

УДК 519.688
DOI: 10.18799/24131830/2025/9/4867
Шифр специальности ВАК: 1.2.1
Научная статья

Прогнозирование притока после гидравлического разрыва нефтяного пласта при помощи искусственного интеллекта

М.А. Ямкин^{1✉}, Е.У. Сафиуллина², А.В. Ямкин³

¹ Университет ИТМО, Россия, г. Санкт-Петербург

² Санкт-Петербургский горный университет императрицы Екатерины II, Россия, г. Санкт-Петербург

³ «ООО Газпром трансгаз Томск», Россия, г. Томск

✉ makson.yamkin@mail.ru

Аннотация. *Актуальность* исследования обусловлена тем, что эффективность проведения гидравлического разрыва пласта зависит от ряда признаков, которые на данный момент не учитываются в традиционных формулах расчета гидроразрыв пласта. Так, в традиционных методиках расчета на основе одномерных геомеханических моделей (в работе рассматривались модели Христиановича–Гиртсма–де Клерка, радиальная модель и модель Перкинса–Керна–Нордгрена) распространения трещины делаются допущения, которые могут снижать точность расчета и, как следствие, точность прогнозирования роста дебита нефти после проведения гидроразрыва пласта. *Целью* данной работы является прогнозирование роста дебита нефти после проведения гидроразрыва пласта с использованием технологий искусственного интеллекта, а именно машинного обучения, и последующее сравнение точности такого прогнозирования с точностью традиционных методик расчетов на основе одномерных моделей для выявления наиболее точной методики. *Объект:* модель роста дебита нефти после проведения гидроразрыва пласта. *Методы:* разработка специального программного обеспечения для прогнозирования дебита нефти после проведения гидроразрыва пласта с использованием традиционных методик на основе одномерных моделей и с помощью методик машинного обучения. *Результаты.* Было установлено, что точность традиционных методик на основе одномерных моделей составляет 84 %, а точность методик с использованием машинного обучения – 87 %. *Заключение.* Показано, что использование технологий искусственного интеллекта, а именно машинного обучения, обеспечивает более точное прогнозирование гидроразрыва пласта по сравнению с традиционными методиками на основе одномерных моделей.

Ключевые слова: дебит нефти, гидроразрыв пласта, машинное обучение, одномерные модели, искусственный интеллект

Для цитирования: Ямкин М.А., Сафиуллина Е.У., Ямкин А.В. Прогнозирование притока после гидравлического разрыва нефтяного пласта при помощи искусственного интеллекта // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2025. – Т. 336. – № 9. – С. 85–96. DOI: 10.18799/24131830/2025/9/4867

UDC 519.688
DOI: 10.18799/24131830/2025/9/4867
Scientific paper

Forecasting inflow after hydraulic fracturing of oil formation using artificial intelligence

M.A. Yamkin^{1✉}, E.U. Safiullina², A.V. Yamkin³

¹ ITMO University, St Petersburg, Russian Federation

² St Petersburg Mining University of Empress Catherine II, St Petersburg, Russian Federation

³ «Gazprom Transgaz Tomsk LLC», Tomsk, Russian Federation

✉ makson.yamkin@mail.ru

Abstract. Relevance. The effectiveness of hydraulic fracturing is determined by many features that are currently not taken into consideration in traditional formulas for calculating hydraulic fracturing. Thus, in traditional calculation methods, which are based on one-dimensional models of crack propagation (the work considered the Khristianovich–Girtsma–de Klerk model, the radial model and the Perkins–Kern–Nordgren model), assumptions are made to calculate various parameters, which in some cases will affect the accuracy of the calculation. Because of this, the effectiveness and necessity of hydraulic fracturing might under question. **Aim.** To predict hydraulic fracturing using artificial intelligence technologies such as machine learning techniques and then compare the accuracy of the prediction with the accuracy of traditional calculations using one-dimensional models to identify the most accurate technique. **Object.** Model of oil production growth after hydraulic fracturing. **Methods.** Development of special software for predicting oil production after hydraulic fracturing using traditional methods based on one-dimensional models and using machine learning techniques. **Results.** Traditional univariate models were 84% accurate, while machine learning models were 87% accurate. **Conclusion.** The use of artificial intelligence technologies, namely machine learning, provides more accurate hydraulic fracturing prediction compared to traditional methods based on one-dimensional models.

Keywords: oil production, hydraulic fracturing, machine learning, one-dimensional models, artificial intelligence

For citation: Yamkin M.A., Safiullina E.U., Yamkin A.V. Forecasting inflow after hydraulic fracturing of oil formation using artificial intelligence. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2025, vol. 336, no. 9, pp. 85–96. DOI: 10.18799/24131830/2025/9/4867

Введение

В настоящее время гидроразрыв пласта (ГРП) часто применяется для увеличения нефтеотдачи и интенсификации притока на месторождениях нефти. Данная операция позволяетратно увеличивать дебит нефти. Основным параметром, характеризующим эффективность ГРП, является прирост дебита нефти [6]. Для расчета дебита нефти после ГРП широко применяются традиционные методики с использованием одно-, двух- и трехмерных моделей распространения трещин ГРП. Несмотря на ряд преимуществ двух- и трехмерных моделей, их применение ограничивается сложностью математического описания, поэтому в данной работе они не рассматриваются [3]. Одномерные модели являются более предпочтительным вариантом за счет простоты математического описания моделей и скорости расчетов. Однако их ограничением является большое расхождение расчетных и фактических параметров ГРП [7]. Для повышения точности прогнозирования дебита нефти после ГРП перспективным представляется использование технологий искусственного интеллекта, а именно машинного обучения. В связи с этим в данной работе рассматривается два подхода:

- 1) расчет дебита нефти с использованием традиционных одномерных моделей распространения трещины ГРП: модель Христиановича–Гиртсма–де Клерка (Khristianovich and Zheltov and Geertsma and de-Klerk – KGD), радиальная модель и модель Перкинса–Керна–Нордгрена (Perkins and Kern and Nordgren – PKN) [8, 9]. Данный расчет осуществлялся путем решения дифференциальных уравнений, закладываемых в традиционных моделях, на языке программирования Python;
- 2) прогнозирование дебита нефти с использованием моделей машинного обучения [10].

Результаты расчета и прогнозирования с использованием вышеуказанных подходов сравнивались с фактическими данными с Ванкорского газонефтяного месторождения.

Актуальность

Ванкорское месторождение, которое выбрано для определения наиболее точной методики по расчету дебита нефти после ГРП, на данный момент находится на третьей стадии разработки [2].

