

Опыт применения алгоритмов машинного обучения для автоматизации процесса поиска скважин-кандидатов для гидравлического разрыва пласта

Галеев А.А., Синицына Т.И.

ООО «Тюменский нефтяной научный центр», Тюмень, Россия
aagaleev@tnnc.rosneft.ru

Аннотация

В условиях постоянно растущих объемов данных по месторождениям актуальным становится вопрос автоматизации процесса обработки и структуризации знаний об изучаемом объекте разработки с последующим выделением основных характеристик, позволяющих выбирать и обосновывать методы повышения нефтеотдачи на месторождениях. Одним из таких методов является проведение геолого-технических мероприятий (ГТМ), в частности, гидравлический разрыв пласта (ГРП).

В данной работе поставлена цель поиска оптимального решения задачи выбора и обоснования скважин-кандидатов для ГРП на примере месторождения имени Малыка. На данном месторождении насчитывается более тысячи скважин по трем эксплуатационным объектам (более двух тысяч пластопересечений). Ежегодно проводится, в среднем, около 50 операций ГРП и поиск оптимальных скважин — это важная задача для инженеров-разработчиков.

Для выбора оптимального метода поиска скважин-кандидатов на ГРП в текущем исследовании выполнен сравнительный анализ существующих подходов к оценке продуктивности скважин. Полученные результаты позволили сделать вывод о применимости гибридной модели, показавшей более высокую точность прогноза в сравнении с другими подходами. В основе такой модели лежит концепция физически-информированного машинного обучения.

Основные этапы работы включали в себя разработку автоматизированной системы сбора и подготовки геолого-промысловой информации, тестирование различных моделей машинного обучения и сравнение с результатами физико-математического моделирования.

В результате выполненной задачи подобран оптимальный алгоритм прогнозирования продуктивности скважин после ГРП, который применим для оперативных расчетов по большому количеству скважин. Разработанный алгоритм включен в систему поиска скважин-кандидатов на ГРП, реализованную на базе языков программирования VBA и Python. На текущий момент выполняется опытно-промышленная апробация результатов работы.

Материалы и методы

Поставленные задачи решаются с применением физико-математического моделирования и алгоритмов машинного обучения. Для автоматизации сбора и предобработки промысловых данных, обучения ML-моделей и расчета

ожидаемых параметров работы скважин используются средства программирования VBA и Python.

Ключевые слова

геолого-технические мероприятия, автоматизация, гидравлический разрыв пласта, машинное обучение

Для цитирования

Галеев А.А., Синицына Т.И. Опыт применения алгоритмов машинного обучения для автоматизации процесса поиска скважин-кандидатов для гидравлического разрыва пласта // Экспозиция Нефть Газ. 2024. № 7. С. 70–75. DOI: 10.24412/2076-6785-2024-7-70-75

Поступила в редакцию: 13.09.2024

OIL PRODUCTION

UDC 622.276 | Original Paper

Lessons learned in applying machine learning algorithms to automate the process of selecting candidate wells for hydraulic fracturing

Galeev A.A., Sinitsyna T.I.

“Tyumen petroleum research center” LLC, Tyumen, Russia
aagaleev@tnnc.rosneft.ru

Abstract

In the context of constantly growing field data volumes, the issue of automating the processing and structuring knowledge about the development target being studied, followed by identifying the main properties that allow selecting and justifying EOR methods, becomes relevant. One of these methods cover well interventions (WI), in particular hydraulic fracturing (frac).

This study sets the goal of finding an optimal solution to the problem of selecting and justifying candidate wells for hydraulic fracturing using a case study of the Malyk field. Over a thousand wells within three production targets (over 2 thousand reservoir intersections) have been drilled in this field. On average, about 50 hydraulic fracturing jobs are carried out annually, and selecting optimal candidate wells is an important task for reservoir engineers.

To find an optimal method for selecting candidate wells for hydraulic fracturing, a comparative analysis of existing approaches to assessing well productivities has been completed within this study. The results pointed out a hybrid model which showed higher forecast accuracy in comparison with other approaches. This model is based on a physics-informed machine learning concept.

The main stages of the study included the development of an automated system for collecting and treatment of geological and field information, testing various machine-learning models and comparison with the results of physical and mathematical modeling.

The study allowed to select an optimal algorithm for predicting post-frac well productivity which is applicable for quick model runs covering large numbers of wells. The developed algorithm is included in the system for selecting candidate wells for hydraulic fracturing, implemented based on the VBA and Python programming languages. Currently, the study results are at the pilot testing phase.

Materials and methods

The tasks are solved via physical and mathematical modeling and machine learning algorithms. VBA and Python programming tools are applied to automate the collection and pre-processing of field

data, to train ML models, and to estimate the expected well operation parameters.

