

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Центр непрерывного образования

Факультета компьютерных наук

ИТОГОВЫЙ ПРОЕКТ

ПРОГНОЗИРОВАНИЕ ГРАДИЕНТА ДАВЛЕНИЯ НА ПРИЕМЕ ЦЕНТРОБЕЖНОГО
НАСОСА ДЛЯ НЕФТЕГАЗОВОЙ ПРОМЫШЛЕННОСТИ

Название темы

Выполнил (а):

Гарбар Феликс Александрович

Ф.И.О.

Руководитель:

Абдуракипов Сергей Сергеевич

Ф.И.О.

Москва 2023

Оглавление

1. Введение
2. Обзор литературы
3. Методология
4. Описание и обработка данных
5. Результаты и выводы
6. Заключение

Список литературы

Приложение

1. Введение

В последние годы тренды на автоматизацию и цифровизацию, искусственный интеллект (ИИ) и машинное обучение приобрели особую популярность. В частности, цифровизация нефтяной отрасли рассматривается как принципиально новая возможность для дальнейшей оптимизации отрасли [1].

Ключевым вопросом при этом становится внедрение вышеописанных технологий таким образом, чтобы учесть существующие риски и проблемы, полученные результаты были достаточно ощутимы с точки зрения изменения прибыльности и, что особенно важно, полученные результаты могли быть устойчиво применимы к выбранным областям.

Исследования, проводившиеся ранее в этой области, заложили основу для дальнейшего развития методов оптимизации добычи нефти. В частности, динамический мониторинг показателей скважин, управление резервуаром [2, 3], оптимизация производства [4], обеспечение нефтепотока [5] и прогнозное техническое обслуживание [6]. Последнее, в частности, фокусируется на использовании машинного обучения для оценки состояния оборудования и последующего планирования сроков по его обслуживанию, избегая простоев.

Одной из наиболее часто используемых систем является электрический центробежный насос (ЭЦН) [7]. Он нужен для добычи сложных смесей, претерпевающих изменение давления, температуры, состава по мере прокачки. Для обеспечения качественной и бесперебойной добычи необходимы постоянные корректировки параметров извлечения. Для достижения этого были вложены значительные ресурсы в так называемое “digital oil field”, фокусирующееся на внедрении машинного обучения в области предиктивного обслуживания ЭЦН [8].

Исходя из вышесказанного, целью данного проекта является обработка большого потока временных рядов центробежного насоса,

разработка модели машинного обучения (виртуальный датчик) для предсказания скорости изменения давления на входе при отсутствии или неисправности собственного датчика насоса.

Проект структурирован следующим образом:

1. Введение
2. Обзор литературы
3. Методология
4. Описание и обработка данных
5. Результаты и выводы
6. Заключение

2. Обзор литературы

Электрические центробежные насосы являются важным и довольно популярным элементом процесса нефтедобычи. За последние 10 лет их доля в добыче выросла от 20% до 80% [7, 9]. Такой рост связан с возможностью ЭЦН достигать высокой скорости добычи в самых разных геологических структурах и скважинах. У подобной популярности есть и обратная сторона, связанная с авариями, особенно на глубоководном шельфе [10]. Аварии на ЭЦН ведут не только к экономическим потерям, но и нарушают нормальное функционирование производственной линии.

Изначально диагностика и принятие решений основывались на экспертной оценке. Однако, даже эксперт, как и любой человек, может допустить ошибку при принятии решений в силу стресса, психологических факторов и уровня профессиональных компетенций.

По мере развития компьютерных технологий была разработана методика условной диагностики, которая отслеживает в режиме реального времени параметры скважины и оказывает поддержку в диагностике, таким образом снижая время принятия решения и уменьшая вероятность поломки. [11] Но при таком подходе вновь ведущая роль остаётся за инженером и его личным опытом, поэтому разработчики должны постоянно валидировать качество и скорость используемых моделей, алгоритмов и методов мониторинга для поддержания высокой точности предсказания работы ЭЦН [12].

В последние годы исследователи нефтяной сферы начали разработку моделей предсказания жизненного цикла насоса на основе data-driven моделей: выявление закономерностей в электрических гармониках [13], мониторинговая система, описанная выше [14], анализ выборки предыдущих ошибок насоса [15].

В 2015 году была разработана модель на основе Метода опорных векторов, которая использовала частотные и электрические данные для предсказания операционных аномалий в ЭЦН, что является довольно устаревшей методикой [16-18]; годом позднее была представлена аналитическая модель отслеживания состояния насоса, отправляющая отчет о состоянии и возможных предстоящих поломках ещё до факта их наступления [19]. В 2019 году модель дерева решений с применением Метода Главных Компонент (РСА в английской литературе) была опубликована. [20] Данная модель определяет наиболее приемлемый диапазон функционирования ЭЦН по измерению температуры, давления, вибраций, тока и частоты работы.

