

2. Кононова К. Ю. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка». Харків: ХНУ імені В. Н. Каразіна, 2020. 301 с. URL: <https://comsys.kpi.ua/upload/Rainforcement%20Learning%20.pdf> (дата звернення: 17.04.2024).

3. What Is Nonlinear Regression? *Nonlinear Regression*. URL: <https://ch.mathworks.com/discovery/nonlinear-regression.html> (дата звернення: 17.04.2024).

4. Метод інтерполяції для прогнозування метрик використання хмарних обчислень в статистичному навчанні / Н. А. Потапова, Л. О. Волонтир, І. П. Частоколенко, М. С. Григоренко. *Наука і техніка сьогодні*. 2024. № 4(32). С. 1192–1205.

УДК 519.6:004.42:004.043

*Козачок А. О., здобувачка 2 курсу спеціальності 122 Комп'ютерні науки, науковий керівник:
Хмелівський Ю. С., асистент кафедри інформаційних технологій*

РОЛЬ МЕТОДІВ ОБЧИСЛЕНЬ У ВИРІШЕННІ ЗАДАЧ АНАЛІЗУ ДАНИХ

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасному світі обробка та аналіз даних є важливим кроком у прийнятті рішень у різних сферах діяльності, зокрема в бізнесі, науці, медицині та соціальних науках. Масштаби зібраної інформації стрімко зростають і потребують відповідних методів обробки та аналізу. Роль обчислювальних методів у вирішенні завдань аналізу даних полягає в їх здатності ефективно обробляти великі обсяги інформації та забезпечувати точність, швидкість і достовірність отриманих результатів [1].

Важливим аспектом обчислювальних методів у вирішенні задач аналізу даних є їх внесок у створення та розвиток алгоритмів обробки даних. Обчислювальні методи, як-от машинне навчання, статистичний аналіз і обробка сигналів, можуть допомогти створити алгоритми, здатні розпізнавати образи і виявлення кореляцій у великих наборах даних. Наприклад, алгоритми класифікації даних, як-от опорні векторні машини та нейронні мережі, можуть відрізнити дані від одного класу до іншого [2].

Другий аспект методів обчислень полягає в їх здатності працювати з великими обсягами даних. Традиційні методи аналізу даних, як-от ручне кодування чи статистичні методи, можуть бути неефективними під час обробки великих обсягів даних через обмежену швидкість та ресурси. Методи обчислень, які базуються на паралельних обчисленнях або використанні великих обчислювальних кластерів, можуть значно прискорити аналіз даних, забезпечуючи швидкий доступ до інформації та високу продуктивність обробки.

На прикладі (рис. 1) показано, як простий метод аналізу даних може значно полегшити життя в сучасному світі, допомагаючи вирішувати навіть складні завдання. Приміром, шляхом аналізу великого обсягу медичних даних можна розробити прості моделі для прогнозу можливого ризику серцево-судинного захворювання у пацієнтів. Це дає змогу вчасно виявляти особливості, що можуть вка-

зувати на загрозу здоров'ю, і вживати відповідних заходів для попередження ускладнень. Такий підхід сприяє покращенню якості медичного обслуговування та збереженню людських життів.

Лістинг програми оцінки ризику серцево-судинних захворювань наведено на рис. 1–2.

```
import urllib.request
# URL для завантаження набору даних про серцево-судинні захворювання
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data"
# Запит на сервер для отримання даних
response = urllib.request.urlopen(url)
# Перевірка статусу відповіді
if response.getcode() == 200:

    data = response.read().decode('utf-8')
    data_lines = data.splitlines()

    # Проаналізуємо дані та прогнозуємо ризик для кожного пацієнта
    for line in data_lines:
        patient_data = line.split(",")
        age = float(patient_data[0])
        sex = float(patient_data[1])
        blood_pressure = int(float(patient_data[3]))

        # Прогнозуємо ризик захворювання на основі віку та артеріального тиску
        if age > 50 and blood_pressure > 140:
            risk = "High"
        else:
            risk = "Low"
```