Keywords

well interventions, automation, hydraulic fracturing, machine learning

For citation

Galeev A.A., Sinityna T.I. Lessons learned in applying machine learning algorithms to automate the process of selecting candidate wells for hydraulic fracturing. Exposition Oil Gas, 2024, issue 7, P. 70–75. (In Russ). DOI: 10.24412/2076-6785-2024-7-70-75

Received: 13.09.2024

Введение

Автоматизация процесса поиска геолого-технических мероприятий (ГТМ) является популярным направлением среди специалистов по разработке нефтяных и газовых месторождений. Данное направление развивается параллельно с ростом вычислительных мощностей компьютеров и накоплением значительных массивов информации, которые требуют принципиально другого уровня обработки. Как следствие, возникает необходимость в развитии систем оперативного принятия решений в условиях постоянно растущих объемов данных.

Существующий подход к поиску и обоснованию скважин-кандидатов для ГРП в большинстве случаев сводится к экспертному анализу. Большой объем геолого-промысловой информации обрабатывается в ручном режиме, скважины рассматриваются точно, а качество оценки геологических рисков напрямую зависит от опыта специалиста. Очевидными недостатками такого подхода являются высокие трудозатраты и необъективность получаемых результатов. Сложность оценки продуктивности скважины после ГРП приводит к необходимости использования гидродинамических симуляторов, которые дают наиболее точный и физически обоснованный результат. Однако высокие трудозатраты на создание и актуализацию гидродинамической модели, а также длительность расчетов не позволяют в полной мере автоматизировать данный процесс и применить его в оперативной работе.

Альтернативным подходом является машинное обучение (ML). В отличие от физического моделирования, ML-модель прогнозирует значение целевой переменной на основе фактических данных, находя

сложные закономерности без предоставления их явной формы. Такой подход значительно упрощает построение модели и имеет высокую скорость расчета. Однако основными недостатками ML-моделей являются их низкая интерпретируемость и отсутствие учета физической составляющей процесса. Для более широкого применения алгоритмов машинного обучения в инженерных расчетах необходимо не только создание интерпретируемых решений, но и интеграция с существующими физико-математическими моделями.

Основная часть

Разработанная система автоматизирует весь цикл сбора и анализа геолого-промысловой информации, а также производит расчет основных показателей разработки в разрезе каждого пласта (рис. 1). Учитываются исторические данные скважины: добыча и закачка, проведенные ГТМ, исследования и т.д. В результате была создана наиболее полная, автоматически обновляемая база данных по всему фонду скважин.

Из полученного массива данных система проводит первичный отбор пар скважина-пласт на основе экспертных ограничений. Критерии отбора могут быть индивидуальны для каждого месторождения в зависимости от геологических параметров продуктивных пластов, конструктивных особенностей скважин, номенклатуры применяемого насосного оборудования, результатов ранее проведенных ГТМ и т.д. Основная задача, решаемая на данном этапе, — это автоматизация принятия типовых решений при отборе скважин-кандидатов.

Для месторождения имени Малыка в качестве критериев отбора задаются минимальные значения остаточных извлекаемых

запасов, расстояния до фронта нагнетания и компенсации отборов. Исключаются пласто-пересечения, по которым ранее проводились неэффективные ГРП (как по целевой скважине, так и по скважинам окружения).

На заключительном этапе производится расчет ожидаемых параметров работы скважин после ГРП: обводненность, коэффициент продуктивности, глубина спуска насосного оборудования, потенциальное забойное давление и дебит жидкости. В результате пользователь получает перечень скважин и пластов с наибольшей потенциальной эффективностью проведения ГРП.

Для прогноза обводненности используется классическая модель машинного обучения на основе градиентного бустинга. Обучающая выборка формируется непосредственно из загруженной базы данных, что позволяет в автоматическом режиме переобучать модель с учетом актуального состояния разработки.

Подбор глубины спуска насоса осуществляется по данным инклинометрии и конструкции скважины, затем рассчитывается целевое забойное давление с учетом максимального содержания свободного газа на приеме насоса.

Для прогноза дебита жидкости после ГРП проведено сравнение различных методов, которые можно разделить на три группы:

1. Физико-математическое моделирование.
2. Модели, основанные на данных (Data Driven-подход).
3. Гибридные модели.

Физико-математическое моделирование — это стандартный подход, основанный на физике рассматриваемого процесса, где точность и скорость расчета напрямую зависят от сложности модели. Для задач автоматизации в данной работе используется упрощенная математическая модель притока к скважине [7], которая обеспечивает удовлетворительную скорость вычислений на больших массивах данных.

Модели, основанные на Data Driven-подходе, характеризуются высокой скоростью расчета и гибкостью настройки на фактические данные, но в некоторых случаях могут давать нефизические результаты. Процесс обучения таких моделей заключается в сборе и подготовке данных, получаемых непосредственно с погружных датчиков, приборов учета, результатов лабораторных исследований и т.д. На основе собранных данных формируется обучающая выборка, которая в дальнейшем подается на вход модели машинного обучения.

