

УДК 622.276  
МРНТИ 52.47.25

DOI: <https://doi.org/10.54859/kjogi108750>

Получена: 22.05.2024.  
Одобрена: 29.11.2024.  
Опубликована: 31.12.2024.

## Научный обзор

# Возможность применения современного метода прогнозирования отказов скважин на месторождениях АО НК «КазМунайГаз»

Л.Г. Утемисова<sup>1</sup>, Т.Ж. Мерембаев<sup>2</sup>, Б.Е. Бекбау<sup>3</sup>

<sup>1</sup>КМГ Инжиниринг, г. Астана, Казахстан

<sup>2</sup>Институт информационных и вычислительных технологий КН МОН РК, г. Алматы, Казахстан

<sup>3</sup>КазНИТУ им. К.И. Сатпаева, г. Алматы, Казахстан

## АННОТАЦИЯ

В процессе разработки зрелых месторождений возникает ряд геологических и технологических осложнений. Для повышения бесперебойной работы скважинного насосного оборудования используются различные методы и приёмы. В данной статье представлен анализ возможности применения предсказания отказов для повышения надёжности подземного оборудования скважин на месторождениях АО НК «КазМунайГаз». Исследование фокусируется на разработке и валидации модели машинного обучения, способной с высокой точностью прогнозировать потенциальные отказы в работе скважинного оборудования. Существующие методики, подходы машинного обучения и их применение в условиях реальной эксплуатации детально проанализированы с выделением ключевых факторов успеха и ограничений. Результаты исследования демонстрируют огромный потенциал применения модели прогнозирования отказов скважин при выборе оптимального подхода машинного обучения с целью уменьшения внеплановых простоев и оптимизации процессов технического обслуживания скважин. Авторами проведена оценка возможности применения прогнозирования отказов глубинного-насосного оборудования скважин, эксплуатируемых ШГН. Применение предсказания отказов глубинно-насосного оборудования скважин позволит обеспечить бесперебойную работу скважин за счет снижения отказов скважин и сокращения времени простоя на ремонт.

**Ключевые слова:** *глубинно-насосное оборудование, наработка на отказ, подземный ремонт скважин, отказы скважин, прогнозирование отказов.*

## Как цитировать:

Утемисова Л.Г., Мерембаев Т.Ж., Бекбау Б.Е. Возможность применения современного метода прогнозирования отказов скважин на месторождениях АО НК «КазМунайГаз» // Вестник нефтегазовой отрасли Казахстана. 2024. Том 6, №4. С. 68–77. DOI: <https://doi.org/10.54859/kjogi108750>.

UDC 622.276  
CSCSTI 52.47.25

DOI: <https://doi.org/10.54859/kjogi108750>

Received: 22.05.2024.

Accepted: 29.11.2024.

Published: 31.12.2024.

## Review article

# Exploring modern methods for predicting well failures in the fields of NC «KazMunayGas» JSC

Laura G. Utemissova<sup>1</sup>, Timur Zh. Merembayev<sup>2</sup>, Bakbergen Ye. Bekbau<sup>3</sup>

<sup>1</sup>KMG Engineering, Astana, Kazakhstan

<sup>2</sup>Institute of Information and Computational Technologies CS MES RoK, Almaty, Kazakhstan

<sup>3</sup>Satbayev University, Almaty, Kazakhstan

## ABSTRACT

In the development of brownfields, various geological and technological complications can arise. To enhance the smooth operation of downhole pumping equipment, companies implement a range of methods and techniques.

This article analyzes the potential of using machine learning to improve the reliability of underground well equipment in the fields of NC KazMunayGas JSC. The research focuses on the development and validation of predictive models that accurately forecast potential downhole equipment failures. It thoroughly analyzes existing machine learning methods, approaches and their real-life application, highlighting key success factors and limitations. The results of the study demonstrate the significant potential for using a well failure prediction model when selecting the optimal machine learning approach to reduce unscheduled downtime and optimize well maintenance processes. The authors assessed the potential for using failure prediction techniques for downhole pumping equipment in wells that utilizes sucker rod pumps. Implementing failure prediction techniques for downhole pumping equipment can help ensure uninterrupted well operation by minimizing well failures and reducing downtime for repairs.

**Keywords:** *downhole pumping equipment; time between failures; underground well workover; well failures; failure prediction.*

## To cite this article:

Utemissova LG, Merembayev TZ, Bekbau BY. Exploring modern methods for predicting well failures in the fields of NC «KazMunayGas» JSC. *Kazakhstan journal for oil & gas industry*. 2024;6(4):68–77.

DOI: <https://doi.org/10.54859/kjogi108750>.

ӨОЖ 622.276  
FTAХР 52.47.25

DOI: <https://doi.org/10.54859/kjogi108750>

Қабылданды: 22.05.2024.  
Мақұлданды: 29.11.2024.  
Жарияланды: 31.12.2024.

