

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ МАТЕМАТИЧНИЙ
КАФЕДРА КОМП'ЮТЕРНИХ НАУК

ЗАТВЕРДЖУЮ

Декан математичного факультету

С. І. Гоменюк

« 01 » вересня 2023 р.



ОБРОБКА ДАНИХ ТА ВІЗУАЛІЗАЦІЯ

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

підготовки магістра
денної (очної) форми здобуття освіти
спеціальності 126 Інформаційні системи та технології
освітня програма Інформаційні системи та штучний інтелект

Укладач: Шило Г. М., д.т.н., доцент, в.о. завідувача кафедри комп'ютерних наук


Обговорено та ухвалено
на засіданні кафедри комп'ютерних наук
Протокол № 1 від «30» серпня 2023 р.
В.о. завідувача кафедри

Ухвалено науково-методичною радою
математичного факультету

Протокол № 1 від «31» серпня 2023 р.
Голова науково-методичної ради
факультету


Г. М. Шило


О. С. Пшенична

Погоджено
гарант освітньо-професійної програми

Г. М. Шило

2023 рік

1. Опис навчальної дисципліни

1	2	3	
Галузь знань, спеціальність, освітня програма рівень вищої освіти	Нормативні показники для планування і розподілу дисципліни на змістові модулі	Характеристика навчальної дисципліни	
		очна (денна) форма здобуття освіти	заочна (дистанційна) форма здобуття освіти
Галузь знань 12 Інформаційні технології	Кількість кредитів – 5	Обов’язкова	
Спеціальність 126 Інформаційні системи та технології	Загальна кількість годин – 150	Цикл професійної підготовки	
		Семестр:	
Освітньо-професійна програма Інформаційні системи та штучний інтелект	Змістових модулів – 8	1-й	1-й
		Лекції	
Рівень вищої освіти: магістерський	Кількість поточних контрольних заходів – 20	10 год.	–
		Лабораторні	
		30 год.	–
Рівень вищої освіти: магістерський	Кількість поточних контрольних заходів – 20	Самостійна робота	
		110 год.	–
		Вид підсумкового семестрового контролю: екзамен	

2. Мета та завдання навчальної дисципліни

Метою вивчення дисципліни «Обробка даних та візуалізація» є набуття студентами знань, умінь і навичок, необхідних для застосування Data Mining та вирішення задач візуалізації даних і розробки інтелектуальних систем.

Основними завданнями вивчення дисципліни «Обробка даних та візуалізація» є:

- оволодіння вміннями зі створення інфографіки та аналітичних звітів;
- набуття вмінь і навичок з обґрунтування вибору конкретного типу моделі та методу інтелектуального аналізу даних;
- опанування вміннями з проведення необхідної попередньої обробки даних, визначення типу задач аналізу та підготовки змістовних висновків та інтерпретації;
- оволодіння вміннями з використання сучасних програмних засобів для проектування та дослідження систем інтелектуального аналізу даних.

У результаті вивчення навчальної дисципліни студент повинен набути таких результатів навчання (знання, вміння тощо) та компетентностей:

Заплановані робочою програмою результати навчання та компетентності	Методи і контрольні заходи, що забезпечують досягнення результатів навчання та компетентностей
Результати навчання	
РН 1. Відшукувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм Контрольні заходи: опитування, тест, захист лабораторних робіт
РН 8. Розробляти моделі інформаційних процесів та систем різного класу, використовувати методи моделювання, формалізації, алгоритмізації та реалізації моделей з використанням сучасних комп'ютерних засобів.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм Контрольні заходи: опитування, тест, захист лабораторних робіт
РН 9. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм Контрольні заходи: опитування, тест, захист лабораторних робіт
Компетентності	
ЗК 1. Здатність до абстрактного мислення, аналізу та синтезу.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт
СК 1. Здатність розробляти та застосувати ІСТ, необхідні для розв'язання стратегічних і поточних задач.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт

Заплановані робочою програмою результати навчання та компетентності	Методи і контрольні заходи, що забезпечують досягнення результатів навчання та компетентностей
СК 4. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт
СК 5. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт
СК 8. Здатність розробляти та використовувати методи штучного інтелекту для аналізу даних і прогнозування в інформаційних системах.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт
СК 9. Здатність використовувати сучасні технології візуалізації даних.	Методи навчання: лекція-візуалізація, пояснення, майстер-клас, лабораторні роботи, практичний метод, мозковий штурм. Контрольні заходи: опитування, тест, захист лабораторних робіт

Міждисциплінарні зв'язки. Курс «Обробка даних та візуалізація» базується на знаннях отриманих під час вивчення курсів бакалаврського рівня. Вивченню дисципліни «Обробка даних та візуалізація» передують курси «Архітектура розподілених систем». Знання, уміння і навички засвоєні при вивченні навчальної дисципліни «Обробка даних та візуалізація» знадобляться при вивченні дисципліни «Сучасні методи машинного навчання», а також при проходженні Виробничої практики та під час виконання кваліфікаційної роботи магістра.