Основными предпосылками для применения технологии ГРП на объекте Нх-І рассматриваемого месторождения являются:

- низкая обводненность объекта – на текущий момент значение обводненности составляет 28 %. За счет этого отсутствует ограничение по применению данной технологии, связанное с высокой обводненностью;
- фонд скважин полностью разбурен: частично скважины эксплуатируются фонтанным способом, а в некоторые скважины из-за прекращения фонтанирования был спущен электроцентробежный насос (ЭЦН). На данный момент количество скважин составляет 110 % от проектного фонда скважин [11]. Несмотря на это, объект характеризуется невыполнением проектного плана разработки по накопленной добыче нефти, текущей добыче нефти и текущему коэффициенту извлечения нефти (КИН). Один из вариантов решения данной проблемы – подключение нижней части пласта, которая не разрабатывается на данный момент по причине низкой проницаемости, к разработке. Более интенсивная разработка верхней части пласта не может привести к необходимому результату, поскольку треть извлекаемых запасов уже извлечена [11], и более активная разработка верхней части может привести к опережающему росту обводненности и снижению добычи нефти;

- объект характеризуется разделением на две части по проницаемости: верхняя часть характеризуется хорошей проницаемостью ($0,031 \text{ мкм}^2$) [11], тогда как нижняя часть характеризуется пониженной проницаемостью ($0,00035 \text{ мкм}^2$) [11]. При этом запасы нижней части составляют треть от всех запасов объекта Нх-І, поэтому подключение нижней части к разработке представляется целесообразным.

На основе вышеизложенного можно предположить, что проведение операции ГРП на объекте Нх-І представляется актуальным. Подробные геолого-физические свойства и показатели разработки по данному объекту для определения возможности проведения ГРП приведены в табл. 1.

Таблица 1. Основные геолого-физические свойства и показатели разработки [12–14]

Table 1. Basic geological and physical properties and development indicators [12–14]

Наименование параметра/Parameter name	Нх-І
Средняя глубина залегания пласта, м Average depth of formation, m	2685
Средняя нефтенасыщенная мощность, м Average oil-saturated capacity, m	6,1
Коэффициент нефтенасыщенности Oil saturation coefficient, %	46,0
Начальная пластовая температура Initial reservoir temperature, °C	50
Начальное пластовое давление, МПа Initial reservoir pressure, MPa	24,6
Проницаемость, мкм^2 /Permeability, mkm^2	0,02
Пористость/Porosity, %	18,9
Коэффициент песчаности, д. ед. Sand content coefficient, units	0,28
Коэффициент расчлененности, д. ед. Partition coefficient, units	3,6
Средняя общая мощность пласта, м Average total thickness of the formation, m	38
Площадь нефтеносности, га/Oil bearing area, ha	27148
Обводненность/Water cut, %	28

С учетом обозначенных выше факторов применение ГРП как метода увеличения нефтеотдачи на рассматриваемом эксплуатационном объекте является эффективным. Исходя из основных критериев применимости ГРП [15], а именно:

- высоких удельных запасов по скважине;
- технически-исправной скважины;
- угла отклонения скважины от вертикали при входе в пласт не более 10° ;
- расстояния до линии нагнетания и ВНК более 500 м;
- текущей обводненности не более 30 %;
- текущего пластового давления более 0,85 от первоначального;
- отсутствия заколонных перетоков;
- пластового давления больше гидростатического;

- нефтенасыщенной толщины более 5 м;
 - глубины скважины не более 3000 м,
- в качестве скважин-кандидатов для проведения ГРП были выбраны скважины № 724 и 728, поскольку они являются наиболее соответствующими вышеперечисленным пунктам. Проницаемость пласта в районе этих скважин является заниженной [12], и, как следствие, данные скважины характеризуются заниженным дебитом против окружающих скважин [13]. Информация по скважинам 724 и 728 приведена в табл. 2.

Таблица 2. Подбор скважин-кандидатов для проведения операции ГРП [12–14]

Table 2. Selection of candidate wells for hydraulic fracturing operations [12–14]

Параметр Parameter	Скважина/Well	
	724	728
Удельные запасы, тыс. т Specific reserves, thousand tons	120	117
Исправность/Serviceability	да/yes	
Угол отклонения от вертикали при входе в пласт, ° Angle of deviation from the vertical at the entrance to the formation, °	7	8
Расстояние до линии водонефтяного контакта, м Distance to oil-water contact line, m	650	640
Текущая обводненность/Current water cut, %	22	18
Начальное пластовое давление, МПа Initial reservoir pressure, MPa	26	
Текущее пластовое давление, МПа Current reservoir pressure, MPa	24,4	
Заколонные перетоки/Behind-the-casing flows	отсутствуют/absent	
Гидростатическое давление, МПа Hydrostatic pressure, MPa	22,7	21
Нефтенасыщенная толщина, м Oil-saturated thickness, m	6,1	
Глубина скважины, м/Well depth, m	2670	2665
Дебит нефти до ГРП, т/сут. Oil flow rate before hydraulic fracturing, t/day	24	44
Дебит нефти после ГРП, т/сут. Oil flow rate after hydraulic fracturing, t/day	60	72

Рассматриваемый объект удовлетворяет всем требованиям, однако расчет дебита нефти с использованием традиционных одномерных моделей распространения трещин ГРП может давать значительную ошибку. Из-за этого фактический дебит нефти после проведения ГРП будет существенно отличаться от расчетного ввиду того, что во всех одномерных моделях делается ряд следующих допущений [8–10]:

- среда распространения трещин предполагает собой однородную изотропную среду;
- жидкость разрыва обладает ньютоновскими свойствами;
- пласт предполагается линейно-упругой средой.

В реальных условиях эти допущения не всегда являются актуальными, поэтому необходим аль-

тернативный, более точный, способ расчета дебита нефти после ГРП. Одним из таких способов является прогнозирование дебита нефти с использованием технологий искусственного интеллекта, а именно машинного обучения, который и рассматривается в данной работе.

Анализ литературы показал, что технологии искусственного интеллекта уже успешно применяются для решения задачи поиска скважин-кандидатов для проведения ГРП [7]. Однако описанная в работе [7] модель является классификационной и позволяет лишь ответить на вопрос: является скважина кандидатом для проведения операции ГРП или нет. Данных по прогнозированию дебита нефти после ГРП с использованием машинного обучения, а также по сравнению точности прогнозирования дебита нефти после ГРП с использованием машинного обучения и с использованием традиционных геомеханических моделей, авторам в открытых источниках информации не встречалось. Вместе с тем прогнозирование дебита нефти после ГРП является актуальной задачей, так как позволяет провести экономическую оценку эффективности проведения ГРП за счет возможности расчета дополнительного прироста дебита нефти после проведения ГРП на скважине. Это обуславливает актуальность результатов, представленных в данной работе.

Методы исследования

Для решения задачи авторами было разработано специальное программное обеспечение (ПО) на языке программирования Python с использованием принципов объектно-ориентированного программирования. Разработанное ПО позволяет рассчитать дебит нефти после проведения ГРП при помощи вышеперечисленных традиционных одномерных моделей, а также спрогнозировать его при помощи методов машинного обучения. Результаты расчета и прогнозирования с использованием разработанного ПО сравнивались с фактическими данными с Ванкорского газонефтяного месторождения. На основании данного сравнения делался вывод: является применение машинного обучения актуальным для определения дебита нефти после ГРП или нет.

