

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ



ЗАТВЕРДЖУЮ

Декан факультету інформаційних технологій

проф. Тетяна ГОВОРУЩЕНКО

2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Інтелектуальний аналіз даних

Галузь знань – 12 Інформаційні технології

Спеціальність – 122 Комп'ютерні науки

Рівень вищої освіти – Перший бакалаврський

Освітньо-професійна програма – Комп'ютерні науки

Обсяг дисципліни – 5 кредитів ЄКТС

Шифр дисципліни – ОПП.08

Мова навчання – українська

Статус дисципліни: обов'язкова (цикл професійної підготовки)

Факультет – Інформаційних технологій

Кафедра – Комп'ютерних наук

Курс	Семестр	Загальне навантаження		Кількість годин						Курсовий проєкт	Курсова робота	Форма семестрового контролю	
		Кредити ЄКТС	Години	Аудиторні заняття				Індивідуальна робота студента	Самостійна робота, у т.ч. ІРС			Залік	Іспит
				Разом	Лекції	Лабораторні роботи	Практичні заняття						
3	5	5	150	68	34	34			82			+	
Разом		5	150	68	34	34			82			1	

Робоча програма складена на основі освітньо-професійної програми «Комп'ютерні науки» та стандарту вищої освіти зі спеціальності 122 – Комп'ютерні науки першого (бакалаврського) рівня вищої освіти.

Програма складена

 док. філ. Павло РАДЮК

 асист. каф. КН Леонід ВОЗНІЮК

Схвалена на засіданні кафедри комп'ютерних наук

Протокол від 30 серпня 2024 р. № 1. Зав. кафедри  проф. Олександр БАРМАК

Робоча програма розглянута та схвалена вченою радою факультету інформаційних технологій

Голова вченої ради факультету  проф. Тетяна ГОВОРУЩЕНКО

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Тип (статус) дисципліни	Обов'язкова професійної підготовки
Рівень вищої освіти	Перший (бакалаврський)
Мова викладання	Українська
Семестр	5
Кількість призначених кредитів ЄКТС	5
Форми здобуття освіти, для яких викладається дисципліна	Денна

Результати навчання. Студент, який успішно завершив вивчення дисципліни, повинен: *освоїти* технологію інтелектуального аналізу даних засобами Python; *вміти проводити* інтелектуальний аналіз даних за допомогою алгоритмів Data Mining засобами Python; *вміло використовувати* понятійний апарат; *уміти виконувати* основні операції над даними, задачі інтелектуального аналізу даних; *визначати* основні властивості предметної області та відповідні методики для розв'язання задач; *характеризувати* оптимальні моделі та фундаментальні принципи дослідження заданої предметної області; *підбирати* алгоритми та програмне забезпечення з метою розв'язання класу задач; *проектувати* за різними методиками типові завдання та способи їхнього розв'язання за класом і призначенням; *виконувати* масштабування та проектування задач, перенесення та узагальнення систем.

Зміст навчальної дисципліни. Первинний аналіз даних з Pandas. Візуальний аналіз даних з Python. Задачі класифікації та регресії. Лінійні моделі класифікації та регресії. Композиції (ансамблі) алгоритмів. Побудова та відбір ознак. Методи машинного навчання без учителя. Онлайн-навчання. Аналіз часових рядів з Python.

Пререквізити – Теорія ймовірності та математична статистика, Методи та системи штучного інтелекту, Об'єктно-орієнтоване програмування.

Кореквізити – Технології розподілених систем та паралельних обчислень.

Запланована навчальна діяльність: лекції – 34 год., лабораторні заняття – 34 год., самостійна робота – 82 год., разом – 150 год.

Форми (методи) навчання: лекції (з використанням методів візуалізації); лабораторні заняття (з використанням майстер-класів), самостійна робота.

Форми оцінювання результатів навчання: захист лабораторних робіт, письмове опитування (контрольна робота), підсумковий контроль.

Вид семестрового контролю: іспит.

Навчальні ресурси:

1. Мокін В.Б. Наука про дані: машинне навчання та інтелектуальний аналіз даних [Електронний навчальний посібник] / В.Б. Мокін, М.В. Дратований. Вінниця : ВНТУ, 2024. 263 с. URL: <https://t.ly/VYK9h>
2. Болюбаш Н.М. Інтелектуальний аналіз даних : навч. посіб. / Н. М. Болюбаш. Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. 320 с. URL: <https://t.ly/Dht4x>
3. Талах М.В. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак – Чернівці: Технодрук, 2022. 367 с. URL: <https://t.ly/B2GLZ>
4. Модульне середовище. Режим доступу : <https://msn.khnu.km.ua/>
5. Електронна бібліотека університету. Режим доступу : http://lib.khmnmu.edu.ua/asp/php_f/plage_lib.php
6. Репозитарій ХНУ. Режим доступу : <https://library.khmnmu.edu.ua/#>

Викладачі: доктор філософії, старший викладач кафедри комп'ютерних наук Павло РАДЮК; викладач-стажист кафедри комп'ютерних наук Леонід ВОЗНЮК.

