

Использование методов машинного обучения для прогнозирования роста обводненности нефтяного месторождения

В.О. Карпушин

*Российский государственный университет нефти и газа
(национальный исследовательский университет)
имени И.М. Губкина, Москва*

Аннотация: В работе рассматриваются два метода машинного обучения для прогнозирования роста обводненности нефтяного месторождения с расчетом основных метрик сходимости.

Ключевые слова: обводненность, статистика, машинное обучение, временной ряд, добыча нефти.

Актуальность

Применение методов машинного обучения в нефтегазовой отрасли становится всё более востребованным, особенно в задачах анализа и прогнозирования технологических показателей скважин. Одним из таких параметров является обводненность, динамика которой напрямую влияет на эффективность добычи. В ряде исследований показано, что модели глубокого обучения, включая LSTM, а также алгоритмы типа Prophet, обладают высоким потенциалом при работе с временными рядами в подобных задачах [1].

В настоящей работе рассматривается прогнозирование изменения обводненности по фонду добывающих скважин на временном интервале в восемь месяцев. Особенностью кейса является нестабильность временного ряда: одна из скважин в определённый момент выводится из работы, а затем снова запускается, что отражается на структуре данных. При этом общий тренд параметра сохраняет восходящий характер.

Для решения задачи были применены модели LSTM и Prophet, позволившие учесть как тренд, так и сезонные и временные колебания. Эти

модели уже успели зарекомендовать себя в решениях задач нефтегазового сектора [2]. Результаты моделирования показали удовлетворительное качество прогноза при наличии неполных и неустойчивых данных, что подтверждается визуальным анализом и расчетом метрик точности. Полученные выводы могут быть полезны при построении систем поддержки принятия решений в области добычи нефти.

Выбор моделей

Для произведения расчетов рассматривались два open-source проекта:

- 1) Prophet – Пророк;
- 2) LSTM – long short-term memory (длинная/короткая память).

Первый вариант самый эффективный: этот проект почти полностью автоматизирован, тренды и сезонности ищутся самостоятельно внутри модели. В основе лежит анализ временного ряда – аддитивная модель, представляющая временной ряд как сумму трех компонентов:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (1)$$

где:

- $g(t)$ — тренд (линейный или логистический),
- $s(t)$ — сезонность (периодические колебания),
- $h(t)$ — влияние событий (например, праздников),
- ε_t — случайный шум.

Вдобавок он хорошо подходит к текущему кейсу тем, что у Prophet гибкая модель тренда с возможностью смены режима роста, моделируется линейно-кусочная функция с точками смены поведения, где скорость роста может меняться:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (2)$$

где:

- k — базовый наклон;
- δ — изменения наклона;

- $a(t)$ — индикатор изменения;
- m — начальное смещение тренда;
- γ - вектор изменений смещения;
- $a(t)^T \delta$ – сумма всех изменений наклона до момента t ;
- $a(t)^T \gamma$ - Сумма всех изменений смещения до момента.

Сезонные компоненты рассчитываются с помощью ряда Фурье, что позволяет использовать различные сезонные циклы.

Prophet можно попробовать использовать для таких задач как:

Прогноз добычи нефти по фонду скважин (нужно учитывать, изменения режимов работы скважин, смену оборудования, желательно брать стационарные отрезки, либо общий дебит по кустовой площадке или по всему добывающему фонду месторождения);

Прогноз потребления спроса на углеводороды. Здесь есть сезонность. Полезно для решения задач по логистике;

Анализ простоя оборудования, можно про-экспериментировать и выявить сезонности, которые не может увидеть первоочередный анализ специалиста-человека.

Второй вариант: LSTM - относится к классу рекуррентных нейронных сетей и предназначена для работы с последовательными данными. В отличие от классических RNN, она эффективно решает проблему исчезающего градиента и способна запоминать информацию на длительных интервалах [3].

