

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ

Декан факультету комп'ютерних
інформаційних технологій

_____ Ігор ЯКИМЕНКО

«___» _____ 2024 р.

ЗАТВЕРДЖУЮ

Проректор з науково-педагогічної роботи

_____ Віктор ОСТРОВЕРХОВ

«___» _____ 2024 р.

РОБОЧА ПРОГРАМА
з дисципліни
„ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ”

ступінь вищої освіти –

бакалавр

галузі знань -

12 «Інформаційні технології»

спеціальність -

124 «Системний аналіз»

освітньо-професійна програма -

«Системний аналіз»

Кафедра економічної кібернетики та інформатики

Форма навчання/ факультет	Курс	Семестр	Лекції	Практичні	ІРС	СРС	Тренінг КПЗ	Разом	Екзамен,
Денна	4	7	46	44	5	73	8	180	Екзамен, 7

Тернопіль 2024

Робоча програма складена на основі освітньо-професійної програми підготовки бакалавра галузі знань 12 Інформаційні технології спеціальності 124 Системний аналіз, затвердженої на засіданні вченою ради ЗУНУ (протокол № від 27 серпня 2024 р.)

Робочу програму склав старший викладач кафедри економічної кібернетики та інформатики АДАМІВ Олег Петрович

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики протокол № __ від _____ 2024 р.

Завідувач кафедри _____ проф. БУЯК Леся Михайлівна

Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол № __ від _____ 2024р.

Голова ГЗС _____ проф. ПАСІЧНИК Р.М.

Гарант ОПП _____ проф. ПАСІЧНИК Р.М.

**СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«Інтелектуальний аналіз даних»**

1 Опис дисципліни «Інтелектуальний аналіз даних»

Дисципліна – Інтелектуальний аналіз даних	Галузь знань, спеціальність, СВО	Характеристика навчальної дисципліни
Кількість кредитів ECTS – 6	Галузь знань 12 «Інформаційні технології»	Нормативна дисципліна Мова навчання - українська
Кількість залікових модулів – 4	Спеціальність 124 «Системний аналіз»	Рік підготовки: <i>Денна – 4</i> Семестр: <i>Денна – 7</i>
Кількість змістових модулів – 3	Ступінь вищої освіти – бакалавр	Лекції: <i>Денна – 46 год.</i> Практичні: <i>Денна – 44 год.</i>
Загальна кількість годин– 180 год. Загальна кількість годин– 90 год.		Самостійна робота: <i>Денна – 73 год. (тренінг (КПІЗ) – 8 год.)</i> Індивідуальна робота – <i>5 год.</i>
Тижневих годин: Денна форма навчання – 12 год., з них аудиторних – 6 год.		Вид підсумкового контролю – екзамен

2. Мета і завдання дисципліни “ Інтелектуальний аналіз даних ”

2.1. Мета вивчення дисципліни – вивчення формальних методів застосування прикладних інформаційних технологій для розв’язання прикладних задач, зокрема щодо аналітики даних та процесів та їх автоматизації.

2.2. Завдання вивчення дисципліни

Завданнями курсу «Інтелектуальний аналіз даних» є підготовка фахівців, здатних застосувати математичні, статистичні основи, алгоритмічні принципи та теорію системного аналізу у застосуванні прикладних інформаційних технологій для розв’язання прикладних задач, зокрема щодо аналітики даних та процесів та їх автоматизації.

В результаті вивчення курсу "Інтелектуальний аналіз даних" студенти повинні:

- знати основні поняття лінійного та квадратичного аналізу вибірок, машин опорних векторів, стохастичного градієнтного спуску, методу найближчих сусідів, сумішей нормальних розподілів, наївних Байєсівських класифікаторів, дерев рішень, ансамблевих методів, напівконтрольованого навчання;
- вміти здійснювати лінійний та квадратичний аналіз вибірок, реалізовувати кластеризацію вибірок за допомогою машин опорних векторів, стохастичного градієнтного спуску, найближчих сусідів, наївних Байєсівських класифікаторів, , дерев рішень, будувати прогнози на основі сумішей нормальних розподілів, ансамблевих методів, напівконтрольованого навчання та ізотонічної регресії.

2.3. Найменування та опис компетентностей, формування котрих забезпечує вивчення дисциплін:

1. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.

2. Здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій з використанням математичних моделей і методів.

3. Уміння здійснювати комплексний статистичний аналіз і прогнозування фізичних та соціально-економічних процесів

2.4. Передумови для вивчення дисципліни.

Для успішного вивчення цієї дисципліни студенти повинні опонувати такі дисципліни: математичний аналіз, теорія імовірностей та математична статистика .

