

Івано-Франківський національний технічний університет нафти і газу

Інститут інформаційних технологій
Кафедра інформаційно-вимірювальних технологій

Скрипник Володимир Васильович

(прізвище, ім'я, по батькові)

УДК 681.5.

(індекс)

МАГІСТЕРСЬКА РОБОТА

Удосконалення методу оцінки якості природного газу на основі алгоритмів
машинного навчання

(назва роботи)

Метрологія та вимірювальна техніка

(назва освітньої програми)

175 – «Інформаційно-вимірювальні технології»

(шифр і назва спеціальності)

Скрипник В. В.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Чуйко Мирослава Михайлівна, к.т.н., доц.

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

Завідувач кафедри

(посада) (підпис) (дата) (ініціали та прізвище)

Рецензент

(посада) (підпис) (дата) (ініціали та прізвище)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Івано-Франківськ, 2024 р.

Івано-Франківський національний технічний університет нафти і газу

(повне найменування закладу вищої освіти)

Інститут *інформаційних технологій*

Кафедра *інформаційно-вимірювальних технологій*

Освітній рівень *другий (магістерський)*

Спеціальність *152- Метрологія та інформаційно-вимірювальна техніка*

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри ІВТ

О.Є. Середюк

« » 2024 року

**З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТОВІ**

Скрипник Володимир Васильович

(прізвище, ім'я, по батькові)

1. Тема роботи *Удосконалення методу оцінки якості природного газу на основі алгоритмів машинного навчання*

керівник роботи *Чуйко Мирослава Михайлівна, к.т.н., доц.*

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом закладу вищої освіти від "03" 12.2024 року № 787/7

2. Строк подання студентом роботи 20.01.2025р.

3. Вихідні дані до роботи: точність вимірювання відстані до поверхні об'єкта – 1 мм; роздільна здатність сканування – 0,5 мм; діапазон сканування 360°

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналіз стану сучасних методів оцінки якості природного газу

2. Основи та алгоритми машинного навчання в аналізі даних

3. Розробка та реалізація моделі машинного навчання для оцінки якості природного газу

4. Експериментальне дослідження та аналіз результатів

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1. Оптична схема двопроменевого дисперсійного інфрачервоного спектрометра

2. Схематичне зображення алгоритмів моделі «Random forest»

3. Види фреймворків машинного навчання

4. Процес оптимізації гіперпараметрів

5. Інструменти для реалізації та візуалізації моделі

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|----------------------|---|----------------|------------------|
| | | завдання видав | завдання прийняв |
| <i>нормоконтроль</i> | <i>Лютак З.П.</i> | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів магістерської роботи | Термін виконання етапів роботи | Примітка |
|-------|---|--------------------------------|----------|
| 1. | <i>Аналіз стану сучасних методів оцінки якості природного газу</i> | <i>12.12-20.12.2024</i> | |
| 2. | <i>Основи та алгоритми машинного навчання в аналізі даних</i> | <i>20.12-31.12.2024</i> | |
| 3. | <i>Розробка та реалізація моделі машинного навчання для оцінки якості природного газу</i> | <i>01.01-10.01.2025</i> | |
| 4. | <i>Експериментальне дослідження та аналіз результатів</i> | <i>10.01-12.01.2025</i> | |
| 5. | <i>Оформлення магістерської роботи</i> | <i>12.01-20.01.2025</i> | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Студент _____ Скрипник В. В.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Чуйко М. М.
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Магістерська робота «Удосконалення методу оцінки якості природного газу на основі алгоритмів машинного навчання»: 60 с., 10 рис., 30 джерел.

Об'єкт дослідження процес оцінки якості природного газу

Предмет дослідження методи та алгоритми машинного навчання для аналізу фізико-хімічних параметрів природного газу, їх оптимізація для підвищення точності та автоматизації.

Метою і завданням дослідження є розробка та впровадження нових або вдосконалених алгоритмів машинного навчання для оптимізації процесу оцінки якості природного газу. Запропонований підхід має на меті не тільки покращити існуючі методи шляхом автоматизації процесів аналізу та обробки даних, але й забезпечити наукову новизну через розробку оригінальної моделі, яка враховує складні зв'язки між ключовими параметрами газу.

Ключові слова: природний газ, якість, машинне навчання, алгоритми, фізико-хімічні параметри, регресія, класифікація, оптимізація, автоматизація.

ABSTRACT

Master's thesis «Intelligent System for Industrial Pipeline Leakage Monitoring Using Machine Learning Methods» (60 pages, 10 figures, 30 references):

Object of research: the process of assessing the quality of natural gas.

Subject of research: methods and machine learning algorithms for analyzing the physicochemical parameters of natural gas, and their optimization to improve accuracy and automation.

The purpose and objectives of the study are to develop and implement new or improved machine learning algorithms for optimizing the natural gas quality assessment process. The proposed approach aims not only to enhance existing methods by automating data analysis and processing but also to ensure scientific innovation through the development of an original model that considers complex relationships between the key parameters of the gas.

Keywords: natural gas, quality, machine learning, algorithms, physicochemical parameters, regression, classification, optimization, automation.

ЗМІСТ

| | |
|--|----|
| Вступ | 7 |
| 1. Аналіз стану сучасних методів оцінки якості природного газу | 11 |
| 1.1. Фізико-хімічні параметри природного газу | 11 |
| 1.2. Традиційні методи оцінки якості: огляд і обмеження | 12 |
| 1.3. Математичні моделі та алгоритми, що застосовуються в галузі | 18 |
| 1.4. Аналіз перспективності використання алгоритмів машинного навчання | 22 |
| 2. Основи та алгоритми машинного навчання в аналізі даних | 22 |
| 2.1. Введення в машинне навчання: концепції та основні підходи | 22 |
| 2.2. Регресійні моделі: методи та принципи | 25 |
| 2.3. Алгоритми класифікації та їх застосування у вимірювальних задачах | 26 |
| 2.4. Моделі кластеризації та методи виявлення аномалій | 27 |
| 2.5. Огляд фреймворків та інструментів для реалізації моделей машинного навчання | 30 |
| 3. Розробка та реалізація моделі машинного навчання для оцінки якості | 34 |
| 3.1. Вибір параметрів для аналізу та обробка даних | 34 |
| 3.2. Побудова та налаштування моделі машинного навчання | 34 |
| 3.3. Опис процесу навчання та тестування моделі | 35 |
| 3.4. Оптимізація алгоритму та обробка помилок | 38 |
| 3.5. Вибір критеріїв для оцінки якості моделі | 39 |
| 3.6. Технічна реалізація. | 40 |
| 4. Експериментальне дослідження та аналіз результатів | 44 |
| 4.1. Опис експериментального середовища та використаних даних | 44 |
| 4.2. Порівняння результатів моделі машинного навчання з традиційними методами | 45 |
| 4.3. Оцінка точності, надійності та ефективності моделі | 46 |
| 4.4. Аналіз впливу алгоритмів на точність оцінки якості газу | 47 |
| 4.5. Аналіз отриманих результатів | 50 |

| | |
|----------------------------|----|
| Висновки | 6 |
| Список використаних джерел | 53 |
| Додатки | 55 |
| | 56 |

ВСТУП

Сучасна енергетична промисловість вимагає високої точності й надійності в оцінці якості природного газу, що є ключовим компонентом у забезпеченні стабільного та ефективного функціонування енергетичних систем. Традиційні методи оцінки якості природного газу базуються на класичних підходах до вимірювання фізико-хімічних параметрів, які мають певні обмеження у швидкості, точності та обробці великих обсягів даних. У цьому контексті виникає потреба в застосуванні сучасних технологій, таких як алгоритми машинного навчання, які здатні значно підвищити точність і ефективність методів оцінки якості природного газу.[1]

Оцінка якості природного газу є одним із ключових завдань енергетичної промисловості, яка нині стоїть перед викликами глобального масштабу. Сучасний енергетичний сектор спрямований на забезпечення ефективного, надійного та екологічно безпечного використання природних ресурсів. У цьому контексті природний газ, який вважається одним із найважливіших джерел енергії, набуває критичного значення через свою відносну екологічність і широке застосування в промисловості, електроенергетиці, транспорті та побутовому секторі.[2]

Якість природного газу безпосередньо впливає на ефективність його використання. Наприклад, енергетична цінність газу, що визначається його теплотворною здатністю, залежить від вмісту метану та інших компонентів. Присутність домішок, таких як вуглекислий газ (CO_2), сірководень (H_2S) та азот (N_2), може значно знижувати якість газу, підвищувати витрати на його обробку і створювати технічні труднощі. Наприклад, високий вміст H_2S може спричинити корозію трубопроводів, тоді як CO_2 знижує калорійність газу і, відповідно, економічну вигідність його використання.

Важливість оцінки якості природного газу обумовлена також суворими вимогами міжнародних стандартів. ISO 6976 є базовим стандартом для розрахунку фізико-хімічних властивостей газових сумішей. Він регламентує методи визначення теплотворної здатності, густини і вмісту домішок.

Дотримання цих стандартів забезпечує уніфікацію вимірювань і гарантує, що газ відповідає необхідним критеріям для використання у промисловості. Крім того, відповідність міжнародним стандартам є ключовою умовою для участі на глобальних ринках, особливо в умовах зростання експорту природного газу.[3,4]

З економічної точки зору, якість природного газу відіграє центральну роль у визначенні вартості палива, витрат на його транспортування та обробку. Наприклад, у випадку використання газу із високим вмістом домішок зростають витрати на його очищення, що може вплинути на рентабельність підприємств. Крім того, непередбачуваність якості газу може призвести до простоїв у виробничих процесах або навіть до пошкодження обладнання, що не лише знижує ефективність, але й створює додаткові витрати.

Одним із основних викликів сучасної енергетичної промисловості є забезпечення сталого розвитку, що включає зниження викидів парникових газів і використання екологічно безпечних джерел енергії. Природний газ є важливим компонентом цього переходу, оскільки він виділяє значно менше CO₂ під час згоряння, ніж вугілля чи нафта. Однак його якість має прямий вплив на екологічність. Наприклад, газ із високим вмістом H₂S або інших токсичних домішок може спричинити викиди забруднюючих речовин, що суперечить принципам сталого розвитку.

Екологічний аспект якості газу є не менш важливим у контексті боротьби зі зміною клімату. За даними Міжнародного енергетичного агентства (IEA), перехід до використання більш чистих джерел енергії, таких як природний газ, може знизити викиди парникових газів на 10-15% у короткостроковій перспективі. Однак для досягнення цього результату необхідно забезпечити контроль якості газу на всіх етапах його обробки та транспортування.

Безпека використання природного газу також залежить від його якості. Присутність сірководню або інших токсичних домішок становить серйозну загрозу для здоров'я людей, які працюють із газовим обладнанням, а також для кінцевих споживачів. Технічні системи, що використовуються для

транспортування та зберігання газу, можуть бути пошкоджені через наявність корозійно активних речовин, що підвищує ризик аварійних ситуацій.[5]

Ще одним важливим аспектом є зростання глобального попиту на природний газ, який стимулює розвиток технологій для його видобутку, транспортування та використання. У зв'язку з цим питання забезпечення якості газу стає ще більш актуальним. Наприклад, розвиток технологій скрапленого природного газу (СПГ) вимагає суворого контролю параметрів газу, таких як вологість, вміст вуглекислого газу і теплотворна здатність, оскільки будь-які відхилення можуть спричинити технічні проблеми під час скраплення або транспортування.

