используется для разработки модели (тренировочная выборка), а оставшаяся часть – в качестве проверки прогнозной способности модели (тестовая выборка). Лучший алгоритм был выбран путем сравнения поведения на тестовых и тренировочных данных.

В статье даются краткие пояснения по теории машинного обучения, ГБ, случайного леса и половинчатого поиска по сетке. Также рассматривается методология и объясняется, как имеющиеся данные были получены, обработаны и использованы при разработке этих моделей. Описывается процедура разработки модели. Затем приводятся результаты, следует их обсуждение и составляются выводы, которые обобщают полученные результаты. В заключении подводятся основные итоги работы.

### Получение и анализ данных

В работе использовались данные, сгенерированные через коммерческий гидродинамический симулятор. Было проведено 300 итераций расчета. На основе геологических неопределенностей (стандартного отклонения и среднего значения) сгенерированы различные варианты петрофизики, а также сняты средние значения входных параметров вдоль стволов скважины. Этот подход позволяет создать достаточное количество данных для обучения модели машинного обучения и дальнейшего прогнозирования дебитов на исследуемом месторождении без применения симуляторов. Система разработки представляет лучевую систему с горизонтальными скважинами, коллектор – одиночный пласт с вертикальной анизотропией проницаемости.

Набор данных включает следующие параметры: временной шаг, пористость, проницаемость, исходную водонасыщенность, толщину пласта, давление в зоне забоя на различных расстояниях от ствола скважины, дебит газа. Данные были случайно разделены на две выборки (тестовая и тренировочная). Соотношения тренировочной выборки к тестовой варьировались и принимали разные значения (80/20, 75/25, 70/30) для дальнейшего анализа и выбора лучшего варианта.

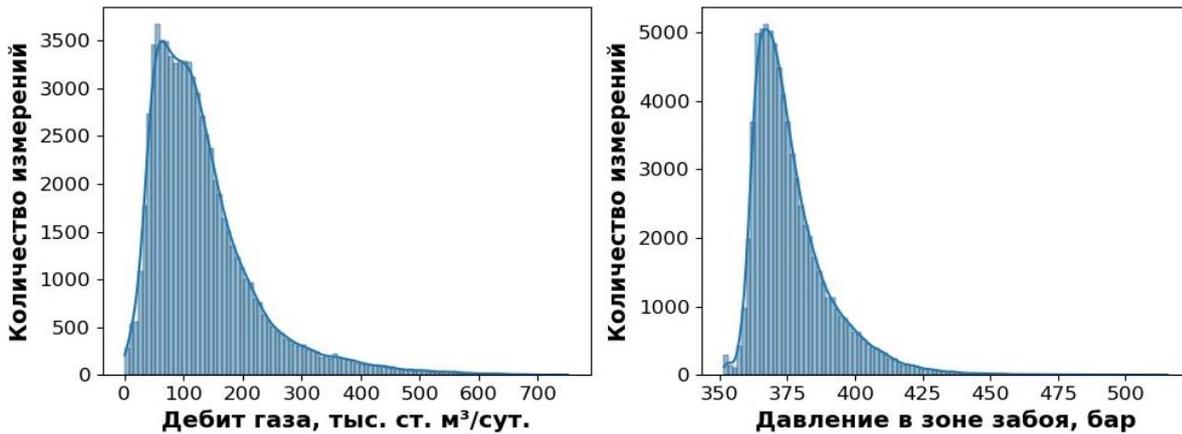
В качестве параметров, характеризующих распределение исходных данных, были выбраны их значения на соответствующих процентилях (25, 50, 75 %), средние значения параметров, их стандартное отклонение и предельное значение в распределении. Параметры представлены в табл. 1.

Распределение давления и дебита газа представлено на рис. 1. Статистика оценена с помощью межквартильного размаха и стандартного отклонения. Распределение дебита показывает большой разбег между максимальным и минимальным значениями. Отношение максимального значения дебита к минимальному составляет 5,65. Процентиль 75 % и максимальное значение различаются в разы. Распределения дебита газа и давления в зоне забоя представляют собой ассиметричные распределения со смещением влево. Коэффициент вариации ( $\text{std}/\text{mean}$ ) равняется 0,68, что говорит о низкой дисперсии распределения.

**Таблица 1.** Статистический анализ входных/выходных данных, используемых в исследовании

**Table 1.** Statistical analysis of the input/output data given in the study

Параметр Parameter	Среднее значение Mean value	Стандартное отклонение Standard deviation	25 %	50 %	75 %	Макс. Max.
Временной шаг, месяц Time step, month	6,5	3,5	3	6	9	12
Пористость, % Porosity, %	0,14	0,01	0,13	0,14	0,14	0,16
Проницаемость, Дарси Permeability, Darcy	0,08	0,03	0,06	0,05	0,1	0,21
Водонасыщенность Water saturation	0,35	0,31	0,03	0,22	0,69	0,88
Эффективная толщина, м Effective thickness, m	4,83	4,55	0,49	2,92	9,40	15,83
Давление вдоль ствола скважины, бар. Pressure along the wellbore, bar	376,7	19,1	366,7	372,9	382,5	515,5
Давление зоны 4 ячеек, бар 4 cells area pressure, bar	401,2	23,8	387,3	397,2	411,4	517,4
Давление зоны 9 ячеек, бар 9 cells area pressure, bar	416,6	14,7	398,7	411,6	430,2	520,1
Дебит газа, м <sup>3</sup> /сут. Gas production rate, m <sup>3</sup> /day	132721	90423	877	70157	111858	166494



**Рис. 1.** Распределение дебита газа и давления в зоне забоя

**Fig. 1.** Distribution of gas flow rate and pressure along the wellbore

Распределение давления коррелирует с распределением дебита газа, что является очевидным, так как данные величины связаны напрямую, согласно закону Дарси [1]. Давление на различных расстояниях от ствола скважины варьируется от 351,69 до 520 бар (соответствует пластовому давлению).

Для наглядной оценки важности каждого компонента для прогнозирования дебита газа была составлена табл. 2. Анализ коэффициентов корреляции показывает, что все признаки взаимосвязаны с дебитом газа для горизонтальной скважины. Наибольший коэффициент корреляции (по модулю) имеет временной шаг – 0,566. Наихудшая корреляция у дебита с эффективной толщиной – –0,197.

Значение корреляции параметра эффективной толщины с целевой переменной может быть отрицательным, хотя он положительно влияет на целевую переменную. Данное явление связано с тем, что другие параметры (временной шаг, петрофизи-

ка и давление) вносили более значимый вклад в результирующий дебит газа. В дополнение было получено распределение для данных геологии с целью оценки характера плотности распределения входных параметров (рис. 2).

