

Турчин О.Б.

Івано-Франківський національний технічний університет нафти і газу

ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ ГНШУ НА ОСНОВІ ДАНИХ RFID ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ

У статті основну увагу зосереджено на прогнозуванні ефективності ГНШУ за допомогою маркерів RFID, інтегрованих із передовими методами машинного навчання. Маркери RFID, що складаються з мікročипа та антени, пропонують надійний збір даних у суворих умовах, підтримуючи як активні, так і пасивні типи. Для активних маркерів потрібне джерело живлення, а для пасивних – енергія сигналу зчитувача. У цьому дослідженні маркери RFID застосовуються до критичних компонентів ГНШУ, таких як крокуюча балка та колінчастий вал, для моніторингу робочих параметрів, таких як цикли руху, сили навантаження та механічні напруги.

Високочастотні датчики RFID, що працюють на частоті 13,56 МГц, використовують індуктивний зв'язок для передачі енергії та даних. Це дослідження використовує дані RFID, взяті з набору даних, який включає години роботи, властивості рідини та механічні навантаження, з нафтового родовища середнього розміру в Техасі за 2020-2023 роки. Дані проходили попередню обробку для вирішення таких проблем, як відсутні значення та викиди, з подальшим масштабуванням для нормалізації.

Використовується двошарова ансамблева модель стекування, що поєднує методи RF, LightGBM, SVR і AdaBoost. Модель інтегрує дані RFID для прогнозування інтервалів технічного обслуговування, підвищуючи точність прогнозування. Оптимізація пошуку в сітці точно налаштовує параметри моделі, при цьому ансамблева модель стекування перевіряє окремі моделі за показниками RMSE, MAE і R^2 . Дослідження демонструє ефективність передових методів машинного навчання в оптимізації прогнозування ефективності насоса, підкреслюючи практичне застосування в нафтопромислових роботах.

Під час тестування запропонованого методу порівняно з окремими ML-методами було виявлено, що ансамблева модель стекування перевершила всі моделі з підвищенням точності на 2,01%, 4,20%, 6,95% і 9,20% порівняно з RF, LightGBM, SVR і AdaBoost, відповідно. Це вказує на те, що ансамблева модель стекування пропонує більш точне та надійне прогнозування ефективності насоса, демонструючи його потенціал для практичного застосування в нафтопромислових роботах.

Ключові слова: маркери RFID, ефективність насоса, машинне навчання, ансамблеві моделі, інтеграція датчиків, прогнозне обслуговування.

Постановка проблеми. Динамічні та постійно мінливі геологічні умови нафтових родовищ представляють серйозну проблему для підтримки ефективності та довговічності глибинно-насосних штангових установок (ГНШУ). Ці умови під впливом підповерхневого потоку рідини та інших факторів можуть призвести до поломок ГНШУ та інших неефективних операцій. З цієї причини постає абсолютна важливість контролювати та прогнозувати різні параметри, такі як вміст води, глибина закачування, максимальні та мінімальні навантаження, тиск масла, тиск в обсадній колоні, довжина ходу та довжина насосної штанги до зносу та корозії насоса, усі з яких тісно пов'язані між собою. Традиційні методи якісного аналізу та статистичного моделювання використовувалися для прогнозування циклів перевірки насосів, але ці підходи мають обмеження щодо точності та надійності.

У контексті нафтопромислових робіт моделі машинного навчання все частіше використовуються

для прогнозування циклів перевірки насосів ГНШУ. Однак багато існуючих досліджень покладаються на окремі моделі машинного навчання, які часто не забезпечують оптимальної прогнозовної ефективності. Щоб усунути це обмеження, було запропоновано більш складний підхід, який передбачає поєднання методів машинного навчання. Зокрема, такі методи, як дерева рішень, алгоритми посилення та опорна векторна регресія (SVR), були використані для покращення вибору функцій і підвищення точності прогнозів.