Гибридные методы сочетают в себе преимущества предыдущих подходов и имеют различные варианты реализации. Совокупность таких методов получила название физически информированного машинного обучения, или PIML (Physics-informed Machine Learning). Гибридные модели можно условно

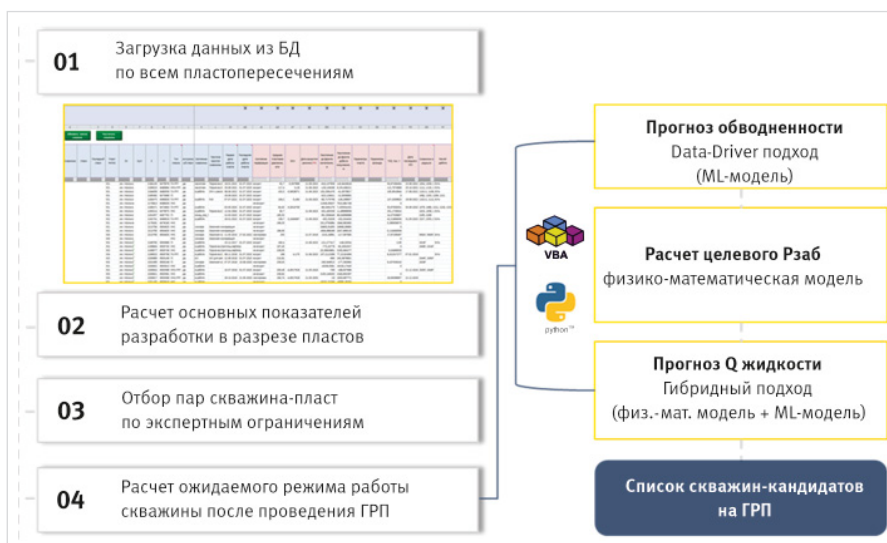


Рис. 1. Общая схема системы поиска скважин-кандидатов на ГРП
Fig. 1. A block diagram for selecting candidate wells for hydraulic fracturing

классифицировать по способу внедрения физических знаний:

1. Обучение физике на основе данных, напрямую воплощающих собой закономерности, присущие реальной системе.
2. Внедрение физики в архитектуру модели.
3. Добавление знаний о физике в функцию потерь модели (Physics-informed Neural Network).

В данной работе рассмотрены два варианта построения гибридных моделей для прогноза дебита жидкости скважин после ГРП, относящиеся к первой группе. Оба варианта принимают за основу решения базовой физико-математической модели, а задачей машинного обучения является снижение ошибки расчета базовой модели.

В первом варианте используется ансамблевая модель на основе метода стекинга (рис. 2). Основная идея данного метода состоит в том, чтобы объединить несколько разнородных моделей в одну линейную метамодель и получить взвешенный прогноз всего ансамбля. Веса настраиваются с помощью алгоритма линейной регрессии. В нашем случае в ансамбль включены три модели машинного обучения (градиентный бустинг, случайный лес, метод опорных векторов), а также базовая модель притока к скважине. Ансамблирование на основе стекинга обеспечивает более высокую точность прогноза в сравнении с отдельными моделями машинного обучения, а также частично решает проблему аномальных прогнозов: если одна из ML-моделей в ансамбле дает аномальный результат, остальные частично корректируют его, при этом общий прогноз ансамбля остается на удовлетворительном уровне.

Из основных недостатков такого подхода стоит отметить сложность интерпретации и внедрения в бизнес-процессы. Кроме

того, базовая физико-математическая модель в ансамбле должна иметь удовлетворительную точность расчета, в противном случае включение такой модели в ансамбль нецелесообразно.

Ввиду указанных ограничений имеет смысл рассмотреть альтернативный вариант построения гибридной модели (рис. 3). В данном варианте за основу принята одна модель машинного обучения, в которой в качестве дополнительных признаков генерируются результаты расчета базовой физико-математической модели. Такой подход значительно сокращает время обучения, имеет хорошую интерпретируемость, а также более прост в реализации и дальнейшей поддержке. При этом оба алгоритма имеют примерно одинаковую точность прогноза.

Генерация дополнительных признаков для гибридной модели

Ввиду того, что основная задача машинного обучения в таком алгоритме — это снижение ошибки расчета базовой модели, то для формирования обучающей выборки необходимо предварительно сгенерировать решения базовой физико-математической модели по действующему добывающему фонду скважин. Из обучающей выборки исключаются скважины с двумя и более работающими пластами для того, чтобы избежать неточности в оценке текущей обводненности и продуктивности по каждому пласту отдельно. В качестве исходных данных для расчета принимаются фактические параметры работающей скважины (объемная обводненность f_w , давление на приеме насоса P_{in} или динамический уровень H_{liq} , пластовое давление P_r , общий скин-фактор S , свойства пласта и флюида, конструкция скважины и т.д.). Таким образом, задача сводится к моделированию текущего режима работы по некоторому

количеству скважин (в зависимости от месторождения) с последующим обучением модели машинного обучения на полученных данных, где целевой переменной будет фактический коэффициент продуктивности. Общий порядок расчета приведен ниже.