Пояснювальна записка

Дисципліна «Інтелектуальний аналіз даних» є обов'язковою дисципліною професійної підготовки здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 122 «Комп'ютерні науки».

Курс призначений для підготовки бакалаврів з галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки. На основі загальних понять теорії ймовірностей та математичної статистики, а також методів та системи штучного інтелекту дисципліна розглядає методи та підходи до отримання та аналізу нової ціннісної інформації з даних.

Пререквізити – Теорія ймовірності та математична статистика, Методи та системи штучного інтелекту, Об'єктно-орієнтоване проєктування.

Кореквізити – Технології розподілених систем та паралельних обчислень.

Відповідно до Стандарту вищої освіти із зазначеної спеціальності та освітньої програми дисципліна має забезпечити:

– **компетентності:**

ЗК07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК11. Здатність приймати обґрунтовані рішення.

ФК 02. 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.

ФК 11. 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.

ФК 17. Здатність проєктувати, розробляти, налагоджувати та вдосконалювати програмне забезпечення експертних систем, з відповідними моделями подання знань, методами, алгоритмами обчислень і структурами даних і знань.

– **програмні результати навчання:**

ПРН 03. 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.

ПРН 04. 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.

ПРН 08. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проєктування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

ПРН 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.

ПРН 19. Проєктувати, розробляти, адаптувати та експлуатувати експертні системи за різними моделями подання знань для вирішення типових та новітніх задач у галузі комп'ютерних наук, що характеризуються комплексністю та невизначеністю умов.

Мета дисципліни – формування в здобувачів фундаментальних знань про принципи роботи з даними; оволодіння студентами математичними та комп'ютерними інструментами для отримання ціннісної інформації зі статистичних даних.

Предмет дисципліни. Дані, основи інтелектуального аналізу даних, методи та алгоритми інтелектуального аналізу даних, методи пониження розмірності, методи класифікації та прогнозування, методи кластеризації, методи пошуку шаблонів даних, мова програмування Python, програмні інструменти мови програмування Python для інтелектуального аналізу даних.

Завдання дисципліни. Надання студентам знань та практичних навичок розв'язування задач, пов'язаних з інтелектуальним обробленням статистичних даних.

Результати навчання. Студент, який успішно завершив вивчення дисципліни, повинен:

освоїти технологію інтелектуального аналізу даних засобами Python; вміти *проводити* інтелектуальний аналіз даних за допомогою алгоритмів Data Mining засобами Python; вміло *використовувати* понятійний апарат; уміти *виконувати* основні операції над даними, задачі інтелектуального аналізу даних; *визначати* основні властивості предметної області та відповідні методики для розв'язання задач; *підбирати* алгоритми та програмне забезпечення з метою розв'язання класу задач; *проектувати* за різними методиками типові завдання та способи їхнього розв'язання за класом і призначенням; *виконувати* масштабування та проектування задач, перенесення та узагальнення систем.

Структура залікових кредитів дисципліни

Назва теми	Кількість годин, відведених на:		
	лекції	лаб. заняття	СРС
Тема 1. Первинний аналіз даних з Pandas.	4	4	8
Тема 2. Візуальний аналіз даних з Python.	4	4	8
Тема 3. Задачі класифікації та регресії.	4	4	8
Тема 4. Лінійні моделі класифікації та регресії.	4	4	12
Тема 5. Композиції (ансамблі) алгоритмів.	4	4	8
Тема 6. Побудова та відбір ознак.	4	4	8
Тема 7. Методи машинного навчання без учителя.	4	4	8
Тема 8. Онлайн-навчання.	4	4	12
Тема 9. Аналіз часових рядів з Python.	2	2	10
Разом за семестр:	34	34	82

