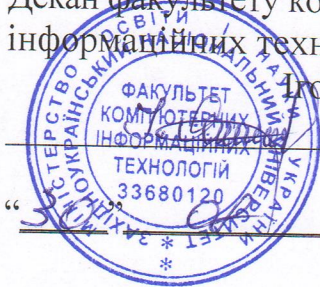


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ:

Декан факультету комп'ютерних  
інформаційних технологій  
Ігор ЯКИМЕНКО



2024 р.

ЗАТВЕРДЖУЮ:

Проректор з  
науково-педагогічної роботи  
Виктор ОСТРОВЕРХОВ



2024 р.

## РОБОЧА ПРОГРАМА

з дисципліни «Інтелектуальний аналіз даних»

ступінь вищої освіти – бакалавр

галузь знань – 12 “Інформаційні технології”

спеціальність – 122 „Комп’ютерні науки”

освітньо-професійна програма – „Штучний інтелект”

### Кафедра інформаційно-обчислювальних систем і управління

Форма навчання	Курс	Семестр	Лекції (год.)	Лабораторні заняття (год.)	ІРС (год.)	Тренінг (год.)	Самост. робота студ. (год.)	Разом (год.)	Екз. (сем.)
Денна	3	5	30	30	4	8	78	150	5

30.08.2024  
*[Signature]*

Тернопіль – ЗУНУ  
2024

Робоча програма складена на основі освітньо–професійної програми «Штучний інтелект» підготовки бакалавра галузі знань 12 “Інформаційні технології” спеціальності 122 «Комп’ютерні науки», затвердженої Вченою радою ЗУНУ (протокол № 9 від 15 червня 2022 р.).

Робочу програму склала доцент кафедри ІОСУ, к.т.н. Христина ЛІП’ЯНІНА-ГОНЧАРЕНКО

Робоча програма затверджена на засіданні кафедри інформаційно-обчислювальних систем і управління, протокол № 1 від 27 серпня 2024 р.

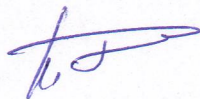
В.о. завідувача кафедри



Надія ВАСИЛЬКІВ

Розглянуто та схвалено групою забезпечення спеціальності „Комп’ютерні науки”, протокол № 1 від 30 серпня 2024 р.

Голова групи  
забезпечення спеціальності,  
д-р техн. наук, професор



Мирослав КОМАР

Гарант освітньо-професійної  
програми "Штучний інтелект",  
к.т.н., доцент



Василь КОВАЛЬ

## СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ " ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ "

### 1. Опис дисципліни " Інтелектуальний аналіз даних "

Дисципліна «Інтелектуальний аналіз даних»	Галузь знань, спеціальність, СВО	Характеристика навчальної дисципліни
Кількість кредитів – 5	Галузь знань – 12 “Інформаційні технології”	<b>Статус дисципліни:</b> обов’язкова дисципліна циклу професійної підготовки <b>Мова навчання:</b> Українська
Кількість залікових модулів – 5	Спеціальність – 122 «Комп’ютерні науки»	Рік підготовки: 3 Семестр: 5
Кількість змістових Модулів – 2	Освітньо- професійна програма «Штучний інтелект»	Лекції - 30 год.  Лабораторні заняття - 30 год.
Загальна кількість годин – 150	Ступінь вищої освіти – бакалавр	Самостійна робота – 78 год., Тренінг – 8 год.  Індивідуальна робота – 4 год.
Тижневих годин – 10, з них аудиторних – 4 год.		Вид підсумкового контролю – екзамен

## **2. Мета і завдання дисципліни " Інтелектуальний аналіз даних "**

### **2.1. Мета вивчення дисципліни**

Метою дисципліни „ Інтелектуальний аналіз даних ” є вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining), а також виявлення в необроблених масивах даних раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень.

### **2.2. Завдання вивчення дисципліни**

Завдання дисципліни “ Інтелектуальний аналіз даних ” полягає у опануванні базових принципів побудови моделей даних, ознайомлення з концепцією Knowledge Discovery in Data (виявлення знань в даних) і Data Mining («видобування» знань), ефективному використанні методів здобуття знань з великих масивів даних, ознайомленні з основними типами задач, що можуть бути вирішені за допомогою методів інтелектуального аналізу даних та отримані практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач і навчитися інтерпретувати отримані результати.

