

Прогнозирование эксплуатационных характеристик электроцентробежных насосов

Выполнила: Анастасия Быкова

Руководитель: Сергей Абдуракипов
PhD, Руководитель лаборатории ИИ

Краткое описание

Создание основы для диагностики состояния ЭЦН в целях повышения эффективности работы скважины и снижения потерь добычи



ЭЦН – наиболее распространенный способ механизированный добычи

Продление ресурса работы оборудования и снижение затрат на ремонт

Фонд до 80 %



Высокая неопределенность геолого-промышленных данных

Получение ценной информации о технологическом процессе, обнаружение трендов

?



Оперативность управления и принятия решений

Принятие превентивных мер, снижение простоя скважин и потерь добычи



Разработка модели для прогнозирования параметров работы ЭЦН позволит своевременно предупреждать о возможных отказах оборудования

Формирование ML-моделей

Задача: прогноз средней скорости изменения давления на приеме ЭЦН на основе исторических данных

Постановка задачи и анализ данных

Формулирование задачи.
Формирование датасета,
анализ корреляций, отбор
признаков

1

Очистка данных, обработка пропущенных значений, кодирование категориальных данных и нормализация количественных признаков при необходимости

2

Построение и обучение моделей

3

Выбор и построение моделей, подходящих для задачи регрессии временных рядов

Валидация, оценка и дообучение

Расчет метрик R², MAE на тестовой выборке. Сравнение результатов. Выбор и донастройка наиболее оптимальной модели

4

5

Внедрение

Предобработка
данных

В рамках работы выполнены загрузка и анализ технологических показателей работы скважин, первичный анализ корреляционных зависимостей, подготовка и запуск моделей, сравнение результатов