## Ғылыми шолу

# «ҚазМұнайГаз» ҰК АҚ кен орындарында ұңғымалардың істен шығуын болжаудың заманауи әдісін қолдану мүмкіндігі

Л.Г. Өтемісова<sup>1</sup>, Т.Ж. Мерембаев<sup>2</sup>, Б.Е. Бекбау<sup>3</sup>

<sup>1</sup>ҚМГ Инжиниринг, Астана қаласы, Қазақстан

<sup>2</sup>Ақпараттық және есептеуіш технологиялар институты ҚР БҒМ ҒК, Алматы қаласы, Қазақстан

<sup>3</sup>Қ.И.Сәтбаев атындағы ҚазҰТЗУ, Алматы қаласы, Қазақстан

## АННОТАЦИЯ

Жетілген кен орындарын игеру барысында бірқатар геологиялық және технологиялық қиындықтар туындайды. Ұңғымаларды айдау жабдықтарының бірқалыпты жұмысын жақсарту үшін кәсіпорындар әртүрлі әдістер мен тәсілдерді қолданады. Бұл мақалада «ҚазМұнайГаз» ҰК АҚ кен орындарындағы жерасты ұңғымалары жабдығының сенімділігін арттыру үшін істен шығуды болжауды қолдану мүмкіндігінің талдауы берілген. Зерттеу ұңғыма жабдығының ықтимал ақауларын жоғары дәлдікпен болжауға қабілетті машиналық оқыту үлгілерін өзірлеуге және тексеруге бағытталған. Қолданыстағы техникалар, машиналық оқыту тәсілдері және оларды нақты өмір жағдайында қолдану табыстың негізгі факторлары мен шектеулерін көрсете отырып, толық талданады. Зерттеу нәтижелері жоспардан тыс тұрып қалу уақытын қысқарту және ұңғымаға техникалық қызмет көрсету процестерін оңтайландыру үшін машинаны оқытудың оңтайлы тәсілін таңдау үшін ұңғыманың істен шығуын болжау үлгісін пайдаланудың орасан зор әлеуетін көрсетеді. Авторлар сорғыш штангаларымен жұмыс істейтін ұңғымаларда ұңғылық сорғы жабдықтарының істен шығуын болжау мүмкіндігін бағалады. Ұңғымалардың ұңғылық сорғы жабдықтарының істен шығуын болжауды пайдалану ұңғымалардың істен шығуын азайту және жөндеу жұмыстарының тоқтап қалу уақытын азайту арқылы ұңғымалардың үздіксіз жұмысын қамтамасыз етеді.

**Негізгі сөздер:** терең сорғы жабдықтары, істен шығуларды атқару, жер асты ұңғымаларын жөндеу, ұңғымалардың істен шығуы, істен шығуды болжау.

## Дәйексөз келтіру үшін:

Өтемісова Л. Г., Мерембаев Т.Ж., Бекбау Б.Е. «ҚазМұнайГаз» ҰК АҚ кен орындарында ұңғымалардың істен шығуын болжаудың заманауи әдісін қолдану мүмкіндігі // Қазақстанның мұнай-газ саласының хабаршысы. 2024. 6 том, №4. 68–77 б. DOI: <https://doi.org/10.54859/kjogi108750>.

**Введение**

В процессе освоения нефтегазового месторождения с момента начала добычи скважинной продукции пластовая энергия уменьшается, в связи с этим после прекращения фонтанирования добывающие компании переходят на механизированный способ эксплуатации. Одним из важных этапов добычи нефти является подъём нефти на поверхность с помощью систем механизированной добычи. В настоящее время в АО НК «КазМунайГаз» (далее – КМГ) основной механизированный фонд скважин эксплуатируется штанговыми глубинными насосами (далее – ШГН). Ключевыми особенностями ШГН, которые объясняют их широкое применение в производстве, являются высокое значение коэффициента полезного действия, возможность использования в скважинах с осложняющими факторами, а также выполнения ремонтных работ в полевых условиях. Основными недостатками ШГН являются ограничения по глубине скважин, невозможность применения в искривленных скважинах, невысокое значение коэффициента подачи. Одним из важных параметров, характеризующих долговременную работу глубинно-насосного оборудования (далее – ГНО), является наработка на отказ оборудования, которая зависит от количества отказов ГНО [1].

Частота преждевременных отказов ГНО месторождений, а также высокая доля часто ремонтируемого фонда скважин (далее – ЧРФ) показывают необходимость применения современных методов диагностики и прогнозирования отказов оборудования (рис. 1) [2]. Авторами были изучены международный опыт и возможности применения прогнозирования отказов внутрискважинного оборудования

