



УДК 622.323

Статья / Article

© ПНИПУ / PNRPU, 2024

Современные методы применения машинного обучения как инструмента прогнозирования добычи нефти

А.Р. Рустамов¹, Г.М. Пеньков¹, Д.Г. Петраков², М.А. Рустамова³¹Университет Хазар (Азербайджанская Республика. А71096. г. Баку, ул. Махсати, д.42)²Санкт-Петербургский горный университет (Российская Федерация, 199106, г. Санкт-Петербург, 21-я линия Васильевского острова, д. 2)³Азербайджанский университет архитектуры и строительства (Азербайджанская Республика, AZ 1073, г. Баку, ул. Айны Султановой, 5)

Modern Methods of Using Machine Learning as a Tool for Oil Production Forecasting

Arif R. Rustamov¹, Grigori M. Penkov¹, Dmitriy G. Petrakov², Mahsati A. Rustamova³¹Khazar University (42 Mahsati st., Baku, AZ1096, Republic of Azerbaijan)²Saint Petersburg Mining University (2 21st Line, Vasilyevsky island, Saint Petersburg, 199106, Russian Federation)³Azerbaijan University of Architecture and Construction (5 Ayna Sultanova st., Baku, AZ1073, Republic of Azerbaijan)

Получена / Received: 31.08.2023. Принята / Accepted: 29.02.2024. Опубликовано / Published: 31.03.2024

Ключевые слова:

прогнозирование добычи, машинное обучение, кривая падения давления, искусственный интеллект, сеть с долговременной и кратковременной памятью.

Прогнозирование добычи нефти играет важную роль в эффективной разработке месторождения нефти. Это помогает скорректировать действующую систему разработки месторождения. Детальное и точное прогнозирование уровня добычи нефти необходимо для оценки экономической и технологической эффективности разработки месторождения нефти. Прогнозирование уровня добычи можно осуществить различными способами. Одним из таких может быть использование специального программного обеспечения (tNavigator и др.). Использование данного программного обеспечения иногда сопряжено с длительными расчетами, поэтому для оперативного прогнозирования уровня добычи возможно использование других инструментов, таких как машинное обучение.

Использование машинного обучения и искусственного интеллекта в нефтегазовой отрасли приобретает все большую популярность в последние годы, поскольку, используя исторические данные по добыче, возможно прогнозирование уровней добычи нефти/жидкости. Кроме того, для аналогичных целей могут быть использованы аналогичные месторождения со схожими геологическими характеристиками и историей эксплуатации.

Помимо использования машинного обучения и искусственного интеллекта, в качестве инструмента прогнозирования возможно применение анализ кривой падения.

Учитывая важность прогнозирования с точки зрения стратегического планирования, предлагается широкий спектр методов для получения точных прогнозов, основанных на характере доступных данных и вычислительной мощности. В данной статье представлен всесторонний анализ инструментов, используемых для долгосрочного прогнозирования добычи нефти, включая алгоритмы машинного обучения и анализ кривой падения добычи (DCA). Представлены результаты применения модели с долговременной и кратковременной памятью и ее практическая применимость на примере ее использования на скважине кандидате.

Keywords:

production forecasting, machine learning, pressure drop curve, artificial intelligence, long short-term memory.

Oil production forecasting plays an important role in efficient oil field development. This helps to adjust the current field development system. Detailed and accurate forecasting of oil production levels is necessary to assess the economic and technological efficiency of oil field development. Forecasting production levels can be done in various ways. One of these may be the use of special software (tNavigator, etc.). The use of this software sometimes involves lengthy calculations, so to quickly predict production levels, it is possible to use other tools, such as machine learning.

The use of machine learning and artificial intelligence in the oil and gas industry has become increasingly popular in recent years, as by using historical production data, it is possible to predict oil/liquid production levels. In addition, similar deposits with similar geological characteristics and exploitation history can be used for similar purposes.

In addition to using machine learning and artificial intelligence as a forecasting tool, it is possible to use decline curve analysis.

Given the importance of forecasting from a strategic planning perspective, a wide range of methods have been proposed to obtain accurate forecasts based on the nature of available data and computing power. This article provides a comprehensive analysis of the tools used for long-term oil production forecasting, including machine learning algorithms and decline curve analysis (DCA), in particular. This article presents the results of applying the long- and short-term memory model and its practical applicability using the example of its use on a candidate well.

© Рустамов Ариф Рашад оглы – магистр кафедры нефтегазового дела (тел.: +994 (50) 324 27 73, e-mail: arif.rustamov2022@khazar.org).

© Пеньков Григорий Михайлович (ORCID: 0000-0001-7291-0757) – кандидат технических наук, доцент кафедры нефтегазового дела (тел.: +994 (50) 336 83 95, e-mail: gpenkov@khazar.org). Контактное лицо для переписки.

© Петраков Дмитрий Геннадьевич (ORCID: 0000-0002-0461-1621) – кандидат технических наук, проректор по образовательной деятельности (тел.: +007 (921) 788 19 62, e-mail: petrakov_dg@pers.spmi.ru).

© Рустамова Махсати Аксиф кызы (ORCID: 0000-0001-5192-1166) – кандидат физико-математических наук, доцент кафедры механики (тел.: +994 (50) 535 06 99, e-mail: mahsati.rustamova@azmiu.edu.az).

© Arif R. Rustamov – Master (tel.: +994-50-324-27-73, e-mail: arif.rustamov2022@khazar.org).

© Grigori M. Penkov (Author ID in Scopus: 57193449722, ORCID: 0000-0001-7291-0757) – PhD in Engineering, Associate Professor at the Department of Petroleum Engineering (tel.: +994-50-336-83-95, e-mail: gpenkov@khazar.org). The contact person for correspondence.

© Dmitriy G. Petrakov (Author ID Scopus: 57015158900, ORCID: 0000-0002-0461-1621) – PhD in Engineering, Associate Professor, Vice-Rector for Educational Activities (tel.: +007 (921) 788 19 62, e-mail: Petrakov_DG@pers.spmi.ru).

© Mahsati A. Rustamova (Author ID in Scopus: 55489058000, ORCID: 0000-0001-5192-1166) – PhD in Physics and Mathematics, Associate Professor at the Department of Mechanics (tel.: +994-50-535-06-99, e-mail: mahsati.rustamova@azmiu.edu.az).

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:

Современные методы применения машинного обучения как инструмента прогнозирования добычи нефти // А.Р. Рустамов, Г.М. Пеньков, Д.Г. Петраков, М.А. Рустамова // Недропользование. – 2024. – Т.24, №1. – С.44–50. DOI: 10.15593/2712-8008/2024.1.6

Please cite this article in English as:

Rustamov A.R., Penkov G.M., Petrakov D.G., Rustamova M.A. Modern methods of using machine learning as a tool for oil production forecasting. *Perm Journal of Petroleum and Mining Engineering*, 2024, vol.24, no.1, pp.44-50. DOI: 10.15593/2712-8008/2024.1.6

Введение

Прогнозирование добычи нефти играет важную роль в энергетическом планировании и принятии решений в нефтяной промышленности [1–3].