У межах поточного дослідження, вищезазначена комбінація аналітичних методів машинного навчання забезпечується даними з маркерів радіочастотної ідентифікації (RFID – Radio Frequency Identification), яка пропонує значний потенціал для підвищення ефективності насосних установок, забезпечуючи точний моніторинг, збір даних у реальному часі та прогнозне обслуговування. Будучи невід'ємною частиною систем автома-

тичної ідентифікації (AutoID), RFID працює через безконтактний обмін сигналами. Передача даних може відбуватися як на вимогу, так і автоматично, коли RFID-маркер наближається до зчитувача. Маркери RFID можуть прикріплюватися до різних компонентів ГНШУ, таких як крокуюча балка, противаги та приводні пристрої. Це дозволяє точно відстежувати ці деталі, контролювати їх знос і прогнозувати, коли потрібне технічне обслуговування. Запропонований підхід може допомогти запобігти аварійним ситуаціям і підтримувати оптимальну ефективність.

Аналіз останніх досліджень і публікацій. У науково-дослідницькому просторі сьогодення з'являються роботи, присвячені винаходу та аналізу методології по розробці моделей прогнозування ефективності ГНШУ на основі аналізу історичних даних і трендів у видобутку нафти.

У роботі [1] досліджується система Rod Pump Controller спрямована на оптимізацію виробництва та зниження експлуатаційних витрат у нафтовій і газовій промисловості, використовуючи машинне навчання і штучний інтелект, щоб допомогти операторам та інженерам-нафтовикам ефективніше керувати свердловинами. Оцінюється дедалі актуальніша модель оплати «як послуга», коли постачальники технологій керують усім обсягом електрики та автоматизації для моніторингу та оптимізації гирла свердловини. Традиційно роботу свердловин контролює персонал шляхом періодичних візитів, що є методом, схильним до неефективності та дорогого ремонту. Система Rod Pump Controller вирішує ці проблеми, забезпечуючи максимальне виробництво рідини при мінімізації експлуатаційних витрат. Оптимізація видобутку досягається шляхом регулювання швидкості насоса у відповідь на коливання рівнів у резервуарі, таким чином максимізуючи вихід без пошкодження обладнання та економії енергії. Граничні обчислення відіграють вирішальну роль, переміщуючи додатки, дані та обчислювальну потужність на свердловини штангових насосів, сприяючи аналітиці в реальному часі завдяки поєднанню ML і хмарного навчання. Ця установка використовує обширні історичні дані та надійну обчислювальну потужність у диспетчерських, виконуючи моделі на межі, щоб забезпечити негайний зворотний зв'язок з операційними групами.

Метою дослідження [2] є перевірка використання високочастотних даних про тиск з високою роздільною здатністю, виміряних на гирлі свердловин ГНШУ, для оцінки та оптимізації продуктивності. У документі викладено обґрунтування

методів, які будуть перевірені під час польових випробувань. У цих випробуваннях високоефективні пристрої моніторингу тиску встановлюються на трубах свердловин штангових насосів, перед зворотним клапаном лінії потоку та на самій лінії потоку. Ці пристрої IoT (Internet of Things – Інтернет Речей), призначені для небезпечних зон класу I, вимірюють тиск з інтервалом в одну секунду з роздільною здатністю 0,006 psi. Щосекундні вимірювання синхронізуються за часом і компенсуються за температурою для точності, а дані надсилаються до хмарної служби протягом кількох тижнів. Різниця в тиску між трубопроводом і потоковою лінією аналізується в п'ятихвилинних вікнах даних, що дає змогу зрозуміти тиск, який чинить насос на поточну лінію протягом кожного циклу. Тиск, який служить проміжним показником для потоку, дозволяє вирахувати заповнення та ефективність вибірного насоса з профілю потоку на поверхні, минаючи штанги та пов'язані з ними помилки в традиційному зборі даних про продуктивність насоса. Ефективність буде порівнюватися з результатами виробничих випробувань, щоб оцінити її життєздатність. Аналізуючи велику вибірку п'ятихвилинних вікон даних протягом місяця, дослідження оцінить метод у різних робочих умовах. Крім того, дослідження спрямоване на виявлення тенденцій тиску, які корелюють із конкретними умовами насоса, потенційно допомагаючи в діагностиці проблем з насосом. Ця методика, яка раніше була багатообіцяючою в тематичному дослідженні (SPE-209253-MS), може запропонувати економічно ефективну альтернативу для моніторингу продуктивності старіння свердловин штангових насосів.

