

Черкаський національний університет імені Богдана Хмельницького
ННІ ІНФОТЕХ
Кафедра прикладної математики

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«МЕТОДИ І ЗАСОБИ ОБРОБКИ ВЕЛИКИХ ДАНИХ»

1. Загальна інформація про курс

Мова викладання	Курс викладається українською мовою
Статус дисципліни	Обов'язкова
Викладачі	Богатирьов О.О., к.ф.-м.н., доцент, доцент кафедри прикладної математики та інформатики
Код класу /	
Корпоративна пошта / E-mail:	a_bogatyrev@vu.cdu.edu.ua
Затвердження та перегляд робочої навчальної програми	Розглянуто та затверджено на засіданні кафедри прикладної математики та інформатики 27.08.2025, протокол № 1

2. Анотація до курсу

Під час навчання дисципліни "Методи та засоби обробки великих даних" студенти зможуть здобути теоретичні знання та практичні навички роботи з великими даними, ознайомляться з основними поняттями, методами та засобами обробки великих даних. Вони навчатимуться виконувати розвідувальний аналіз даних, підготовку даних та здійснювати машинне навчання на великих даних.

Передбачено контроль якості отриманих знань у вигляді модульної контрольної роботи.

3. Мета та цілі курсу

Мета курсу - формування у студентів теоретичних знань і практичних умінь застосовувати методи і засоби збирання, зберігання, аналізу та обробки великих даних.

Завданнями вивчення навчальної дисципліни “Методи і засоби обробки великих даних” передбачено:

Знання

- методів та засобів агрегації та інтегрування великих даних різних типів;
- методів та засобів проектування сховищ для зберігання великих даних;
- методів, засобів та інструментів інтелектуальної обробки великих даних;
- методів та засобів візуалізації великих даних;
- методів та засобів отримання знання з великих даних.

Уміння

- використовувати інструментальні засоби інтеграції різнотипових великих даних;
- правильно обирати та використовувати методи та алгоритми для фільтрації, валідації та зберігання великих даних;
- правильно обирати найбільш інформативні способи візуалізації великих даних;
- отримувати знання шляхом аналізу великих даних;
- проектувати та розробляти прикладні інформаційні програми для аналізу великих даних та прийняття рішень на основі отриманої інформації.

4. Компетентності та очікувані результати навчання

Навчальна дисципліна “Методи і засоби обробки великих даних” забезпечує формування таких компетентностей, передбачених освітньою програмою підготовки магістрів спеціальності: F1 Прикладна математика.

Загальні компетентності:

ЗК03. Здатність оволодівати сучасними знаннями, формулювати та вирішувати проблеми.

ЗК07. Здатність спілкуватися та здійснювати професійну діяльність державною мовою та мовою країн ЄС.

Спеціальні (фахові, предметні) компетентності:

СК01. Здатність розв’язувати задачі й проблеми, які можуть бути формалізовані, потребують оновлення й інтеграції знань, зокрема в умовах неповної інформації.

СК07. Здатність проектувати та розробляти програмне забезпечення для розв’язування формалізованих задач, зокрема систем з великими обсягами даних.

СК08. Здатність формалізувати та будувати моделі даних або знань, одержувати релевантні знання з великих обсягів даних, володіти методами обробки експериментальних даних, обирати методи інтелектуального аналізу даних для розв’язання задач.

Згідно з вимогами освітньо-професійної програми, **результатами вивчення** дисципліни “Методи і засоби обробки великих даних” є такі:

РН01. Спілкуватися й комунікувати в межах професійних компетенцій однією з мов країн ЄС

РН02. Здійснювати збір, систематизацію та аналіз науково-технічної інформації з питань професійної діяльності.

РН05. Обґрунтовувати та за необхідності розробляти нові алгоритми і програмні засоби для розв’язання наукових та прикладних задач, застосовувати, модифікувати і досліджувати аналітичні та обчислювальні методи їх розв’язування.

РН07. Розв’язувати задачі комп’ютерного моделювання шляхом використання і розробки сучасних програмних засобів, зокрема методами розподіленого, паралельного та хмарного програмування.

РН09. Вміти аналізувати та проектувати системи з великими обсягами даних, застосовувати та адаптувати методи здобуття знань, методи оцінки та інтерпретації знайдених закономірностей.

РН10. Розробляти та застосовувати сучасні концепції машинного навчання та інтелектуального аналізу даних.