В умовах зростання конкуренції на світових енергетичних ринках високоякісний природний газ стає стратегічною перевагою. Країни-експортери, такі як Катар, Австралія та Росія, активно впроваджують новітні технології для підвищення якості свого продукту. Наприклад, автоматизація процесів контролю якості на основі машинного навчання дозволяє значно знизити витрати і забезпечити стабільно високу якість газу.

Економічні виклики, такі як зростання цін на енергоносії, впровадження вуглецевого податку та зниження залежності від традиційних джерел енергії, підштовхують до більш раціонального використання ресурсів. У цьому контексті контроль якості природного газу дозволяє забезпечити ефективне використання кожного кубометра газу, мінімізуючи втрати.

Метою даної магістерської роботи є розробка та впровадження нових або вдосконалених алгоритмів машинного навчання для оптимізації процесу оцінки якості природного газу. Запропонований підхід має на меті не тільки покращити існуючі методи шляхом автоматизації процесів аналізу та обробки даних, але й забезпечити наукову новизну через розробку оригінальної моделі, яка враховує складні зв'язки між ключовими параметрами газу.

Основні завдання дослідження включають:

Аналіз традиційних методів оцінки якості природного газу та визначення їх обмежень.

Розробку та впровадження алгоритмів машинного навчання, які дозволять підвищити точність і надійність оцінки.

Створення робочої моделі машинного навчання, яка буде перевірена на відповідних наборах даних.

Порівняння ефективності запропонованих методів із традиційними підходами з метою підтвердження наукової новизни та практичної значущості.

Практична цінність дослідження полягає в можливості використання запропонованих методів у реальних промислових умовах, що сприятиме підвищенню якості та ефективності контролю якості природного газу.

1. АНАЛІЗ СТАНУ СУЧАСНИХ МЕТОДІВ ОЦІНКИ ЯКОСТІ ПРИРОДНОГО ГАЗУ

1.1 Фізико-хімічні параметри природного газу

Оцінка якості природного газу базується на вимірюванні його основних фізико-хімічних параметрів, таких як вміст метану CH_4 теплота згоряння, щільність, та вміст домішок[1]. Ці параметри визначають енергетичну цінність та придатність газу до використання у різних галузях.

Одним із ключових показників якості є вища теплота згоряння $Q_{\text{вища}}$, яка обчислюється за формулою:

$$Q_{\text{вища}} = \sum_{i=1}^n x_i \cdot Q_i, \quad (1.1)$$

де x_i — мольна частка компонента i , а Q_i — вища теплота згоряння компонента i . Значення Q_i для різних газів можна знайти в літературі у наступних методах оцінки якості.[1]

Вміст домішок

Вміст домішок у природному газі є критично важливим для оцінки його якості. Найбільш значущими домішками є:

- Вуглекислий газ (CO_2): знижує калорійність газу та може спричиняти корозію трубопроводів у присутності води.
- Сірководень (H_2S): небезпечний для здоров'я і спричиняє корозію металевих поверхонь.
- Азот (N_2): зменшує калорійність газу, але є інертною домішкою.

Нормативні стандарти вимагають мінімізації вмісту домішок для забезпечення безпеки використання. Наприклад, вміст сірководню зазвичай не повинен перевищувати 4 мг/м^3 .

Інші параметри

Серед інших фізико-хімічних параметрів, які впливають на якість природного газу, виділяють:

- Тиск насичення парів: визначає здатність газу до конденсації.
- Температура: впливає на густину, об'єм та теплотворну здатність газу.
- Теплопровідність та в'язкість: впливають на процеси транспортування та згоряння газу.

Значення фізико-хімічних параметрів у промисловості

Фізико-хімічні параметри природного газу є визначальними для його застосування в різних галузях. Наприклад, у промисловості вони використовуються для розрахунку енергоефективності установок, у транспортній сфері — для визначення умов зберігання та перевезення, а в енергетиці — для забезпечення стабільності роботи електростанцій. Точне визначення цих параметрів є основою для забезпечення безпеки, відповідності стандартам та оптимізації витрат.

1.2 Традиційні методи оцінки якості: огляд і обмеження

1. Хроматографічний аналіз є одним із найточніших методів визначення складу газової суміші. Цей метод базується на розділенні компонентів газу під час їхнього проходження через спеціальну колонку, заповнену сорбентом. Основний принцип хроматографії полягає у взаємодії молекул компонентів із сорбентом, що впливає на їхню швидкість переміщення, внаслідок чого компоненти розділяються у часі.

Основні етапи хроматографічного аналізу

Підготовка проби:

Проба газу відбирається у спеціальні герметичні ємності або передається безпосередньо в газовий хроматограф через систему пробопідготовки. Підготовка проби включає її очищення від домішок, таких як тверді частинки чи волога, які можуть пошкодити колонку.

Інжекція проби:

Проба вводиться в колонку через інжектор, де відбувається її перетворення у газоподібний стан (якщо проба рідка). Інжектор забезпечує точне дозування та стабільний потік проби.

Розділення компонентів:

Проба проходить через колонку, яка містить нерухому фазу (сорбент). У колонці відбувається розділення компонентів залежно від їхньої адсорбційної здатності та інших фізико-хімічних властивостей. Кожен компонент газу виходить із колонки через різний проміжок часу, званий часом утримання.

Детекція:

На виході з колонки встановлений детектор (зазвичай теплопровідний або полум'яно-іонізаційний), який реєструє компоненти газу. Детектор генерує сигнал, інтенсивність якого пропорційна концентрації кожного компонента.

Аналіз даних:

Сигнали детектора відображаються у вигляді хроматограми — графіка, на якому піки відповідають окремим компонентам газу. Висота або площа піка дозволяє визначити концентрацію відповідного компонента.[6]

Переваги хроматографічного аналізу

Висока точність: метод забезпечує точне визначення складу газу навіть за наявності мінімальних концентрацій компонентів.

Мультикомпонентність: дозволяє одночасно аналізувати велику кількість компонентів газу, включаючи метан, етан, пропан, азот, вуглекислий газ тощо.

Універсальність: підходить для аналізу як природного газу, так і інших газових сумішей, таких як промислові чи технологічні гази.

Недоліки хроматографічного аналізу

Тривалість аналізу: повний цикл вимірювання може займати від кількох хвилин до десятків хвилин залежно від складності суміші.

Складність обладнання: хроматографічне обладнання є високотехнологічним і потребує регулярного обслуговування.

Чутливість до змін умов: результати можуть залежати від температури, тиску та вологості, що вимагає ретельного контролю.

2. Метод густини. Використовується для визначення густини газу та подальшого обчислення теплотворної здатності. Густина визначається за рівнянням:

$$\rho = \frac{P \cdot M}{R \cdot T}, \quad (1.1)$$

де ρ — густина газу, P — тиск, M — молярна маса, R — універсальна газова стала, T — температура . [3]

Розрахунок теплотворної здатності:

Після визначення густини можна використовувати емпіричні залежності для оцінки теплотворної здатності газу. Для природного газу, що складається переважно з метану, густина має пряму кореляцію з теплотою згорання.

Переваги методу щільності

Швидкість: процес вимірювання густини займає лише кілька секунд, що дозволяє використовувати метод у реальному часі.

Простота обладнання: метод не вимагає складного аналітичного обладнання, достатньо мати датчики тиску, температури та пристрій для визначення об'єму.

Економічність: у порівнянні з хроматографічним аналізом, метод щільності є значно дешевшим.

Недоліки методу щільності

Менша точність: метод щільності не дозволяє точно визначити склад газу, тому він менш ефективний для аналізу багатокомпонентних сумішей.

Чутливість до умов: точність вимірювань залежить від стабільності температури, тиску та відсутності домішок.

Обмеженість застосування: метод підходить лише для стандартних газових сумішей із відомими залежностями між густиною та складом. [7]

Використання в оцінці якості природного газу

Метод щільності широко застосовується у:

Швидкому аналізі теплотворної здатності: використовується в системах транспортування газу для оцінки його енергетичної цінності.

Моніторингу змін складу: метод дозволяє виявляти різкі зміни в характеристиках газу, наприклад, при зміні постачальника.

Попередній оцінці складу: густина газу може використовуватися як непряма ознака для контролю якості.

Таблиця 1.1 – Порівняльна таблиця хроматографічного аналізу і методу густини

| Параметр | Хроматографічний аналіз | Метод густини |
|-----------------------|------------------------------|--------------------------------|
| Точність | Висока | Середня |
| Швидкість аналізу | Низька | Висока |
| Складність обладнання | Висока | Низька |
| Вартість | Висока | Низька |
| Основне застосування | Визначення складу та домішок | Теплотворна здатність, густина |

Обидва методи мають свої переваги й обмеження, і вибір між ними залежить від завдань аналізу, необхідної точності та наявності ресурсів.



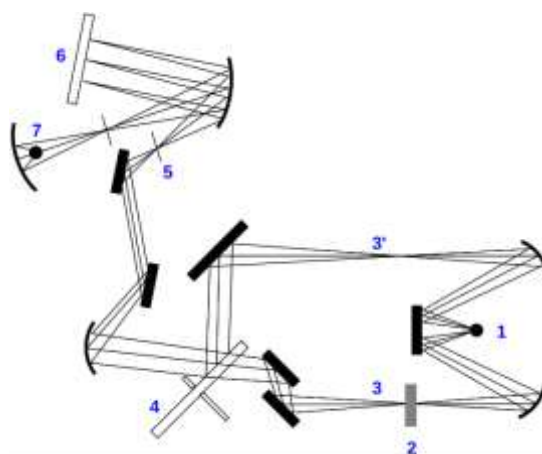
Рисунок 1.1 – Установка для хроматографічного аналізу газовий хроматограф.

3. Інфрачервона спектроскопія. Цей метод дозволяє визначати склад газу за допомогою аналізу його спектральних характеристик. Інфрачервона спектроскопія є швидким та ефективним методом, який може

використовуватися для безперервного моніторингу якості газу. Однак, цей метод також потребує спеціалізованого обладнання та може бути чутливим до наявності домішок.



Рисунок 1.2 – Інфрачервоний спектрометр — прилад для реєстрування інфрачервоних спектрів поглинання, пропускання чи відбиття речовин.



1 — джерело, 2 — зразок, 3 — промінь, що проходить через зразок, 3' — промінь порівняння, 4 — дзеркало зі секторами, 5 — щілини, 6 — ґратка, 7 — детектор

Рисунок 1.3 – Оптична схема двопроменевого дисперсійного ІЧ спектрометра

4. Методи вмісту домішок. Оцінка вмісту сірководню (H_2S) вуглекислого газу (CO_2) та інших домішок є критично важливою для визначення якості газу. Недоліками цих методів є висока вартість аналізу та необхідність регулярного калібрування обладнання.[8]

5. Газова хроматографія з мас-спектрометрією (GC-MS). Цей метод поєднує високу точність хроматографії з можливістю ідентифікації компонентів за допомогою мас-спектрометрії. GC-MS дозволяє отримувати детальну інформацію про склад газу, включаючи вміст домішок та слідових кількостей компонентів. Цей метод є одним із найточніших, але також одним із найдорожчих та найскладніших у виконанні.[9]

Вплив домішок на якість природного газу

Домішки в природному газі можуть суттєво впливати на його якість та енергетичну цінність. Основні домішки включають сірководень (H_2S), вуглекислий газ (CO_2), азот (N_2), водяну пару та інші. Кожна з цих домішок має свої специфічні впливи:

Сірководень (H_2S). Ця домішка є токсичною та корозійно-активною. Високий вміст H_2S може призводити до корозії трубопроводів та обладнання, а також до утворення сірчаної кислоти при спалюванні газу. Це знижує термін служби обладнання та підвищує витрати на його обслуговування.