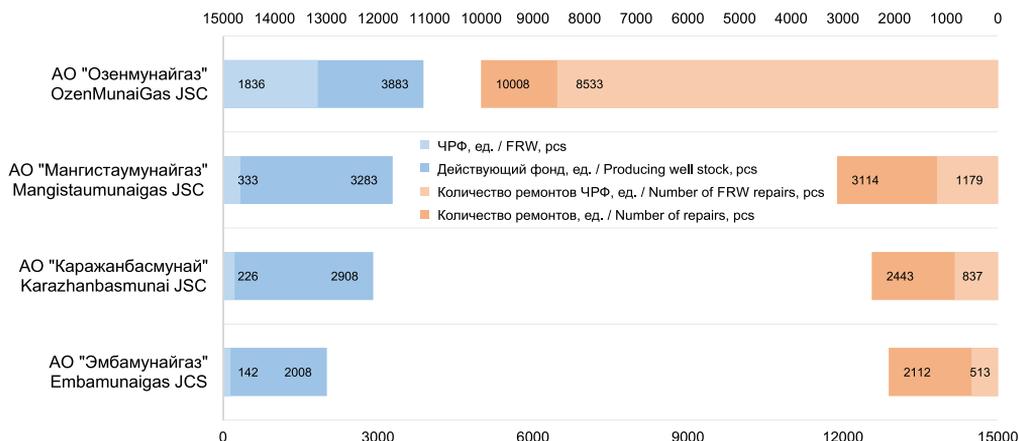
Целью исследования является построение и дальнейшее применение на промысловых данных модели прогнозирования отказов ГНО скважин, эксплуатируемых ШГН. Данная работа выполняется с целью уменьшения внеплановой остановки оборудования, перевода внепланового отказа в плановый за счёт раннего заблаговременного прогноза отказа оборудования и планирования необходимых мероприятий на скважинах. Для достижения цели решаются следующие задачи:

- анализ и подготовка качественных данных;
- подготовка и определение признаков, которые влияют на прогноз;
- построение и выбор модели предсказания отказов;
- обучение модели для задачи прогнозирования;
- валидация модели на реальных данных.

**Область применения машинного обучения**

В процессе эксплуатации установки ШГН могут отказывать по ряду причин, среди которых основными являются: геологические причины, такие как снижение притока; технологические причины; некорректный подбор ГНО. Анализ причин отказов при подземных ремонтах скважин (далее – ПРС) отражает ряд скважин с высокой частотой ремонтов в год (3 и более ремонтов в год).

Зачастую причинами отказов на скважинах, оборудованных ШГН, является обрыв штанг, протирание штанг, негерметичность насосно-компрессорных труб (далее – НКТ) и др. На рис. 2 представлена статистика причин отказов ГНО [2]. Как показывает международный опыт, основываясь на исторических данных



**Рисунок 1. Действующий фонд и количество ремонтов, включая ЧРФ, в разрезе ДЗО КМГ**  
**Figure 1. Producing well stock and the number of repairs, including the frequently repaired well stock, in the context of the subsidiaries and affiliates of the KMG**

о работе скважин и отказах с учётом их причин, можно построить модель прогнозирования отказов, с помощью которой на начальном уровне становится возможным предсказать следующую дату отказа, а при более усложнённой модели – причины отказа.

Статистика отказов скважин КМГ показывает необходимость применения предиктивной аналитики скважин (рис. 2).

Подход к построению модели прогнозирования имеет общую структуру, где основным шагом является сбор и подготовка качественных данных. Данные для построения модели можно разделить на статистические и динамические: данные, которые в процессе жизни скважины не меняются, и те, которые меняются с течением времени. Так, авторами работы [3] подробно описана группировка параметров работы скважин, оборудованных электроцентробежными насосами (далее – ЭЦН). В работе авторами уделяется особое внимание подготовке данных для построения модели прогнозирования отказов, т.к. успешность модели зависит от качества используемых данных.

Высокое качество входных данных имеет решающее значение для надёжных результатов анализа, поэтому важно, чтобы подготовка данных осуществлялась с применением надёжных методов проверки доступности данных и их работоспособности. Это позволяет гарантировать, что модели обучаются на основе соответствующих характеристик данных.

Данные, необходимые для построения модели прогнозирования отказов ГНО, включают статические данные, такие как профиль скважины, установленное оборудование, история отказов и технического обслуживания, а также данные, полученные в режиме реального времени по скважинам (режим работы скважины).

Большую часть данных, используемых для машинного обучения, можно разделить на 4 основных типа: числовые, или количественные, данные, которые могут быть непрерывными и дискретными; категориальные данные; данные временных рядов и текст.

Предварительная обработка данных – это метод интеллектуального анализа данных, который включает преобразование необработанных данных в корректный формат, поскольку реальные данные часто неполны, противоречивы и/или не содержат определённых моделей поведения или тенденций, а также могут иметь множество ошибок. Предварительная обработка данных – проверенный метод решения проблем некорректности данных, поскольку она напрямую влияет на способность модели прогнозирования к обучению.

Для прогнозирования отключений и сбоев системы ГНО создаются различные алгоритмы машинного обучения, тесно связанные с моделями больших данных, которые описывают

рабочее состояние ГНО на основе собранных параметров.

### Методика машинного обучения

Машинное обучение – это методология, в которой компьютерные системы обучаются на основе исторических данных, чтобы делать прогнозы и принимать решения [4, 5]. Для построения модели предсказания отказов необходимо выполнить следующие шаги:

1. Сбор данных: сначала собираются данные о работе подземного оборудования.

2. Подготовка данных: данные анализируются, очищаются от выбросов и пропущенных значений, а также могут быть созданы новые признаки.

3. Обучение модели: на основе обработанных данных создаётся модель машинного обучения, которая может предсказывать вероятность отказа насоса на основе текущих параметров.