Учитывая важность прогнозирования добычи нефти, разрабатываются различные методы, позволяющие давать прогнозы на основе имеющихся исторических данных.

Одним из таких методов может быть анализ кривой падения (Decline curve analysis – DCA). Этот метод базируется на том, что в течение разработки нефтяного месторождения происходит снижение уровня добычи и что данный тренд возможно описать, используя закономерности.

Анализ кривых падения давления позволяет осуществить достаточно точное прогнозирование в связи с тем, что существует большое количество аналитических кривых, описывающих характер снижения добычи. Одним из плюсов данного способа прогнозирования является небольшое количество необходимых входных данных.

Передовые подходы, основанные на базе данных, такие как методы машинного обучения (Machine learning – ML) и искусственного интеллекта (Artificial intelligence – AI), приобретают все большую популярность в последние годы. Данные подходы позволяют на основе изученных закономерностей и взаимосвязей параметров разработки осуществлять прогнозирование этих параметров.

Рассматривается и предлагается одномерная модель долгой краткосрочной памяти (Long short-term memory – LSTM), основной целью которой является расчет одноэтапных и многоэтапных прогнозов добычи.

В настоящее время проведено большое количество работ по использованию машинного обучения при выполнении различных нефтегазовых задач. Например, в работе М. Берихун и др. [4] использовались нейронные сети прямого распространения для моделирования течения нефти, газа и воды. Для оценки эффективности модели использовались данные, собранные с месторождения, расположенного в Малайском бассейне. Чтобы установить взаимосвязь между параметрами добывающей и нагнетательной скважины, подбирались случайные комбинации входных параметров (например, скорость закачки воды, давление нагнетания воды и др.). Результаты показали, что извлечение признаков таким способом значительно повышает производительность модели и дает наименьшую среднеквадратическую ошибку и самый высокий коэффициент детерминации, когда модель обучается с помощью байесовской регуляризации.

Ш.М. Бернети и М. Шахбазян [5] предложили модель, основанную на алгоритме империалистической конкуренции (Imperialist competitive algorithm – ICA), для оптимизации начального веса нейронной сети прямого распространения и для прогнозирования дебита нефти из скважин на одном из нефтяных месторождений Ирана на севере Персидского залива. Авторы предложили использовать модель ICA-ANN (artificial neural network) в качестве более дешевой и быстрой альтернативы многофазного расходомера [6]. Модель ICA-ANN с двумя входными параметрами (температурой и давлением) дала наиболее точные прогнозы расхода со среднеквадратической ошибкой 0,0123 и коэффициентом эффективности R^2 0,97. В аналогичном исследовании А. Пайамана, С. Салавати [7] в качестве входных параметров для нейронной сети использовались: давление, размер штуцера и соотношение добываемого газа к нефти (ГФ). Позже прогнозы сравнивались с известными эмпирическими зависимостями, которые используются [8–12] для прогнозирования течения двухфазной жидкости через устьевого штуцер.

Авторы П. Чжан, М. Чжао [13] в своих исследованиях применяли нейронную сеть с обучением по методу обратного распространения ошибки (Back propagation neural network – BP) вместе с данными каротажа и историей добычи для прогнозирования дебита нефти и воды. Набор входных данных изначально был разделен на три части (статические данные, динамические данные и пространственно-временные зависимости) для учета различных эффектов, возникающих при тении жидкости в породе-коллекторе и насосно-компрессорных трубах. Для анализа пространственно-временных зависимостей между скважинами на уровне месторождения авторами была разработана диаграмма Вороного. Результат показал, что погрешность прогнозирования дебита нефти составляет менее 7 %, а воды – в пределах 5 %.

Модель LSTM была построена и обучена для прогнозирования добычи нефти на одном из месторождений Китая с 5 добывающими и 4 нагнетательными скважинами [14]. Согласно анализу, наибольшее влияние на модель оказал объем остаточных извлекаемых запасов и устьевое давление. Средняя квадратичная ошибка (RMSE) и средняя абсолютная ошибка (MAPE) модели LSTM оказались небольшими со значениями 0,985 и 0,035 соответственно.

Методика исследования

Выбор нейронных сетей LSTM для долгосрочного прогнозирования добычи нефти в данном исследовании объясняется тем, что такая нейронная сеть позволяет улавливать временные зависимости в последовательных данных.

Новый класс нейронных сетей, рекуррентные нейронные сети (Recurrent Neural Networks – RNN), был представлен в 1980-х гг. и специально разработан для решения проблем, связанных с временными рядами [16–18]. Уникальная архитектура, принадлежащая RNN, позволяет сохранять информацию на протяжении последовательных временных шагов, что делает их пригодными для последовательных данных. В процессе обучения RNN используют алгоритм обратного распространения ошибки во времени, позволяющий корректировать вес путем расчета градиентов. Несмотря на преимущества, предлагаемые моделями RNN для задач, включающих последовательности, проблема исчезновения градиента делает обучение все более неэффективным на длительном временном интервале зависимостей [19].

С целью уменьшить ограничения моделей RNN был представлен более сложный вариант – LSTM [20]. В отличие от аналогов RNN, сети LSTM могут фиксировать долгосрочные зависимости в последовательных данных и эффективно запоминать расширенные временные отношения, избегая проблем с производными. Внутреннее решение проблемы заключается в использовании в модели карусели постоянных ошибок (Constant error carousel – CEC), которая обеспечивает сохранение сигналов ошибок в каждой элементарной ячейке, тем самым позволяя градиентам сохраняться в длинных последовательностях [21, 22].

Перекрестная проверка временных рядов и настройка гиперпараметров

Для оценки производительности модели в машинном обучении используется отдельный набор данных [23–25]. Чтобы понять, насколько хорошо модель обобщает и рассчитывает прогнозируемые

величины, производительность модели оценивается на наборе проверочных данных во время обучения.

В машинном обучении существуют различные типы методов проверки [26–29], используемых для оценки поведения модели. Критерии выбора заключаются в конкретном типе обрабатываемых данных и их размере. Традиционные методы проверки, такие как стандартная K-кратная перекрестная проверка и простые методы проверки с удержанием, не всегда могут использоваться с данными, которые зависят от времени. В отличие от этих методов, перекрестная проверка временных рядов сохраняет временной порядок данных и хорошо подходит для данных хронологического типа [30].