Крім того, варто зазначити праці наступних науковців: Мехтієва З. [3] Некрасов В.І., Зіганшин Р.А. Саваст'їн М.Ю., Зіганшина О.В., Шпітько Г.М. [4], Алієв А., Алієва С., Фаталієв Ст. [5], Нашед С., Зідан Х., Ельхаді А., Хелмі А., Єхія Т., Ельнагар Х. [6], Чан Т., Гальєгос А., Понд Би., Дуек Ч. [7], Акілдіз Т., Ку Р., Хардер Н., Ебрахімі Н., Махдавіфар Х. [8], Соуза А., Флорес Дж., Міллендес Л., Філіо М., Мо Ю., Жако К., Делані М. [9], Пекгор М., Араблуді Р., Нікзад М., Масуд С. [10], Суганьядеві К., Сангіта Д. [11], Мандхаре П., Аутаде Р., Гутал Ст., Шрібахадуркар Д., Упадхе П. [12], Кукреджа Д., Менон Р., Дхачнеш М., Радж А., Субаш А., Манодж К., Мадху Л. [13], Сенгупта С., Рей С., Сен Д., Саркар С., Рой С. [14], Бакір З., Мотлак Х. [15] та інших.

Проте, беручи до уваги вище зазначену наукову документацію, питання, пов'язане з методологією

по розробці моделей прогнозування ефективності ГНШУ на основі аналізу історичних даних і трендів у видобутку нафти, все ще залишається недостатньо дослідженим та потребує подальшого опрацювання.

Постановка завдання. Метою статті є розробка моделі прогнозування ефективності ГНШУ на основі ML-аналізу даних і трендів у видобутку нафти за допомогою маркерів RFID.

Виклад основного матеріалу. Ядром систем RFID є транспондери/маркери. Ці маркери складаються з мікročіпа та антени, призначених для радіозв'язку із зовнішніми зчитувачами. Переваги транспондерів RFID включають високу надійність в екстремальних умовах навколишнього середовища, безконтактний збір даних, значну ємність для зберігання даних, а також можливість зчитувати кілька носіїв даних одночасно в режимі реального часу.

Пристрої RFID широко поділяються на активні та пасивні типи. Активні маркери потребують джерело живлення, підключене до джерела інфраструктури або живлення від внутрішньої батареї, що обмежує термін їх служби на основі накопиченої енергії та кількості операцій зчитування. Навпаки, пасивні маркери RFID споживають енергію від сигналу зчитувача, що робить їх невимагаючими в обслуговуванні, але обмеженими в діапазоні та функціональності порівняно з активними маркерами.

Оснащення компонентів ГНШУ RFID має сенс лише в тому випадку, якщо ці компоненти можуть сприймати певні властивості навколишнього середовища. Ця можливість зазвичай досягається за допомогою датчиків, які записують фізичні, хімічні або біологічні параметри. Потім зібрані дані повинні бути належним чином передані, як правило, через кабелі або систему шини до центрального процесора. Складні системи або процеси, такі як ті, що пов'язані з видобутком нафти, часто контролюються численними датчиками, інтегрованими в сенсорні мережі.

Високочастотні (HF – High frequency) датчики RFID передають і отримують як енергію, так і дані за допомогою індуктивного зв'язку, що досягається змінним магнітним полем між котушковими антенами зчитувача та маркера. Цей індуктивний зв'язок генерується резонансною ланцюгом LC-бака, що містить котушки індуктивності та конденсатори, які резонують на певній частоті, як правило, 13,56 МГц для ВЧ-систем RFID. Індуквана напруга в антені маркера прямо пропорційна швидкості зміни магнітного потоку, на яку впли-

ває кількість витків котушки, напруженість магнітного поля та відносна орієнтація між зчитувачем і маркером.