**Таблица 2.** Анализ коэффициентов корреляции с выходным параметром (дебит газа)

**Table 2.** Analysis of correlation coefficients with the output parameter (gas flow rate)

Временной шаг, месяц/Time step, month	-0,566
Пористость/Porosity, %	0,353
Проницаемость, Дарси/Permeability, Darcy	0,375
Водонасыщенность/Water saturation	-0,248
Эффективная толщина, м/Effective thickness, m	-0,197
Давление вдоль ствола скважины, бар. Pressure along the wellbore, bar	0,311
Давление зоны 4 ячеек, бар./4 cells area pressure, bar	0,298
Давление зоны 9 ячеек, бар./9 cells area pressure, bar	0,299

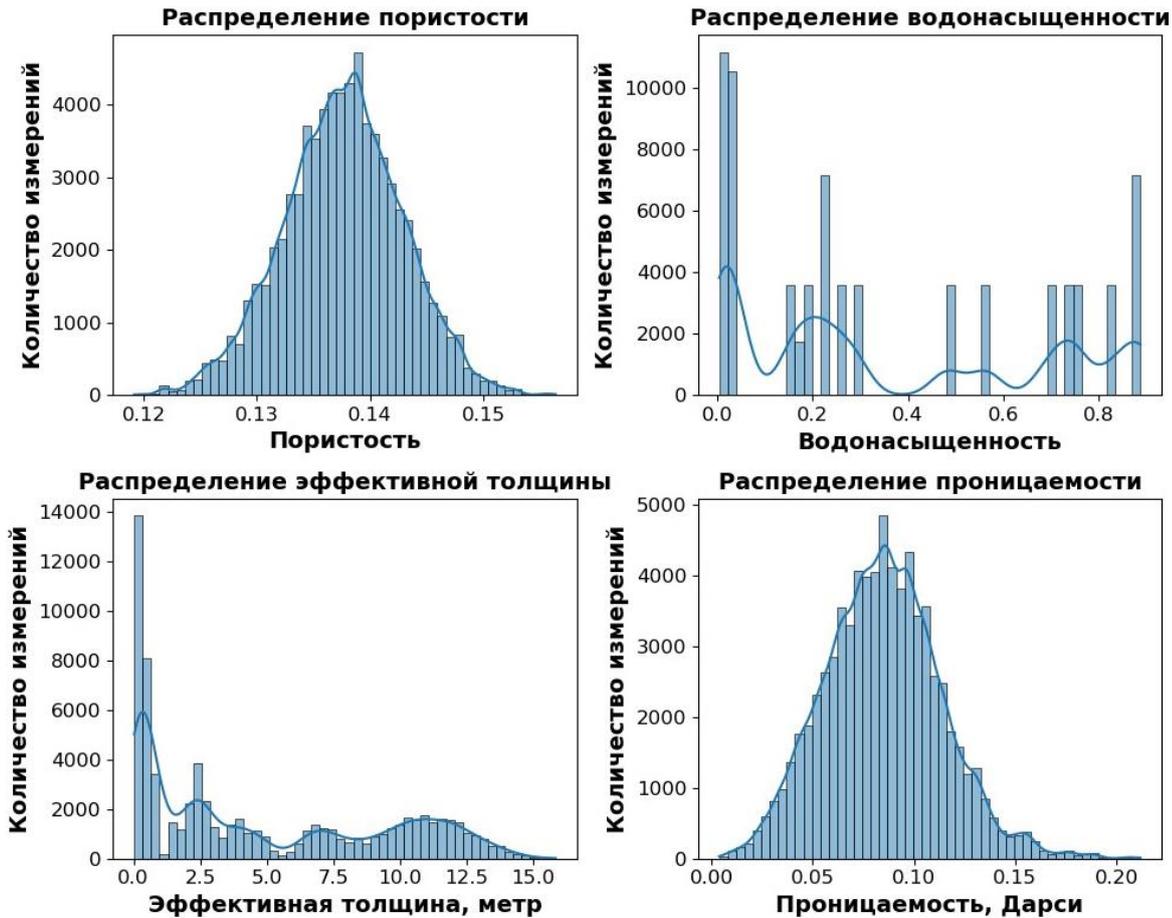


Рис. 2. Распределение геологических свойств  
Fig. 2. Distribution of geological properties

В данном случае пористость и проницаемость на основе петрофизической модели носят характер нормального распределения. Значения варьируются от 11,9 до 15,6 % и от 4 до 212 мД, соответственно. Распределения носят схожий характер, так как данные геологические свойства имеют петрофизическую зависимость. Распределения равномерные. Характер распределения для водонасыщенности и эффективной толщины значительно отличается от характера нормального распределения. Имеется большая дисперсия, что является объяснимым фактом – данные параметры часто носят большую неопределенность при моделировании месторождений и разбор закономерно может принимать большие значения. Коэффициент вариации данных параметров находится в районе 1,5.

#### Материалы и методы

Несколько моделей машинного обучения с различным соотношением тестовой выборки к тренировочной применено для получения оптимального решения задачи прогнозирования дебита газа. Данное решение имеет наименьший показатель средней квадратичной ошибки (Root Mean Square

Error – RMSE), средней абсолютной процентной ошибки (Mean Absolute Percentage Error – MAPE) и наибольший показатель коэффициента детерминации (R2) для тестовой и тренировочной выборки.

Тренировочные данные не требовали нормализации или стандартизации, так как методы случайного леса и ГБ не требуют данных операций [20, 21].

МО – область исследования в искусственном интеллекте, связанная с созданием и изучением статических алгоритмов, способных эффективно обобщать систему и выполнять задачи без явных инструкций. Другими словами, это способность охарактеризовать систему, находя зависимость между исходными и выходными данными. Существует множество систем, в которых зависимости между входом и выходом могут быть настолько сложными, что установить их аналитическим путем бывает просто невозможно. Также расчет усложняется большим количеством данных, количество значений входных параметров может исчисляться тысячами, поэтому решение данной системы становится нетривиальной задачей. Система пласт–скважина является также трудно прогнозируемой системой, что делает применение машинного обучения актуальным [22].

Зачастую, используя лишь одну модель машинного обучения, получить хороший прогноз бывает сложно. Результат модели зависит от начального шума данных, соотношения разделения стартового набора на тестовую и обучающую выборки, сам выбор значений в обучающую выборку несет долю случайности и создает разный разброс. Для того чтобы минимизировать данную ошибку создаются ансамбли моделей. Ансамблевое обучение – это общий мета-подход к машинному обучению, целью которого является улучшение предсказательной эффективности путем объединения прогнозов нескольких моделей [23]. Такие ансамбли и будут применяться в данной работе для прогнозирования показателей добычи скважины, в частности, случайный лес и ГБ на основе деревьев решений.

