

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ

БАКАЛАВР
(освітній ступінь)

Кафедра інформаційних технологій

«Затверджую»

Голова НМР факультету автоматизації і
інформаційних технологій

Олександр ТЕРЕНТЬЄВ/

«_____» 2024 року



РОБОЧА ПРОГРАМА ОСВІТНЬОЇ КОМПОНЕНТИ

ОК 26 «ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ»

(шифр та назва освітньої компоненти)

шифр	назва спеціальності, освітньої програми
122	Комп'ютерні науки Інформаційні управляючі системи і технології

Мова викладання: українська

Розробник:

Світлана БіЛОЩИЦЬКА, доктор технічних наук, доцент

(ім'я та прізвище, науковий ступінь, звання)

(підпис)

Робоча програма затверджена на засіданні кафедри інформаційних технологій,
протокол № 13 від «25» червня 2024 року

Завідувачка кафедри ІТ

(підпис)

Тетяна ГОНЧАРЕНКО

(ім'я та прізвище)

Схвалено гарантом освітньої програми

Гарант ОП

(підпис)

Олександр ПОПЛАВСЬКИЙ


(ім'я та прізвище)

Розглянуто на засіданні науково-методичної комісії спеціальності

122 «Комп'ютерні науки».

Протокол № 3 від «28» червня 2024 року

ВИТЯГ З РОБОЧОГО НАВЧАЛЬНОГО ПЛАНУ 2024-2025 рр.

Шифр	Назва спеціальності, освітньої програми	Форма здобуття ВО: денна										Погодження заступником декана факультету			
		Кількість кредитів	Кількість годин				Кількість індивідуальних робіт				Форма контролю		Семестр		
			Всього	Аудиторних		Само	КП	КР	РГР	Конт. роб.					
				Ра	у тому числі									Ла	П
с	зо	м	ці	ї	бо	рат	ор	ні	ра	кт	ич	ні	бот	а	
122	Комп'ютерні науки, Інформаційні управляючі системи і технології	3,0	90	50	20	30	40				1		Залік	5	

Анотація. Мета та завдання освітньої компоненти

Пререквізити: освітні компоненти «Програмування та алгоритмічні мови», «Інструментальні засоби програмування», «Чисельні методи в інформатиці», «Теорія алгоритмів».

Посилання на сторінку електронного навчально-методичного комплексу освітньої компоненти:
<https://org2.knuba.edu.ua/course/view.php?id=447>

Метою освітньої компоненти – теоретична та практична підготовка у майбутніх бакалаврів з комп'ютерних наук у напрямку розв'язання задач обробки великих масивів інформації, проектування інформаційного забезпечення інформаційних систем та розробки сценаріїв можливих дій в умовах невизначеності з використанням інтелектуальних методів обчислень.

Завдання освітньої компоненти полягає у формуванні в здобувачів здатності до виявлення статистичних закономірностей недетермінованих явищ, застосування методів статистичної обробки даних, машинного навчання, інтелектуального аналізу даних на основі методів обчислювального інтелекту, оперативної обробки структурованих та слабкоструктурованих даних та візуалізації результатів аналізу в процесі розв'язання прикладних задач.

Вивчення освітньої компоненти «Інтелектуальний аналіз даних» сприяє формуванню у здобувачів **наступних компетентностей**.

Компетенції здобувачів, що формуються в результаті засвоєння освітньої компоненти

Код	Зміст компетентності
Інтегральна компетентність	
ІК	Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук або у процесі навчання, що передбачає застосування теорій та методів інформаційних технологій і характеризується комплексністю та невизначеністю умов.
Загальні компетентності	
ЗК1	Здатність до абстрактного мислення, аналізу та синтезу.
ЗК3	Знання та розуміння предметної області та розуміння професійної діяльності
ЗК6	Здатність вчитися і оволодівати сучасними знаннями
ЗК7	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
ЗК11	Здатність приймати обґрунтовані рішення.
СПЕЦІАЛЬНІ (ФАХОВІ) КОМПЕТЕНТНОСТІ	
СК1	Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування

СК2	Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.
СК3	Здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем.
СК6	Здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризики.
СК7	Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.
СК11	Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

Це забезпечує досягнення *програмних результатів навчання*, згідно з якими **Здобувач повинен мати знання з питань:**

Програмні результати здобувачів освітньої програми, що формуються в результаті засвоєння освітньої компоненти

Код	Програмні результати
ПР 1	Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук
ПР 3	Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.
ПР 4	Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.
ПР 5	Проектувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій.