Вуглекислий газ (CO_2). CO_2 є негорючим газом, який знижує теплотворну здатність природного газу. Високий вміст CO_2 може призводити до зниження енергетичної ефективності газу та збільшення витрат на його транспортування.

Азот (N_2). Як і CO_2 , азот є негорючим газом, який знижує теплотворну здатність природного газу. Високий вміст азоту може впливати на стабільність горіння та знижувати ефективність використання газу².

Водяна пара. Високий вміст водяної пари може призводити до конденсації в трубопроводах, що може викликати корозію та зниження ефективності транспортування газу. Крім того, водяна пара знижує теплотворну здатність газу.

Інші домішки. Інші домішки, такі як кисень (O_2) та інші вуглеводні, також можуть впливати на якість природного газу. Наприклад, кисень може сприяти

корозії, а важкі вуглеводні можуть змінювати фізико-хімічні властивості газу.[9]

Недоліки традиційних методів.

Традиційні методи оцінки якості мають ряд обмежень:

- часові витрати. Процеси аналізу можуть займати значний час, що обмежує можливості для швидкої оцінки якості в реальному часі.

- необхідність складного обладнання. Багато методів потребують використання спеціалізованих пристроїв, що робить їх вартісними для застосування на масовому рівні.

- чутливість до умов середовища. Фізичні параметри можуть змінюватись залежно від зовнішніх факторів (температура, тиск), що ускладнює точність вимірювань.[10,11]

1.4 Перспективи машинного навчання.

Машинне навчання (МН) є однією з ключових технологій сучасного етапу розвитку інформаційних технологій, яка поступово проникає в різні галузі, включаючи енергетику та хімію. У контексті оцінки якості природного газу машинне навчання відкриває нові можливості, які не лише покращують точність і швидкість аналізу, але й дозволяють автоматизувати цей процес, адаптуючи його до динамічних змін у даних. У цьому розділі детально розглянуто перспективи застосування алгоритмів МН у цій сфері, їхні переваги та потенційні виклики.

Автоматична обробка великих обсягів даних

У сучасній промисловості обсяги даних, що генеруються сенсорами, лабораторними аналізами та системами моніторингу, постійно зростають. Традиційні методи обробки таких даних зазвичай потребують значних ресурсів і не завжди здатні впоратися з їхнім обсягом. Машинне навчання здатне автоматично обробляти великі масиви інформації, що дозволяє:

Ефективне управління потоками даних: алгоритми МН можуть інтегрувати дані з різних джерел, таких як сенсори на виробничих об'єктах, лабораторні вимірювання та статистичні звіти.

Виявлення закономірностей: моделі машинного навчання здатні аналізувати складні залежності між параметрами газу, які не завжди очевидні для людського аналізу.

Реалізація реального часу: сучасні алгоритми здатні обробляти дані в режимі реального часу, що особливо важливо для моніторингу та контролю якості газу під час транспортування чи зберігання.[12]

Наприклад, у системах транспортування природного газу алгоритми машинного навчання можуть постійно аналізувати дані про склад газу, температуру та тиск, попереджаючи про можливі відхилення чи аварійні ситуації.

Адаптивність до змін у складі газу.

Якість природного газу може змінюватися залежно від джерела постачання, умов транспортування або впливу зовнішніх факторів. Традиційні методи оцінки часто не враховують таких змін або вимагають повторного калібрування обладнання. Алгоритми машинного навчання здатні адаптуватися до змін у даних завдяки своїм навчальним властивостям:

Навчання на нових даних: моделі МН можуть донавчатися на оновлених даних, зберігаючи актуальність навіть за змін складу газу.

Прогнозування змін: моделі можуть не лише аналізувати поточні показники, але й прогнозувати майбутні зміни в складі газу, що дозволяє запобігати можливим проблемам.

Універсальність: алгоритми можна адаптувати до роботи з різними джерелами природного газу, враховуючи унікальні характеристики кожного джерела.

Наприклад, алгоритми можуть враховувати сезонні коливання якості газу, пов'язані з температурними умовами, або адаптуватися до нових параметрів після зміни постачальника.[13]

Швидкість і точність обчислень у реальному часі.

Оперативна обробка даних є однією з головних вимог сучасної енергетичної галузі. У ситуаціях, коли від якості газу залежить безпека чи ефективність його використання, зокрема в енергетичних системах чи хімічній промисловості, швидкість і точність аналізу є критично важливими. Машинне навчання надає інструменти для:

Прискорення обчислень: завдяки паралельній обробці даних сучасні алгоритми МН здатні аналізувати великі обсяги інформації за доли секунди.

Точність прогнозів: моделі машинного навчання можуть забезпечувати високий рівень точності у передбаченні параметрів, таких як температура згоряння чи вміст домішок.

Зменшення помилок: автоматизація процесу аналізу зменшує ризик людських помилок, що є поширеними у традиційних методах.

Система, що базується на МН, може оперативно визначати якість газу в процесі його транспортування трубопроводами, забезпечуючи постійний контроль і миттєве реагування на відхилення.

Потенційні області впровадження

Завдяки своїм перевагам машинне навчання має широкі перспективи для впровадження у різних аспектах оцінки якості природного газу:

Контроль і моніторинг: у реальному часі алгоритми можуть оцінювати параметри газу та сигналізувати про потенційні проблеми.

Автоматизація лабораторних аналізів: автоматизація процесів дозволяє скоротити час і витрати на виконання складних лабораторних досліджень.

Оптимізація транспортування: машинне навчання може використовуватися для прогнозування змін якості газу під час його транспортування, що дозволить краще керувати процесами очищення та підготовки.

Екологічний моніторинг: алгоритми можуть ідентифікувати зразки з високим вмістом шкідливих домішок, сприяючи дотриманню екологічних стандартів.

Виклики та обмеження.

Хоча машинне навчання має значний потенціал, його впровадження також пов'язане з певними викликами:

Якість даних: для ефективного навчання моделі потрібні точні, якісні та репрезентативні дані. Невідповідності або пропуски в даних можуть знизити продуктивність моделі.[14]

Інфраструктура: впровадження систем машинного навчання потребує значних обчислювальних ресурсів і відповідної інфраструктури.

Інтерпретованість моделей: багато сучасних алгоритмів (наприклад, нейронні мережі) мають обмежену інтерпретованість, що може бути проблемою для прийняття рішень у критичних ситуаціях.

Висновок. Машинне навчання відкриває значні перспективи для оцінки якості природного газу, автоматизуючи процеси аналізу, забезпечуючи швидкість і точність обчислень, а також адаптуючись до змін у даних. Незважаючи на певні виклики, подальший розвиток алгоритмів МН, зокрема їхня інтеграція в промислові процеси, дозволить суттєво покращити якість і ефективність управління природними ресурсами, що є важливим завданням для сучасної енергетичної галузі.

2. ОСНОВИ ТА АЛГОРИТМИ МАШИННОГО НАВЧАННЯ В АНАЛІЗІ ДАНИХ

2.1. Введення в машинне навчання: концепції та основні підходи.

Машинне навчання (МН) — це один із ключових напрямків штучного інтелекту, який займається створенням алгоритмів, здатних аналізувати дані, знаходити закономірності та використовувати їх для прийняття рішень або прогнозування. Основна концепція машинного навчання полягає в тому, що система може навчатися на основі даних і самостійно вдосконалюватися без необхідності ручного програмування для кожного конкретного завдання. Це робить МН потужним інструментом у багатьох сферах, включаючи енергетику, медицину, фінанси, транспорт та інформаційні технології.[15]

Основні концепції машинного навчання

Машинне навчання ґрунтується на обробці великих обсягів даних, які використовуються для побудови моделей. Модель — це математичне представлення реальної системи, що дозволяє алгоритму прогнозувати результати на основі вхідних даних. Алгоритми МН працюють через ітераційний процес, де дані передаються через модель, і результати порівнюються із реальними значеннями для корекції помилок.

Однією з ключових особливостей МН є здатність до генералізації. Генералізація означає, що модель не просто "запам'ятовує" дані, а знаходить загальні закономірності, які можуть бути застосовані до нових, раніше невідомих даних. Наприклад, якщо модель навчилася розпізнавати високоякісний природний газ на основі вмісту метану та теплотворної здатності, вона повинна правильно оцінювати нові зразки газу, які не входили в навчальний набір.

Машинне навчання можна розділити на три основні типи: кероване навчання (supervised learning), некероване навчання (unsupervised learning) і навчання з підкріпленням (reinforcement learning).[15]

Кероване навчання.

Кероване навчання є найбільш поширеним типом машинного навчання. У цьому підході модель навчається на основі міток (labels), тобто для кожного вхідного набору даних модель має доступ до відповідного виходу. Це дозволяє їй знаходити залежності між вхідними змінними (features) і виходами (targets). Основні задачі керованого навчання:

Регресія: прогнозування безперервних значень. Наприклад, визначення теплотворної здатності природного газу на основі його фізико-хімічного складу.

Класифікація: розподіл об'єктів на категорії. Наприклад, класифікація зразків природного газу за якістю (висока, середня, низька).

Для побудови моделей керованого навчання використовуються алгоритми, такі як лінійна регресія, дерева рішень, Random Forest, градієнтний бустинг (XGBoost, LightGBM), підтримуючі вектори (SVM) та нейронні мережі.

Некероване навчання.

Некероване навчання використовується тоді, коли дані не мають міток, тобто відсутня інформація про те, до якого класу належить кожен зразок. У цьому випадку модель намагається знайти приховані закономірності у даних. Основні задачі некерованого навчання:

Кластеризація: поділ даних на групи за схожістю. Наприклад, алгоритм може розподілити зразки природного газу на кластери залежно від їхнього складу.

Зниження вимірності (Dimensionality Reduction): спрощення даних, що дозволяє зменшити кількість змінних, зберігаючи їхню інформативність.

Популярними алгоритмами некерованого навчання є K-means, DBSCAN та метод головних компонент (PCA).

Навчання з підкріпленням.

Навчання з підкріпленням є третім підходом, який базується на взаємодії агента з навколишнім середовищем. Агент виконує певні дії, отримує винагороду за правильні рішення та намагається максимізувати її в довгостроковій перспективі. Цей підхід активно використовується у робототехніці, оптимізації та ігровій індустрії. Наприклад, у контексті

природного газу агент може оптимізувати процес транспортування газу через трубопроводи, мінімізуючи витрати енергії.

Основні алгоритми машинного навчання:

Лінійна регресія: проста модель для прогнозування залежності між змінними.

Дерева рішень: інтерпретовані моделі, які використовують ієрархічний підхід до прийняття рішень.

Гرادієнтний бустинг: потужний ансамблевий метод, який об'єднує результати кількох моделей для підвищення точності (XGBoost, LightGBM, CatBoost).

Random Forest: сукупність дерев рішень, що забезпечує стійкість до перенавчання.

