

Київський національний університет
будівництва і архітектури
Кафедра інформаційних технологій

Шифр спеціальності	Назва спеціальності, освітньої програми	Освітній рівень
122	Комп'ютерні науки, Інформаційні управляючі системи і технології	Бакалавр

«Затверджую»

Завідувачка кафедри
Тетяна ГОНЧАРЕНКО



Розробник силабусу
Світлана БІЛОЩИЦЬКА



СИЛАБУС

OK22 «Інтелектуальний аналіз даних»

(назва, шифр освітньої компоненти (дисципліни))

1) Статус освітньої компоненти: обов'язкова	
2) Контактні дані викладача: (д.т.н, доцент Білощицька С.В., biloshchytska.sv@knuba.edu.ua , +380 (93) 825 35 93, https://www.knuba.edu.ua/elementor-161588/)	
3) Пререквізити: «Програмування та алгоритмічні мови», «Інструментальні засоби програмування», «Чисельні методи в інформатиці», «Теорія алгоритмів».	
4) Коротка анотація дисципліни Дисципліна здійснює теоретичну та практичну підготовку 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.	
5) Структура курсу: лекції, лабораторні роботи, самостійні роботи, розрахунково-графічна робота, залік	
Загальна кількість кредитів ECTS	3,0
Сума годин:	90
Вид індивідуального завдання	РГР
Форма контролю	Залік
6) Зміст курсу: <i>Змістовний модуль 1. Основи статистичного аналізу та обробки даних</i> <i>Лекція 1.</i> Методи первісної обробки даних. Класифікація ознак за шкалами вимірювання. Описова статистика. Числові характеристики вибірки <i>Лекція 2.</i> Постановка задачі побудови статистичного зведення. Статистичне групування, його задачі та види. Принципи побудови статистичних групувань. Ряди розподілу. Статистичні таблиці. Статистичні графіки. <i>Лекція 3.</i> 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 багатомірного 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 занять

Лабораторна робота №1. Основи роботи в системі R.

Лабораторна робота №2. Перевірка статичних гіпотез.

Лабораторна робота №3. Задачі відновлення регресії.

Лабораторна робота №4. Кластеризація в R.

Лабораторна робота №5. Побудова дерев рішень.

Лабораторна робота №6. Метод опорних векторів.

Лабораторна робота №7. Використання бустінг алгоритму в задачах класифікації.

Лабораторна робота №8. Дослідження вибірки за допомогою функції bootstrap.

Самостійна робота:

1. Дослідження видів помилок вибірки та їх вплив на числові характеристики.
2. Принципи побудови статистичних зведень і групувань. Огляд рядів розподілу, таблиць і графіків.
3. Основи кореляційного і регресійного аналізу. Теоретичні аспекти методу найменших квадратів.
4. Дослідження методів оцінки якості кластеризації (індекс Дейвіса-Болдіна, силуетний аналіз).
5. Вивчити вплив збалансованості вибірки на точність класифікації.
6. Ознайомитися з методами побудови ансамблів дерев рішень (Random Forest, Boosting).
7. Дослідження використання алгоритму ECLAT у задачах пошуку частих наборів елементів.
8. Ознайомлення з концепцією кубів даних у реляційних базах (ROLAP-моделі).
9. Ознайомлення з методами очищення та трансформації даних перед Data Mining.
10. Вивчення методики оцінки точності прогнозів часових рядів (MAPE, RMSE).

Індивідуальна робота: Виконання РГР згідно індивідуального варіанту для закріплення теоретичних знань і практичних навичок з інтелектуального аналізу даних за 5 семестр. На виконання РГР відводиться 12 годин самостійної роботи.

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

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