Програма навчальної дисципліни

Зміст лекційного курсу*

Номер лекції	Перелік тем лекцій, їх анотації	Кількість годин
	<i>П'ятий семестр</i>	
1	Первинний аналіз даних з Pandas. Розгляд проблеми опрацювання даних. Демонстрація основних методів Pandas. Читання з файлу і первинний аналіз. Індексуння та вилучення даних. Літ.: [1], розділи 1, 4 та 6.	2
2	Об'єкти та класи в Python. Використання спеціалізованих функцій. Групування даних. Зведені таблиці. Перетворення датафреймів. Розв'язування задачі прогнозування відтоку клієнтів. Літ.: [1], розділи 4; [2], розділ 1.	2
3	Візуальний аналіз даних з Python. Розгляд проблеми інтерпретування аналізу даних. Візуалізація даних. Основні бібліотеки Python для візуального аналізу даних. Matplotlib. Seaborn. Plotly. Літ.: [1], розділ 5.	2
4	Візуальна аналітика для задач інтелектуального аналізу даних. Гістограми. Теплові мапи. Коробковий графік. Метод t-SNE. Літ.: [1], розділ 5.	2
5	Метод дерева рішень для задач класифікації та регресії. Розгляд проблеми отримання ціннісної інформації. Ентропія. Алгоритм побудови дерева рішень. Критерії якості розбиття для задачі класифікації. Дерево рішень з кількісними ознаками. Літ.: [1], розділ 9; [2], розділ 3; [7]; розділ 1.3.	2
6	Метод найближчих сусідів для задач класифікації та регресії. Вибір параметрів моделі найближчих сусідів. Крос-валідація. Переваги та недоліки методу найближчих сусідів. Літ.: [2], розділ 3; [5]; розділ 1.4.	2

7	Лінійні моделі регресії. Розгляд задачі передбачення. Метод найменших квадратів. Метод максимальної правдоподібності. Розкладання помилки на зміщення і розкид. Регуляризація лінійної регресії. L2-регуляризація логістичної функції втрат. Літ.: [1], розділи 6 та 8; [2], розділ 2; [7]; розділи 1 та 2.	2
8	Лінійні моделі класифікації. Формальна модель класифікації. Логістична регресія, як лінійний класифікатор. Логістична регресія з поліноміальними ознаками. Задача класифікації відгуків IMDB до фільмів. Криві валідації й навчання. Літ.: [1], розділи 6 та 8; [2], розділ 3.	2
9	Ансамблі алгоритмів: беггінг. Розгляд проблеми низького узагальнення моделі оброблення даних. Формалізація ансамблів. Бутстреп. Беггінг. Помилка out-of-bag. Перетворення ознак в багатовимірний простір. Зміщення. Літ.: [2], розділ 7.	2
10	Ансамблі алгоритмів: випадковий ліс. Алгоритм випадкового лісу. Порівняння випадкового лісу з деревом рішень і беггінгом. Параметри випадкового лісу. Варіація й декореляційний ефект. Важливість ознак на прикладі випадкового лісу. Літ.: [2], розділ 7.	2
11	Побудова та відбір ознак. Типи даних. Виявлення ознак. Перетворення ознак. Нормалізація та зміна розподілу. Заповнення пропусків. Статистичні підходи до відбору ознак. Літ.: [2], розділ 3, [7], розділ 2.	2
12	Основні метрики якості класифікації. Матриця невідповідностей. Характеристики бінарного класифікатора. ROC-крива. Площа під кривою AUC. Дисбаланс класів. Приклади застосувань метрик класифікації. Багатокласова класифікація. Літ.: [2], розділ 3, [7], розділ 2.	2
13	Методи пониження розмірності. Розгляд проблеми великої розмірності даних. Метод головних компонент. Приклад використання методу пониження розмірності. Метод t-SNE. Сингулярне розкладання матриці, SVD. Літ.: [3], розділ 4.	2
14	Методи кластерного аналізу. Розгляд проблеми оброблення неструктурованих даних. Метод k -середніх. Метод поширення спорідненості. Метод ієрархічної кластеризації. Вибір кількості кластерів. Метод ліктя. Метрики якості кластеризації. Літ.: [3], розділ 4.	2
15	Онлайн-підхід до навчання. Робота з категоріальними ознаками. Label Encoding. One-Hot Encoding. Hashing trick. Бібліотека Vowpal Wabbit. Літ.: [3], розділ 2.	2
16	Стохастичний градієнтний спуск. Бінарна класифікація. Багатокласова класифікація методом опорних векторів. Задача класифікації рецензій до фільмів IMDB. Задача класифікації питань на StackOverflow. Літ.: [3], розділ 3.	2
17	Аналіз часових рядів з Python. Розгляд задачі передбачення з трендом. Авторегресійні моделі. Експоненційне згладжування. Подвійне експоненційне згладжування. Модель Хольта-Вінтерса. Крос-валідація на часових рядах. Вилучення ознак на часових рядах. Літ.: [3], розділ 4.	2
	Разом за семестр:	34

Зміст лабораторних занять
Перелік лабораторних занять для студентів денної форми навчання