Метод дерева решений – семейство моделей МО, суть которого заключается в принятии большого количества простых решений, которые должны привести к ответу. Данный метод не обладает высокой обобщающей способностью из-за простого математического аппарата, но в то же время его расчёт очень прост, он служит фундаментом для построения ансамбля моделей [24]. Принцип работы дерева решений представлен на рис. 3.

Одним из самых распространённых и наглядных примеров применения ансамбля моделей МО является случайный лес [25–27]. Суть метода заключается в создании множества деревьев решений ограниченной глубины и дальнейшем получении ответа путем усреднения отдельных ответов на деревьях решений (задача регрессии) или выбором наиболее популярного класса (задача классификации). Основные параметры, которые необходимо уточнить: количество признаков для обучения дерева, количество деревьев в лесу, глубина деревьев в лесу. Особенность данного метода заключается в том, что он практически не подвержен переобучению,

он позволяет уменьшить дисперсию, способствуя уменьшению значения функции потерь [24].

Модель случайного леса была применена для прогнозирования дебита газа с учетом геологических параметров и давления в зоне забоя и пласте. Модель получает набор данных и выдает наиболее вероятный дебит газа, основываясь на результате множества деревьев решений. Данное дерево представлено на рис. 3.

ГБ представляет собой высокоэффективный и широко используемый алгоритм машинного обучения [29–31]. ГБ заключается в использовании объединения базовых алгоритмов (как правило, упрощенных) в единую систему. Обучение происходит последовательно (что отличается от методик с беггингом). Каждая новая итерация пытается вычислить отклонение уже обученной модели на обучающей выборке. Создавая такой ансамбль моделей, можно получить на выходе минимальные отклонения [32]. В целом можно использовать различные алгоритмы в качестве базового. Как правило, градиентный бустинг хорошо себя показывает при работе с деревьями решений [33]. Функция для оптимизации градиентного бустинга может выглядеть как в выражении (1):

$$L(t) = \sum_{i=1}^n l(y_i, \hat{y}(x_i)^{t-1} + f_t(x_i) + \Omega(f_t)), \quad (1)$$

где  $L(t)$  – функция оптимизации;  $l(g(t))$  – функция потерь;  $y_i, \hat{y}(x_i)^{t-1}$  – значение  $i$ -ого элемента обучающей выборки и сумма значений первых  $t$  базовых функций соответственно (в нашем случае деревьев);  $x_i$  – набор признаков  $i$ -ого элемента обучающей выборки;  $f_t$  – функция, которую хотим обучить (в нашем случае дерево) на шаге  $t$ ;  $f_t(x_i)$  – значение модели на  $i$ -м элементе обучающей выборки;  $\Omega(f_t)$  – функция регуляризации (не позволяет модели переобучиться).

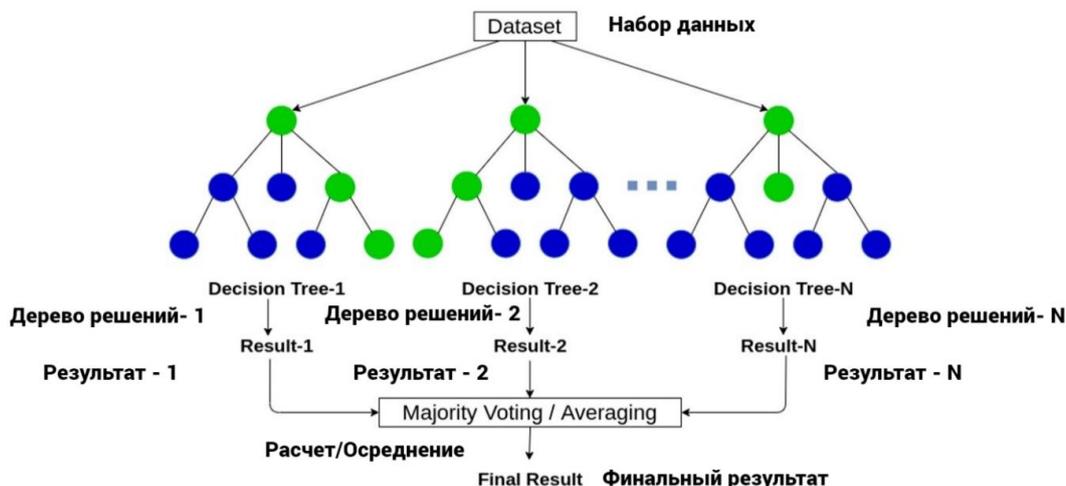


Рис. 3. Ансамбль деревьев решений (случайный лес) [28]  
 Fig. 3. Ensemble of decision trees (random forest) [28]

ГБ в наше время является одной из передовых моделей. Он активно применяется в нефтегазовой отрасли и позволяет решать различные задачи, имея возможность адаптироваться под конкретную задачу [34]. Именно по этой причине данная модель была выбрана для оценки целевого показателя (дебита газа горизонтальной скважины) в данной работе.

Для прогнозирования дебита газа также применялся ГБ, который позволяет обучаться, предсказывать ошибку исходной модели машинного обучения и вносить в итоговый дебит корректировки. Для поиска оптимальных параметров применялся половинчатый поиск по сетке. Данные разбивались в аналогичном соотношении, как в итерации со случайным лесом.

Гиперпараметры – параметры, которые характеризуют модель машинного обучения. Именно эти параметры определяют, как модель будет подбирать выходные параметры (целевую переменную). Часто к таким параметрам относятся: шаг итерации, глубина деревьев, количество объектов ансамбля, коэффициент регуляризации и т. д. [35].

Подбор гиперпараметров является важным этапом к оптимизации модели машинного обучения. Имеются различные подходы к подбору гиперпараметров: поиск по сетке (Grid Search), случайный поиск по сетке (Random Search), Байесовская оптимизация и множество других видов оптимизаторов. В целом можно создать произвольный оптимизатор, все ограничено лишь навыками разработчика и сферой использования алгоритма МО [35].

Половинчатый поиск по сетке заключается в переборе комбинаций гиперпараметров (заранее заданный набор). Суть метода заключается в уменьшении изначальной выборки до  $n$ -ого количества элементов ( $n$  задает пользователь) и дальнейшей оценке комбинаций гиперпараметров этой выборки. После этого выборка увеличивается в  $k$  раз, а количество комбинаций гиперпараметров уменьшается в  $k$  раз (убираются кандидаты с наихудшим показателем ошибки). В итоге остается один наилучший набор гиперпараметров для данной модели МО. Выбранный алгоритм позволяет выполнять подбор параметров быстрее, с небольшими потерями в точности [36].

Кросс валидация – один из подходов к подбору наилучших показателей гиперпараметров. Она разбивает данные на  $k$ -количество частей, используя 1 часть для оценки модели, а  $k-1$  часть для обучения. Затем выбирается другая часть для оценки и другие  $k-1$  части для обучения. Суммируется ошибка всех полученных моделей – данное число характеризует итоговую ошибку исследуемой модели МО. Этот подход позволяет использовать все данные и уменьшить долю случайности в определении наилучшей модели [37].