Сбор и подготовка данных

Первичный анализ исходных данных

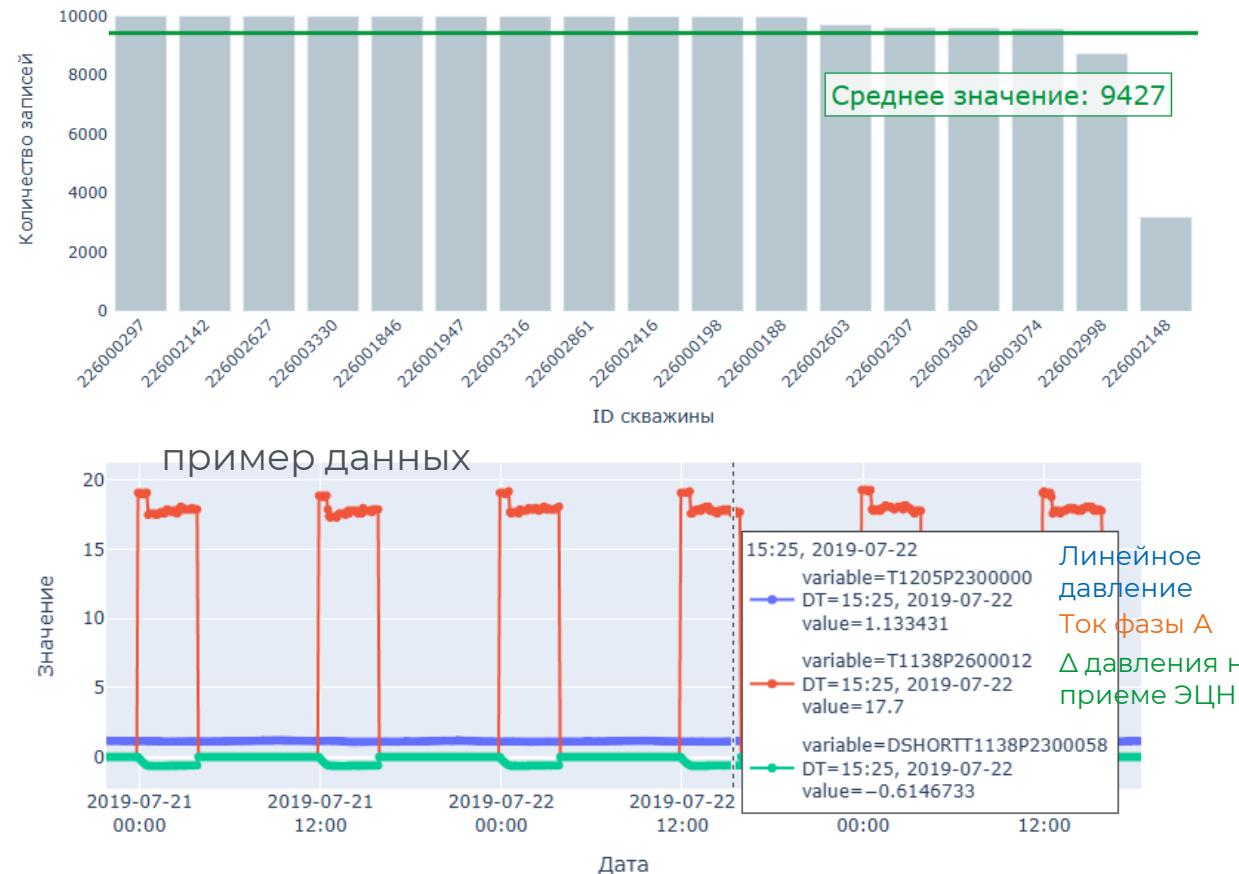
17 периодических скважин

26 параметров

160 262 записей
за период: 27.06.2019-31.07.2019
с частотой: 5 минут



Исключение параметра из
обучающей выборки?