У контексті ГНШУ маркери RFID розміщуються на різних критичних компонентах для моніторингу конкретних робочих аспектів. Наприклад, прикріплення маркерів RFID до крокуючої балки дозволяє відстежувати кількість циклів, швидкість і кути руху, полегшуючи виявлення нерівностей, таких як зміщення або знос. Подібним чином розміщення маркерів на кривошипі та противагах допомагає контролювати прикладені зусилля, забезпечуючи збалансовану роботу та запобігаючи надмірному напруженню, яке може знизити ефективність.

Крім того, RFID-маркери можуть бути встановлені на рукоятці сошки та редукторі насоса, щоб аналізувати механічні навантаження та рівні змащення, забезпечуючи безперебійну роботу та прогножуючи можливі несправності. Полірований шток і сальник, необхідні для запобігання витокам масла, отримують переваги від моніторингу RFID шляхом відстеження зносу та цілісності ущільнення, таким чином підтримуючи оптимальну ефективність насоса. Крім того, маркери RFID на двигуні можуть контролювати такі параметри, як енергоспоживання, температура та вібрація, що дозволяє завчасно виявляти неефективність або проблеми з двигуном.

Інтеграція RFID-маркерів із датчиками рівня рідини в свердловині забезпечує дані в реальному часі про рівні нафти та ефективність перекачування. Ці дані, якщо їх співвіднести з робочими показниками насоса, пропонують комплексне уявлення про процес екстракції.

У контексті цього дослідження, використовується протокол передачі, «reader-talks-first», який забезпечує надійний зв'язок між зчитувачем і маркерами. Цей протокол використовує амплітудну модуляцію та кодування позиції імпульсу для передачі даних із механізмами точної синхронізації для точного декодування отриманих сигналів. У системах ГНШУ це гарантує, що дані з різних маркерів точно збираються та обробляються без перешкод, навіть у середовищах із кількома активними маркерами.

Удосконалені системи RFID можуть включати додаткові датчики, такі як акселерометри або гіроскопи, для визначення орієнтації та руху компонентів насоса. Наприклад, акселерометр, приєднаний до RFID-маркера, може вимірювати кути нахилу, надаючи розуміння механічних навантажень і робочої динаміки насоса. Ці вимірювання,

оброблені мікроконтролером у маркері RFID, передаються на зчитувач, що дозволяє точно обчислювати відстань і орієнтацію на основі даних про силу сигналу та нахил.

В якості наступного кроку, передбачається проведення інтеграції даних RFID у двошарову ансамблеву модель стекування, призначену для прогнозування стану та потреби в обслуговуванні компонентів ГНШУ. Ця модель використовує сильні сторони методів машинного навчання, включаючи Random Forest (RF), LightGBM, Support Vector Regression (SVR) і Adaptive Boosting (AdaBoost), для аналізу складних взаємозв'язків між різними робочими параметрами, записаними маркерами RFID. Ансамблеву модель стекування оптимізовано за допомогою ретельного процесу пошуку сітки для точного налаштування її параметрів, гарантуючи точну інтерпретацію даних і прогнозування потенційних збоїв або вимог до обслуговування.

Так як даним, які містяться в системах із підтримкою RFID для моніторингу насосів, властива велика розмірність, модель включає застосування RF. Метод виявляє сильну стійкість до переобладнання, що робить його придатним для складних наборів даних, що фіксуються датчиками RFID, прикріпленими до компонентів насоса. Для реалізації RF-методу створено кілька дерев рішень, при цьому кожне дерево рішень незалежно навчається на випадково вибраних підвбірках даних. Остаточний результат прогнозування для ефективності ГНШУ отримується шляхом зваженого середнього прогнозування з цих кількох дерев.

LightGBM є методом, який відомий своїм ефективним паралельним навчанням, високою швидкістю обробки, меншим споживанням пам'яті та підвищеною точністю, які є критично важливими атрибутами для обробки даних у реальному часі з маркерів RFID у системах ГНШУ. LightGBM використовує кілька обчислювальних методів, у тому числі на основі гістограми, прискорення різниці гістограм та обмеження глибини поверхшинного зростання (Leaf-wise growth). Стратегія поверхшинного зростання з додатковим обмеженням максимальної глибини допомагає збалансувати ефективність і запобігти переобладнанню під час моделювання зв'язку між даними RFID і продуктивністю насоса.