2.5. Результати навчання:

1. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп’ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

2. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.

2.6. Завдання лекційних занять

Мета проведення лекцій полягає у тому, щоб ознайомити студентів із головними питаннями курсу "Інтелектуальний аналіз даних".

Завдання проведення лекцій полягає у:

- викладенні студентам у відповідності з програмою та робочим планом основних питань курсу "Інтелектуальний аналіз даних";
- сформуванні у студентів цілісної системи теоретичних знань з курсу "Інтелектуальний аналіз даних".

2.7. Завдання проведення практичних занять

Мета проведення практичних занять полягає у тому, щоб виробити у студентів практичні навички використання теоретичного матеріалу.

Завдання проведення практичних занять полягає у глибшому засвоєнні та закріпленні теоретичних знань, одержаних на лекціях.

3. Програма навчальної дисципліни “ Інтелектуальний аналіз даних ”

Змістовий модуль 1. Задачі аналізу даних.

Тема 1. Технології аналізу даних.

Принципи аналізу даних. Структуровані дані. Технології Knowledge Discovery in Database і Data Mining. Аналітичні платформи. Задачі аналізу даних. Сховища даних. Очистка, трансформація та попередня обробка даних.

Тема 2. Консолідація даних.

Сховища даних: багатомірні, реляційні, гібридні, віртуальні. Вступ в ETL. Завантаження даних в сховище. Завантаження даних з локальних джерел. Збагачення даних.

Тема 3. OLAP системи.

Багатомірна модель даних. Тест FASMI. Архітектура OLAP систем.

Змістовий модуль 2. Базові методи аналізу даних.

Тема 4. Класифікація і регресія.

Постановка задачі та представлення результатів. Методи побудови правил класифікації: «1-правило» («1-rule»), Naive Bayes. Методи побудови дерев рішень («розділяй і владарюй», ID3, C4.5). Алгоритм покриття. Методи побудови математичних функцій: лінійні, нелінійні, Support Vector Machine. Прогнозування часових рядів.

Тема 5. Пошук асоціативних правил.

Постановка задачі пошуку асоціативних правил. Сиквенціальний аналіз. Представлення результатів при пошуку асоціативних правил. Алгоритм APRIORI.

Тема 6. Кластеризація.

Постановка задачі та представлення результатів задачі кластеризації. Алгоритми кластеризації: ієрархічні (алгомеративні, дивізімні). Неієрархічні алгоритми кластеризації: Fuzzy K-Means, кластеризація по Гюстафсону-Кеселю. Адаптивні методи кластеризації. Показники якості кластеризації.

Змістовий модуль 3. Прикладні аспекти аналізу даних

Тема 7. Візуальний аналіз даних.

Виконання візуального аналізу даних. Характеристики засобів візуалізації даних.

Методи візуалізації: методи геометричних перетворень, відображення іконок, методи, які орієнтовані на пікселі, ієрархічні висновки.

Тема 8. Аналіз текстової інформації (Text Mining).

Постановка задачі аналізу текстів. Вилучення ключових понять з тексту. Класифікація текстових документів. Методи кластеризації текстових документів. Задачі анотування текстів. Web-mining. Інструментальні засоби аналізу текстової інформації (Oracle Text, IBM Intelligent Miner for Text, SAS Text Miner та інші).

Тема 9. Ансамблі моделей.

Беггінг. Бустінг. Оцінка моделей (оцінка ефективності та похибки). Криві Lift та Profit. ROC-аналіз. Навчання в умовах незбалансованості класів.

Тема 10. Інтелектуальний аналіз даних в реальному часі.

Адаптивне здобування знань. Рекомендаційні системи. Інструменти інтелектуального аналізу даних в реальному режимі.

Тема 11. Здобування даних через Web.

Методи вилучення Web-контенту та Web-структур. Дослідження використання Web-ресурсів.

Тема 12. Методи аналізу бізнес процесів.

Автоматизація виконання процесів управління. Аналіз процесів. Методи Process Mining. Алгоритми аналізу процесів.

Тема 13. Стандарти аналізу даних.

Стандарт CWM. Стандарт CRISP. Стандарти PMML. Інші стандарти ІАД: SQL/MM, OLE DB для Data Mining та JDM API.