5. Обсяг і характеристика курсу

Найменування показників	Характеристика навчального курсу	
	денна форма навчання	заочна форма навчання
Освітня програма, спеціальність	Прикладна математика, F1 Прикладна математика	
Рік навчання	1	
Семестр вивчення	2	
обов'язкова /вибіркова	обов'язкова	
Кількість кредитів ЄКТС	4	
Загальний обсяг годин	120	
Кількість годин навчальних занять	40	
Лекційні заняття	12	
Практичні заняття	0	
Семінарські заняття	0	
Лабораторні заняття	24	
Самостійна та індивідуальна робота	84	
Форма підсумкового контролю	залік	

6. Пререквізити курсу

Для вивчення курсу студенти потребують базових знань з дисциплін “Програмування”, “Алгоритми та структури даних”, “Алгебра та геометрія”, “Математичний аналіз”, “Теорія ймовірностей та математична статистика”, “Методи оптимізації”.

7. Технічне забезпечення

Вивчення курсу потребує використання хмарних служб Google Cloud Platform, доступ до яких забезпечується через безкоштовні курси від компанії Google, що розташовані на платформі Coursera.

8. Політика курсу

Система вимог, які ставляться перед студентом:

- відвідування лекційних та лабораторних занять є обов'язковою складовою вивчення матеріалу;
- на лекції викладач користується власним презентаційним матеріалом; використовує гугл-диск для викладання матеріалу поточної лекції, додаткової інформації, завдань до лабораторних робіт та інше;
- питання на лекції задаються у відведений для цього час;
- звіти з лабораторних робіт завантажуються напередодні захисту; для захисту лабораторної роботи необхідно продемонструвати роботу відповідної завданню програми та відповісти на питання щодо програми та контрольні питання;
- додаткові бали виставляються за самостійне опанування теоретичним матеріалом та розв'язання задач;
- модульна контрольна роботи пишеться у вигляді тесту і проводиться без застосування допоміжних засобів (мобільні телефони, планшети та ін.); результат завантажується у файлі через гугл-форму до відповідної директорії гугл-диску;
- штрафні бали виставляються за несвоєчасну здачу лабораторних робіт без поважних причин; переписування модульної контрольної роботи. Кількість штрафних балів не більше 10.

9. Схема курсу

Тема, основні питання / завдання	Розподіл годин за темами та формам и занять (денна/за очна)	Форми та методи проведення	Література. Ресурси в інтернеті	Завдання для самостійної роботи, год	Форма контролю, бали
Лекція 1. Прийняття кращих рішень на основі даних. Розвідка даних 1. Мета збору даних - приймати на їх основі рішення 2. Багато схожих рішень 3. Роль аналітика даних та науковця з даних 4. Співпраця різних фахівців 5. Завантаження даних 6. Рекомендована архітектура даних	2	Лекція-візуалізація (з використанням презентації)	Основна: 1, 2 Додаткова: 1-7	1. Опрацювати лекційний матеріал (3 год) 2.* Виконати вправи 1-7 [2, с.96-102] (3,5 год)	Опитування на лекції Перевірка виконаних завдань (0-7 додаткових балів)

7. Розділення обчислень і зберігання даних 8. Вхідні дані 9. Архітектура Google Cloud 10. Звіт про якість даних 11. Ознайомлення з даними 12. Визначення проблем, які пов'язані з даними 13. Розв'зування проблем, які пов'язані з якістю даних 14. Поглиблене дослідження даних 15. Підготовка даних					
Лекція 2. Інформаційне навчання 1. Основна ідея 2. Основи 2.1. Дерева рішень 2.2. Ентропійна модель Шеннона 2.3. Інформаційний виграш 3. Стандартний підхід: алгоритм ID3 3.1. Працюючий приклад: Прогнозування розподілу рослинності 4. Розширення та варіації 4.1. Альтернативний вибір ознак та показники неоднорідності 4.2. Обробка неперервних описових ознак 4.3. Прогнозування цільових міток, які є неперервними 4.4. Ансамблі моделей	2	Лекція-візуалізація (з використанням презентації)	Основна: 2 Додаткова: 2-7	1. Опрацювати лекційний матеріал (3 год) 2.* Виконати вправи 1-5 [2, с.172-175] (2,5 год)	Опитування на лекції Перевірка виконаних завдань (0-5 додаткових балів)
Лекція 3. Навчання на основі схожості 1. Основна ідея 2. Основи 2.1. Простір ознак 2.2. Вимірювання подібності за допомогою метрики відстані 3. Стандартний підхід: алгоритм найближчих сусідів 3.1. Працюючий приклад	2	Лекція-візуалізація (з використанням презентації)	Основна: 2 Додаткова: 2-7	1. Опрацювати лекційний матеріал (3 год) 2.* Самостійно опрацювати теоретичні питання теми: - 4.1. Обробка зашумлених даних - 4.2. Ефективний пошук у пам'яті - 4.3. Нормалізація даних - 4.5. Інші міри подібності	Опитування на лекції Опитування на лабораторній роботі (0-4 додаткових балів)