## Результаты исследования

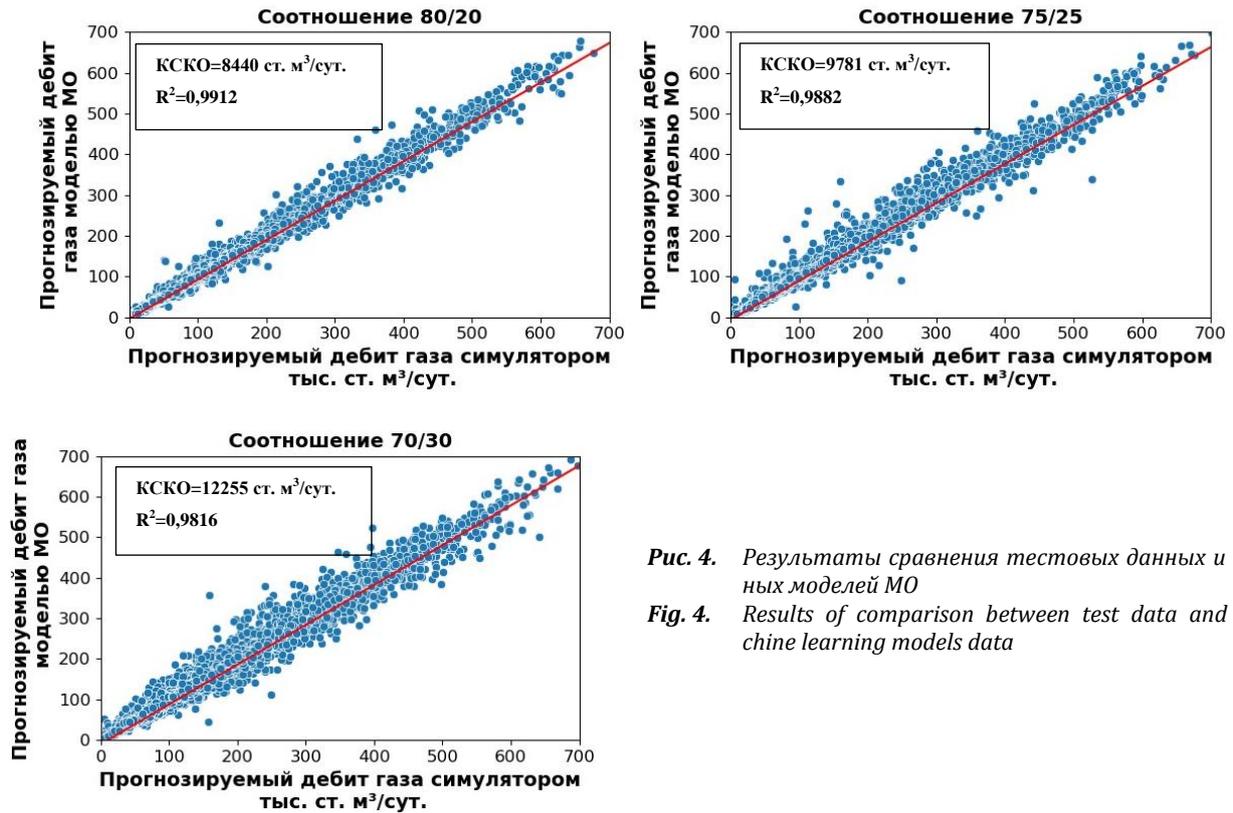
Выбор соотношения тестовой выборки к тренировочной является очень важным аспектом обучения и дальнейшей валидации моделей МО. Правильность данного соотношения играет важную роль в оценке, контроле переобучения и выборе наилучшей модели в машинном обучении [20]. Рис. 4 демонстрирует результаты работы моделей градиентного бустинга на тестовых данных с разным соотношением тестовой выборки к тренировочной.

Сравнение ошибки этих моделей на рис. 4 показывает, что с увеличением количества обучающих данных ошибка тестовой подгруппы понижается. Для соотношения 80/20 показатели ошибок меньше по сравнению с соотношениями 75/25 или 70/30. Это говорит о том, что модель, полученная с использованием соотношения 80/20, обладает более высокой способностью к обобщению. Для моделей случайного леса ситуация аналогична и лучшее соотношение – 80/20.

Результаты, полученные в процессе обучения, могут сильно зависеть от конкретного разбиения набора данных на обучающий и тестовый. Для решения этой проблемы использовался половинчатый поиск по сетке – пятикратная кросс-валидация для заранее заданной комбинации гиперпараметров. Происходило изначальное уменьшение тренировочной выборки (80 % от общего объема данных, 58929 записей данных) до 230 элементов (желательно подбирать число, близкое к кратному выборке, что следует из принципа работы алгоритма [36]) и дальнейшая оценка комбинаций гиперпараметров этой выборки через кросс валидацию с разделением на пять равных частей. Это дает хорошую оценку эффективности модели, поскольку для обучения и оценки модели используется пять различных разделений данных обучения и оценки модели. Более того, это гарантирует, что каждое наблюдение используется как для обучения, так и для тестирования, что снижает риск переобученности. После этого выборка увеличивается в три раза, а количество комбинаций гиперпараметров уменьшается в три раза (убираются кандидаты с наихудшим показателем ошибки). В итоге остается один наилучший кандидат, он и будет наилучшим набором гиперпараметров для данной модели машинного обучения. Подобный подход позволяет сэкономить время с незначительными потерями в точности [36].

В табл. 3, 4 представлены варианты гиперпараметров, комбинации которых использовались при половинчатом поиске по сетке.

Ключевыми параметрами для определения модели ГБ (в случае, если базовая модель – дерево) являются: количество моделей, темп обучения, максимальная глубина дерева, минимальное количество листьев и минимальное количество разделений.