№ з/п	Тема лабораторного заняття	Кількість годин
<i>П'ятий семестр</i>		
1	Налаштування робочого середовища Python та Pandas. Літ.: [1], розділи 1, 4 та 6.	2
2	Аналіз даних за доходом населення UCI Adult. Літ.: [1], розділи 4; [2], розділ 1.	2
3	Налаштування бібліотек візуалізації даних середовища Python. Літ.: [1], розділ 5; [4], розділи 13 та 14.	2
4	Візуальний аналіз даних публікацій соціальних мереж. Літ.: [1], розділ 5; [4], розділи 13 та 14.	2
5	Функції для розрахунку ентропії та приросту інформації. Літ.: [1], розділ 9; [2], розділ 3; [5]; розділ 1.3.	2
6	Дерево рішень та метод найближчих сусідів на наборі даних UCI Adult. Літ.: [2], розділ 3; [2]; розділ 1.4.	2
7	Побудова моделі класифікації на основі лінійної регресії. Літ.: [1], розділи 6 та 8; [2], розділ 2; [5]; розділ 1.2.	2
8	Прогнозування популярності статей на TechMedia за допомогою лінійних регресій. Літ.: [1], розділи 6 та 8; [2], розділ 3.	2
9	Беггінг логістичних регресій для задачі кредитного скорингу. Літ.: [2], розділ 7; [4], розділ 7.	2
10	Випадковий ліс для задачі кредитного скорингу. Літ.: [2], розділ 7; [4], розділ 7.	2
11	Лінійна регресія для задачі визначення якості вина. Літ.: [2], розділ 2; [9], розділ 8.	2
12	Lasso-регресія та випадковий ліс для задачі визначення якості вина. Літ.: [2], розділ 2; [9], розділ 9.	2
13	Розв'язування задачі пониження розмірності. Літ.: [3], розділ 4; [4], розділ 10.	2
14	Розв'язування задачі визначення видів активності людини. Літ.: [3], розділ 4; [4], розділ 6.	2
15	Реалізація алгоритмів онлайн-навчання. Літ.: [3], розділ 2; [9], розділ 6.	2
16	Прогнозування затримок вильотів з допомогою градієнтного бустингу. Літ.: [3], розділ 3; [9], розділ 9.	2
17	Аналіз часових рядів з допомогою Python. Літ.: [3], розділ 4; [9], розділ 9.	2
Разом за семестр:		34

Зарахування результатів неформальної освіти

Окремі результати вивчення курсу можуть бути зараховані у випадку отримання студентом результатів навчання у неформальній освіті, що підтверджені відповідним документом (сертифікат, свідоцтво, освітня програма тощо):

– Як результат виконання ЛР №1–4 зараховується онлайн курс:

<https://pythoninstitute.org/python-essentials-1>

– Як результат виконання ЛР №5–8 зараховується онлайн курс:

<https://www.kaggle.com/learn/data-visualization>

– Як результат виконання ЛР №9–12 зараховується онлайн курс:

https://prometheus.org.ua/course/course-v1:IRF+ML101+2016_T3

– Як результат виконання ЛР №9–10 зараховується онлайн курс:

<https://www.coursera.org/learn/machine-learning#modules>

– Як результат виконання ЛР №11–17 зараховується онлайн курс:
<https://bit.ly/3LVYzR5>

Зміст самостійної (у т.ч. індивідуальної) роботи

Самостійна робота студентів денної форми навчання полягає у систематичному опрацюванні програмного матеріалу з відповідних джерел інформації, підготовці до виконання і захисту лабораторних робіт.

Самостійна робота студентів денної форми навчання полягає у систематичному опрацюванні програмного матеріалу, підготовці до виконання і захисту лабораторних робіт. Керівництво самостійною роботою та контроль за виконанням індивідуального завдання здійснює викладач, відповідно до розкладу консультацій у позаурочний час.

Самостійна робота дисципліни «Інтелектуальний аналіз даних» також передбачає виконання Індивідуальної самостійної роботи (ІСР) – покрокового створення проєкту в вигляді розробленого програмного забезпечення та звіту, відповідно до дев'яти індивідуальних завдань, що має на меті закріплення та розвиток вмінь і навичок, набутих під час виконання циклу лабораторних робіт. Оцінка, яка виставляється за виконання ІСР, виводиться як середня з оцінювання таких елементів (опис критеріїв наведено вище):

- якість та рівень розв'язання задач роботи;
- якість наповнення й оформлення звіту.