Основна ідея SVR, застосована в цьому контексті, полягає в мінімізації відстані найдальших точок вибірки від гіперплощини. SVR має на меті передбачити ефективність насоса шляхом розміщення найкращої можливої гіперплощини в межах є-нечутливої трубки, яка представляє допуск до

помилки. Зосереджуючись на мінімізації помилок за межами цієї зони допуску, SVR ефективно моделює варіації продуктивності ГНШУ, згідно з показниками у RFID-отриманих даних.

AdaBoost ітеративно додає слабкі класифікатори, доки не досягне достатньо низького рівня помилок у прогнозуванні ефективності насоса. Спочатку AdaBoost призначає рівні ваги всім спостереженням у навчальній вибірці. Потім він ітеративно оновлює ці ваги, навчаючи слабкі класифікатори на зважених даних для покращення точності прогнозу. Остаточний прогноз робиться шляхом об'єднання результатів цих слабких класифікаторів шляхом зваженої більшості голосів, гарантуючи, що модель точно відображає показники ефективності насосів на основі вхідних даних RFID.

Ансамблева модель перевершує межі продуктивності окремих моделей і підвищує загальну точність прогнозування, представляючи найкращий досяжний результат завдяки використанню сильних сторін багатьох підходів машинного навчання. Однак реалізація ансамблевих моделей у області поточного дослідження є складною через припущення сильної незалежності між моделями, що не завжди може бути вірним у практичних застосуваннях. Отже, продуктивність ансамблевої моделі стекування залежить від конкретних характеристик використовуваних даних RFID.

Зважене усереднення в рамках процесу ансамблю може призначати різні ваги під час усереднення, щоб підвищити точність кінцевого прогнозу ефективності насосної станції, використовуючи різноманітні вихідні дані включених моделей.

Стекування ефективно поєднує кілька слабких учбових моделей у сильнішу прогностичну модель. Для прогнозування ефективності ГНШУ це передбачає навчання моделі стекування, яка використовує прогнози з кількох базових моделей для створення більш надійної метамоделі. Спочатку кілька моделей машинного навчання навчаються з використанням одного набору даних на основі даних RFID маркерів. Кожна базова модель надає прогнози щодо даних навчання, які потім об'єднуються учбовими моделями. Метамодель, навчена визначати найкращу комбінацію цих прогнозів, сприяє більш точній загальній моделі для прогнозування ефективності насосної станції. Алгоритми RF і XGBoost використовуються для вибору функцій, відсіюючи найважливіші функції. Базові моделі оптимізовані за допомогою дослідження перевірки, під час якого на чотирьох обраних моделях виконується п'ятикратне перехресне навчання. Згодом ридж-регресія навча-

ється як інструмент для вивчення нових функцій і міток, згенерованих із набору перевірки, що призводить до остаточної добре навченої моделі для прогнозування ефективності насосної станції.

У цьому дослідженні набір даних походить з оперативної бази даних насосів на нафтовому родовищі середнього розміру, розташованому в Пермському басейні, штат Техас, США, і охоплює період з 2020 по 2023 рік. Мета полягає в тому, щоб передбачити ефективність насосів шляхом аналізу часового інтервалу між двома циклами технічного обслуговування, використовуючи маркери RFID, розміщені на ключових компонентах насосів для збору робочих даних у реальному часі. Основною змінною реакції є інтервал часу між циклами технічного обслуговування, виміряний у днях.