4. Мониторинг и обновление: с целью сохранения актуальности и точности прогнозов необходимо регулярно мониторить и обновлять модель на основе новых данных.

Для обучения модели предсказания отказов могут использоваться различные методы, включая:

а) методы временных рядов, которые могут использоваться для анализа и прогнозирования изменений параметров скважин со временем. Это помогает выявлять паттерны, которые предшествуют отказам, и принимать меры предосторожности заблаговременно;

б) методы классификации, которые могут определять состояние оборудования как «нормальное» или «неисправное» на основе данных о его работе;

в) методы кластеризации, которые могут группировать скважины с похожими характеристиками и историей, что позволяет выявлять общие аномалии и прогнозировать отказы на групповом уровне.

Одним из главных приоритетов использования машинного обучения в предсказании отказов является возможность проводить профилактическое обслуживание. Модели машинного обучения определяют, когда оборудование приближается к отказу, и позволяют инженерам и техническим специалистам предпринимать меры для предотвращения отказов и, как следствие, остановки скважин.

Анализ международного опыта применения прогнозирования отказов показывает высокие результаты модели предсказания отказов скважин, оборудованных ЭЦН [5, 6]. В таких скважинах машинное обучение позволяет создавать системы мониторинга, которые следят за работой насосов в режиме реального времени. Датчики, установленные на насосах, собирают данные о давлении, расходе, вибрации и других

параметрах, которые затем анализируются с использованием машинного обучения для выявления аномалий и предсказания возможных отказов.

На сегодняшний день авторами статьи проводятся работы по построению и оценке моделей предиктивной аналитики отказов на скважинах, оборудованных штанговыми ГНО, в связи с тем, что основной механизированный фонд скважин КМГ эксплуатируется ШГН.

На рис. 3 показана схема работы штанговой скважиной насосной установки [7], расположенной на месторождениях КМГ. Мониторинг и оптимизация работы ШГН являются ключевыми факторами для повышения производительности скважин и снижения затрат. Применение моделей машинного обучения в данной области обосновано и имеет потенциал для достижения высокого эффекта.

В основе построения модели прогнозирования отказов лежит выбор и подготовка исторических данных работы скважины. Авторами для создания модели использованы данные по кривизне ствола скважин (инклинометрия скважин – темп набора кривизны, зенитный угол ствола скважины), режим работы скважин (дебит жидкости, обводненность), режим откачки, дата и причина отказа ПРС.

Для увеличения класса точности предсказания авторами проводится моделирование различными методами для дальнейшего выбора успешной модели и применения её на месторождениях с целью уменьшения количества отказов скважин, оборудованных ШГН. Как результат, модели предсказания отказов ГНО представляют следующие преимущества [8]:

1) проактивное (заблаговременное) обслуживание): проведение обслуживания насосов до того, как произойдет отказ, сокращение количества ремонтов и потерь добычи жидкости;

2) оптимизация затрат на ремонт и замену оборудования;

3) увеличение надежности – регулярный мониторинг и обновление модели;

4) снижение рисков, связанных с нештатными ситуациями, такими как протечки или аварии.

Модели машинного обучения имеют общепризнанный метод поэтапного формирования модели обучения:

- обнаружение аномалий;
- диагностика отказов;
- прогнозирование;
- уменьшение последствий (смягчение).

В процессе подготовки данных были собраны исторические данные фонда скважин, эксплуатируемых ШГН. По итогам сбора были проведены предварительная подготовка и верификация данных путём очистки и интерполяции имеющихся данных скважины. За основу анализа были приняты следующие данные скважины: режим эксплуатации, кривизна ствола

скважины, история ПРС (дата и причина отказа), спущенное ГНО, режим откачки (число качаний, длина хода станка-качалки). Исторические данные работы скважины были сгруппированы в статистические и динамические (постоянные и переменные в процессе эксплуатации скважины соответственно) [9].

В процессе формирования модели предсказания отказов ГНО аномалией является отказ подземного оборудования: дата отказа, как первая ступень предсказания, и более усложнённая модель – предсказание причины отказа.

Диагностика отказов подземного оборудования, дальнейшее формирование модели прогнозирования и её использование позволит недропользователям проводить смягчение последствий отказов оборудования путём реализации заблаговременных мероприятий на скважинах с целью уменьшения простоев или же оснащения промысла необходимым подземным оборудованием с целью минимизации простоев скважин из-за ПРС.

Успешность модели прогнозирования зависит от выбора корректного метода построения модели предиктивной аналитики.

Основные методы модели машинного обучения следующие [10]:

1. Метод опорных векторов (Support vector machine). Используется для задач классификации и регрессии.

2. Случайный лес (Random Forest). Деревья решений строятся на основе вопросов о признаках и позволяют делать прогнозы. Случайные леса – это ансамбль деревьев, что улучшает обобщающую способность модели [13].