Подробное описание и алгоритм расчета дебита нефти после ГРП при помощи традиционных методов представлены в работе [16] и в настоящей статье не приводятся.

Прогнозирование дебита нефти после проведения ГРП с использованием технологий искусственного интеллекта осуществлялось при помощи общедоступных моделей машинного обучения [17]:

- RandomForestRegressor;
- LinearRegressor;
- LightGBMRegressor.

Рассматриваемые модели машинного обучения относятся к библиотеке sklearn языка программирования Python. Задача прогнозирования дебита нефти после проведения ГРП является задачей регрессии – прогнозирование количественного значения на основе других параметров.

Весь процесс работы с данными состоял из следующих основных этапов:

Этап 1. Подбор признаков для обучения модели прогнозирования дебита после проведения операции ГРП. При этом для обучения использовались следующие параметры:

- дебит нефти до ГРП;
- эффективная нефтенасыщенная толщина;
- пластовое давление;
- забойное давление;
- обводненность до ГРП;
- проницаемость;
- вязкость нефти;
- количество проппанта;
- плотность нефти в пластовых условиях;
- дебит нефти после проведения операции ГРП.

Вышеуказанные параметры были выбраны, так как они используются для анализа эффективности ГРП специалистами нефтегазовых компаний. При этом целевым признаком является дебит нефти после проведения операции ГРП.

Обучение модели машинного обучения проводилось следующим образом:

- в рассматриваемом случае каждая скважина до проведения ГРП характеризуется вышеперечисленными параметрами;
- после проведения ГРП на скважине изменяется дебит;
- суть модели машинного обучения заключается в нахождении определенных зависимостей скважинных параметров до ГРП с дебитом нефти после ГРП;
- на основе этих зависимостей модель обучается и может прогнозировать дебит нефти после проведения операции ГРП на основе параметров, которые были выбраны для обучения модели.

Выбранные данные выгружались из различных отчетов по месторождениям, а также из открытых источников. Необходимо отметить, что для обучения модели использовались не только данные Ванкорского месторождения, но и данные других месторождений России из разных регионов: Муравленковское, Приобское, Бавлинское, Игольское-Тавровое, Вынгапуровское и др. В целом модель обучалась на выборке данных из 500 ГРП, проведенных на месторождениях в разных регионах России.

Все данные прошли предобработку: были удалены аномалии и выбросы данных. Аномалии – значения, которые не могут существовать в реаль-

ном мире (к примеру, отрицательный дебит нефти), а выбросы – это значения, которые сильно выбиваются из общего распределения данных (например, дебит нефти по скважине – 500000 т/сут.). Имеющиеся пропуски в данных были заполнены по следующему принципу: заполнение средним значением по конкретному месторождению по отдельному пласту, на который пробурена скважина.

Так как прогнозирование дебита нефти после проведения операции ГРП является задачей регрессии и порядок значений признаков сильно различается, была проведена стандартизация признаков при помощи функции Python MinMaxScaler [18]. Стандартизация признаков – это масштабирование функций таким образом, чтобы они находились между заданным минимальным и максимальным значением, часто между нулем и единицей, или так, чтобы максимальное абсолютное значение каждой функции масштабировалось до размера единицы [18].

Этап 2. Разделение выборки на тренировочную и тестовую в отношении 90:10. Тренировочная выборка служила для обучения модели машинного обучения. Тестовая выборка служила для тестирования модели машинного обучения. Далее проводился поиск наилучшей модели машинного обучения.

Необходимо отметить, что данные из тестовой и тренировочной выборки не должны пересекаться, или будет наблюдаться переобучение модели, которое заключается в том, что модель уже обучалась на данных, которые она должна предсказывать, соответственно, прогнозирование не будет являться объективным.

Для поиска наилучшей модели машинного обучения проводилась ручная кросс-валидация с подбором наилучших гиперпараметров для каждой модели машинного обучения с использованием библиотеки GridSearchCV [19, 20]. Гиперпараметры модели машинного обучения – параметры машинного обучения, значения которых использовались для управления процессом обучения [21]. В качестве таких параметров выступали: время обучения модели, количество итераций, которые происходит обучение модели, и другие. Для каждой модели гиперпараметры отличались. Использование кросс-валидации было обусловлено тем, что она позволяет более точно оценить модели машинного обучения, чем обычная валидация. Суть кросс-валидации заключалась в разделении тренировочной выборки на определенное количество частей, именуемых подвыборками, которое выбирали авторы. После такого разделения модель машинного обучения обучалась несколько раз на различных подвыборках из тренировочной выборки и тестировалась также на различных подвыборках из тренировочной выборки. Каждый раз вычислялась

точность модели по выбранной метрике. После этого вычислялось среднее значение метрики модели. В данной работе тренировочная выборка была разделена на четыре подвыборки.

Кросс-валидации подвергались следующие модели машинного обучения:

- RandomForestRegressor. Этот метод относится к ансамблевым. Цель ансамблевых методов – объединить прогнозы нескольких базовых оценок, построенных с заданным алгоритмом обучения, чтобы улучшить надежность по сравнению с одной оценкой [22].
- LinearRegression. Суть данной модели состоит в том, что модель линейризует набор признаков [23]. Затем на основе линейной зависимости целевого признака от обучающих признаков выполняется прогнозирование для нового набора обучающих признаков, целевой признак которых не известен.
- LightGBMRegressor является библиотекой градиентного бустинга [24]. Суть данной библиотеки заключается в том, что модели задается определенное количество итераций, в течение которых модель будет обучаться. После каждой итерации она вычисляет ошибку по заданной метрике. После этого модель не улучшает прогнозирование, а стремится уменьшить ошибку. Модели градиентного бустинга являются наиболее современными и точными моделями среди моделей машинного обучения [24].

Для оценки результатов обучения использовалась метрика «средняя абсолютная ошибка» (Mean Absolute Error – MAE) из библиотеки sklearn языка программирования Python [25]. Данная метрика использовалась, так как она менее чувствительна к выбросам в данных.

Метрика MAE определяется по формуле (1) в соответствии с работой [25]:

$$\sigma = \frac{\sum_{i=1}^n (X_i - \bar{X})}{n}, \quad (1)$$

где X_i – спрогнозированное значение; $\bar{X}$ – реальное значение; n – количество анализируемых значений.

Этап 3. Выбор наилучшей модели машинного обучения. Выбранная модель была заново обучена на всем тренировочном наборе данных, после чего по данной модели было осуществлено прогнозирование дебита нефти с использованием специального ПО, написанного авторами на языке программирования Python.

Для оценки правильности расчетных данных вычислялась погрешность расчетных и реальных данных по формуле относительной погрешности (2):

$$x = \frac{Q_{\text{реальн}} - Q_{\text{расч}}}{Q_{\text{реальн}}} * 100\%, \quad (2)$$

где x – относительная погрешность вычислений, %; $Q_{\text{реальн}}$ – практический дебит по всем скважинам, т/сут; $Q_{\text{расч}}$ – расчетный дебит по всем скважинам, т/сут.