ПР 12	Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.
--------------	--

ЗМІСТ КУРСУ

Змістовний модуль 1. Основи статистичного аналізу та обробки даних

Лекція 1. Методи первісної обробки даних. Класифікація ознак за шкалами вимірювання. Описова статистика. Числові характеристики вибірки

Лекція 2. Постановка задачі побудови статистичного зведення. Статистичне групування, його задачі та види. Принципи побудови статистичних групувань. Ряди розподілу. Статистичні таблиці. Статистичні графіки.

Лекція 3. Кореляційний і регресійний аналіз. Метод найменших квадратів. Парна регресія. Множинна регресійна модель. Автокореляція.

Змістовний модуль 2. Методи кластеризації, класифікації та розвідувального аналізу

Лекція 4. Методи багатовимірного розвідувального аналізу. Задача кластеризації. Ієрархічні методи кластеризації.

Лекція 5. Постановка задачі класифікації. Подання результатів розв'язання задачі класифікації. Класифікаційні правила. Методи побудови класифікаційних правил.

Лекція 6. Дерева рішень. Методи побудови дерев рішень. Алгоритм ID3. Алгоритм C4.5. Алгоритм CART.

Змістовний модуль 3. Пошук шаблонів, багатовимірний аналіз і часові ряди

Лекція 7. Пошук асоціативних правил. Секвенційний аналіз. Різновиди задач пошуку асоціативних правил. Метод Apriori. Побудова FP-дерев пошуку шаблонів даних. Min-max асоціації у базах даних. Побудова hash-дерев.

Лекція 8. Розробка OLAP-кубів під час аналізу багатовимірних даних у великих базах даних. Поняття OLAP-систем. Правила Кодда.

Лекція 9. OLAP і Data Mining. Етапи розв'язання задачі аналізу даних. Ентропія і кількість інформації. Нормалізація і стандартизація вихідних значень.

Лекція 10. Часові ряди. Моделі часових рядів. Методи аналізу часових рядів. Екстраполяція та інтерполяція.

Теми лабораторних робіт

№	Назва теми	К-сть годин	К-сть балів
Змістовний модуль 1. Основи статистичного аналізу та обробки даних			
1	Лабораторна робота №1. Основи роботи в системі R.	2	6
2	Лабораторна робота №2. Перевірка статичних гіпотез.	4	7
3	Лабораторна робота №3. Задачі відновлення регресії.	4	7
Змістовний модуль 2. Методи кластеризації, класифікації та розвідувального аналізу			
4	Лабораторна робота №4. Кластеризація в R.	4	6
5	Лабораторна робота №5. Побудова дерев рішень.	4	7
6	Лабораторна робота №6. Метод опорних векторів.	4	7
Змістовний модуль 3. Пошук шаблонів, багатовимірний аналіз і часові ряди			
7	Лабораторна робота №7. Використання бустінг алгоритму в задачах класифікації.	4	10
8	Лабораторна робота №8. Дослідження вибірки за допомогою функції bootstrap.	4	10
	Разом	30	60

