



Нефтесервисная компания

17 марта 2025 г

# Презентация решения

Разработка решения для обнаружения бинарных признаков в данных гидродинамических исследований скважин

# Команда: Дайте Два

Трек: Разработка решения для обнаружения бинарных признаков в данных гидродинамических исследований скважин



Нефтесервисная компания

# Проблема

Перед нами поставлена задача:

Поиск бинарных признаков на данных  
диагностических графиков с определением  
их численных характеристик



Нефтесервисная компания

# Гипотеза

Автоматизация анализа диагностических графиков скважин позволит значительно ускорить процесс интерпретации данных, снизить субъективность и повысить точность выявления ключевых режимов течения и особенностей пласта.

Разработка алгоритма для определения бинарных признаков и их численных характеристик обеспечит единый стандарт качества анализа и станет основой для более сложных моделей оценки параметров скважин и пластов.



Нефтесервисная компания

# Исходные данные

1. Данные представляют собой временные ряды давления и его производной для более 45 000 исследований скважин, каждый файл содержит три колонки:

- время,
- изменение давления ( $\Delta P$ ),
- производную давления ( $P'$ ).

2. Разметка включает 8 бинарных признаков, характеризующих режимы течения и особенности пласта, а также численные коэффициенты, указывающие на ключевые моменты проявления этих признаков.



Нефтесервисная компания



# Дизайн исследования

1. Исследование включает анализ временных рядов давления и его производной для автоматического выявления 8 ключевых режимов течения и особенностей пласта на основе диагностических графиков.
2. Для каждого режима определяется бинарный признак (наличие/отсутствие) и численный коэффициент, характеризующий момент его проявления, что позволит стандартизировать процесс интерпретации данных.



Нефтесервисная компания

# Обработка исходных данных

1. Исходные данные включают файл с разметкой в формате CSV, содержащий информацию о бинарных признаках и численных коэффициентах, а также отдельный каталог «/data» с более 45 000 текстовых файлов, каждый из которых содержит временные ряды давления и его производной для конкретного исследования.
2. Для обработки данных файлы из каталога «/data» были объединены с разметкой по уникальным идентификаторам (UUID), что позволило создать единый датасет для анализа и обучения модели.