Номер тижня	Вид самостійної роботи	Кількість годин
<i>П'ятий семестр</i>		
1–2	Опрацювання теоретичного матеріалу з Т1, підготовка до захисту лабораторних робіт №1–2. Знайдіть реальний набір даних, виконайте його первинний аналіз з використанням Pandas: очистіть його від пропусків, проаналізуйте основні статистики і створіть індекси для ефективного доступу. Поясніть обраний підхід до обробки даних.	8
3–4	Опрацювання теоретичного матеріалу з Т2, підготовка до захисту лабораторних робіт №3–4. Створіть клас для обробки даних, який включає функції для групування та створення зведених таблиць; на основі даних про відтік клієнтів проведіть дослідження трендів. Запропонуйте рекомендації для прогнозування відтоку.	8
5–6	Опрацювання теоретичного матеріалу з Т3, підготовка до захисту лабораторних робіт №5–6. Застосуйте Matplotlib, Seaborn, та Plotly для візуального аналізу вибраного набору даних, створивши різні типи графіків. Проаналізуйте основні залежності та аномалії в даних, поясніть результати.	8
7–8	Опрацювання теоретичного матеріалу з Т4, підготовка до захисту лабораторних робіт №7–8. Побудуйте гістограми, теплові мапи та коробкові графіки для інтелектуального аналізу, досліджуючи кореляції між ознаками. Використайте метод t-SNE для візуалізації та інтерпретації багатовимірних даних.	12
9–10	Опрацювання теоретичного матеріалу з Т5, підготовка до захисту лабораторних робіт №9–10. Створіть та налаштуйте модель дерева рішень для вирішення задачі класифікації з власного набору даних, проаналізуйте критерії якості розбиття та побудуйте графік ентропії. Оцініть переваги та обмеження використання дерева рішень.	8
11–12	Опрацювання теоретичного матеріалу з Т6, підготовка до захисту лабораторних робіт №11–12. Проведіть класифікацію даних з використанням методу найближчих сусідів, протестуйте різні параметри та крос-валідацію для оптимізації моделі. Вивчіть складні випадки та оцініть обмеження моделі.	8
13–14	Опрацювання теоретичного матеріалу з Т7, підготовка до захисту	8

	лабораторних робіт №13–14. Побудуйте модель лінійної регресії для передбачення даних, застосовуючи методи регуляризації та оцінюючи точність. Проведіть аналіз зміщення і розкиду для отримання надійної моделі.	
15–16	Опрацювання теоретичного матеріалу з Т8, підготовка до захисту лабораторних робіт №15–16. Створіть логістичну регресію для класифікації відгуків з IMDb, використайте поліноміальні ознаки та побудуйте криві навчання для визначення узагальненості моделі.	12
17	Опрацювання теоретичного матеріалу з Т9, підготовка до захисту лабораторної роботи №17. Реалізуйте беггінг для моделі класифікації, оцініть ефективність методу бутстрепінгу і побудуйте графік out-of-bag помилки для зменшення розкиду моделі. Підготовка до підсумкового контрольного заходу.	10
	Разом за семестр:	82

Технології та методи навчання

Процес навчання з дисципліни ґрунтується на використанні традиційних та сучасних технологій, зокрема: лекції (з використанням методів візуалізації); лабораторні заняття (з використанням майстер-класів), самостійна робота (індивідуальна робота, курсове проектування).

Необхідні інструменти, обладнання, програмне забезпечення: комп'ютер (надається для використання в лабораторіях кафедри КН), Visual Studio Code (безоплатне), Anaconda Software (безоплатне) або інші безоплатні рішення.

Методи контролю

Поточний контроль здійснюється під час лекційних та лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу. Зокрема використовуються такі методи поточного контролю:

- усне опитування перед допуском до лабораторних занять;
- захист лабораторних робіт і формування звіту;
- тестовий контроль теоретичного матеріалу з теми;
- підсумковий контрольний захід.

При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу, який проводиться методом тестування з усього матеріалу дисципліни. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід (іспит), вважається невстигаючим.

Оцінювання результатів навчання студентів у семестрі

Оцінювання академічних досягнень здобувача вищої освіти здійснюється відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ». Кожний вид роботи з дисципліни оцінюється за інституційною чотирибальною шкалою і виставляється в електронному журналі обліку успішності.

Засвоєння студентом теоретичного матеріалу з дисципліни оцінюється під час захисту лабораторних робіт та підсумкового контрольного заходу. Поточний контроль передбачає захист лабораторних робіт. Поточний контроль здійснюється під час лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком навчального процесу.

Політика проведення лабораторних робіт з навчальної дисципліни «Інтелектуальний аналіз даних»:

1. Студент має бути присутнім на кожному занятті.
2. Студент може бути відсутнім на занятті, якщо здав відповідну лабораторну роботу або з поважної причини: хвороба, повітряна тривога тощо.
3. В разі потреби під час виконання лабораторної роботи студент може скористатися засобами зв'язку для пошуку інформації в інтернеті тощо.

4. На одному занятті захищається тільки одна лабораторна робота. В разі бажання студента здати ще одну роботу, студент має стати в кінець черги. За одне заняття можна здати не більше 2 робіт.

5. Перескладання лабораторної роботи можливе на наступному занятті.

6. Можливе перезарахування деяких лабораторних робіт елементами неформальної освіти, відповідно до Робочої програми дисципліни.

У процесі виконання лабораторних робіт студент має дотримуватись положень академічної доброчесності; використання напрацювань інших студентів (компоненти програмного забезпечення та програмні коди, графічні матеріали лабораторних робіт та фрагменти звітів з лабораторних робіт) є підставою для зниження оцінки чи перескладання лабораторної роботи.