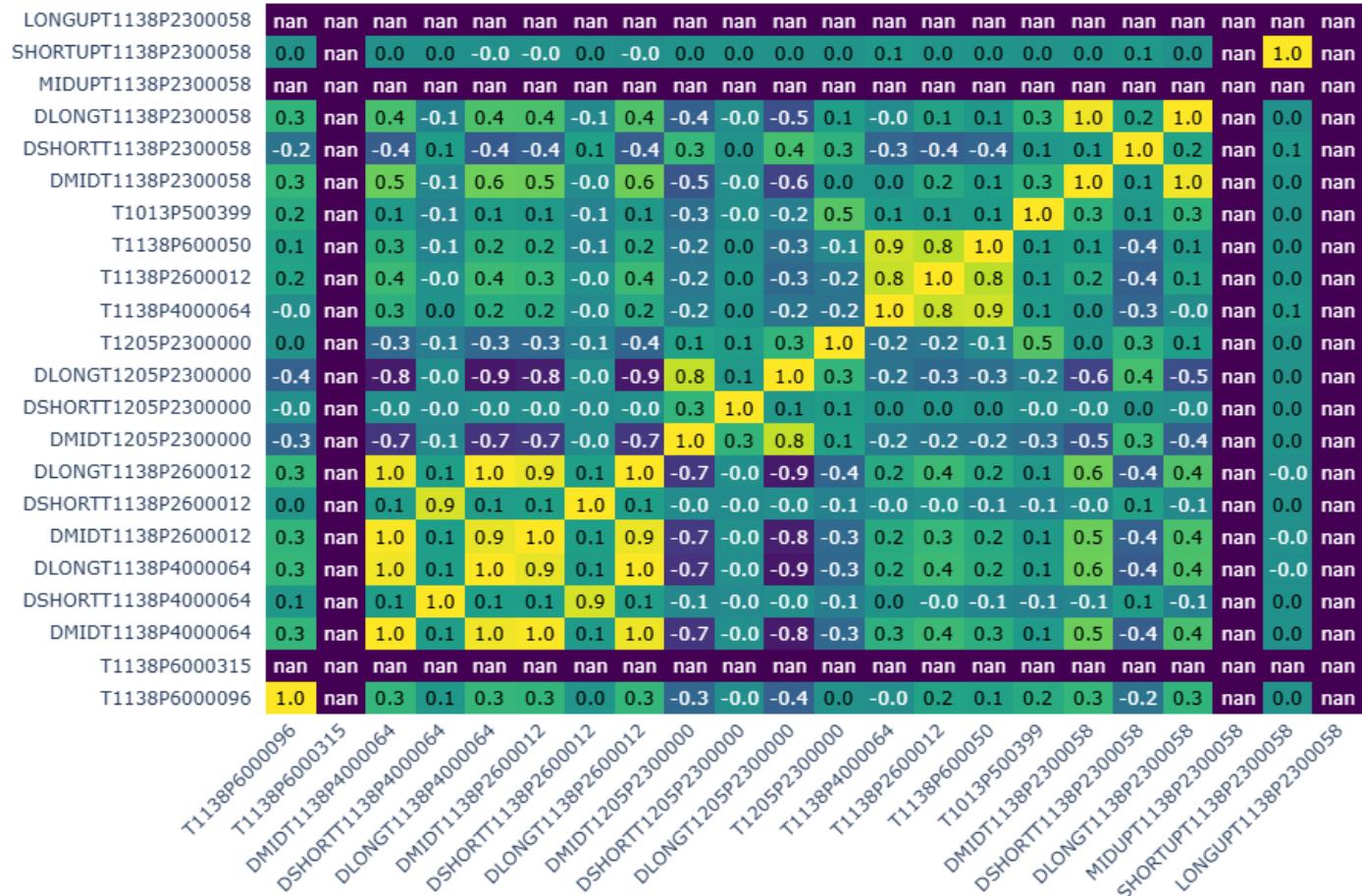


Исходные данные даны для 17 периодических скважин за 1 месяц

Сбор и подготовка данных

Анализ корреляций

Матрица коэффициентов корреляции Пирсона



Признаки

Код	Описание
T1205P2300000	Линейное давление, МПа
T1138P4000064	Загрузка двигателя, %
T1138P2600012	Ток фазы А двигателя, А (I)
T1138P600050	Коэффициент мощности ($\cos \varphi$)
T1013P500399	Суточный объем жидкости, м ³
T1138P2300058	Давление на приеме насоса, МПа - таргет
T1138P6000096	Наработка двигателя с последнего включения, с
T1138P6000315	Время простоя двигателя с последнего выключения, с
DMID-, DSHORT-, DLONG-	Часовая скорость изменения, усредненная в СУТ, ЧАС, НЕДЕЛЯ
MIDUPT-, SHORTUPT-, LONGUPT-	?

Сила тока коррелирует с загрузкой и коэффициентом мощности двигателя, что может быть обусловлено паспортными характеристиками двигателя

Выбор ML-моделей

Обучение разных типов моделей: от классических алгоритмов до современных архитектур

Категория	Выбранные модели	Примечание
Классические ансамбли	Random Forest XGBoost LightGBM	Использование в качестве Baseline. Оценка достаточности признаков и временного окна для прогнозирования Особенность: Поиск простых, статистических связей, высокая скорость обучения
Рекуррентные и сверточные сети	LSTM TCN (Temporal Convolutional Network)	Выявление и учет долго- и краткосрочных локальных зависимостей Особенность: Прямая работа с последовательностями, автоматическое извлечение временных паттернов
Современные архитектуры	Transformer N-BEATS	Обнаружение скрытых и нелинейных зависимостей Особенность: Разложение временного ряда на базовые компоненты (N-BEATS), выявление сложных взаимосвязей между разнородными потоками данных (Transformer – рекомендуется при наличии большого пакета данных и вычислительных ресурсов)

Применение различных типов моделей обусловлено различиями в условиях эксплуатации: различный набор параметров и их доступность, отличия в характеристиках насосов

Сопоставление результатов

Выбор оптимальной модели

Принцип: Создание индивидуальной модели для скважины / 226003330 /
Python-библиотека: darts



Модель	R2	MAE
Random Forest	0.9791	0.0104
XGBoost	0.9822	0.0100
LightGBM	0.9515	0.0463
RNN	-59.5832	2.0522
TCN	0.9868	0.0091
N-BEATS	0.9472	0.0429
Transformer	0.9117	0.0459

Оценка выполнена на тестовой
выборке (20 % данных)

Наилучшим образом показала себя модель TCN, способная выявлять локальные временные причинно-следственные связи и обладающая оптимальным балансом сложности и производительности

Дальнейшее развитие проекта

Предложения



Дополнить текущую работу алгоритмом подбора наилучших гиперпараметров модели (Grid Search CV)



Построить общую модель для всех скважин / для группы скважин, с дополнительным учетом статических данных о ID скважины, месторождения, насоса и др.



Рассмотреть возможность создания гибридной модели (физическая + математическая)



Продолжить исследование по направлению поиска аномалий в данных датчиков



Спасибо за внимание!