Далее вычислялась сходимость результатов в соответствии с формулой (3):

$$\beta = 100 - x, \quad (3)$$

где β – сходимость расчетных и практических данных, %.

Результаты

Разработанное ПО состоит из двух частей: первая часть выполняла расчет дебита нефти после ГРП по традиционным одномерным моделям, вторая часть прогнозировала дебит нефти после ГРП при помощи методик машинного обучения.

Результаты расчета дебита нефти после проведения операции ГРП при помощи одномерных моделей представлены в табл. 3.

Таблица 3. Результаты моделирования притока нефти после проведения операции ГРП при помощи одномерных моделей

Table 3. Results of traditional 1D-modeling of oil inflow after hydraulic fracturing operation

Параметр Parameter	Модель расчета Computation model	Скважина/Well	
		724	728
Дебит нефти после операции ГРП, т/сут. Oil flow rate after hydraulic fracturing operation, t/day	PKN	105	126
	KGD	69	83
	Радиальная модель Radial Model	72	86

Прогнозирование дебита нефти после ГРП с использованием методик машинного обучения.

Результаты ручной кросс-валидации представлены в табл. 4.

Таблица 4. Значение метрики MAE

Table 4. MAE metric value

Модель/Model	Значения метрики MAE/MAE metric value
LinearRegression	17,1
LightGBMRegressor	10,5
RandomForestRegressor	15,3

Из табл. 4 очевидно, что модель LightGBMRegressor обладает наилучшим значением метрики. В связи с этим модель LightGBMRegressor была заново обучена на всем тренировочном наборе данных с использованием подобранных гиперпараметров. После этого было выполнено прогнозирование дебита нефти после ГРП на тестовом наборе данных. Итоговое значе-

ние метрики MAE модели LightGBMRegressor составило 9,5. Качество предсказания выросло в 1,1 раз за счет увеличения количества данных для обучения вследствие добавления валидационной выборки для обучения. Время обучения модели составило 90 секунд.

Для проверки адекватности модели была построена константная модель. Константная модель – это модель, которая вне зависимости от входных признаков для прогнозирования всегда выдает неизменный ответ [26]. В рассматриваемом случае константная модель всегда выдавала среднее значение дебита после ГРП по 500 скважинам, на которых обучалась модель. Это делалось для проверки целесообразности применения модели машинного обучения. Если значение метрики MAE константной модели ниже значения метрики модели машинного обучения, то применение последней не считается целесообразным. Значение метрики MAE составило 25, что указывает на актуальность применения модели машинного обучения.

После окончательной проверки модели было выполнено прогнозирование дебита нефти после ГРП на рассматриваемых в данной работе скважинах-кандидатах № 724 и 728, которое составило 67 и 81 т/сут. соответственно.

Для верификации и оценки точности разработанных моделей данные по расчету дебита нефти после ГРП, полученные при помощи всех используемых традиционных моделей, а также данные, полученные с использованием модели машинного обучения LightGBMRegressor сравнивались с фактическими данными по дебиту нефти с Ванкорского месторождения, представленным в открытых источниках [12–14] и приведенными в табл. 2. Отметим, что другие исследованные модели машинного обучения не участвовали в сравнении по причине того, что данные модели использовались с целью подбора наиболее адекватной универсальной модели для прогнозирования дебита нефти после ГРП, обучение которой базируется на различных пластах и различных условиях залегания.

Обсуждение

Традиционный расчет дебита нефти после проведения ГРП с использованием одномерных моделей

Сравнение фактической и расчетной суммы дебитов нефти по двум скважинам для традиционного расчета дебита нефти после проведения ГРП с использованием одномерных моделей приведено на рис. 1. Практические данные по дебиту нефти были взяты из открытых источников [12–14].

В табл. 5 приведены значения относительной погрешности и сходимости для традиционных расчетов дебита нефти после проведения ГРП.

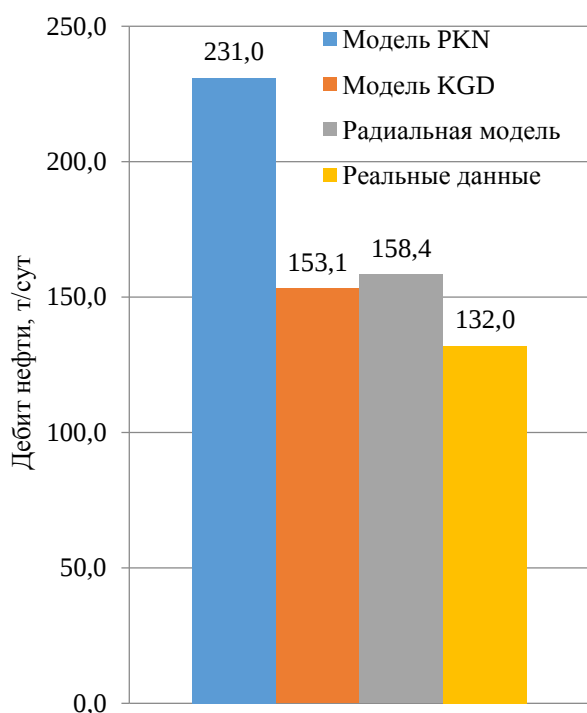


Рис. 1. Сравнение фактических и расчетных значений суммарного дебита нефти в скважинах № 724 и 728 для традиционного расчета дебита нефти после проведения ГРП с использованием одномерных моделей

Fig. 1. Comparison of actual and calculated values of the total oil flow rate in wells no. 724 and 728 for traditional calculation of oil flow rate after hydraulic fracturing using one-dimensional models

Таблица 5. Сходимость фактических и расчетных значений суммарного дебита нефти в скважинах № 724 и 728 для традиционного расчета дебита нефти после проведения ГРП с использованием одномерных моделей

Table 5. Convergence of actual and calculated values of the total oil flow rate in wells no. 724 and 728 for traditional calculation of oil flow rate after hydraulic fracturing using one-dimensional models

Модель Model	Относительная погрешность вычислений Relative error of calculations	Сходимость вычислений Convergence of calculations
	%	
PKN	75	25
KDG	16	84
Радиальная Radial	20	80

По данным, представленным в табл. 5, можно сделать вывод о том, что две из трех исследованных одномерных модели обладают невысокой точностью при расчете дебита нефти после ГРП. По мнению ряда авторов, это может быть связано с допущениями в расчетах, которые в реальных

условиях невозможно воплотить [8–10]. Например, линейно-упругий пласт с однородными свойствами является лишь моделью, как и ньютоновский характер жидкости разрыва. При этом модель KGD все же обеспечила высокую сходимость с фактическими данными. Это указывает на возможность использования модели KGD в разработанном авторами ПО для предварительной оценки дебита нефти после ГРП.

