

ПРАКТИКА ПО КУРСУ «PYTHON: ОТ ОСНОВ ДО МАШИННОГО ОБУЧЕНИЯ»

Бахмутов Алексей Викторович

Борисов Владимир Вячеславович

Волков Егор Дмитриевич

Охтин Никита Сергеевич

07.06.2025



Цель работы

Разработать инструмент для оперативного прогнозирования динамики дебита нефтяных скважин. Обеспечить высокую точность прогноза при минимальном времени вычислений для поддержки принятия решений в управлении добычей.

Предлагаемое решение

LSTM-модель, обученная на исторических данных добычи с предобработкой (нормализация, формирование временных последовательностей). Архитектура: два LSTM-слоя (256 нейронов) с Dropout для регуляризации, выходной Dense-слой. Обучение на 80% данных с окном в 100 дней.



Почему именно LSTM-модель?

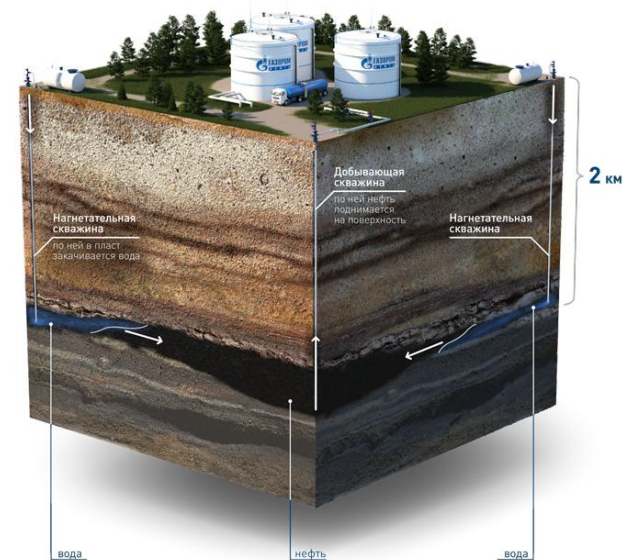
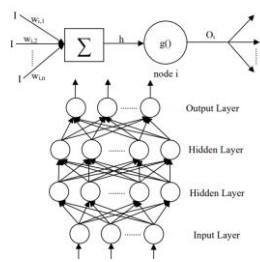
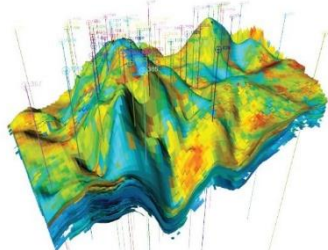
LSTM эффективно обрабатывает долгосрочные зависимости во временных рядах благодаря механизму "забывания/запоминания". Это критично для прогнозирования дебита, где ключевые паттерны могут быть разделены месяцами/годами (сезонность, изменение динамики после проведения геолого-технических мероприятий).

АКТУАЛЬНОСТЬ ПРОБЛЕМЫ

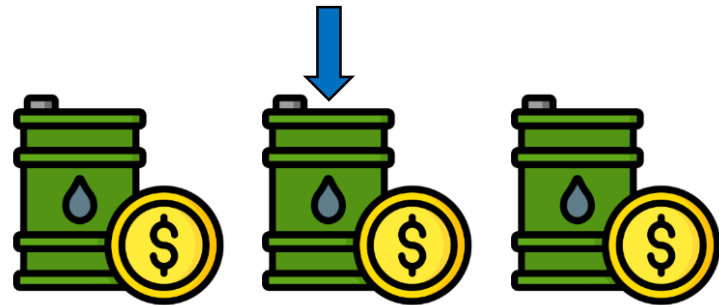
- Методы искусственного интеллекта все активнее проникают в различные области, связанные с анализом, прогнозированием и оптимизацией различных технологических процессов
- Не является исключением и область разработки месторождений нефти газа
- Использование искусственных нейронных сетей позволяет выявлять закономерности поведения и взаимное влияние одних параметров на другие
- Традиционные ГДМ сложно использовать в задачах оперативного управления из-за ограничений, связанных с длительностью расчетов и высокой ресурсоемкостью

