



Силабус освітнього компонента Програма навчальної дисципліни



Інтелектуальний аналіз даних

Шифр та назва спеціальності

F3 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Спеціалізація

–

Кафедра

Системного аналізу та інформаційно-аналітичних технологій (322)

Освітня програма

Комп'ютерні науки

Тип дисципліни

Спеціальна (фахова), обов'язкова

Рівень освіти

Магістр

Форма освіти

Денна, заочна

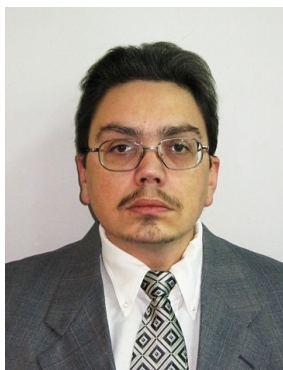
Семестр

9

Мова викладання

Українська

Викладачі, розробники

**Мельников Олег Станіславович**

Oleg.Melnikov@khpі.edu.ua

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

Досвід роботи – понад 30 років. Автор біля 140 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.

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 проектувати, 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.

РН2. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур.

РН3. Зрозуміло і недвозначно доносити власні знання, висновки та аргументацію у сфері комп'ютерних наук до фахівців і нефахівців, зокрема до осіб, які навчаються.

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 великим).

РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

РН10. Проектувати 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

Лекції

Теми лекцій

Кількість годин

Тема 1. Предмет та задачі інтелектуального аналізу даних (ІАД) Предмет інтелектуального аналізу даних. Основні задачі Data Mining. Типи вихідних даних для ІАД. Шкали вимірювання. Інформаційне середовище ІАД	2
Тема 2. Імовірнісний підхід до аналізу даних Випадкові події та їх імовірності. Умовні ймовірності. Теорема Байєса. Випадкові величини та їх характеристики. Системи випадкових величин. Визначення залежностей між випадковими величинами	2
Тема 3. Основні поняття математичної статистики Вибірki і способи їх подання. Описові характеристики даних. Точкові оцінки параметрів. Інтервальні оцінки параметрів. Перевірка статистичних гіпотез. Визначення залежностей між даними. Таблиці спряженості. Коваріація та кореляція. Дисперсійний аналіз	4
Тема 4. Лінійний регресійний аналіз Специфікація лінійної моделі. Матрична форма лінійної моделі. Метод найменших квадратів. Теорема Гауса-Маркова. Оцінка статистичної значущості параметрів моделі. Оцінка адекватності моделі. Перетворення змінних в регресійному аналізі	4
Тема 5. Задача класифікації: загальні положення Математична постановка задачі класифікації. Способи побудови функції класифікації. Поняття про класифікаційні правила. Алгоритм «1-RULE». Способи оцінки якості класифікації	2
Тема 6. Ординарні алгоритми класифікації Метод опорних векторів. Метод найближчого сусіду. Дерева рішень. Інформація та ентропія. Алгоритм ID3 для побудови дерев рішень	4
Тема 7. Імовірнісні алгоритми класифікації Байєсівський підхід до аналізу даних. Наївний байєсівський класифікатор. Технічні деталі реалізації алгоритмів байєсівської класифікації. Біноміальна логістична регресія. Мультиноміальна логістична регресія	4
Тема 8. Задачі кластеризації Поняття кластеризації. Відмінності між задачами класифікації та кластеризації. Метричні простори. Способи визначення відстані між об'єктами та кластерами. Алгоритм k-середніх. Ієрархічна кластеризація. Алгоритм максимізації очікувань. Визначення кількості кластерів	4
Тема 9. Пошук шаблонів та асоціативні правила Поширені підмножини в базах даних транзакцій. Асоціативні правила. Критерії оцінки значущості виявлених правил. Алгоритм Apriori. Алгоритм Eclat	2
Тема 10. Інструменти Python для інтелектуального аналізу даних Візуалізація даних з бібліотеками Matplotlib та Seaborn. Операції з табличними даними з бібліотекою Pandas. Операції з векторами, матрицями та тензорами з бібліотекою NumPy. Статистичний аналіз даних з пакетом statsmodels. Машинне навчання з пакетами scikit-learn, TensorFlow та PyTorch.	4
Загальна кількість годин	32

Лабораторні заняття

Теми лабораторних робіт	Кількість годин	Вагові коефіцієнти а
Тема 1. Перетворення даних у різні шкали виміру	2	1
Тема 2. Визначення залежностей між даними	2	1
Тема 3. Розв'язання задачі побудови класифікаційних правил за допомогою алгоритму «1-RULE»	2	1
Тема 4. Побудова дерев рішень за алгоритмом ID3	2	1
Тема 5. Класифікація методом опорних векторів	2	1
Тема 6. Побудова наївного байєсівського класифікатора	2	1
Тема 7. Кластеризація методом k-середніх	2	1
Тема 8. Побудова асоціативних правил для аналізу споживчого кошику за алгоритмом Apriori	2	1
Загальна кількість годин	16	$\sum_{i=1}^n a_i = 8$