4. Розширення та варіації 4.4. Прогнозування неперервних цільових ознак 4.6. Відбір ознак				(4 год) 3.* Виконати вправи 1-3 [2, с.236-239] (1,5 год)	Перевірка виконаних завдань (0-3 додаткових балів)
Лекція 4. Ймовірнісне навчання 1. Основна ідея 2. Основи 2.1. Теорема Баєса 2.2. Байєсівське прогнозування 2.3. Умовна незалежність і факторизація 3. Стандартний підхід: наївна модель Баєса 3.1. Працюючий приклад 4. Розширення та варіації 4.1. Згладжування 4.2. Неперервні ознаки на основі функції густини ймовірностей 4.3. Неперервні ознаки: групування	2	Лекція-візуалізація (з використанням презентації)	Основна: 2 Додаткова: 2-7	1. Опрацювати лекційний матеріал (3 год) 2* Самостійно опрацювати теоретичні питання теми: - 4.4. Байєсівські мережі (1 год) 3* Виконати вправи 1-4 [2, с.305-307] (2 год)	Опитування на лекції Опитування на лабораторній роботі (0-1 додаткових балів) Перевірка виконаних завдань (0-4 додаткових балів)
Лекція 5. Навчання на основі помилок 1. Основна ідея 2. Основи 2.1. Проста лінійна регресія 2.2. Вимірювання помилок 2.3. Поверхні помилок 3. Стандартний підхід: множинна лінійна регресія з градієнтним спуском 3.1. Багатовимірна лінійна регресія 3.2. Градієнтний спуск 3.3. Вибір темпів навчання та початкових вагових коефіцієнтів 3.4. Працюючий приклад 4. Розширення та варіації 4.1. Інтерпретація багатовимірних лінійних регресійних моделей 4.4. Робота з категоріальними	2	Лекція-візуалізація (з використанням презентації)	Основна: 2 Додаткова: 2-7	1. Опрацювати лекційний матеріал (3 год) 2* Самостійно опрацювати теоретичні питання теми: - 4.2. Налаштування швидкості навчання за допомогою затухання вагових коефіцієнтів - 4.3. Обробка категоріальних описових ознак - 4.5. Моделювання нелінійних залежностей - 4.6. Мультиномінальна логістична регресія (4 год) 3* Виконати вправи 1-4 [2, с.371-374] (2 год)	Опитування на лекції Опитування на лабораторній роботі (0-4 додаткових балів) Перевірка виконаних завдань (0-4 додаткових

ознаками цілі: Логістична регресія 4.7. Машина опорних векторів (SVM)					балів)
Лекція 6. Глибоке навчання 1. Основна ідея 2. Основи 2.1. Штучні нейрони 2.2. Штучні нейронні мережі 2.3. Нейронні мережі як матричні операції 2.4. Навіщо потрібні нелінійні функції активації? 2.5. Чому важлива глибина мережі? 3. Стандартний підхід: алгоритм зворотного поширення помилки та градієнтний спуск 3.1. Алгоритм зворотного поширення помилки: Загальна структура алгоритму 3.3. Алгоритм 3.5. Працюючий приклад: Використання алгоритму зворотного поширення помилки для навчання мережі прямого поширення для задачі регресії 4. Розширення та варіації 4.3. Обробка цільових категорійних ознак цілі: вихід softmax та крос-ентропійна функція втрат 4.5. Згорткові нейронні мережі 4.6. Рекурентні нейронні мережі та мережі з довгою короткодіючою пам'яттю	2	Лекція-візуалізація (з використанням презентації)	Основна: 2 Додаткова: 2-7	1. Опрацювати лекційний матеріал (3 год) 2* Самостійно опрацювати теоретичні питання теми: - 3.2. Метод зворотного поширення помилки: Розповсюдження градієнтів похибок - 3.3. Метод зворотного поширення помилки: Оновлення ваг у мережі - 4.1. Зникаючі градієнти та ReLU - 4.2. Ініціалізація вагових коефіцієнтів та нестабільні градієнти - 4.4. Раннє припинення та відсів нейронів: Як запобігти перенавчанню (5 год) 3* Виконати вправи 1-7 [2, с.524-529] (3,5 год)	Опитування на лекції Опитування на лабораторній роботі (0-5 додаткових балів балів) Перевірка виконаних завдань (0-7 додаткових балів)
Лабораторне заняття 1. Основи роботи з великими даними в Google Cloud - 1 <i>Великі дані та машинне навчання в Google Cloud</i>	6	Практична робота з Google Cloud	Основна: 1, 3	Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи	Перевірка виконання