Рассчитывается эффективная вязкость жидкости в пластовых условиях μ_{liq} , учитывающая различие в проводимости нефти и воды:

$$\mu_{liq} = \frac{\mu_w \times \mu_o}{k_{rw}(S_w) \times \mu_o + k_{ro}(S_w) \times \mu_w}. \quad (1)$$

Общая сжимаемость системы c_t задается выражением:

$$c_t = c_o \times (1 - S_w) + c_w \times S_w + c_r. \quad (2)$$

Эффективный объемный коэффициент жидкости B_{liq} определяется следующим образом:

$$B_{liq} = B_o \times (1 - f_w) + B_w \times f_w, \quad (3)$$

где μ_w — вязкость воды в пластовых условиях, мПа·с; μ_o — вязкость нефти в пластовых условиях, мПа·с; k_{ro} — ОФП по нефти, д.ед.; k_{rw} — ОФП по воде, д.ед.; S_w — текущая водонасыщенность, д.ед.; C_o — сжимаемость нефти в пластовых условиях, 1/атм; C_w — сжимаемость воды в пластовых условиях, 1/атм; C_r — сжимаемость породы, 1/атм; B_o — объемный коэффициент расширения нефти, м³/м³; B_w — объемный коэффициент расширения воды, м³/м³; f_w — текущая объемная обводненность продукции, д.ед.

Далее производится расчет коэффициента продуктивности J с использованием математической модели притока к скважине [7]. Для учета текущего состояния призабойной зоны пласта используется скин-фактор, полученный по результатам гидродинамических исследований.

Для расчета текущего забойного давления механизированной скважины P_{wf} требуется проведение гидравлического расчета [2–4], который в итоге сводится к построению профиля давления в скважине:

$$P_{wf} = P_c + \Delta P_0 + \Delta P_1 + \Delta P_2, \quad (4)$$

где P_c — затрубное давление на устье скважины, атм; ΔP_0 — давление столба газа в затрубном пространстве, атм; ΔP_1 — давление газированного столба нефти в затрубном пространстве над приемом насоса, атм; ΔP_2 — потери давления при движении газожидкостной смеси под насосом, атм.

Дебит жидкости Q_{liq} определяется исходя из рассчитанных значений коэффициента продуктивности J и забойного давления P_{wf} в соответствии с [6]:

дебит в точке насыщения:

$$Q_b = J \times (P_r - P_b), \quad (5)$$

максимальный дебит при 100 % нефти:

$$Q_{omax} = Q_b + \frac{J \times P_b}{1,8}, \quad (6)$$

давление для комбинированного уравнения притока при дебите Q_{omax} :

$$P_{wfg} = f_w \times \left(P_r - \frac{Q_{omax}}{J} \right), \quad (7)$$

если $f_w = 1$ и $P_{wfg} > P_b$, то:

$$Q_{liq} = J \times (P_r - P_{wfg}). \quad (8)$$

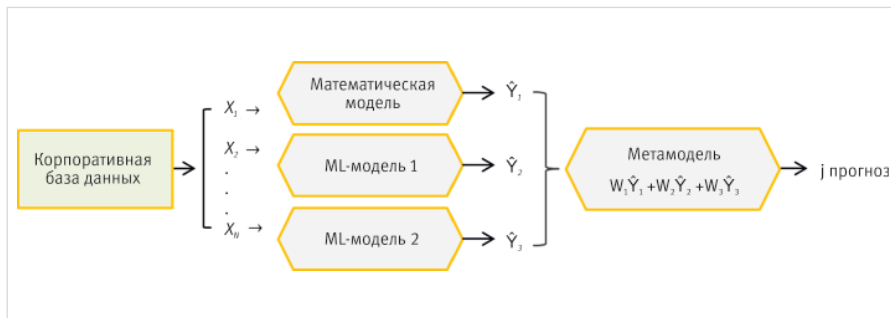


Рис. 2. Ансамблевая модель на основе стекинга
Fig. 2. Stacking-based ensemble model

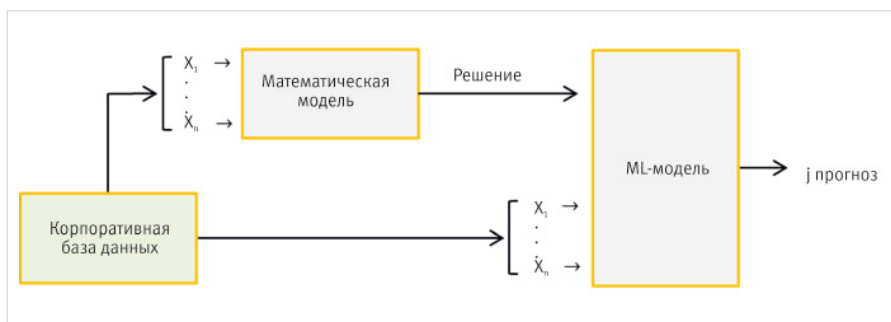


Рис. 3. Модель с генерацией дополнительных признаков с помощью физико-математической модели
Fig. 3. A model with the generation of additional features using a physical and mathematical model

