

**Бембетова Алина Джунгаровна**, студент, МИРЭА — Российский технологический университет, г. Москва

## **ГИБРИДНЫЙ ПОДХОД К ПРОГНОЗИРОВАНИЮ ВРЕМЕННЫХ РЯДОВ ДЛЯ ОЦЕНКИ ПОКАЗАТЕЛЕЙ НЕФТЕДОБЫЧИ**

**Аннотация.** Мировая нефтегазовая промышленность испытывает постоянную потребность в точном прогнозировании добычи нефти, поскольку от качества таких прогнозов напрямую зависят технологические и экономические решения. Управленческие меры по оптимизации работы месторождений требуют количественных оценок динамики изменения дебита, учета внешних факторов и их взаимосвязи с геолого-техническими показателями. В последнее время на первый план выходит интеллектуальный анализ данных, позволяющий расширить возможности классических методов прогнозирования за счет современных подходов машинного обучения и глубоких нейронных сетей. В работе предложена комплексная методика, объединяющая классические методы анализа кривых снижения добычи (Decline Curve Analysis) и современные алгоритмы машинного обучения, включая градиентный бустинг (XGBoost), рекуррентные нейронные сети (LSTM) и архитектуры Transformer. При этом особая роль отводилась физико-информированным нейросетевым решениям, реализованным посредством PINN, которые, интегрируя дифференциальные операторы, обеспечивают соответствие модели фундаментальным законам динамики потоков нефти в пласте. Проведенные эксперименты включали тестирование методов как на синтетических данных, так и на реальных данных. Результаты демонстрируют значительное улучшение точности прогнозов по сравнению с традиционными методами.

**Annotation.** The global oil and gas industry has a constant need for accurate oil production forecasting, as the quality of such forecasts directly influences technological and economic decisions. Management strategies for optimizing field

operations require quantitative assessments of production rate dynamics, consideration of external factors, and their relationship with geological and technical parameters. Recently, data-driven analytics has come to the forefront, enhancing traditional forecasting methods through modern machine learning and deep neural network approaches. This study proposes a comprehensive methodology that combines classical decline curve analysis (DCA) with advanced machine learning algorithms, including gradient boosting (XGBoost), recurrent neural networks (LSTM), and Transformer architectures. A special emphasis is placed on physics-informed neural networks (PINN), which incorporate differential operators to ensure the model's compliance with fundamental laws of oil flow dynamics in reservoirs. The conducted experiments involved testing the developed models on both synthetic data and real-world data. The results demonstrate a significant improvement in forecast accuracy compared to traditional methods.

**Ключевые слова:** прогнозирование нефтедобычи, гибридные модели, машинное обучение, временные ряды, градиентный бустинг

**Keywords:** oil production forecasting, hybrid models, machine learning, time series, gradient boosting

### **Введение**

Нефтедобыча — сложный технологический процесс извлечения нефти из недр Земли, включающий разведку, разработку месторождений и эксплуатацию скважин. Это ключевая отрасль топливно-энергетического комплекса, требующая применения современных методов моделирования для оптимизации добычи. Опубликованные исследования [1, 3] указывают на ряд ограничений при использовании традиционных уравнений снижения (Decline Curve Analysis – DCA): они не всегда справляются с выраженными нелинейностями и резкими скачками, возникающими из-за изменений состава продукции, колебаний в режиме закачки и ремонтно-технических операций. В таких ситуациях классические методики теряют универсальность и требуют адаптации для получения достоверных результатов на длительных временных интервалах.

