

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

МЕТОДИЧНІ ВКАЗІВКИ ТА ЗАВДАННЯ ДЛЯ
САМОСТІЙНОЇ РОБОТИ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»
для студентів спеціальності 122 «Комп'ютерні науки»

м. Тернопіль – ЗУНУ
2024 р.

Методичні вказівки та завдання для самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для студентів освітньо-кваліфікаційного рівня «бакалавр» спеціальності 122 «Комп'ютерні науки» / Укладач: Лип'яніна-Гончаренко Х.В. — Тернопіль, ЗУНУ, 2024, – с.11

Укладач:

Лип'яніна-Гончаренко Х.В., доцент к.т.н.

ЗМІСТ

Мета і завдання дисципліни	4
Програма навчальної дисципліни	5
Структура залікового кредиту з дисципліни	7
Самостійна робота.....	8
Рекомендовані джерела інформації.....	13

Мета і завдання дисципліни " Інтелектуальний аналіз даних "

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.

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 отриманні практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач і навчитися інтерпретувати отримані результати.

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.

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».

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 багатовимірної аналізу даних на основі технологій DataMining, TextMining, WebMining.

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 вибірових 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 вибірковими 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.

Структура залікового кредиту з дисципліни «Інтелектуальний аналіз даних»

Денна форма навчання

Тема	Кількість годин					
	Лекції	Лабора-торні заняття	Індивідуа-льна робота	Тренінг	Самостійна робота	Контрольні заходи
Змістовий модуль 1 – Основи інтелектуального аналізу даних						
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	
Змістовий модуль 2 – Статистичні методи та моделювання						
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	
Разом	30	30	4	8	78	

Заочна форма навчання

Тема	Кількість годин			
	Лекції	Лабораторні заняття	Тренінг	Самостійна робота
Змістовий модуль 1 – Основи інтелектуального аналізу даних				
Тема 1. Концепції інтелектуального аналізу даних та DataMining	4	2	-	14
Тема 2. Попередня обробка та аналіз даних				10
Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу				10
Тема 4. Аналіз великих наборів даних та методи зменшення розмірності				10
Тема 5. Методи візуалізації				14
Змістовий модуль 2 – Статистичні методи та моделювання				
Тема 6. Статистичний аналіз	4	2	-	10
Тема 7. Методи оцінки результатів				10
Тема 8. Лінійний дискримінантний аналіз				10
Тема 9. Регресійні моделі				20
Тема 10. Прогнозування часових рядів				20
Тема 11. TextMining і WebMining				10
Разом	8	4	-	138

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.
27.	Виявлення тенденцій у сфері праці та безробіття за допомогою аналізу великих даних.
28.	Прогнозування тенденцій на фондовому ринку за допомогою моделей машинного навчання.
29.	Оцінка впливу кліматичних змін на врожайність сільськогосподарських культур із використанням алгоритмів прогнозування.
30.	Аналіз даних споживання електроенергії для виявлення неефективних користувачів із використанням кластеризації.
31.	Дослідження впливу соціальних медіа на виборчі процеси з використанням методів текстового аналізу.
32.	Оцінка ефективності заходів щодо зменшення забруднення повітря за допомогою регресійних моделей.
33.	Вивчення факторів, що впливають на рентабельність малого бізнесу за допомогою методів кластеризації.
34.	Прогнозування поведінки покупців в електронній комерції із використанням моделей машинного навчання.
35.	Оцінка ризиків інвестування в криптовалюти з використанням моделей прогнозування.
36.	Аналіз впливу пандемій на економіку країн із використанням статистичних методів та моделей прогнозування.
37.	Прогнозування попиту на медичні послуги в умовах демографічних змін із використанням моделей прогнозування.
38.	Аналіз факторів, що впливають на якість питної води, з використанням методів кластеризації та регресії.
39.	Прогнозування відтоку клієнтів у телекомунікаційній сфері за допомогою алгоритмів машинного навчання.
40.	Виявлення шахрайських транзакцій у банківських операціях з використанням методів класифікації.
41.	Аналіз впливу соціальних мереж на споживчі тенденції за допомогою текстового аналізу.
42.	Оцінка ефективності різних методів лікування захворювань з використанням статистичних моделей.
43.	Дослідження впливу цін на нерухомість на ринок оренди за допомогою регресійного аналізу.
44.	Визначення факторів, що впливають на успішність фільмів у прокаті, з використанням багатовимірного аналізу.
45.	Аналіз поведінки клієнтів банку для пропозиції персоналізованих продуктів із застосуванням методів кластеризації.
46.	Прогнозування попиту на електромобілі в різних регіонах з використанням моделей прогнозування.
47.	Вивчення впливу демографічних факторів на вибір страхових продуктів із застосуванням методів класифікації.
48.	Аналіз взаємозв'язків між фізичною активністю та здоров'ям населення з використанням статистичних методів.
49.	Виявлення тенденцій у споживанні музики за допомогою аналізу великих даних.
50.	Оцінка ефективності рекламних кампаній у соціальних медіа з використанням алгоритмів