Рис. 1. Лістинг програми оцінки ризику серцево-судинних захворювань. Частина 1

```
# Виводимо результати аналізу для кожного пацієнта
print(f"Пацієнт: Вік={age}, Стать={sex}, Артеріальний тиск={blood_pressure}")
print(f"Ризик серцево-судинного захворювання: {risk}")
print("="*50) # Для розділення результатів між пацієнтами
else:
    print("Помилка отримання даних:", response.getcode())
```

Рис. 2. Лістинг програми оцінки ризику серцево-судинних захворювань. Частина 2

Результат роботи цієї програми наведено на рис. 3.

```
Пацієнт: Вік=63.0, Стать=1.0, Артеріальний тиск=145
Ризик серцево-судинного захворювання: High
=====
Пацієнт: Вік=67.0, Стать=1.0, Артеріальний тиск=160
Ризик серцево-судинного захворювання: High
=====
Пацієнт: Вік=67.0, Стать=1.0, Артеріальний тиск=120
Ризик серцево-судинного захворювання: Low
=====
Пацієнт: Вік=37.0, Стать=1.0, Артеріальний тиск=130
Ризик серцево-судинного захворювання: Low
=====
Пацієнт: Вік=41.0, Стать=0.0, Артеріальний тиск=130
Ризик серцево-судинного захворювання: Low
```

Рис. 3. Результати роботи програми оцінки ризику серцево-судинних захворювань

Обчислювальні методи дають змогу покращити валідацію та тестування моделей аналізу даних. Широкий спектр обчислювальних методів допомагає проводити різні експерименти з даними, перевіряти різні гіпотези та визначати найбільш ефективний підхід до аналізу набору даних. Це підвищує достовірність результатів аналізу та робить процес прийняття рішень більш обґрунтованим [3].

Можна зазначити, що роль методів обчислень у вирішенні задач аналізу даних є критичною та незамінною. Сучасні підходи до обробки даних ґрунтуються на їх використанні, і вони відіграють ключову роль у всіх аспектах аналізу даних. Методи обчислень допомагають виявляти закономірності, які важливі для прийняття обґрунтованих рішень у різних сферах життя, від медицини до економіки. До того ж вони стимулюють розвиток нових напрямів досліджень у сфері аналізу даних, що відкриває нові можливості для вирішення складних проблем та досягнення нових знань. Тому важливо продовжувати розвивати та вдосконалювати методи обчислень, щоб забезпечити подальший прогрес в аналізі даних та в їх практичному застосуванні.

Список використаних джерел

1. Опрацювання даних як інформаційний процес. *Основні поняття з інформатики. UA5.org*. URL: <https://ua5.org/osnovi/2167-opraczyvannya-danyh-yak-informacziynuj-proczes.html>
2. Сем Фокс. OpenAI – нові можливості для бізнесу – розумовий аналіз даних та виявлення патернів. *Mediacom: світ новин*. 30.03.2024. URL: <https://mediacom.com.ua/openai-rozumovij-analiz-danix-ta-viyavlennya-paterniv-novi-mozhливosti-dlya-biznesu/>
3. Верес О. М. Класифікація методів аналізу великих даних. *Вісник Національного університету «Львівська політехніка»*. Серія: *Інформаційні системи та мережі*. Львів: Видавництво Львівської політехніки, 2017. № 872. С. 84–92.

УДК 519.2

*Сапожнікова В. Є., здобувачка 2 курсу спеціальності 122 Комп'ютерні науки,
Сеник І. О., асистент кафедри
інформаційних технологій*

МЕТОД НАЙМЕНШИХ КВАДРАТІВ

Донецький національний університет імені Василя Стуса, м. Вінниця

У галузі статистики часто виникає потреба знайти зв'язок між змінними, і метод найменших квадратів є широко використовуваною технікою для підгонки математичних моделей до даних. Завдяки своїй універсальності та широкому спектру застосувань він відіграє важливу роль у розв'язанні різноманітних задач.

Цей метод використовується для знаходження найкращої прямої (або деколи кривої), що представляє взаємозв'язок між цими змінними. Англійською мовою лінія має назву *least squares line*, що означає лінія найменших квадратів, або *regression line*, тобто лінія регресії, чи також *best fit line*, що перекладається як найбільш підходяща лінія; до прикладу, це може бути $y = mx + b$ (або $y = b_0 + b_1x$).