Прогнозирование дебита нефти с использованием моделей машинного обучения

При анализе результатов прогнозирования дебита нефти после проведения ГРП при помощи методик машинного обучения необходимо отметить большую разницу сходимости моделей при кросс-валидации. В первую очередь это объясняется различным качеством моделей. Так, модели LinearRegression и RandomForestRegressor являются классическими моделями, которые, как правило, не отличаются высокой точностью. Тогда как модель LightGBMRegressor является современной моделью градиентного бустинга, характеризующейся высокой точностью прогнозирования [24].

Связь параметров и признаков в модели для обучения в ПО показана в табл. 6.

Таблица 6. Связка параметров и признаков в модели для обучения

Table 6. Linking parameters and features in the model for training

Параметр Sign	Признак в модели Sign in model
Скважина/Well	well
Дебит до операции ГРП Flow rate before hydraulic fracturing operation	q_before
Эффективная нефтенасыщенная толщина Effective oil-saturated thickness	h_neft
Пластовое давление/Reservoir pressure	p_plast
Забойное давление/Bottomhole pressure	p_down
Обводненность/Water cut	watercut
Проницаемость/Permeability	perm
Вязкость нефти /Oil viscosity	viscosity
Количество проппанта/Proppant quantity	proppant
Плотность нефти в пластовых условиях Oil density in reservoir conditions	density
Дебит нефти после ГРП Oil production after hydraulic fracturing	q_after

Для прогнозирования на тестовой выборке была выбрана модель LightGBMRegressor со следующими подобранными гиперпараметрами [27]:

- random_state=12345 (данный параметр отвечает за случайность подобранных для обучения данных);
- n_estimators=30 (данный параметр указывает количество итераций в обучении, то есть сколь-

ко раз модель будет стремиться уменьшить ошибку прогнозирования);

- $\text{max_depth}=15$ (характеризует глубину дерева, то есть через сколько «условий» проходит набор признаков);
- $\text{num_leaves}=20$ (основной параметр для характеристики сложности модели, дословно максимальное количество листьев на дереве, фактически является функцией, зависящей от параметра max_depth);
- $\text{learning_rate}=0,01$ (характеризует скорость обучения модели).

Следует отметить, что, если признаки при машинном обучении сильно коррелируют между собой, это может негативно сказаться на обучении модели машинного обучения, то есть снизится качество прогнозирования [28]. В связи с этим предварительно для анализа корреляции между признаками был проведен heatmap анализ [28]. Результаты данного анализа приведены на рис. 2.

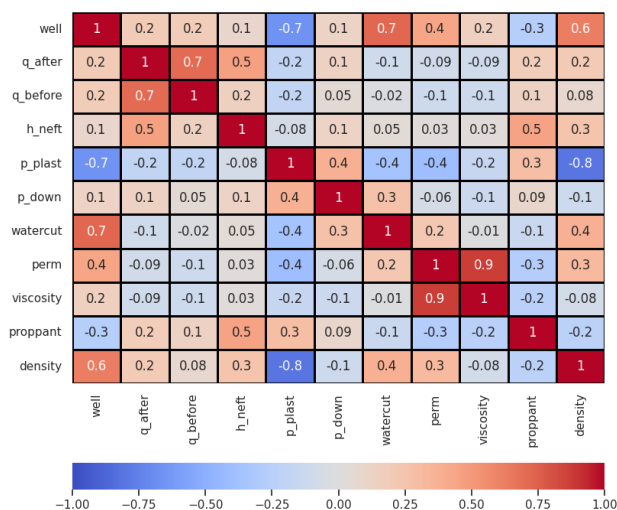


Рис. 2. Heatmap-анализ
Fig. 2. Heatmap analysis

Исходя из анализа рис. 2, было выявлено, что признак «плотность нефти в пластовых условиях» (density) сильно коррелирует с признаком «пластовое давление» (p_plast), степень корреляции составляет 0,8. Поэтому для улучшения качества обучения признак «плотность нефти в пластовых условиях» был удален. Таким образом, модель обучалась на восьми признаках: дебит нефти до ГРП, эффективная нефтенасыщенная толщина, пластовое давление, забойное давление, обводненность до ГРП, проницаемость, вязкость нефти, количество проппанта.

После прогнозирования дебитов нефти после ГРП на выбранных скважинах была вычислена погрешность и сходимость результатов по формулам (2, 3) для сравнения сходимости фактических и рас-

четных значений суммарного дебита нефти в скважинах № 724 и 728 при использовании моделей машинного обучения. Результаты приведены в табл. 7.

Таблица 7. Сходимость фактических и расчетных значений суммарного дебита нефти в скважинах № 724 и 728 при использовании моделей машинного обучения

Table 7. Convergence of actual and calculated values of the total oil flow rate in wells no. 724 and 728 for machine learning models

Модель Model	Относительная погрешность вычислений Relative error of calculations	Сходимость вычислений Convergence of calculations
	%	
LightGBMRegressor	13	87

Также перед обучением проводилось исследование важности признаков при помощи функции `feature_importances_` библиотеки `sklearn` в соответствии с работой [29]. Гистограмма данного исследования представлена на рис. 3.

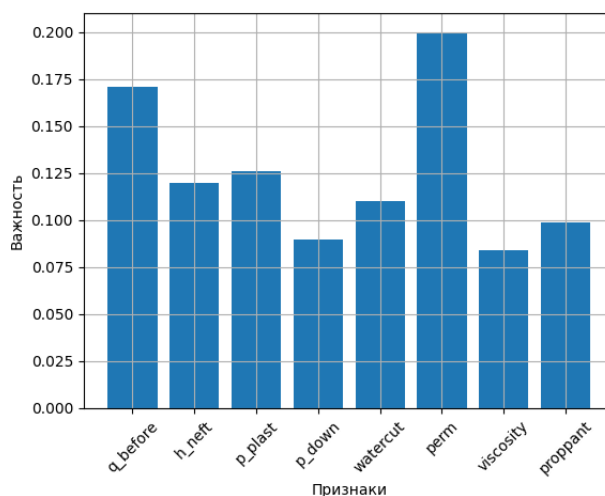


Рис. 3. Исследование важности признаков
Fig. 3. Research of the importance of features

Анализируя рис. 3, можно сделать вывод о том, что наиболее важными признаками для обучения являются дебит нефти до ГРП и проницаемость, а остальные признаки имеют менее важную роль. Наименее важным признаком является вязкость нефти. В будущем данный признак будет удален из модели машинного обучения для увеличения ее производительности.

При исследовательском анализе данных, как было отмечено выше, было сделано допущение о том, что некоторые признаки с пропущенными зна-

чениями заменены средними значениями. Таким образом пропуски заполнялись в столбцах с количеством проппанта, проницаемостью пласта, а также забойным давлением.