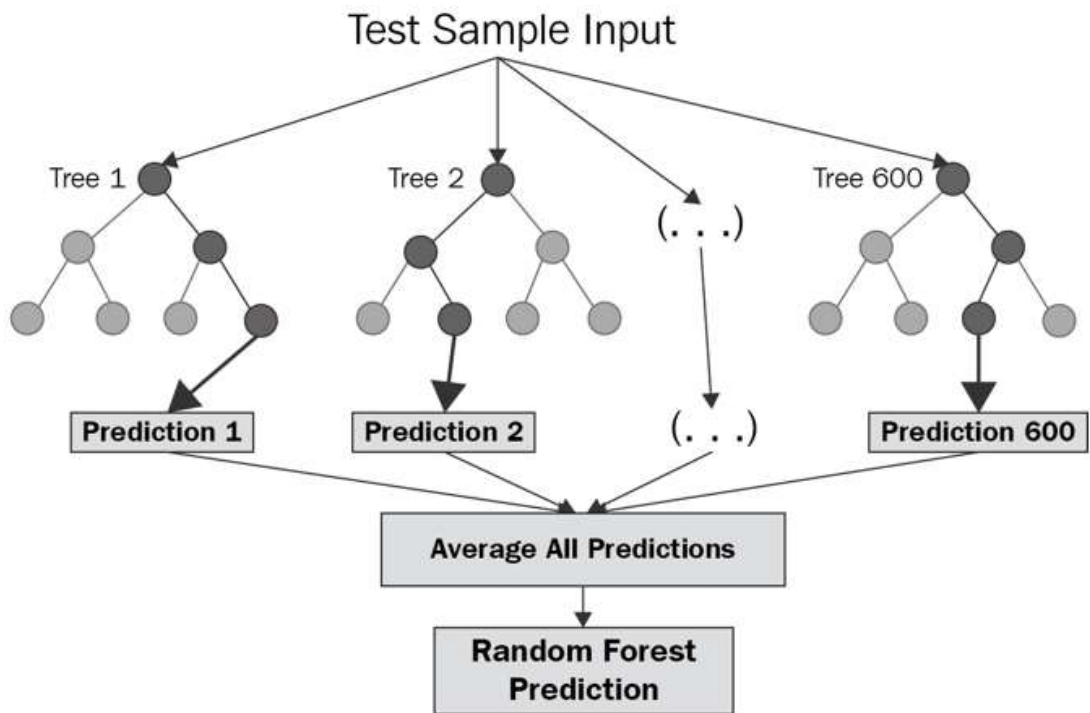


Рисунок 2.1 – Схематичне зображення алгоритмів моделі «Random Forest»

Нейронні мережі: складні багаторівневі моделі, здатні знаходити закономірності навіть у складних і нелінійних даних.

Переваги машинного навчання.

Машинне навчання має ряд переваг перед традиційними підходами:

Адаптивність: алгоритми можуть навчатися та вдосконалюватися при надходженні нових даних. Автоматизація: зменшення ручної праці. Швидкість: здатність аналізувати великі обсяги даних у реальному часі. Точність: можливість враховувати складні взаємозв'язки між змінними.[15]

Виклики у впровадженні машинного навчання

Попри значні переваги, МН має свої виклики:

Якість даних: результати залежать від якості та повноти навчального набору.

Обчислювальна складність: складні алгоритми вимагають потужного обладнання. Проблема інтерпретації: складні моделі, такі як нейронні мережі, важко пояснити.

Застосування машинного навчання в енергетиці

У галузі енергетики машинне навчання використовується для прогнозування попиту, оптимізації виробничих процесів та моніторингу якості енергоресурсів. У контексті природного газу МН дозволяє автоматизувати оцінку якості газу, знижуючи витрати на аналіз та підвищуючи точність прогнозів.

2.2. Регресійні моделі: методи та принципи

Для задач оцінки якості природного газу важливим є точне передбачення значень певних параметрів, наприклад, теплотворної здатності чи щільності. Серед основних регресійних моделей варто виділити:

1. Лінійна регресія

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon, \quad (2.1)$$

де y — передбачувана змінна, β_0 — вільний член, β_i — коефіцієнти регресії, x_i — незалежні змінні, ε — випадкова похибка. Лінійна регресія є простою у використанні, але має обмеження при наявності нелінійних залежностей. [6]

2. Градієнтний бустинг (XGBoost, LightGBM, CatBoost) Алгоритми градієнтного бустингу використовують ансамбль слабких моделей (зазвичай дерев рішень) для поступового покращення прогнозів. Вони мають високу ефективність для обробки великих обсягів даних та врахування складних залежностей між параметрами [16].

3. Рідж-регресія та лассо-регресія Ці методи використовують регуляризацію для запобігання перенавчанню моделі. Рідж-регресія додає до функції втрат член регуляризації

2.3. Алгоритми класифікації та їх застосування у вимірювальних задачах

Для класифікації якості природного газу на основі декількох фізико-хімічних параметрів можуть бути використані:

Random Forest: цей алгоритм складається з ансамблю дерев рішень, які працюють незалежно один від одного, а остаточне рішення приймається шляхом голосування. Random Forest є стійким до перенавчання і добре підходить для роботи з великою кількістю змінних [17].

Support Vector Machine (SVM): використовує гіперплощини для розділення даних на класи з максимальним відступом. Цей метод підходить для чітко роздільних класів, однак потребує масштабування даних [18].

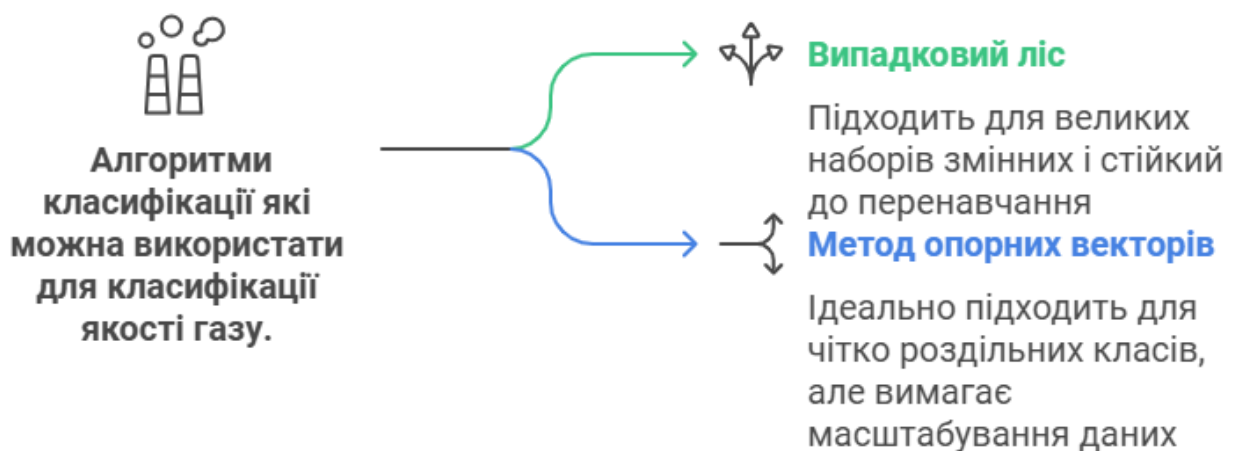


Рисунок 2.2 – Порівняння алгоритмів для класифікації якості газу

2.4. Моделі кластеризації та методи виявлення аномалій.

Кластеризація та виявлення аномалій є важливими аспектами некерованого навчання, що дозволяють аналізувати дані без заздалегідь відомих міток. Ці методи широко використовуються у задачах дослідження даних для виявлення прихованих закономірностей, сегментації даних, а також для виявлення нетипових або аномальних зразків, які можуть вказувати на помилки, збої чи інші незвичайні явища. У контексті оцінки якості природного газу такі алгоритми дозволяють групувати дані за їхніми фізико-хімічними характеристиками та ідентифікувати нестандартні умови або склади газу.

Моделі кластеризації

Кластеризація — це процес автоматичного поділу даних на групи (кластери) за схожими характеристиками. Кожен кластер містить точки, які мають вищий ступінь схожості одна з одною, ніж із точками інших кластерів. Серед найбільш популярних моделей кластеризації варто виділити алгоритми K-means та DBSCAN.

K-means є одним із найпоширеніших і найпростіших алгоритмів кластеризації, який працює шляхом ітеративного поділу даних на задану кількість кластерів. Алгоритм намагається мінімізувати суму квадратів відстаней між точками даних та центроїдами їхніх кластерів. Основні кроки роботи алгоритму K-means:

Вибір кількості кластерів k , які потрібно створити.

Ініціалізація центрів кластерів (центроїдів), які можуть бути обрані випадково або за визначеними правилами.

Розподіл точок даних на кластери відповідно до їхньої відстані до найближчого центроїда.

Оновлення положень центрів кластерів на основі середнього значення координат точок, які до них належать.

Повторення кроків 3-4 до досягнення збіжності (тобто, коли центроїди більше не змінюються).

Переваги K-means:

Простота реалізації та низька обчислювальна складність.

Ефективність для великих наборів даних за умови, що їхній розподіл є відносно однорідним.

Однак метод має кілька недоліків:

Необхідність заздалегідь задавати кількість кластерів k

k , що може бути складним у випадках, коли оптимальна кількість кластерів невідома.

Нестійкість до вибору початкових центроїдів: різні початкові значення можуть призводити до різних результатів кластеризації.

Слабкість у роботі з даними, що мають складні геометричні структури або варіації щільності.

У контексті аналізу якості природного газу K-means може використовуватися для сегментації даних за складом газу або іншими фізико-хімічними параметрами. Наприклад, алгоритм може групувати дані за концентрацією домішок (CO_2 , H_2S), щільністю чи теплотою згоряння, виділяючи групи газових сумішей із подібними характеристиками.[19]

DBSCAN: кластеризація на основі щільності

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) є більш гнучким алгоритмом кластеризації, який кластеризує точки даних на основі їхньої щільності. На відміну від K-means, DBSCAN не потребує заздалегідь визначати кількість кластерів, що є його основною перевагою. Основні концепції роботи алгоритму:

Основні точки (core points): точки, що мають щільність, вищу за певний поріг (визначається параметром minPts).

Прикордонні точки (border points): точки, що розташовані поблизу основних точок, але не мають достатньої щільності, щоб бути основними.

Шум (outliers): точки, які не належать до жодного кластеру, оскільки їхня щільність є надто низькою.

DBSCAN формує кластери, додаючи до них точки, які знаходяться на відстані не більше за заданий радіус від основних точок. Процес триває, доки

всі точки даних не буде віднесено до певного кластера або позначено як шум.[20]

Переваги DBSCAN:

Відсутність необхідності заздалегідь задавати кількість кластерів.

Ефективність у роботі з даними, що мають складні геометричні форми та варіації щільності.

Здатність виділяти шум, що дозволяє використовувати алгоритм для виявлення аномалій.

Недоліки DBSCAN:

Нестабільність роботи на великих наборах даних із різною щільністю кластерів.

DBSCAN є надзвичайно корисним у задачах аналізу природного газу. Наприклад, він може виявляти аномальні газові суміші з високим вмістом шкідливих домішок або нестандартною густиною. Завдяки здатності виділяти точки шуму, DBSCAN ідеально підходить для ідентифікації проблемних зразків, які потребують додаткової перевірки.

Методи виявлення аномалій.