Стоит отметить, что применение РСА в исследованиях нефтяной отрасли очень популярно. Данный метод используется для трансформации, уменьшения размерности, большого потока данных с датчиков и дальнейшее соотнесение полученных результатов с отчетами сенсоров. Суть процесса заключается в группировке данных по параметрам скважин. Большая часть таких исследований проводилась как обучение без учителя на данных в реальном времени или с применением наблюдением методом исключения (surveillance by exception), но ни один из этих методов не был предсказательным. Последний метод применяется с использованием нормализованного диапазона данных для обучения алгоритмов. Затем алгоритм применяется на тестовых данных для определения точек вне предсказательного доверительного интервала [21, 22].

В данной же работе мы сосредоточимся на использовании различных методов машинного обучения (Random forest, Gradient Boosting, Ridge Regression) и сузим задачу до предсказания скорости изменения давления насоса как наиболее важного параметра работы. Исследования с

применением подобных моделей не новы [23, 24], но они сосредоточены на прогнозировании срока службы и ошибок, а также рассматривают регионы отличные от России, где существуют свои геологические особенности добычи.

3. Методология

Машинное обучение (Machine Learning или ML в англоязычной литературе) является подразделом Искусственного интеллекта (ИИ). Оно основывается на сборе данных и самообучающихся системах. ML это метод анализа данных, который автоматизирует построение аналитической модели. Итеративная сторона ML особенно важна, потому что при попадании новых данных в уже существующую модель она самостоятельно/независимо адаптируется к ним. Таким образом, модель может обучаться на предыдущих вычислениях для получения воспроизводимых и надежных результатов.

3.1 Случайный лес (Random forest)

Для рассмотрения данного алгоритма необходим краткий экскурс в важный составляющий его элемент: Дерево решений (Decision tree или DT). Оно представляет собой иерархическую структуру, которую можно использовать для классификации объектов.

Случайный лес как алгоритм строится на составлении набора большого количества вышеупомянутых Деревьев решений и затем по принципу максимизации вероятности (наиболее частого результата Деревьев) принимается решение.

Математически данный алгоритм можно описать следующим образом: Случайный лес включает в себя большое количество деревьев, где каждое дерево случайно образовано случайным выбором части данных и имеющихся переменных (Рис. 1). Благодаря такому механизму удастся достичь высокого результата, но сам алгоритм представляет собой “черный ящик”, что в конечном счёте можно рассматривать как компромисс между объясняющей способностью и интерпретируемостью.

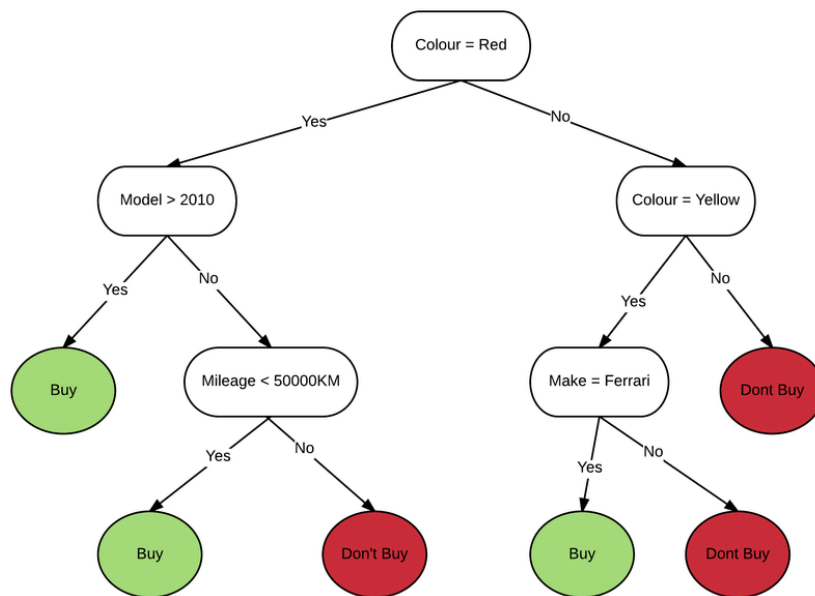


Рис. 1 Пример случайного леса

3.2 Light Gradient Boosting (LGB)

LGBBoost - это ансамблевая модель на основе дерева решений. Аесамблевое обучение сочетает в себе предсказательные способности множества моделей, комбинируя их в единую модель. При этом результаты отдельных моделей сами по себе не сильно выше случайного угадывания (из-за высокого смещения), но тем не менее обладают хотя бы небольшой информацией. Поэтому их совокупный вклад показывает ощутимо лучшие результаты при обучении.