На рис. 1 поясняется основной принцип работы перекрестной проверки. Точками является добыча нефти в определенный месяц. Синие точки модель использует для обучения. Красные точки – это результат прогнозирования добычи нефти. Изначально используется набор данных из первой итерации (1-я итерация – модель обучается до 01.11.2017). После расчета всех прогнозных точек в модели увеличивается количество тестовых точек на одну (2-я итерация – модель обучается до 01.12.2018), и производится расчет значений заново. Такая последовательность действий продолжается для всех точек в горизонте прогнозирования, и на каждом этапе для оценки разницы между реальным и прогнозируемым моментом используется среднеарифметическое отклонение (MAE). В конце все ошибки усредняются, чтобы оценить поведение модели по мере поступления новых данных. Процедура имитирует реальный сценарий, когда новые ежемесячные данные о добыче нефти регулярно добавляются к историческим данным, и модель должна адаптироваться меняющиеся закономерности по мере того, как новые данные становятся доступными.

Эта методология позволяет точно настроить критические гиперпараметры, включая размер окна, количество эпох и единиц в архитектуре LSTM. Доступны различные алгоритмы для выбора оптимальной комбинации гиперпараметров, чтобы значительно улучшить прогнозирующую эффективность модели [31–36].

Показатели оценки производительности

Оценка эффективности применяемой модели важна для понимания качества используемой модели прогнозирования [37, 38].

Среднеквадратическая ошибка (RMSE) – это стандартный критерий, используемый для оценки степени соответствия путем количественного определения средней величины ошибок между наблюдаемыми и прогнозируемыми значениями [39]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}. \quad (1)$$

Среднеарифметическое отклонение – еще один полезный показатель при оценке регрессионной модели:

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i|. \quad (2)$$

Еще одним важным показателем, который следует использовать при оценке регрессионной модели, является среднеквадратичная логарифмическая ошибка (MSLE), особенно полезная для наборов данных, следующих экспоненциальному тренду [40].

$$MSLE = \frac{1}{n} \sum_{i=1}^n (\log(O_i + 1) - \log(P_i + 1))^2. \quad (3)$$

Результаты проведенных исследований

В этом исследовании использовались данные добывающей скважины нефтяного месторождения X Западной Сибири. На рис. 2 показаны профиль добычи нефти и воды за период с 1 января 2003 г. по 1 ноября 2020 г.

Красные точки на кривой добычи соответствуют периодам капитального/текущего ремонта скважины (снижение изоляции ЭЦН, отключение напряжения ЭЦН, выход на проектную мощность и т.д.). Периоды простоя варьировались от 11 до 481 ч и, как видно из графика, они обычно совпадают с провалами по добыче жидкости.

Для начала прогнозирования с помощью модели LSTM необходимо было определить нужное количество гиперпараметров, таких как количество сегментов, количество эпох и размер окна.

Для выбора оптимального числа эпох, размера окна и количества сегментов был произведен расчет среднеарифметического отклонения для различных комбинаций этих параметров.

На рис. 3 показаны значения среднеарифметического отклонения для каждой комбинации гиперпараметров в виде вертикальных полос, а значение среднеарифметического отклонения указано на горизонтальной оси. Стоит отметить, что уменьшение отклонения происходит с увеличением сложности модели LSTM. Исходя из графика, можно сделать вывод, что при увеличении количества гиперпараметров точность прогнозирования модели возрастает, также с увеличением размера окна происходит уменьшение среднеарифметического отклонения для нескольких комбинаций количества единиц и эпох.

Анализ графика показал, что для получения наиболее точного прогноза необходимо использовать модель с количеством эпох, равным 100, а количество сегментов должно быть равно 128. После определения комбинации этих гиперпараметров было необходимо определить оптимальный размер окна. Для этого для каждого размера окна на каждом временном промежутке была рассчитана абсолютная ошибка (рис. 4).

Определив необходимые гиперпараметры, данная LSTM модель использовалась для прогнозирования уровня добычи нефти на 36 месяцев вперед. Для моделирования уровня добычи применялись три разных подхода. Первый метод (одномерная одношаговая модель) заключался в том, что, спрогнозировав добычу на один месяц вперед, при следующем шаге прогнозирования не использовалось значение, которое было рассчитано с помощью LSTM-модели, а задействовалось реальное значение. На основании этого можно сделать следующий вывод, что при прогнозировании уровня добычи таким способом необходимо наличие реальных данных (рис. 5).

Отличие второго подхода (рекурсивное прогнозирование) от первого было в том, что при прогнозировании, например на втором шаге, модель обучалась на значении, которое было спрогнозировано на первом временном шаге. Таким образом, спустя несколько временных шагов, модель обучалась на своих ранее спрогнозированных значениях (рис. 6).

При прогнозировании уровня добычи с помощью третьего метода (одномерная многоступенчатая модель) расчет делался не на каждый месяц, а был рассчитан профиль добычи сразу на 36 месяцев (рис. 7).

