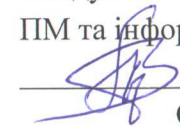


Черкаський національний університет імені Богдана Хмельницького  
Навчально-науковий інститут інформаційних та освітніх технологій  
Кафедра прикладної математики та інформатики

ЗАТВЕРДЖЕНО

Завідувач кафедри

ПМ та інформатики



О.В.Піскун

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
“Методи і засоби обробки великих даних”

1. Загальна інформація про курс

Назва курсу, мова викладання	Методи і засоби обробки великих даних Курс викладається українською з елементами англійської мови.
Статус дисципліни	Обов'язкова
Викладачі	Богатирьов Олександр Олегович, канд. фіз.-мат. 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 28.08.2024, 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 навчаються 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: 126 Інформаційні системи та технології.

**Інтегральна компетентність:** здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у сфері обробки великих даних, що передбачають застосування відповідних методів та засобів.

**Загальні компетентності:**

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність спілкуватися іноземною мовою.

ЗК05. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ЗК06. Здатність застосовувати знання у практичних ситуаціях.

**Спеціальні (фахові, предметні) компетентності:**

СК01. Здатність розробляти та застосувати ІСТ, необхідні для розв'язання стратегічних і поточних задач.

СК04. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.

СК05. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.

СК08. Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями, у першу чергу, орієнтованими на роботу у локальній та глобальній мережі.

СК09. Здатність розв'язувати практичні завдання, використовуючи знання систем хмарних обчислень, архітектури та стандартів комунікаційних засобів розподілених обчислень, концепцій паралельної обробки інформації.

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

РН01. **Відшукувати** необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, **аналізувати** та **оцінювати** цю інформацію.

РН02. Вільно **спілкуватись** державною та іноземною мовами в науковій, виробничій та соціально-суспільній сферах діяльності.

РН08. **Розробляти** моделі інформаційних процесів та систем різного класу, **використовувати** методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.

РН09. **Розробляти** і **використовувати** сховища даних, **здійснювати** аналіз даних для підтримки прийняття рішень.

РН11. **Розв'язувати** задачі цифрової трансформації у нових або невідомих середовищах на основі спеціалізованих концептуальних знань, що включають сучасні наукові здобутки у сфері інформаційних технологій, досліджень та інтеграції знань з різних галузей.

РН12. **Демонструвати** знання сучасного рівня технологій інформаційних систем з метою їх запровадження у професійній діяльності; **знати** принципи функціонування та технології віртуалізації серверних систем, архітектури, та стандарти комунікаційних засобів розподілених обчислень; **вміти** розробляти програмне забезпечення різного рівня складності, що входить до складу інформаційних систем та технологій, при розв'язанні прикладних науково-виробничих задач і задач бізнесу.

## 5. Обсяг і характеристика курсу

Найменування показників	Характеристика навчального курсу	
	денна форма навчання	заочна форма навчання
Освітня програма, спеціальність	Веб-орієнтовані інформаційні системи, 126 Інформаційні системи та технології	
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	14	
Практичні заняття	0	
Семінарські заняття	0	
Лабораторні заняття	26	
Самостійна та індивідуальна робота	80	
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
<b>Лекція 1. Прийняття кращих рішень на основі даних</b> 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 7. Розділення обчислень і зберігання даних 8. Вхідні дані 9. Архітектура Google Cloud	2	Лекція-візуалізація (з використанням презентації)	Основна: 1 Додаткова: 1	1. Опрацювати лекційний матеріал (3 год)	Опитування на лекції
<b>Лекція 2. Розвідка даних</b> 1. Звіт про якість даних	2	Лекція-візуалізація (з використанням	Основна: 2	1. Опрацювати лекційний матеріал (3 год)	Опитування на лекції