Основные компоненты архитектуры:

Каждая LSTM-ячейка содержит четыре гейта/точки. Они управляют потоком информации, решая, что сохранить, что забыть и что передать дальше:

forget gate – точка «забывания»:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (3)$$

input gate – входная точка – внесение данных в память:

$$i_t = \sigma(W_i \cdot [h_t^{-1}, x_t] + b_i), \quad (4)$$

output gate – выходная точка – выдача данных + новое состояние ячейки.

$$o_t = \sigma(W_o \cdot [h_t^{-1}, x_t] + b_o), \quad (5)$$

Обновленное состояние памяти – точка, где суммируется старые данные + новые данные.

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

где:

- σ — сигмоида (функция активации);
- \tanh — гиперболический тангенс (ещё одна функция активации);
- W — веса для каждого гейта (например, W_f , W_i , W_c , W_o);
- b — соответствующие смещения (например, b_f , b_i и т. д.);
- x_t — вход в момент времени t ;
- h_t — скрытое состояние на шаге t ;
- C_t — состояние ячейки памяти (*cell state*).

Расчет прогноза

Исходные данные: одномерный массив по динамике обводненности месторождения Западной Сибири «Х», измеряющийся в долях единиц, покрывающий временной период за 250 дней или за восемь месяцев. Со сложной динамикой и локальными провалами, но с общим восходящим трендом – ряд не стационарен.



Рис. 1. – Исторические данные по изменению обводненности, д. ед.

Расчет по первому варианту

Для создания и обучения модели были предложены следующие настроечные параметры:

- Прямой, восходящий тренд в виде $kx+b$;
- Сезонность: поиск повторяющихся паттернов по дням недели. Здесь используются ряды Фурье (сумма синусов и косинусов), чтобы описать циклы;
- Шкалы изменения предыдущей точки/шага - вес параметра, который контролирует плавность смены направления тренда, в основе используется Байесовская регуляризация, штрафы за резкие скачки тренда;
- Добавлена месячная сезонность;

Для прогнозирования были сделаны следующие шаги:

- Создан дополнительный временной ряд на 30 дней вперед от последней даты;
- Для каждой даты модель вычисляет прогноз из тренда и сезонности;
- Границы ошибки.

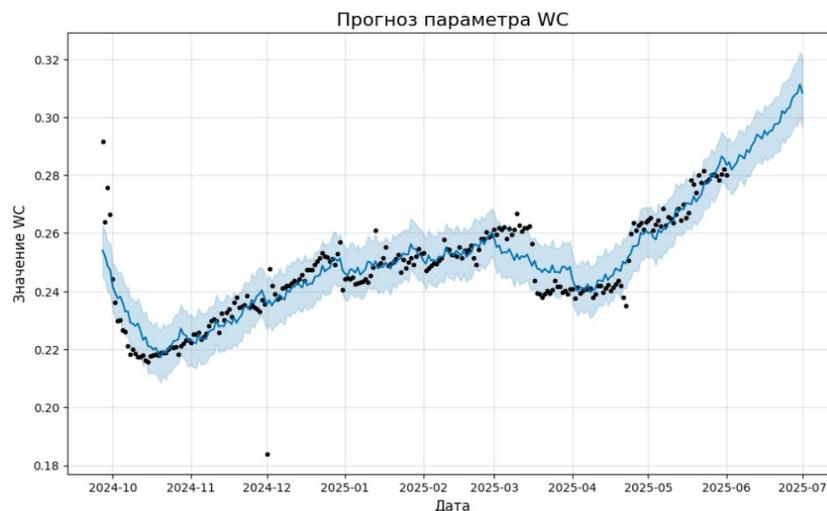


Рис. 2. – Адаптация модели на исторические данные.

Прогноз = Тренд + Недельная сезонность + месячная сезонность.

На рис.3 с осью trend/тренд – показывается общая долгосрочная динамика данных: рост, спад или стабильность.