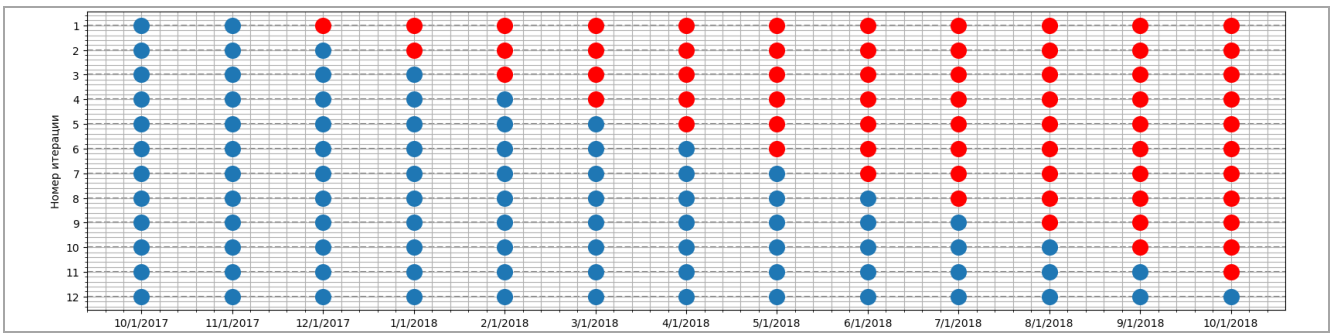


Рис. 1. Перекрестная проверка временных рядов с увеличением размера окна

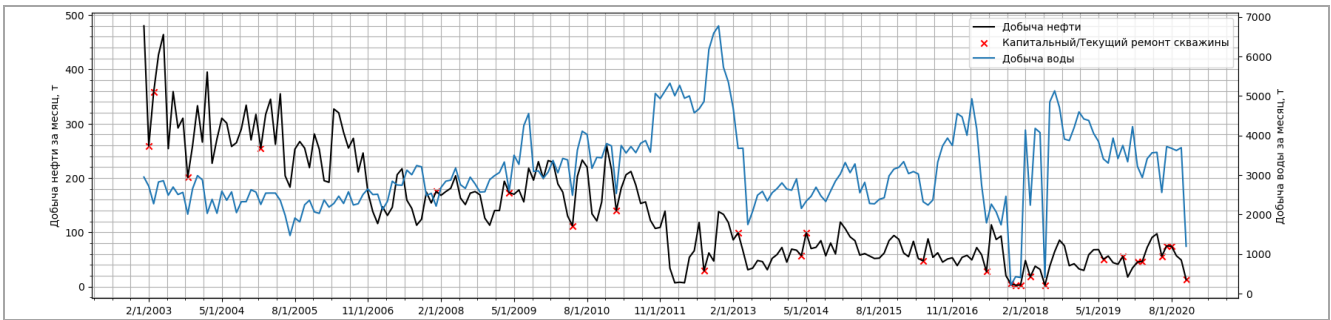


Рис. 2. Профиль добычи нефти/воды исследуемой скважины

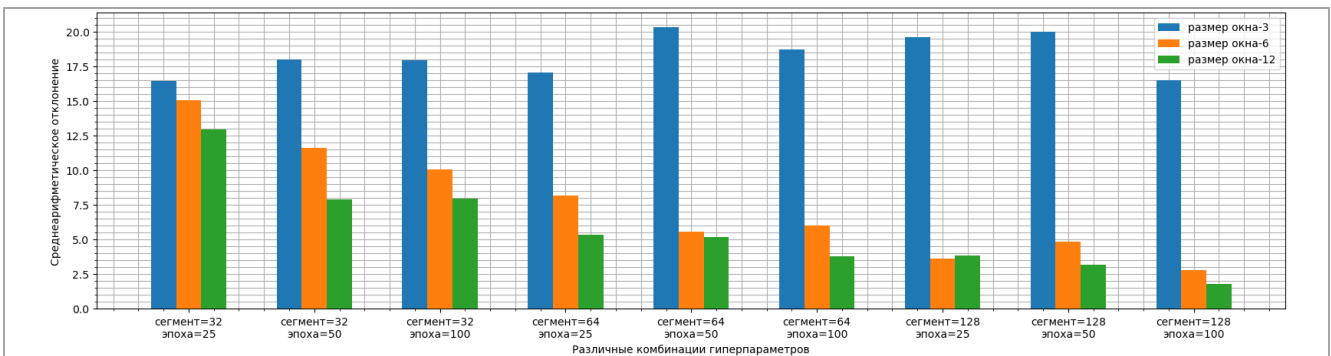


Рис. 3. Среднеарифметического отклонения для каждой комбинации гиперпараметров

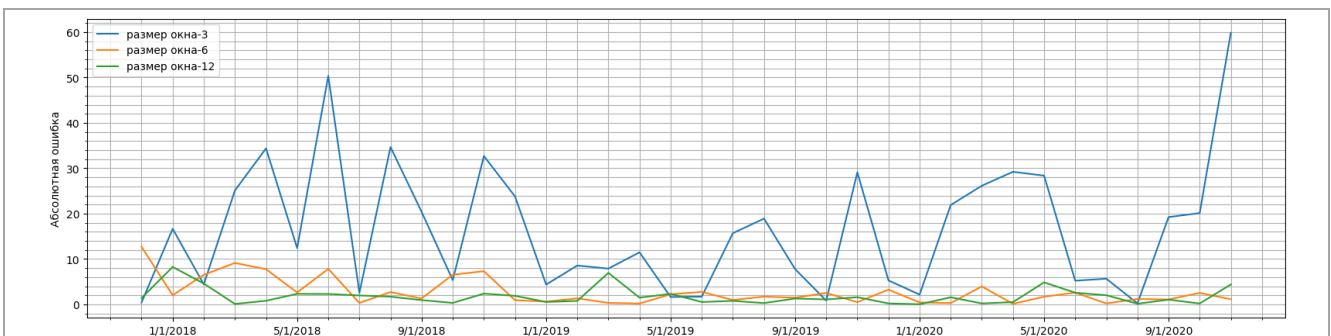


Рис. 4. Подбор размера окна

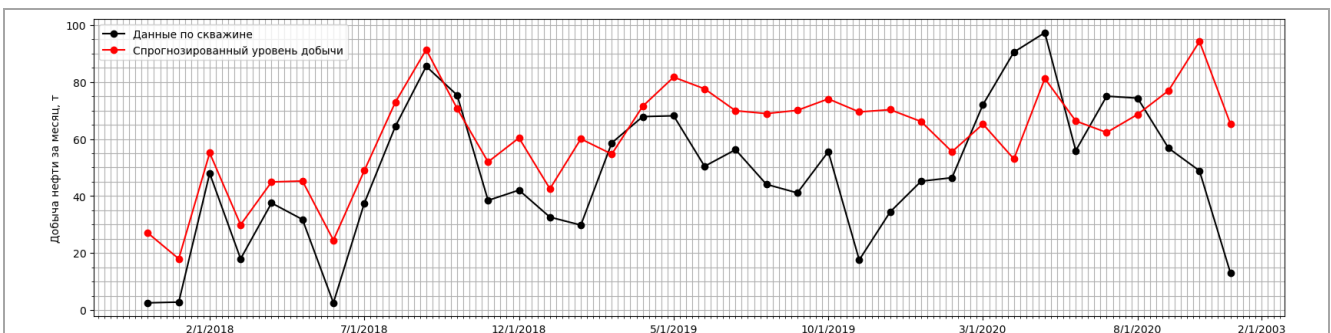


Рис. 5. Прогнозирование с помощью одномерной одношаговой модели

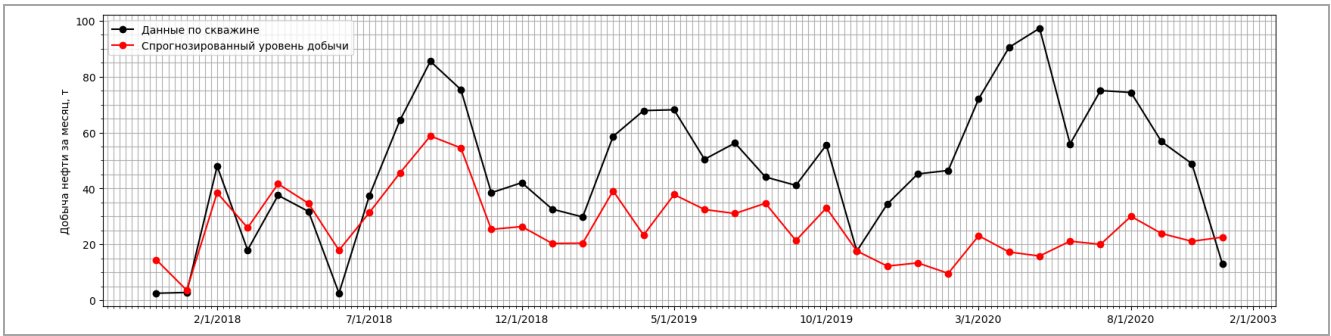


Рис. 6. Прогнозирование с помощью рекурсивного прогнозирования

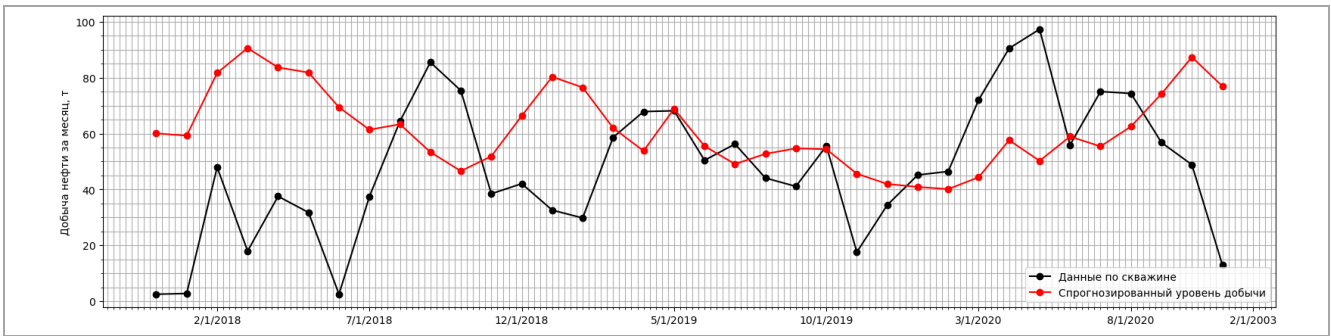


Рис. 7. Прогнозирование с помощью одномерной многоступенчатой модели