**Рис. 4.** Результаты сравнения тестовых данных и данных моделей МО

**Fig. 4.** Results of comparison between test data and machine learning models data

**Таблица 3.** Набор гиперпараметров для поиска лучшего решения при градиентном бустинге

**Table 3.** Set of hyperparameters while searching for the best solution in a gradient boosting

Гиперпараметр Hyperparameter	Значение/Value											
Количество моделей Number of estimators	1000			1500			3000			5000		
Темп обучения Learning rate	0,01			0,1				0,2				
Максимальная глубина Max depth	3		4		5		6		7		8	9
	10	20	30	40	50	60	70	80	90	100	Нет None	
Минимальное количество листьев Min samples leaf	1		2			4		6			8	
Минимальный объём разделения Min samples split	2		5			10		20			40	

**Таблица 4.** Набор гиперпараметров для поиска лучшего решения при случайном лесе

**Table 4.** Set of hyperparameters while searching for the best solutions in a random forest

Гиперпараметр Hyperparameter	Значение/Value											
Бутстреп/Bootstrap	Истина/True						Ложь/False					
Максимальная глубина Max depth	1	20	30	40	50	60	70	80	90	100	Нет None	
Минимальное количество листьев Min samples leaf	1		2			4		6			8	
Минимальный объём разделения Min samples split	2		5			10		20			40	
Количество моделей Number of estimators	20	400	600	800	1000	1200	1400	1600	1800	2000		

**Таблица 5.** Наилучшие гиперпараметры для моделей МО  
**Table 5.** The best hyperparameters for ML models

Случайный лес/Random forest		Градиентный бустинг/Gradient boosting	
Гиперпараметр Hyperparameter	Значение Value	Гиперпараметр Hyperparameter	Значение Value
Бутстреп Bootstrap	Истина True	Темп обучения Learning rate	0,1
Максимальная глубина Max depth	40	Максимальная глубина Max depth	10
Минимальное количество листьев Min samples leaf	1	Минимальное количество листьев Min samples leaf	8
Минимальный объём разделения Min samples split	2	Минимальный объём разделения Min samples split	40
Количество моделей Number of estimators	2000	Количество моделей Number of estimators	5000

Ключевыми параметрами для определения модели случайного леса являются: количество моделей, наличие бутстрепа, максимальная глубина дерева, минимальное количество листьев и минимальное количество разделений.

В табл. 5 представлены наилучшие комбинации гиперпараметров, которые были получены на тестовую выборку с соотношением тренировочной выборки к тестовой 80/20 для моделей случайного леса и градиентного бустинга.

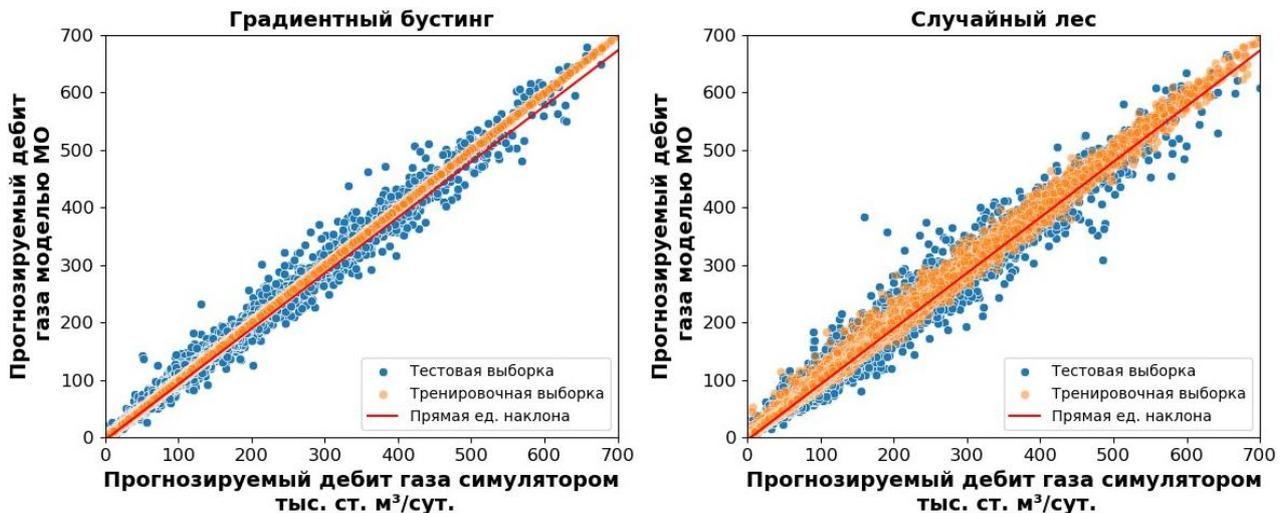
Половинчатый поиск по сетке определил, что при комбинациях гиперпараметров, указанных в табл. 5, получаются модели с наименьшими показателями ошибок (КСКО, САПО и  $R^2$ ).

Кросс плот дебита газа, прогнозируемого симулятором и моделью машинного обучения, представлен на рис. 5. На данном рисунке показана работа модели как на тренировочной выборке, так и на тестовой. Сосредоточение большинства точек данных вблизи линии с наклоном 1 для обучающих и тестовых данных указывает на то, что модель ГБ

обеспечивает более эффективное прогнозирование по сравнению с моделью случайного леса.

В табл. 6 приведены значения ошибок двух рассматриваемых алгоритмов на различном наборе данных. Данный анализ позволяет сделать вывод о переобученности и недообученности модели, оценить ее обобщающую способность [20].

Модель градиентного бустинга показывает лучшие результаты на всех наборах данных, что представлено в табл. 6 (тренировочная выборка, тестовая выборка, все данные), по сравнению с моделью случайного леса. Значения всех ошибок меньше в случае использования первого алгоритма. Низкие показатели ошибок градиентного бустинга на тестовой выборке показывают, что получена модель с хорошей обобщающей способностью. Применение перекрестной проверки позволило добиться хороших показателей на всех наборах данных. В случае тестовой выборки значение КСКО отличается в 1,66 раз,  $R^2$  на 1,5 %, САПО в 1,87.



**Рис. 5.** Результаты сравнения моделей

**Fig. 5.** Results of comparison between the models

**Таблица 6.** Сравнение ошибок прогнозирования дебита газа с помощью моделей машинного обучения на тренировочных, тестовых выборках и на всем наборе данных

**Table 6.** Comparison of gas flow rate prediction errors by the applied machine learning models to the training, test samples and the full dataset

Модель Model	Градиентный бустинг Gradient boosting			Случайный лес Random forest		
	Ошибка Error	САПО, % MAPE, %	КСКО, ст. м <sup>3</sup> /сут RMSE, std. m <sup>3</sup> /day	КД R <sup>2</sup>	САПО, % MAPE, %	КСКО, ст. м <sup>3</sup> /сут RMSE, std. m <sup>3</sup> /day
Тренировочная выборка Train sample 58929 записей 58929 records	0,536	704	0,9999	2,76	5125	0,9967
Тестовая выборка Test sample 14733 записей 14733 records	3,95	8440	0,9912	7,4	13994	0,9761
Все данные Full data 73662 записей 73662 records	1,21	3826	0,9982	3,67	7758	0,9926

### Заключение

В данном исследовании использовалось несколько моделей машинного обучения с целью определения дебита горизонтальной газовой скважины. Дебит газа был спрогнозирован путем применения моделей МО случайного леса и градиентного бустинга. Модель требует геологических параметров (пористость, проницаемость, эффективную толщину, насыщенность), давления на

забое и вблизи скважины, а также временного шага, на который необходимо установить дебит газа. Было произведено 300 итераций гидродинамических расчетов с целью создания исходного набора данных для дальнейшей разработки и проверки модели. Лучший результат показала модель градиентного бустинга: КД=0,9912, R<sup>2</sup>=3,96 %, КСКО=8440 ст. м<sup>3</sup>/сут.

