

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

ЗАТВЕРДЖУЮ
Володарчук Ірина Іванівна, комп'ютерних
інформаційних технологій
проф. ЯКИМЕНКО
" 24 " березня 2023 р.

ЗАТВЕРДЖУЮ
Директор навчально -
наукового інституту новітніх
освітніх технологій
Святослав ПИТЕЛЬ
" 24 " березня 2023 р.

ЗАТВЕРДЖУЮ
В.о.проєктора з науково-педагогічної
роботи
Віктор ОСТРОВЕРХОВ
" 24 " березня 2023 р.

РОБОЧА ПРОГРАМА
з дисципліни
«Інтелектуальний аналіз даних»

ступінь вищої освіти – магістр
галузь знань 12 «Інформаційні технології»
спеціальність – 121 «Інженерія програмного забезпечення»
Освітньо-професійна програма «Інженерія програмного забезпечення»

Кафедра комп'ютерних наук

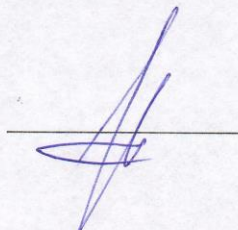
Форма навчання	Курс	Семестр (сесія)	Лекції (год.)	Практ. (год.)	ІРС (год.)	Тренінг (год.)	СРС (год.)	Разом (год.)	Залік. (сем)
денна	I	2	30	14	4	4	98	150	2
заочна	I	2,3	8	4	-	-	138	150	3

Тернопіль – 2023

Робочу програму склала доцент кафедри комп'ютерних наук, к.е.н.
Гончар Людмила Іванівна

Робоча програма затверджена на засідання кафедри комп'ютерних наук,
протокол № 1 від 28 серпня 2023 р.

Завідувач кафедри



Андрій Пукас

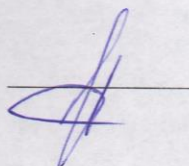
Розглянуто та схвалено групою забезпечення спеціальності Інженерія
програмного забезпечення, протокол № 1 від 28 серпня 2023 р.

Голова групи
забезпечення спеціальності



Микола Дивак

Гарант ОП



Андрій Пукас

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «Інтелектуальний аналіз даних»

1. Опис дисципліни «Інтелектуальний аналіз даних»

Дисципліна «Інтелектуальний аналіз даних»	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів – 5	Галузь знань – 12 «Інформаційні технології»	Статус дисципліни: вибіркова Мова навчання: українська
Кількість залікових модулів – 3	Спеціальність – 121 «Інженерія програмного забезпечення»	Рік підготовки: денна – 1 заочна – 2 Семестр: денна – 2 заочна – 2,3
Кількість змістових модулів – 2	Ступінь вищої освіти – магістр	Лекції: денна – 30 год. заочна – 8 год. Практичні заняття: денна – 14 год. заочна – 4 год.
Загальна кількість годин – 150 год		Самостійна робота: денна – 98 год. заочна – 138 год. Тренінг: денна – 4 год. Індивідуальна робота (КПЗ): денна – 5 год.
Тижневих годин – 10 год., з них аудиторних – 3 год.		Вид підсумкового контролю – залік

2. Мета і завдання вивчення дисципліни «Інтелектуальний аналіз даних»

2.1. Мета вивчення дисципліни

Метою викладання навчальної дисципліни «Інтелектуальний аналіз даних» є формування у студентів знань про основи інтелектуальної технології Data Mining, Data Science, моделей Data Mining, статистичних і кібернетичних методів Data Mining, машинного навчання, інтелектуальної технології Text Mining і Big Data, а також, умінь та використання доступного програмного забезпечення інтелектуального аналізу даних та сучасних інструментів Data Science.

2.2 Завдання вивчення дисципліни

Основними завданнями вивчення дисципліни «Інтелектуальний аналіз даних» є надання базових відомостей із основ інтелектуального аналізу даних (ІАД), світового досвіду та сучасного стану програмного забезпечення Data Mining, Big Data. Значна увагу приділяється вивченню технології машинного навчання, генетичних алгоритмів, набуття практичних навиків у правильному підборі сучасного програмного забезпечення для здійснення ІАД.

