Дніпровський національний університет імені Олеся Гончара

«ЗАТВЕРДЖУЮ»
Голова науково-методичної ради
Факультету прикладної математики
_____ Лілія БОЖУХА
« 28 » червня 2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ОК 2.29 Інформаційні системи аналізу даних великого обсягу

шифр із ОПП і повна назва навчальної дисципліни

для здобувачів вищої освіти

рівень вищої освіти перший (бакалаврський)

галузь знань 12 Інформаційні технології

спеціальність 126 Інформаційні системи та технології

спеціалізація

освітня програма Інформаційні системи та технології

рік набору 2021/2022 форма навчання денна термін навчання 3 роки 10 місяців

вид дисципліни *обов'язкова*

Розробник (-и)
Ємел'яненко Т.Г., доцент, к.т.н., доцент _____
підпис

Погоджено гарант ОП _____ Тетяна ЄМЕЛ'ЯНЕНКО
підпис

Робоча програма схвалена на засіданні кафедри математичного забезпечення ЕОМ

Протокол від « 21 » червня 2024 року № 14

Ухвалено на засіданні науково-методичної ради факультету прикладної математики
Протокол від « 28 » червня 2024 року № 15

Дніпро
2024

Опис навчальної дисципліни

Навчальний рік (роки*) викладання дисципліни	Курс	Семестр	Підсумковий контроль				Індивіду-альні завдання		Кредитів ECTS	Обсяг роботи студента (години)						
											аудиторні					
			екзамен	диф.залік	залік	курсова робота	форма	кількість		всього	всього аудиторних	лекції	практичні заняття	семінарські заняття	лабораторні заняття	самостійна робота
2024/25	4	8	8				кмр	1	5,0	150	64	32	8		24	86

1. Мета дисципліни

Метою навчальної дисципліни є формування у студентів теоретичних знань і практичних здатностей застосовувати методи і засоби збирання, зберігання, аналізу та обробки великих даних.

Вивчення дисципліни забезпечує формування компетентностей за ОП:

ФК12 Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).
ФК16 Здатність обґрунтовано обирати, застосовувати і реалізовувати інформаційні технології інтелектуального аналізу даних та підходи до побудови штучних нейронних мереж.

2. Попередні вимоги до опанування навчальної дисципліни.

Необхідно знати фундаментальні принципи розробки програмного забезпечення, структури даних та алгоритмів, вищої математики, теорії ймовірностей та математичної статистики, статистичних методів аналізу даних, штучного інтелекту.

3. Результати навчання за дисципліною та їх співвідношення із програмними результатами навчання.

№	Результати навчання за дисципліною	Програмні результати навчання за ОП	Номери тем
1	Здійснювати системний аналіз об'єктів проєктування інформаційних систем для роботи з великими даними, враховуючи історію розвитку технологій, виклики та особливості роботи з великими даними. Обґрунтовувати вибір структури, алгоритмів та способів обробки великих даних у реальному часі, з урахуванням специфіки потоків даних і вимог до продуктивності.	**ПР04.** Проводити системний аналіз об'єктів проєктування та обґрунтовувати вибір структури, алгоритмів та способів передачі інформації в інформаційних системах та технологіях.	1, 6
2	Демонструвати знання екосистеми Hadoop та сучасних хмарних сервісів, зокрема Google Cloud Platform, для роботи з великими даними. Застосовувати практичні навички обробки великих даних у реальному часі, використовуючи BigQuery та інші інструменти Google Cloud. Розробляти, аналізувати та візуалізувати дані за допомогою продуктів Google Cloud, таких як Data Studio чи Looker. Будувати моделі машинного навчання за допомогою AutoML і Vertex AI у Google Cloud для розв'язання прикладних завдань.	**ПР06.** Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.	4, 5, 6, 7, 8, 9, 10
3	Знати підходи до побудови моделей машинного навчання за допомогою AutoML та Vertex AI у Google Cloud. Вміти обґрунтовано обирати відповідні інструменти та технології Google Cloud для створення і налаштування моделей інтелектуального аналізу даних. Реалізовувати та використовувати штучні нейронні мережі в хмарному середовищі, використовуючи можливості Vertex AI та AutoML для вирішення практичних завдань. Застосовувати сучасні технології Google Cloud для автоматизації створення моделей та аналізу їх ефективності.	**ПР13.** Знати і вміти обґрунтовано обирати, застосовувати і реалізовувати інформаційні технології інтелектуального аналізу даних та підходи до побудови штучних нейронних мереж.	9, 10

4. Структура навчальної дисципліни.