3. Програма навчальної дисципліни

Змістовий модуль 1

Тема 1. Основи візуалізації даних. Сучасні інструменти візуалізації даних. Типи даних. Мітки та візуальні канали. Вибір техніки візуалізації. Візуалізація даних у Google Data Studio.

Змістовий модуль 2-3

Тема 2. Інфодизайн: графіки та ефективність візуального кодування. Основи бізнес-аналітики. Основні поняття графічного дизайну. Основи композиції. Контраст. Візуальна вага. Рух. Ритм. Баланс. Психологічні принципи у графічному дизайні. Табличні дані і графіки. Географічні дані. Аналіз прикладів візуалізацій. Інструменти веб-аналітики. Google Analytics. Планування KPI.

Змістовий модуль 4-5

Тема 3. Основи інтелектуального аналізу даних. Задача класифікації. Визначення Data Mining і область застосування. Задачі, моделі та методи Data Mining. Методи, стадії, задачі Data Mining. Поняття Business Intelligence. Цикл одержання, попередньої обробки, аналізу даних, інтерпретації результатів та їхнього використання. Етапи процесу Data Mining, пов'язані з побудовою, перевіркою, оцінкою, вибором и корекцією моделей. Методи первісної обробки даних. Інструментальні засоби Data Mining. Методи дослідження структури даних: візуалізація даних.

Постановка задачі класифікації та представлення результатів. Методи побудови правил класифікації. Методи побудови дерев рішень. Методи побудови математичних функцій. Методи опорних векторів, «найближчого сусіда», Байеса. Аналіз багатомірних угруповань. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації.

Змістовий модуль 6

Тема 4. Прогнозування. Задача регресії. Методи аналізу часових рядів. Задача кластеризації. Сутність задачі прогнозування. Методи вирішення задачі регресії. Поняття нечітких часових рядів. Методи моделювання часових рядів. Методи аналізу та прогнозування поведінки часових рядів. Постановка задачі кластеризації та представлення результатів. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації

Змістовий модуль 7–8

Тема 5. Розв'язання задачі пошуку асоціативних правил. Сховища даних та оперативний аналіз даних (OLAP). Постановка задачі пошуку асоціативних правил та представлення результатів. Секвенціальний аналіз. Різновиди задач пошуку асоціативних правил. Методи подання результатів. Алгоритми пошуку асоціативних правил. Методи пошуку асоціативних правил: метод Apriori, побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.

Визначення сховища даних, порівняння з базами даних, використання.. Архітектура сховища даних. ETL-процеси (добування, перетворення й завантаження даних). Вітрини даних, куби даних, багатомірна модель даних.

Розгортання OLAP-кубів. Операції над OLAP-кубами (зріз, обертання, консолідація, деталізація). Архітектура OLAP-систем: MOLAP, ROLAP, HOLAP.

4. Структура навчальної дисципліни

Зміст. модуль	Усього годин	Аудиторні (контактні) години						Самостійна робота, год		Система накопичення балів		
		Усього годин		Лекційні заняття, год.		Лабораторні заняття, год.		о/д ф.	з/дист ф.	Теор. завд., к-ть балів	Практ. завд., к-ть балів	Усього балів
		о/д ф.	з/дист т. ф.	о/д ф.	з/дист т. ф.	о/д ф.	з/дист т. ф.					
1	2	3	4	5	6	7	8	9	10	11	12	13
1	15	6		2		4		9		3	3	6
2	15	6		2		4		9		6	6	12
3	15	4				4		11		3	3	6
4	15	4		2		2		11		1	2	3
5	15	4				4		11		3	3	6
6	15	6		2		4		9		3	3	6
7	15	6		2		4		9		3	3	6
8	15	4				4		11		3	12	15
Усього за змістові модулі	120	40		10		30		80		25	35	60
Підсумковий семестровий контроль екзамен	30							30		20	20	40
Загалом	150	40		10		30		110		45	55	100

5. Теми лекційних занять

№ змістового модуля	Назва теми	Кількість годин
1	Основи візуалізації даних. Сучасні інструменти візуалізації даних	2
2–3	Інфодизайн: графіки та ефективність візуального кодування. Основи бізнес-аналітики	2
4–5	Основи інтелектуального аналізу даних. Задача класифікації	2
6	Прогнозування. Задача регресії. Методи аналізу часових рядів. Задача кластеризації	2
7–8	Розв'язання задачі пошуку асоціативних правил. Сховища даних та оперативний аналіз даних (OLAP)	2
Разом		10