4. Структура залікового кредиту дисципліни «Інтелектуальний аналіз даних»

Назва теми	Кількість годин					
	Лекції	Практичні заняття	Самостійна робота	Індивідуальна робота	Тренінг та КПІЗ	Контроль заходів
Змістовий модуль 1. Задачі аналізу даних.						
Тема 1. Технології аналізу даних.	4		6	1	2	поточне опит.
Тема 2. Консолідація даних.	4	6	6			поточне опит.
Тема 3. OLAP системи.	4		6			поточне опит.
Змістовий модуль 2. Базові методи аналізу даних.						
Тема 4. Класифікація і регресія.	4	6	8		4	поточне опит.
Тема 5. Пошук асоціативних правил.	4	8	8			поточне опит.
Тема 6. Кластеризація.	4	7	8			поточне опит.
Змістовий модуль 3. Прикладні аспекти аналізу даних.						
Тема 7. Візуальний аналіз даних.	2	6	4		2	поточне опит.
Тема 8. Аналіз текстової інформації (Text Mining).	4	6	4	2		поточне опит.
Тема 9. Ансамблі моделей.	2		4			поточне опит.
Тема 10. Інтелектуальний аналіз даних в реальному часі.	4		4			поточне опит.
Тема 11. Здобування даних через Web.	2		5			поточне опит.

Тема 12.Методи аналізу бізнес процесів.	4	5	5	2		поточне опит.
Тема 13. Стандарти інтелектуального аналізу даних.	4		5			ректорс. контр.
Всього	46	44	73	5	8	

5. Тематика практичних занять.

Практична робота № 1.

Тема: Попередня обробка даних: очистка та трансформація.

Мета: Вивчити основні можливості оболонки Data Transformer і оволодіти основними навичками програмування.

Практична робота №2

Тема: Побудова правил класифікації в системі WizWhy.

Мета: Вивчити основні можливості графічного інтерфейсу WizWhy та побудувати систему класифікації.

Практична робота №3

Тема: Вирішення класифікаційних задач ймовірнісними оцінками засобами HuginExpert.

Мета: Навчитися створювати Баєсівські мережі довіри для задач класифікації.

Практична робота № 4.

Тема: Побудова дерев рішень

Мета: Провести розпізнавання і регресійний аналізу даних з допомогою дерев рішень.

Практична робота № 5.

Тема: Проведення кластеризації засобами штучних нейронних мереж в системі Дедуктор.

Мета: Ознайомитись з можливостями пакету Дедуктор для аналізу даних.

Практична робота № 6.

Тема: Дослідження текстової інформації.

Мета: Ознайомитись з концепцією добування знань через парсинг текстової інформації.

Практична робота № 7.

Тема: Аналіз бізнес процесів на основі системи Business Studio.

Мета: Побудова моделей бізнес процесів підприємств.

6. Комплексне практичне індивідуальне завдання з дисципліни «Інтелектуальний аналіз даних»

Індивідуальні завдання з дисципліни “Інтелектуальний аналіз даних” студенти виконують самостійно на основі сформованого випадковим чином вибіркового завдання. КППЗ охоплює усі основні теми дисципліни “Інтелектуальний аналіз даних”. Метою виконання КППЗ є оволодіння навиками застосування систем бізнес аналітики для виконання задач аналізу даних та управління. КППЗ оформлюється у відповідності із встановленими вимогами; при його виконанні та оформленні КППЗ студент використовує комп’ютерну техніку.

Виконання КПЗ є одним із обов'язкових складових модулів залікового кредиту з дисципліни “Інтелектуальний аналіз даних” і складається з наступних етапів:

1. Вивчити питання джерел інформації та консолідації даних з обраної предметної області.

2. Сформувані сховища даних з допомогою ETL інструментів обраної предметної області:

2.1. Провести трансформацію даних:

- трансформацію впорядкованих даних;
- групування даних;
- поєднання даних;
- квантування;
- нормалізація;
- формалізація даних.

2.2. Провести очистку та попередню обробку даних:

- оцінити якість даних;
- провести очистку і попередню обробку даних;
- провести фільтрацію даних;
- здійснити обробку дублікатів і протиріч;
- виявити аномальні значення;
- відновити пропущені дані;
- зменшити розмірність;
- зменшити кількість ознак;
- зменшити кількість значень характеристик та записів;
- провести семплінг.

На основі проведених етапів провести побудову сховища даних та проаналізувати можливі методи бізнес аналізу, які можна застосувати для подальшої роботи.

7. Самостійна робота студентів

№ п/п	ЗМІСТ
1	Візуалізатори загального призначення.
2	Візуалізатори для оцінки якості моделей.
3	Візуалізатори, що використовуються для інтерпретації результатів аналізу.
4	Ієрархічні асоціативні правила.
5	Послідовні шаблони.
6	Проблеми алгоритмів кластеризації.
7	Проста лінійна регресія.
8	Множинна лінійна регресія.
9	Регресія з категоріальними вхідними характеристиками об'єктів.
10	Методи відбору вхідних змінних в регресійні моделі.
11	Обмеження застосування регресійних моделей.