## Основная часть

За последние почти сто лет методы анализа добычи в нефтегазовой отрасли претерпели существенную эволюцию [2]. Первоначально, начиная с 1920 года, акцент делался на эмпирическом поиске функций снижения добычи (метод Арпса, 1945) для прогнозирования будущей добычи при условии постоянного давления. В основе метода лежит эмпирическая зависимость между дебитом скважины и временем. Основная идея заключается в том, что, если известна текущая динамика дебита, можно спрогнозировать его изменения в будущем. В традиционной модели Арпса предполагается, что скорость снижения дебита скважины зависит от времени, и описывается следующим уравнением:

$$q_t = \frac{q_i}{(1 + bD_i t)^{\frac{1}{b}}}, \quad (1)$$

где  $q_t$  — дебит скважины в момент времени  $t$ ;  $q_i$  — начальный дебит;  $D_i$  — начальный темп спада;  $b$  — гиперболический коэффициент;  $t$  — время.

Формула (1) идеально описывает поведение скважины на фазе, когда давление на забое остается постоянным и истощение резервуара происходит в условиях, близких к установившемуся потоку. Важно заметить, что этот метод применим в основном для анализа данных на поздних стадиях эксплуатации, когда процесс добычи стабилизируется. Однако на начальных этапах, когда скважины находятся в переходном или нестабильном состоянии, данный подход может давать погрешности.

Математическая формулировка для расчета скорости снижения  $D$ , основанная на производной натурального логарифма дебита.

$$D = -\frac{d(\ln q)}{dt} = -\frac{1}{q} \frac{dq}{dt}. \quad (2)$$

Формула (2) позволяет вычислить скорость изменения дебита скважины на основе изменений дебита во времени, что необходимо для оценки будущего состояния скважины. В случае стабильного процесса ее использование

позволяет получать точные прогнозы, однако для нестабильных скважин требуется учитывать дополнительные параметры.

Геолого-физические особенности формируют крайне неоднородную картину насыщения, а также влияют на проницаемость и коллекторские свойства. По мнению авторов, рассматривавших применение мультискважинных моделей, высокие различия в параметрах пористой среды и существенная разница в глубине залегания приводят к тому, что единой простейшей зависимости для всех участков не существует [6, 9]. Наблюдения, полученные при помощи измерительных приборов на устье и забое, содержат шумы, пропуски и аномальные всплески, поскольку реальная нефтедобыча сопряжена с плановыми остановками, ремонтами и разнообразными технологическими операциями [3,10]. Сложность интерпретации этих сигналов увеличивается еще и потому, что вмешательства в работу нагнетательных скважин приводят к непредсказуемым изменениям водонефтяного контакта, влияя на результат по всей сети добычи [12] ситуациями, когда прогноз, построенный по классическим кривым снижения без учета внешних воздействий, начинает заметно расходиться с фактическими показателями [13]. Подобная проблема особенно остра в регионах с многолетней историей эксплуатации, где основные запасы уже вышли на фазу падающей добычи, а поддержание пластового давления требует постоянных корректировок. [7, 8]

Недавние исследования [1,4] показывают, что для улучшения качества прогнозов необходим более гибкий подход, способный учитывать дополнительные факторы, которые не всегда могут быть учтены в рамках традиционных методов. С этой целью активно используются методы машинного обучения, включая нейронные сети, деревья решений, и модели на основе градиентного бустинга. XGBoost используется для решения различных задач с контролируемыми обучающими данными  $x_i$ , для прогнозирования целевой переменной  $y_i$ . При выполнении обучения модель совершает математические операции, в результате которых выполняется

прогнозирование  $y_i$ . Задача обучения модели XGBoost заключается в нахождении наилучшего подхода к обучению целевой функции  $\theta$  (Формула 3). Для целевых функций характерны две части: потери при обучении (loss function) и регуляция (regularization term):

$$obj(\theta) = \sum_{i=1}^n l(y_i, y_i^{(t)} + f_i(x_i)) \sum_{i=1}^r \omega(f_i), \quad (3)$$

где  $l$  — функция потерь;  $y_i, y_i^{(t)}$  — значение  $i$ -го элемента обучающей выборки, сумма предсказаний первых  $t$  деревьев;  $f_i$  — функция которую мы хотим обучить на шаге  $i$ ;  $f_i(x_i)$  — предсказание на  $i$ -ом элементе обучающей выборки;  $\omega(f_i)$  — функция регуляризации.

