



Силабус освітнього компонента Програма навчальної дисципліни



Інтелектуальний аналіз даних

Шифр та назва спеціальності

122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Комп'ютерні науки.

Кафедра

Системного аналізу та інформаційно-аналітичних технологій

Рівень освіти

Магістр

Тип дисципліни

Спеціальна (фахова), Обов'язкова

Семестр

9

Мова викладання

Українська,

Викладачі, розробники



Мельников Олег Станіславович

Oleg.Melnikov@khpi.edu.ua

Кандидат економічних наук, доцент, доцент кафедри системного аналізу та інформаційно-аналітичних технологій НТУ "ХПІ".

Досвід роботи – понад 30 років. Автор біля 130 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Основи теорії систем та системного аналізу», «Випадкові процеси», «Інтелектуальний аналіз даних», «Інформаційні технології бізнес-аналітики»

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Предметом дисципліни є інтелектуальні технології виявлення в даних раніше невідомих, нетривіальних, практично корисних та доступних інтерпретації знань, необхідних для прийняття рішень в різних сферах людської діяльності.

Мета та цілі дисципліни

Оволодіння теоретичним апаратом інтелектуального аналізу даних та практичними навичками його використання для вирішення практичних задач. Знайомство з методикою проведення, основними алгоритмами та програмними засобами інтелектуального аналізу даних.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК05. Здатність вчитися й оволодівати сучасними знаннями.
ЗК06. Здатність бути критичним і самокритичним.
ЗК07. Здатність генерувати нові ідеї (креативність).
СК01. Усвідомлення теоретичних засад комп'ютерних наук.
СК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.
СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.
СК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень.
СК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.
СК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.
СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.
СК13. Здатність проєктувати, розробляти та використовувати складні інформаційні системи для вирішення практичних задач у галузі комп'ютерних наук, в тому числі з використанням систем штучного інтелекту.

Результати навчання

РН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.
РН6. Розробляти концептуальну модель інформаційної або комп'ютерної системи.
РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.
РН8. Розробляти математичні моделі та методи аналізу даних (включно з великими).
РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).
РН10. Проєктувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.
РН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.
РН19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.

Обсяг дисципліни

Загальний обсяг 120 год. (4 кредити ECTS): лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Базові знання теорії ймовірностей та математичної статистики, навички алгоритмізації.

Особливості дисципліни, методи та технології навчання

Лекції проводяться інтерактивно з використанням мультимедійних технологій. На лабораторних заняттях використовуються командні методи роботи; акцентується увага на використанні інструментарію дисципліни для аналізу реальних даних. Навчальні матеріали, в тому числі відеозаписи лекцій, доступні студентам через засоби Microsoft Teams.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Предмет та задачі інтелектуального аналізу даних (ІАД)

1. Предмет інтелектуального аналізу даних
2. Основні задачі Data Mining

3. Типи вихідних даних для ІАД
4. Шкали вимірювання
5. Інформаційне середовище ІАД

Тема 2. Імовірнісний підхід до аналізу даних

1. Випадкові події та їх імовірності
2. Умовні ймовірності. Теорема Байєса
3. Випадкові величини та їх характеристики
4. Системи випадкових величин
5. Визначення залежностей між випадковими величинами

Тема 3. Основні поняття математичної статистики

1. Вибірки і способи їх подання.
2. Описові характеристики даних
3. Точкові оцінки параметрів
4. Інтервальні оцінки параметрів
5. Перевірка статистичних гіпотез
6. Визначення залежностей між даними
7. Таблиці спряженості
8. Дисперсійний аналіз

Тема 4. Лінійний регресійний аналіз

1. Специфікація лінійної моделі
2. Матрична форма лінійної моделі
3. Метод найменших квадратів
4. Теорема Гауса-Маркова
5. Оцінка статистичної значущості параметрів моделі
6. Оцінка адекватності моделі
7. Перетворення змінних в регресійному аналізі

Тема 5. Задача класифікації: загальні положення

1. Математична постановка задачі класифікації
2. Способи побудови функції класифікації
3. Поняття про класифікаційні правила
4. Алгоритм «1-RULE»
5. Способи оцінки якості класифікації