1. Виконати завдання <u>лабораторної роботи</u> “Дослідження публічного набору даних BigQuery” (0-3,5 балів) 2. Пройти <u>тест</u> (6 запитань) (0-1,5 балів) <i>Інженерія даних для потокової передачі даних</i> 1. Виконати завдання <u>лабораторної роботи</u> “Створення потокового конвеєра даних для інформаційної панелі в режимі реального часу за допомогою Dataflow” (0-3,5 балів) 2. Пройти <u>тест</u> (5 запитань) (0-1,25 бали) <i>Великі дані за допомогою BigQuery</i> 1. Виконати завдання <u>лабораторної роботи</u> “Прогнозування покупок відвідувачів за допомогою BigQuery ML” (0-3,5 балів) 2. Пройти <u>тест</u> (8 запитань) (0-2 балів) <i>Варіанти машинного навчання в Google Cloud</i> 1. Пройти <u>тест</u> (5 запитань) (0-1,25 балів) <i>Прогнозування кредитного ризику за допомогою AutoML</i> 1. Виконати завдання <u>лабораторної роботи</u> “Прогнозування кредитного ризику за допомогою AutoML” (0-3,5 балів) 2. Пройти <u>тест</u> (6 запитань) (0-1,5 балів)		Platform Онлайн тестування Практична робота з Google Cloud Platform Онлайн тестування Практична робота з Google Cloud Platform Онлайн тестування Онлайн тестування Практична робота з Google Cloud Platform Онлайн тестування	Додаткова: 2 Основна: 1, 3 Додаткова: 2 Основна: 1, 3 Додаткова: 2 Основна: 1, 3 Додаткова: 2 Основна: 1, 3 Додаткова: 2	(4 год) Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год) Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год) Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год) Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год)	лабораторної роботи та результатів тестування (0-5 балів) Перевірка виконання лабораторної роботи та результатів тестування (0-4,75 балів) Перевірка виконання лабораторної роботи та результатів тестування (0-5,5 балів) Перевірка результатів тестування (0-1,25 балів) Перевірка виконання лабораторної роботи та результатів тестування
--	--	--	---	--	---

					(0-5 балів)
Лабораторне заняття 2. Основи машинного навчання в Google Cloud - 1 <i>Організаційні ноу-хау для впровадження машинного навчання</i> 1. Пройти <u>тест</u> “Як Google робить ML” (5 запитань) (0-1,25 балів) 2. Пройти <u>тест</u> “Що означає бути першими в галузі штучного інтелекту” (4 запитання) (0-1 балів)	1	Онлайн тестування	Основна: 1, 4 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (4 год)	Перевірка результатів тестування (0-2,25 балів)
Лабораторне заняття 3. Основи машинного навчання в Google Cloud - 2 <i>Розробка машинного навчання з Vertex AI</i> 1. Виконати завдання <u>лабораторної роботи</u> “Використання набору даних зображень для навчання AutoML моделі” (0-3,5 балів) 2. Виконати завдання <u>лабораторної роботи</u> “Навчання моделі класифікації відео за допомогою AutoML” (0-3,5 балів) 3. Пройти <u>тест</u> “Розробка машинного навчання з Vertex AI” (7 запитань) (0-2,75 балів)	4,5	Практична робота з Google Cloud Platform Онлайн тестування	Основна: 1, 4 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (6 год)	Перевірка виконання лабораторної роботи та результатів тестування (0-9,75 балів)
Лабораторне заняття 4. Основи машинного навчання в Google Cloud - 3 <i>Розробка машинного навчання з ноутбуками Vertex</i> 1. Пройти <u>тест</u> “Розробка машинного навчання з ноутбуками Vertex” (6 запитань) (0-1,5 балів) 2. Пройти <u>тест</u> “Кращі практики впровадження машинного навчання на Vertex AI” (4 запитання)	1,5	Онлайн тестування	Основна: 1, 4 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (6 год)	Перевірка результатів тестування (0-3,5 балів)