Модели машинного обучения, основанные на анализе временных рядов, становятся особенно полезными, когда речь идет о предсказаниях на основе долгосрочных данных. Например, рекуррентные нейронные сети (RNN), включая LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit), продемонстрировали свою эффективность при анализе и предсказании сложных временных зависимостей в данных добычи. Эти сети способны учитывать как краткосрочные колебания в производственных показателях, так и долгосрочные тренды, что делает их незаменимыми в задачах прогнозирования на различных этапах разработки месторождений. Применение таких сетей позволяет моделировать изменение дебита скважин, которое зависит не только от текущих технологических параметров, но и от более глубоких процессов, таких как колебания давления в пластах или изменения состава добываемой нефти. В частности, работы, посвященные применению LSTM и GRU в нефтегазовой сфере, показали, что данные методы значительно увеличивают точность прогнозов, особенно в случае сложных многопараметрических систем [5, 11, 15]. LSTM используют специальные блоки с ячейками памяти и тремя воротами (входным, выходным и забывания), что позволяет эффективно сохранять и использовать информацию на длительных интервалах времени.

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f). \quad (4)$$

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i). \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot C_t. \quad (6)$$

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o), \quad (7)$$

где  $x_t$  — входной вектор;  $h_t$  — выходной вектор;  $C_t$  — вектор состояний;  $W, b$  — матрицы параметров и вектор;  $f_t, i_t, o_t$  — векторы вентилей.

Нейронные сети, использующие механизмы внимания и Transformer-архитектуру, становятся перспективными для более сложных задач, когда необходимо учесть большое количество входных параметров и их взаимные зависимости. В Transformer-моделях используется механизм внимания, который позволяет эффективно работать с длинными последовательностями данных и учитывать их глобальную структуру, что делает этот метод особенно подходящим для задач, связанных с прогнозированием динамических процессов, таких как нефтедобыча. Каждый механизм внимания параметризован матрицами весов запросов  $W_Q$ , весов ключей  $W_K$ , весов значений  $W_V$ .

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q K^T}{\sqrt{d_k}}\right) V, \quad (8)$$

где  $d_k$  — размерность ключей и значений;  $Q, K, V$  — матрицы запросов, ключей и значений.

Нейросетевые методы позволяют эффективно обрабатывать большое количество данных, включая не только временные ряды, но и другие типы информации, такие как геофизические характеристики месторождений, результаты сейсморазведки и текущее состояние нагнетательных систем. Рекуррентные сети с механизмами внимания позволяют моделировать динамические процессы, в которых одна переменная может оказывать влияние на несколько других в долгосрочной перспективе, что критически важно для прогнозирования в нефтедобыче [14, 16]. Сложные RNN-модели способны более точно предсказывать поведение скважин в условиях динамически

меняющихся внешних факторов, что значительно улучшает процесс принятия решений, например, в отношении закачки воды или оптимизации работы насосов.

В рамках данной работы приведем результаты двух экспериментов на синтетических данных и на реальных данных месторождения Volve [1.7]. Исторические материалы месторождения, рассекреченные для исследовательских целей, содержат идентификатор скважины (WELL\_ID), дату (Month), объем добычи ( $q_{obs}$ ), а также рассчитанные признаки двух компонент закачки ( $inj_1$ ,  $inj_2$ ). Особенностью такой таблицы оказывается наличие пропусков (NaN) в столбцах закачки: не все месяцы имели зарегистрированную операцию нагнетания, а на некоторых участках скважины функционировали исключительно в режиме отбора.

Старт начинается со сбора массива данных, который содержит либо реальную эксплуатационную статистику (Volve), либо синтетические кривые.