Набір даних складається з 40 512 записів, кожен з яких містить двадцять характеристик, отриманих з інформації, наданої маркерами RFID та іншими показаннями датчиків:

- загальна кількість годин, відпрацьованих безперервно кожним насосом;
- відсоток води в видобутій сирій нафті;
- відстань від поверхні до насоса, записана в метрах;
- проникність породи в місці свердловини, виміряна в мілідарсі;
- пікове навантаження в кілоньютонах, що діє на гніздо під час роботи;
- мінімальне навантаження на жолобі, яке відчуває насос;
- максимальне відхилення стовбура свердловини від вертикальності в градусах;
- тиск на рівні пласта, виміряний у мегапаскалях;
- тиск у колоні НКТ (насосоно-компресорна труба), у мегапаскалях;
- довжина ходу насоса, у метрах;
- кількість ударів за хвилину;
- середньодобовий видобуток у барелях;
- в'язкість видобутої нафти, виміряна в сантитуазах;
- температура на рівні водойми, записана в градусах Цельсія;
- загальна довжина насосної штанги в метрах;
- швидкість потоку нафти, виміряна в барелях на день;
- загальна довжина колони НКТ;
- електрична енергія, споживана насосом, у кіловат-годинах;
- амплітуда вібрації, виміряна на насосній станції, у міліметрах на секунду;

– потужність отриманого сигналу RFID, що вказує на стан компонентів.

Враховуючи велику кількість свердловин і різноманітні робочі умови, набір даних містить такі помилки, як відсутні записи, дублікати та неправильні записи через несправності датчиків або помилки ручного введення. Попередня обробка даних має вирішальне значення для вирішення цих проблем. Лише записи без пропущених значень були збережені для навчання моделі, щоб забезпечити точність.

Щоб усунути аномалії, використовувалися коробкові діаграми для виявлення та видалення викидів. Цей метод ефективно фіксує розподіл даних, гарантуючи, що екстремальні значення не спотворюють продуктивність моделі. Останній крок включав масштабування даних для нормалізації діапазонів ознак, покращуючи порівняльність різних показників. Ця нормалізація є важливою для прискорення конвергенції алгоритмів градієнтного спуску, які використовуються в процесі навчання моделі.

Оптимізацію параметрів для моделей машинного навчання проводили за допомогою пошуку в сітці з перехресною перевіркою. Цей вичерпний метод пошуку перевіряє всі можливі комбінації параметрів моделі, щоб визначити оптимальний набір, який забезпечує найкращу продуктивність. Систематично оцінюючи конфігурацію кожного параметра, пошук у сітці гарантує, що остаточно модель точно налаштована на точність і надійність.

(RF) було налаштовано з початковим параметром 'maxfeatures = 12' і 'n_estimators = 1000' дерев. Після застосування оптимізації пошуку в сітці найкращими параметрами були 'maxfeatures = 15' і 'n_estimators = 180'.

LightGBM передбачав більш складне налаштування параметрів через численні параметри за замовчуванням. Спочатку вищий рівень навчання використовувався для прискорення конвергенції. Пошук у сітці уточнив параметри до 'n_estimators = 120', 'num_leaves = 40' і 'learning_rate = 0,05', щоб збалансувати складність і точність моделі.

SVR використовувала ядро радіальної базисної функції (RBF – radial basis function). Коефіцієнт штрафу C змінювався між 0,5 і 2000, а гамма-коефіцієнт ядра був оптимізований, в результаті чого C = 800 і гамма = 0,02.

AdaBoost вимагав налаштування кількості оцінювачів і швидкості навчання. Оптимальними параметрами було визначено 'n_estimators = 80' і 'learning_rate = 0,25', з налаштованим вручну 'random_state = 42' для забезпечення узгодженості під час ітерацій навчання.

Форматовані дані були розділені на 80% для навчання та 20% для тестування. Показники продуктивності, включаючи R^2 , MAE (Mean Absolute Error – середня абсолютна похибка) і RMSE (Root Mean Squared Error – середньоквадратична похибка), використовувалися для оцінки точності прогнозування кожної моделі. Показники для чотирьох моделей на основі даних навчання наведено в таблиці 1.

Таблиця 1

Показники ефективності моделі

Метод	RMSE	MAE	R^2
RF	2.50	1.00	91.50%
LightGBM	3.90	1.80	89.00%
SVR	4.10	1.20	87.00%

Метамоделі на основі ридж-регресії разом з RF і SVR були оптимізовані за допомогою налаштування гіперпараметрів для досягнення найкращої продуктивності. Остаточні результати точності для метамodelей відображаються в таблиці 2.