Студент до виконання лабораторної роботи має бути підготовлений, зокрема: ознайомлений із темою, порядком виконання та відповідними теоретичними відомостями. Для встановлення рівня готовності студента до виконання роботи викладач здійснює опитування.

Пропущене лабораторне заняття студент зобов'язаний відпрацювати не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Семестрова підсумкова оцінка визначається як середньозважена з усіх видів навчальної роботи, виконаних і зданих студентом позитивно, з урахуванням коефіцієнта вагомості і розраховується в автоматизованому режимі за відповідною програмою. Вагові коефіцієнти змінюються залежно від структури дисципліни і важливості окремих видів її робіт. При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контрольного заходу, який проводиться письмово з усього матеріалу дисципліни. Студент, який набрав позитивний середньозважений бал за поточну роботу і не здав підсумковий контрольний захід, вважається нестигаючим.

Оцінювання знань студентів здійснюється за такими критеріями:

Оцінка за національною шкалою	Узагальнений критерій
Відмінно	Студент глибоко і у повному обсязі опанував зміст навчального матеріалу, легко в ньому орієнтується і вміло використовує понятійний апарат; уміє пов'язувати теорію з практикою, вирішувати практичні завдання, впевнено висловлювати і обґрунтовувати свої судження. Відмінна оцінка передбачає грамотний, логічний виклад відповіді (як в усній, так і у письмовій формі), якісне зовнішнє оформлення роботи. Студент не вагається при видозміні запитання, вміє робити детальні та узагальнюючі висновки. При відповіді допустив дві-три несуттєві похибки.
Добре	Студент виявив повне засвоєння навчального матеріалу, володіє понятійним апаратом, орієнтується у вивченому матеріалі; свідомо використовує теоретичні знання для вирішення практичних задач; виклад відповіді грамотний, але у змісті і формі відповіді можуть мати місце окремі неточності, нечіткі формулювання закономірностей тощо. Відповідь студента має будуватися на основі самостійного мислення. Студент у відповіді допустив дві-три несуттєві помилки.
Задовільно	Студент виявив знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь студента будується на рівні репродуктивного мислення, студент має слабкі знання структури курсу, допускає неточності і суттєві помилки у відповіді, вагається при відповіді на видозмінене запитання. Разом з тим набув навичок, необхідних для виконання нескладних практичних завдань, які відповідають мінімальним критеріям оцінювання і володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.

Незадовільно	Студент виявив розрізнені, безсистемні знання, не вміє виділяти головне і другорядне, допускається помилок у визначенні понять, перекручує їх зміст, хаотично і невпевнено викладає матеріал, не може використовувати знання при вирішенні практичних завдань. Як правило, оцінка "незадовільно" виставляється студенту, який не може продовжити навчання без додаткової роботи з вивчення дисципліни.
--------------	--

Структурування дисципліни за видами робіт і оцінювання результатів навчання студентів у п'ятому семестрі за ваговими коефіцієнтами

Аудиторна робота								Самостійна робота		Семестровий контроль, іспит
Лабораторна роботи №:								Тестовий контроль:		Підсумковий контрольний захід
1	2	3	4	5	6	7	8	Т 1-4	Т 5-9	1
9	10	11	12	13	14	15	16			
ВК:				0,4				0,2		0,4

Умовні позначення: Т – тема дисципліни; ВК – ваговий коефіцієнт.

Оцінювання тестових завдань

Тематичний тест для кожного студента складається з двадцяти п'яти тестових завдань, кожне з яких оцінюється одним балом. Максимальна сума балів, яку може набрати студент, складає 25.

Оцінювання здійснюється за чотирибальною шкалою.

Відповідність набраних балів за тестове завдання оцінці, що виставляється студенту:

Сума балів за тестові завдання	1–13	14–16	17–22	23–25
Оцінка за 4-бальною шкалою	2	3	4	5

На тестування відводиться 25 хвилин. Тестування проводиться в онлайн режимі у модульному середовищі для навчання MOODLE.

При отриманні негативної оцінки тест варто перездати до терміну наступного контролю.

Підсумкова семестрова оцінка за інституційною шкалою і шкалою ЄКТС встановлюється в автоматизованому режимі після внесення викладачем усіх оцінок до електронного журналу. Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС наведені у таблиці.

Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС

Оцінка ЄКТС	Інституційна інтервальна шкала балів	Інституційна оцінка, критерії оцінювання		
A	4,75–5,00	5	Зарховано	Відмінно – глибоке і повне опанування навчального матеріалу і виявлення відповідних умінь та навичок
B	4,25–4,74	4		Добре – повне знання навчального матеріалу з кількома незначними помилками
C	3,75–4,24	4		Добре – в загальному правильна відповідь з двома-трьома суттєвими помилками
D	3,25–3,74	3		Задовільно – неповне опанування програмного матеріалу, але достатнє для практичної діяльності за професією
E	3,00–3,24	3		Задовільно – неповне опанування програмного матеріалу, що задовольняє мінімальні критерії оцінювання
FX	2,00–2,99	2	Незарховано	Незадовільно – безсистемність одержаних знань і неможливість продовжити навчання без додаткових знань з дисципліни

F	0,00–1,99	2	<i>Незадовільно</i> – необхідна серйозна подальша робота і повторне вивчення дисципліни
---	-----------	---	---

Перелік питань для самоконтролю

1. Вкажіть переваги та недоліки використання логістичної регресії, як лінійного класифікатора.
2. Вкажіть переваги та недоліки використання методу випадкового лісу.
3. Вкажіть переваги та недоліки використання методу дерева рішень.
4. Вкажіть переваги та недоліки використання методу найближчих сусідів.
5. Вкажіть та коротко опишіть основні параметри дерева рішень.
6. Коротко опишіть алгоритм побудови дерева рішень.
7. Коротко опишіть використання дерева рішень для задачі регресії.
8. Коротко опишіть задачі інтелектуального аналізу даних: регресія, кластеризація, виявлення аномалій.
9. Коротко опишіть задачу бінарної класифікації.
10. Опишіть алгоритм вибору параметрів моделі за допомогою крос-валідації.
11. Опишіть алгоритм роботи дерева рішень із кількісними ознаками.
12. Опишіть критерії якості розбиття для завдання класифікації.
13. Опишіть логістичну регресію, як лінійний класифікатор.
14. Опишіть модель логістичної регресії Lasso. В чому її різниця проти інших типів логістичної регресії?
15. Опишіть метод ансамблей для завдання класифікації.
16. Опишіть метод беггінг для завдання класифікації.
17. Опишіть метод k -середніх.
18. Опишіть метод алгомеративної кластеризації.
19. Опишіть метод афінного розширення (Affinity Propagation).
20. Опишіть метод вибору числа кластерів для k -середніх.
21. Опишіть метод випадкового лісу для задачі класифікації.
22. Опишіть метод головних компонент.
23. Опишіть метод дерева рішень.
24. Опишіть метод максимальної правдоподібності.
25. Опишіть метод найближчих сусідів.
26. Опишіть метод найменших квадратів.
27. Опишіть підхід до оброблення даних під назвою Feature Selection.
28. Опишіть підхід до перетворення даних під назвою Feature Extraction.
29. Опишіть підхід до трансформації даних під назвою Feature Transformations.
30. Коротко опишіть підхід перетворення ознак в багатовимірний простір.
31. Опишіть поняття варіації та декореляційного ефекту.
32. Розпишіть формулу ентропії Шеннона.
33. Охарактеризуйте складність моделі інтелектуального аналізу даних.
34. Охарактеризуйте задачу кластеризації.
35. Дайте визначення та коротко охарактеризуйте поняття «категоріальні ознаки».
36. Опишіть «проблему мультиколінеарності». В яких задачах і за використання яких алгоритмів вона виникає? Які є підходи до її вирішення?
37. Охарактеризуйте L2-регуляризація логістичної функції втрат.
38. Охарактеризуйте ROC-криву та наведіть її визначення. Як ефективно обчислити ROC-криву?
39. Охарактеризуйте використання кривої навчання та валідування.
40. Охарактеризуйте лінійний класифікатор.
41. Охарактеризуйте регуляризацію лінійної регресії.
42. Охарактеризуйте спектральну кластеризацію.
43. Наведіть два приклади використання методу «Ensemble» для розв'язання практичних задач.

44. Наведіть два приклади використання методу «Bagging» для розв'язання практичних задач.
45. Наведіть два приклади використання методу «Random forest» для розв'язання практичних задач.
46. Наведіть два приклади використання методу «Principle component analysis» для розв'язання практичних задач.
47. Наведіть два приклади використання методу «Decision trees» для розв'язання практичних задач.
48. Наведіть два приклади використання методу « k -nearest neighbors» для розв'язання практичних задач.
49. Наведіть два приклади використання методу « k -means» для розв'язання практичних задач.
50. Наведіть формулу та опишіть критерій якості розбиття в задачі класифікації Gini impurity.