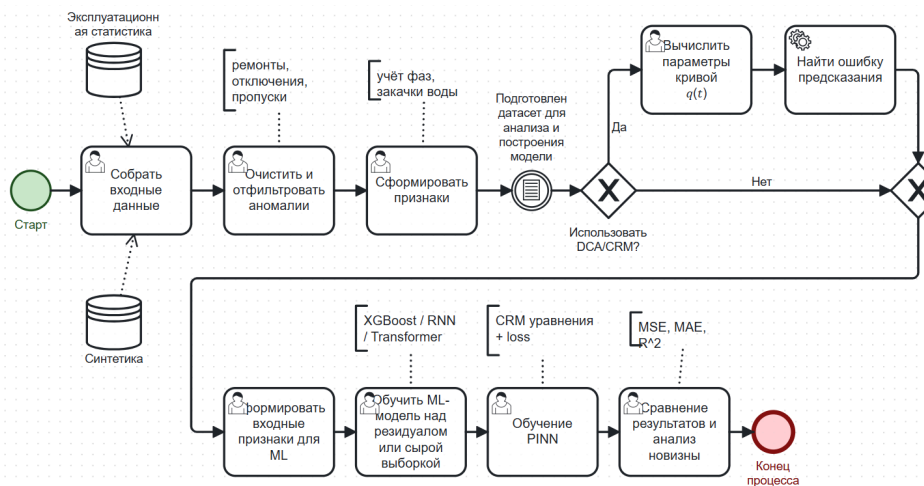


Рисунок 1 — Диаграмма проектирования и анализа гибридных моделей

В ходе начальной очистки выявляются anomальные эпизоды, связанные с временными отключениями оборудования или крупными ремонтom, и формируется окончательный пул необходимых признаков, учитывающих закачку воды и изменения фазового состава. Затем принимается решение о применении классического анализа кривых снижения (fit DCA/CRM), причем этот шаг можно пропустить, если характер временных рядов настолько нерегулярный, что DCA бессмысленно. Если принят путь кривых снижения,

то на выходе получается некая аппроксимация, дающая базовый профиль  $q(t)$ , а также можно сформировать «ошибку», которую и пытается исправить машинное обучение. В случае неустойчивых процессов или явных нелинейностей, формируем набор входных признаков и подаем их в ML-модель (XGBoost, RNN, Transformer). Выученные параметры алгоритма позволяют обновлять предсказания при появлении новых данных, что делает гибридный подход интерактивным и способным адаптироваться к технологическим изменениям. Завершающий этап – обучение PINN.

Первые результаты получены для каждого метода на синтетических данных и представляют сравнение MSE, MAE и R2 для DCA+XGBoost, LSTM, Transformer и PINN-CRM.

	Model	MSE	MAE	R2
0	DCA+XGB	16.103436	2.401164	0.991260
1	Attention-LSTM	1114.708504	28.306115	-0.096640
2	Transformer	1028.945402	27.985491	-0.012267
3	PINN-CRM	172.692052	11.482542	0.906268