Алгоритм начинается с одного корня (содержит все тренировочные данные), затем происходит итерация по всем объясняющим переменным и данным и, впоследствии, оценивается каждое возможное уменьшение потерь при split. Уравнения 1 и 2 представляют собой целевую функцию (ф-ия потерь и регуляризации, соответственно):

$$\mathcal{L}^t = \sum_{i=1}^n l(y_i, p_i) + O_{value} + \frac{1}{2} \lambda O_{value}^2 (1)$$

$$l(y_i, p_i) = -[y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (2)$$

где y_i - истинное значение i -го наблюдения; p_i - предсказанная вероятность для i -го наблюдения; $l(y_i, p_i)$ - функция потерь задачи классификации; O_{value} - выход на новом дереве; $\frac{1}{2}\lambda O_{value}^2$ - параметр регуляризации.

В одной из статей было показано, что XGBoost функция не может быть оптимизирована в методах Евклидова пространства (конечномерное вещественное пространство, где задано скалярное произведение) [25]. Поэтому для решения задачи минимизации ошибки нужно привести функцию в Евклидово пространство: через разложение в ряд Тейлора до второго порядка (Уравнения 3 и 4).

$$(L)^t \cong \left[\sum_{i=1}^n l(y_i, p_i) + g_i O_{value} + \frac{1}{2} h_i O_{value}^2 \right] + \frac{1}{2} \lambda O_{value}^2 \quad (3)$$

где g_i - градиент, вычисленный как $g_i = \frac{\partial}{\partial p_i} l(y_i, p_i)$ (второе слагаемое в разложении в ряд Тейлора) и h_i - Гессиян, вычисленный как $h_i = \frac{\partial^2}{\partial p_i^2} l(y_i, p_i)$ (третье слагаемое в разложении в ряд Тейлора).

В итоге, после избавления от констант функция для минимизации выглядит следующим образом:

$$(L)^t = \sum_{i=1}^n g_i O_{value} + \frac{1}{2} h_i O_{value}^2 + \frac{1}{2} \lambda O_{value}^2 \quad (4)$$

Для нахождения минимума функции нужно 1) взять производную и 2) приравнять к 0, откуда можно выразить значение для выхода дерева через изменяемые вводные (первые и вторые производные):

$$\frac{d}{dO_{value}} \left(\sum_{i=1}^n g_i O_{value} + \frac{1}{2} h_i O_{value}^2 + \frac{1}{2} \lambda O_{value}^2 \right) = 0 \quad (5)$$

$$O_{value} = - \frac{\sum_{i=1}^n g_i}{\sum_{i=1}^n h_i + \lambda} \quad (6)$$

Комбинируя Уравнение 6 с первой и второй производными функции потерь (Уравнение 2), g_i и h_i , соответственно, функция подобия (similarity) равна:

$$similarity = \frac{\sum residual_i}{\sum previous_probability_i \cdot (1 - previous_probability_i) + \lambda} \quad (7)$$

Функция подобия вычисляется для “ветки дерева”. Различные пороги используются для разбивки дерева на большее число листьев. Функция подобия вычисляется сначала для каждой ветки и потом складывается в , так называемый, прирост (gain):

$$gain = left_{similarity} + right_{similarity} \quad (8)$$

Пороги продолжают устанавливаться пока достигается больший прирост и дерево продолжает увеличиваться. Существует минимальный порог остатков в каждой ветке для дальнейшего роста дерева. Это число определяется параметром cover (знаменатель в Уравнении 6). В процессе бустинга пересчет параметров осуществляется так, что деревья строятся последовательно. Каждое следующее уменьшает ошибку прошлого и обучается у него же, динамически обновляя остатки.

3.3 Ridge regression

Последней моделью является ридж-регрессия или линейная регрессия с L2 регуляризацией. Регрессия добавлена для сравнения результатов с менее интерпретируемыми моделями.

По своей сути данная модель отличается от классической линейной регрессии только параметром регуляризации, который, с одной стороны, немного добавляет смещение, но, с другой, сильно уменьшает дисперсию. Иначе говоря, контролирует переобучение.

$$L_2 = \sum_i (y_i - y(t_i))^2 + \lambda \sum_i a_i^2$$

3.4 Кросс-валидация (Cross-Validation)

Данная процедура применяется для валидации модели через разбиение имеющейся выборки на подвыборки. В нашем случае имеется дело с временными рядами, поэтому случайное разбиение нам не подойдёт: данные четко ранжированы. В таком случае отсортированные по времени данные равномерно добавляются к текущему батчу (Рис. 2).