Проведенная работа показала, что использование моделей машинного обучения для прогнозирования нефти после ГРП является актуальным, так как модели машинного обучения обладают более высокой точностью по сравнению с традиционными расчетами по одномерным моделям. Модели машинного обучения также позволяют при помощи анализа важности признаков (библиотека `features_importances`) делать заключение о том, какие признаки были важны для обучения, а какие нет [29]. Это позволяет впоследствии использовать лишь те признаки, которые улучшают качество модели для более корректной оценки.

Сравнение результатов прогнозирования дебита при помощи методик машинного обучения и расчетов при помощи традиционных моделей

Итак, были выбраны наилучшие модели машинного обучения (`LightGBMRegressor`) и наилучшая одномерная модель (KGD). Расчетные результаты дебита представлены в табл. 8.

Таблица 8. Сравнение модели машинного обучения и одномерной модели [14]

Table 8. Comparison of machine learning model and one-dimensional model [14]

Название модели Model	Машинное обучение, модель LightGBMRegressor Machine learning, model LightGBMRegressor		Традиционный расчет с использованием одномерной модели KGD Traditional calculation using one-dimensional KGD model		Фактические данные Actual data	
	724	728	724	728	724	728
Номер скважины Well no						
Дебит нефти после ГРП, т/сут. Oil production after hydraulic fracturing, t/day	67	81	69	83	60	72

Сходимость результатов описана выше. В соответствии с моделью KGD сходимость результатов составляет 84 %, в соответствии с моделью `LightGBMRegressor` – 87 %.

Исходя из сходимостей и табл. 8, можно сделать вывод о том, что большей сходимостью обладает модель машинного обучения `LightGBMRegressor`. Однако одномерная модель KGD обладает сходи-

мостью, которая ниже всего на 3 %, что тоже является хорошим результатом.

Также в обеих методиках расчета были сделаны некоторые допущения. Если их устранить, в дальнейшем удастся улучшить качество моделей.

В рамках прогнозирования дебита нефти при помощи методик машинного обучения планируется сделать следующее для улучшения показателя сходимости реальных и смоделированных данных:

1. Необходимо использовать большее количество данных для обучения. На данный момент модель `LightGBMRegressor` обучается на 500 ГРП, что является небольшим количеством данных.
2. Для улучшения качества можно использовать другие методы для заполнения пропусков в данных. Например, использовать метод `knn-ближайших соседей` [28], который базируется на том, что пропуски заполняются с помощью поиска наиболее похожих строк, в которых пропущенные значения являются заполненными.
3. Кроме того, для улучшения качества прогнозирования можно использовать и другие модели градиентного бустинга, к примеру `XGBoost` [24], а также нейронные сети.

В рамках расчетов при помощи традиционных моделей планируется сделать следующее для улучшения показателя сходимости реальных и смоделированных данных:

4. Исключить предположение об отсутствии фильтрационных утечек, которое было сделано в соответствии с низкой проницаемостью пласта.
5. Кроме того, для более точного описания ГРП при помощи традиционных моделей необходимо использовать не только одномерные модели, но и двух- и трехмерные модели распространения трещин.

Подводя итог, следует остановиться на том, что на данный момент модель машинного обучения `LightGBMRegressor` и модель KGD обладают высокой сходимостью и высокой производительностью, поэтому на данный момент обе модели могут использоваться в качестве инструмента для прогнозирования дебита нефти на скважине после проведения операции ГРП.