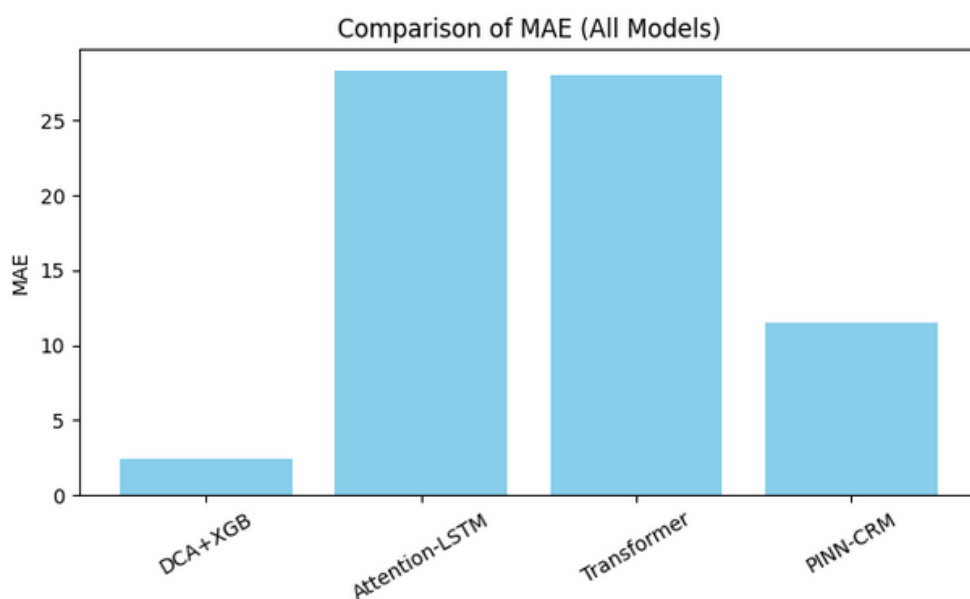


Рисунок 2 — Итоговое сравнение моделей (синтетические данные)

Параметры зависят от случайного шума и конкретной конфигурации сетей, но в целом тенденция следующая: гибриды (DCA+XGBoost или PINN-CRM) воспроизводят тренд точнее, чем отдельные модели. LSTM и Transformer иногда превосходят XGBoost при большом объеме данных, но в сжатом или «коротком» временном ряде XGBoost оказывается вполне

конкурентен. PINN-модель дает более равномерное поведение на длительных интервалах. Специалисты, анализирующие подобные итоги, отмечают, что результат на искусственном датасете дает первичное подтверждение состоятельности подхода, а самые интересные детали обычно проявятся на реальных полевых сведениях.

Вторые результаты получены для каждого метода на данных Volve. MAE PINN-CRM превосходит LSTM и Transformer, а также готовый гибрид (DCA+XGBoost). Интерпретация заключается в том, что классическая кривая снижения остается слишком упрощенной для таких случаев, а рекуррентные сети и сети со слоями внимания натываются на сложность — им приходится самостоятельно угадывать физические механизмы закачки без жесткого контурного уравнения.

	Model	MSE	MAE	R2
0	DCA+XGB	1.188929e+09	24162.312332	0.145818
1	Attn-LSTM	1.986715e+09	33314.972262	-0.023255
2	Transformer	1.988365e+09	33288.166217	-0.024105
3	PINN-CRM	7.255171e+08	15655.685609	0.478755