<p>2. Ознайомлення з даними</p> <p>3. Визначення проблем, які пов'язані з даними</p> <p>4. Розв'язування проблем, які пов'язані з якістю даних</p> <p>5. Поглиблене дослідження даних</p> <p>6. Підготовка даних</p>		<p>презентації)</p>	<p>Додаткова: 2-7</p>	<p>2.* Виконати вправи 1-7 [2, с.96-102] (3,5 год)</p>	<p>Перевірка виконаних завдань (0-7 додаткових балів)</p>
<p><b>Лекція 3. Інформаційне навчання</b></p> <p>1. Основна ідея</p> <p>2. Основи</p> <p>    2.1. Дерева рішень</p> <p>    2.2. Ентропійна модель Шеннона</p> <p>    2.3. Інформаційний виграш</p> <p>3. Стандартний підхід: алгоритм ID3</p> <p>    3.1. Працюючий приклад: Прогнозування розподілу рослинності</p> <p>4. Розширення та варіації</p> <p>    4.1. Альтернативний вибір ознак та показники неоднорідності</p> <p>    4.2. Обробка неперервних описових ознак</p> <p>    4.3. Прогнозування цільових міток, які є неперервними</p> <p>    4.4. Ансамблі моделей</p>	<p>2</p>	<p>Лекція-візуалізація (з використанням презентації)</p>	<p>Основна: 2 Додаткова: 2-7</p>	<p>1. Опрацювати лекційний матеріал (3 год)</p> <p>2.* Виконати вправи 1-5 [2, с.172-175] (2,5 год)</p>	<p>Опитування на лекції</p> <p>Перевірка виконаних завдань (0-5 додаткових балів)</p>
<p><b>Лекція 4. Навчання на основі схожості</b></p> <p>1. Основна ідея</p> <p>2. Основи</p> <p>    2.1. Простір ознак</p> <p>    2.2. Вимірювання подібності за допомогою метрики відстані</p> <p>3. Стандартний підхід: алгоритм найближчих сусідів</p> <p>    3.1. Працюючий приклад</p> <p>4. Розширення та варіації</p> <p>    4.4. Прогнозування неперервних цільових ознак</p> <p>    4.6. Відбір ознак</p>	<p>2</p>	<p>Лекція-візуалізація (з використанням презентації)</p>	<p>Основна: 2 Додаткова: 2-7</p>	<p>1. Опрацювати лекційний матеріал (3 год)</p> <p>2.* Самостійно опрацювати теоретичні питання теми:</p> <p>- 4.1. Обробка зашумлених даних</p> <p>- 4.2. Ефективний пошук у пам'яті</p> <p>- 4.3. Нормалізація даних</p> <p>- 4.5. Інші міри подібності (4 год)</p> <p>3.* Виконати вправи 1-3 [2, с.236-239] (1,5 год)</p>	<p>Опитування на лекції</p> <p>Опитування на лабораторній роботі (0-4 додаткових балів)</p> <p>Перевірка виконаних</p>

					завдань (0-3 додаткових балів)
<b>Лекція 5. Ймовірнісне навчання</b> 1. Основна ідея 2. Основи <ul style="list-style-type: none"> <li>2.1. Теорема Баєса</li> <li>2.2. Байєсівське прогнозування</li> <li>2.3. Умовна незалежність і факторизація</li> </ul> 3. Стандартний підхід: наївна модель Баєса <ul style="list-style-type: none"> <li>3.1. Працюючий приклад</li> </ul> 4. Розширення та варіації <ul style="list-style-type: none"> <li>4.1. Згладжування</li> <li>4.2. Неперервні ознаки на основі функції густини ймовірностей</li> <li>4.3. Неперервні ознаки: групування</li> </ul>	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)
<b>Лекція 6. Навчання на основі помилок</b> 1. Основна ідея 2. Основи <ul style="list-style-type: none"> <li>2.1. Проста лінійна регресія</li> <li>2.2. Вимірювання помилок</li> <li>2.3. Поверхні помилок</li> </ul> 3. Стандартний підхід: множинна лінійна регресія з градієнтним спуском <ul style="list-style-type: none"> <li>3.1. Багатовимірна лінійна регресія</li> <li>3.2. Градієнтний спуск</li> <li>3.3. Вибір темпів навчання та початкових вагових коефіцієнтів</li> <li>3.4. Працюючий приклад</li> </ul> 4. Розширення та варіації <ul style="list-style-type: none"> <li>4.1. Інтерпретація багатовимірних лінійних регресійних моделей</li> <li>4.4. Робота з категоріальними ознаками цілі: Логістична регресія</li> <li>4.7. Машини опорних векторів (SVM)</li> </ul>	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 балів)
<b>Лекція 7. Глибоке навчання</b>	2	Лекція-візуалізація	Основна:	1. Опрацювати лекційний матеріал	Опитування на