если $f_w < 1$ и $P_{wf} > P_{wfg}$, то:

$$A = 1 + \frac{P_{wf} - (f_w \times P_r)}{0,125 \times f_o \times P_b}, \quad (9)$$

$$B = \frac{f_w}{0,125 \times f_o \times P_b \times J}, \quad (10)$$

$$C = 2 \times A \times B + \frac{80}{Q_{o\max} - Q_b}, \quad (11)$$

$$D = A^2 - \left(80 \times \frac{Q_b}{Q_{o\max} - Q_b} \right) - 81. \quad (12)$$

Если $B = 0$, то:

$$Q_{liq} = \left| \frac{D}{C} \right|. \quad (13)$$

Если $B \neq 0$, то:

$$Q_{liq} = \frac{-C + \sqrt{C^2 - 4 \times B^2 \times D}}{2 \times B^2}. \quad (14)$$

Если $f_w < 1$ и $P_{wf} \leq P_{wfg}$, то:

$$CG = 0,001 \times Q_{o\max}, \quad (15)$$

$$CD = f_w \times \left(\frac{CG}{J} \right) + f_o \times 0,125 \times \times P_b \times \left(-1 + \sqrt{1 + 80 \times \frac{0,001 \times Q_{o\max}}{Q_{o\max} - Q_b}} \right), \quad (16)$$

$$Q_{liq} = \frac{P_{wfg} - P_{wf}}{\frac{CD}{CG} + Q_{o\max}}, \quad (17)$$

где: P_b — давление насыщения нефти в пластовых условиях, атм; f_o — текущая доля нефти в продукции, д.ед.

Расчеты выполняются по всем скважинам, входящим в обучающую выборку, затем производится обучение ML-модели для прогноза коэффициента продуктивности. В качестве независимых параметров модели используются как данные, загружаемые из корпоративной базы данных, так и рассчитанные параметры (μ_{liq} , P_{wf} , J , Q_{liq}). Зависимая переменная — это коэффициент продуктивности, рассчитанный от фактических параметров работы скважины.

Для построения гибридной модели в данной работе используется алгоритм градиентного бустинга над решающими деревьями, который обычно показывает наилучшую точность на выборках с неоднородными табличными данными. Данный алгоритм представляет собой ансамблевую модель, в основе которой лежит итеративное обучение деревьев решений с целью минимизировать функцию потерь.

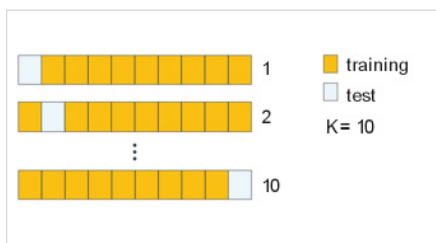


Рис. 4. Разделение выборки по методу k-Fold кросс-валидации
Fig. 4. Sample separation using the k-Fold cross-validation method

Прогноз параметров работы скважины после ГРП

Для получения прогнозных параметров работы скважины после проведения ГРП производятся аналогичные расчеты μ_{liq} , c_p , B_{liq} в пересчете на ожидаемую обводненность по выражениям (1, 2, 3), а также коэффициента продуктивности J с учетом новых значений эффективной мощности пласта и скин-фактора после ГРП.

Далее по данным инклинометрии и конструкции скважины подбирается оптимальная глубина спуска насоса, рассчитывается целевое забойное давление P_{wf} и дебит жидкости Q_{liq} с учетом допустимого содержания свободного газа на приеме насоса. Рассчитанные параметры подаются на вход обученной модели машинного обучения, в результате чего будет получено новое значение коэффициента продуктивности J . Для согласования всех параметров системы расчеты целевого забойного давления P_{wf} , дебита жидкости Q_{liq} и коэффициента продуктивности по ML-модели проводятся итеративно до достижения заданной точности.

Поскольку обучение модели производится на действующих скважинах, по которым также впоследствии необходимо получить прогноз дебита жидкости после ГРП, то для исключения «утечки данных» используется последовательное разделение выборки на заданное количество частей по методу k-Fold кросс-валидации (рис. 4):

- пользователь задает целое число k , меньшее числа строк в обучающей выборке;
- обучающая выборка разбивается на k одинаковых частей (фолдов);

- производится k итераций, во время каждой из которых по одному фолду рассчитывается прогнозный дебит жидкости после ГРП по описанному выше алгоритму, а оставшаяся часть данных выступает в качестве обучающей выборки для k -ой модели.

Количество фолдов выбирается пользователем, исходя из требований к скорости и точности расчета, а также размера обучающей выборки. Более высокая точность прогноза достигается при максимальном значении k ($k = \text{размер выборки} - 1$), но при этом будет произведено k итераций обучения модели, что может значительно увеличить время расчета на больших выборках.

Для месторождения имени Малыка общий размер выборки составил 460 строк и 18 признаков. На рисунке 5 приведена оценка важности параметров обученной модели с использованием Python-библиотеки SHAP [10]. Наибольшее влияние на прогноз модели оказывают параметры, сгенерированные с помощью базовой физико-математической модели: «Дебит жидкости (базовая модель)», «Кэф. продуктивности (базовая модель)».