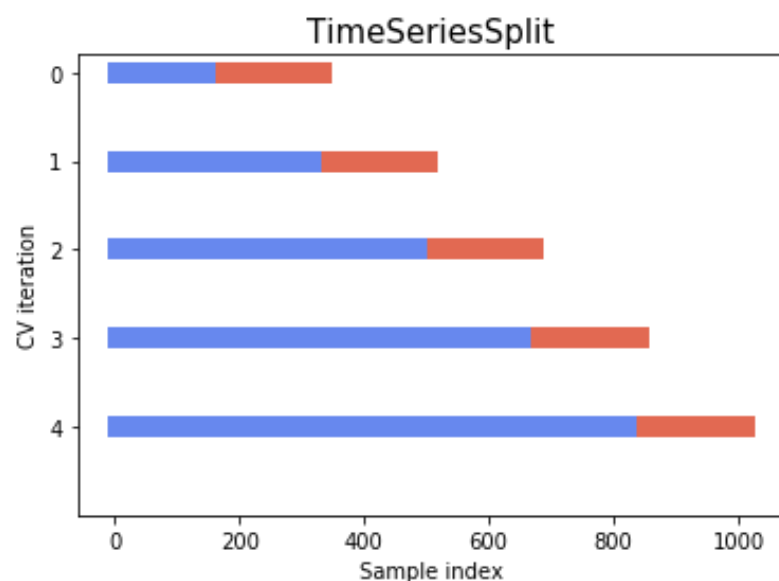


Рис. 2 Кросс-валидация временного ряда

Однако у данного подхода есть недостатки.

- Первый синий фолд очень короткий по времени;
- Большая дисперсия по метрикам в тестовых данных (из-за сильно разного количества данных в обучающей выборке);
- При большом количестве скважин нужно сначала группировать по отдельным источникам, а потом делить по времени – это вне стандартного функционала функции `sklearn.TimeSeriesSplit`.

Но первые два недостатка можно обойти следующим образом: сразу отсекаются 50% наблюдений, а остальные делятся на n частей и добавляются к предыдущим ($50\% < n$ наблюдений).

3.5 Подбор параметров

Подбор оптимальных параметров является одним из важных шагов при построении модели анализа данных.

В случае LGBost и Random Forest параметры можно разделить на общие (тип алгоритма, число попыток), обучающие (минимальное изменение ошибки) и бустинга (скорость обучения, минимальная сумма весов для выделения новой ноды). Последнее применимо только к LGBost. Таблица ниже показывает использованные параметры и их диапазон значений.

Параметр	Определение	Диапазон
LightGBM		
colsample_bytree	доля параметров для train каждого дерева	0.3-0.9
learning_rate	скорость обучения при град. спуске	0.05-.05
max_depth	глубина дерева решения	5-32, int
min_child_weight	кол-во элементов нужных для образования “листа”	1-10, int
subsample	доля данных, отбираемая для построения ветки дерева	0.5-0.99, uniform
eval_metric	среднеквадратичная ошибка	l2
RandomForestRegressor		
n_estimators	общее число деревьев	2-20, int
max_depth	макс. глубина дерева	5-32, int
max_features	макс. число features для наилучшего разделения	0.1-1
min_samples_leaf	мин. число наблюдений в ноде	1-5, int
min_samples_split	мин. число наблюдений для образования ноды	2-10, int
eval_metric	площадь под кривой ROC	auc

4. Описание и обработка данных

Набор данных представляет собой операционные показатели работы скважин (визуализация в Приложении 1) . Данные получены от компании Сургутнефтегаз.

- Период наблюдений Июль 2019 года;
- Количество скважин: 17;
- Частота обновления данных: 5 мин;
- Количество параметров: 9.

Признак	Описание	Интервал значений
T1138P6000096	Наработка двигателя с момента последнего включения, с	0-17100
T1138P6000315	Время простоя двигателя с момента последнего выключения, с	0-86100
DSHORTT1138P4000064	Средняя скорость изменения загрузки двигателя ЧАС, %/час	-49.2 - 79.2
DSHORTT1138P2600012	Средняя скорость изменения тока фазы А двигателя в ЧАС, А/час	-21.12 - 21.6
DSHORTT1205P2300000	Средняя скорость изменения давления в коллекторе ИУ в ЧАС, МПа/час	-0.38 - 3.45
T1138P4000064	Загрузка двигателя, %	0 - 108
T1138P2600012	Ток фазы А двигателя, А	0 - 37.60
T1205P2300000	Давление в коллекторе измерительной установки, МПа	0.69 - 3.99
DSHORTT1138P2300058	Средняя скорость изменения давления на приеме насоса в ЧАС, МПа/час	0.30 - 0.31

Корреляционная матрица (Приложение 2) показала наибольшую связь (по абсолютной величине) у целевой переменной