Розподіл годин самостійної роботи здобувачів

№	Назва теми	К-сть годин
Змістовний модуль 1		7
1	Дослідження видів помилок вибірки та їх вплив на числові характеристики.	2
2	Принципи побудови статистичних зведень і групувань. Огляд рядів розподілу, таблиць і графіків.	2
3	Основи кореляційного і регресійного аналізу. Теоретичні аспекти методу найменших квадратів.	3
Змістовний модуль 2		7
4	Дослідження методів оцінки якості кластеризації (індекс Дейвіса-Болдіна, силуетний аналіз).	2
5	Вивчити вплив збалансованості вибірки на точність класифікації.	2
6	Ознайомитися з методами побудови ансамблів дерев рішень (Random Forest, Boosting).	3
Змістовний модуль 3		8
7	Дослідження використання алгоритму ECLAT у задачах пошуку частих наборів елементів.	2
8	Ознайомлення з концепцією кубів даних у реляційних базах (ROLAP-моделі)	2
9	Ознайомлення з методами очищення та трансформації даних перед Data Mining.	2
10	Вивчення методики оцінки точності прогнозів часових рядів	2

	(MAPE, RMSE).	
	Виконання РГР	12
	Підготовка до заліку	6
	Всього	40

Індивідуальне завдання: РГР

Виконання РГР згідно індивідуального варіанту для закріплення теоретичних знань і практичних навичок з інтелектуального аналізу даних за 5 семестр.

Структура розрахунково-графічна роботи, розподіл часу та балів

Мета розрахунково-графічної роботи – закріплення теоретичних знань і набуття практичних навичок роботи з методами аналізу даних, класифікації, кластеризації та візуалізації результатів, з використанням середовища R.

Структура розрахунково-графічної роботи	Години	Бали
Завдання повинно включати: 1) Вступ ○ Мета і завдання роботи. ○ Опис обраної теми та її значення для інтелектуального аналізу даних. 2) Теоретична частина ○ Короткий огляд методів і алгоритмів, які використовуються в обраній темі. ○ Теоретичні основи реалізації задачі. 3) Практична частина ○ Опис обраного набору даних (джерело, структура, особливості). ○ Покрокова реалізація завдання: ● Попередня обробка даних. ● Застосування методів/алгоритмів для вирішення задачі. ● Аналіз отриманих результатів. 4) Результати та висновки ○ Основні результати роботи, їх аналіз та інтерпретація. ○ Оцінка ефективності застосованих методів.	12	20
Разом	12	20

Теми для розрахунково-графічної роботи

1. Дослідження та попередня обробка даних
2. Статистичний аналіз і регресійні моделі
3. Методи кластеризації
4. Класифікація даних

5. Пошук шаблонів у даних
6. Аналіз часових рядів
7. Багатовимірний аналіз даних
8. Використання функцій bootstrap у дослідженні вибірки
9. Застосування бустінг-алгоритмів для класифікації
10. Візуалізація багатовимірних даних
11. Аналіз даних із різними типами змінних
12. Аналіз впливу змінних у моделі
13. Використання методів скорочення розмірності
14. Оптимізація гіперпараметрів моделей
15. Аномалії у даних та їх виявлення

Методи контролю та оцінювання знань

Загальне оцінювання здійснюється через вимірювання результатів навчання у формі проміжного (модульного) та підсумкового контролю (залік, захист індивідуальної роботи тощо) відповідно до вимог зовнішньої та внутрішньої системи забезпечення якості вищої освіти.

Політика щодо академічної доброчесності

Тексти індивідуальних завдань (в т.ч. у разі, коли вони виконуються у формі презентацій або в інших формах) можуть перевірятись на плагіат. Для цілей захисту індивідуального завдання оригінальність тексту має складати не менше 70%. Виключення становлять випадки зарахування публікацій Здобувачів у матеріалах наукових конференціях та інших наукових збірниках, які вже пройшли перевірку на плагіат.

Списування під час тестування та інших опитувань, які проводяться у письмовій формі, заборонені (в т.ч. із використанням мобільних девайсів). У разі виявлення фактів списування з боку здобувача він отримує інше завдання. У разі повторного виявлення призначається додаткове заняття для проходження тестування.

Політика щодо відвідування

Здобувач, який пропустив аудиторне заняття з поважних причин, має продемонструвати викладачу та надати до деканату факультету документ, який засвідчує ці причини.