### **2.3. Найменування та опис компетентностей, формування яких забезпечує вивчення дисципліни:**

СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв’язування прикладних задач.

### **2.4. Передумови для вивчення дисципліни**

Курс базується на знаннях, одержаних при вивченні дисциплін «Вища математика», «Дискретна математика», «Чисельні методи та програмування», «Теорія ймовірності і математична статистика», «Основи комп’ютерних наук», «Алгоритми і структури даних», «Бази і сховища даних» та «Сучасні парадигми програмування».

### **2.5. Результати навчання:**

ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп’ютерних наук.

ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв’язування задач статистичної обробки даних і побудови прогнозних моделей.

ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

### **3. Програма навчальної дисципліни «Інтелектуальний аналіз даних»**

#### ***Змістовий модуль 1 – Основи інтелектуального аналізу даних***

##### **Тема 1. Концепції інтелектуального аналізу даних та DataMining**

Визначення необхідності аналізу даних. Цілі та основні завдання процесу інтелектуального аналізу даних. Технології інтелектуального аналізу даних. Ітераційний характер процесу інтелектуального аналізу даних та його основні кроки. Вплив якості даних на процес інтелектуального аналізу даних. Зв'язок між зберіганням даних та інтелектуальним аналізом даних. Концепції великих даних та науки про дані..

##### **Тема 2. Попередня обробка та аналіз даних**

Важливість належного представлення необроблених даних. Методи збору та введення даних до системи. Характеристики необроблених даних: типи, обсяг, структура. Процеси перетворення необроблених даних. Методи нормалізації та згладжування даних. Встановлення відмінностей та зв'язків між даними. Методи оцінки адекватності та повноти даних, зокрема методи поведінки та аугментації даних при обмежених вибірках.

##### **Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу**

Проблема відсутності даних і стратегії їх обробки. Методи заповнення пропусків у даних. Аналіз даних з неповним набором інформації. Виявлення та опрацювання викидів у даних.

##### **Тема 4. Аналіз великих наборів даних та методи зменшення розмірності**

Значущість великих обсягів даних. Проблеми збереження та використання великих датасетів. Стратегії для скорочення обсягу даних. Методи вибору та вилучення функцій. Підходи до зниження значення для збереження інформативності даних. Застосування міри ентропії для ранжування функцій. Визначення важливості функцій на основі інформаційної цінності. Аналіз головних компонентів. Методи пошуку головних компонентів.

##### **Тема 5. Методи візуалізації**

Різноманітні методи та підходи до візуалізації даних. Інструменти для графічного відображення інформації, які сприяють візуальному аналізу залежностей, шаблонів та трендів у даних. Графіки, діаграми, розподіли та інші засоби для ефективного представлення результатів аналізу даних.

#### ***Змістовий модуль 2 – Статистичні методи та моделювання***

##### **Тема 6. Статистичний аналіз**

Основні концепції та методи статистичного аналізу даних. Метод описової статистики та інференційної статистики. Статистичних тестів для зроблення

висновків на основі вибіркового даних. Ця тема охоплює методи порівняння та оцінки статистичних різниць між різними наборами даних. Досліджуються підходи для виявлення значущих різниць між групами даних та їхній вплив на аналіз результатів. Методи оцінки результатів.

### **Тема 7. Методи оцінки результатів**

Вступ до методів оцінки результатів. Метрики точності. F-міра та інші гармонічні метрики. Матриця помилок (Confusion Matrix). ROC-крива та AUC (Area Under the Curve). Перехресна перевірка (Cross-Validation).

### **Тема 8. Лінійний дискримінантний аналіз**

Лінійний дискримінантний аналіз як метод класифікації та розділення категорій на основі лінійних комбінацій змінних. Підходи до побудови та використання лінійного дискримінантного аналізу. Баєсівський підхід до класифікації даних. Метрики точності класифікації: F-Measure, Harmonic Mean, матриця помилок, ROC-крива та AUC.

### **Тема 9. Регресійні моделі**

Різні підходи до регресійного аналізу, які використовуються для передбачення залежних змінних на основі незалежних. Проста лінійна регресія, множинна лінійна регресія, нелінійна регресія, регресія з вибілковими точками, регресія з поліномами, регресія зі згладжуванням, регуляризована регресія. Методи оцінки результатів у регресійних моделях.