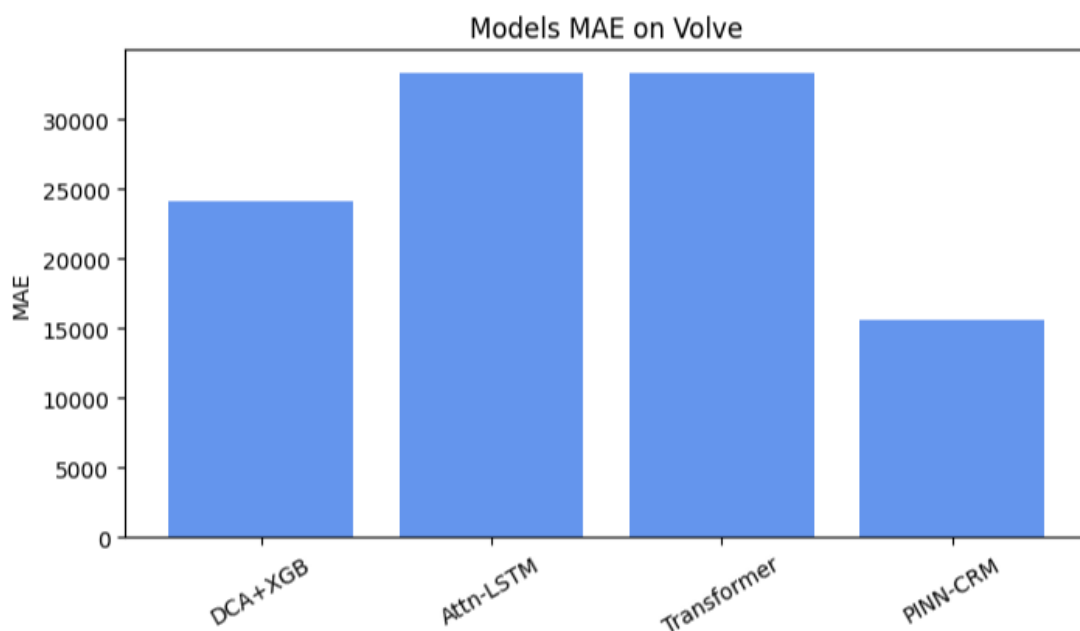


Рисунок 3 — Итоговое сравнение моделей (данные Volve)

Когда участвуют модели CRM и PINN, формальные ошибки становятся значительно меньше, а коэффициент детерминации ( $R^2$ ) поднимается на

уровни около 0.48. Отдельные пики все равно оказываются за гранью точного воспроизведения, потому что, возможно, нагнетательная активность принимала непредсказуемые значения, а исторические файлы не передают всей полноты технологических деталей (таких как перенастройки оборудования, перфорационные работы и т. д.). Анализ результатов подтверждает, что гибридная система DCA + XGBoost выдает неплохое решение, поскольку XGBoost гибко строит деревья по различным признакам, но его недостаток проявляется при больших резких провалах и выбросах. Заполнять пропуски зачастую нужно более искусно, применяя интерполяцию по смежным месяцам или дополнительные технологии очистки. Однако даже такая реализация, говорит о возможности применения алгоритмов: ошибка не становится запредельной (как у чистого DCA). Линейка LSTM/Transformer неплохо схватывает локальные закономерности, но остается уязвимой к неравномерному паттерну пластово-технологических воздействий, потому что нет встроенной физики. CRM + PINN уже на небольшой серии эпох при значительной вычислительной нагрузке демонстрирует большее соответствие физическим предпосылкам.

### **Заключение**

Исследование, направленное на сопоставление традиционных и современных алгоритмов прогнозирования нефтедобычи, выявило существенные преимущества комплексного гибридного подхода, в котором классические эмпирические уравнения снижения объединяются с методами машинного обучения и физико-информированными нейронными сетями. Результаты экспериментов, проведенных как на синтетических данных, так и на реальных исторических данных месторождения Volve, показали, что классический подход, основанный исключительно на эмпирических кривых снижения, не способен адекватно отразить сложные изменения технологических и геолого-физических параметров. Включение методов машинного обучения, таких как XGBoost, позволяет корректировать первичные прогнозы, однако лишь комплексное объединение с PINN-

решением обеспечивает достаточную устойчивость модели к резким технологическим колебаниям и внешним воздействиям. Достигнутый результат подтверждает, что интегрированная система, построенная на основе последовательной цепочки этапов – от первичной обработки временных рядов до обучения гибридной модели – демонстрирует заметное улучшение точности прогнозов даже в условиях нестабильности производственного процесса.

### **Список литературы**

1. Басииев К. С, Кочина И. Н., Максимов В. М. Подземная гидромеханика: Учебник для вузов. – М.: Недра, 1993. 416 с: ил
2. Брагин Ю.И., Кузнецова Г.П. Нефтегазовая геология. Статическое геологическое моделирование залежей углеводородов: Учебное пособие для вузов. – М.: ООО «Издательский дом Недра», 2013.
3. Вознюк С.А., Демин Е.В., Хабибуллин Р.А., Ющенко Т.С. – Динамическое моделирование механизированной добычи малодебитными скважинами с протяженным ГС и МГРП [Электронный ресурс] URL [https://www.oil-industry.net/SD\\_Prezent/2022/09/Demin.pdf](https://www.oil-industry.net/SD_Prezent/2022/09/Demin.pdf)
4. Лейбензон Л.С. – М. ; Л. : ОГИЗ. Гос. изд-во технико-теорет. лит., 1947. – 244 с.
5. Чарный И.А. Издание: Гостоптехиздат, Москва, 1963 г.
6. Щелкачев В.Н. Избранные труды. – М.: Недра, 1990. – 399 с: ил
7. Эквинор. Обмен данными по месторождению Volve [Электронный ресурс]. URL: <https://www.equinor.com/energy/volve-data-sharing>