2.3. Результати навчання

У результаті вивчення дисципліни “Інтелектуальний аналіз даних” студент повинен **знати**:

- сутність та призначення технології Data Mining;
- характеристики процесів та активностей дейтамайнінгу;
- статистичні та кібернетичні методи Data Mining ;
- основні підходи до розв’язання задач машинного навчання;
- основні поняття та принципи роботи штучних нейронних мережі;
- доступне програмне забезпечення генетичних алгоритмів;
- технології та тенденції роботи з Big Data

У результаті вивчення дисципліни “Інтелектуальний аналіз даних” студент повинен **вміти**:

- застосовувати технологію Data Mining ;
- здійснювати кластерний аналіз засобами дейтамайнінгу ;
- розв’язувати задачі пошуку асоціативних правил;
- розв’язувати задачі, що виникають у практичній діяльності, за допомогою методів машинного навчання;
- застосовувати сучасні технології Big Data;
- використовувати онтологічний підхід до використання знань на підприємствах.

3. Програма навчальної дисципліни «Інтелектуальний аналіз даних»

Змістовний модуль 1. Дейтамайнінг (DataMining) – засоби інтелектуального аналізу даних у СППР.

Тема 1. Розвиток та призначення DataMining

Поняття ІАД. Моделі DataMining. Статистичні та кібернетичні методи DataMining. Інтеграція DataMining і OLAP. Відкриття. Пророче моделювання. Аналіз аномалій.

Література: [1,2,7,]

Тема 2. Дерево методів дейтамайнінгу

Класифікація. Кластеризація. Асоціація. Упорядкування. Прогнозування.

Література: [4,8,]

Тема 3. Збережені дані. Дистильовані дані

Метод “nearestneighbor”. Міркування за аналогією. Кластерний аналіз. Крос-

табуляція. Задача пошуку асоціативних правил.

Література: [1,5,6]

Тема 4. Характеристика процесів та активностей дейтамайнінгу

Умовна логіка. Тренди і варіації. Прогнозування. Виявлення ідевітації. Характеристика процесів і активностей дейтамайнінгу. Процеси дейтамайнінгу. Типи процесів дейтамайнінгу. Онтологічний підхід до використання знань на IT-підприємстві.

Література: [1,7]

Тема 5. Генетичні алгоритми

Генетичні успадкування - концептуальна засада генетичних алгоритмів. алгоритми. Загальна схема генетичних алгоритмів. Генетичні успадкування. Доступне програмне забезпечення генетичних алгоритмів.

Література: [4,8]

Тема 6. Програмні агенти в СППР

Призначення і основні характеристики програмних агентів. DSS Agent. Агенто-орієнтовані системи на підприємстві.

Література: [1,3,5]

Змістовий модуль 2. Доступне програмне забезпечення дейтамайнінгу

Тема 7. Poly Analyst. Mine Set – візуальний інструмент аналітика. Knowledge STUDIO. Інтелектуальна технологія Text Mining

Література: [2,4,10]

Тема 8. Засоби штучного інтелекту в СППР. Нейронні мережі

Визначення та еволюція нейронних мереж. Застосування нейронних мереж. Математична модель штучного нейронна. Архітектура нейромереж. Навчання та використання нейромереж. Готове програмне забезпечення нейромереж (нейропакети).

Література: [1,3,5,6,10]

Тема 9. Нейронні мережі з TensorFlow.

Функції активації. Оптимізатори. Згортковий шар. Обробка природної мови за допомогою RNN. Створення ігрового генератора.

Тема 10. Машинне навчання

Машинне навчання з Python. Основи машинного навчання. Тренування та тестування даних. Класифікація. Побудова моделі. Кластеризація. Приховані моделі Маркова.

Тема 11. Основні технології Big Data. Сучасні інструменти Data Science

Характеристика Big Data. Приклади задач Big Data. Сучасні технології Big Data. Парадигма MapReduce. Платформа для розподілених обчислень Hadoop.