Таблиця 2

Результати навчання метамodelей

Метамодель	Точність навчання	Точність тестування
Ридж-регресія	97.50%	92.30%
RF	98.20%	91.40%
SVR	96.75%	90.50%

Ансамблева модель стекування перевершила всі окремі моделі з точки зору RMSE, MAE і R^2 , з підвищенням точності на 2,01%, 4,20%, 6,95% і 9,20% порівняно з RF, LightGBM, SVR і AdaBoost відповідно. Це вказує на те, що ансамблева модель стекування пропонує більш точне та надійне прогнозування ефективності ГНШУ, демонструючи її потенціал для практичного застосування в нафтопромислових роботах.

Висновки. У підсумку необхідно відзначити, що вибір функцій і ретельні етапи попередньої обробки гарантують, що модель навчена на високоякісних даних, мінімізуючи вплив шуму та помилок. Використання методів вибору функцій, таких як RF і XGBoost, підкреслює необхідність фіксації складних взаємодій між змінними, які часто присутні в даних нафтопромислу.

Пошук у сітці додатково вдосконалює модель, оптимізуючи її для ефективної роботи в конкретних умовах нафтового родовища Пермського басейну. Цей методичний підхід не тільки підвищує точність моделі, але також дає змогу зрозуміти ключові фактори, що впливають на ефективність насоса, наприклад години роботи, вміст води та рівні вібрації.

Загалом, запропонований метод забезпечує чітку структуру для застосування машинного навчання для прогнозування ефективності ГНШУ, пропонуючи структуровану методологію, яку можна відтворити або змінити для інших нафтових родовищ або промислових застосувань.

Список літератури:

1. Susanto T. Leverage Telemetry Scada and Machine Learning on Pumpjack Wellhead Production Facilities. *Paper presented at the ADIPEC*. October 2022. DOI: <https://doi.org/10.2118/211194-MS>.
2. Vekved K., Ito T., Gordon R. Optimizing Rod Pump Performance Using High Frequency, High Resolution Pressure Data at the Wellhead. *SPE Oklahoma City Oil and Gas Symposium*. April 2023. DOI: <https://doi.org/10.2118/213052-MS>.
3. Mehdiyeva Z. Study Of Optimization Of The Structural Design Of Support Units Of Pumpjacks. *ETM - Equipment, Technologies, Materials*. 2023. № 18. P. 66-75. DOI:10.36962/ETM18062023-66.
4. Pumpjack of the telescope-mechanics type / Nekrasov V.I., Ziganshin R.A., Savast'in M.Y.U., Ziganshina A.V., Shpitko G.N. *International Journal of Mechanical Engineering and Technology*. 2018. № 9. P. 1447-1454.
5. Aliyev A., Aliyeva S., Fataliyev V. Influence of Technological Factors on the Performance and Aging Process of Oil Pumpjack. *Advances in Science and Technology*. 2024. Vol. 148. P. 111-116. DOI:10.4028/p-rgm0Zj.
6. Application of Machine Learning and Deep Learning to Predict Production Rate of Sucker Rod Pump Wells / Nashed S., Zidan H., Elhadydy A., Helmy A., Yehia T., Elnaggar H. *Paper presented at the GOTECH (Dubai, UAE)*. May 2024. DOI:10.2118/219231-MS.
7. Harnessing Machine Learning for Proactive Detection and Predictive Maintenance of Sticking Pumps in Rod Lift Oil Wells / Chan T., Gallegos A., Pond B., Dueck Ch. *Paper presented at the GOTECH (Dubai, UAE)*. May 2024. DOI:10.2118/219563-MS.
8. ML-Aided Collision Recovery for UHF-RFID Systems / Akyildiz T., Ku R., Harder N., Ebrahimi N., Mahdaviifar H. *Proceedings of the 2022 IEEE International Conference on RFID (RFID) (Las Vegas, NV, USA, May 17–19, 2022)*. P. 41-46. DOI: 10.1109/RFID54732.2022.9795968
9. Radiofrequency identification tracking system (RFID) significantly improves blood bank inventory management and decreases staff work effort / Souza A., Flores J., Millendez L., Filio M., Mo Y., Jacquot C., Delaney M. *Transfusion*. 2024. № 64. DOI:10.1111/trf.17771.

