Дніпровський національний університет імені Олеся Гончара

«ЗАТВЕРДЖУЮ»
Голова науково-методичної ради
Факультету прикладної математики
_____ Лілія БОЖУХА
«_30_» ___червня___ 2023 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ОК 2.25 Технології Data Mining

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)
Мацуга О.М., доцент, канд.тех.наук, доцент _____
підпис

Погоджено гарант ОП _____ Тетяна ЄМЕЛ'ЯНЕНКО
підпис

Робоча програма схвалена на засіданні кафедри математичного забезпечення ЕОМ

Протокол від «_25_» червня 2023 року № 19

Ухвалено на засіданні науково-методичної ради факультету прикладної математики
Протокол від «_30_» червня 2023 року № 15

Дніпро
2023

Опис навчальної дисципліни

Навчальний рік (роки*) викладання дисципліни	Курс	Семестр	Підсумковий контроль				Індивіду-альні завдання		Кредитів ECTS	Обсяг роботи студента (години)						
			екзамен	диф.залік	залік	курсова робота	форма	кількість		всього	аудиторні					самостійна робота
											всього аудиторних	лекції	практичні заняття	семінарські заняття	лабораторні заняття	
2023/24	3															
		6	6						3	90	40	24			16	50
20__/__																
20__/__																
20__/__																

1. Мета дисципліни

Мета викладання дисципліни – ознайомити здобувачів вищої освіти з основними задачами Data Mining, навчити методам та технологіям їх розв'язання, прищепити практичні навички їх розв'язання, надати досвід створення програмного забезпечення для інформаційних систем та технологій інтелектуального аналізу даних.

Вивчення дисципліни забезпечує формування компетентностей за ОП:

ЗК05. Здатність вчитися і оволодівати сучасними знаннями.

ФК06. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.

ФК13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.

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.

Знання та навички, отримані під час вивченні освітніх компонентів «Інформаційні технології обробки даних», «Теорія ймовірностей та математична статистика», «Основи програмування».

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

№	Результати навчання за дисципліною	Програмні результати навчання за ОП	Номери тем
1	Знати основні поняття та постановки задач Data Mining.	**ПР01.** Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації. **ПР13.** Знати і вміти обґрунтовано обирати, застосовувати і реалізовувати інформаційні технології інтелектуального аналізу даних та підходи до побудови штучних нейронних мереж.	1.1, 1.2
2	Знати базові етапи та методи препроцессінгу даних.		1.3
3	Знати основи кореляційного аналізу даних.		1.4
4	Знати сутність та методи навчання базових моделей регресії та класифікації.		1.5–1.7, 1.10–1.13
5	Знати метрики якості моделей регресії та класифікації, підходи до їх оцінювання, сутність та способи вирішення проблем перенавчання.		1.8, 1.9
6	Вміти обґрунтовано обирати і застосовувати на практиці технології Data Mining, виконувати попередню обробку та очищення даних, навчати на підготовлених даних моделі регресії та класифікації, робити висновки щодо якості навчених моделей.	**ПР02.** Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проєктування і використання інформаційних систем та технологій. **ПР13.** Знати і вміти обґрунтовано обирати, застосовувати і реалізовувати інформаційні технології інтелектуального аналізу даних та підходи до побудови штучних нейронних мереж.	1.1–1.13
7	Вміти використовувати базові знання інформатики, навички програмування, створення алгоритмів та комп'ютерних програм мовами високого рівня, включаючи мову Python, у процесі розв'язання задач Data Mining та розроблення інформаційних систем і технологій інтелектуального аналізу даних.	**ПР03.** Використовувати базові знання інформатики й сучасних інформаційних систем та технологій, навички програмування, технології безпечної роботи в комп'ютерних мережах, методи створення баз даних та інтернет-ресурсів, технології розроблення алгоритмів і комп'ютерних програм мовами високого рівня із застосуванням об'єктно-орієнтованого програмування для розв'язання задач проєктування і використання інформаційних систем та технологій.	1.3, 1.4, 1.7, 1.8, 1.9
8	Мати навички розроблення програмного забезпечення для інформаційних систем і		1.3–1.10