	класифікації та регресії.
51.	Дослідження впливу рівня освіти на рівень безробіття за допомогою регресійного аналізу.
52.	Аналіз факторів, що впливають на ефективність роботи персоналу, з використанням методів кластеризації.
53.	Прогнозування ризиків дорожньо-транспортних пригод із застосуванням моделей машинного навчання.
54.	Оцінка впливу глобалізації на локальні ринки за допомогою статистичних моделей та прогнозування.
55.	Аналіз даних про споживання води для оптимізації ресурсів із застосуванням методів кластеризації.
56.	Вивчення впливу онлайн-рецензій на продажі продуктів за допомогою текстового аналізу.
57.	Оцінка впливу кібербезпеки на довіру користувачів до онлайн-сервісів з використанням статистичних методів.
58.	Аналіз факторів, що впливають на споживання онлайн-контенту, з використанням методів кластеризації та регресії.
59.	Прогнозування попиту на електроенергію в пікові години із застосуванням моделей часових рядів.
60.	Виявлення шахрайства в страхових вимогах за допомогою алгоритмів класифікації та аномалій.
61.	Аналіз впливу цін на паливо на транспортні звички населення з використанням статистичних методів.
62.	Оцінка ефективності програм лояльності в роздрібній торгівлі із застосуванням методів кластеризації.
63.	Дослідження факторів, що впливають на успішність мобільних ігор, з використанням багатовимірної аналізу.
64.	Прогнозування рівня забруднення повітря в містах із використанням моделей машинного навчання.
65.	Аналіз поведінки користувачів в онлайн-банкінгу для підвищення безпеки з використанням кластеризації та аномалій.
66.	Оцінка впливу маркетингових стратегій на продажі електронних гаджетів із застосуванням регресійного аналізу.
67.	Вивчення впливу соціальних факторів на результати спортивних змагань за допомогою статистичних моделей.
68.	Аналіз факторів, що впливають на вибір університету абітурієнтами, з використанням методів кластеризації.
69.	Прогнозування ризиків природних катастроф з використанням алгоритмів машинного навчання та геопросторових даних.
70.	Виявлення трендів у модній індустрії за допомогою аналізу соціальних мереж та текстового аналізу.
71.	Оцінка ефективності програм з енергозбереження в житлових будинках із використанням статистичних методів.
72.	Аналіз впливу відеоігрової залежності на успішність студентів за допомогою регресійного аналізу.
73.	Дослідження факторів, що впливають на поширення вірусних відео, з використанням методів класифікації.
74.	Прогнозування продажів книг на основі аналізу попередніх даних та відгуків читачів.
75.	Визначення факторів, що впливають на якість обслуговування в готелях, з використанням багатовимірної аналізу.
76.	Аналіз даних про використання паркувальних місць для оптимізації міської інфраструктури із застосуванням кластеризації.
77.	Оцінка впливу освіти на здоров'я населення із використанням статистичних методів та регресійного аналізу.
78.	Аналіз факторів, що впливають на продуктивність сільськогосподарських тварин із застосуванням методів кластеризації та регресії.
79.	Прогнозування попиту на послуги хмарних обчислень з використанням моделей машинного