ГДМ vs ИИ

Характеристика	Гидродинамическое моделирование (ГДМ)	Нейронная сеть (ИИ)
Скорость расчета	Очень низкая	Очень высокая
Требования к выч. ресурсам	Очень высокие	Умеренные
Пригодность для оперативного управления	Непрактично из-за длительного времени расчета	Идеально подходит благодаря скорости прогноза
Анализ больших данных	Ограничена фокусом на физической модели	Высокая эффективность
Точность (в идеале)	Очень высокая (если модель хорошо адаптирована)	Зависит от качества и объема данных для обучения
Основа прогноза	Физические законы	Закономерности, выявленные в данных
Генерация обучающих данных	Источник (создает данные для обучения нейросети)	Потребитель (требуется готовых данных для обучения)
Стоимость (эксплуатация)	Высокая	Умеренная/Низкая (после обучения; меньше ресурсов)
Гибкость (адаптация)	Сложная и долгая	Относительно простая



💡 Для задач оперативного управления добычей и быстрой оптимизации режимов разработки нейронные сети, обученные на данных ГДМ, предлагают решающее преимущество в скорости при сохранении приемлемой точности, преодолевая главные ограничения традиционного ГДМ.



ХОД РАБОТЫ

1. Подготовка данных:

Загрузка данных: исторические данные в формате CSV (столбцы Date и Production)

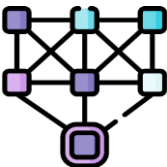
Предобработка: масштабирование значений Production в диапазон [0, 1] для ускорения сходимости обучения LSTM и предотвращение «исчезающих градиентов»

Разделение данных: 80% тренировочные, 20% тестовые



2. Архитектура нейронной сети:

Слой	Параметры	Назначение
LSTM-1	256 нейронов	Извлечение первичных временных закономерностей с сохранением последовательности
Dropout	20%	Случайное отключение 20% нейронов для предотвращения переобучения
LSTM-2	256 нейронов	Анализ долгосрочных зависимостей
Dropout	20%	Дополнительная регуляризация
Dense	25 нейронов	Извлечение высокоуровневых признаков
Dense	1 нейрон, линейная активация	Финальное преобразование(прогноз добычи нефти на следующий день)

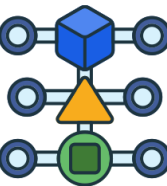


3. Обучение модели:

Основная задача: Оптимизация параметров модели (весов) для минимизации ошибки прогноза

Адаптация к данным, выявление закономерностей, обобщение: обеспечение работоспособности на новых данных (не участвовавших в обучении)

Оптимизатор: Adam (адаптивная скорость обучения) - автоматическая адаптация, устойчивость к шуму в реальных данных добычи, баланс между скоростью сходимости и итоговым качеством



4. Прогнозирование и оценка:

Обратное масштабирование

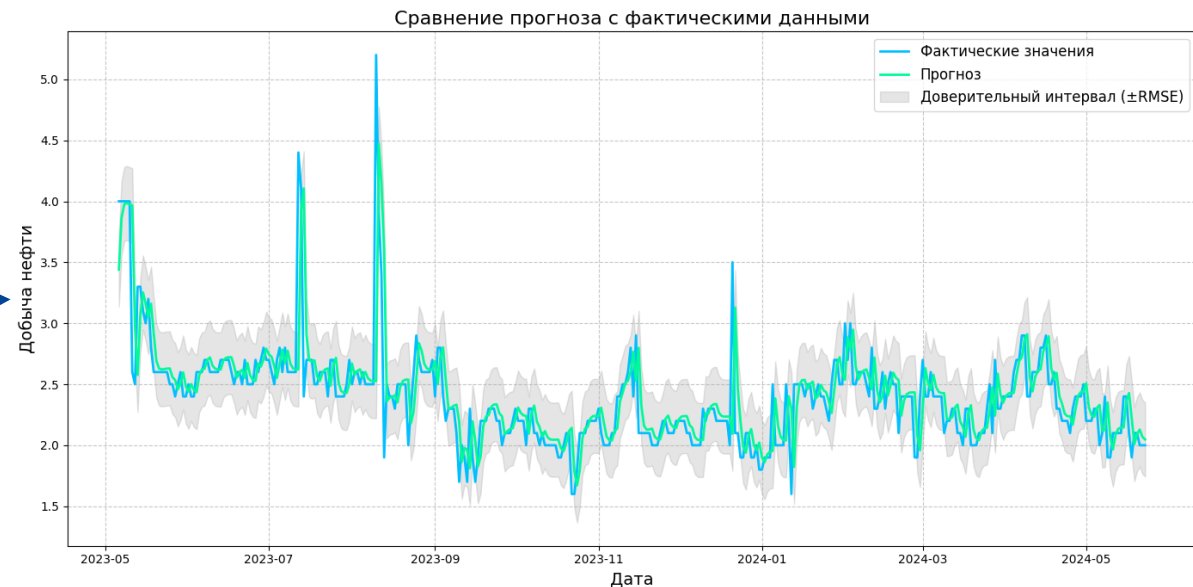
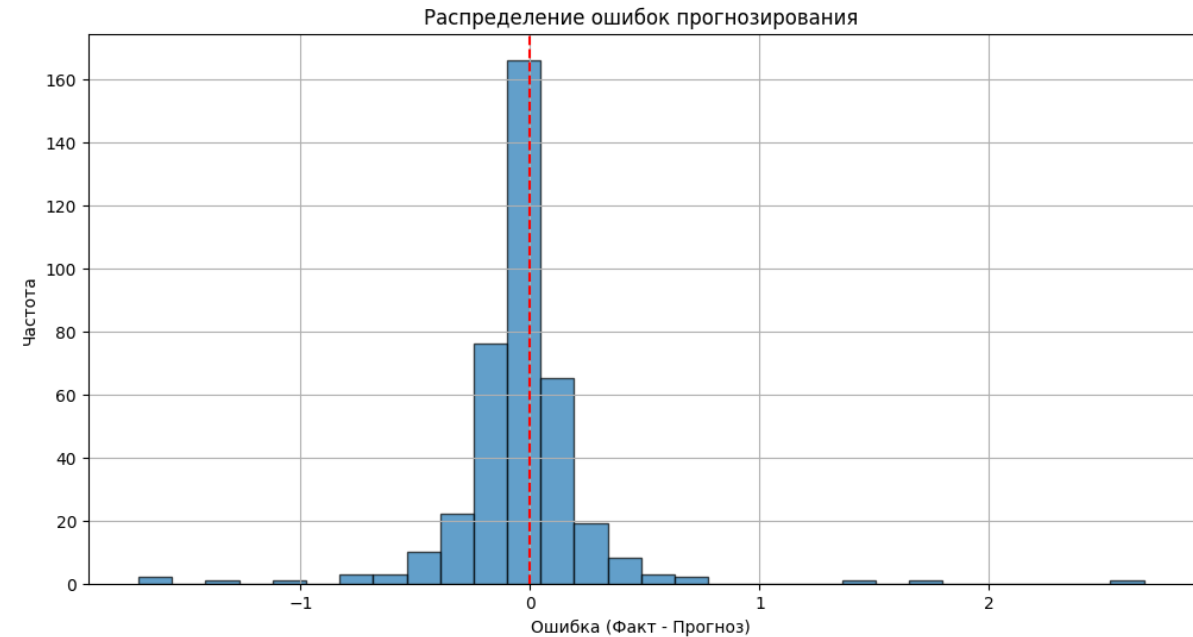
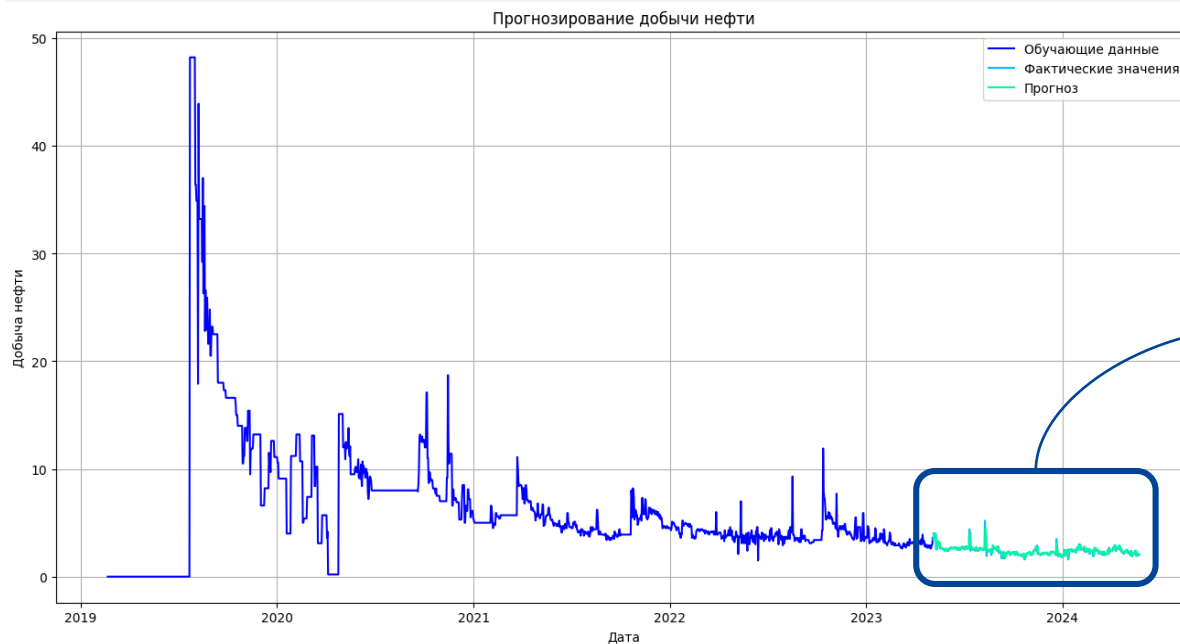
Наложение прогноза на реальные значения тестовой выборки