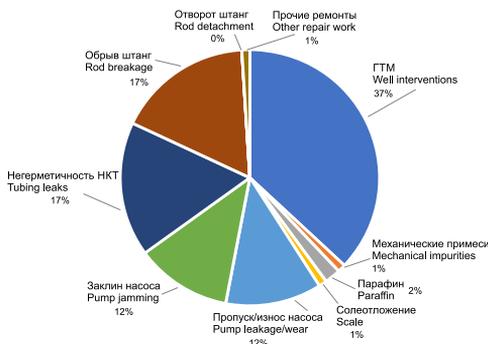
3. Градиентный бустинг (XGBoost). Ансамблевые методы, которые комбинируют несколько слабых моделей для создания более сильной и устойчивой модели [14].

4. Нейронные сети (Artificial Neural Network). Моделирование, состоящее из искусственных нейронов, применяется в задачах на распознавание изображений и обработке естественного языка.

5. Долгосрочная память (Long short-term memory). Моделирование последовательных данных учитывает долгосрочные зависимости в данных [15].

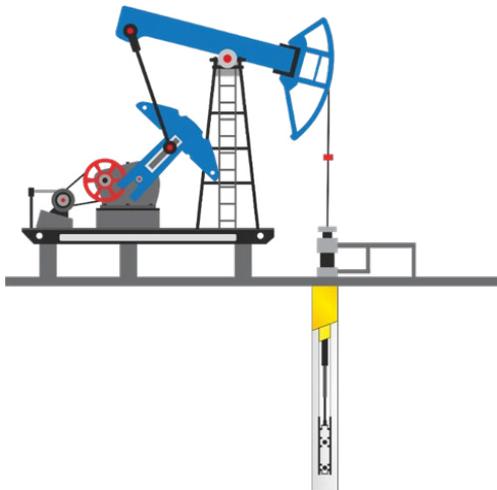
6. Вариационные автоэнкодеры (Variational Autoencoder). Эффективный способ генерации новых данных, изучения структуры скрытого пространства. Инструмент для обучения вероятностных моделей и генерации новых данных с сохранением структуры входных данных.

Корректность модели предсказания можно определить с помощью обнаружения аномалий. Целью обнаружения аномалий является определение верной работы модели в активном режиме. Есть несколько способов обнаружения



**Рисунок 2. Основные причины отказов ГНО**  
**Figure 2. Main causes of downhole pumping equipment failures**

ГТМ – геолого-технические мероприятия



**Рисунок 3. Схема работы штанговой скважинной насосной установки**  
**Figure 3. Schematic diagram of sucker rod**

аномалий с использованием моделей машинного обучения: классификация (доступны помеченные данные разных классов), одноклассовая классификация (доступны помеченные данные только одного класса) и кластеризация (доступны непомеченные данные).

Обнаружение аномалий является одним из важных шагов в выявлении необычных и аномальных паттернов в данных, которые могут указывать на наличие проблем. На рис. 4 указаны несколько методов, которые часто используются для обнаружения аномалий с использованием глубокого обучения.

В работе по предсказанию отказов подземного оборудования скважин авторами был использован один из методов градиентного бустинга LightGBM [4, 9, 12], который помог модели определить нормальные

и аномальные данные. Поскольку отказы скважины сопровождаются на всем протяжении эксплуатации, для прогнозирования отказов были использованы исторические данные выбранного пилотного проекта, и задача сводилась к прогнозированию последовательности отказов за счёт обучения модели на нормальных данных, которые модель использовала в процессе обучения для понимания обычных закономерностей и последовательности отказов скважин. На рис. 5 показано количество аварий в течение одного года. Были рассмотрены 6 типов событий (классов): 0 – нет поломки, 1 – ГТМ, 2 – обрыв штанги, 3 – заклинивание насоса, 4 – протирание штанги, 5 – негерметичность НКТ, 6 – не работает насос.

Визуализация на рис. 5 и анализ данных производился в специализированном пакете Seaborn [11]. Данная библиотека создана для построения статистических графиков в Python и представляет собой высокоуровневый оболочку для matplotlib библиотеки и тесно интегрируется со структурами данных pandas.

Как видно из рис. 5, общее количество аварий не превышает 5% общего количества событий на скважинах, а также наблюдается большое количество аварий типа ГТМ. Данную задачу авторы рассматривали как задачу многоклассовой классификации.

Модель предсказания отказов была обучена на промышленных данных выбранного месторождения. Впоследствии модель была проверена на выявление отклонений и аномалий в данных, которые не соответствуют обычным закономерностям. Были использованы данные пилотного месторождения в группе компаний КМГ. Данные собраны по 2131 скважине за один год. Обучение модели было выполнено на случайной выборке данных. Обучающая выборка представляет собой массив данных из 16477 записей и 24 разных показателей. Валидирующая выборка состоит из 4191 записей и 24 показателя работы скважины.