Описанный алгоритм реализован на базе Microsoft Excel с использованием языков программирования VBA и Python. Загрузка геолого-промысловой информации из корпоративной базы данных осуществляется в автоматическом режиме с помощью SQL-запросов. При разработке моделей машинного обучения применяется стандартный набор Python-библиотек: «Numpy», «Pandas», «Scikit-learn», «CatBoost», «Optuna».

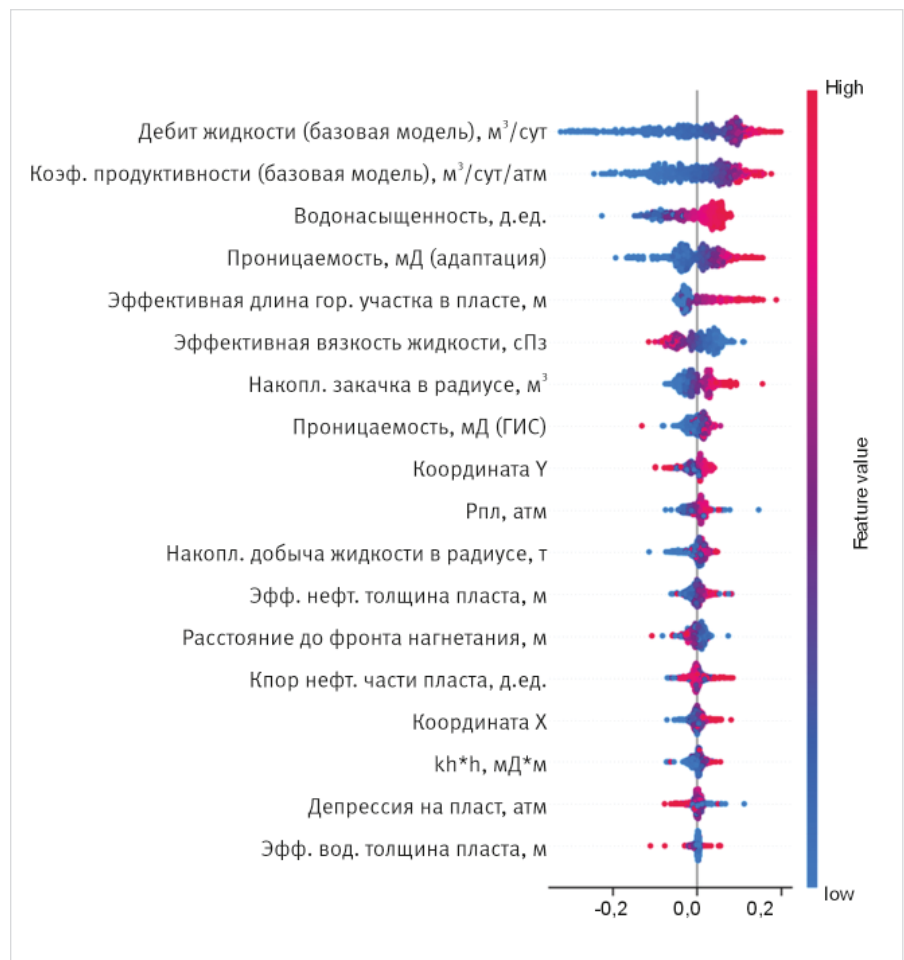


Рис. 5. Важность признаков модели
Fig. 5. Importance of model features

Табл. 1. Средние значения метрик по результату поэлементной кросс-валидации
 Tab. 1. Average metrics based on the leave-one-out cross-validation

МОДЕЛЬ	RMSE	R2
Гибридная модель	53,303 (+\– 49,009)	0,785
Gradient Boosting Machine	59,988 (+\– 53,653)	0,734
Random Forest	65,105 (+\– 56,059)	0,697
Support Vector Machine	65,838 (+\– 57,920)	0,684
Математическая модель	84,082 (+\– 112,611)	0,637

Оценка качества модели

Для сравнения качества прогноза проведены тестовые расчеты дебита жидкости с применением рассмотренных методов (рис. 6) на действующих скважинах. Оценка проведена с применением поэлементной перекрестной проверки (k = размер выборки – 1) по следующей метрике:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (18)$$

где N – число наблюдений в выборке, y_i – фактическое значение целевого параметра, \hat{y}_i – прогнозное значение целевого параметра.

По результатам тестирования гибридная модель показала более высокую точность прогноза в сравнении с другими подходами.

