

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

ЗАТВЕРДЖЕНО

на засіданні кафедри
інформатики та комп'ютерної
техніки
Протокол № 1 від 29.08.2023 р.

ПОГОДЖЕНО

Проректор з навчально-методичної
роботи



Каріна НЕМАШКАЛО

**ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ
робоча програма навчальної дисципліни (ІАД)**

Галузь знань **12 "Інформаційні технології"**
Спеціальність **126 "Інформаційні системи та технології "**
Освітній рівень **перший (бакалаврський)**
Освітня програма **"Інформаційні системи та технології"**

Статус дисципліни **вибіркова**
Мова викладання, навчання та оцінювання **українська**

Розробник:
д.т.н., проф.

Сергій УДОВЕНКО

Завідувач кафедри
інформатики та комп'ютерної
техніки

Сергій УДОВЕНКО

Гарант програми

Ольга ТЮТЮНИК

**Харків
2023**

ВСТУП

Навчальна дисципліна "Інтелектуальний аналіз даних" є вибірковою навчальною дисципліною та вивчається згідно з навчальним планом підготовки здобувачів вищої освіти за спеціальністю 126 "Інформаційні системи та технології" першого (бакалаврського) рівня усіх форм навчання. Програму навчальної дисципліни розроблено у відповідності до вимог галузевого стандарту вищої освіти на базі освітньо-професійної програми підготовки бакалавра.

Метою викладання навчальної дисципліни є формування у майбутніх фахівців теоретичних і практичних знань з основ застосування методів інтелектуального аналізу даних для розв'язання спеціалізованих задач попередньої обробки, перетворення, класифікації даних та дослідження процесів у сфері використання інформаційних технологій.

Завданням навчальної дисципліни "Інтелектуальний аналіз даних" знайомить здобувачів з основними принципами застосування сучасних підходів до інтелектуальної обробки даних. Увагу приділено способам представлення даних, їх попередньої обробки та перетворення (Data Mining). Розглянуто методи опису сигналів та процесів, а також методи моделювання та аналізу властивостей об'єктів інформаційних систем з використанням програмних пакетів. Особливу увагу приділено методам аналізу даних з використанням засобів обчислювального інтелекту, а саме: інтелектуального аналізу неструктурованої текстової інформації (Text Mining) та інформаційного пошуку релевантних документів за запитом користувача (Web Mining), а також використання генетичних алгоритмів та еволюційної оптимізації в інформаційних системах.

Розглянуто приклади вирішення практичних задач з використанням методів інтелектуального аналізу даних.

Предметом дисципліни є програмні продукти, сучасні аналітичні методи обробки аналізу даних та обробки інформації.

Об'єктом дисципліни є процеси інтелектуального аналізу даних в інформаційних системах.

Програма навчальної дисципліни передбачає навчання у формі лекцій, лабораторних занять та самостійної роботи здобувачів. Лекції, лабораторні заняття, індивідуальна робота та консультації проводяться з застосуванням персональних комп'ютерів, локальної мережі та мережі Інтернет у комп'ютерних класах (або з використанням дистанційних технологій навчання в режимі он-лайн). Всі види занять забезпечуються необхідними електронними методичними матеріалами.

З метою підвищення ефективності вивчення навчальної дисципліни здобувачі мають змогу користуватись системою дистанційного навчання ХНЕУ ім. С. Кузнеця.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
ПР 2	КЗ 2
	КЗ 3
	КС 1
	КС 11
ПР 3	КЗ 1
	КС 6
ПР 5	КЗ 5
	КЗ 6
ПР 6	КЗ 3
	КС 12
	КС 13
ПР 7	КЗ 1
	КЗ 2
	КЗ 5
	КС 7
	КС 12

де:

КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу.

КЗ 2. Здатність застосовувати знання у практичних ситуаціях.

КЗ 3. Здатність до розуміння предметної області та професійної діяльності.

КЗ 5. Здатність вчитися і оволодівати сучасними знаннями.

КЗ 6. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.

КС 1. Здатність аналізувати об'єкт проектування або функціонування та його предметну область.

КС 6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.

КС 7. Здатність застосовувати інформаційні технології у ході створення, впровадження та експлуатації системи менеджменту якості та оцінювати витрати на її розроблення та забезпечення.

КС 11. Здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій з використанням математичних моделей і методів.

КС 12. Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).

КС13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.

ПР 2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проектування і використання інформаційних систем та технологій.