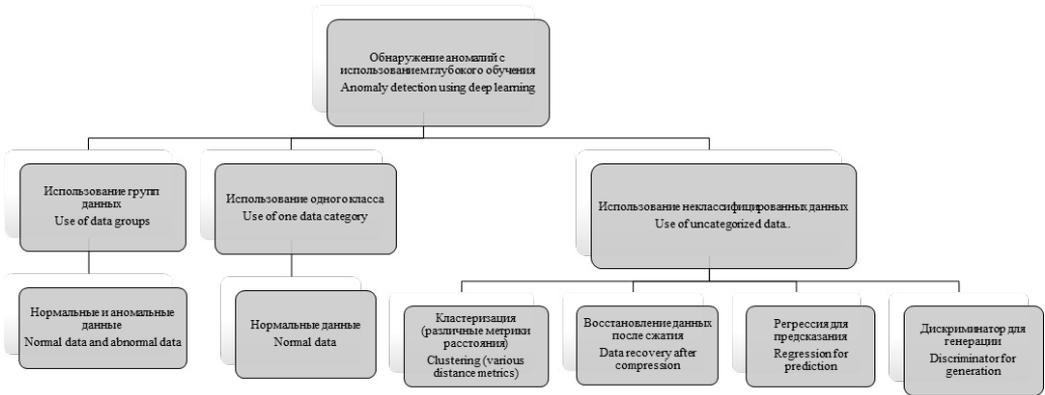
Показателями работы скважины являются:

1) характеристики режима работы скважины: забойное давление, дебит нефти, дебит жидкости, обводнённость, газовый фактор, вязкость нефти, воды и жидкости в пластовых условиях, объёмный коэффициент нефти;

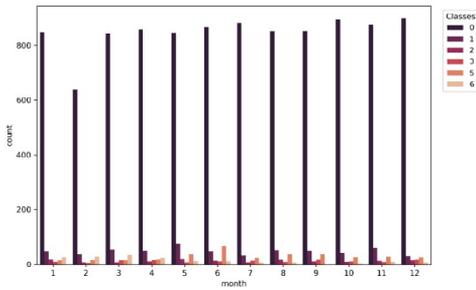
2) режим откачки: тип насоса (диаметр), глубина спуска насоса, число качаний, длина хода;

3) инклинометрия ствола скважины (температура набора пространственной кривизны, зенитный угол).

Теоретически и практически создание модели машинного обучения включает в себя несколько шагов, таких как выбор модели, обучение и оценка модели (рис. 6). Набор исторических данных работы скважин месторождения после прохождения предобработки данных был разделён на две части: первая часть –



**Рисунок 4. Эффективные инструменты для обнаружения аномалий**  
**Figure 4. Effective tools for anomaly detection**



**Рисунок 5. Распределение аварий по типам событий скважин пилотного месторождения**  
**Figure 5. Breakdown of accidents by types of events at wells of the pilot field**

для обучения модели, вторая часть – для проверки / тестирования производительности модели. Авторами для создания модели предсказания отказов подземного оборудования было использовано отношение 80%/20% – для обучения и тестирования соответственно. Для оценки точности модели мы используем метрику ассигасу [4]. Ассигасу (точность) – это доля правильно классифицированных примеров (1):

$$Accuracy = \frac{\sum_{i=1}^C TP_i}{N} \tag{1}$$

где  $C$  — количество классов;  $TP_i$  — количество истинно положительных результатов для класса  $i$ ;  $N$  — общее количество случаев.

В то время как точность вычисляет, сколько положительно идентифицированных примеров на самом деле являются положительными, а полнота – это отношение того, сколько фактических положительных примеров были идентифицированы правильно.

Модель LightGBM на валидирующем датасете показала результат по метрике ассигасу, равный 0,63 (максимальное значение по данной метрике является 1). Авторы считают, что полученные результаты являются недостаточно точными, но это является первым шагом по решению нетривиальной задачи. Для принятия объективного решения о превентивной остановке и ремонте скважины, по мнению авторов, модель должна иметь метрику выше 0,75. Предложенная метрика является усреднением точности модели по всем классам, поскольку авторы имеют датасет с высокой степенью разбалансированности данных в классах. В дальнейших исследованиях планируется изучить направления по решению проблемы разбалансированности данных в классах.

**Заключение**

Машинное обучение для предсказания отказов ГНО представляет собой важный шаг в бесперебойной работе подземного оборудования и, как результат, скважины. Предсказание отказов в ШГН является необходимостью для обеспечения надёжности, эффективности и безопасности процессов в нефтяной и газовой промышленности. Это позволяет сократить риски, оптимизировать производительность, снизить операционные расходы и увеличить срок службы подземного оборудования. Применение методов прогнозирования отказов в ШГН приводит к улучшению всей добычной системы и снижению эксплуатационных рисков за счёт снижения количества проведённых ПРС и уменьшения сроков простоя скважин на ПРС.

В данной статье авторы провели оценку возможности применения подходов машинного обучения модели для прогнозирования отказов ГНО и поделились дальнейшими планами развития исследования.