10. Displacement Estimation via 3D-Printed RFID Sensors for Structural Health Monitoring: Leveraging Machine Learning and Photoluminescence to Overcome Data Gaps / Pekgor M., Arablouei R., Nikzad M., Masood S. *Sensors*. 2024. № 24. 1233 p. DOI:10.3390/s24041233.
11. Suganyadevi K., Sangeetha D. Intelligent Fuel Pump System with RFID and the Internet of Things. *2023 International Conference on Networking and Communications (ICNWC)*. 2023. P. 1–6.
12. RFID Based Petrol Pump Automation System. *International Journal of Advanced Research in Science, Communication and Technology* / Mandhare P., Autade R., Gutal V., Shribahadurkar D., Upadhye P. 2021. P. 788-792. DOI:10.48175/IJARSCT-1479.
13. The Startup Analysis Canvas: Smartfill RFID System / Kukreja D., Menon G., Dhatchnesh M., Raj A., Subash A., Manoj K., Madhu L. *International Journal for Research in Applied Science and Engineering Technology*. 2024. № 12. P. 255-263. DOI:10.22214/ijraset.2024.59744.
14. Analysis of Four-Stage Charge Pump Circuit for UHF RFID Tag Design / Sengupta S., Ray S., Sen D., Sarkar S., Roy S. *2019 Devices for Integrated Circuit (DevIC)* (Kalyani, India, 23-24 March 2019). IEEE, 2019. DOI:10.1109/DEVIC.2019.8783538.
15. Baqir Z., Motlak H. Smart Automatic Petrol Pump System Based on RFID and ESP8266. *Journal of Physics: Conference Series*. 1933. 2021. № 012109. DOI:10.1088/1742-6596/1933/1/012109.

Turchyn O.B. PREDICTING THE EFFICIENCY OF PUMPJACKS BASED ON RFID DATA AND MACHINE LEARNING METHODS

This study focuses on predicting the efficiency of pumpjacks using RFID markers integrated with advanced machine learning methods. RFID markers, consisting of a microchip and antenna, offer reliable data collection in harsh environments, supporting both active and passive types. Active markers require a power source, while passive ones harness energy from the reader signal. In this research, RFID markers are applied to critical pumpjack components, such as the walking beam and crankshaft, to monitor operational parameters like movement cycles, load forces, and mechanical stresses.

High-frequency RFID sensors, operating at 13.56 MHz, use inductive coupling for energy and data transmission. This research leverages RFID data, including operational hours, fluid properties, and mechanical loads, from a medium-sized oil field in Texas over 2020-2023. The data undergoes preprocessing to address issues like missing values and outliers, followed by scaling for normalization.

A two-layer ensemble stacking model is utilized, combining Random Forest (RF), LightGBM, Support Vector Regression (SVR), and AdaBoost. The model integrates RFID data to predict maintenance intervals, enhancing prediction accuracy. The grid search optimization fine-tunes model parameters, with the stacking ensemble model outperforming individual models in RMSE, MAE, and R² metrics. The study demonstrates the effectiveness of advanced machine learning techniques in optimizing pumpjack efficiency prediction, highlighting the practical applicability in oil field operations.

During testing of the proposed method compared to individual ML-methods, it was revealed that stacking ensemble model outperformed all models with improvements in accuracy of 2.01%, 4.20%, 6.95%, and 9.20% over RF, LightGBM, SVR, and AdaBoost, respectively. This indicates that the stacking ensemble model offers a more precise and reliable prediction for pumpjack efficiency, showcasing its potential for practical applications in oilfield operations.

Key words: RFID markers, pumpjack efficiency, machine learning, ensemble models, sensor integration, predictive maintenance.