### СПИСОК ЛИТЕРАТУРЫ

- Ritzi Jr R.W., Bobeck P. Comprehensive principles of quantitative hydrogeology established by Darcy (1856) and Dupuit (1857) // Water Resources Research. – 2008. – Vol. 44. – № 10. – W 10402. DOI: 10.1029/2008WR007002.
- Fu Y. A critical productivity equation of horizontal wells in a bottom water drive reservoir with low-permeability interbeds // Arabian Journal Geosciences. – 2019. – Vol. 12. – Ar. Numb. 758. – № 24. DOI: 10.1007/s12517-019-4930-y.
- Ahmed T. Reservoir engineering handbook. 5th ed. – Oxford: Gulf Professional Publ., 2019. – 492 p. DOI: 10.1016/C2016-0-04718-6.
- Arps J.J. Analysis of decline curves // Transactions of the AIME. – 1945. – Vol. 160. – № 1. – P. 228–247. DOI: 10.2118/945228-G.
- Tracy G.W. Simplified form of the material balance equation // Transactions of the AIME. – 1955. – Vol. 204. – № 1. – P. 243–246. DOI: 10.2118/438-g.
- Fuentes-Cruz G., Vasquez-Cruz M.A. Reservoir performance analysis through the material balance equation: an integrated review based on field examples // Journal of Petroleum Science and Engineering. – 2022. – Vol. 208. – 109377, Part A. DOI: 10.1016/j.petrol.2021.109377.
- Coats K.H. Use and misuse of reservoir simulation models // Journal of Petroleum Technology. – 1969. – Vol. 21. – № 11. – P. 1391–1398. DOI: 10.2118/2367-pa.
- Fanchi J. Principles of applied reservoir simulation. 3rd ed. – Edinburgh: Gulf Professional, 2006. – 510 p.
- Hossain M.S. Assessment of prospective gas resources and develop a production view of eastern foldbelt (block 18), Bangladesh using Monte Carlo simulation method // Natural Gas Industry B. – 2022. – Vol. 9. – № 6. – P. 588–594. DOI: 10.1016/j.ngib.2022.11.001.
- Approach to quantify productivity of fishbone multilateral well / A. Hassan, A. Abdulraheem, S. Elkatatny, M. Ahmed // SPE Annual Technical Conf. and Exhibition. – San Antonio, Texas, USA, October 9–11, 2017. – P. 1–11. DOI: 10.2118/187458-ms.
- Zhong R., Salehi C., Johnson R. Machine learning for drilling applications: a review // Journal of Natural Gas Science and Engineering. – 2022. – Vol. 108. – 104807. DOI: 10.1016/j.jngse.2022.104807.
- Hybridized machine-learning for prompt prediction of rheology and filtration properties of water-based drilling fluids / S. Davoodi, M. Mehrad, D.A. Wood, H. Ghorbani, V.S. Rukavishnikov // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 123. – 106459, Part C. DOI: 10.1016/j.engappai.2023.106459.
- Dramschi J.S. 70 years of machine learning in geoscience in review // Advances in Geophysics. – 2020. – Vol. 61. – P. 1–55. DOI: 10.1016/bs.agph.2020.08.002.

14. A new approach to mechanical brittleness index modeling based on conventional well logs using hybrid algorithms / T.M. Zamanzadeh, S. Davoodi, B. Larki, M. Mehrad, S. Rashidi, Vasfi M. // *Earth Science Informatics*. – 2023. – Vol. 16. – P. 3387–3416. DOI: 10.1007/s12145-023-01098-1.
15. Wang H., Chen S. Insights into the application of machine learning in reservoir engineering: current developments and future trends // *Energies*. – 2023. – Vol. 16. – 1392. – № 3. DOI: 10.3390/en16031392.
16. Zhang R., Jia H. Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs // *Petroleum Exploration and Development*. – 2021. – Vol. 48. – № 1. – P. 201–211. DOI: 10.1016/S1876-3804(21)60016-2.
17. A systematic review of data science and machine learning applications to the oil and gas industry / Z. Tariq, M.S. Aljawad, A. Hasan, M. Murtaza, E. Mohammed, A. El-Husseiny, A.A. Sulaiman, M. Mohamed, A. Abdulazeez // *Petroleum Exploration and Production Technology*. – 2021. – Vol. 11. – P. 4339–4374. DOI: 10.1007/s13202-021-01302-2.
18. Multilateral wells evaluation utilizing artificial intelligence / A. Buhulaigah, A.S. Al-Mashhad, S.A. Al-Arifi, M.S. Al-Kadem, M.S. Al-Dabbous // *SPE Middle East Oil and Gas Show and Conf.* – Manama, Kingdom of Bahrain, March 6–9, 2017. – P. 1–12. DOI: 10.2118/183688-ms.
19. Wui Ng C.S., Ghahfarokhi A.J., Amar M.N. Well production forecast in Volve field: application of rigorous machine learning techniques and metaheuristic algorithm // *Journal of Petroleum Science and Engineering*. – 2022. – Vol. 208. – 109468, Part B. DOI: 10.1016/j.petrol.2021.109468.
20. Gopal M. *Applied machine learning*. 1st ed. – New York: McGraw-Hill Education, 2019. – 623 p.
21. Filho M. Do decision trees need feature scaling or normalization? // *Forecasting*. – 2023. URL: <https://forecasting.com/posts/do-decision-trees-need-feature-scaling-or-normalization/> (дата обращения 06.12.2023).
22. Pugliese R., Regondi S., Marini R. Machine learning-based approach: Global trends, research directions, and regulatory standpoints // *Data Science and Management*. – 2021. – Vol. 4. – P. 19–29. DOI: 10.1016/j.dsm.2021.12.002.
23. Mohammed A., Kora R. A comprehensive review on ensemble deep learning: opportunities and challenges // *Journal of King Saud University – Computer and Information Sciences*. – 2023. – Vol. 35. – № 2. – P. 757–774. DOI: 10.1016/j.jksuci.2023.01.014.
24. Leondes C.T. *Expert systems: the technology of knowledge management and decision making for the 21st century*. – Cambridge: Academic Press, 2002. – 1947 p.
25. Gordon J.B., Sanei H., Pedersen P.K. Predicting hydrogen and oxygen indices (HI, OI) from conventional well logs using a Random Forest machine learning algorithm // *International Journal of Coal Geology*. – 2022. – Vol. 249. – 103903. DOI: 10.1016/j.coal.2021.103903.
26. Rahimi M., Riahi M.A. Reservoir facies classification based on random forest and geostatistics methods in an offshore oilfield // *Journal of Applied Geophysics*. – 2022. – Vol. 201. – 104640. DOI: 10.1016/j.jappgeo.2022.104640.
27. Darijani M., Farquharson C.G., Perrouy S.A. Random Forest approach to predict geology from geophysics in the Pontiac subprovince, Canada // *Canadian Journal of Earth Sciences*. – 2022. – Vol. 59. – № 8. – P. 489–503. DOI: 10.1139/cjes-2021-0089.
28. Brital A. Random forest algorithm explained. – 2021. URL: <https://anasbrital98.github.io/blog/2021/Random-Forest/> (дата обращения 05.12.2023).
29. Advanced well stimulation selection with gradient boosting / V. Kniazev, A. Erofeev, A. Demidov, D. Orlov, D. Koroteev // *Geoenergy Science and Engineering*. – 2023. – Vol. 228. – 212026. DOI: 10.1016/j.geoen.2023.212026.
30. Lee H., Lee H.P. Formation lithology predictions based on measurement while drilling (MWD) using gradient boosting algorithms // *Geoenergy Science and Engineering*. – 2023. – Vol. 227. – 211917. DOI: 10.1016/j.geoen.2023.211917.
31. A physics-guided eXtreme gradient boosting model for predicting the initial productivity of oil wells / Y. Dong, L. Song, Q. Zhao, Z. Ding, L. Qiu, C. Lu, G. Chen // *Geoenergy Science and Engineering*. – 2023. – Vol. 231. – 212402, Part B. DOI: 10.1016/j.geoen.2023.212402.
32. Friedman J.H. Greedy function approximation: a gradient boosting machine // *The Annals of Statistics*. – 2001. – Vol. 29. – № 5. – P. 1189–1232. DOI: 10.1214/aos/1013203451.
33. Chen T., Guestrin C. XGBoost: a scalable tree boosting system // *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*. – San Francisco, CA, USA, August 7–8, 2016. – P. 785–794. DOI: 10.1145/2939672.2939785.
34. Belyadi H., Haghightat A. *Supervised learning // Machine Learning Guide for Oil and Gas Using Python*. – Oxford: Gulf Professional Publ., 2021. – P. 169–295. DOI: 10.1016/B978-0-12-821929-4.00004-4.
35. Yang L., Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice // *Neurocomputing*. – 2020. – Vol. 415. – P. 295–316. DOI: 10.1016/j.neucom.2020.07.061.
36. Sklearn.model\_selection.HalvingGridSearchCV // scikit-learn, 2023. URL: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.HalvingGridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingGridSearchCV.html) (дата обращения 06.12.2023).
37. Елистратова Е. Кросс-валидация // *Школа анализа данных*. URL: <https://education.yandex.ru/handbook/ml/article/kross-validaciya> (дата обращения 05.12.2023).