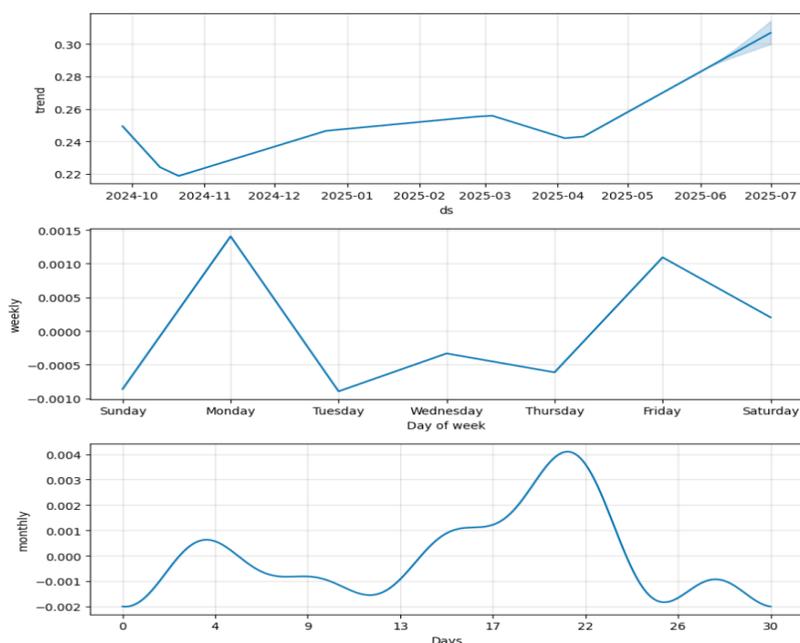


Рис. 3. – Структура временного ряда из модели.

С осью weekly/недельный – изменение данных в течение недели, данные группируются по дням недели (все понедельники, все вторники и т. д.). Для каждого дня вычисляется отклонение от общего тренда (не просто среднее, а с учетом динамики). Аппроксимируются эти отклонения суммой синусов и косинусов (ряды Фурье). Пример для понедельника: если в понедельники значения обычно на $+0.004$ выше тренда — это и будет недельной сезонностью.

Последний график на рис.3 monthly/месяц – изменение длиной в 30 дней. Ищутся циклы длиной в 30.5 дней - средняя длина месяца в выборке. Анализируется, как данные отклоняются от тренда каждые ~ 30 дней (не привязываясь к календарным месяцам). Используются те же ряды Фурье, но с другим периодом. Пример: если каждые 30 дней значения растут на $+0.0015$ — это месячный паттерн. Также учитываются периоды с перекрытием (например, с 15 января по 15 февраля, затем с 20 февраля по 20 марта и т. д.).

Черные точки на рис.2 – это фактические значения из выборки. Синим цветом отмечена модель и доверительный интервал с синей заливкой.

Что интересно – модель нашла недельную сезонность, рост показателя как в понедельник, так и в воскресенье несмотря на то, что добычные предприятия работают бесперебойно.



Рис. 4. – Прогноз роста обводненности с помощью модели «пророка».

Метрики ошибок модели:

- MAE (средняя абсолютная ошибка): 0.0032
- RMSE (среднеквадратичная ошибка): 0.0039
- MAPE (средняя абсолютная процентная ошибка): 1.200 %

Был произведен расчет прогноза значения обводненности примерно в диапазоне 0.26 – 0.27.

Выводы по первому варианту расчета:

- Средняя ошибка составила 0.003– 0.004, максимум около 0.007;
- Это соответствует ошибке менее 3% от абсолютных значений (MAPE \approx 1.2%);
- Прогноз близко следует за трендом фактических данных.

Расчет по второму варианту

По загрузке и обработке данных текущий вариант не отличается от первого за исключением дополнительно рассчитанной переменной "Trend" — нормированное количество дней от начала измерений.

Для обеспечения корректного обучения нейронной сети значения параметра WC и тренда были масштабированы с использованием нормализации на основе минимума-максимума в диапазоне [0, 1].

Для обучения модели LSTM из нормализованных данных были сформированы обучающие выборки методом скользящего окна. Размер окна был выбран равным 90, что позволяет модели учитывать информацию о динамике показателя за последние три месяца при каждом прогнозе.

На основе последовательностей из признаков были сгенерированы массивы признаков и целевых значений для обучения модели.