	технологій інтелектуального аналізу даних.	**ПР13.** Знати і вміти обґрунтовано обирати, застосовувати і реалізовувати інформаційні технології інтелектуального аналізу даних та підходи до побудови штучних нейронних мереж.	
9	Вміти аргументувати вибір програмних та технічних засобів, необхідних для розв'язання задач Data Mining та розроблення інформаційних систем і технологій інтелектуального аналізу даних.	**ПР05.** Аргументувати вибір програмних та технічних засобів для створення інформаційних систем та технологій на основі аналізу їх властивостей, призначення і технічних характеристик з урахуванням вимог до системи і експлуатаційних умов; мати навички налагодження та тестування програмних і технічних засобів інформаційних систем та технологій.	1.3, 1.4, 1.7, 1.8, 1.9
10	Мати навички використання бібліотек Python у ході розв'язання задач Data Mining.	**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.	1.3, 1.4, 1.7, 1.8, 1.9

4. Структура навчальної дисципліни.

№ п/п	Номер і назва теми	Кількість годин*				
		лекції	практичні заняття	семінарські заняття	лабораторні заняття	самостійна робота
6 семестр						
Розділ 1. Технології Data Mining						
1	Тема 1.1 Вступ до Data Mining	2				2
2	Тема 1.2 Задачі Data Mining	2				2
3	Тема 1.3 Попередня обробка та очищення даних. Використання бібліотек Python	2			4	4
4	Тема 1.4 Кореляційний аналіз даних. Використання бібліотек Python	4			4	6
5	Тема 1.5 Лінійна одновимірна регресія	2			2	8
6	Тема 1.6 Нелінійна одновимірна регресія	2				4
7	Тема 1.7 Лінійна багатовимірна регресія. Використання бібліотек Python	2				8
8	Тема 1.8 Метрики якості моделі регресії. Підходи до оцінювання якості. Проблема перенавчання. Використання бібліотек Python	2			2	4
9	Тема 1.9 Задача класифікації. Оцінювання якості моделі класифікації. Використання бібліотек Python	2			2	4
10	Тема 1.10 Правило *k* найближчих сусідів	2			2	6
11	Тема 1.11 Логістична регресія	2				2
Всього		**24**			**16**	**50**

Тематика лабораторних занять

№ Теми	Тема лабораторного заняття	Кількість годин	Рекомендована література (№ з переліку)
Розділ 1.			
Тема 1.3	Лабораторна робота 1 на тему «Попередня обробки та очищення даних»	4	4
Тема 1.4	Лабораторна робота 2 на тему «Кореляційний аналіз»	4	2, 3
Теми 1.5, 1.7, 1.8	Лабораторна робота 3 на тему «Регресійний аналіз»	4	2, 3
Теми 1.9, 1.10	Лабораторна робота 4 на тему «Класифікація»	4	1, 4, 7, 16
Всього годин		**16**	**-**

Тематика самостійної роботи

№ Теми	Тема самостійної роботи	Кількість годин	Рекомендована література (№ з переліку)
Тема 1.3	Виконання лабораторної роботи 1 поза межами лабораторних занять	2	4
Тема 1.4	Виконання лабораторної роботи 2 поза межами лабораторних занять	4	2, 3
Теми 1.5, 1.7, 1.8	Виконання лабораторної роботи 3 поза межами лабораторних занять	8	2, 3
Теми 1.9, 1.10	Виконання лабораторної роботи 4 поза межами лабораторних занять	6	1, 4, 7, 16
Теми 1.1–1.13	Підготовка до екзамену	30	1–5
Всього годин		**50**	**-**

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
Оцінювання виконання та захисту лабораторних робіт		
Лабораторна робота 1	4	7
Лабораторна робота 2	8	14
Лабораторна робота 3	12	20
Лабораторна робота 4	16	19
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 1, 2, 3, 4	
Бали	**Критерій**
90-100 % від максимальної кількості балів	Здобувач виконав роботу самостійно, якісно, в повному обсязі, надає повні і ґрунтовні відповіді на всі питання щодо тематики роботи і процесу її виконання.
75-89 % від максимальної кількості балів	Здобувач виконав роботу самостійно і якісно, але деякі пункти роботи виконані з неточностями та/або відповіді студента є неповними, іноді фрагментарними.
60-74 % від максимальної кількості балів	Здобувач виконав роботу самостійно, але виконано не всі пункти роботи (проте більше визначеного постановкою мінімуму), 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.
50-59 % від максимальної кількості балів	Здобувач виконав роботу самостійно, але обмежився виконанням лише мінімального набору пунктів, здобувач демонструє дуже погане розуміння тематики роботи, при цьому він в змозі пояснити як саме він виконав той обсяг роботи, що захищає.
0-49 % від максимальної кількості балів	Виконано менше передбаченого мінімуму, або 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.
Екзамен	
Бали	**Критерій**
40	40 тестових питань по 1 балу за кожну правильну відповідь.