№ п/п	Номер і назва теми	Кількість годин*				
		лекції	практичні заняття	семінарські заняття	лабораторні заняття	самостійна робота
1 семестр						
1	**Тема 1.** Історія великих даних та продуктів машинного навчання. Виклики під час роботи з великими даними. Загальні поняття та особливості великих даних	2				
2	**Тема 2.** Основні етапи життєвого циклу. Збір, консолідація і очищення даних, побудова моделей	2				
3	**Тема 3.** Розподілена обробка великих даних та метод Map-Reduce	4			2	8
4	**Тема 4.** Екосистема Hadoop	4	2		2	8
5	**Тема 5.** Огляд сервісів Google Cloud Platform	2			2	4
6	**Тема 6.** Обробка великих даних у реальному часі	2	2		2	8
7	**Тема 7.** Візуалізація даних за допомогою продуктів Google Cloud	4	2		4	12
8	**Тема 8.** Використання BigQuery	4	2		4	14
9	**Тема 9.** Побудова моделей в Google Cloud за допомогою AutoML	4			4	16
10	**Тема 10.** Побудова моделей в Google Cloud за допомогою VertexAI	4			4	16
Всього		**32**	**8**		**24**	**86**

Тематика практичних занять

№ Теми	Тематика практичного заняття	Кількість годин	Рекомендована література (№ з переліку)
Тема 4	Екосистема Hadoop	2	
Тема 6	Обробка великих даних у реальному часі	2	
Тема 7	Візуалізація даних за допомогою продуктів Google Cloud	2	
Тема 8	Використання BigQuery	2	
Всього годин		**8**	

Тематика лабораторних занять

№ Теми	Тематика лабораторного заняття	Кількість годин	Рекомендована література (№ з переліку)
Тема 3	Використання методу Map-Reduce	2	1, 2, 3
Тема 4	Огляд екосистеми Hadoop	2	1, 2, 7

Тема 5	Огляд сервісів Google Cloud Platform	2	6
Тема 6	Створення потокових пайплайнів в Cloud Dataflow	2	6
Тема 7	Візуалізація даних за допомогою Looker та Data Studio	4	6
Тема 8	Використання BigQuery ML для прогнозування продажу товарів	4	6
Тема 9	Побудова моделі для прогнозування повернення позики за допомогою AutoML	4	6
Тема 10	Запуск пайплайнів у VertexAI	4	6
Всього годин		**24**	

Тематика самостійної роботи

№ Теми	Тема самостійної роботи	Кількість годин	Рекомендована література (№ з переліку)
Тема 3	Закінчення лабораторної роботи «Розподілена обробка великих даних та метод Map-Reduce»	8	1, 2
Тема 4	Підготовка до практичного заняття «Екосистема Hadoop»	8	1, 2, 7
Тема 5	Самостійний огляд сервісів Google Cloud Platform поза межами лабораторних занять	4	6
Тема 6	Підготовка до практичного заняття «Обробка великих даних у реальному часі»	8	6
Тема 7	Виконання лабораторної роботи «Візуалізація даних за допомогою Looker та Data Studio» поза межами лабораторних занять	12	6
Тема 8	Виконання лабораторної роботи «Використання BigQuery ML для швидкої побудови моделей» поза межами лабораторних занять	14	6
Тема 9	Виконання лабораторної роботи «Побудова моделі для прогнозування повернення позики за допомогою AutoML» поза межами лабораторних занять	16	6
Тема 10	Виконання лабораторної роботи «Запуск пайплайнів у VertexAI» поза межами лабораторних занять	16	6
Всього годин		**86**	

5. Схема формування оцінки.