DSHORTT1138P2300058

1. **T1138P6000096**: Нарботка двигателя с момента последнего включения, с
2. **T1138P6000315**: Время простоя двигателя с момента последнего выключения, с
3. **T1138P4000064**: Загрузка двигателя, %
4. **T1138P2600012**: Ток фазы А двигателя, А

Для учета факта, что данные имеют разные единицы измерения и диапазоны необходимо применить нормировку. В данной работе был применен min-max scaler:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

5. Результаты и выводы

Результаты оценки моделей приведены ниже. Анализ проводился по каждой скважине отдельно, чтобы учесть индивидуальные особенности. Ниже приведена усредненная ошибка по каждой модели. В Приложении 3 находятся результаты для каждой скважины в отдельности.

Модель	MAPE average, %
LightGBM	0.0803
RandomForestRegressor	0.0781
Ridge Regression	0.0855

И оптимальные параметры для одной из скважин (14-ой):

Параметр	Описание	Значение
LightGBoost		
colsample_bytree	доля параметров для trian каждого дерева	0.9
learning_rate	скорость обучения при град. спуске	0.09
max_depth	глубина дерева решения	23
min_child_weight	кол-во элементов нужных для образования “листа”	2
subsample	доля данных, отбираемая для построения ветки дерева	0.81
RandomForestRegressor		
n_estimators	общее число деревьев	4

max_depth	макс. глубина дерева	10
max_features	макс. число features для наилучшего разделения	0.1
min_samples_leaf	мин. число наблюдений в ноде	1
min_samples_split	мин. число наблюдений для образования ноды	5

Первое, что можно отметить, - это очень низкое значение ошибки ($<0.1\%$). Наиболее хороший результат показала модель Случайного леса, однако было замечено в процессе обучения, что в некоторых случаях наилучший результат был у LGBBoost.

В Приложении 4 находятся графики предсказанного и фактического значения скорости изменения давления. По данным графикам можно увидеть, что модель, хотя и хорошо предсказывает направление, недосоздаёт вакуум в насосе. Возможное объяснение этому отсутствие внешних параметров окружающей среды: температура воздуха, разности давлений на поверхности и внутри скважины, относительная влажность и иные параметры. Также стоит отметить не непрерывность получаемой информации (лаг данных), что также приводит к неточности в прогнозировании.

И тем не менее стоит отметить, что эмпирическая ошибка параметров расчетов в инженерных дисциплинах составляет 5%, что значительно выше полученных результатов.

Из модели линейной регрессии получено значение и направление коэффициентов. Для наглядности рассмотрим результат для той же 14-ой скважины. Наибольшее влияние имеют следующие переменные:

- **T1205P2300000** - Давление в установке, МПа (Знак +)

- **T1138P4000064** - Загрузка двигателя, % (Знак -)
- **DSHORTT1205P2300000** - Средняя скорость изменения давления, МПа/час (Знак -)
- **DSHORTT1138P2600012** - Средняя скорость изменения тока фазы А, А/час (Знак -)

С точки зрения работы насоса при уменьшении давления внутри (создание вакуума) большой объем нефти начинает прокачиваться (создается движущая сила) за счет прироста в скорости вращения лопаток насоса, то есть отрицательное значение коэффициента говорит в пользу увеличения скорости прокачки.

Таким образом полученные знаки коэффициентов совпадают с физическим смыслом их переменных.

Среди переменных, показавших наименьшее влияние оказались:

- **T1138P6000096** - Время простоя двигателя с момента последнего выключения, с
- **T1138P6000096**: Нарботка двигателя с момента последнего включения, с

Хотя данные параметры показали высокую корреляцию с целевой переменной (не равно зависимость), их непосредственный вклад в предсказание скорости изменения давления может быть мал ввиду малой информативности после выхода насоса на рабочие параметры добычи, т.е вклад только на этапе запуска-остановки.

6. Заключение

Работа центробежного насоса сопряжена с влиянием большого количества факторов, а основным его выходным параметром является скорость прокачки или создаваемое давление. Способность проактивно и точно прогнозировать его значение помогает увеличить срок службы системы, снизить затраты и минимизировать экологический вред от возможной аварии.

В данной работе был предложен метод предсказания скорости изменения давления на основе методов машинного обучения.

Использовались модели LGBost, Random Forest и Ridge regression.

Наилучший, в среднем, результат показала первая модель.

В дальнейшем улучшением полученной модели может стать применение нелинейной модели (Нейронной сети), выделения новых параметров из имеющихся (библиотека tsfresh) при наличии достаточных вычислительных мощностей и рассмотрение влияния новых внешних факторов.