Базові відомості про Data Science. Методи використання навчальної інформації. Методи багатомірного розвідувального аналізу (багатовимірний не графічний та багатовимірний графічний). Метрики оцінки якості моделей машинного навчання. Методи прогнозування.

**4. Структура залікового кредиту дисципліни «Інтелектуальний аналіз даних»
денна форма навчання**

денна форма навчання	Кількість годин					
	Лекції	Лабораторна робота	СРС	ІРС	Тренінг КПЗ	Контрольні заходи
Змістовий модуль 1. Дейтамайнінг (DataMining) – засоби інтелектуального аналізу даних у СППР						
Тема 1. Розвиток та призначення DataMining.	2	1	8	2	2	Усне опитування/тестування
Тема 2. Дерево методів дейтамайнінгу	3	1	8			Усне опитування/тестування
Тема 3. Збережені дані. Дистильовані дані.	2	1	8			Усне опитування/тестування
Тема 4. Характеристика процесів та активностей дейтамайнінгу	4	1	8			Усне опитування/тестування
Тема 5. Генетичні алгоритми	4	1	10			Усне опитування/тестування
Тема 6. Програмні агенти в СППР	3	1	10			Усне опитування/тестування
Змістовий модуль 2. Доступне програмне забезпечення дейтамайнінгу						
Тема 7. Poly Analyst. Mine Set – візуальний інструмент аналітика. Knowledge STUDIO. Інтелектуальна технологія Text Mining	2	2	10	3	2	Усне опитування/тестування
Тема 8. Засоби штучного інтелекту в СППР. Нейронні мережі	2	1	10			Усне опитування/тестування
Тема 9. Нейронні мережі з TensorFlow	2	1	10			Усне опитування/тестування
Тема 10. Машинне навчання	2	2	10			Усне опитування/тестування
Тема 11. Основні технології Big Data. Сучасні інструменти Data Science	4	2	10			Усне опитування/тестування
Разом	30	14	102	5	4	

заочна форма навчання	Кількість годин		
	Лекції	Лабораторна робота	Самостійна робота
Тема 1. Розвиток та призначення DataMining.	2	1	14
Тема 2. Дерево методів дейтамайнінгу			10
Тема 3. Характеристика процесів та активностей дейтамайнінгу			10
Тема 4. Генетичні алгоритми			20
Тема 5. Програмні агенти в СППР			10
Тема 6 Poly Analyst. Mine Set – візуальний інструмент аналітика. Knowledge STUDIO. Інтелектуальна технологія Text Mining	2		20
Тема 7. Засоби штучного інтелекту в СППР. Нейронні мережі			20
Тема 8. Нейронні мережі з TensorFlow	2	1	10
Тема 9. Машинне навчання .			10
Тема 10. Основні технології Big Data. Сучасні інструменти Data Science	2		24
Разом	8	2	138

5. Тематика практичних занять

Практичне заняття № 1. (1 год.)

Тема: Поняття ІАД. Моделі DataMining.

Мета: Ознайомитись із ІАД, моделями, знати інтеграцію DataMining і OLAP.

Питання для обговорення:

1. Статистичні та кібернетичні методи DataMining.
2. Інтеграція DataMining і OLAP
3. Інтелектуальна фаза процесу прийняття рішень.

Література: 1, 4, 9,14.

Практичне заняття № 2. (1 год.)

Тема: PolyAnalyst– сімейство продуктів DataMining

Мета: Навчитись практично застосовувати ПЗ PolyAnalyst для машин досліджень (згідно варіанту)

Питання для обговорення:

1. Архітектура PolyAnalyst.
2. Машини досліджень PolyAnalyst.
3. Підтримка OLE DB for Data Mining.

Література: 1,8,15,21.

Практичне заняття № 3. (1 год.)

Тема: Метод “nearestneighbor. Міркування за аналогією. Кластерний аналіз

Мета: Вивчити метод “nearestneighbor” і вміти проводити кластерний

аналіз засобами PolyAnalyst.

Питання для обговорення:

1. Програмне забезпечення CBR.
2. Поняття кластерного аналізу.
3. Міркування за аналогією.

Література:1,4,9,17,24.