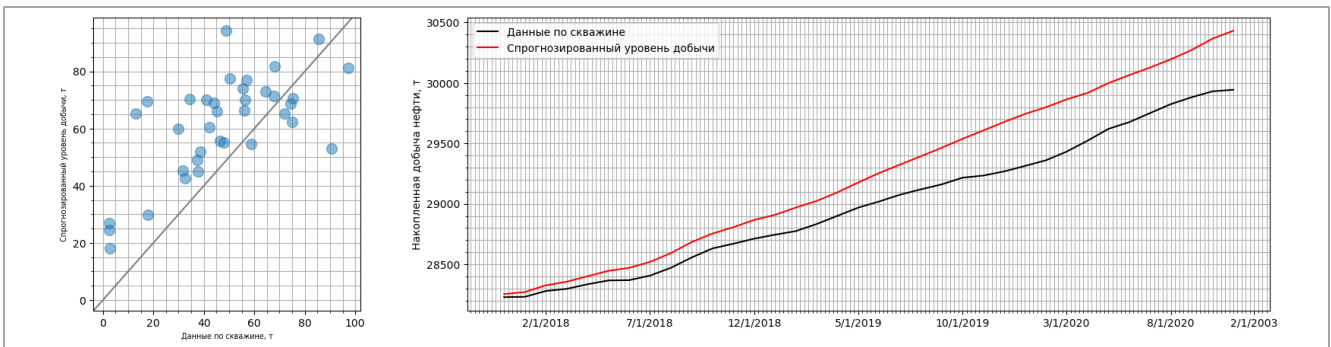


Рис. 8. Диаграмма разброса в одномерной одношаговой модели

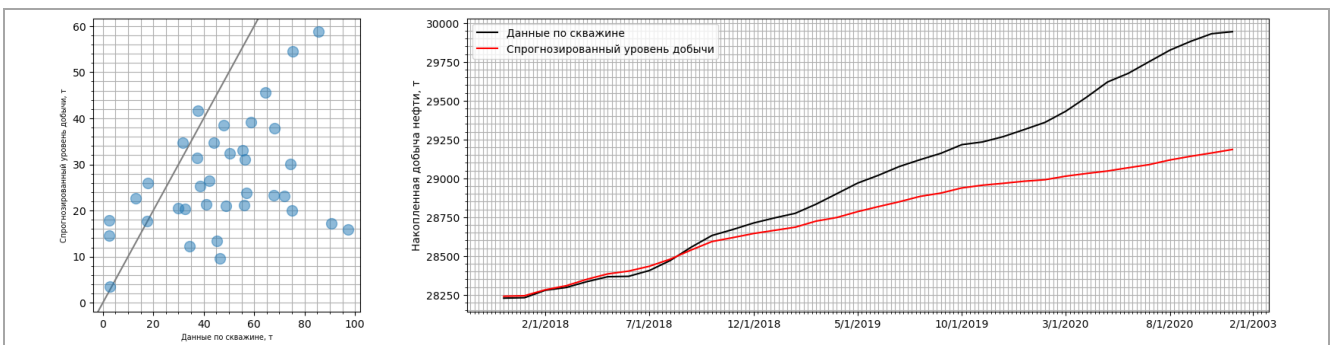


Рис. 9. Диаграмма разброса при рекурсивном прогнозировании

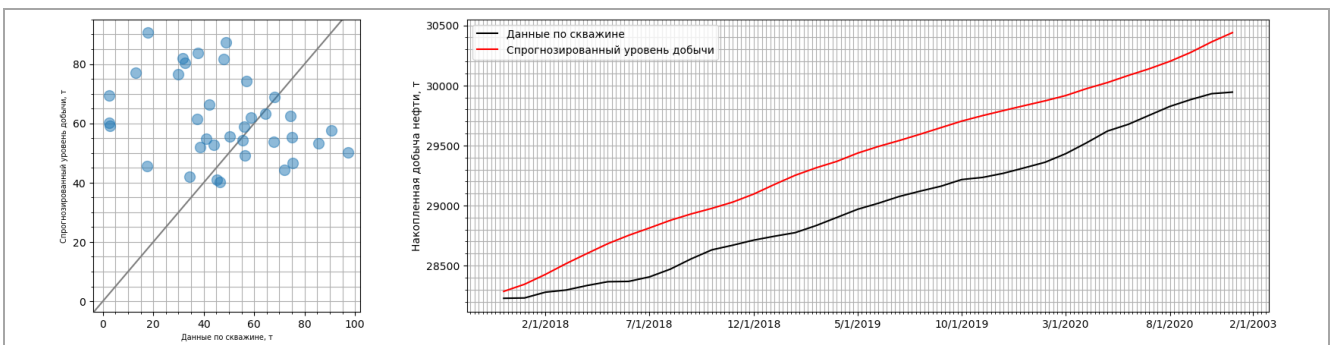


Рис. 10. Диаграмма разброса в одномерной многоступенчатой модели

Каждый спрогнозированный профиль добычи сравнивался с реальными данными, и был произведен расчет ошибок каждого метода: 1-й метод: RMSE = 22,48, MAE = 18,43, и MSLE = 0,52; 2-й метод: RMSE = 30,46, MAE = 24,00, и MSLE = 0,71; 3-й метод: RMSE = 31,72, MAE = 25,21, и MSLE = 1,03. Также каждая модель была оценена на предрасположенность завышать или занижать значения.