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



<p>2. Пройти <u>тест</u> (6 запитань) (0-1,5 балів)</p> <p><i>Інженерія даних для потокової передачі даних</i></p> <p>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)</p> <p>2. Пройти <u>тест</u> (5 запитань) (0-1,25 бали)</p> <p><i>Великі дані за допомогою BigQuery</i></p> <p>1. Виконати завдання <u>лабораторної роботи</u> “Прогнозування покупок відвідувачів за допомогою BigQuery ML” (0-3,5 балів)</p> <p>2. Пройти <u>тест</u> (8 запитань) (0-2 балів)</p>		<p>Практична робота з Google Cloud Platform Онлайн тестування</p> <p>Практична робота з Google Cloud Platform Онлайн тестування</p>	<p>Основна: 1, 3 Додаткова: 2</p> <p>Основна: 1, 3 Додаткова: 2</p>	<p>Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год)</p> <p>Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год)</p>	<p>(0-5 балів)</p> <p>Перевірка виконання лабораторної роботи та результатів тестування (0-4,75 балів)</p> <p>Перевірка виконання лабораторної роботи та результатів тестування (0-5,5 балів)</p>
<p><b>Лабораторне заняття 2. Основи роботи з великими даними в Google Cloud - 2</b></p> <p><i>Варіанти машинного навчання в Google Cloud</i></p> <p>1. Пройти <u>тест</u> (5 запитань) (0-1,25 балів)</p> <p><i>Прогнозування кредитного ризику за допомогою AutoML</i></p> <p>1. Виконати завдання <u>лабораторної роботи</u> “Прогнозування кредитного ризику за допомогою AutoML” (0-3,5 балів)</p> <p>2. Пройти <u>тест</u> (6 запитань) (0-1,5 балів)</p>	3	<p>Онлайн тестування</p> <p>Практична робота з Google Cloud Platform Онлайн тестування</p>	<p>Основна: 1, 3 Додаткова: 2</p> <p>Основна: 1, 3 Додаткова: 2</p>	<p>Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год)</p> <p>Ознайомитися з відеоматеріалом стосовно виконання лабораторної роботи (4 год)</p>	<p>Перевірка результатів тестування (0-1,25 балів)</p> <p>Перевірка виконання лабораторної роботи та результатів тестування (0-5 балів)</p>
<p><b>Лабораторне заняття 3. Основи машинного навчання в Google Cloud - 1</b></p>	1				

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

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

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)					результатів тестування (0-4,75 балів)
<b>Лабораторне заняття 8. Робота з машинним навчанням в Google Cloud - 3</b>					
<i>Оптимізація</i> 1. Пройти <u>тест</u> “Оптимізація” (5 запитань) (0-1,25 балів)	1	Онлайн тестування	Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год)	Перевірка результатів тестування (0-1,25 балів)
<i>Узагальнення та вибірка</i> 1. Пройти <u>тест</u> “Узагальнення та вибірка” (5 запитань) (0-1,25 балів)		Онлайн тестування	Основна: 1, 5 Додаткова: 2	Ознайомитися з відеоматеріалом стосовно лабораторної роботи (2 год)	Перевірка результатів тестування (0-1,25 балів)
<b>Модульна контрольна робота</b>				Повторити навчальний матеріал (5 год)	Перевірка результатів тестування (0-40 балів)
<b>Всього годин</b>	<b>120</b>				
Лекцій	<b>14</b>			<b>21 год</b>	
Лабораторних занять	<b>26</b>			<b>54 год</b>	<b>60 балів</b>
Самостійна робота	<b>80</b>				

### 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 кожну 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 вииграш
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. Багатовимірною лінійною регресією
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*

### 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-univiersitiet-imieni-boghdana-khmiel-nits-kogho/skills/big-data?collectionId=skill~big-data&productId=CD090o2qEeuLhXhFTw&productType=s12n&showMiniModal=true&source=browse>

5. Course “Launching into Machine Learning”

<https://www.coursera.org/programs/chierkas-kii-natsional-nii-univiersitiet-imieni-boghdana-khmiel-nits-kogho/skills/big-data?collectionId=skill~big-data&productId=CD090o2qEeuLhXhFTw&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