ПР 3. Використовувати базові знання інформатики й сучасних інформаційних систем та технологій, навички програмування, технології безпечної роботи в комп'ютерних мережах, методи створення баз даних та інтернет-ресурсів, технології розроблення алгоритмів і комп'ютерних програм мовами високого рівня із застосуванням об'єктно-орієнтованого програмування для розв'язання задач проектування і використання інформаційних систем та технологій.

ПР 5. Аргументувати вибір програмних та технічних засобів для створення інформаційних систем та технологій на основі аналізу їх властивостей, призначення і технічних характеристик з урахуванням вимог до системи і експлуатаційних умов; мати навички налагодження та тестування програмних і технічних засобів інформаційних систем та технологій.

ПР 6. 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.

ПР 7. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Попередній аналіз та методи перетворення сигналів в інформаційних системах (ІС)

Тема 1. Основні поняття та визначення курсу. Попередній аналіз сигналів

Визначення терміну «Інтелектуальний аналіз даних» (Data Mining). Основні моделі і методи інтелектуального аналізу даних. Способи представлення даних, їх передобробки та перетворення. Галузі використання інтелектуального аналізу даних.

Тема 2. Методи перетворення сигналів ІС. Опис об'єктів ІС

Загальні принципи перетворення сигналів в ІС. Z-перетворення сигналів. Методи опису сигналів та об'єктів ІС. Різницеві рівняння, дискретні передавальні функції. Аналіз характеристик об'єктів ІС. Моделювання та аналіз властивостей об'єктів ІС з використанням програмних пакетів.

Тема 3. Фрактали та хаос в інтелектуальному аналізі даних

Загальна характеристика фрактальних властивостей процесів. Регулярні фрактали. Розмірність фракталів. Приклади побудови фракталів. Системи функцій, що ітеруються. Метод випадкових ітерацій. Стискуючі афінні перетворення. L-системи побудови фракталів. Аналітичні методи побудови фракталів. Зв'язок фракталів та хаосу. Аналіз фрактальних та хаотичних властивостей процесів в ІС.

Змістовий модуль 2. Методи інтелектуального аналізу даних в ІС

Тема 4. Інтелектуальний аналіз електронних текстів

Загальна характеристика завдань інтелектуального аналізу неструктурованої текстової інформації (Text Mining). Інформаційний пошук релевантних документів за запитом користувача (Web Mining). Показники оцінювання якості пошуку документів. Текстові та посилальні методи ранжування. Попередня обробка тексту. Методи класифікації текстових документів. Семантичний аналіз текстів. Методи машинного перекладу та анотування текстів.

Тема 5. Методи обчислювального інтелекту в ІАД. Генетична оптимізація

Загальна характеристика використання методів обчислювального інтелекту в ІАД. Принципи генетичної оптимізації. Генетичні алгоритми (ГА) пошуку оптимальних рішень. Формати представлення даних в ГА. Приклади практичного використання ГА в завданнях ІАД.

Тема 6. Методи еволюційної оптимізації в ІАД

Загальна концепція пошуку еволюційних стратегій (ЕС) в ІАД. Основні параметри ЕС. Самоадаптація ЕС. Моделі колективного інтелекту в схемах пошуку ЕС. Метод мурашиних колоній. Імунні методи. PSO-метод. Оцінка результатів еволюційного пошуку. Приклади використання еволюційної оптимізації в завданнях ІАД.

Перелік практичних (семінарських) та / або лабораторних занять / завдань за навчальною дисципліною наведено в табл. 2

Таблиця 2

Перелік практичних (семінарських) та / або лабораторних занять / завдань

Назва теми та завдання	Зміст
Тема 1. Завдання 1	Попередній аналіз сигналів (визначення статистичних властивостей)
Тема 2. Завдання 2	Методи z-перетворення сигналів інформаційної системи
Тема 2. Завдання 3	Аналіз перехідних характеристик об'єктів ІС
Тема 2. Завдання 4	Дослідження процесів квантування за часом та рівнем сигналів ІС
Тема 3. Завдання 5	Фрактальний аналіз процесів в ІС
Тема 4. Завдання 6	Попередня обробка текстових документів
Тема 4. Завдання 7	Класифікація текстів
Тема 5. Завдання 8	Апроксимація функцій з використанням генетичного алгоритму
Тема 6. Завдання 9	Аналіз результатів еволюційної оптимізації

Перелік самостійної роботи за навчальною дисципліною наведено в табл. 3

Таблиця 3

Перелік самостійної роботи

Назва теми	Зміст
Тема 1 - 6	Вивчення лекційного матеріалу
Тема 1 - 6	Підготовка до лабораторних занять
Тема 1 - 6	Підготовка до екзамену

Кількість годин лекційних, практичних (семінарських) та / або лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

МЕТОДИ НАВЧАННЯ

Методи навчання – взаємодія між викладачем і здобувачем, під час якої відбувається передача та засвоєння знань, умінь і навичок від викладача до здобувача, а також самостійна та індивідуальна робота здобувача.