*** (*за кожною формою оцінювання, зазначеною у п. 5.2*)**

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

Методи навчання:
У ході викладання дисципліни передбачається використання таких методів навчання як словесний (лекції, пояснення), наочний (презентації, демонстрації), практичний (виконання лабораторних робіт) та самостійне навчання. За потреби застосовуються сучасні методи дистанційного навчання (відеолекції, відеоконференції).

Інструменти та обладнання:
Під час проведення навчальних занять передбачається використання мультимедійного обладнання та персональних комп'ютерів, обладнаних 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 проєктів, групової динаміки і комунікації.

7. Рекомендована література:

Основна:

1. Мацуга О.М., Архангельська Ю.М., Єрещенко Н.М. Навчальний посібник до вивчення курсу «Інформаційні технології розпізнавання образів». Д.: РВВ ДНУ, 2016. 60 с.
2. Методичні вказівки до виконання практичних робіт із дисципліни «Математична статистика та обробка геологічної інформації» / Укладачі Мацуга О.М., Ємел'яненко Т.Г. Дніпро: РВВ ДНУ, 2019. 64 с.
3. Приставка П.О., Мацуга О.М. Аналіз даних. Д.: РВВ ДНУ, 2008. 92 с.
4. Aggarwal C.C. Data Mining: The Textbook. Springer, 2015. 763 p.
5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. 2009. 745 p.

Додаткова:

6. Бабак В.П., Білецький А.Я., Приставка О.П., Приставка П.О. Статистична обробка даних. К.: МІВВЦ, 2001. 388 с.
7. Мацуга О.М., Дудукіна С.О., Григорук С.П. Побудова моделі прогнозування результату лікування на прикладі однієї медичної задачі. *Актуальні проблеми автоматизації та інформаційних технологій*: зб. наук. пр. Дніпро, 2020. Т. 24. С. 47-56.
8. Олійник А.О., Субботін С.О., Олійник О.О. Інтелектуальний аналіз даних: навчальний посібник. Запоріжжя: ЗНТУ, 2012. 278 с.
9. Приставка О.П., Приставка П.О., Смирнов С.О. Статистичний аналіз в АСОД: Кореляція. Регресія. Д.: РВВ ДДУ, 2000. 120 с.
10. Albon C. Machine Learning with Python Cookbook. O'Reilly, 2018. 366 p.
11. Bishop C.M. Pattern Recognition and Machine Learning. Springer, 2006. 738 p.
12. Draper N., Smith H. Applied Regression Analysis. Wiley-Interscience, 1998. 736 p.
13. Duda R., Hart P., Stork D. Pattern Classification. Wiley. 2001.
14. Rokach L., Maimon D. Data mining with decision trees. Theory and applications. 2nd ed. Singapore : World Scientific Publishing Company. 2015. 305 p.
15. Lutz M. Learning Python. O'Reilly, 2013. 1643 p.
16. Matsuga O., Nashylnyk B. Decision tree ensemble constructing algorithms and their comparative analysis. *Актуальні проблеми автоматизації та інформаційних технологій*: зб. наук. пр. Дніпро, 2018. Т. 22. С. 62-69.
17. McKinney W. Python for Data Analysis. O'Reilly, 2022. 579 p.
18. VanderPlas J. Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly, 2017. 546 p.
19. Zaki M.J., Meira Jr. W. Data Mining and Analysis. Fundamental Concepts and Algorithms. Camridge University Press, 2014. 593 p.

8. Інформаційні ресурси:

1. Репозиторій ДНУ: http://repository.dnu.dp.ua:1100/?page=faculty
2. Репозиторій з даними: https://archive.ics.uci.edu/ml/datasets.html
3. Репозиторій з даними: https://www.kaggle.com
4. Документація бібліотеки scikit-learn: https://scikit-learn.org/stable/index.html
5. Data Science Tutorials: https://www.datacamp.com/tutorial/category/python, https://www.datacamp.com/tutorial/data-preparation-with-pandas