12	Інтерпретація моделі логічної регресії.
13	Простий байєсівський інтерпретатор.
14	Спрощення дерев рішень.
15	Алгоритм CART.
16	Моделі прогнозування.
17	Застосування моделей прогнозування в бізнес аналітиці
18	Оцінка ефективності та аналіз моделей.
19	Оцінка похибки моделі.
20	Недоліки класифікації за похибкою.
21	Криві Lift та Profit.
22	ROC-аналіз.
23	Навчання в умовах незбалансованих класів.

8. Тренінг з дисципліни

Тематика Дослідження основних можливостей OLAP систем (модулів).

Приклади OLAP систем: SAP NetWeaver BW, SAP NetWeaver BW, Oracle Database OLAP Option, Mondrian OLAP server, MicroStrategy Intelligence Server, Microsoft Analysis Services, Jedox OLAP Server, Infor BI OLAP Server, icCube, IBM Cognos TM1, DB2 OLAP Server, Oracle/Hyperion EssBase, SAP BW, BusinessObjects, SAS OLAP Server тощо.

1. Загальна характеристика OLAP системи (технічні характеристики (обмеження), підтримка архітектур, APIs та мови запитів, безпека, підтримка операційних систем тощо).

2. Сфери застосування (приклад конкретних задач).

3. Приклад роботи OLAP системи на основі даних з п.2.

4. Демонстрація основних характеристик встановленої OLAP системи.

9. Засоби оцінювання та методи демонстрування результатів навчання

У процесі навчання дисципліни «Інтелектуальний аналіз даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- стандартизовані тести;
- поточне опитування;
- залікове модульне тестування та опитування;
- презентації результатів виконаних завдань та досліджень;
- оцінювання результатів КПЗ;
- студентські презентації та виступи на наукових заходах;
- завдання на лабораторному обладнанні, тренажерах, реальних об'єктах тощо;
- ректорська контрольна робота;
- екзамен.

10. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Інтелектуальний аналіз» визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

Заліковий модуль 1	Заліковий модуль 2 (ректорська контрольна робота)	Заліковий модуль 3 (підсумкова оцінка за КПЗ, враховуючи поточне опитування)	Заліковий модуль 4(Екзамен)	Разом
20%	20%	20%	40%	100%
1. Усне опитування під час заняття (50 балів) 2. Письмова робота = 50 балів	1. Усне опитування під час заняття (50 балів) 2.Письмова робота = 50 балів	1.Написання та захист КПЗ = 60 балів. 3. Виконання завдань під час тренінгу = 40 балів	1. 3 запитання по 20 балів =60 балів 2. Задача = 40 балів	

11. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№ з/п	Найменування	Номер теми
1.	Персональний комп'ютер	1-13
2.	Прикладні пакети WizWhy, HuginLite	1-13

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Scikit-learn. User Guide. https://scikit-learn.org/stable/user_guide.html
2. О. І. Шеремет, О. В. Садовой. Метод опорних векторів (SVM). <https://www.dstu.dp.ua/Portal/Data/74/72/3st13-17.pdf>
3. Scikit-learn. Stochastic Gradient Descent. <https://scikitlearn.org/stable/modules/sgd.html>
4. Scikit-learn. SGD: Maximum margin separating hyperplane. https://scikitlearn.org/stable/auto_examples/linear_model/plot_sgd_separating_hyperplane.html#sphx-glr-auto-examples-linear-model-plot-sgd-separating-hyperplane-py
5. SGD: Weighted samples. https://scikitlearn.org/stable/auto_examples/linear_model/plot_sgd_weighted_samples.html#sphx-glr-auto-examples-linear-model-plot-sgd-weighted-samples-py
6. Nearest Neighbors regression. https://scikitlearn.org/stable/auto_examples/neighbors/plot_regression.html#sphx-glr-autoexamples-neighbors-plot-regression-py

7. Nearest Neighbors. <https://scikit-learn.org/stable/modules/neighbors.html>
8. Kaixin Wang. Introduction to Gaussian process regression. <https://medium.com/data-science-at-microsoft/introduction-to-gaussian-processregression-part-1-the-basics-3cb79d9f155f>
9. Naive Bayes Classifiers. <https://www.geeksforgeeks.org/naive-bayes-classifiers/>.
10. Decision Tree. <https://www.geeksforgeeks.org/decision-tree/>