### **Тема 10. Прогнозування часових рядів**

Основні терміни та поняття, пов'язані з аналізом та прогнозуванням часових рядів. Методи визначення стаціонарності. Методи аналізу та моделювання часових рядів. Прогнозування сезонних та трендових компонентів. Авторегресія (AR). Ковзне середнє (MA). Авторегресія з ковзним середнім (ARMA). Авторегресія з інтегрованим ковзним середнім (ARIMA). Сезонна авторегресія з інтегрованим ковзним середнім (SARIMA). Авторегресія з екзогенними факторами (ARX). Екзогенна модель для ідентифікації (X12-ARIMA). Авторегресія з виправленням аномалій (ARIMA-X). Експоненційне згладжування (Exponential Smoothing). Модель Гарча (GARCH). Методи оцінки результатів.

### **Тема 11. TextMining і WebMining**

Специфіка WebMining. Класифікація основних підзадач WebMining. Можливості WebMining за допомогою HITS, LOGSOM та алгоритмів проходження шляхів. Незалежний від запитів ранжування веб-сторінок та основні характеристики алгоритму PageRank. Структура TextMining, що визначає етапи переробки та дистиляції. Методологія семантичної індексації.

#### 4. Структура залікового кредиту з дисципліни «Інтелектуальний аналіз даних»

##### Денна форма навчання

Тема	Кількість годин					
	Лекції	Лабораторні заняття	Індивідуальна робота	Тренінг	Самостійна робота	Контрольні заходи
<b>Змістовий модуль 1 – Основи інтелектуального аналізу даних</b>						
Тема 1. Концепції інтелектуального аналізу даних та DataMining	2	-	2	4	10	Опитування під час заняття
Тема 2. Попередня обробка та аналіз даних	2	2			8	
Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу	2	2			6	
Тема 4. Аналіз великих наборів даних та методи зменшення розмірності	2	4			6	
Тема 5. Методи візуалізації	2	2			10	
<b>Змістовий модуль 2 – Статистичні методи та моделювання</b>						
Тема 6. Статистичний аналіз	2	4	2	4	6	Опитування під час заняття
Тема 7. Методи оцінки результатів	2				6	
Тема 8. Лінійний дискримінантний аналіз	4	4	6			
Тема 9. Регресійні моделі	4	2	8			
		2				
Тема 10. Прогнозування часових рядів	4	4	6			
Тема 11. TextMining і WebMining	4	4	6			
<b>Разом</b>	30	30	4	8	78	

#### 5. Тематика лабораторних занять

**Лабораторна робота 1:** Введення до обробки та аналізу даних з використанням Python

**Мета:** Ознайомити студентів з основами використання мови програмування Python для обробки та аналізу даних, навчити роботі з Jupyter Notebook, завантаженню та читанню даних з різних форматів, а також надати загальний огляд бібліотеки Pandas для ефективних маніпуляцій з даними.

**Питання для обговорення:**

- Вступ до Python для аналізу даних.
- Робота з Jupyter Notebook та основними командами.
- Завантаження та читання даних з різних джерел (CSV, Excel, тощо).
- Огляд бібліотеки Pandas для маніпуляцій з даними.

**Лабораторна робота 2:** Дослідницький аналіз даних у Python.

**Мета:** Метою даної лабораторної роботи є вивчення основних кроків дослідницького аналізу даних з використанням мови програмування Python та відповідних бібліотек. У ході виконання лабораторної роботи студенти будуть ознайомлені з ключовими етапами обробки та аналізу даних, такими як збереження

та завантаження даних, видалення зайвої інформації, очищення від пропущених значень, виявлення викидів та візуалізація даних.

**Питання для обговорення:**

- Збереження і та визначення контрольної точки
- Вивчення набору даних
- Видалення неактуальних стовпців
- перейменування стовпців
- Видалення повторюваних рядків, пропущених або нульових значень, викидів.

**Лабораторне заняття 3:** Обробка та аналіз неструктурованими даних, включаючи великі дані.

**Мета:** Розглянути способи роботи з неструктурованими даними, такими як зображення, аудіо та текст, навчити студентів використовувати відповідні бібліотеки та техніки для обробки цих даних.

**Питання для обговорення:**

- Робота з неструктурованими даними: роль та значення неструктурованих даних у сучасному аналізі інформації.
- Використання бібліотек для обробки неструктурованих даних: переваги та особливості бібліотек для роботи з різними типами даних (зображення, аудіо, текст).