Список литературы

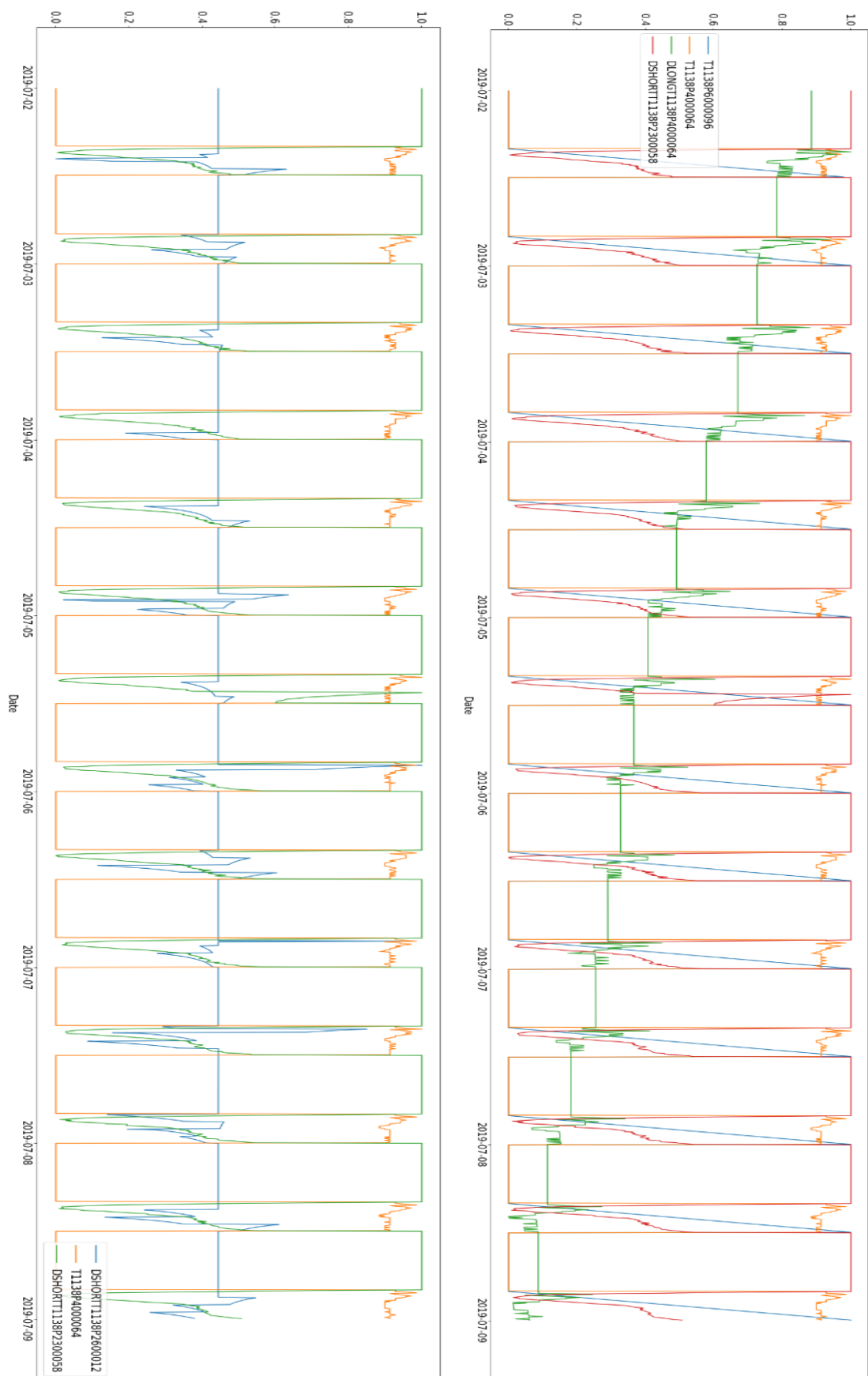
1. Göktug D, Frühbauer H, Edna Michelle Bisso Bi Mba. "Development of a Digital ESP Performance Monitoring System Based on Artificial Intelligence.", 2021.
2. Bai, Y.; Li, J.; Zhou, J.; Li, Q. Sensitivity Analysis of the Dimensionless Parameters in Scaling a Polymer Flooding Reservoir. *Transp Porous Med.* 2008, 73 (1), 21–37.
3. Hou, J.; Zhou, K.; Zhang, X.-S.; Kang, X.-D.; Xie, H. A review of closed-loop reservoir management. *Pet. Sci.* 2015, 12 (1), 114–128.
4. Hidayat, A.; Prakoso, N. F.; Sujai, A.; Medco, P. T. Production and Cost Optimization in a Complex Onshore Operation Using Integrated Production Model. *SPE Symposium: Production Enhancement and Cost Optimisation 2017*, SPE-189223-MS
5. Wylde, J. J.; Fell, D. Scale Inhibitor Solutions For High Temperature ESP Lifted Wells In Northern California: A Case History Of Failure Followed By Success. *SPE International Oilfield Scale Conference 2008*, SPE-113826-MS
6. Yanaliev V, Tagirova K, Salahov T, Shubin S etc 2011 Diagnostics of electric centrifugal pumps during their operation by using neural network technologies *Oil and Gas Business* vol 9 p 68
7. Pavlihina A 2019 An expert approach to mechanized mining *Business magazine Neftegaz* vol 4 pp 92-96
8. Alamu, O. A.; Pandya, D. A.; Warner, O.; Debacker, I. ESP Data Analytics: Use of Deep Autoencoders for Intelligent Surveillance of Electric Submersible Pumps. *Offshore Technology Conference 2020*, OTC-30468-MS.
9. Pragale, R. and Shipp, D.D. (2012) Investigation of Premature ESP Failures and Oil Field Harmonic Analysis. *2012 Petroleum and Chemical*