(0-1 балів) 3. Пройти <u>тест</u> “Відповідальна розробка штучного інтелекту” (4 запитання) (0-1 балів)					
Лабораторне заняття 5. Робота з машинним навчанням в Google Cloud - 1 <i>Покращення даних за допомогою дослідницького аналізу даних</i> 1. Виконати завдання <u>лабораторної роботи</u> “Покращення якості даних” (0-3,5 балів) 2. Виконати завдання <u>лабораторної роботи</u> “Розвідувальний аналіз даних за допомогою Python та BigQuery” (0-3,5 балів) 3. Пройти <u>тест</u> “Покращення даних за допомогою дослідницького аналізу даних” (5 запитань) (0-1,25 бали)	7	Практична робота з Google Cloud Platform Онлайн тестування	Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (6 год)	Перевірка виконання лабораторної роботи та результатів тестування (0-8,25 балів)
<i>Машинне навчання на практиці</i> 1. Виконати завдання <u>лабораторної роботи</u> “Вступ до лінійної регресії” (0-3,5 балів) 2. Пройти <u>тест</u> “Машинне навчання на практиці” (5 запитань) (0-1,25 балів)		Практична робота з Google Cloud Platform Онлайн тестування	Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год)	Перевірка виконання лабораторної роботи та результатів тестування (0-4,75 балів)
Лабораторне заняття 6. Робота з машинним навчанням в Google Cloud - 2 <i>Навчання AutoML-моделей за допомогою Vertex AI</i> 1. Пройти <u>тест</u> “Навчання AutoML-моделей за допомогою Vertex AI” (9 запитань) (0-2,25 балів)	2	Онлайн тестування	Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год)	Перевірка результатів тестування (0-2,25 балів)
<i>Машинне навчання за допомогою BigQuery</i>		Практична робота з	Основна:	Ознайомитися з відеоматеріалом стосовно	Перевірка

<i>ML</i> 1. Виконати завдання <u>лабораторної роботи</u> “Використання BigQuery ML для прогнозування ваги пінгвінів” (0-3,5 балів) 2. Пройти <u>тест</u> “Машинне навчання з BigQuery: розроблюйте ML-моделі там, де живуть ваші дані” (5 запитань) (0-1,25 балів)		Google Cloud Platform Онлайн тестування	1, 5 Додаткова: 2	виконання лабораторної роботи (4 год)	виконання лабораторної роботи та результатів тестування (0-4,75 балів)
Лабораторне заняття 7. Робота з машинним навчанням в Google Cloud - 3 <i>Оптимізація</i> 1. Пройти <u>тест</u> “Оптимізація” (5 запитань) (0-1,25 балів) <i>Узагальнення та вибірка</i> 1. Пройти <u>тест</u> “Узагальнення та вибірка” (5 запитань) (0-1,25 балів)	2	Онлайн тестування Онлайн тестування	Основна: 1, 5 Додаткова: 2 Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год) Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год)	Перевірка результатів тестування (0-1,25 балів) Перевірка результатів тестування (0-1,25 балів)
Модульна контрольна робота				Повторити навчальний матеріал (5 год)	Перевірка результатів тестування (0-40 балів)
Всього годин	120				
Лекцій	12				
Лабораторних занять	24				60 балів
Самостійна робота	84				

10. Система оцінювання та вимоги

Навчальні досягнення студентів оцінюються за 100-бальною шкалою Університету, чотирибальною шкалою (5 «відмінно», 4 «добре», 3 «задовільно», 2 «незадовільно»), і шкалою оцінок ЄКТС. На поточний контроль відводиться 60 балів, на підсумковий контроль (екзамен) – 40 балів.

Кількість балів за роботу з теоретичним матеріалом, на лабораторних заняттях залежить від дотримання таких вимог:

- своєчасність виконання навчальних завдань;

- повний обсяг їх виконання;
- якість виконання навчальних завдань;
- самостійність виконання;
- творчий підхід у виконанні завдань.