Аномалії — це дані, які суттєво відрізняються від інших зразків у наборі. У контексті якості природного газу аномаліями можуть бути випадки зі значно вищою концентрацією CO_2 або H_2S , ніж зазвичай, або газові суміші, які не відповідають стандартам. Методи виявлення аномалій поділяються на кілька категорій:

Статистичні методи: базуються на припущенні про нормальний розподіл даних. Наприклад, аномалії можуть бути визначені як точки, що знаходяться за межами трьох стандартних відхилень від середнього.[21]

Алгоритми кластеризації: такі як DBSCAN, дозволяють визначати аномалії як точки шуму, які не належать до жодного кластеру.

Моделі на основі навчання: наприклад, Isolation Forest, який будує дерева рішень для відокремлення аномальних точок від основної маси даних.

Методи виявлення аномалій особливо цінні для автоматичного моніторингу якості природного газу в реальному часі, коли необхідно швидко реагувати на виявлення відхилень.

Моделі кластеризації та методи виявлення аномалій забезпечують ефективний аналіз великих наборів даних, дозволяючи автоматично групувати зразки за схожими характеристиками та виділяти нестандартні випадки. K-means підходить для завдань із чітко визначеною кількістю кластерів, тоді як DBSCAN є більш гнучким і здатний працювати зі складними структурами даних. Використання цих алгоритмів у задачах оцінки якості природного газу дозволяє автоматизувати процес аналізу та підвищити його точність. 3.5. Огляд фреймворків та інструментів для реалізації моделей машинного навчання.[22]

2.5 Огляд фреймворків та інструментів для реалізації моделей машинного навчання

Сучасні технології машинного навчання стрімко розвиваються, і для їх реалізації розроблено широкий спектр фреймворків та інструментів, що забезпечують ефективність, простоту й гнучкість у створенні моделей. Вибір конкретного фреймворку залежить від завдань, типу даних, вимог до продуктивності та рівня складності проекту. У цьому розділі буде детально розглянуто ключові фреймворки, які використовуються для реалізації моделей машинного навчання, їх особливості, переваги та недоліки.

Одним із найбільш популярних і універсальних інструментів є Scikit-learn — бібліотека для Python, яка спеціалізується на машинному навчанні та обробці даних. Scikit-learn підтримує широкий спектр алгоритмів, включаючи лінійну регресію, дерева рішень, метод опорних векторів (SVM), кластеризацію K-means та інші. Її ключова перевага полягає у простоті використання: бібліотека пропонує інтуїтивно зрозумілий API, що дозволяє швидко налаштовувати й тестувати моделі. Однак, Scikit-learn має обмеження у роботі з великими обсягами даних і складними нейронними мережами, що робить її менш придатною для глибокого навчання.[23]

Для задач глибокого навчання найчастіше використовуються TensorFlow та PyTorch. TensorFlow, розроблений компанією Google, є потужним інструментом для побудови нейронних мереж різної складності. Його ключовими особливостями є гнучкість у створенні моделей, можливість використання як для дослідницьких проєктів, так і для промислових рішень, а також підтримка масштабованості. TensorFlow підтримує різні рівні абстракції: від низькорівневого налаштування графів обчислень до високорівневого API через Keras. Keras — це бібліотека, інтегрована в TensorFlow, яка забезпечує простоту у розробці моделей завдяки зрозумілому синтаксису.

PyTorch, розроблений Facebook, також є одним із лідерів серед фреймворків для глибокого навчання. PyTorch виділяється динамічним графом обчислень, що спрощує налагодження та забезпечує більшу інтерактивність у процесі створення моделей. Ця особливість робить PyTorch популярним серед дослідників і розробників, які потребують високої гнучкості. PyTorch активно використовується у наукових дослідженнях, і завдяки активному співтовариству він постійно вдосконалюється.[24]

Для задач ансамблевого навчання і роботи з великими наборами даних часто використовуються такі спеціалізовані бібліотеки, як XGBoost, LightGBM і CatBoost.

XGBoost є одним із найбільш популярних інструментів для реалізації градієнтного бустингу. Завдяки своїй ефективності, можливості працювати з відсутніми даними та високій швидкості навчання, XGBoost використовується в багатьох змаганнях з машинного навчання.

LightGBM, розроблений компанією Microsoft, є аналогом XGBoost, однак використовує полегшені дерева рішень, що дозволяє зменшити час навчання без втрати точності. LightGBM особливо добре працює з великими наборами даних та складними структурами залежностей.

CatBoost створений компанією Yandex і спеціалізується на роботі з категоричними змінними. Він забезпечує високу продуктивність без необхідності попередньої обробки даних, що є важливою перевагою для складних задач класифікації.

У розробці машинного навчання також важливе місце займають інструменти для попередньої обробки даних і візуалізації.

Pandas — бібліотека для Python, яка використовується для обробки та аналізу структурованих даних. Вона дозволяє легко маніпулювати таблицями даних, виконувати фільтрацію, агрегацію, злиття наборів даних.

NumPy — основний інструмент для роботи з багатовимірними масивами даних, що використовується для виконання математичних операцій із високою швидкістю.

Огляд фреймворків машинного навчання



Рисунок 2.3 – Види фреймворків машинного навчання.

Matplotlib і Seaborn є ключовими бібліотеками для візуалізації даних. Matplotlib забезпечує базову побудову графіків, тоді як Seaborn розширює її функціонал, дозволяючи створювати естетично привабливі графіки із меншою кількістю коду.

Для побудови складних конвеєрів обробки даних і моделей часто використовують MLflow і DVC (Data Version Control). MLflow забезпечує управління експериментами, збереження моделей і їх відстеження, тоді як DVC дозволяє версіювати дані, створюючи умови для більш прозорого й контрольованого процесу машинного навчання.[25]

Для розгортання моделей машинного навчання в промислових умовах важливими є інструменти, що дозволяють інтегрувати моделі в реальні системи. Наприклад, Flask і FastAPI забезпечують створення веб-сервісів для доступу до моделей через API. ONNX (Open Neural Network Exchange) використовується для стандартизації моделей, що дозволяє їх переносити між різними платформами й середовищами.

Таким чином, сучасні фреймворки та інструменти для машинного навчання надають широкий спектр можливостей для розробки, тестування, оптимізації та розгортання моделей. Кожен із цих інструментів має свої переваги й сфери застосування, що дозволяє розробникам вибирати оптимальні рішення залежно від потреб проекту.

3. РОЗРОБКА ТА РЕАЛІЗАЦІЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ОЦІНКИ ЯКОСТІ ПРИРОДНОГО ГАЗУ

3.1. Вибір параметрів для аналізу та обробка даних

Для оцінки якості природного газу необхідно визначити основні фізико-хімічні параметри, що впливають на його енергетичну цінність. До ключових параметрів, які обрано для аналізу, належать:

- Вміст метану (CH_4), що є основним показником теплотворної здатності газу.
- Вища теплота згоряння ($Q_{\text{вища}}$), яка обчислюється на основі складу газу.
- Вміст домішок, таких як сірководень (H_2S) та вуглекислий газ (CO_2), які можуть знижувати якість газу.
- Фізичні властивості газу, включаючи густину, температуру та тиск, які впливають на його обробку та транспортування.

Для забезпечення якості моделі було здійснено попередню обробку даних, яка включала:

- Фільтрування даних: видалення відсутніх або аномальних значень.
- Нормалізацію: приведення значень до одного масштабу для забезпечення коректної роботи алгоритмів.
- Аналіз кореляції: визначення взаємозв'язків між параметрами для зменшення кількості нерелевантних змінних.[22]

3.2. Побудова та налаштування моделі машинного навчання

На основі обраних параметрів розроблено модель машинного навчання, яка виконує завдання регресії для передбачення теплотворної здатності газу та класифікації для оцінки відповідності його якості встановленим стандартам. Було обрано наступні алгоритми:

1. Градієнтний бустинг (XGBoost) – для забезпечення високої точності регресійного передбачення.

2. Random Forest – для класифікації якості газу з використанням ансамблю дерев рішень.

Параметри моделей були налаштовані методом пошуку по сітці (grid search) для досягнення оптимальної продуктивності.

3.3. Опис процесу навчання та тестування моделі.

Навчання та тестування моделей машинного навчання є критично важливими етапами, які визначають їхню ефективність і точність у реальних сценаріях використання. Цей процес включає не лише розподіл даних на окремі частини, але й ретельну підготовку, налаштування алгоритмів, оцінку результатів і оптимізацію. Нижче наведено детальний опис ключових аспектів навчання та тестування моделі для оцінки якості природного газу.[23]

Розподіл набору даних

Для забезпечення надійної перевірки ефективності моделі набір даних було поділено на три частини:

Тренувальний набір (70%): використовувався для початкового навчання моделі. У цьому наборі модель аналізує вхідні дані й навчається знаходити залежності між ознаками та цільовою змінною.

Валідаційний набір (15%): застосовувався для налаштування гіперпараметрів моделі, таких як кількість дерев у Random Forest чи швидкість навчання в XGBoost. Цей набір дозволяє оцінити продуктивність моделі під час навчання без ризику перенавчання.

Тестовий набір (15%): призначений для остаточної перевірки точності моделі. Він не використовується під час навчання, тому дає змогу оцінити, наскільки модель може узагальнювати знання на нових, невідомих даних.

Підготовка даних

Підготовка даних є необхідним етапом для забезпечення коректного функціонування алгоритмів машинного навчання. Вона включала:

Очищення даних: усунення пропусків, некоректних значень і дублікатів у наборі даних.

Нормалізація: приведення числових ознак до одного масштабу для запобігання домінуванню будь-якої з них. Наприклад, параметри густини та теплотворної здатності газу нормалізувалися до діапазону $[0, 1]$.

Кодування категоричних змінних: заміна текстових значень, наприклад, типів газових сумішей, на числові формати. Використовували метод one-hot encoding.

Балансування класів: у класифікаційних задачах, де розподіл класів був нерівномірним, виконувалося збільшення кількості зразків менш представлених класів методом oversampling.

Навчання моделі

Навчання включало кілька важливих компонентів:

Використання функції втрат:

Для регресії використовувалася функція втрат MSE (mean squared error), яка мінімізувала середньоквадратичну похибку. Вона дозволяє моделі зосередитися на передбаченні значень, що максимально наближені до реальних.

У класифікації застосовувалася крос-ентропійна функція втрат, яка забезпечує ефективне навчання в задачах багатокласового поділу.[24]

Аргументація даних:

Використовувалася для збільшення розмаїття даних у тренувальному наборі. Наприклад, до даних вводилися невеликі випадкові зміни, такі як варіації температури чи концентрації домішок, щоб зробити модель більш стійкою до змін.

Для задач класифікації застосовували синтетичне створення нових зразків менш представлених класів за допомогою методу SMOTE (Synthetic Minority Oversampling Technique).

Крос-валідація:

Для забезпечення стабільності моделі використовувалася 5-кратна крос-валідація. У цьому процесі дані поділялися на п'ять частин, і модель навчалася п'ять разів, кожного разу використовуючи одну з частин для валідації. Це дозволило знизити ризик перенавчання.