- Industry Conference, New Orleans, 24-26 September 2012, 1-8.
10.1109/PCICON.2012.6549650
- 10.G. Supriya, S. Luigi, and N. Michael, “Big data analytics workflow to safeguard ESP operations in real-time,” in Proceedings of SPE North America Artificial Lift Conference and Exhibition (Society of Petroleum Engineers, Woodlands, 2016).
 - 11.K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech. Syst. Signal Process.* 20, 1483 (2006).
 - 12.W. Y. Liu, B. P. Tang, J. G. Han, X. N. Lu, N. N. Hu, and Z. Z. He, “The structure healthy condition monitoring and fault diagnosis methods in wind turbines: A review,” *Renewable Sustainable Energy Rev.* 44, 466 (2015).
 - 13.Van Rensburg, N. J. (2019) Autonomous Well Surveillance for ESP Pumps Using Artificial Intelligence. SPE Oil and Gas India Conference and Exhibition, Mumbai, 9-11 April, Article No. SPE-194587-MS.
 - 14.Takacs, G, (2009) Electric Submersible Pumps Manual: Design Operation, and Maintenance, Gulf Professional Publishing
 - 15.Al Maghlouth, A., Cumings, M., Al Awajy, M. and Amer, A. (2013) ESP Surveillance and Optimization Solutions: Ensuring Best Performance and Optimum Value. the SPE Middle East Oil and Gas Show and Conference, Manama, 10-13 March 2013.
 - 16.Guo, D., Raghavendra,C.S., Yao, K.-T., Harding, M., Anvar, A. and Patel, A. (2015) Data Driven Approach to Failure Prediction for Electric Submersible Pump Systems. SPE Western Regional Meeting, Garden Grove, 27-30 April 2015.
 - 17.Xi, W. J. Research on fault diagnosis of electric submersible pumps based on vibration detection. Master’s Thesis, China University of Petroleum (East China), Dongying, China, 2008.