### Информация об авторах

**Сергей Александрович Пискунов**, инженер лаборатории геологии месторождений нефти и газа Национального исследовательского Томского политехнического университета, Россия, 634050, г. Томск, пр. Ленина, 30. [piskunovsa@hw.tpu.ru](mailto:piskunovsa@hw.tpu.ru); <https://orcid.org/0009-0007-8491-4596>

**Шадфар Давуди**, инженер-исследователь Центра подготовки и переподготовки специалистов нефтегазового дела при Национальном исследовательском Томском политехническом университете, Россия, 634050, г. Томск, пр. Ленина, 30. [davoodis@hw.tpu.ru](mailto:davoodis@hw.tpu.ru); <https://orcid.org/0000-0003-1733-1677>

Поступила в редакцию: 29.01.2024

Поступила после рецензирования: 20.02.2024

Принята к публикации: 19.04.2024

## REFERENCES

1. Ritz Jr R.W., Bobeck P. Comprehensive principles of quantitative hydrogeology established by Darcy (1856) and Dupuit (1857). *Water Resources Research*, 2008, vol. 44, no. 10, W 10402. DOI: 10.1029/2008WR007002.
2. Fu Y. A critical productivity equation of horizontal wells in a bottom water drive reservoir with low-permeability interbeds. *Arabian Journal Geosciences*, 2019, vol. 12, Ar. Numb. 758, no. 24. DOI: 10.1007/s12517-019-4930-y.
3. Ahmed T. *Reservoir engineering handbook*. 5<sup>th</sup> ed. Oxford, Gulf Professional Publ., 2019. 492 p. DOI: 10.1016/C2016-0-04718-6.
4. Arps J.J. Analysis of decline curves. *Transactions of the AIME*, 1945, vol. 160, no. 1, pp. 228–247. DOI: 10.2118/945228-G.
5. Tracy G.W. Simplified form of the material balance equation. *Transactions of the AIME*, 1955, vol. 204, no. 1, pp. 243–246. DOI: 10.2118/438-g.
6. Fuentes-Cruz G., Vasquez-Cruz M.A. Reservoir performance analysis through the material balance equation: an integrated review based on field examples. *Journal of Petroleum Science and Engineering*, 2022, vol. 208, 109377, Part A. DOI: 10.1016/j.petrol.2021.109377.
7. Coats K.H. Use and misuse of reservoir simulation models. *Journal of Petroleum Technology*, 1969, vol. 21, no. 11, pp. 1391–1398. DOI: 10.2118/2367-pa.
8. Fanchi J. *Principles of applied reservoir simulation*. 3<sup>rd</sup> ed. Edinburgh, Gulf Professional, 2006. 510 p.
9. Hossain M.S. Assessment of prospective gas resources and develop a production view of eastern foldbelt (block 18), Bangladesh using Monte Carlo simulation method. *Natural Gas Industry B*, 2022, vol. 9, no. 6, pp. 588–594. DOI: 10.1016/j.ngib.2022.11.001.
10. Hassan A., Abdurraheem A., Elkatatny S., Ahmed M. Approach to quantify productivity of fishbone multilateral well. *SPE Annual Technical Conf. and Exhibition*. San Antonio, Texas, USA, October 9–11, 2017. pp. 1–11. DOI: 10.2118/187458-ms.
11. Zhong R., Salehi C., Johnson R. Machine learning for drilling applications: a review. *Journal of Natural Gas Science and Engineering*, 2022, vol. 108, 104807. DOI: 10.1016/j.jngse.2022.104807.
12. Davoodi S., Mehrad M., Wood D.A., Ghorbani H., Rukavishnikov V.S. Hybridized machine-learning for prompt prediction of rheology and filtration properties of water-based drilling fluids. *Engineering Applications of Artificial Intelligence*, 2023, vol. 123, 106459, Part C. DOI: 10.1016/j.engappai.2023.106459.
13. Dramsch J.S. 70 Years of machine learning in geoscience in review. *Advances in Geophysics*, 2020, vol. 61, pp. 1–55. DOI: 10.1016/bs.agph.2020.08.002.
14. Zamanzadeh T.M., Davoodi S., Larki B., Mehrad M., Rashidi S., Vasfi M. A new approach to mechanical brittleness index modeling based on conventional well logs using hybrid algorithms. *Earth Science Informatics*, 2023, vol. 16, pp. 3387–3416. DOI: 10.1007/s12145-023-01098-1.
15. Wang H., Chen S. Insights into the application of machine learning in reservoir engineering: current developments and future trends. *Energies*, 2023, vol. 16, 1392, no. 3. DOI: 10.3390/en16031392.
16. Zhang R., Jia H. Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs. *Petroleum Exploration and Development*, 2021, vol. 48, no. 1, pp. 201–211. DOI: 10.1016/S1876-3804(21)60016-2.
17. Tariq Z., Aljawad M.S., Hasan A., Murtaza M., Mohammed E., El-Husseiny A., Sulaiman A.A., Mohamed M., Abdulazeez A. A systematic review of data science and machine learning applications to the oil and gas industry. *Petroleum Exploration and Production Technology*, 2021, vol. 11, pp. 4339–4374. DOI: 10.1007/s13202-021-01302-2.
18. Buhulaigah A., Al-Mashhad A.S., Al-Arifi S.A., Al-Kadem M.S., Al-Dabbous M.S. Multilateral wells evaluation utilizing artificial intelligence. *SPE Middle East Oil and Gas Show and Conf.* Manama, Kingdom of Bahrain, March 6–9, 2017. pp. 1–12. DOI: 10.2118/183688-ms.