На рис. 8 показана диаграмма разброса ежемесячных прогнозов и реальных данных за указанный период с коэффициентом корреляции Пирсона, равным 0,66. Большое скопление точек над линией, проходящей через начало координат с наклоном 1, является явным признаком завышения добычи нефти. Кроме того, на этом же рисунке можно увидеть сравнение исторической и прогнозируемой накопленной добычи нефти за период. В конце периода модель завышает совокупные реальные данные на 480 т.

В отличие от предыдущего случая, рекурсивное прогнозирование явно недооценивает накопленную добычу. При рассмотрении рис. 9 можно отметить 758 т заниженной накопленной добычи. Кроме того, диаграмма рассеяния показывает разбросанные точки данных с коэффициентом корреляции Пирсона 0,4.

Одномерная многоступенчатая модель завышает накопленную добычу нефти на 530 т. В этом случае можно наблюдать более разбросанную диаграмму разброса с коэффициентом корреляции Пирсона $-0,22$ (рис. 10), что указывает на отрицательную корреляцию между переменными.

Библиографический список

- Standards for Selection of Surfactant Compositions used in Completion and Stimulation Fluids / D.G. Petrakov [et al.] // International Journal of Engineering. – 2023. – Т. 36, № 9. – P. 1605–1610. DOI: 10.5829/IJE.2023.36.09C.03
- Density-dependent relationship between changes in ultrasonic wave velocities, effective stress, and petrophysical-elastic properties of sandstone / H. Aghaei [et al.] // Ultrasonics. – 2023. – Т. 132. – P. 106985. DOI: 10.1016/j.ultras.2023.106985
- Petrakov, D.G. Experimental study on the effect of rock pressure on sandstone permeability / D.G. Petrakov, G.M. Penkov, A.B. Zolotukhin // Записки Горного института. – 2022. – Т. 254. – С. 244–251. DOI: 10.31897/PMI.2022.24
- Negash, B.M. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection / B.M. Negash, A.D. Yaw // Petroleum Exploration and Development. – 2020. – Т. 47, № 2. – P. 383–392. DOI: 10.1016/S1876-3804(20)60055-6
- Berneti, S.M. An imperialist competitive algorithm artificial neural network method to predict oil flow rate of the wells / S.M. Berneti, M. Shahbazian // International journal of computer applications. – 2011. – Т. 26, № 10. – P. 47–50. DOI: 10.5120/3137-4326
- Falcone, G. The challenges of multiphase flow metering: today and beyond / G. Falcone, C. Alimonti // International Conference on Offshore Mechanics and Arctic Engineering. – 2007. – Т. 42681. – P. 823–834. DOI: 10.1115/OMAE2007-29527
- Mirzaei-Paiaman, A. The application of artificial neural networks for the prediction of oil production flow rate / A. Mirzaei-Paiaman, S. Salavati // Energy Sources, Part A: Recovery, utilization, and environmental effects. – 2012. – Т. 34, № 19. – P. 1834–1843. DOI: 10.1080/15567036.2010.492386
- Gilbert, W.E. Flowing and Gas-Lift Well Performance / W.E. Gilbert // Drilling and Production Practice. – 1954. – Т. 13. – P. 126–157.
- Baxendell, P.B. Bean performance-lake wells. Shell Internal Report / P.B. Baxendell. – Houston, TX: Shell Oil. – 1957.
- Ros, N.C.J. An analysis of critical simultaneous gas/liquid flow through a restriction and its application to flowmetering / N.C.J. Ros // Applied Scientific Research. – 1960. – Т. 9. – P. 374–388. DOI: 10.1007/BF00382215
- Achong, L.B. Revised bean and performance formula for Lake Maracaibo wells. Shell Internal Report / L.B. Achong. – Houston, TX: Shell Oil Co. – 1961.
- Mirzaei-Paiaman, A. A new empirical correlation for sonic simultaneous flow of oil and gas through wellhead chokes for Persian oil fields / A. Mirzaei-Paiaman, S. Salavati // Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. – 2013. – Т. 35, № 9. – P. 817–825. DOI: 10.1080/15567031003773304
- Zhang, P.Y. The Application of BP Neural Network In Oil-Field / P.Y. Zhang, Z. Meng-Meng // TELKOMNIKA Indonesian Journal of Electrical Engineering. – 2013. – Т. 11. – №. 9. – P. 527–5283. DOI: 10.11591/telkomnika.v11i9.3280
- Liu, W. Forecasting oil production using ensemble empirical mode decomposition based Long Short-Term Memory neural network / W. Liu, W.D. Liu, J. Gu // Journal of Petroleum Science and Engineering. – 2020. – Т. 189. – С. 107013. DOI: 10.1016/j.petrol.2020.107013
- Data normalization and standardization: a technical report / P.J.M. Ali [et al.] // Mach Learn Tech Rep. – 2014. – Т. 1, № 1. – P. 1–6. DOI: 10.13140/RG.2.2.28948.04489
- Rumelhart, D.E. Learning representations by back-propagating errors / D.E. Rumelhart, G.E. Hinton, R.J. Williams // Nature. – 1986. – Т. 323, № 6088. – P. 533–536. DOI: 10.1038/323533a0
- Werbos, P.J. Generalization of backpropagation with application to a recurrent gas market model / P.J. Werbos // Neural networks. – 1988. – Т. 1, №. 4. – P. 339–356. DOI: 10.1016/0893-6080(88)90007-X
- Elman, J.L. Finding structure in time / J.L. Elman // Cognitive science. – 1990. – Т. 14, № 2. – P. 179–211. DOI: 10.1016/0364-0213(90)90002-E
- Bengio, Y. Learning long-term dependencies with gradient descent is difficult / Y. Bengio, P. Simard, P. Frasconi // IEEE transactions on neural networks. – 1994. – Т. 5, № 2. – P. 157–166. DOI: 10.1109/72.79181
- Hochreiter, S. Long short-term memory / S. Hochreiter, J. Schmidhuber // Neural computation. – 1997. – Т. 9, № 8. – P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735
- Gers, F.A. Learning to forget: Continual prediction with LSTM / F.A. Gers, J. Schmidhuber, F. Cummins // Neural computation. – 2000. – Т. 12, № 10. – P. 2451–2471. DOI: 10.1162/089976600300015015
- Staudemeyer, R.C. Understanding LSTM—a tutorial into long short-term memory recurrent neural networks / R.C. Staudemeyer, E.R. Morris // arXiv preprint arXiv:1909.09586. – 2019. DOI: 10.48550/arXiv.1909.09586
- Understanding LSTM networks by Colah accessed at Understanding LSTM Networks -- colah's blog.
- Werbos, P.J. Backpropagation through time: what it does and how to do it / P.J. Werbos // Proceedings of the IEEE. – 1990. – Т. 78, № 10. – P. 1550–1560. DOI: 10.1109/5.58337
- A study of cross-validation and bootstrap for accuracy estimation and model selection / R. Kohavi [et al.] // Ijcai. – 1995. – Т. 14, № 2. – P. 1137–1145.
- Cross-Validation / Berrar D. [et al.]. – 2019. DOI: 10.1016/B978-0-12-809633-8.20349-X
- A Comparative Analysis of Cross-Validation Techniques for a Smart and Lean Pick-and-Place Solution with Deep Learning / E. Kee [et al.] // Electronics. – 2023. – Т. 12, № 11. – P. 2371. DOI: 10.3390/electronics12112371
- Using and understanding cross-validation strategies. Perspectives on Saeb et al / M.A. Little [et al.] // GigaScience. – 2017. – Т. 6, № 5. – С. gix020. DOI: 10.1093/gigascience/gix020
- On the need of preserving order of data when validating within-project defect classifiers / D. Falessi [et al.] // Empirical Software Engineering. – 2020. – Т. 25. – P. 4805–4830. DOI: 10.1007/s10664-020-09868-x
- Time series forecasting of univariate agrometeorological data: a comparative performance evaluation via one-step and multi-step ahead forecasting strategies / S. Suradhaniwar [et al.] // Sensors. – 2021. – Т. 21, № 7. – P. 2430. DOI: 10.3390/s21072430.