6. Теми лабораторних занять

№ змістового модуля	Назва теми	Кількість годин
1	Лабораторна робота №1. Візуалізація даних у Google Data Studio	4
2	Лабораторна робота № 2. Створення аналітичного звіту	4
3	Лабораторна робота №3. Веб-аналітика	2
	Лабораторна робота №4. Підготовка даних для інтелектуального аналізу даних. Використання Weka	2
4	Лабораторна робота №4. Підготовка даних для інтелектуального аналізу даних. Використання Weka	2
5	Лабораторна робота №5. Задача класифікації	4
6	Лабораторна робота №6. Прогнозування. Задача регресії	4
7	Лабораторна робота № 7. Задача кластеризації	4
8	Лабораторна робота № 8. Пошук асоціативних правил	4
Разом		30

7. Види і зміст поточних контрольних заходів

№ змістового модуля	Вид поточного контрольного заходу	Зміст поточного контрольного заходу	Критерії оцінювання	Усього балів
1	Лабораторна робота №1	Завдання: Візуалізація даних у Google Data Studio	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
		Захист лабораторної роботи, опитування	Захист - 2 бали, опитування - 1	3
2	Лабораторна робота №2	Завдання: Створення аналітичного звіту	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
		Захист лабораторної роботи, опитування	Захист – 2 бали, опитування – 1	3
3	Лабораторна робота №3	Завдання: Розробка веб-аналітики	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
	Лабораторна робота №4	Завдання: Підготовка даних для інтелектуального аналізу даних з використанням Weka	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
4	Лабораторна робота №4	Завдання: Підготовка даних для інтелектуального аналізу даних з використанням Weka	Захист – 2 бали, опитування – 1	3
	Тест 1	Тестування (12 тестових завдань)	Правильна відповідь на кожне завдання – 0,5 бали	6
Усього за змістові модулі 1–4				27
5	Лабораторна робота №5	Завдання: Реалізувати задачу класифікації	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
	Лабораторна робота №5	Завдання: Реалізувати задачу класифікації	Захист – 2 бали, опитування – 1	3
6	Лабораторна робота №6	Завдання: Здійснення прогнозування з використанням регресійного аналізу	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бал	3
	Лабораторна робота №6	Завдання: Здійснення прогнозування з використанням регресійного аналізу	Захист – 2 бали, опитування – 1	3

№ змістового модуля	Вид поточного контрольного заходу	Зміст поточного контрольного заходу	Критерії оцінювання	Усього балів
7	Лабораторна робота №7	Завдання: Реалізувати задачу кластеризації.	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
	Лабораторна робота №7	Завдання: Реалізувати задачу кластеризації.	Захист – 2 бали, опитування – 1	3
8	Лабораторна робота №8	Завдання: Пошук асоціативних правил	Повне виконання завдання оцінюється в 3 бали, за наявності помилок у 1,5 бали	3
	Лабораторна робота №8	Завдання: Пошук асоціативних правил	Захист – 2 бали, опитування – 1	3
	Тест 2	Тестування (18 тестових завдань)	Правильна відповідь на кожне завдання – 0,5 бали	9
Усього за змістові модулі 5–8				33

8. Підсумковий семестровий контроль

Форма	Види підсумкових контрольних заходів	Зміст підсумкового контрольного заходу	Критерії оцінювання	Усього балів
Екзамен	Теоретична частина тест	Відповідь на 40 тестових завдань	Правильна відповідь на кожне завдання оцінюється в 0,5 бали	20
	Практична частина	Індивідуальне завдання: розв'язання задачі інтелектуального аналізу даних у системі WEKA	Виконання оцінюється максимально у 20 балів	20
Усього				40

9. Рекомендована література

Змістові модулі 1–8:

Основна:

1. Муляр В. П. Візуалізація даних та інфографіка. Харків : ФОП Панов А. М., 2020. 200 с.
2. Гороховатський В. О., Творошенко І. С. Методи інтелектуального аналізу та оброблення даних : навч. посібник. Харків : ХНУРЕ, 2021. 92 с.
3. Лупан І. В. Інтелектуальний аналіз даних Data Mining : навчально-методичний посібник. Кропивницький : ФОП Піскова М. А., 2022. 112 с.

Додаткова:

4. Сергесв-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних : комп'ютерний практикум. Електронні текстові данні (1файл: 1,72 Мбайт). Київ : КПІ ім. Ігоря Сікорського, 2018. 73 с.

Інформаційні джерела:

1. Weka 3: Machine Learning Software in Java. URL: <http://www.cs.waikato.ac.nz/ml/weka/>