Заключение

В ходе исследования был рассчитан дебит нефти после проведения ГРП с использованием технологий искусственного интеллекта, а именно методики машинного обучения, а также расчет при помощи традиционных моделей расчета. В ходе произведенного анализа были получены следующие результаты: сходимость дебита нефти после операции ГРП с реальными данными наилучшей модели машинного обучения составила 87 %, а сходимость наилучшей традиционной модели KGD – 84 %. Таким образом более точной методикой является методика прогнозирования дебита при помощи машинного обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Computer simulation of hydraulic fractures / J. Adachi, E. Siebrits, A. Peirce, J. Desroches // *International Journal of Rock Mechanics & Mining Sciences*. – 2007. – Vol. 44. – P. 739–757.
2. Байкин А.Н. Динамика трещины гидроразрыва пласта в неоднородной пороупругой среде: дис. ... канд. физ.-мат. наук. – Новосибирск, 2016. – 94 с.
3. Economides M.J. Modern fracturing – enhancing natural gas production. – Houston: Energy Tribune Publ. Inc., 2007. – 531 p.
4. Dontsov E.V., Peirce A.P. An enhanced pseudo-3D model for hydraulic fracturing accounting for viscous height growth, non-local elasticity, and lateral toughness // *Engineering Fracture Mechanics*. – 2015. – Vol. 142. – P. 116–139.
5. Ямкин М.А., Сафиуллина Е.У. Моделирование механических свойств геологической среды для создания оптимальной трещины гидроразрыва пласта // *Нефтяная провинция*. – 2022. – № 3 (31). – С. 104–118.
6. Интенсификация притока нефти из карбонатных коллекторов для условий месторождений Западной Сибири / А.Д. Румянцев, А.М. Машкова, Н.В. Соловьев, К.О. Щербакова, Б.А. Овезов // *Молодые – наукам о земле: Материалы X Международной научной конференции молодых ученых*. – М.: Российский государственный геологоразведочный университет им. С. Орджоникидзе, 2022. – Т. 4. – С. 279–282.
7. Jincai Zhang, Yushuai Zhang, Shangxian Yin. PKN solution revisit: 3-D hydraulic fracture size and stress anisotropy effects // *Rock Mechanics and Rock Engineering*. – 2017. – Vol. 51. – P. 653–660.
8. Воспроизведение пластового давления методами машинного обучения и исследование его влияния на процесс образования трещин при гидравлическом разрыве пласта / Е.В. Филиппов, Л.А. Захаров, Д.А. Мартюшев, И.Н. Пономарева // *Записки Горного института*. – 2022. – Т. 258. – С. 924–932.
9. Основы технологии гидравлического разрыва пластов / В.Г. Салимов, Ш.Ф. Тахаутдинов, А.В. Насыбуллин, О.В. Салимов. – М.: Изд-во «Фэн» АН РТ, 2021. – 386 с.
10. Юмангуллова Н.З., Ганиева Г.Р., Минханов И.Ф. Эффективность проведения гидроразрыва пласта на Поточном месторождении // *Нефтяная провинция*. – 2021. – № 3 (27). – С. 95–111.
11. Анализ и перспективы эффективной разработки Ванкорского месторождения / Т.В. Всеволодов, М.В. Липаев, Р.Р. Мукминов, М.С. Сасина, Е.В. Егорова // *Новейшие технологии освоения месторождений углеводородного сырья и обеспечение безопасности экосистем Каспийского шельфа: XII Международная научно-практическая конференция: сб. науч. ст.* – Астрахань: Изд-во АГТУ, 2021. – С. 11–16.
12. Арестов А.А. Пути оптимизации разработки месторождений на поздней стадии эксплуатационного цикла // *Проблемы геологии и освоения недр: Материалы XXVII Международного молодежного научного симпозиума имени академика М.А. Усова, посвященного 160-летию со дня рождения академика В.А. Обручева и 140-летию академика М.А. Усова, основателей Сибирской горно-геологической школы*. – Томск: ТПУ, 2022. – С. 34–35.
13. Ямкин М.А., Сафиуллина Е.У., Ямкин А.В. Анализ результатов моделирования притока жидкости к трещине гидроразрыва пласта // *Известия Томского политехнического университета. Инжиниринг георесурсов*. – 2024. – Т. 335. – № 4. – С. 14–21.
14. Григорьев Г.С., Салищев М.В., Сенчина Н.П. О применимости способа электромагнитного мониторинга гидроразрыва пласта // *Записки Горного института*. – 2021. – Т. 250. – С. 492–500.
15. Economides M., Oligney R., Valkó P. Unified fracture design (bridging the gap between theory and practice). – Texas: Orsa Press Alvin, 2004. – 316 p.
16. Ямкин М.А., Сафиуллина Е.У. Оценка соответствия результатов компьютерного моделирования притока жидкости к трещине гидроразрыва пласта реальным данным // *Известия Томского политехнического университета. Инжиниринг георесурсов*. – 2023. – Т. 334. – № 3. – С. 210–217.
17. Hartmann F.G., Kopp J., Lois D. Social science data analysis. – Wiesbaden: Springer Publ., 2023. – 191 p.
18. Miranda F.M., Kohnecke N., Renard B.Y. Hiclass: a python library for local hierarchical classification compatible with scikit-learn // *Journal of Machine Learning Research*. – 2023. – № 24 (29). URL: <https://www.jmlr.org/> (дата обращения 19.09.2024).
19. Hyperparameter tuning using GridsearchCV on the comparison of the activation function of the ELM method to the classification of pneumonia in toddlers / Pirjatullah, Dwi Kartini, Dodon Turianto Nugrahadi, Muliadi, Andi Farmadi. – Jakarta: IEEE Xplore, 2021. – 110 p.
20. Ranjan G.S.K., Verma A.K., Radhika S. K-nearest neighbors and grid search CV based real time fault monitoring system for industries. – Pune: International Conference on Natural Computation, 2019. – 160 p.
21. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques / Z.M. Alhakeem, Y.M. Jebur, S.N. Henedy, H. Imran, L.F.A. Bernardo, H.M. Hussein // *Materials*. – 2022. – Vol. 2. – № 15 (21). URL: <https://www.mdpi.com/1996-1944/15/21/7432> (дата обращения 19.09.2024).
22. Yanli Liu, Yourong Wang, Jian Zhang. New machine learning algorithm: Random Forest // *Information Computing and Applications: Materials International Conference on Information Computing and Applications*. – 2012. – Vol. 7473. – P. 246–252.
23. An introduction to decision tree modeling / A.J. Myles, R.N. Feudale, Y. Liu, N.A. Woody, S.D. Brown // *Journal of Chemometrics*. – 2004. – Vol. 18. – P. 275–285. URL: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.873> (дата обращения 19.09.2024).
24. Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth / A. Humphrey, W. Kuberski, J. Bialek, N. Perrakis, W. Cools, N. Nuytens, H. Elakhrass, P.A.C. Cunha // *Monthly Notes*. – 2022. – Vol. 517. – P. 116–120. URL: <https://academic.oup.com/mnras/article-abstract/517/1/L116/6761704> (дата обращения 19.09.2024).
25. Stop oversampling for class imbalance learning: a review / A.S. Tarawneh, A.B. Hassanat, G.A. Altarawhen, A. Almuhammad // *IEEE Access*. – 2022. – Vol. 10. – P. 47643–47659.
26. Abd Elrahman S.M., Abraham A. A review of class imbalance problem // *Journal of Network and Innovative Computing*. – 2008. – Vol. 1. – P. 332–340.
27. On the role and the importance of features for background modeling and foreground detection / Th. Bouwmans, C. Silva, C. Marghes, M.S. Zitouni, H. Bhaskar, C. Frelicot // *Computer Science Review*. – 2018. – Vol. 28. – P. 26–91.

28. Машанский А. Метод k-ближайших соседей (k-nearest neighbour). URL: <https://proglab.io/p/metod-k-blizhayshih-sosedey-k-nearest-neighbour-2021-07-19> (дата обращения 01.10.2024).
29. Ямкин М.А., Сафиуллина Е.У., Ямкин А.В. Отбор скважин-кандидатов при обработке призабойной зоны пласта методами машинного обучения // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2024. – Т. 335. – № 5. – С. 7–16.

Информация об авторах

Максим Александрович Ямкин, студент, факультет программной инженерии, Университет ИТМО, Россия, 197101, г. Санкт-Петербург, Кронверкский пр., 49, лит. А, 2. makson.yamkin@mail.ru

Елена Улубековна Сафиуллина, кандидат технических наук, доцент, кафедра разработки и эксплуатации нефтяных и газовых месторождений, Санкт-Петербургский горный университет императрицы Екатерины II, Россия, 199106, г. Санкт-Петербург, 21-я лин. В.О., 2. safiullinaeu@yandex.ru

Александр Владимирович Ямкин, заместитель начальника технического отдела, «ООО Газпром трансгаз Томск», Россия, 634029, г. Томск, пр. Фрунзе, 9. A.Yamkin@gtt.gazprom.ru