За об'єктивних причин (хвороба, міжнародне стажування, наукова та науково-практична конференція (круглий стіл) тощо) навчання може відбуватись в онлайн формі за погодженням з керівником курсу.

Методи контролю

Основні форми участі здобувачів у навчальному процесі, що підлягають поточному контролю: виступ на практичних заняттях; доповнення, опонування до виступу, рецензія на виступ; участь у дискусіях; аналіз першоджерел; письмові завдання (тестові, індивідуальні роботи у формі рефератів); та інші письмові роботи, оформлені відповідно до вимог. Кожна тема курсу, що винесена на лекційні та практичні заняття, відпрацьовується Здобувачами у тій чи іншій формі, наведеній вище. Обов'язкова присутність на лекційних

заняттях, активність впродовж семестру, відвідування/відпрацювання усіх аудиторних занять, виконання інших видів робіт, передбачених навчальним планом з цієї дисципліни.

Під час оцінювання рівня знань здобувача аналізують підлягають:

- характеристики відповіді: цілісність, повнота, логічність, обґрунтованість, правильність;
- якість знань (ступінь засвоєння фактичного матеріалу): осмисленість, глибина, гнучкість, дієвість, системність, узагальненість, міцність;
- ступінь сформованості уміння поєднувати теорію і практику під час розгляду ситуацій, практичних завдань;
- рівень володіння розумовими операціями: вміння аналізувати, синтезувати, порівнювати, абстрагувати, узагальнювати, робити висновки з проблем, що розглядаються;
- досвід творчої діяльності: вміння виявляти проблеми, розв'язувати їх, формувати гіпотези;
- самостійна робота: робота з навчально-методичною, науковою, допоміжною вітчизняною та зарубіжною літературою з питань, що розглядаються, вміння отримувати інформацію з різноманітних джерел (традиційних; спеціальних періодичних видань, ЗМІ, Internet тощо).

Тестове опитування може проводитись за одним або кількома змістовими модулями. В останньому випадку бали, які нараховуються Здобувачу за відповіді на тестові питання, поділяються між змістовими модулями.

Індивідуальне завдання підлягає у виконанні РГР згідно індивідуального варіанту для закріплення теоретичних знань і практичних навичок з інтелектуального аналізу даних за 5 семестр.

Також як виконання індивідуального завдання за рішенням викладача може бути зарахована участь Здобувача у міжнародній або всеукраїнській науково-практичній конференції з публікацією у матеріалах конференції тез виступу (доповіді) на одну з тем, дотичних до змісту освітньої компоненти, або публікація статті на одну з таких тем в інших наукових виданнях.

Результати **поточного контролю** заносяться до журналу обліку роботи. Позитивна оцінка поточної успішності Здобувачів за відсутності пропущених та невідпрацьованих практичних занять та позитивні оцінки за індивідуальну роботу є підставою для допуску до підсумкової форми контролю. Бали за аудиторну роботу відпрацьовуються у разі пропусків.

Підсумковий контроль здійснюється під час проведення залікової сесії з урахуванням підсумків поточного та модульного контролю. Під час семестрового контролю враховуються результати задачі усіх видів навчальної роботи згідно зі структурою кредитів.

Оцінювання проводиться за 100-бальною шкалою.

Розподіл балів для дисципліни з формою контролю залік

Поточне оцінювання				Залік	Сума балів
Змістовий модуль 1	Змістовий модуль 2	Змістовий модуль 3	Індивідуальне завдання		
20	20	20	20	20	100