Рейтинг студента формується з:

1. Балів за виконання та захист лабораторних робіт.
2. Штрафних балів за несвоєчасне виконання.
3. Балів за проходження тестів під час виконання лабораторних робіт.
4. Додаткових балів за розв'язання теоретичних задач (0-44 додаткових балів).
5. Балів, отриманих за виконання модульної контрольної роботи.

Студент, який не підготувався до лабораторного заняття, не отримує жодного бала. Поточну заборгованість, пов'язану з невідповідністю або недостатньою підготовленістю до навчальних занять, студент повинен ліквідувати шляхом виконання у визначений термін завдань, передбачених програмою. За завдання, виконані невчасно, нараховуються штраф у вигляді зарахування вдвічі меншої кількості балів.

Студенти, які за результатами поточного контролю набрали менше 20 балів, вважаються такими, що мають академічну заборгованість, ліквідація якої є обов'язковою. Студенти, які не мають академічної заборгованості за результатами поточного контролю, допускаються до екзамену.

11. Критерії оцінювання успішності навчання

Система рейтингових балів та критерії оцінювання

Лабораторні завдання

- «відмінно», повна відповідь на питання під час захисту (не менше ніж 90% потрібної інформації), повне виконання завдання лабораторної роботи – 3,5 бали;
- «добре», достатньо повна відповідь на питання під час захисту (не менше ніж 75% потрібної інформації), повне виконання завдання лабораторної роботи – 2,5 бали;
- «задовільно», неповна відповідь на питання під час захисту (не менше ніж 60% потрібної інформації), незначні помилки у виконанні завдання лабораторної роботи – 1,5 бали;

- «незадовільно», незадовільна відповідь та/або значні помилки у виконання завдання лабораторної роботи – 0 балів.

Додаткові бали

Додаткові бали нараховуються студенту за самостійне вивчення окремих параграфів підручника та за розв'язання задач з нього:

- 1 бал за кожен опрацьований параграф (14 параграфів)
- 1 бал за кожен розв'язану задачу (30 задач)

Модульна контрольна робота

Завданням модульної контрольної роботи є комплексна діагностика результатів навчання, глибини засвоєння студентом програмного матеріалу з навчальної дисципліни, логіки та взаємозв'язків між окремими його змістовими модулями, здатності до творчого використання набутих знань.

Екзамен проводиться у формі тестування. Для виконання екзаменаційної роботи виділяється 2 години (120 хвилин). У роботі 28 завдань. Вони розділені на 3 частини.

Частина А містить 20 завдань (A1-A20) з розділів “Розвідка даних” (A1-A4), “Інформаційне навчання” (A5, A6), “Навчання на основі схожості” (A7, A8), “Ймовірнісне навчання” (A9, A10), “Навчання на основі помилок” (A11, A12), “Глибоке навчання” (A13, A14), “Основи роботи з великими даними в Google Cloud” (A15, A16), “Основи машинного навчання в Google Cloud” (A17, A18), “Робота з машинним навчанням в Google Cloud” (A19, A20). До кожного з них дані 4 варіанти відповіді, з яких лише один вірний. Кожне завдання цієї частини оцінюється в 1 бал.

Частина В містить 6 більш складних завдань (B1-B6) з розділів “Розвідка даних” (B1), “Інформаційне навчання” (B2), “Навчання на основі схожості” (B3), “Ймовірнісне навчання” (B4), “Навчання на основі помилок” (B5), “Глибоке навчання” (B6). Завдання даної частини є питаннями відкритого типу і оцінюються в 2 бали.

Частина С містить 2 найскладніших завдання (C1, C2), при виконанні яких потрібно записати повний розв'язок. Кожне завдання оцінюється в 4 бали.