Практичне заняття № 4. (1 год.)

Тема :Крос-табуляція

Мета: Знати поняття крос-табуляції і вміти будувати крос-таблицю при здійсненні ІАД.

Питання для обговорення:

- 1.Візуалізація даних.
- 2.Дерева рішень.
3. Процес побудови крос-таблиці в Data Mining.

Література:3,4,9,18,22.

Практичне заняття № 5. (2 год.)

Тема: Аналіз даних за допомогою ПЗ Wizwhy

Мета: Навчитись аналізувати дані за допомогою **Wizwhy**

Питання для обговорення:

1. Завантаження даних
2. Налаштування параметрів
3. Передбачувана змінна
4. Характеристики результуючих правил
5. Ціна помилки
6. Параметри виводу результатів
7. Додаткові параметри налаштування
8. Запуск аналізатора
9. Інтерпретуючі результати
- 10.Збереження результатів

Література: 1,15,16,23.

Практичне заняття № 6. (2 год.)

Тема: Доступне програмне забезпечення генетичних алгоритмів. Генетичні алгоритми. Генетичні успадкування. GeneHunter 1.0

Мета: Ознайомитись із поняттям генетичного алгоритму, доступним програмним забезпеченням генетичних алгоритмів, пакетом Evolver 4.0.,

Питання для обговорення:

1. Пакет Evolver 4.0.
2. GeneHunter 1.0.

3. Побудова генетичного алгоритму засобами Matlab.

4. Література:4,9,13,17,20.

Практичне заняття № 7 (2 год.)

Тема: Основні технології Big Data

Мета : Ознайомитись із технологіями і тенденціями роботи з Big Data.

Питання для обговорення:

1. Характеристика Big Data.
2. Приклади задач Big Data.
3. Сучасні технології Big Data.
4. Парадигма MapReduce.
5. Платформа для розподілених обчислень Hadoop.
6. Методи і техніка аналізу великих даних.

Література:2,3, 11, 17.

Практичне заняття № 8. (2 год.)

Тема: Машинне навчання

Мета: Ознайомитись із технологією машинного навчання з Python.

Питання для обговорення:

1. Основи машинного навчання.
2. Тренування та тестування даних.
3. Класифікація.
4. Побудова моделі.
5. Кластеризація.
6. Приховані моделі Маркова.

Література:3,12,19,22.

Практичне заняття № 9. (2 год.)

Тема: Сучасні інструменти Data Science

Мета: Вивчити базові відомості про Data Science.

Питання для обговорення:

1. Методи використання навчальної інформації.
2. Методи багатомірного розвідувального аналізу (багатовимірний не графічний та багатовимірний графічний).
3. Метрики оцінки якості моделей машинного навчання.
4. Методи прогнозування.

Література:14, 16, 17,21,24.

6. Комплексне практичне індивідуальне завдання

Індивідуальне завдання з дисципліни «Інтелектуальний аналіз даних» виконується самостійно кожним студентом, а саме:

Завдання до КПЗ:

1. Вивчити роботу програмного продукту Poly Analyst та інших ПЗ для реалізації обраного алгоритму.
2. Розробити за допомогою пакетів Poly Analyst та інших

реалізацію вибраного алгоритму.

3. Зробити порівняльну характеристику розробленого алгоритму із використанням різних програмних продуктів.
4. Навести приклад практичної реалізації алгоритмів для різного ПЗ.
5. Зробити висновки.
6. Створити презентацію для захисту індивідуального завдання

Індивідуальне завдання оцінюється за 100-бальною шкалою. Виконання індивідуального завдання є одним із обов'язкових складових модулів залікового кредиту з дисципліни «Інтелектуальний аналіз даних».