Шкала оцінювання індивідуальної роботи

Оцінка за національною шкалою	Кількість балів	Критерії
відмінно	20	відмінне виконання (розкриття теми, посилання та цитування сучасних наукових джерел (не старше 2019 року), дотримання норм доброчесності)
	15	відмінне виконання з незначною кількістю помилок виконання (розкриття теми, посилання та цитування сучасних наукових джерел (більшість з яких не старше 2019 року), дотримання норм доброчесності)
добре	12	виконання вище середнього рівня з кількома помилками (розкриття теми в межах об'єкта та завдань роботи, посилання та цитування сучасних наукових джерел (серед яких є такі, що не старше 2019 року), дотримання норм доброчесності)
	10	виконання з певною кількістю помилок (розкриття теми в межах об'єкта та завдань роботи, наявність посилань та цитувань наукових джерел, дотримання норм доброчесності)
задовільно	8	виконання роботи задовольняє мінімальним критеріям помилок (розкриття теми в основному в межах об'єкта роботи, наявність концептуального апарату роботи, присутність не менше 5 посилань та цитувань наукових джерел, дотримання норм доброчесності)

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
90-100	A	Відмінно
82-89	B	Добре
74-81	C	
64-73	D	Задовільно
60-63	E	
35-59	FX	Не зараховано з можливістю повторного складання
0-34	F	Не зараховано з обов'язковим повторним вивченням дисципліни

Умови допуску до підсумкового контролю

Здобувачу, який має підсумкову оцінку за дисципліну від 35 до 59 балів, призначається додаткова залікова сесія. В цьому разі він повинен виконати додаткові завдання, визначені викладачем.

Здобувач, який не виконав вимог робочої програми за змістовими модулями, не допускається до складання підсумкового контролю. В цьому разі він повинен виконати визначене викладачем додаткове завдання за темами відповідних змістових модулів в період між основною та додатковою сесіями.

Здобувач має право на опротестування результатів контролю (апеляцію). Правила подання та розгляду апеляції визначені внутрішніми документами КНУБА, які розміщені на сайті КНУБА та зміст яких доводиться Здобувачам до початку вивчення дисципліни.

Методичне забезпечення дисципліни

Основна: (Базова)

1. М.В. Талах, В.В. Дворжак Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак – Чернівці: Технодрук, 2022. – 367 с.
2. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 2021. 92 с
3. Іванов С.М., Максишко Н.К., Бречко Д.О. Інтелектуальний аналіз даних: конспект лекцій. Запоріжжя: ЗНУ, 2020, 156 с.

Додаткова:

4. Ліщинська Л. Б., Добровольська Н. В. Перспективні програмні інструменти для аналізу даних у бізнесі Вісник Хмельницького національного університету Серія: «Технічні науки» №1, 2022. С.78-83 DOI: <https://www.doi.org/10.31891/2307-5732-2022-305-1>
5. Болюбаш Н. М. Методичні вказівки до виконання курсової роботи з дисципліни «Інтелектуальний аналіз» даних для студентів спеціальності 122 «Комп'ютерні науки» : методичні вказівки / Н. М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2021. – 28 с. – (Методична серія ; вип. 347). Режим доступу: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/478>
6. Yanchang Zhao and Yonghua Cen (Eds.). Data Mining Applications with R.
7. К. Chupryna, I. Ivakhnenko, S. Biloshchytska, С. Mykhaylo, D. Ryzhakov and D. Sobol, "Formalized Management of Changes at the Enterprise by Means of Fuzzy Logic," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023, pp. 490-494, <https://doi.org/10.1109/SIST58284.2023.10223567>
8. Biloshchytskyi, A., Omirbayev, S., Mukhatayev, A., Biloshchytska, S., Toxanov, S., & Faizullin, A. (2024). The concept of the Internet of Things in the development of information and analytical systems based on the method of constructing a scalar assessment of the results of research activities of scientists. Procedia Computer Science, 231, 684–690. <https://doi.org/10.1016/j.procs.2023.12.161>

Додаткові ресурси:

9. Evolutionary Complexity Research Group (Eplex): <http://eplex.cs.ucf.edu/>
10. R: <https://www.r-project.org/>
11. The Comprehensive R Archive Network – <https://cran.r-project.org/>
12. RStudio - <https://posit.co/download/rstudio-desktop/>