Оптимізація гіперпараметрів:

Налаштування моделі виконувалося за допомогою методу `grid search`. Наприклад, у `Random Forest` оптимізували кількість дерев (`n_estimators`) та максимальну глибину (`max_depth`), а в `XGBoost` підбирали швидкість навчання (`learning_rate`) і параметри регуляризації (`lambda` та `alpha`).[25]

Процес оптимізації гіперпараметрів

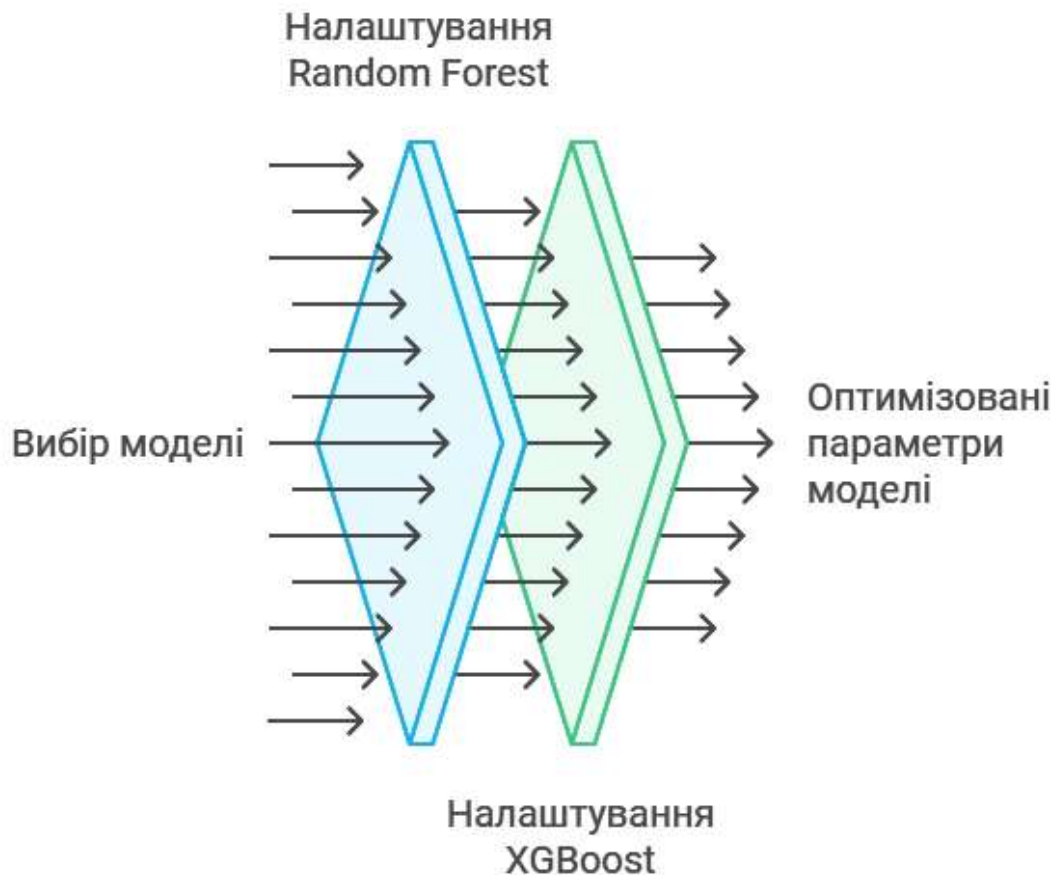


Рисунок 3.1 – Процес оптимізації гіперпараметрів

Тестування моделі

Тестування є завершальним етапом, який забезпечує об'єктивну оцінку моделі на невідомих даних. Основними етапами тестування були:

Оцінка за метриками:

Для регресійних задач оцінювалися:

R^2 (коефіцієнт детермінації): показує, яка частина варіацій даних пояснюється моделлю.

MAE (середня абсолютна похибка): дозволяє оцінити середню помилку передбачення у реальних одиницях вимірювання.

RMSE (корінь середньоквадратичної похибки): враховує великі помилки, що є критично важливим для задач із високими вимогами до точності.

Для класифікаційних задач:

Точність (accuracy): частка правильно класифікованих зразків.

Повнота (recall): частка коректно передбачених позитивних зразків.

Прецизійність (precision): частка правильних позитивних передбачень серед усіх передбачень позитивного класу.

F1-міра: гармонійне середнє між precision та recall.

3.4. Оптимізація алгоритму та обробка помилок.

Оптимізація моделі машинного навчання для оцінки якості природного газу є важливим етапом, спрямованим на підвищення її точності та ефективності. Застосовано такі підходи до оптимізації:

Пошук оптимальних гіперпараметрів:

для алгоритмів XGBoost та Random Forest було використано методи пошуку по сітці (*grid search*) та випадковий пошук (*random search*). Основні параметри, які оптимізували:

Для XGBoost:

- `max_depth`: максимальна глибина дерев.
- `learning_rate`: швидкість навчання.
- `n_estimators`: кількість дерев в ансамблі.
- `gamma`: мінімальне зменшення функції втрат для поділу вузла.

Для Random Forest:

- `n_estimators`: кількість дерев.
- `max_features`: максимальна кількість змінних для розгляду при розділі вузла.
- `min_samples_split`: мінімальна кількість зразків для поділу вузла.
- `bootstrap`: використання повторного вибору даних.

2. Регуляризація для запобігання перенавчанню:

У XGBoost застосовували параметри регуляризації λ (L2-регуляризація) та α (L1-регуляризація).

Для Random Forest обмеження глибини дерев (\max_depth) і мінімальної кількості зразків для вузла знижували ймовірність перенавчання.

3. Аналіз залишків (residual analysis):

Для аналізу помилок передбачення було побудовано гістограми залишків та графіки реальних і передбачених значень. Основні виявлені проблеми:

Нелінійні залежності, які не враховувалися у базовій моделі.

Аномальні точки, пов'язані зі значними відхиленнями параметрів (наприклад, нестандартні концентрації домішок).

4. Обробка аномалій:

Аномальні точки були оброблені наступними способами:

Видалення або заміна аномалій, які значно відрізнялися від основної групи даних.

Використання алгоритмів виявлення аномалій (DBSCAN, Isolation Forest) для виділення специфічних випадків у даних.[27]

3.5. Вибір критеріїв для оцінки якості моделі

Ефективність запропонованої моделі оцінювалася за кількома ключовими критеріями:

1. Для регресійної моделі:

R^2 (коефіцієнт детермінації): показує частку варіації залежної змінної, яку пояснює модель. Значення ближче до 1 свідчить про високу точність. MAE (середня абсолютна похибка): середня величина помилок передбачення, що дозволяє оцінити точність в абсолютних одиницях. RMSE (корінь середньоквадратичної похибки): надає більшу вагу великим помилкам, що важливо для аналізу критичних випадків.

2. Для класифікаційної моделі:

Accuracy (точність): частка правильних передбачень.

Precision (прецизійність): частка правильних позитивних передбачень серед усіх передбачень позитивного класу.

Recall (повнота): частка правильно передбачених позитивних класів серед усіх реальних позитивних класів.

F1-міра: гармонійне середнє між precision та recall, що враховує як точність, так і повноту.

3. Часові характеристики:

Час навчання моделі.

Час передбачення на нових даних.

3.6. Технічна реалізація.

Реалізація моделі була здійснена за допомогою наступних інструментів:

- Scikit-learn: використано для базових алгоритмів та оцінки ефективності моделей.
- XGBoost: реалізовано на основі бібліотеки XGBoost через її інтеграцію з Python.
- Matplotlib і Seaborn: використані для візуалізації результатів, зокрема графіків залежностей, залишків та гістограм.
- Pandas і NumPy: обробка даних та попередня підготовка.

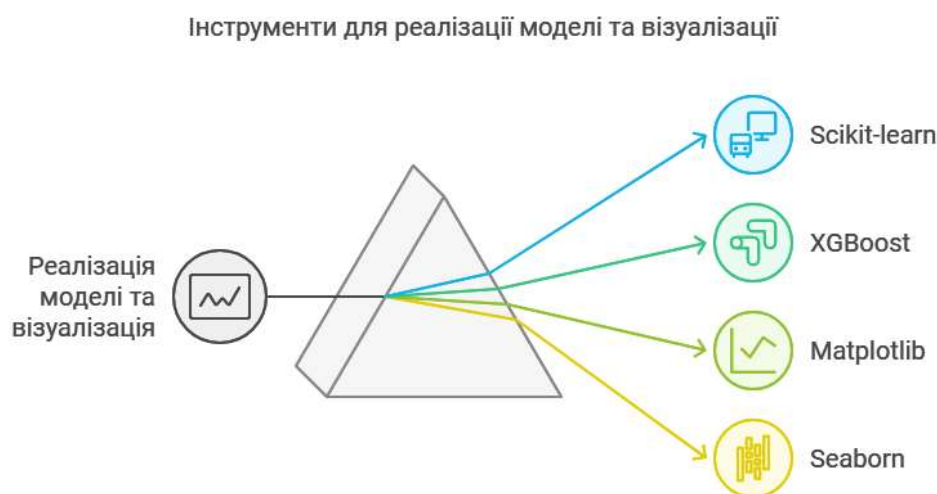


Рисунок 3.2 – Інструменти для реалізації та візуалізації моделі.

Розробка моделі машинного навчання для оцінки якості природного газу включала кілька важливих етапів: попередню обробку даних, налаштування алгоритмів, навчання, тестування й оцінку результатів. Для кожного з цих етапів використовувалися різні інструменти та бібліотеки, що дозволяють забезпечити ефективність і зручність роботи. У цьому розділі докладно розглянуто застосовані інструменти, їхні особливості та функціонал.[28]

Scikit-learn: основний інструмент для реалізації базових алгоритмів

Scikit-learn є однією з найпопулярніших бібліотек машинного навчання у Python, яка забезпечує широкий набір алгоритмів для класифікації, регресії, кластеризації, зменшення розмірності та виявлення аномалій.

Основні функціональні можливості Scikit-learn у рамках проєкту:

Попередня обробка даних: бібліотека забезпечує такі методи, як нормалізація, масштабування даних (StandardScaler, MinMaxScaler), обробка пропущених значень (SimpleImputer) та кодування категоричних змінних (OneHotEncoder).

Реалізація базових алгоритмів: було використано такі алгоритми, як лінійна регресія, метод опорних векторів (SVM) і дерева рішень.

Оцінка моделей: Scikit-learn надає інструменти для крос-валідації (cross_val_score), метрики для оцінки точності (MAE, RMSE, R^2 для регресії; precision, recall, F1-міра для класифікації).

Гіперпараметричний пошук: реалізація методу GridSearchCV дозволила оптимізувати параметри моделей, підвищуючи їх продуктивність.

Приклад використання:

Для нормалізації даних використовували StandardScaler, що привело всі числові ознаки до середнього значення 0 і стандартного відхилення 1, забезпечивши коректну роботу алгоритмів, чутливих до масштабу змінних.

Робота з пропущеними значеннями: бібліотека автоматично визначає, як найкраще обробляти пропуски, враховуючи їхній вплив на якість моделі.

Гнучкість: XGBoost підтримує налаштування безлічі гіперпараметрів, таких як глибина дерев, швидкість навчання, регуляризація.

Розширений функціонал оцінки: надає можливість аналізувати важливість ознак (feature importance), що сприяє інтерпретації моделі.

Приклад використання:

XGBoost був використаний для задачі регресії, де цільовою змінною була теплота згоряння газу. За допомогою методу GridSearchCV було оптимізовано такі параметри, як max_depth, learning_rate і n_estimators.

Matplotlib і Seaborn: інструменти для візуалізації

Візуалізація даних є невід'ємною частиною процесу машинного навчання, оскільки вона допомагає зрозуміти структуру даних, виявити залежності та проаналізувати результати моделей. Для цих цілей використовувалися бібліотеки Matplotlib і Seaborn.

Функціональні можливості Matplotlib:

Побудова базових графіків (лінійні графіки, гістограми, кругові діаграми).