**Лабораторна робота 4:** Візуалізація даних

**Мета:** Вивчити методи візуалізації даних, ознайомити студентів з бібліотеками Matplotlib та Seaborn для створення графіків, діаграм та більш складних графічних представлень, а також вивчити способи візуалізації залежностей між різними ознаками.

**Питання для обговорення:**

- Використання бібліотеки Matplotlib для створення графіків та діаграм.
- Використання бібліотеки Seaborn для більш складних графічних представлень.
- Візуалізація залежностей між даними.

**Лабораторна робота 5:** Основи статистичного аналізу даних

**Мета:** Ознайомити студентів з основними поняттями описової статистики, розподілів даних, методами візуалізації, а також навчити їх проводити перевірку гіпотез та визначати р-значення для статистичних висновків.

**Питання для обговорення:**

- Описові статистики: середнє, медіана, мода, дисперсія, тощо.
- Розподіли даних та їх візуалізація.
- Перевірка гіпотез та р-значення.

**Лабораторна робота 6:** Лінійний дискримінантний аналіз

**Мета:** Ознайомити студентів з методом лінійного дискримінантного аналізу (LDA), навчити їх застосовувати LDA для зменшення розмірності та класифікації даних, а також порівнювати його з методом головних компонент (PCA).

**Питання для обговорення:**

- Вступ до методу лінійного дискримінантного аналізу (LDA).
- Застосування LDA для зменшення розмірності та класифікації.
- Порівняння з методом головних компонент (PCA).



### **Лабораторна робота 7:** Регресійний аналіз та прогнозування

**Мета:** Вивчити базові поняття регресійного аналізу, ознайомити студентів з методами простої та багатофакторної регресії, навчити оцінювати якість моделей та використовувати бібліотеку Scikit-learn для побудови прогнозуючих моделей.

#### **Питання для обговорення:**

- Вступ до регресійного аналізу.
- Проста та багатофакторна регресія.
- Оцінка якості моделі.
- Використання бібліотеки Scikit-learn для побудови моделей.

### **Лабораторна робота 8:** Баєсівський аналіз

**Мета:** Представити студентам введення до баєсівської статистики, навчити їх використовувати баєсівський підхід для моделювання невизначеності, а також познайомити з баєсівською регресією та класифікацією.

#### **Питання для обговорення:**

- Введення до баєсівської статистики.
- Використання баєсівського підходу для моделювання невизначеності.
- Баєсівська регресія та класифікація.

### **Лабораторна робота 9:** Прогнозування часових рядів

**Мета:** Вивчити методи аналізу та прогнозування часових рядів, навчити студентів застосовувати моделі ARIMA та експоненційного згладжування для прогнозування майбутніх значень часових рядів.

#### **Питання для обговорення:**

- Огляд методів аналізу та прогнозування часових рядів.
- Застосування моделей ARIMA та експоненційного згладжування.
- Оцінка точності прогнозів.

### **Лабораторна робота 10:** Text Mining і Web Mining

**Мета:** Ознайомити студентів з методами витягнення та аналізу даних з веб-сторінок, навчити обробляти та аналізувати текстові дані, а також використовувати методи аналізу настроїв тексту та виявлення ключових слів.

#### **Питання для обговорення:**

- Витягнення та аналіз даних з веб-сторінок.
- Обробка та аналіз текстових даних.
- Аналіз настроїв тексту та виявлення ключових слів.

## **6. Самостійна робота**

Самостійна робота студентів у рамках дисципліни передбачає виконання індивідуальних проектів, які спрямовані на оволодіння навичками застосування методів інтелектуального аналізу даних для вирішення прикладних проблем. Завдання на проект обираються студентом з переліку тем, наведеного нижче, або тема дослідження пропонується самостійно студентом (узгоджується з викладачем). Студенти мають представити та захистити свій проект у встановлені терміни.

Кожен проект оцінюється на основі наступних **критеріїв:**

- Якість аналізу та обробки даних.
- Оригінальність та інноваційність підходу.
- Глибина дослідження вибраної теми.
- Чіткість та логічність представлення результатів.

– Відповідність оформлення проекту встановленим нормам.