В следующих исследованиях авторы продолжат развивать данное направление и решать нетривиальную практическую задачу

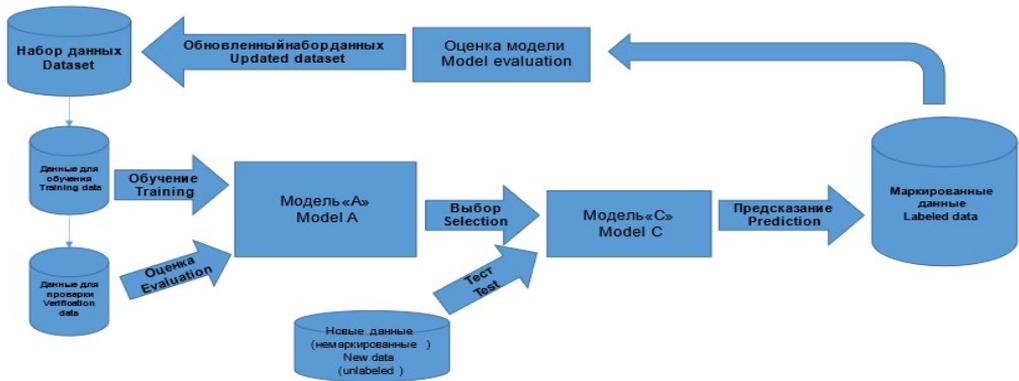


Рисунок 6. Карта создания модели машинного обучения  
Figure 6. Machine learning model creation map

предсказания отказов ГНО. Авторы планируют применять новые подходы по аугментации

данных, которые должны улучшить качество модели машинного обучения.

## ДОПОЛНИТЕЛЬНО

**Источник финансирования.** Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: Утемисова Л.Г. – концептуализация исследования, валидация и подготовка данных; Мерембаев Т.Ж. – выбор методологии исследования, построение и тестирование модели; Бекбау Б.Е. – руководство и направление исследования.

## ADDITIONAL INFORMATION

**Funding source.** This study was not supported by any external sources of funding.

**Competing interests.** The authors declare that they have no competing interests.

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The greatest contribution is distributed as follows: Laura G. Utemissova – research conceptualization, data validation and preparation; Timur Zh. Merembayev – selection of research methodology, model building and testing; Bakbergen Ye. Bekbaу – research administration and supervision.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Михайлович Н.Н. Повышение эксплуатационных характеристик нефтепромышленного оборудования в территориально-производственном предприятии : дис. канд. техн. наук. Москва, 2009. Режим доступа: <https://www.disscat.com/content/povysheniye-effektivnosti-ekspluatatsii-nasosnykh-skvazhin-optimizatsiei-raboty-shtangovykh-k>. Дата обращения: 12.03.2014.
2. 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. Russian Petroleum Technology Conference; October 26–29, 2020; Virtual. Available from: <https://onepetro.org/SPERPTC/proceedings-abstract/20RPTC/1-20RPTC/D013S027R001/450252>.
3. Merembayev T., Kurmangaliyev D., Bekbaуov B., Amanbek Y. A Comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan // Energies. 2021. Vol. 14, N 7. doi: 10.3390/en14071896.
4. Merembayev T., Amanbek Y. Time-series event prediction for the uranium production wells using machine learning algorithms. 56th U.S. Rock Mechanics/Geomechanics; June 26, 2022; Santa Fe, USA. Available from: <https://research.nu.edu.kz/en/publications/time-series-event-prediction-for-the-uranium-production-wells-usi-2>.
5. American Petroleum Institute. API TR 11L Design Calculations for Sucker Rod Pumping Systems (Conventional Units), 5th edition. Washington : API TR 11L, 2008.
6. Abdelaziz M., Lastra R., Xiao J.J. ESP data analytics: Predicting failures for improved production performance. Abu Dhabi International Petroleum Exhibition & Conference; November 13–16 2017; Abu Dhabi, UAE. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=2938839>.
7. Ивановский В.И. Сквaziнные насосные установки для добычи нефти. Москва, 2002.

8. Seradilla O., Zugastu E., Rodriguez J., Zurutuza U. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect // Appl Intell. 2020. Vol. 52. P. 10934–10964. doi: [10.1007/s10489-021-03004-y](https://doi.org/10.1007/s10489-021-03004-y).
9. Takacs G. Sucker-Rod Pumping Handbook. Production engineering fundamentals and long-stroke rod pumping. Hungary, 2015.
10. Rudakov V., Merembayev T., Amirgaliyev Ye., Omarova P. Time Series Analysis of Biogas Monitoring with Deep Learning Approaches. 5th International Conference on Problems of Cybernetics and Informatics; 28–30 August 2023; Baku, Azerbaijan. Available from: <https://ieeexplore.ieee.org/document/10325955>.
11. Waskom M.L. Seaborn: statistical data visualization // Journal of Open Source Software. 2021. Vol. 6, N 60. doi: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021).
12. Ke G., Meng Q., Finley T., et al. Lightgbm: A highly efficient gradient boosting decision tree // Advances in neural information processing systems. 2017. Vol. 30.
13. Breiman L. Random forests // Machine learning. 2001. Vol. 45. P. 5–32.
14. Chen T., He T., Benesty M., et al. Xgboost: extreme gradient boosting // R package version 0.4-2. 2015. Vol. 1, N 4. doi: [10.32614/cran.package.xgboost](https://doi.org/10.32614/cran.package.xgboost).
15. Hochreiter S., Schmidhuber J. Long Short-term Memory // Neural Computation MIT-Press. 1997. Vol. 9, N 8. P. 1735–1780. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).