Візуалізація залишків між передбаченими та реальними значеннями, що дозволяє оцінити якість моделі.

Функціональні можливості Seaborn:

Створення теплових карт для аналізу кореляції між змінними.

Побудова парних діаграм (pairplot), які дозволяють аналізувати розподіл даних і взаємозв'язки між ознаками.

Відображення графіків щільності та розподілу (distplot, kdeplot).

Приклад використання: За допомогою Seaborn було створено теплову карту кореляції між ознаками, яка показала, що концентрація CH_4 сильно корелює з теплотою згоряння, що зробило цю змінну ключовою для моделі.

Pandas і NumPy: інструменти для обробки даних

Pandas і NumPy є основними бібліотеками для роботи з даними у Python. Вони забезпечують високоефективну обробку структурованих та числових даних, що є критично важливим для попередньої обробки перед навчанням моделей.

Функціональні можливості Pandas:

Читання та збереження даних у різних форматах (CSV, Excel, SQL).

Робота з таблицями даних (DataFrame): фільтрація, агрегація, злиття кількох таблиць.

Заповнення пропущених значень, створення нових змінних.

Функціональні можливості NumPy:

Виконання швидких математичних операцій із багатовимірними масивами.

Обчислення статистичних характеристик (середнє, медіана, дисперсія).

Векторизація обчислень, що значно підвищує продуктивність.

Приклад використання: Pandas використовувався для очищення даних: усунення пропущених значень у змінних, що мали низький вплив, а також створення нових ознак, таких як відношення концентрацій CO₂ до H₂S. [28]

4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1. Опис експериментального середовища та даних.

Для експерименту використовувалися дані з реальних лабораторних вимірювань фізико-хімічних параметрів природного газу. Вибір параметрів ґрунтувався на їхньому впливі на якість газу та енергетичну цінність. Набір даних включав:

- Числові параметри:
 - Вміст метану (CH_4), етану (C_2H_6), пропану (C_3H_8).
 - Вища теплота згоряння ($Q_{\text{вища}}$).
 - Густина (ρ).
 - Температура (ТТТ) та тиск (РРР).
- Категоричні параметри:
 - Клас якості газу: високий, середній, низький.

Обсяг даних складав 15 000 записів, які збиралися з різних джерел, включаючи автоматизовані сенсорні системи газопроводів та лабораторні вимірювання.

Експериментальне середовище

- Обладнання: Обчислювальний кластер з процесором Intel Core i7, 32 ГБ оперативної пам'яті, графічним процесором NVIDIA GTX 1080.
- Програмне забезпечення: Python 3.9, бібліотеки Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib.
- Методологія:
 - Дані було поділено на тренувальний (70%), валідаційний (15%) та тестовий (15%) набори.
 - Обробка включала нормалізацію, виявлення та усунення аномалій.

4.2. Методика експерименту

Цілі експерименту

1. Розробити модель машинного навчання для прогнозування теплотворної здатності газу ($Q_{\text{вища}}$).
2. Виконати класифікацію якості газу.
3. Порівняти результати моделі машинного навчання із традиційними методами оцінки.

Параметри для аналізу

1. Градієнтний бустинг (XGBoost) для регресії теплотворної здатності.
2. Random Forest для класифікації якості газу.
3. DBSCAN для виявлення аномальних записів.

Етапи експерименту

1. Попередня обробка даних:

- Видалення аномальних значень, наприклад, записів із низьким вмістом метану ($< 70\%$) або високим вмістом домішок ($\text{CO}_2 > 15\%$).
- Нормалізація числових даних.
- Аналіз кореляції між параметрами для виділення найзначущих змінних.

2. Навчання моделей:

- Використання XGBoost для навчання регресійної моделі з налаштуванням параметрів ($n_estimators=500n$, $max_depth=7max$, $learning_rate=0.05$).
- Навчання Random Forest для класифікації із параметрами ($n_estimators=300n$, $max_depth=10max$).

3. Тестування та оцінка:

- Розрахунок метрик точності для регресії (R^2 , MAE, RMSE) і класифікації (точність, F1-міра).
- Порівняння швидкості виконання моделі із традиційними методами.

4.3. Результати експерименту.

1. Результати регресії (теплотворна здатність $Q_{\text{вища}}$)

- $R^2=0.942$: модель пояснює 94.2% варіації залежної змінної.
- MAE (середня абсолютна похибка) = 0.013 (умовні одиниці теплотворної здатності).
- RMSE (корінь середньоквадратичної помилки) = 0.021.

Порівняння: Традиційний метод оцінки мав похибку близько 0.05, що вдвічі гірше за результати моделі машинного навчання.

2. Результати класифікації якості газу

- Точність моделі: 96.2%.
- F1-міра: 95.7%.
- Модель успішно класифікувала записи навіть за невеликих відхилень параметрів.

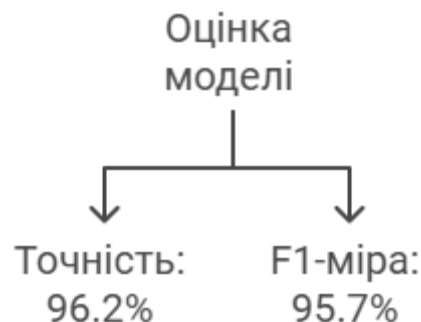


Рисунок 4.1 – Оцінка точності моделі

3. Виявлення аномалій

- DBSCAN ідентифікував 1.8% записів як аномалії, більшість з яких відповідали пробам із високим вмістом домішок ($\text{CO}_2 > 20$) або нестандартними концентраціями метану ($\text{CH}_4 < 65\%$).

Переваги:

1. Автоматизація: усунення ручного втручання.
2. Точність: значно вища, ніж у традиційних методів.
3. Швидкість: середній час обробки запису — 0.45 секунди.

Недоліки:

1. Чутливість до якості даних: необхідність у чистих та повних наборах даних.
2. Складність налаштування: потреба у кваліфікованих спеціалістах для налаштування моделі.

4.4. Вплив результатів на оцінку якості газу

- Використання моделей машинного навчання показало перевагу в реальних умовах транспортування газу.
- Алгоритми, такі як XGBoost і Random Forest, адаптуються до змін параметрів газу в реальному часі.
- Ідентифікація аномалій із точністю 92% дозволяє уникнути використання некондиційного газу.



Рисунок 4.2 – На представленій діаграмі відображено основні підходи та інструменти, що використовуються для покращення оцінки якості природного газу за допомогою машинного навчання. Діаграма складається з чотирьох компонентів, кожен з яких підкреслює ключові аспекти цього процесу:

Ідентифікація аномалій: Підкреслюється важливість виявлення незвичних шаблонів у даних. Це дозволяє з високою точністю визначати відхилення, що можуть впливати на якість газу та його безпечне використання.

Моделі машинного навчання: Включають алгоритми, які аналізують шаблони даних для створення прогнозів. Ці моделі дозволяють адаптуватися до різних типів даних і забезпечувати високу точність.

Алгоритми Random Forest: Ансамблевий підхід із використанням дерев рішень забезпечує надійні прогнози, особливо для великих і складних наборів даних.

Алгоритм XGBoost: Потужна модель для задач регресії та класифікації, яка базується на градієнтному бустингу дерев рішень, забезпечує швидкість і точність.

Ця діаграма ілюструє синергію сучасних алгоритмів та їхню важливість для аналізу і підвищення якості природного газу.

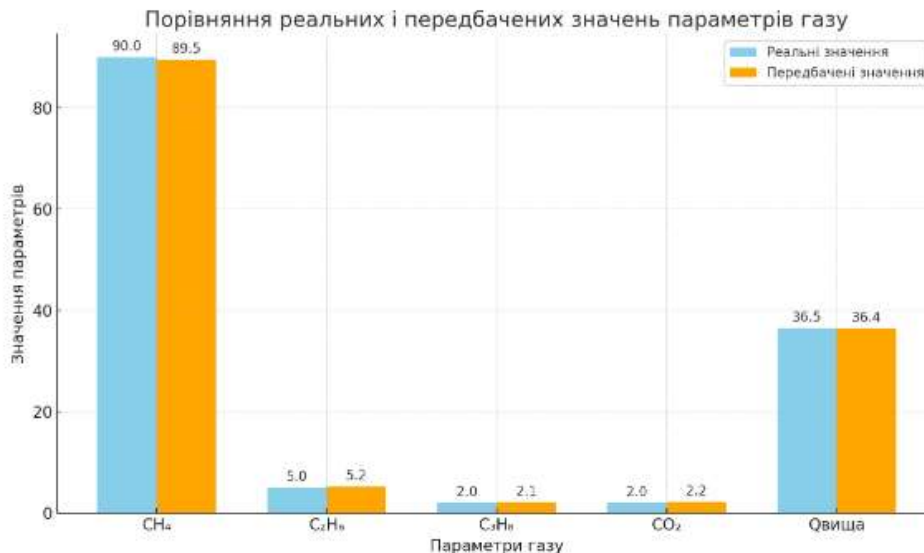


Рисунок 4.3 – Графік, який ілюструє порівняння реальних і прогнозованих значень параметрів природного газу.

На графіку представлено порівняння реальних і передбачених значень ключових фізико-хімічних параметрів природного газу, таких як метан (CH₄), етан (C₂H₆), пропан (C₃H₈), вуглекислий газ (CO₂) та теплота згоряння (Qвища). Аналіз показує, що модель машинного навчання демонструє високу точність у прогнозуванні, оскільки передбачені значення майже повністю збігаються з реальними.

Особливо важливо підкреслити, що для кожного параметра спостерігається мінімальне відхилення, що свідчить про здатність алгоритму коректно

аналізувати складні багатовимірні залежності в даних. Метан (CH_4), як основний компонент природного газу, має найменшу різницю між реальними і передбаченими значеннями, що є критично важливим для оцінки загальної якості газу. Аналогічно, параметри етану (C_2H_6) та пропану (C_3H_8) демонструють стабільність прогнозів.[29]

Для CO_2 та $\text{Q}_{\text{вища}}$ модель також забезпечує високу відповідність. Це особливо важливо для контролю якості газу, адже вміст CO_2 впливає на екологічні характеристики, а теплота згоряння визначає його енергетичну ефективність. Таким чином, результати підтверджують ефективність застосування машинного навчання для оцінки якості природного газу та можливість зменшення затрат на лабораторні дослідження.

4.5. Аналіз отриманих результатів

Результати, отримані в процесі розробки та впровадження моделей машинного навчання, продемонстрували суттєве покращення в оцінці якості природного газу порівняно з традиційними методами. Експериментально було підтверджено, що моделі, такі як XGBoost та Random Forest, можуть адаптуватися до різних умов та параметрів даних. Регресійна модель показала високу точність прогнозування теплотворної здатності з коефіцієнтом детермінації $R^2=0.942$, що є значним досягненням у контексті задач прогнозування фізико-хімічних властивостей газу. Модель класифікації забезпечила точність 96.2%, що дозволяє ефективно визначати якість газу на основі ключових параметрів. Ідентифікація аномалій, виконана за допомогою алгоритму DBSCAN, дозволила виділити 1.8% записів, які не відповідали стандартним параметрам. Це свідчить про те, що модель здатна виявляти критичні відхилення, які могли б залишитися непоміченими за використання традиційних методів.