### Перелік тем:

№	Тема
1.	Аналіз ефективності заходів безпеки на підприємствах з використанням методів класифікації та прогнозування.
2.	Визначення факторів, що впливають на вірогідність банкрутства компаній, за допомогою алгоритмів кластерного аналізу.
3.	Аналіз текстових даних для виявлення фейкових новин за допомогою методів TextMining.
4.	Вивчення впливу різних джерел забруднення на стан довкілля із застосуванням багатовимірного аналізу даних.
5.	Прогнозування попиту на продукти харчування в різних регіонах за допомогою аналізу часових рядів.
6.	Оцінка ефективності різних методів маркетингових кампаній із використанням алгоритмів класифікації та регресії.
7.	Аналіз факторів, що впливають на якість освіти в різних країнах, з використанням статистичних моделей.
8.	Визначення патернів у даних про захворюваність населення з використанням методів кластерного аналізу.
9.	Оцінка впливу соціальних програм на рівень життя населення із застосуванням методів багатовимірного аналізу.
10.	Аналіз факторів, що впливають на затримки в авіаперевезеннях, з використанням методів класифікації.
11.	Вивчення взаємозв'язків між доходами населення та рівнем кримінальності в різних регіонах за допомогою регресійного аналізу.
12.	Оцінка ефективності методів попередження епідемій з використанням алгоритмів класифікації та прогнозування.
13.	Виявлення ключових факторів, що впливають на рівень задоволеності клієнтів у різних сферах послуг, з використанням багатовимірного аналізу.
14.	Аналіз взаємозв'язків між характеристиками продуктів харчування та їх популярністю серед споживачів із використанням кластерного аналізу.
15.	Оцінка впливу економічних криз на фінансові ринки з використанням моделей прогнозування.
16.	Аналіз даних про продажі автомобілів для визначення ключових факторів, що впливають на вибір споживачів.
17.	Визначення патернів у поведінці користувачів мобільних додатків із застосуванням методів кластеризації.
18.	Оцінка впливу сезонних факторів на попит на туристичні послуги за допомогою аналізу часових рядів.
19.	Аналіз факторів, що впливають на успішність онлайн-курсів, з використанням методів класифікації.
20.	Виявлення закономірностей у даних про використання енергії для оптимізації енергоспоживання на підприємствах.
21.	Оцінка впливу різних факторів на тривалість життя населення за допомогою регресійного аналізу.
22.	Аналіз даних про використання громадського транспорту для поліпшення його ефективності із застосуванням методів кластеризації.
23.	Визначення факторів, що впливають на лояльність клієнтів у банківському секторі, з використанням багатовимірного аналізу.
24.	Оцінка ризиків у страхуванні життя із використанням методів класифікації та прогнозування.
25.	Аналіз даних про результати виборів для визначення ключових факторів, що впливають на політичні уподобання.
26.	Вивчення взаємозв'язків між споживанням медичних послуг та рівнем доходів населення із застосуванням регресійного аналізу.

### 7. Тренінг з дисципліни

Завданням тренінгу є розробка програмного забезпечення із застосуванням методів інтелектуального аналізу даних. Тематика проектів буде базуватися на індивідуальних завданнях самостійної роботи, з особливим акцентом на практичне використання аналітичних методів для вирішення реальних задач. Основна мета тренінгу — розробка програмних рішень, що демонструють вміння студентів аналізувати, проектувати та впроваджувати ефективні інструменти аналізу даних у різних прикладних сферах.

Етап тренінгу	Завдання
Інструментарії розробки ПЗ для	- Налаштувати середовище розробки та встановити

аналізу даних	необхідні інструменти. - Створити базовий проект.
Використання методів інтелектуального аналізу даних у розробці ПЗ	Реалізувати програмне рішення з використанням обраного алгоритму інтелектуального аналізу даних.
Презентація та обговорення результатів	Підготувати презентацію результатів роботи та захистити проект перед аудиторією.

### 8. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Інтелектуальний аналіз даних» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

Модуль 1		Модуль 2		Модуль 3	Модуль 4	Модуль 5
10 %	10 %	10 %	10 %	5%	15%	40%
Поточне оцінювання	Модульний контроль 1	Поточне оцінювання	Модульний контроль 2	Тренінг.	Самостійна робота.	Екзамен.
	25 тестових питань.		- Одне теоретичне питання. - Одна задача.			- 10 тестових питань. - Одне теоретичне питання. - Одна задача.

Оцінка за «Поточне оцінювання» визначається як середнє арифметичне з оцінок, отриманих під час занять (кожен здобувач має бути оцінений не рідше як раз на два заняття). Пропуски практичних (лабораторних) занять обов'язково відпрацьовуються в години консультацій, в іншому випадку вони вважаються оцінкою «0» та враховуються при визначенні середнього арифметичного. Для здобувачів, які навчаються за індивідуальним графіком, поточне оцінювання проводиться під час консультацій, та шляхом виконання завдань в системі Moodle.