5.1 Шкала відповідності оцінювання:

Відмінно/Excellent	Зараховано/Passed	90-100
Добре/Good		82-89
		75-81
Задовільно/Satisfactory		64-74
		60-63
Незадовільно/Fail	Не зараховано/Fail	0-59

5.2 Форми та організація оцінювання:

Поточний контроль:

Форма оцінювання	Строки проведення оцінювання (тижні викладання)	Максимальна кількість балів
1. Оцінювання виконання лабораторних робіт: письмовий звіт та захист		
Тема 3	27	5
Тема 4	28	5
Тема 5	29	5
Тема 6	30	5
Тема 7	31	5
Тема 8	32	5
Тема 9	33	10
Тема 10	34	10
2. КМР: письмове опитування за темами № 1-10 (тест, MS Forms)	останній тиждень теоретичного навчання	10
Максимальна кількість балів за поточне оцінювання		**60**

Семестровий контроль:

Форма оцінювання	**Максимальна кількість балів**
Екзамен	40

5.3 Критерії оцінювання:

Критерії оцінювання знань здобувачів	
Оцінювання виконання та захисту лабораторних робіт	
Бали	**Критерій**
90-100% від максимальної кількості балів	Робота виконана самостійно, якісно, в повному обсязі, здобувач надає повні і ґрунтовні відповіді на всі питання щодо тематики роботи і процесу її виконання.
75-89% від максимальної кількості балів	Робота виконана самостійно і якісно, але деякі пункти роботи виконані з неточностями та/або відповіді студента є неповними.
50-74% від максимальної кількості балів	Робота виконана самостійно, але виконано не всі вимоги до роботи, здобувач дає неправильні або дуже неповні відповіді на більшість питань, демонструє лише часткове розуміння термінів. При цьому

	здобувач в змозі пояснити як саме він виконав той обсяг роботи, що захищає.
0-49% від максимальної кількості балів	Робота виконана не самостійно, або здобувач не дає правильних відповідей на більшість питань щодо тематики роботи і не в змозі захистити роботу. Робота повертається на доопрацювання.
Форма 2 КМР: письмове опитування за темами № 1- 8 (тест, MS Forms)	
Бали	**Критерій**
10	20 теоретичних питань по темам лекцій, лабораторних занять та самостійної роботи (0.5 бала за вірну відповідь на питання)

6. Методи навчання, інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна:

Методи навчання:

- Словесні методи навчання – лекції, пояснення;
- Наочні – демонстрації, ілюстрації;
- Практичні – лабораторні роботи;
- Самостійне навчання – опанування завдань для самостійної роботи у результаті аналізу навчальної та наукової літератури.

Традиційні методи навчання поєднуються зі сучасними методами з використанням дистанційних технологій навчання у системі Office 365.

Інструменти та обладнання:

Мультимедійне обладнання (проєктор, ноутбук), персональна обчислювальна техніка та/або власні ПК з доступом до мережі Internet, обладнаних стандартними пакетами прикладних програм.та засобами для розроблення програмного забезпечення.

У разі дистанційного навчання передбачається використання платформи MS Office 365.

Програмне забезпечення:

Програмно-апаратні інструментальні засоби моделювання, документування та управління вимогами, компіляції, налагодження коду, аналізу програмного коду, підтримки процесу тестування, верифікації та валідації програмного забезпечення, менеджменту проектів, групової динаміки і комунікації.

7. Рекомендована література:

Основна (Базова)

1. Akerkar R. Models of Computation for Big Data / R. Akerkar // Springer, 2018, 110 p.
2. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics / P. Ghavami // CreateSpace Independent Publishing Platform, 2016, 204 p.
3. Wiktorski T. Data-intensive Systems Principles and Fundamentals using Hadoop and Spark / Tomasz Wiktorski. – Springer Nature Switzerland AG. : Springer International Publishing, 2019. – 524 p.
4. Zgurovsky M.Z. Big Data: Conceptual Analysis and Applications/ M.Z. Zgurovsky, Y.P. Zaychenko // Springer, 2020, 298 p.

Додаткова

5. Джеймс У. Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени / У. Джеймс, Н. Марц. – Киев: Manning Publications, 2018. – 368 с.

8. Інформаційні ресурси:

6. Google Cloud. https://cloud.google.com/
7. Hadoop https://hadoop.apache.org/