Использовались следующие настроечные параметры:

- Два слоя LSTM: первый с 128 нейронами, второй с 64 нейронами;
- Полносвязный слой с 32 нейронами и функцией активации ReLU (Rectifier linear unit);
- Финальный слой с 1 нейроном для регрессии.

Модель компилировалась с функцией потерь MSE (Mean Squared Error) и оптимизатором. Для предотвращения переобучения применялся метод ранней остановки (EarlyStopping) с параметром patience (*с англ. терпение*) = 15, отслеживающим валидационную ошибку.

Для прогнозирования значений на 30 дней вперёд была использована последняя обучающая последовательность. Чтобы обеспечить устойчивость прогноза и сохранить гладкость перехода, использовалось сглаживание прогноза, сочетающее выход модели LSTM и линейный тренд, полученный посредством регрессии по исходным данным.

Сначала преобладает вклад LSTM, но по мере продвижения во времени усиливается влияние линейного тренда. Это снижает риск расползания предсказаний за пределы физически разумного диапазона и обеспечивает устойчивость долгосрочного прогноза.

Были построены два графика:



Рис. 5. – Общий график, включающий исторические данные по обводненности и прогноз на месяц вперёд.



Рис. 6. – Фрагмент, фокусирующий на последних 60 днях и прогнозируемом отрезке, что позволяет визуально оценить плавность перехода между обучающей и прогнозной частью.

Выводы по второму варианту расчета

В рамках данной работы была реализована модель на основе долгой/краткосрочной памяти (LSTM) для прогнозирования роста обводнённости. В отличие от традиционных методов, использующих гидродинамические модели (ГДМ) или специализированные программные комплексы, прогноз осуществлялся исключительно на основе временного ряда без привлечения промышленных симуляторов.

По результатам обучения модели и последующего прогнозирования получены следующие метрики качества на обучающем множестве:

MAE (средняя абсолютная ошибка): 0.0068

RMSE (среднеквадратичная ошибка): 0.0097

MAPE (средняя абсолютная процентная ошибка): 2.75 %

Полученные значения ошибок указывают на удовлетворительную точность модели при воспроизведении общей динамики временного ряда. Визуальный анализ тренда показывает, что прогноз сохраняет линейный восходящий характер, однако не полностью отражает резкое увеличение обводнённости, наблюдаемое в конце исторических данных. Это указывает на потенциальную необходимость интеграции дополнительных факторов (например, производственных режимов, ввода новых скважин – дополнительных признаков) для повышения чувствительности модели к локальным изменениям [6].

Общие выводы

«Пророк» показал более высокую точность на ретроспективном отрезке, однако его прогностическая способность ограничена предположением о строго аддитивном поведении тренда и сезонности.

Таким образом, модель LSTM показала приемлемую точность в условиях отсутствия инженерных данных, сохранив при этом потенциал к улучшению за счёт включения дополнительных признаков и последующей калибровки архитектуры.

Литература

1. Хохрева А. А. Рекуррентные нейронные сети и их применение в задачах обработки последовательностей // Информационные технологии и вычислительные системы. — 2018. — № 1. — с. 65–74.

2. Рустамов А. Р., Пеньков Г. М., Петраков Д. Г., Рустамова М. А. Современные методы применения машинного обучения как инструмента прогнозирования добычи нефти // Недропользование. 2024. № 1, том 24 URL:



cyberleninka.ru/article/n/sovremennyye-metody-primeneniya-mashinnogo-obucheniya-kak-instrumenta-prognozirovaniya-dobychi-nefti

3. Топольников А. С. Машинное обучение для механизированной добычи нефти // Деловой журнал neftegaz.Ru. – 2021. – №. 5. – с. 14-19. URL: magazine.neftegaz.ru/articles/dobycha/682013-mashinnoe-obuchenie-dlya-mekhanizirovannoy-dobychi-nefti/

4. Галушка В. В., Фатхи В. А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных // Инженерный вестник Дона. – 2013. — №. 2. –URL: ivdon.ru/uploads/article/pdf/IVD_7_Galushka.pdf_1597.pdf

5. Клевцов С. И. Моделирование алгоритма краткосрочного прогнозирования изменения быстроизменяющейся физической величины в реальном времени //Инженерный вестник Дона. – 2012. – №. 3. – URL: ivdon.ru/uploads/article/pdf/2012_3_33.pdf_920.pdf

6. Еникеев М. Р. и др. Прогноз обводненности на проектируемых к бурению скважинах методами машинного обучения // Сборник трудов ИТНТ-2019. – 2019. – С. 21-24.