Заключение

1. Целью исследования было использование алгоритмов машинного обучения, в частности LSTM, для проведения долгосрочных прогнозов добычи нефти. Методика включала использование перекрестной проверки временных рядов с изменяющимся размером окна для определения оптимальной комбинации гиперпараметров для модели LSTM.

2. Результаты исследования показали, что одномерная одношаговая двунаправленная модель LSTM продемонстрировала значительно более низкую частоту ошибок по сравнению с моделью рекурсивного одношагового прогнозирования и с моделью многошагового прогнозирования. Однако важно понимать, что одномерная одношаговая двунаправленная модель LSTM имеет практические ограничения. Во время прогнозирования пары входных и выходных данных генерируются с использованием реальных данных. Этот аспект следует учитывать при оценке его применимости в реальных сценариях.

3. С другой стороны, рекурсивная модель продемонстрировала тенденцию к занижению значений за трехлетний период. Эта характеристика подразумевает склонность к риску, что делает ее потенциально надежным выбором для обоснования будущих решений нефтегазовой компании.

4. Будущее направление может включать изучение моделей многомерного прогнозирования, которые включают динамические данные. Это расширение могло бы повысить точность прогнозов за счет включения дополнительных соответствующих факторов.

31. Zhang, X. Optimal model averaging based on forward-validation / X. Zhang, X. Zhang // Journal of Econometrics. – 2022. DOI: 10.1016/j.jeconom.2022.03.010
32. Hyperparameter tuning and pipeline optimization via grid search method and tree-based autoML in breast cancer prediction / S.F.M. Radzi [et al.] // Journal of personalized medicine. – 2021. – Т. 11, № 10. – P. 978. DOI: 10.3390/jpm11100978
33. Belete, D.M. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results / D.M. Belete, M.D. Huchaiah // International Journal of Computers and Applications. – 2022. – Т. 44, № 9. – P. 875–886. DOI: 10.1080/1206212X.2021.1974663
34. Shekar, B.H. Grid search-based hyperparameter tuning and classification of microarray cancer data / B.H. Shekar, G. Dagnev // 2019 second international conference on advanced computational and communication paradigms (ICACCP). – IEEE, 2019. – С. 1–8. DOI: 10.1109/ICACCP.2019.8882943
35. Bergstra, J. Random search for hyper-parameter optimization / J. Bergstra, Y. Bengio // Journal of machine learning research. – 2012. – Т. 13, № 2. DOI: 10.5555/2188385.2188395
36. Liashchynskiy, P. Grid search, random search, genetic algorithm: a big comparison for NAS / P. Liashchynskiy, P. Liashchynskiy // arXiv preprint arXiv:1912.06059. – 2019. DOI: 10.48550/arXiv.1912.06059
37. St-Aubin, P. Precision and Reliability of Forecasts Performance Metrics / P. St-Aubin, B. Agard // Forecasting. – 2022. – Т. 4, № 4. – P. 882–903. DOI: 10.3390/forecast4040048
38. Cerqueira, V. Evaluating time series forecasting models: An empirical study on performance estimation methods / V. Cerqueira, L. Torgo, I. Mozetič // Machine Learning. – 2020. – Т. 109. – P. 1997–2028. DOI: 10.1007/s10994-020-05910-7
39. Chai, T. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature / T. Chai, R.R. Draxler // Geoscientific model development. – 2014. – Т. 7, № 3. – P. 1247–1250. DOI: 10.5194/gmd-7-1247-2014
40. Park, Y.H. Concise logarithmic loss function for robust training of anomaly detection model / Y.H. Park // arXiv preprint arXiv:2201.05748. – 2022. DOI: 10.48550/arXiv.2201.05748