### **References**

1. Abdullayeva F., Imamverdiyev Y. Development of oil production forecasting method based on deep learning // Statistics, Optimization & Information Computing. – 2019. – Vol. 7, № 3. – P. 263–277.
2. Arps J.J. Analysis of Decline Curves // Trans. AIME. – 1945. – Vol. 160. – P. 228–247.

3. Busby D. Deep-DCA: a new approach for well hydrocarbon production forecasting // ECMOR XVII – 17th European Conference on the Mathematics of Oil Recovery. – 2020. – 10 p.
4. Cao C., Jia P., Cheng L., Qi S. A review on application of data-driven models in hydrocarbon production forecast // Journal of Petroleum Science and Engineering. – 2022. – Vol. 212. – P. 110296.
5. Lin, H., Sun, Q. Crude oil prices forecasting: An approach of using CEEMDAN-based multi-layer gated recurrent unit networks // Energies. – 2020. – Vol. 13, № 7.
6. Mahzari P., Emambakhsh M., Temizel C., Jones A.P. Oil production forecasting using deep learning for shale oil wells under variable gas-oil and water-oil ratios // Petroleum Science and Technology. – 2022. – Vol. 40. – P. 445–468
7. Manasipov R., Nikolaev D., Didenko D., Abdalla R., Stundner M. Physics informed machine learning for production forecast // SPE Reservoir Characterisation and Simulation Conference and Exhibition. – 2023. – SPE-212666-MS. – 8 p.
8. Maniglio M., Fighera G., Dovera L., Stabile C.C. Physics informed neural networks based on a capacitance-resistance model for reservoirs under water flooding conditions // Abu Dhabi International Petroleum Exhibition and Conference. – 2021. – SPE-207800-MS. – 10 p.
9. Negash B., Yaw A.D. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection // Petroleum Exploration and Development. – 2020. – Vol. 47, № 2. – P. 383–392.
10. Ng C.S.W., 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. – P. 109468.
11. Qin, Q., Huang, Z., Zhou, Z., Chen, C., Liu, R. Crude oil price forecasting with machine learning and google search data: An accuracy comparison of

single-model versus multiple-model // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 123. – P. 106266.  
<https://doi.org/10.1016/j.engappai.2023.106266>

12. Sayarpour M., Kabir C.S., Lake L.W. Field applications of capacitance-resistive models in waterfloods // SPE Annual Technical Conference and Exhibition. – 2008. – SPE-114983-MS. – 10 p.
13. Tadjer A., Hong A., Bratvold R.B. Machine learning based decline curve analysis for short-term oil production forecast // Energy Exploration & Exploitation. – 2021. – Vol. 39, № 5. – P. 1747–1769.
14. Vaswani A., Shazeer N., Parmar N., Uszkoreit J. [et al.] Attention is all you need // Advances in Neural Information Processing Systems (NeurIPS). – 2017. – Vol. 30. – P. 5999–6009.
15. Wu, H., Levinson, D. The ensemble approach to forecasting: A review and synthesis // Transportation Research Part C: Emerging Technologies. – 2021. – Vol. 132. – P. 103357. <https://doi.org/10.1016/j.trc.2021.103357>
16. Zhen Y., Fang J., Zhao X., Ge J., Xiao Y. Temporal convolution network based on attention mechanism for well production prediction // Journal of Petroleum Science and Engineering. – 2022. – Vol. 218. – P. 111043. [49]