Тема 6. Ординарні алгоритми класифікації

1. Метод опорних векторів
2. Метод найближчого сусіду
3. Дерева рішень
4. Інформація та ентропія
5. Алгоритм ID3 для побудови дерев рішень

Тема 7. Імовірнісні алгоритми класифікації

1. Байєсівський підхід до аналізу даних
2. Наївний байєсівський класифікатор
3. Технічні деталі реалізації алгоритмів байєсівської класифікації
4. Біноміальна логістична регресія
5. Мультиноміальна логістична регресія

Тема 8. Пошук шаблонів та асоціативні правила

1. Поширені підмножини в базах даних транзакцій
2. Асоціативні правила
3. Критерії оцінки значущості виявлених правил
4. Алгоритм Apriori
5. Алгоритм Eclat

Тема 9. Задачі кластеризації

1. Поняття кластеризації. Відмінності між задачами класифікації та кластеризації
2. Метричні простори. Способи визначення відстані між об'єктами та кластерами
3. Алгоритм k-середніх
4. Ієрархічна кластеризація
5. Алгоритм максимізації очікувань
6. Визначення кількості кластерів

Теми лабораторних робіт

Тема 1. Виявлення взаємозв'язків в статистичних даних

Тема 2. Оцінка параметрів багатовимірної регресійної моделі

Тема 3. Розв'язання задачі побудови класифікаційних правил за допомогою алгоритму «1-RULE»

Тема 4. Побудова дерев рішень за допомогою алгоритму ID3

Тема 5. Класифікація методом опорних векторів

Тема 6. Побудова наївного байєсівського класифікатора

Тема 7. Побудова асоціативних правил для аналізу споживчого кошику за алгоритмом Apriori

Тема 8. Кластеризація методом k -середніх

Самостійна робота

Курс передбачає виконання індивідуального розрахункового завдання з розробки регресійної моделі для аналізу даних. Результати розрахунків оформлюються у письмовий звіт.

Студентам також рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення окремих тем дисципліни.

Література та навчальні матеріали

Основна література

1. Мельников О.С. Интеллектуальный анализ данных: навчально-методичний посібник. Харків: Вид-во НТУ «ХПІ», 2023. 196 с. <https://repository.kpi.kharkov.ua/handle/KhPI-Press/72877>.
2. Олійник А.О. Субботін С.О., Олійник О.О. Интеллектуальный анализ данных : навчальний посібник. Запоріжжя : ЗНТУ, 2012. 278 с.
3. Ситник В.Ф., Краснюк М.Т. Интеллектуальный анализ данных (дейтамайнінг) : Навч. посібник. Київ : КНЕУ, 2007. 376 с.
4. Черняк О.І., Захарченко П.В. Интеллектуальный анализ данных: Підручник. Київ : Знання, 2014. 599 с.
5. Ozdemir, S. , Kakade, S., Tibaldeschi, M. Principles of Data Science. 2nd edition. Birmingham-Mumbai: Packt Publishing, 2018. 420 p.

Допоміжна література

1. Грін В. Економетричний аналіз. (Green W. Econometric Analysis. New York: Macmillan, 2000). Переклад і наукова редакція О.В.Комашка, передмова О.І.Черняка. Київ: Основи, 2005.
2. Огірко, О. І., Галайко Н.В. Теорія ймовірностей та математична статистика: навчальний посібник. Львів: ЛьвДУВС, 2017. 292 с.
3. Zaki M.J., Meira W. Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, 2014. □ 595 p.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді екзамену (40%) та поточного оцінювання (60%).

Екзамен: письмове завдання (2 запитання з теорії + розв'язання задачі) та усна доповідь.

Поточне оцінювання: 8 лабораторних робіт (40%) та розрахункове завдання (20%).

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

Дата погодження, підпис

Завідувач кафедри
Юрій ДОРОФЄЄВ

Дата погодження, підпис

Гарант ОП
Марія ПЕТРЕНКО