19. Wui Ng C.S., Ghahfarokhi A.J., Amar M.N. Well production forecast in Volve field: application of rigorous machine learning techniques and metaheuristic algorithm. *Journal of Petroleum Science and Engineering*, 2022, vol. 208, 109468, Part B. DOI: 10.1016/j.petrol.2021.109468.
20. Gopal M. *Applied machine learning*. 1<sup>st</sup> ed. New York, McGraw-Hill Education, 2019. 623 p.
21. Filho M. Do decision trees need feature scaling or normalization? *Forecastegy*, 2023. Available at: <https://forecastegy.com/posts/do-decision-trees-need-feature-scaling-or-normalization/> (accessed 6 December 2023).
22. Pugliese R., Regondi S., Marini R. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, 2021, vol. 4, pp. 19–29. DOI: 10.1016/j.dsm.2021.12.002.
23. Mohammed A., Kora R. A comprehensive review on ensemble deep learning: opportunities and challenges. *Journal of King Saud University – Computer and Information Sciences*, 2023, vol. 35, no. 2, pp. 757–774. DOI: 10.1016/j.jksuci.2023.01.014.
24. Leondes C.T. *Expert systems: the technology of knowledge management and decision making for the 21<sup>st</sup> century*. Cambridge, Academic Press, 2002. 1947 p.
25. Gordon J.B., Sanei H., Pedersen P.K. Predicting hydrogen and oxygen indices (HI, OI) from conventional well logs using a Random Forest machine learning algorithm. *International Journal of Coal Geology*, 2022, vol. 249, 103903. DOI: 10.1016/j.coal.2021.103903.
26. Rahimi M., Riahi M.A. Reservoir facies classification based on random forest and geostatistics methods in an offshore oilfield. *Journal of Applied Geophysics*, 2022, vol. 201, 104640. DOI: 10.1016/j.jappgeo.2022.104640.
27. Darijani M., Farquharson C.G., Perrouy S.A. Random Forest approach to predict geology from geophysics in the Pontiac subprovince, Canada. *Canadian Journal of Earth Sciences*, 2022, vol. 59, no. 8, pp. 489–503. DOI: 10.1139/cjes-2021-0089.

28. Brital A. *Random Forest Algorithm Explained*. 2021. Available at: <https://anasbrital98.github.io/blog/2021/Random-Forest/> (accessed 5 December 2023).
29. Kniazev V., Erofeev A., Demidov A., Orlov D., Koroteev D. Advanced well stimulation selection with gradient boosting. *Geoenergy Science and Engineering*, 2023, vol. 228, 212026. DOI: 10.1016/j.geoen.2023.212026.
30. Lee H., Lee H.P. Formation lithology predictions based on measurement while drilling (MWD) using gradient boosting algorithms. *Geoenergy Science and Engineering*, 2023, vol. 227, 211917. DOI: 10.1016/j.geoen.2023.211917.
31. Dong Y., Song L., Zhao Q., Ding Z., Qiu L., Lu C., Chen G. A physics-guided eXtreme gradient boosting model for predicting the initial productivity of oil wells. *Geoenergy Science and Engineering*, 2023, vol. 231, 212402, Part B. DOI: 10.1016/j.geoen.2023.212402.
32. Friedman J.H. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 2001, vol. 29, no. 5, pp. 1189–1232. DOI: 10.1214/aos/1013203451.
33. Chen T., Guestrin C. XGBoost: a scalable tree boosting system. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*. San Francisco, CA, USA, August 7–8, 2016. pp. 785–794. DOI: 10.1145/2939672.2939785.
34. Belyadi H., Haghghat A. *Supervised learning. Machine Learning Guide for Oil and Gas Using Python*. Oxford, Gulf Professional Publ., 2021. pp. 169–295. DOI: 10.1016/B978-0-12-821929-4.00004-4.
35. Yang L., Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 2020, vol. 415, pp. 295–316. DOI: 10.1016/j.neucom.2020.07.061.
36. Sklearn.model\_selection.HalvingGridSearchCV. *scikit-learn*, 2023. Available at: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.HalvingGridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.HalvingGridSearchCV.html) (accessed 6 December 2023).
37. Elistratova E. *Cross-validation. School of Data Analysis*. (In Russ.) Available at: <https://education.yandex.ru/handbook/ml/article/kross-validaciya> (accessed 5 December 2023).

### Information about the authors

**Sergey A. Piskunov**, Engineer, National Research Tomsk Polytechnic University, 30, Lenin avenue, Tomsk, 634050, Russian Federation. [piskynovsa@hw.tpu.ru](mailto:piskynovsa@hw.tpu.ru); <https://orcid.org/0009-0007-8491-4596>.

**Shadfar Davoodi**, Research Engineer, National Research Tomsk Polytechnic University, 30, Lenin avenue, Tomsk, 634050, Russian Federation. [davoodis@hw.tpu.ru](mailto:davoodis@hw.tpu.ru); <https://orcid.org/0000-0003-1733-1677>.

Received: 29.01.2024

Revised: 20.02.2024

Accepted: 19.04.2024