Поступила в редакцию: 15.10.2024

Поступила после рецензирования: 25.10.2024

Принята к публикации: 30.06.2025

REFERENCES

1. Adachi J., Siebrits E., Peirce A., Desroches J. Computer simulation of hydraulic fractures. *International Journal of Rock Mechanics & Mining Sciences*, 2007, vol. 44, pp. 739–757.
2. Baykin A.N. *Dynamics of hydraulic fracture in an inhomogeneous poroelastic medium*. Cand. Diss. Abstract. Novosibirsk, 2016. 94 p. (In Russ.)
3. Economides M.J. *Modern fracturing – enhancing natural gas production*. Houston, Energy Tribune Publ. Inc., 2007. 531 p.
4. Dontsov E.V., Peirce A.P. An enhanced pseudo-3D model for hydraulic fracturing accounting for viscous height growth, non-local elasticity, and lateral toughness. *Engineering Fracture Mechanics*, 2015, vol. 142, pp. 116–139.
5. Yamkin M.A., Safiullina E.U. Modeling the mechanical properties of the geological environment to create an optimal hydraulic fracture. *Oil province*, 2022, no. 3 (31), pp. 104–118. (In Russ.)
6. Rumyantsev A.D., Mashkova A.M., Solovyov N.V., Shcherbakova K.O., Ovezov B.A. Intensification of oil inflow from carbonate reservoirs for the conditions of fields in Western Siberia. *Young for geosciences. Proceedings of the X International Scientific Conference of Young Scientists*. Moscow, Russian State Geological Prospecting University named after. S. Ordzhonikidze Publ., 2022. Vol. 4, pp. 279–282. (In Russ.)
7. Jincai Zhang, Yushuai Zhang, Shangxian Yin. PKN solution revisit: 3-D hydraulic fracture size and stress anisotropy effects. *Rock Mechanics and Rock Engineering*, 2017, vol. 51, pp. 653–660.
8. Fillipov E.V., Zakharov L.A., Martyushev D.A., Ponomareva I.N. Reproducing reservoir pressure using machine learning methods and studying its influence on the process of crack formation during hydraulic fracturing. *Notes of the Mining Institute*, 2022, vol. 258, pp. 924–932. (In Russ.)
9. Salimov V.G., Takhautdinov Sh.F., Nasybullin A.V., Salimov O.V. *Fundamentals of hydraulic fracturing technology*. Kazan, Feng AN RT Publ., 2021. 386 p. (In Russ.)
10. Yumangullova N.Z., Ganiyeva G.R., Minkhanov I.F. Efficiency of hydraulic fracturing at the Potochnoye field. *Neftyanaya provintsia*, no. 3 (27), pp. 95–111. (In Russ.)
11. Vsevolodov T.V., Lipayev M.V., Mukminov R.R., Sasina M.S., Egorova E.V. Analysis and prospects for the effective development of the Vankor field. *Latest technologies for the development of hydrocarbon deposits and ensuring the safety of the ecosystems of the Caspian shelf. XII International Scientific and Practical Conference*. Astrakhan, ASTU Publ. House, 2021. pp. 11–16. (In Russ.)
12. Arestov A.A. Ways to optimize field development at the late stage of the operational cycle. *Problems of geology and subsoil development. Proc. of the XXVII International Youth Scientific Symposium named after Academician M.A. Usov, dedicated to the 160th anniversary of the birth of Academician V.A. Obruchev and the 140th anniversary of Academician M.A. Usov, the founders of the Siberian Mining and Geological School*. Tomsk, TPU Publ., 2022. pp. 34–35. (In Russ.)
13. Yamkin M.A., Safiullina E.U., Yamkin A.V. Analysis of the results of modeling fluid inflow to a hydraulic fracturing fracture. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2024, vol. 335, no. 4, pp. 14–21. (In Russ.)
14. Grigoriev G.S., Salishchev M.V., Senchina N.P. About the applicability of the method of electromagnetic monitoring of hydraulic fracturing. *Notes of the Mining Institute*, 2021, vol. 250, pp. 492–500. (In Russ.)
15. Economides M., Oligney R., Valkó P. *Unified fracture design (bridging the gap between theory and practice)*. Texas, Orsa Press Alvin, 2004. 316 p.
16. Yamkin M.A., Safiullina E.U. Estimation of compliance of the results of computer simulation of fluid inflow to the fracture of hydraulic fracturing with real data. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2023, vol. 334, no. 3, pp. 210–217. (In Russ.)
17. Hartmann F.G., Kopp J., Lois D. *Social science data analysis*. Wiesbaden, Springer Publ., 2023. 191 p.
18. Miranda F.M., Kohnecke N., Renard B.Y. Hiclass: a python library for local hierarchical classification compatible with scikit-learn. *Journal of Machine Learning Research*, 2023, no. 24 (29). Available at: <https://www.jmlr.org/> (accessed 19 September 2024).
19. Pirjatullah, Dwi Kartini, Dodon Turianto Nugrahadhi, Muliadi, Andi Farmadi. *Hyperparameter tuning using GridsearchCV on the comparison of the activation function of the ELM method to the classification of pneumonia in toddlers*. Jakarta, IEEE Xplore Publ., 2021. 110 p.

20. Ranjan G.S.K., Verma A.K., Radhika S. K-nearest neighbors and grid search CV based real time fault monitoring system for industries. *International Conference on Natural Computation*. Pune, IEEE Xplore Publ., 2019. 160 p.
21. Alhakeem Z.M., Jebur Y.M., Henedy S.N., Imran H., Bernardo L.F.A., Hussein H.M. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with GridSearchCV hyperparameter-optimization techniques. *Materials*, 2022, vol. 2, no. 15 (21). Available at: <https://www.mdpi.com/1996-1944/15/21/7432> (accessed 19 September 2024).
22. Yanli Liu, Yourong Wang, Jian Zhang. New machine learning algorithm: Random Forest. *Information Computing and Applications. Materials International Conference on Information Computing and Applications*. China, Springer Publ., 2012. Vol. 7473, pp. 246–252.
23. Myles A.J., Feudale R.N., Liu Y., Woody N.A., Brown S.D. An introduction to decision tree modeling. *Journal of Chemometrics*, 2004, vol. 18, pp. 275–285. Available at: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.873> (accessed 19 September 2024).
24. Humphrey A., Kuberski W., Bialek J., Perrakis N., Cools W., Nuytens N., Elakhrass H., Cunha P.A.C. Machine-learning classification of astronomical sources: estimating F1-score in the absence of ground truth. *Monthly Notes*, 2022, vol. 517, pp. 116–120. Available at: <https://academic.oup.com/mnrasl/article-abstract/517/1/L116/6761704> (accessed 19 September 2024).
25. Tarawneh A.S., Hassanat A.B., Altarawhen G.A. Stop oversampling for class imbalance learning: a review. *IEEE Access*, 2022, vol. 10, pp. 47643–47659.
26. Shaza M. Abd Elrahman, Ajith Abraham. A review of class imbalance problem. *Journal of Network and Innovative Computing*, 2008, vol. 1, pp. 332–340.
27. Bouwmans Th., Silva C., Marghes C., Zitouni M.S., Bhaskar H., Frelicot C. On the role and the importance of features for background modeling and foreground detection. *Computer Science Review*, 2018, vol. 28, pp. 26–91.
28. Mashansky A. *K-nearest neighbour*. (In Russ.) Available at: <https://proglib.io/p/metod-k-blizhayshih-sosedey-k-nearest-neighbour-2021-07-19> (accessed 01 October 2024).
29. Yamkin M.A., Safiullina E.U., Yamkin A.V. Selection of candidate wells when processing the bottomhole formation zone using machine learning methods. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2023, vol. 335, no. 5, pp. 7–16. (In Russ.)

Information about the authors

Maxim A. Yamkin, Student, ITMO University, 49, lit. A, 2, Kronverksky avenue, St Petersburg, 197101, Russian Federation. makson.yamkin@mail.ru

Elena U. Safiullina, Cand. Sc., Associate Professor, St Petersburg Mining University of Empress Catherine II, 2, 21st lin. V.O., St Petersburg, 199106, Russian Federation. safiullinaeu@yandex.ru

Alexander V. Yamkin, Deputy Head of Technical Department, «Gazprom Transgaz Tomsk LLC», 9, Frunze avenue, Tomsk, 634029, Russian Federation. A.Yamkin@gtt.gazprom.ru

Received: 15.10.2024

Revised: 25.10.2024

Accepted: 30.06.2025