## REFERENCES

1. Mihaylovich NN. *Povysheniye ekspluatatsionnykh kharakteristik neftepromyslovogo oborudovaniya v territorial'no-proizvodstvennom predpriyatii* [dissertation]. Moscow; 2009. Available from: <https://www.dissercat.com/content/povyshenie-effektivnosti-ekspluatatsii-nasosnykh-skvazhin-optimizatsiei-raboty-shtangovykh-k>. (In Russ).
2. Khabibullin RA, Shabonas AR, Gurbatov NS, Timonov AV. Prediction of ESPs Failure Using ML at Western Siberia Oilfields with Large Number of Wells. Russian Petroleum Technology Conference; 2020 Oct 26–29; Virtual. Available from: <https://onepetro.org/SPERPTC/proceedings-abstract/20RPPTC/1-20RPPTC/D013S027R001/450252>.
3. Merembayev T, Kurmangaliyev D, Bekbauov B, Amanbek Y. A Comparison of machine learning algorithms in predicting lithofacies: Case studies from Norway and Kazakhstan. Energies. 2021;14(7):1896. doi: [10.3390/en14071896](https://doi.org/10.3390/en14071896).
4. Merembayev T, Amanbek Y. Time-series event prediction for the uranium production wells using machine learning algorithms. 56th U.S. Rock Mechanics/Geomechanics; 2022 June 26; Santa Fe, United States. Available from: <https://research.nu.edu.kz/en/publications/time-series-event-prediction-for-the-uranium-production-wells-usi-2>.
5. American Petroleum Institute. *API TR 11L Design Calculations for Sucker Rod Pumping Systems (Conventional Units), 5th edition*. Washington : API TR 11L; 2008.
6. Abdelaziz M, Lastra R, Xiao JJ. ESP data analytics: Predicting failures for improved production performance. Abu Dhabi International Petroleum Exhibition & Conference; 2017 Nov 13–16; Abu Dhabi, UAE. Available from: <https://www.scrip.org/reference/referencespapers?referenceid=2938839>.
7. Ivanonovskii VI. *Skvazhinnye nasosnye ustnovki dlya dobychi nefli*. Moscow; 2002. (In Russ).
8. Seradilla O, Zugastu E, Rodriguez J, Zurutuza U. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect. *Appl Intell*. 2020;52:10934–10964. doi: [10.1007/s10489-021-03004-y](https://doi.org/10.1007/s10489-021-03004-y).
9. Takacs G. *Sucker-Rod Pumping Handbook. Production engineering fundamentals and long-stroke rod pumping*. Hungary; 2015.
10. Rudakov V, Merembayev T, Amirgaliyev Y, Omarova P. Time Series Analysis of Biogas Monitoring with Deep Learning Approaches. 5th International Conference on Problems of Cybernetics and Informatics; 2023 Aug 28–30; Baku, Azerbaijan. Available from: <https://ieeexplore.ieee.org/document/10325955>.
11. Waskom ML. Seaborn: statistical data visualization. *Journal of Open Source Software*. 2021;6(60):3021. doi: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021).
12. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*. 2017;30.
13. Breiman L. Random forests. *Machine learning*. 2001;45:5–32.
14. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*. 2015;1(4). doi: [10.32614/cran.package.xgboost](https://doi.org/10.32614/cran.package.xgboost).
15. Hochreiter S, Schmidhuber J. Long Short-term Memory. *Neural Computation MIT-Press*. 1997;9(8):1735–1780. doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).

## ИНФОРМАЦИЯ ОБ АВТОРАХ

### \*Утемисова Лаура Гибратовна

ORCID [0000-0003-4194-6727](https://orcid.org/0000-0003-4194-6727)

e-mail: [l.utemissova@kmge.kz](mailto:l.utemissova@kmge.kz).

### Мерембаев Тимур Жумақанович

PhD

ORCID [0000-0001-8185-235X](https://orcid.org/0000-0001-8185-235X)

e-mail: [timur.merembayev@gmail.com](mailto:timur.merembayev@gmail.com).

### Бекбау Бакберген Ермекбайұлы

PhD

ORCID [0000-0003-2410-1626](https://orcid.org/0000-0003-2410-1626)

e-mail: [bakbergen.bekbau@gmail.com](mailto:bakbergen.bekbau@gmail.com).

## AUTHORS' INFO

### \*Laura G. Utemissova

ORCID [0000-0003-4194-6727](https://orcid.org/0000-0003-4194-6727)

e-mail: [l.utemissova@kmge.kz](mailto:l.utemissova@kmge.kz).

### Timur Zh. Merembayev

PhD

ORCID [0000-0001-8185-235X](https://orcid.org/0000-0001-8185-235X)

e-mail: [timur.merembayev@gmail.com](mailto:timur.merembayev@gmail.com).

### Bakbergen Ye. Bekbau

PhD

ORCID [0000-0003-2410-1626](https://orcid.org/0000-0003-2410-1626)

e-mail: [bakbergen.bekbau@gmail.com](mailto:bakbergen.bekbau@gmail.com).

\*Автор, ответственный за переписку/Corresponding Author