	навчання.
80.	Виявлення шахрайства в онлайн-торгівлі за допомогою алгоритмів класифікації та виявлення аномалій.
81.	Аналіз впливу медіа-контенту на громадську думку з використанням методів аналізу тексту та соціальних мереж.
82.	Оцінка ефективності державних програм підтримки малого бізнесу із застосуванням статистичних методів.
83.	Дослідження факторів, що впливають на успішність краудфандингових кампаній, з використанням багатовимірного аналізу.
84.	Прогнозування ризиків кредитування фізичних осіб з використанням алгоритмів машинного навчання.
85.	Аналіз даних про транспортні потоки для оптимізації дорожнього руху із застосуванням методів кластеризації.
86.	Оцінка впливу онлайн-реклами на поведінку споживачів за допомогою регресійного аналізу.
87.	Вивчення впливу різних стилів управління на продуктивність команд із використанням статистичних методів.
88.	Аналіз факторів, що впливають на успішність подкастів, з використанням багатовимірного аналізу.
89.	Прогнозування тенденцій розвитку ринку електронної комерції з використанням моделей прогнозування.
90.	Визначення факторів, що впливають на задоволеність пасажирів авіакомпаній, із застосуванням методів кластеризації.
91.	Аналіз впливу гендерних стереотипів на працевлаштування з використанням методів аналізу тексту та статистики.
92.	Оцінка ефективності програм по боротьбі зі зміною клімату за допомогою статистичних моделей та прогнозування.
93.	Дослідження впливу цін на продукти харчування на споживчі звички населення з використанням регресійного аналізу.
94.	Аналіз даних про використання мобільних пристроїв для підвищення безпеки з використанням методів класифікації.
95.	Прогнозування успішності стартапів на основі аналізу попередніх даних із застосуванням машинного навчання.
96.	Виявлення трендів у використанні відновлюваних джерел енергії за допомогою аналізу великих даних.
97.	Оцінка впливу технологій штучного інтелекту на ринок праці з використанням статистичних моделей та прогнозування.
98.	Аналіз факторів, що впливають на вибір туристичних напрямків, з використанням методів кластеризації та регресії.
99.	Прогнозування попиту на житлову нерухомість із використанням моделей машинного навчання та аналізу ринку.
100.	Виявлення шахрайства в медичному страхуванні за допомогою алгоритмів класифікації та виявлення аномалій.
101.	Аналіз впливу зміни клімату на морські екосистеми з використанням статистичних методів та моделей прогнозування.
102.	Оцінка ефективності онлайн-освіти порівняно з традиційною за допомогою багатовимірного аналізу та статистичних моделей.
103.	Дослідження факторів, що впливають на якість повітря в міських умовах із застосуванням методів кластеризації та регресії.
104.	Прогнозування успіху музичних альбомів на основі аналізу попередніх даних та трендів у музичній індустрії.
105.	Аналіз поведінки користувачів соціальних мереж для покращення алгоритмів рекомендацій із застосуванням машинного навчання.
106.	Оцінка впливу транспортних заторів на економіку міста з використанням статистичних моделей та аналізу великих даних.
107.	Вивчення факторів, що впливають на рішення про вакцинацію населення, з використанням

	методів класифікації та аналізу соціальних медіа.
108.	Аналіз взаємозв'язків між заняттями спортом та академічною успішністю студентів із застосуванням регресійного аналізу.
109.	Виявлення трендів у споживанні здорової їжі за допомогою аналізу споживчих даних та кластеризації.
110.	Прогнозування рівня безробіття в країні із застосуванням моделей часових рядів та економічних індикаторів.
111.	Аналіз впливу кібератак на фінансові установи з використанням статистичних методів та аналізу великих даних.
112.	Оцінка ефективності стратегій SEO в цифровому маркетингу за допомогою методів класифікації та регресії.
113.	Дослідження факторів, що впливають на вибір мов програмування розробниками, із застосуванням багатовимірної аналізу.
114.	Прогнозування успішності кандидатів на ринку праці на основі аналізу їхніх резюме та профілів у соціальних мережах.
115.	Аналіз даних про споживання алкоголю для виявлення ризиків для здоров'я населення із застосуванням статистичних методів.
116.	Виявлення факторів, що впливають на задоволеність клієнтів у сфері електронної комерції з використанням методів кластеризації.
117.	Аналіз впливу цифрової трансформації на продуктивність підприємств із використанням статистичних методів та моделей прогнозування.
118.	Аналіз факторів, що впливають на ефективність дистанційної роботи, з використанням методів кластеризації та регресії.
119.	Прогнозування тенденцій розвитку фінтех-індустрії із застосуванням моделей машинного навчання та аналізу ринку.
120.	Виявлення шахрайських дій у сфері електронного голосування за допомогою алгоритмів класифікації та виявлення аномалій.

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 ; КПП ім. Ігоря Сікорського. – Київ : КПП 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/>