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ

Навчальний процес з дисципліни «Інтелектуальний аналіз даних» повністю і в достатній кількості забезпечений необхідною навчально-методичною літературою.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Мокін В.Б. Наука про дані: машинне навчання та інтелектуальний аналіз даних [Електронний навчальний посібник] / В.Б. Мокін, М.В. Дратований. Вінниця : ВНТУ, 2024. 263 с. URL: <https://t.ly/VYK9h>
2. Болюбаш Н. М. Інтелектуальний аналіз даних : навч. посіб. / Н. М. Болюбаш. Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. 320 с. URL: <https://t.ly/Dht4x>
3. Талах М.В. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак. Чернівці: Технодрук, 2022. 367 с. URL: <https://t.ly/B2GLZ>

Допоміжна

4. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посіб. / за рец. В.О. Філатова, О.А. Винокурової. Харків: ХНУРЕ, 2021. 92 с. URL: <https://openarchive.nure.ua/server/api/core/bitstreams/2e55d639-52fd-48d9-b7b7-14989f49f291/content>
5. Melnychenko O., Scislo L., Savenko O., Sachenko A., Radiuk P. Intelligent integrated system for fruit detection using multi-UAV imaging and deep learning. *Sensors*. 2024. Vol. 24. No. 6. P. 1913. (*Scopus, Q1; Web of Science, Q2*). URL: <https://doi.org/10.3390/s24061913>
6. Radiuk P., Barmak O., Manziuk E., Krak Iu. Explainable deep learning: A visual analytics approach with transition matrices. *Mathematics*. 2024. Vol. 12. No. 7. P. 1024. (*Scopus, Q2; Web of Science, Q1*). URL: <https://doi.org/10.3390/math12071024>
7. Kalyta O., Barmak O., Radiuk P., Krak I. Facial emotion recognition for photo and video surveillance based on machine learning and visual analytics. *Applied Sciences*. 2023. Vol. 13. No. 17. P. 9890. (*Scopus, Q2; Web of Science, Q2*). URL: <https://doi.org/10.3390/app13179890>
8. Analysis of deep learning methods in adaptation to the small data problem solving / I. Krak, V. Kuznetsov, S. Kondratiuk, L. Azarova, O. Barmak, P. Radiuk. In: Babichev, S., Lytvynenko, V. (eds) *Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making*. ISDMCI–2022. Springer, Cham. 2023. Vol. 149. P. 333–352. (*Scopus, Q3*). URL: https://doi.org/10.1007/978-3-031-16203-9_20
9. Raschka S., Liu Y., Mirjalili V. Machine learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. Birmingham, UK: Packt Publishing Ltd., 2022. 770 p. URL: <https://github.com/rasbt/machine-learning-book>
10. Banachewicz K., Massaron L. The Kaggle book: Data analysis and machine learning for competitive data science. Birmingham, UK: Packt Publishing Ltd., 2022. 531 p. URL: <https://github.com/PacktPublishing/The-Kaggle-Book>
11. Radiuk P.M., Skrypnik T.K., Karlechuk D.T. Applying mental models to making controlled critically safe decisions in IT project management. *Herald of Khmelnytskyi National University*.

Technical sciences. 2021. Vol. 301, No. 5. P. 32–35. URL: <https://www.doi.org/10.31891/2307-5732-2021-301-5-32-35>

12. Radiuk P.M., Mazurets O.V., Skrypnyk T.K., Moroz O.V. Intelligent data analysis using artificial neural networks for decision making in the education domain. *Herald of Khmelnytskyi National University. Technical sciences*. 2021. Vol. 303, No. 6. P. 111–114. URL: <https://www.doi.org/10.31891/2307-5732-2021-303-6-111-114>

13. Zhang A., Lipton Z.C., Li M., Smola A.J. Dive into deep learning: 3rd ed. arXiv preprint arXiv:2106.11342, 2021. 1197 p. URL: <https://arxiv.org/abs/2106.11342>

14. Сторова О.В. Методичні рекомендації до самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для здобувачів освітнього ступеня «бакалавр» зі спеціальностей 122 Комп'ютерні науки та 126 – Інформаційні системи та технології усіх форм навчання [Електронний ресурс] / за рец. А.О. Лавданського. М-во освіти і науки України, Черкас. держ. технол. ун-т. Черкаси: ЧДТУ, 2020. 13 с. URL: <https://er.chdtu.edu.ua/bitstream/ChSTU/1605/1/19-5-25-%D0%9F.pdf>

15. Кононова К.Ю. Машинне навчання: Методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка» / за рец. А.В. Матвійчука, Л.С. Гур'янової. Харків: ХНУ імені В. Н. Каразіна, 2020. 301 с. URL:

https://www.researchgate.net/publication/345765254_MASINNE_NAVCANNA_METODI_TAJ_MOD_ELI

16. Deisenroth M.P., Faisal A.A., Ong C. S. Mathematics for machine learning. Cambridge: Cambridge University Press, 2020. 412 p. URL: <https://mml-book.github.io/book/mml-book.pdf>

Інформаційні ресурси

17. Модульне середовище. Режим доступу : <https://msn.khnu.km.ua/>

18. Електронна бібліотека університету. Режим доступу : http://lib.khmnu.edu.ua/asp/php_f/plage_lib.php.

19. Репозитарій ХНУ. Режим доступу : <https://library.khmnu.edu.ua/#>