Модуль «Тренінг» визначається як середнє арифметичне з оцінок, за використання завдань тренінгу.

Модуль «Самостійна робота» оцінюється як середнє арифметичне з двох оцінок, отриманих під час виконання наскрізного проекту: за якість викладу матеріалу та його представлення, відповідно до критеріїв, визначених у розділі 6 «Самостійна робота» цієї робочої програми.

### 9. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№	Найменування	Номер теми
1.	Середовище Python	1-11
2.	Мультимедійне обладнання	1-11

### РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних

: навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.

2. Бахрушин В. Є. Методи аналізу даних: Навч. посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2019. – 268 с.

3. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2020. – 73 с.

4. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2020. 152 с.

5. Jiajun, Z., Zong, C., & Xia, R. (2022). Text Data Mining. Springer.

6. Li, B., Yue, L., Jiang, J., Chen, W., Li, X., Long, G., Fang, F., & Yu, H. (Ред.). (2022). Advanced Data Mining and Applications. Springer International Publishing. <https://doi.org/10.1007/978-3-030-95408-6>

7. Park, L. A. F., Gomes, H. M., Doborjeh, M., Boo, Y. L., Koh, Y. S., Zhao, Y., Williams, G., & Simoff, S. (Ред.). (2022). Data Mining. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8746-5>

8. Tan, Y., & Shi, Y. (Ред.). (2022). Data Mining and Big Data. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8991-9>

9. Ufuk Baytar, C. (Ред.). (2022). The Future of Data Mining. Nova Science Publishers. <https://doi.org/10.52305/kcin5931>

10. Cao, J. (2023). E-Commerce Big Data Mining and Analytics. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-3588-8>

11. Konys, A., & Nowak-Brzezińska, A. (2023). Knowledge Engineering and Data Mining. MDPI. <https://doi.org/10.3390/books978-3-0365-6789-1>

12. Mai, T. T., Crane, M., & Bezbradica, M. (2023). Educational Data Mining und Learning Analytics. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-39607-7>

13. Olson, D. L., & Lauhoff, G. (2023). Deskriptives Data-Mining. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-21274-1>

14. Shah, K., Shah, N., Sawant, V., & Parolia, N. (2023). Practical Data Mining Techniques and Applications. Auerbach Publications. <https://doi.org/10.1201/9781003390220>

15. Zhang, H. (2023a). Handbook of Mobility Data Mining, Volume 1: Data Preprocessing and Visualization. Elsevier.

16. Zhang, H. (2023b). Handbook of Mobility Data Mining, Volume 2: Mobility Analytics and Prediction. Elsevier.

17. Zhang, H. (2023c). Handbook of Mobility Data Mining, Volume 3: Mobility Data-Driven Applications. Elsevier.

### **Перелік баз даних, які можуть бути використані для реалізації практичних задач:**

1. World Bank Open Data - <https://data.worldbank.org/>
2. Kaggle Datasets - <https://www.kaggle.com/datasets>

3. UCI Machine Learning Repository - <https://archive.ics.uci.edu/ml/>
4. European Social Survey (ESS) - <https://www.europeansocialsurvey.org/>
5. National Climatic Data Center (NCDC) - <https://www.ncei.noaa.gov/>
6. Global Terrorism Database (GTD) - <https://www.start.umd.edu/gtd/>
7. Google Trends - <https://trends.google.com/trends/>
8. OECD Data - <https://data.oecd.org/>
9. IMDb Datasets - <https://www.imdb.com/interfaces/>
10. HealthData.gov - <https://www.healthdata.gov/>
11. OpenStreetMap - <https://www.openstreetmap.org/>
12. US Census Bureau - <https://www.census.gov/data.html>
13. Yelp Open Dataset - <https://www.yelp.com/dataset>
14. Eurostat - <https://ec.europa.eu/eurostat>
15. Financial Data from Yahoo Finance - <https://finance.yahoo.com/>
16. CDC Wonder - <https://wonder.cdc.gov/>
17. GDELT Project - <https://www.gdeltproject.org/>
18. FRED Economic Data - <https://fred.stlouisfed.org/>
19. Airbnb Data - <https://www.airbnb.com/data>
20. NASA Earth Observing System Data and Information System (EOSDIS) - <https://earthdata.nasa.gov/>