References

1. Petrakov D.G. et al. Standards for Selection of Surfactant Compositions used in Completion and Stimulation Fluids. *International Journal of Engineering*, 2023, vol. 36, no. 9, pp. 1605-1610. DOI: 10.5829/IJE.2023.36.09C.03
2. Aghaei H. et al. Density-dependent relationship between changes in ultrasonic wave velocities, effective stress, and petrophysical-elastic properties of sandstone. *Ultrasonics*, 2023, vol. 132, 106985 p. DOI: 10.1016/j.ultras.2023.106985
3. Petrakov D.G., Penkov G.M., Zolotukhin A.B. Experimental study on the effect of rock pressure on sandstone permeability. *Zapiski Gornogo institute*, 2022, vol. 254, pp. 244-251. DOI: 10.31897/PMI.2022.24
4. Negash B.M., Yaw A.D. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection. *Petroleum Exploration and Development*, 2020, T. 47, no. 2, pp. 383-392. DOI: 10.1016/S1876-3804(20)60055-6
5. Berneti S.M., Shabbazian M. An imperialist competitive algorithm artificial neural network method to predict oil flow rate of the wells. *International journal of computer applications*, 2011, vol. 26, no. 10, pp. 47-50. DOI: 10.5120/3137-4326
6. Falcone G., Alimonti C. The challenges of multiphase flow metering: today and beyond. *International Conference on Offshore Mechanics and Arctic Engineering*, 2007, vol. 42681, pp. 823-834. DOI: 10.1115/OMAE2007-29527
7. Mirzaei-Paiaman A., Salavati S. The application of artificial neural networks for the prediction of oil production flow rate. *Energy Sources, Part A: Recovery, utilization, and environmental effects*, 2012, vol. 34, no. 19, pp. 1834-1843. DOI: 10.1080/15567036.2010.492386
8. Gilbert W.E. Flowing and Gas-Lift Well Performance. *Drilling and Production Practice*, 1954, vol. 13, pp. 126-157.
9. Baxendell P.B. Bean performance-lake wells. Shell Internal Report. Houston, TX: Shell Oil, 1957.
10. Ros N.C.J. An analysis of critical simultaneous gas/liquid flow through a restriction and its application to flowmetering. *Applied Scientific Research*, 1960, vol. 9, pp. 374-388. DOI: 10.1007/BF00382215
11. Achong L.B. Revised bean and performance formula for Lake Maracaibo wells. Shell Internal Report. Houston, TX: Shell Oil Co, 1961.
12. Mirzaei-Paiaman A., Salavati S. A new empirical correlation for sonic simultaneous flow of oil and gas through wellhead chokes for Persian oil fields. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2013, vol. 35, no. 9, pp. 817-825. DOI: 10.1080/15567031003773304
13. Zhang P.Y., Meng-Meng Z. The Application of BP Neural Network In Oil-Field. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 2013, vol. 11, no. 9, pp. 527-75283. DOI: 10.11591/telkomnika.v11i9.3280
14. Liu W., Liu W.D., Gu J. Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network. *Journal of Petroleum Science and Engineering*, 2020, vol. 189, 107013 p. DOI: 10.1016/j.petrol.2020.107013
15. Ali P.J.M. et al. Data normalization and standardization: a technical report. *Mach Learn Tech Rep*, 2014, vol. 1, no. 1, pp. 1-6. DOI: 10.13140/RG.2.2.28948.04489
16. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors. *Nature*, 1986, vol. 323, no. 6088, pp. 533-536. DOI: 10.1038/323533a0
17. Werbos P.J. Generalization of backpropagation with application to a recurrent gas market model. *Neural networks*, 1988, vol. 1, no. 4, pp. 339-356. DOI: 10.1016/0893-6080(88)90007-X
18. Elman J.L. Finding structure in time. *Cognitive science*, 1990, vol. 14, no. 2, pp. 179-211. DOI: 10.1016/0364-0213(90)90002-E
19. Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 1994, vol. 5, no. 2, pp. 157-166. DOI: 10.1109/72.279181
20. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural computation*, 1997, vol. 9, no. 8, pp. 1735-1780. DOI: 10.1162/neco.1997.9.8.1735
21. Gers F.A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM. *Neural computation*, 2000, vol. 12, no. 10, pp. 2451-2471. DOI: 10.1162/089976600300015015
22. Staudemeyer R.C., Morris E.R. Understanding LSTM—a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019. DOI: 10.48550/arXiv.1909.09586
23. Understanding LSTM networks by Colah accessed at Understanding LSTM Networks -- colah's blog.
24. Werbos P.J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 1990, vol. 78, no. 10, pp. 1550-1560. DOI: 10.1109/5.58337
25. Kohavi R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 1995, vol. 14, no. 2, pp. 1137-1145.
26. Berrar D. et al. Cross-Validation, 2019. DOI: 10.1016/B978-0-12-809633-8.20349-X
27. Kee E. et al. A Comparative Analysis of Cross-Validation Techniques for a Smart and Lean Pick-and-Place Solution with Deep Learning. *Electronics*, 2023, vol. 12, no. 11, 2371 p. DOI: 10.3390/electronics12112371
28. Little M.A. et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience*, 2017, vol. 6, no. 5, gix020 p. DOI: 10.1093/gigascience/gix020
29. Falessi D. et al. On the need of preserving order of data when validating within-project defect classifiers. *Empirical Software Engineering*, 2020, vol. 25, pp. 4805-4830. DOI: 10.1007/s10664-020-09868-x
30. Suradhaniwar S. et al. Time series forecasting of univariate agrometeorological data: a comparative performance evaluation via one-step and multi-step ahead forecasting strategies. *Sensors*, 2021, vol. 21, no. 7, pp. 2430. DOI: 10.3390/s21072430
31. Zhang X., Zhang X. Optimal model averaging based on forward-validation. *Journal of Econometrics*, 2022. DOI: 10.1016/j.jeconom.2022.03.010
32. Radzi S.F.M. et al. Hyperparameter tuning and pipeline optimization via grid search method and tree-based autoML in breast cancer prediction. *Journal of personalized medicine*, 2021, vol. 11, no. 10, 978 p. DOI: 10.3390/jpm11100978
33. Belete D.M., Huchaiah M.D. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers and Applications*, 2022, vol. 44, no. 9, pp. 875-886. DOI: 10.1080/1206212X.2021.1974663
34. Shekar B.H., Dagnev G. Grid search-based hyperparameter tuning and classification of microarray cancer data. *2019 second international conference on advanced computational and communication paradigms (ICACCP)*, IEEE, 2019, pp. 1-8. DOI: 10.1109/ICACCP.2019.8882943
35. Bergstra J., Bengio Y. Random search for hyper-parameter optimization. *Journal of machine learning research*, 2012, vol. 13, no. 2. DOI: 10.5555/2188385.2188395
36. Liashchynskiy P., Liashchynskiy P. Grid search, random search, genetic algorithm: a big comparison for NAS. *arXiv preprint arXiv:1912.06059*, 2019. DOI: 10.48550/arXiv.1912.06059
37. St-Aubin P., Agard B. Precision and Reliability of Forecasts Performance Metrics. *Forecasting*, 2022, vol. 4, no. 4, pp. 882-903. DOI: 10.3390/forecast4040048
38. Cerqueira V., Torgo L., Mozetič I. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 2020, vol. 109, pp. 1997-2028. DOI: 10.1007/s10994-020-05910-7
39. Chai T., Draxler R.R. Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific model development*, 2014, vol. 7, no. 3, pp. 1247-1250. DOI: 10.5194/gmd-7-1247-2014
40. Park Y.H. Concise logarithmic loss function for robust training of anomaly detection model. *arXiv preprint arXiv:2201.05748*, 2022. DOI: 10.48550/arXiv.2201.05748

Финансирование. Исследование не имело спонсорской поддержки.
 Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.
 Вклад авторов равноценен.