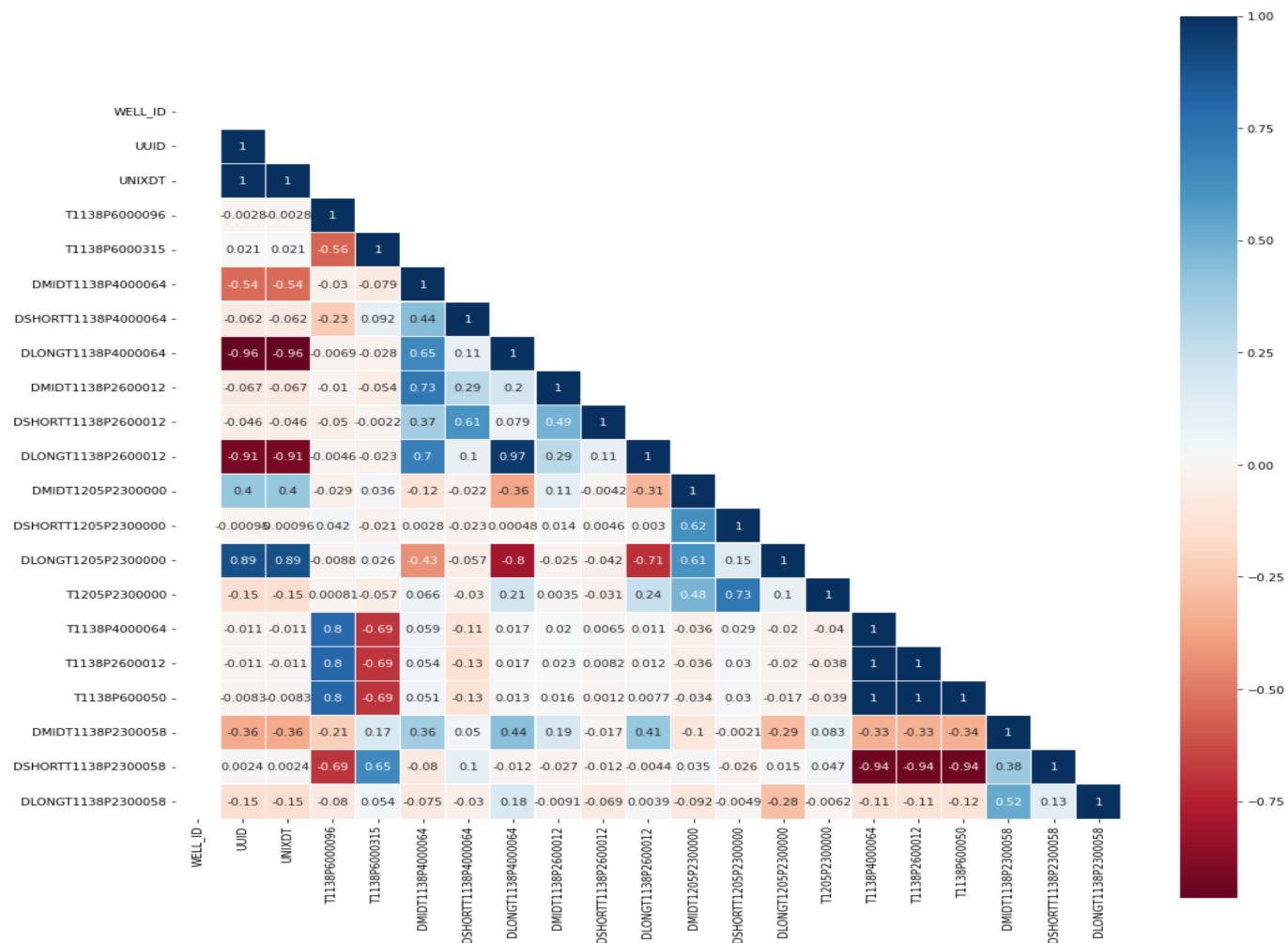
18. Rajakarunakaran, S.; Devaraj, D.; Rao, K. S. Fault detection in centrifugal pumping systems using neural networks. *Inter. J. Model. Ident. Cont.* 2008, 3, 131.
19. Gupta, S., Nikolaou, M., Saputelli, L. and Bravo, C. (2016) ESP Health Monitoring KPI: A Real Time Predictive Analytics Application. SPE Intelligent Energy International Conference and Exhibition, Aberdeen, 6-8 September 2016.
20. Sherif, S., Adenike, O., Obehi, E., Funso, A. and Eyituyo, B. (2019) Predictive Data Analytics for Effective Electric Submersible Pump Management. SPE Nigeria Annual International Conference and Exhibition, Lagos, 5-7 August 2019.
21. Peng, L.; Han, G.; Pagou, A. L.; Zhu, L.; Ma, H.; Wu, J.; Chai, X. A Predictive Model to Detect the Impending Electric Submersible Pump Trips and Failures. SPE Annual Technical Conference and Exhibition 2021, SPE-206150-MS.
22. Khabibullin, R. A.; Shabonas, A. R.; Gurbatov, N. S.; Timonov, A. V. Prediction of ESPs Failure Using ML at Western Siberia Oilfields with Large Number of Wells. SPE Russian Petroleum Technology Conference 2020, SPE-201881-MS.
23. Sneed, J. (2017) Predicting ESP Lifespan with Machine Learning, Austin, Texas, USA: Unconventional Resources Technology Conference. SPE/AAPG/SEG Unconventional Resources Technology Conference, Austin, 24-26 July.
24. Van Rensburg, N.J. (2018) Usage of Artificial Intelligence to Reduce Operational Disruptions of ESPs by Implementing Predictive Maintenance. Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, 12-15 November 2018.

25. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016, 785–794.

Приложение 1



Приложение 2



Приложение 3

MAPE

Номер скважины	MAPE LGBost, %	MAPE RF, %	MAPE L2, %
1	0.026	0.036	0.024
2	0.102	0.083	0.066
3	0.015	0.024	0.051
4	0.215	0.293	0.128
5	0.058	0.065	0.190
6	0.05	0.022	0.216
7	0.064	0.053	0.023
8	0.134	0.159	0.092
9	0.098	0.125	0.137
10	0.1	0.139	0.206
11	0.279	0.285	0.057
12	0.026	0.024	0.058
13	0.044	0.113	0.032
14	0.033	0.027	0.029
15	0.04	0.030	0.033
16	0.024	0.024	0.056
17	0.045	0.079	0.056

Приложение 4