12. Перелік питань /завдань для підсумкового контролю

1. Мета збору даних - приймати на їх основі рішення
2. Роль аналітика даних та науковця з даних
3. Завантаження даних
4. Рекомендована архітектура даних
5. Розділення обчислень і зберігання даних
6. Архітектура Google Cloud
7. Звіт про якість даних

8. Ознайомлення з даними
9. Визначення проблем, які пов'язані з даними
10. Розв'язування проблем, які пов'язані з якістю даних
11. Поглиблене дослідження даних
12. Підготовка даних
13. Основна ідея інформаційного навчання
14. Основи інформаційного навчання. Дерева рішень. Ентропійна модель Шеннона. Інформаційний виграш
15. Стандартний підхід інформаційного навчання: алгоритм ID3. Приклад
16. Основна ідея навчання на основі схожості
17. Основи навчання на основі схожості. Простір ознак. Вимірювання подібності за допомогою метрики відстані
18. Стандартний підхід навчання на основі схожості: алгоритм найближчих сусідів. Приклад
19. Основна ідея ймовірнісного навчання
20. Основи ймовірнісного навчання. Теорема Баєса. Байєсівське прогнозування. Умовна незалежність і факторизація
21. Стандартний підхід ймовірнісного навчання: наївна модель Баєса. Приклад
22. Основна ідея навчання на основі помилок
23. Основи навчання на основі помилок. Проста лінійна регресія. Вимірювання помилок. Поверхні помилок
24. Стандартний підхід навчання на основі помилок: множинна лінійна регресія з градієнтним спуском. Багатовимірна лінійна регресія
25. Навчання на основі помилок. Градієнтний спуск. Вибір темпів навчання та початкових вагових коефіцієнтів
26. Навчання на основі помилок. Приклад
27. Основна ідея глибокого навчання
28. Основи глибокого навчання. Штучні нейрони. Штучні нейронні мережі. Нейронні мережі як матричні операції. Нелінійні функції активації
29. Стандартний підхід глибокого навчання: алгоритм зворотного поширення помилки та градієнтний спуск. Загальна структура алгоритму
30. Глибоке навчання. Приклад: використання алгоритму зворотного поширення помилки для навчання мережі прямого поширення для задачі регресії
31. *Великі дані та машинне навчання в Google Cloud*
32. *Інженерія даних для потокової передачі даних*
33. *Робота з великими даними за допомогою BigQuery*
34. *Варіанти машинного навчання в Google Cloud*
35. *Організаційні ноу-хау для впровадження машинного навчання*
36. *Розробка машинного навчання з Vertex AI*
37. *Розробка машинного навчання з ноутбуками Vertex*
38. *Покращення даних за допомогою дослідницького аналізу даних*
39. *Особливості застосування машинного навчання на практиці*
40. *Навчання AutoML-моделей за допомогою Vertex AI*
41. *Машинне навчання за допомогою BigQuery ML*

13. Список рекомендованої літератури / інтернет-ресурси / нормативні документи

Основна

1. V.Lakshmanan and J.Tigani. Google BigQuery: The Guide. Data Warehousing, Analytics, and Machine Learning at Scale, 2019
2. J.Kelleher, B.Namee, A.D'Arcy. Fundamentals of Machine Learning for Predictive Data Analytics, Second Edition, 2020
3. Course “Google Cloud Big Data and Machine Learning Fundamentals”
<https://www.coursera.org/programs/chierkas-kii-natsional-nii-universitiet-imieni-boghdana-khmiel-nits-kogho/skills/big-data?collectionId=skill~big-data&productId=CD090o2qEeuLhxLgyXhFTw&productType=s12n&showMiniModal=true&source=browse>
4. Course “How Google does Machine Learning”
<https://www.coursera.org/programs/chierkas-kii-natsional-nii-universitiet-imieni-boghdana-khmiel-nits-kogho/skills/big-data?collectionId=skill~big-data&productId=CD090o2qEeuLhxLgyXhFTw&productType=s12n&showMiniModal=true&source=browse>
5. Course “Launching into Machine Learning”
<https://www.coursera.org/programs/chierkas-kii-natsional-nii-universitiet-imieni-boghdana-khmiel-nits-kogho/skills/big-data?collectionId=skill~big-data&productId=CD090o2qEeuLhxLgyXhFTw&productType=s12n&showMiniModal=true&source=browse>

Додаткова

1. V.Lakshmanan and J.Tigani. Google BigQuery: The Guide. Data Warehousing, Analytics, and Machine Learning at Scale, 2019
2. A.Marrandino. Machine Learning with BigQuery ML, 2021
3. E.Bisong. Building Machine Learning and Deep Learning Models on Google Cloud Platform, 2019
4. S.Raschka, V.Mirjalili. Python Machine Learning, Third Edition, 2019
5. F.Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Third Edition, 2023
6. A.Müller, S.Guido. Introduction to Machine Learning with Python, 2016
7. A.Ng, K.Soo. NumSense! Data Science for the Layman, 2017