7. Самостійна робота

1. Класи процесів дейтамайнінгу: відкриття
2. Класи процесів дейтамайнінгу: аналіз аномалій.
3. Поняття класифікації в ІАД
4. Асоціація в ІАД
5. Упорядкування в ІАД
6. Прогнозування в ІАД
7. Класифікаційне дерево методів DataMining
8. Збережені дані. Дистильовані дані
9. Міркування за аналогією в ІАД
10. Програмне забезпечення CBR
11. Логічні методи (підходи)
12. Нечіткі запити і аналізи
13. Схема використання правил індукції в системі дейтамайнінгу
14. Багатовимірний індикативний аналіз
15. Архітектура системи багатовимірного інтелектуального аналізу даних
16. Класифікація технологічних методів ІАД
17. Knowledge STUDIO – нова версія дейтамайнінгу
18. Предметно-орієнтовані аналітичні системи
19. Метод „найближчого сусіда”
20. Критерії оцінки існуючих продуктів ІАД
21. Поняття генетичних алгоритмів
22. Еволюційне програмування
23. Модулі для побудови числових моделей і прогнозу числових змінних в Poly Analyst
24. Застосування Data Science, Big Data, Data Mining в бізнесі
25. Поняття краудсорсинг
26. Методи і техніка аналізу великих даних
27. Нейронні мережі з TensorFlow.
28. Функції активації

- 29.. Згортковий ша
- 30.Обробка природної мови за допомогою RNN
- 31.Створення ігрового генератора.
- 32.Машинне навчання
- 33.Машинне навчання з Python
- 34.Тренування та тестування даних
- 35.Приховані моделі Маркова.
- 36.Основні технології Big Data
- 37.Сучасні інструменти Data Science
- 38.Характеристика Big Data
- 39.Приклади задач Big Data
- 40.Парадигма MapReduce
- 41.Платформа для розподілених обчислень Hadoop.
- 42.Базові відомості про Data Science
- 43.Методи використання навчальної інформації
- 44.Методи багатомірного розвідувального аналізу (багатовимірний не графічний та багатовимірний графічний).
- 45.Метрики оцінки якості моделей машинного навчання.
- 46.Засоби штучного інтелекту в СППР. Нейронні мережі
- 47.Визначення та еволюція нейронних мереж.
48. Застосування нейронних мереж.
- 49.Математична модель штучного нейронна. Архітектура нейромереж.
- 50.Навчання та використання нейромереж. Готове програмне забезпечення нейромереж (нейропакети).

8. Тренінг із дисципліни « Інтелектуальний аналіз даних»

Тематика: Text Mining в системі управління знаннями

Порядок проведення:

1. Технологія аналізу тексту
2. Структурування та розбір тексту
3. Опис проблеми та варіант вирішення завдання
4. Text Mining для аналізу неструктурованих даних
5. Пошук, витяг, аналіз
6. Застосування пошукових систем
7. Автоматизація добування інформації та автоматизація аналітичних процедур
8. Основні технології Text Mining
9. Перспективи розвитку технологій ТМ
10. Пакети рішень Text Mining
11. Приклад роботи з Text Analysis у програмному продукті PolyAnalyst

Література: 1, 4, 8, 9,14, 20,24.

9. Засоби оцінювання та методи демонстрування результатів навчання

У процесі вивчення дисципліни «Інтелектуальний аналіз даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне опитування;
- залікове модульне тестування та опитування;
- реферати;
- презентації результатів виконання завдань;
- оцінювання результатів КПЗ;
- ректорська контрольна робота;
- тренінги;
- залік.

10. 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:

Заліковий модуль 1	Заліковий модуль 2 (РКР)	Заліковий модуль 3 (КПЗ)	Разом
30%	40%	30%	100 %
Виконання лабораторних робіт (6 робіт по 10 балів – 60 балів) Написання модульної роботи – 40 балів	Виконання лабораторних робіт (6 робіт по 10 балів – 60 балів) Написання ректорської контрольної роботи – 40 балів	Виконання завдань під час тренінгу (20 балів) Написання та захист КПЗ (80 балів)	100

Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90–100	відмінно	A (відмінно)
85–89	добре	B (дуже добре)
75–84		C (добре)
65–74	задовільно	D (задовільно)
60–64		E (достатньо)
35–59	незадовільно	FX (незадовільно з можливістю повторного складання)
1–34		F (незадовільно з обов'язковим повторним курсом)