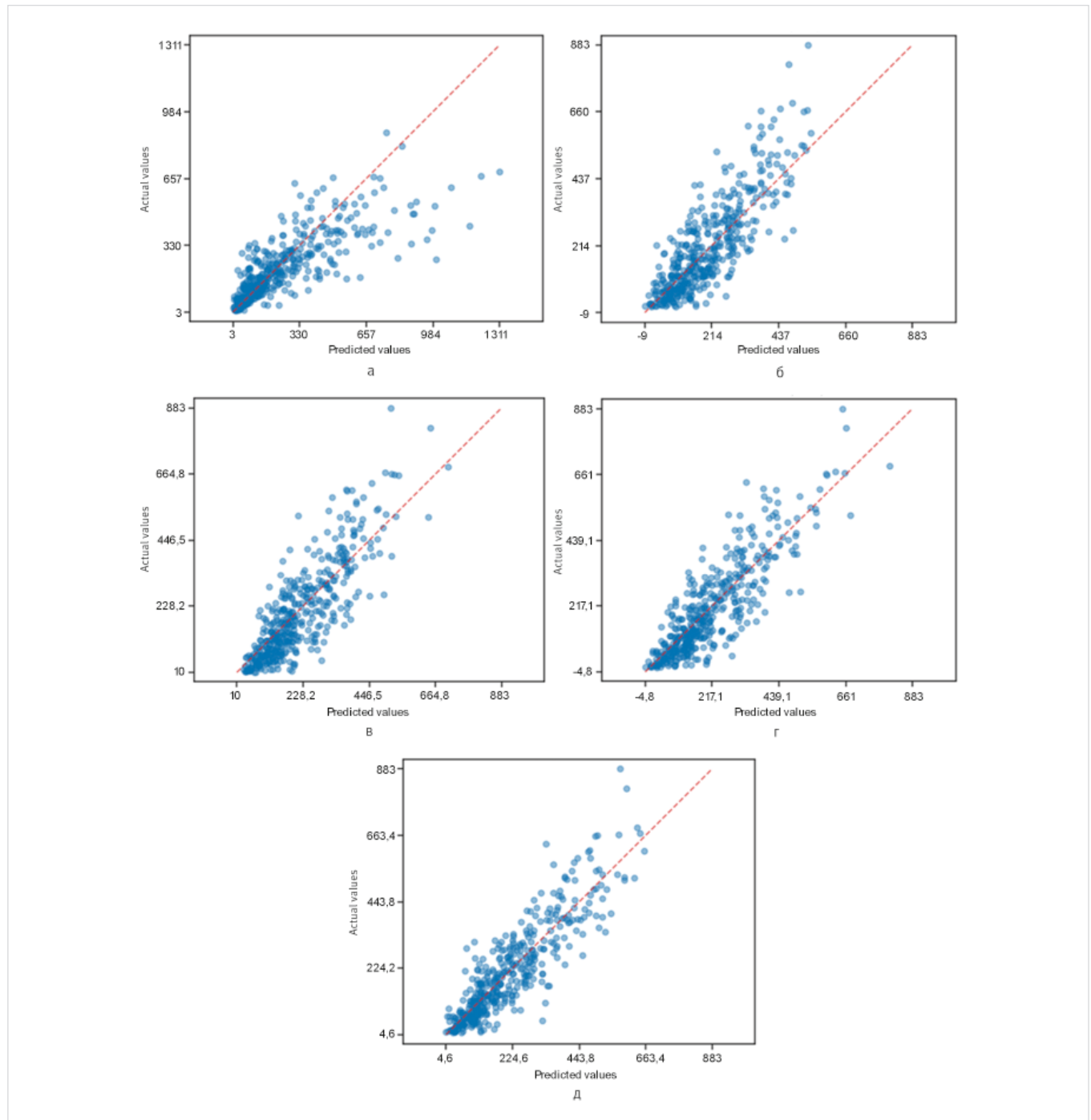


Рис. 6. Результат тестирования различных моделей для прогноза дебита жидкости: а – физико-математическая модель; б – support vector machine; в – random forest; г – gradient boosting machine; д – гибридная модель

Fig. 6. The result of testing various models for predicting liquid rates: а – physical-mathematical model; б – support vector machine; в – random forest; г – gradient boosting machine; д – hybrid model

Полученные средние значения метрик качества приведены в таблице 1.

Итого

На текущий момент разработанная система проходит апробацию на месторождении имени Малыка. Подобрано четыре скважины-кандидата для ГРП, которые находятся в стадии реализации. Практический результат применения подобных решений показывает не только значительное сокращение ручного труда, но и дает комплексную оценку потенциала объектов разработки к проведению ГРП.

Выводы

На данный момент применение алгоритмов машинного обучения в области физического моделирования является перспективным направлением для исследований. Особый интерес представляют гибридные модели, в которые тем или иным образом могут быть внедрены знания о физических законах. В данной работе представлена программная реализация автоматизированной системы поиска скважин-кандидатов для гидравлического разрыва пласта с применением гибридной модели. Приведено сравнение качества прогнозирования гибридной модели и классических методов на основе

физико-математического моделирования и алгоритмов машинного обучения. По результату тестирования гибридная модель показала более высокое качество прогнозирования продуктивности скважины в сравнении с классическими подходами.