РЕЗУЛЬТАТЫ

В ходе выполнения работы были получены следующие результаты:

- В ходе проведения многократных расчетов метрики качества принимали следующие значения: **RMSE = 0.35 - 0.65**, **MAE = 0.3 - 0.6**, что свидетельствует о высокой прогнозной способности;
- Значения ошибок прогнозирования **минимальные** и при хорошем подборе параметров они распределяются **симметрично около нуля**, следовательно модель стабильна;
- Количество эпох **перестает сильно влиять** на точность расчета начиная с 30;
- С помощью модели LSTM получается достаточно точно прогнозировать временные ряды, в нашем случае дебит нефти.



ВЫВОДЫ И РЕКОМЕНДАЦИИ

Разработанная в ходе практики модель доказала свою эффективность в решении конкретной производственной задачи - точном краткосрочном прогнозировании дебита нефти. Ее внедрение и дальнейшее развитие имеют значительный потенциал для повышения операционной эффективности и обоснованности управленческих решений в области добычи углеводородов.

Развитие модели:

Хотя модель эффективна для краткосрочного прогноза, рекомендуется продолжить работу для повышения точности среднесрочных прогнозов. Это может включать:

1. Углубленный Feature Engineering (создание более сложных признаков, учитывающих геологию, историю ГТМ).
2. Эксперименты с другими архитектурами моделей (например, ансамбли, более сложные нейросети) или гибридными подходами.
3. Привлечение дополнительных данных (геофизические данные, информация о плановых ГТМ, данные по закачке).