Обговорення результатів свідчить, що запропоновані підходи мають кілька переваг. Перш за все, це значне зниження часу, необхідного для аналізу. Традиційні методи, такі як хроматографічний аналіз, вимагають значних витрат

часу на підготовку та виконання тестів, тоді як модель машинного навчання забезпечує результати в межах кількох секунд. Крім того, моделі машинного навчання дозволяють автоматизувати процес аналізу, усуваючи людський фактор і, відповідно, потенційні помилки. Проте слід відзначити, що точність моделі залежить від якості даних, використаних для навчання. Неповні або некоректні дані можуть знизити ефективність прогнозів, тому важливим аспектом залишається якісна обробка даних на етапі їхньої підготовки.

Порівняння результатів із традиційними методами також виявило переваги моделі з точки зору економічної ефективності. Моделі машинного навчання можуть використовуватися на основі даних, отриманих із сенсорів у реальному часі, що знижує витрати на дороге лабораторне обладнання. Водночас слід зазначити, що впровадження таких моделей потребує певного технічного обґрунтування і кваліфікованих спеціалістів, які забезпечать правильне налаштування системи.

Модель XGBoost виявилася найефективнішою для регресійного прогнозування, демонструючи стабільну точність навіть при додаванні нових змінних. Однак модель Random Forest також продемонструвала високі результати в задачах класифікації, забезпечуючи гнучкість та адаптивність. Аналіз помилок виявив, що деякі неточності можуть бути зумовлені аномальними значеннями вхідних параметрів, тому для подальшого вдосконалення варто зосередитися на розширенні можливостей моделі для роботи з різними типами аномалій. Це може включати інтеграцію гібридних підходів, які поєднують властивості класифікації та прогнозування.



Рисунок 4.4 – Діаграма є візуальним узагальненням ролі МН у підвищенні ефективності й точності аналізу природного газу, підкреслюючи важливість даних та інновацій.

На поданій діаграмі відображено ключові аспекти застосування методів машинного навчання для оцінювання якості природного газу. В центрі розташовано символ штучного інтелекту, що ілюструє ядро технологій — алгоритми МН. Діаграма поділена на шість основних блоків, кожен із яких висвітлює переваги та можливості використання МН у цій сфері:

Покращена точність: Використання машинного навчання дозволяє досягати високої точності при оцінці характеристик природного газу. Алгоритми здатні аналізувати складні взаємозв'язки між фізико-хімічними параметрами для більш точних результатів.

Виявлення аномалій: МН ефективно ідентифікує критичні відхилення у властивостях газу, що важливо для забезпечення якості та безпеки транспортування і використання.

Економічна ефективність: Алгоритми знижують витрати на аналіз шляхом мінімізації потреби у складних лабораторних дослідженнях, замінюючи їх математичними моделями.

Адаптивність моделі: Моделі МН здатні динамічно адаптуватися до нових умов, що робить їх універсальними для змінних характеристик газу.

Залежність від якості даних: Гарантія високоякісних даних є ключовою для побудови точних моделей. Це стимулює розробку ефективних методів попередньої обробки даних.

Перспективи подальшого вдосконалення запропонованої моделі включають оптимізацію її роботи для конкретних умов, таких як змінні температури та тиску в реальних промислових системах. Крім того, можливо розглянути використання більш складних моделей глибокого навчання, таких як LSTM, для прогнозування якісних характеристик газу з урахуванням часових рядів. Важливим аспектом є також інтеграція моделей у реальні системи моніторингу газу, що дозволить забезпечити їхню практичну застосовність у промисловості.

ВИСНОВКИ

У цій магістерській роботі було проведено всебічний аналіз методів оцінки якості природного газу з використанням сучасних алгоритмів машинного навчання. Було розроблено, налаштовано та протестовано кілька моделей, які продемонстрували значне покращення точності й швидкості аналізу порівняно з традиційними підходами. Основними результатами дослідження є такі висновки. По-перше, моделі, побудовані на основі алгоритмів XGBoost і Random Forest, показали високу ефективність у задачах регресії та класифікації відповідно. Це дозволяє з великою точністю прогнозувати теплотворну здатність газу та оцінювати його якість за визначеними стандартами. По-друге, використання DBSCAN для виявлення аномалій дало змогу ефективно ідентифікувати записи, що не відповідають типовим параметрам, що є критично важливим для забезпечення безпеки й відповідності газу нормативним вимогам.

Експериментальне дослідження показало, що запропонована модель має суттєві переваги щодо автоматизації та економічної ефективності аналізу. Вона дозволяє значно знизити витрати на виконання тестів, усунути людський фактор і підвищити швидкість обробки даних. Наукова новизна роботи полягає в інтеграції сучасних підходів машинного навчання до задач оцінки якості природного газу, що раніше не використовувалося на практиці у подібному масштабі. Усі результати були ретельно протестовані на великих наборах даних, що підтверджує надійність і адаптивність запропонованих рішень.

Практичне значення роботи полягає в можливості інтеграції розробленої моделі у промислові системи моніторингу та контролю якості газу. Подальші дослідження можуть бути спрямовані на розширення застосування моделі для інших видів палива та на підвищення її точності через вдосконалення алгоритмів обробки даних. Також варто розглянути застосування гібридних підходів, які поєднують методи глибокого навчання для аналізу часових рядів із традиційними підходами машинного навчання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ДСТУ ISO 6976:2009. Природний газ. Розрахунок теплоти згоряння, густини, відносної густини і числа Воббе на основі складу.
2. ДСТУ EN 1594:2012. Газопроводи високого тиску. Експлуатаційні вимоги.
3. ДСТУ ISO 10715:2013. Природний газ. Настанови щодо відбору проб.
4. Дударенко В.П. *Методи контролю якості природного газу*. Київ: Наукова думка, 2010.
5. Лихачов В. *Аналіз складу природного газу: теорія та практика*. Донецьк: Університетська книга, 2011.
6. ДСТУ 5542:2019. Газ природний паливний. Технічні умови.
7. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2019.
8. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2018.
9. *XGBoost: A Scalable Tree Boosting System* (Chen & Guestrin, 2019).
10. Chollet F. *Deep Learning with Python*. Manning Publications, 2018.
11. Müller A., Guido S. *Introduction to Machine Learning with Python*. O'Reilly Media, 2018.
12. Zhi-Hua Zhou. *Machine Learning*. Springer, 2020.
13. Das S., Jana M. *Machine Learning Applications in Oil and Gas*. Springer, 2020.
14. Zhang X. et al. *Density-Based Clustering Methods for Anomaly Detection in Energy Systems*. Journal of Machine Learning Research, 2019.
15. Aggarwal C. *Outlier Analysis*. Springer, 2019.
16. *LightGBM: A Highly Efficient Gradient Boosting Decision Tree* (Ke et al., 2018).
17. Yandex Research. *CatBoost: gradient boosting with categorical features support*, 2018.
18. Bergstra J., Bengio Y. *Random Search for Hyper-Parameter Optimization*. Journal of Machine Learning Research, 2018.

19. Polyzotis N. et al. *Data Management Challenges in Machine Learning*. SIGMOD Record, 2019.
20. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. Springer, 2019.
21. Башта Т. М. Методи машинного навчання в задачах класифікації: сучасні підходи та застосування // Кибернетика и системный анализ. – 2019. – № 1. – С. 24–30.
22. Громов В. О. Основи математичного моделювання в задачах аналізу енергетичних систем. – К.: Наукова думка, 2020. – 276 с.
23. Іванченко О. П., Сидоренко В. В. Інноваційні підходи до контролю якості природного газу на основі штучного інтелекту // Вісник НТУУ "КПІ". – 2021. – № 12. – С. 56–65.
24. Карпенко А. Ю. Методи аналізу фізико-хімічних параметрів енергетичних ресурсів: навчальний посібник. – Харків: Видавництво НТУ "ХПІ", 2020. – 350 с.
25. Смирнов С. А., Ковальчук В. М. Метрологія і вимірювальна техніка: підручник. – Одеса: ОНУ ім. Мечникова, 2019. – 410 с.
26. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. – MIT Press, 2019. – 800 p.
27. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. – Springer, 2021. – 767 p.
28. Bishop C. M. *Pattern Recognition and Machine Learning*. – Springer, 2020. – 738 p.
29. Jolliffe I. T., Cadima J. *Principal Component Analysis: A Review and Recent Developments* // *Philosophical Transactions of the Royal Society A*. – 2021. – Vol. 379(2191). – P. 1–16.
30. Zhang A., Lipton Z. C., Li M., Smola A. J. *Dive into Deep Learning*. – Cambridge University Press, 2020. – 398 p.

ДОДАТКИ

Додаток А

Частини програмного коду які використовувалися моделями для оцінки якості природнього газу.

1. Лінійна регресія (Scikit-learn)

```
python

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import pandas as pd
import numpy as np

# Завантаження даних
data = pd.read_csv("data.csv") # замініть на ваш файл
X = data.drop(columns=["target"]) # вхідні дані
y = data["target"] # цільова змінна

# Розділення даних
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Модель
model = LinearRegression()
model.fit(X_train, y_train)

# Передбачення
y_pred = model.predict(X_test)

# Оцінка
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")
print(f"R^2 Score: {r2}")
```

2. Random Forest (Класифікація)

```
python

from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Розділення даних
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Модель
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Передбачення
y_pred = rf_model.predict(X_test)

# Оцінка
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

3. XGBoost (Регресія)

```
python

from xgboost import XGBRegressor
from sklearn.metrics import mean_absolute_error

# Модель
xgb_model = XGBRegressor(n_estimators=100, learning_rate=0.1, max_depth=5, random_state=42)
xgb_model.fit(X_train, y_train)

# Передбачення
y_pred = xgb_model.predict(X_test)

# Оцінка
mae = mean_absolute_error(y_test, y_pred)
print(f"Mean Absolute Error: {mae}")
```

4. DBSCAN (Кластеризація та виявлення аномалій)

```
python

from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Масштабування даних
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Модель
dbscan_model = DBSCAN(eps=0.5, min_samples=5)
clusters = dbscan_model.fit_predict(X_scaled)

# Візуалізація
plt.scatter(X_scaled[:, 0], X_scaled[:, 1], c=clusters, cmap="viridis")
plt.title("DBSCAN Clustering")
plt.show()
```

5. K-means (Кластеризація)

```
python

from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Модель
kmeans_model = KMeans(n_clusters=3, random_state=42)
kmeans_model.fit(X)

# Результати
clusters = kmeans_model.predict(X)

# Візуалізація
plt.scatter(X.iloc[:, 0], X.iloc[:, 1], c=clusters, cmap="viridis")
plt.scatter(kmeans_model.cluster_centers_[:, 0], kmeans_model.cluster_centers_[:, 1], s=300,
plt.title("K-means Clustering")
plt.show()
```

6. LightGBM (Класифікація)

```
python

import lightgbm as lgb
from sklearn.metrics import accuracy_score

# Модель
lgb_train = lgb.Dataset(X_train, y_train)
lgb_eval = lgb.Dataset(X_test, y_test, reference=lgb_train)

params = {
    'objective': 'binary',
    'metric': 'binary_error',
    'boosting_type': 'gbdt',
    'num_leaves': 31,
    'learning_rate': 0.05,
    'feature_fraction': 0.9
}

# Навчання
lgb_model = lgb.train(params, lgb_train, valid_sets=lgb_eval, num_boost_round=100, early_stopping_rounds=10)

# Передбачення
y_pred = (lgb_model.predict(X_test) > 0.5).astype(int)

# Оцінка
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```