Литература

1. Qin S.J., Chiang L.H. Advances and opportunities in machine learning for process data analytics. *Computers & Chemical Engineering*, Vol. 126, P. 465–473. (In Eng).
2. Ansari A.M., Sylvester N.D., Sarica C., Shoham O., Brill J.P. A comprehensive mechanistic model for upward two-phase flow in wellbores. *SPE Production & Facilities*, 1994, Vol. 9, issue 2, P. 143–151. (In Eng).
3. Beggs H.D., Brill J.P. A study of two-phase flow in inclined pipes. *Journal of Petroleum Technology*, 1973, Vol. 25, issue 5, P. 607–617. (In Eng).
4. Hasan A.R., Kabir C.S. A study of multiphase flow behavior in vertical wells. *SPE Production Engineering*, 1988, Vol. 3, issue 2, P. 263–272. (In Eng).
5. Vogel J.V. Inflow performance relationships for solution gas drive wells. *Journal of Petroleum Technology*, 1968, Vol. 20, issue 1, P. 83–92. (In Eng).
6. Brown K.E. *The Technology of artificial lift methods*. Tulsa: PennWell Books, 1984, 474 p. (In Eng).
7. ПАО НК «Роснефть». Методические указания Компании «Расчет геологических показателей программ эксплуатационного бурения и резарезки боковых стволов». 2022.
8. Bikmukhametov T., Jäschke J. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers and Chemical Engineering*, 2020, Vol. 138, 106834. (In Eng).
9. Бикбулатов С.М., Пашали А.А. Анализ и выбор методов расчета градиента давления в стволе скважины // Нефтегазовое дело. 2005. № 2. С. 12. URL: https://ogbus.ru/files/ogbus/authors/Bikbulatov/Bikbulatov_1.pdf (дата обращения: 15.09.2024).
10. Lundberg S.M., Erion G.G., Lee S.I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv*, 2018, URL: <https://arxiv.org/pdf/1802.03888> (accessed: 15.09.2024). (In Eng).

ENGLISH

Results

At the moment, the developed system is being tested at the Malyk Field. Four candidate wells have been selected for hydraulic fracturing jobs which are under implementation. The practical outcomes of such solutions show not only a significant reduction in manual efforts, but also provide a comprehensive assessment of the development potential of hydraulic fracturing targets.

Conclusions

At the moment, the machine learning algorithms applied in physical modeling are a promising area for research. Hybrid models are of

particular interest where the knowledge about physical laws can be introduced in one way or another.

This paper presents a software implementation of an automated system for selecting candidate wells for hydraulic fracturing using a hybrid model. The prediction quality of the hybrid model and classical methods based on physical and mathematical modeling and machine learning algorithms was compared. The test results showed that the hybrid model has a higher quality of predicting well productivity in comparison with classical approaches.

References

1. Qin S.J., Chiang L.H. Advances and opportunities in machine learning for process data analytics. *Computers & Chemical Engineering*, Vol. 126, P. 465–473. (In Eng).
2. Ansari A.M., Sylvester N.D., Sarica C., Shoham O., Brill J.P. A comprehensive mechanistic model for upward two-phase flow in wellbores. *SPE Production & Facilities*, 1994, Vol. 9, issue 2, P. 143–151. (In Eng).
3. Beggs H.D., Brill J.P. A study of two-phase flow in inclined pipes. *Journal of Petroleum Technology*, 1973, Vol. 25, issue 5, P. 607–617. (In Eng).
4. Hasan A.R., Kabir C.S. A study of multiphase flow behavior in vertical wells. *SPE Production Engineering*, 1988, Vol. 3, issue 2, P. 263–272. (In Eng).
5. Vogel J.V. Inflow performance relationships for solution gas drive wells. *Journal of Petroleum Technology*, 1968, Vol. 20, issue 1, P. 83–92. (In Eng).
6. Brown K.E. *The Technology of artificial lift methods*. Tulsa: PennWell Books, 1984, 474 p. (In Eng).
7. “NK “Rosneft” PJSC. Calculation of geological indicators of programs of operational drilling and cutting of side shafts. 2022. (In Russ).
8. Bikmukhametov T., Jäschke J. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers and Chemical Engineering*, 2020, Vol. 138, 106834. (In Eng).
9. Bikbulatov S.M. Pashali A.A. Analysis and selection of wellbore pressure gradient estimation methods. *Oil and Gas Business*, 2005, issue 2, P. 12, URL: https://ogbus.ru/files/ogbus/authors/Bikbulatov/Bikbulatov_1.pdf (accessed: 15.09.2024). (In Russ).
10. Lundberg S.M., Erion G.G., Lee S.I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv*, 2018, URL: <https://arxiv.org/pdf/1802.03888> (accessed: 15.09.2024). (In Eng).

ИНФОРМАЦИЯ ОБ АВТОРАХ | INFORMATION ABOUT THE AUTHORS

Галеев Айрат Алмазович, главный специалист управления автоматизации процессов и менеджмента данных, ООО «Тюменский нефтяной научный центр», Тюмень, Россия
Для контактов: aagaleev@tnnc.rosneft.ru

Синицына Татьяна Ивановна, начальник управления инжиниринга добычи, ООО «Тюменский нефтяной научный центр», Тюмень, Россия

Galeev Airat Almazovich, chief specialist of the process automation and data management division, “Tyumen petroleum research center” LLC, Tyumen, Russia
Corresponding author: aagaleev@tnnc.rosneft.ru

Sinitsyna Tatyana Ivanovna, head of the production engineering division, “Tyumen petroleum research center” LLC, Tyumen, Russia