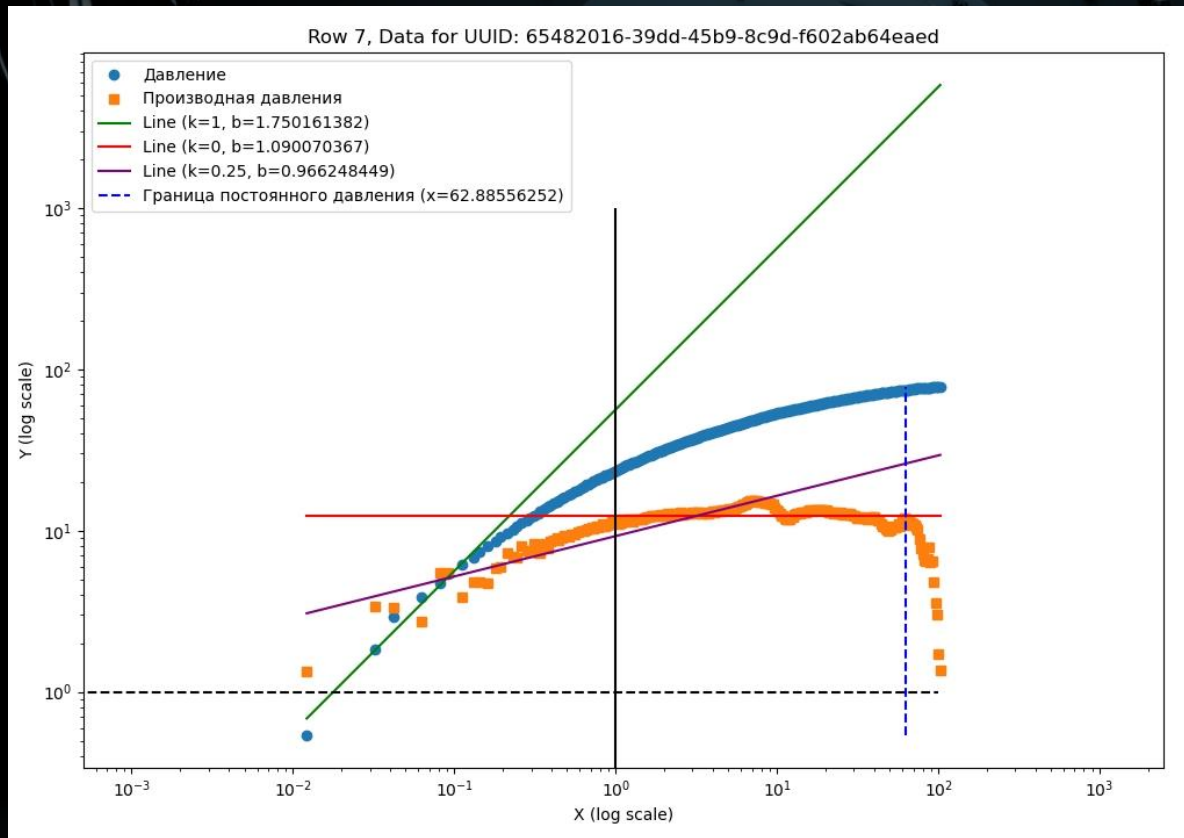
Нефтесервисная компания

# Анализ данных (EDA)

Было  
просмотрено  
много графиков  
для выявления  
особенностей  
работы скважин



Нефтесервисная компания





# Feature engineering

1. Для подготовки признаков были рассчитаны базовые статистики (среднее, медиана, стандартное отклонение) для каждого временного ряда, а также их разности (первые производные) и соответствующие статистики для этих разностей.

2. Для рядов давления  $dP$  и  $P$  дополнительно вычислялось скользящее среднее и проверялся монотонный рост или плато, а также определялось процентное соотношение выбросов в пределах 1, 2 и 3 стандартных отклонений от среднего.

3. Модуль `tsfresh` для генерации признаков.



Нефтесервисная компания

# Апробированные подходы

## Алгоритмы/Модели

1. Для решения задачи multi-label классификации были апробированы CatBoostClassifier и LGBMClassifier, обёрнутый в MultiOutputClassifier, а для регрессии — LGBMRegressor обёрнутый в MultiOutputRegressor.
2. Также тестировался подход с использованием библиотеки LightAutoML (TabularAutoML), однако его результаты оказались менее точными по сравнению с CatBoost и LightGBM.
3. Для характеристики 'Влияние ствола скважины\_details' была построена математическая модель на основе статистик, которая выдает результат лучше регрессора. Для построения моделей по остальным численным признакам не хватило времени.



Нефтесервисная компания

# Финальная модель

1. Для предсказания бинарных признаков использовался `LGBMClassifier`, обученных на 5 фолдах со стратификацией по метке составленной из бинарных признаков. Результирующая метка определялась голосованием.
2. Для предсказания численных характеристик к исходным признакам были добавлены бинарные признаки, полученные классификатором и подавались в `LGBMRegressor`. Регрессор обучался на 5 фолдах со стратификацией по метке из бинарных признаков. Результирующая характеристика бралась как среднее 5 моделей.



Нефтесервисная компания

# Полученные результаты

№	Модель	Результат
1	LightAutoML предсказание одного признака	0.3125
2	LightAutoML предсказание всех признаков	0.3599
3	CatBoostClassifier + LGBMRegressor	0.6302
4	LGBMClassifier + LGBMRegressor	0.6499
5	LGBMClassifier + LGBMRegressor + BIN targets	0.6583



Нефтесервисная компания

# Инфраструктура и инструменты используемые для проведения экспериментов

Опционально

1. Python 3.11 + Pandas + Sklearn + tsfresh
2. Обучение классификаторов на 5 фолдах  
выполняется за 1ч 12мин.
3. Обучение регрессоров на 5 фолдах  
выполняется за 43мин.



Нефтесервисная компания



# Презентация решения

1. Установить библиотеки из requirements.txt
2. Склонировать себе репо.
3. Задать путь к тестовым временным рядам в скрипте model\_inference.py:

```
test_data_path = '/valid'
```

4. Выполнить скрипт model\_inference.py
5. Дождаться результата :)



Нефтесервисная компания



Нефтесервисная компания

Спасибо  
за внимание!