Самостійна робота

До самостійної роботи відноситься опрацювання лекційного матеріалу (8 годин), підготовка до лабораторних робіт (8 годин), самостійне вивчення тем, які не викладаються на лекційних заняттях (40 годин), та виконання індивідуального розрахункового завдання (16 годин). Студентам надаються додаткові матеріали (відео, статті) для самостійного вивчення окремих тем дисципліни

Опрацювання теоретичного матеріалу

Теми для самостійного вивчення	Кількість годин
Тема 1. Основні розподіли математичної статистики та їх властивості	4
Тема 2. Z-,T-,F- та хі-квадрат статистичні тести	6
Тема 3. Основи попередньої обробки та очищення даних	4
Тема 4. Методи зниження розмірності даних	6
Тема 5. Методи щільнісної кластеризації	6
Тема 6. Методи згладжування даних	4
Тема 7. Методи аналізу часових рядів	6
Тема 8. Методи виявлення в даних аномалій	4
Загальна кількість годин	40

Індивідуальне завдання

Індивідуальне завдання містить комплекс завдань, пов'язаних з побудовою регресійних моделей для аналізу даних. Завдання призначається за варіантами; номер варіанта збігається з номером здобувача освіти у списку групи. За результатами завдання оформлюється письмовий звіт відповідно до вимог, наведених у джерелі [1]. Завдання виконується протягом 3 навчальних тижнів і подається на перевірку до залікового тижня.

Неформальна освіта

Здобувач має можливість зарахувати окремі теми або курс шляхом: проходження професійних курсів чи тренінгів, онлайн-освіти, професійних стажувань у сфері, що відповідає навчальним цілям дисципліни.

Для зарахування необхідно надати: сертифікат (електронний або друкований) про проходження курсу/стажування, опис програми тренінгу із зазначенням змісту тем, обсягу та тривалості.

Рекомендовані курси, тренінги, стажування

1. CS50: Вступ до штучного інтелекту з Python
<https://prometheus.org.ua/prometheus-free/cs50-introduction-to-ai-with-python/>
2. Introduction to Data Mining (англ. мовою)
<https://www.simplilearn.com/free-introduction-to-data-mining-course-skillup>
3. Data Mining Methods (англ. мовою)
<https://www.coursera.org/learn/data-mining-methods>

Література та навчальні матеріали

Основна література

1. Мельников О.С. Інтелектуальний аналіз даних: 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. Інтелектуальний аналіз даних : навчальний посібник. Запоріжжя : ЗНТУ, 2012. 278 с.
3. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг) : Навч. посібник. Київ : КНЕУ, 2007. 376 с.
4. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: 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. Мазманішвілі О. С., Мельников О. С. Практикум з теорії ймовірностей : навч. посібник. Харків : НТУ "ХПІ", 2025. – 320 с. – URI: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/88582>
2. Мазманішвілі О. С., Мельников О. С. Практикум з математичної статистики : навч. посібник. Харків : НТУ "ХПІ", 2025. – 288 с. – URI: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/88593>
3. Zaki M.J., Meira W. Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press, 2014. - 595 p.

Система оцінювання

Підсумкова оцінка з освітнього компонента визначається відповідальним лектором за темами, видами занять, тощо відповідно до силабусу і є інтегральною оцінкою результатів усіх вид навчальної діяльності здобувача вищої освіти. Підсумкова оцінка повинна відображати всі оцінки за складовими навчального процесу з урахуванням їх вагових показників k :

Поточний контроль (лабораторні роботи), k_1	Контрольні роботи (за наявності), k_2	Індивідуальне завдання (за наявності), k_3	Підсумковий контроль (для ОК з іспитом), k_4
0,4	0,0	0,3	0,3

Сума коефіцієнтів повинна складати одиницю: $k_1+k_2+k_3+k_4=1$. Підбір вагових коефіцієнтів підсумкової оцінки здійснює розробник курсу.

Розрахунок підсумкової оцінки проводиться за формулою:

$$O = П \cdot k_1 + K \cdot k_2 + I \cdot k_3 + Пк \cdot k_4,$$

де: $П$ – середньозважена середня оцінка за поточний контроль,

I – оцінка за виконання індивідуального завдання,

K – середньозважена оцінка за контрольні роботи,

$Пк$ – оцінка за підсумковий контроль.

$$П = \frac{П_1 \cdot a_1 + П_2 \cdot a_2 + \dots + П_n \cdot a_n}{\sum_{i=1}^8 a_i},$$

де: a_i - ваговий коефіцієнт за лабораторне заняття.

$$K = \frac{K_1 \cdot b_1}{\sum_{i=1}^1 b_i},$$

де: b_i - ваговий коефіцієнт за контрольну роботу.

Поточні оцінки за кожну складову ($П, K, I, \dots$) виставляються за 100-бальною шкалою згідно з положенням «Про критерії та систему оцінювання знань та вмінь і про рейтинг здобувачів вищої освіти» НТУ «ХПІ».

Підсумкова оцінка виставляється відповідно до розрахованої O з округленням до найближчого цілого числа в більшу сторону.

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

Здобувач вищої освіти 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 ДОРОФЕЄВ

Дата погодження, підпис

Гарант ОП
Максим МАЛЬКО