7. Легостаев Д. Ю., Косяков В. П. Моделирование добычи жидкости и динамики обводненности при совместном использовании фильтрационной модели и нейронных сетей // Вестник Тюменского государственного университета. Серия: Физико-математическое моделирование. Нефть, газ, энергетика.—2023.—Т. 9, №2(34). – 2023 URL: vestnik.utmn.ru/upload/iblock/bc8/ouf0c0dzmp2vghs6ud1xhgramlmosl65/75_92.pdf

8. Рыжов Р. В. Классификация и обзор методов определения причин обводнения скважин. Нефтяная провинция. 2024. № 3(39). С. 270-291 URL: vkro-raen.com/files/009/770/601/9770601/original/15-R.V.Ryzhov.pdf

9. Пономарев Р. Ю. и др. Оптимизация системы заводнения: преимущества гибридного подхода к разработке нейросетевых моделей фильтрации //Актуальные проблемы нефти и газа. – 2024. – Т. 15. – №. 4. – С. 349-363.

10. Кочнев А. А., Козырев Н. Д., Кочнева О. Е., Галкин С. В. Разработка комплексной методики прогноза эффективности геолого-технических мероприятий на основе алгоритмов машинного обучения // Георесурсы. 2020. №3. URL: cyberleninka.ru/article/n/razrabotka-kompleksnoy-metodiki-prognoza-effektivnosti-geologo-tehnicheskikh-meropriyatiy-na-osnove-algoritmov-mashinnogo-obucheniya

References

1. Hohreva A. A. Informacionnye tehnologii i vychislitel'nye sistemy. 2018. № 1. p. 65–74.

2. Rustamov A. R., Pen'kov G. M., Petrakov D. G., Rustamova M. A. Nedropol'zovanie. 2024. № 1, том 24 URL: cyberleninka.ru/article/n/sovremennye-metody-primeneniya-mashinnogo-obucheniya-kak-instrumenta-prognozirovaniya-dobychi-nefti

3. Topol'nikov A. S. Delovoj zhurnal neftegaz. Ru. 2021. №. 5. pp. 14-19.

4. Galushka V. V., Fathi V. A. Inzhenernyj vestnik Dona. 2013. №. 2. URL: ivdon.ru/uploads/article/pdf/IVD_7_Galushka.pdf_1597.pdf

5. Klevcov S. I. Inzhenernyj vestnik Dona. 2012. №. 3. URL: ivdon.ru/uploads/article/pdf/2012_3_33.pdf_920.pdf

6. Enikeev M. R. i dr. Sbornik trudov ITNT-2019. 2019. p. 21-24.

7. Legostaev D. Ju., Kosjakov V. P. Fiziko-matematicheskoe modelirovanie. Neft', gaz, jenergetika. 2023 T. 9, №2(34). 2023. URL: vestnik.utmn.ru/upload/iblock/bc8/ouf0c0dzmp2vghs6ud1xhgramlmosl65/75_92.pdf



8. Ryzhov R. V. Neftjanaja provincija. 2024. № 3(39). pp. 270-291. URL: vkro-raen.com/files/009/770/601/9770601/original/15-R.V.Ryzhov.pdf

9. Ponomarev R. Ju. Aktual'nye problemy nefti i gaza. 2024. Т. 15. №. 4. pp. 349-363.

10. Kochnev A. A., Kozyrev N. D., Kochneva O. E., Galkin S. V. Georesursy. 2020. №3. URL: cyberleninka.ru/article/n/razrabotka-kompleksnoy-metodiki-prognoza-effektivnosti-geologo-tehnicheskikh-meropriyatij-na-osnove-algoritmov-mashinnogo-obucheniya.

Дата поступления: 15.07.2025

Дата публикации: 25.08.2025