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

№	Найменування	Номер теми
---	--------------	------------

1.	Мультимедійний проектор	1-11
2.	Проекційний екран	1-11
3.	Комунікаційне програмне забезпечення (Internet Explorer, Google Chrome, Firefox)	1-11
4.	Операційна система Windows, наявність доступу до мережі Internet	1-11
5.	Персональні комп'ютери	1-11
6.	Комунікаційне програмне забезпечення (Zoom) для проведення занять у режимі онлайн (за необхідності)	1-11
7.	Комунікаційна навчальна платформа (Moodle) для організації дистанційного навчання (за необхідності)	1-11
8.	Базове програмне забезпечення Microsoft Office	1-11
9.	Спеціалізоване програмне забезпечення Poly Analyst, Text Mining, Wizwhy, Deductor, мова програмування C++, Python, Matlab.	1-11

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Andrew Moore. Statistical Data Mining Tutorials [http://www.autonlab.org/tutorials/]
2. Великі дані. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/Великі_дані
3. Гаврилюк О. М., Круп'якевич В. В. Методи штучного інтелекту у видобуванні знань з баз даних. Науковий вісник НЛТУ України, 28(7), 2018, с.108-116. Посилання: http://nbuv.gov.ua/UJRN/Nv_nltu_2018_28
4. Big Data і блокчейн. [Електронний ресурс]. – Режим доступу: <https://forklog.com/big-data-i-blokchejn-proryv-v-oblasti-analiza-dannyh/>
5. 515K Hotel Reviews Data in Europe [Режим електронного доступу] https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe?select=Hotel_Reviews.csv
6. The Python Package Index (PyPI) is a repository of software for the Python programming language [Режим електронного доступу] <https://pypi.org/>
7. NumPy and Pandas Tutorial – Data Analysis with Python <https://cloudxlab.com/blog/numpy-pandas-introduction/>
8. PolyAnalyst 6.5 Technical capabilities and system requirements, Megaputer intelligence, 2019.
9. "Ukrainian Data Mining Group" - <http://www.udmg.com.ua/>
10. "Блог Data Mining в Україні" - <https://datamining.com.ua/>
11. "Data Mining: моделі та методи" - <http://study.uaitlab.com/datamining/>

12. "Data Science UA" - <https://datascienceua.com/>
13. Pierre Baldi and Søren Brunak. Chapter 6. Neural networks: applications. In *Bioinformatics: The Machine Learning Approach*. MIT press, 2020.
14. Machine Learning [Електронний ресурс] Автор brilliant.org
<https://brilliant.org/wiki/machine-learning/#techniques-in-machine-learning>
15. <https://www.python.org/about/apps/>
16. <https://www.kaggle.com/docs/notebooks>
17. <https://numpy.org/doc/stable/user/whatisnumpy.html>
18. Keras Models – Types and Examples [Електронний ресурс] Автор data-
19. flair.training <https://data-flair.training/blogs/keras-models/>
20. Fruits 360 [Електронний ресурс] 2020 Автор Mihai Oltean
<https://www.kaggle.com/moltean/fruits>
21. Наївний баєсів класифікатор - Вікіпедія. *Вікіпедія*.
URL: https://uk.wikipedia.org/wiki/%D0%9D%D0%B0%D1%97%D0%B2%D0%BD%D0%B8%D0%B9_%D0%B1%D0%B0%D1%94%D1%81%D1%96%D0%B2_%D0%BA%D0%BB%D0%B0%D1%81%D0%B8%D1%84%D1%96%D0%BA%D0%B0%D1%82%D0%BE%D1%80 (дата звернення: 30.05.2023).
22. 10 найкращих інструментів аналізу даних
URL: https://uk.myservername.com/10-best-data-analysis-tools#1_Microsoft_Power_BI (дата звернення: 30.05.2023).
23. RapidMiner studio - rapidminer documentation. *Home - RapidMiner Documentation*. URL: <https://docs.rapidminer.com/latest/studio/> (дата звернення: 1.08.2023).
24. Power BI Tutorial - Microsoft
URL: <https://learn.microsoft.com/uk-ua/power-bi/create-reports/desktop-excel-stunning-report> (дата звернення: 15.07.2023).