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

Словесні (лекція (Тема 1-6)).

Наочні (демонстрація (Тема 1-6)).

Практичні (лабораторна робота (Тема 1 – 6)).

У разі здобуття освіти за дистанційною формою або з використанням дистанційних технологій навчання лекційні заняття проводяться в режимі он-лайн (відео-конференції ZOOM та GoogleMeet).

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

ХНЕУ ім. С. Кузнеця використовує накопичувальну (100-бальну) систему оцінювання.

Оцінювання здійснюється за такими видами контролю:

– поточний контроль, що здійснюється протягом семестру під час проведення лабораторних занять та контрольних робіт і оцінюється сумою набраних балів (максимальна сума – 60 балів; мінімальна сума, що дозволяє здобувачу отримати допуск до екзамену – 35 балів);

– підсумковий/семестровий контроль здійснюється у формі екзамену.

Поточний контроль включає: оцінювання здобувачів під час лабораторних занять – за умови виконання здобувачем завдань з 18 лабораторних робіт, що згруповані у 9 занять, об'єднаних спільними темами; оцінювання результатів виконання студентами 2 контрольних робіт. Захист кожної подвійної лабораторної роботи оцінюється у 5 балів. Оцінка за лабораторну роботу отримується здобувачем за наявності звіту з лабораторної роботи, виконаних

завдань лабораторної роботи, розгорнутої відповіді на запитання та виконання контрольних прикладів. Максимальна кількість балів за виконання та захист лабораторних робіт складає 45 балів.

Контрольні роботи 1 та 2 (за відповідними змістовими модулями) виконуються на комп'ютері з застосуванням системи дистанційного навчання. Кожна контрольна робота містить два або три практичні завдання (задачі) та оцінюється у: 8 балів (контрольна робота №1) та 7 балів (контрольна робота №2). Оцінка за контрольну роботу знижується при відсутності виконаного завдання, припущенні помилок у розрахунках, неповному виконанні завдань. Максимальна кількість за виконання та захист контрольних робіт складає 15 балів.

Самостійна робота здобувача включає ознайомлення із літературними джерелами за відповідними темами, підготовку до: виконання і захисту лабораторних та контрольних робіт, а також до складання іспиту.

Підсумковий контроль:

Екзамен виконується на комп'ютері з застосуванням системи дистанційного навчання. Екзамен передбачає необхідність виконання трьох завдань: одного стереотипного (оцінюється у 15 балів), одного діагностичного (оцінюється у 15 балів) та одного евристичного (оцінюється у 10 балів). Оцінка за екзамен знижується при відсутності виконаного завдання, припущенні помилок у розрахунках, неповному виконанні завдань. Максимальна оцінка за екзамен складає 40 балів. Мінімальна сума, за якою екзамен (іспит) вважається складеним – 25 балів.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час поточного та підсумкового контролю за накопичувальною системою.

Підсумкова оцінка за навчальною дисципліною визначається для дисциплін з формою семестрового контролю екзамен (іспит) – сумуванням балів за поточний та підсумковий контроль.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: захист лабораторних робіт (45 бали), письмові контрольні роботи (15 балів).

Семестровий контроль: Екзамен (40 балів)

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

Приклад екзаменаційного білета та критерії оцінювання для навчальної дисципліни:

Приклад екзаменаційного білета

Харківський національний економічний університет ім. С. Кузнеця

Освітній рівень перший (бакалаврський)

Спеціальність 126 "Інформаційні системи та технології"

Освітня програма "Інформаційні системи та технології"

Навчальна дисципліна "Інтелектуальний аналіз даних"

Екзаменаційний білет № 1

Завдання 1 (стереотипне, 15 балів).

Поняття фракталу. Типи фракталів. Приклади геометричних фракталів. Визначення фрактальної розмірності.

Завдання 2 (евристичне, 15 балів).

Визначити зворотнє z -перетворення (для $n = 1, 2, 3, \dots, 5$):

$$Z^{-1} \left\{ F(z) = \frac{1,5z}{z^3 - 3,5z^2 + 5} \right\}$$

за методами: а) ділення чисельника $F(z)$ на знаменник $F(z)$; б) за рекурентним методом.

Завдання 3 (діагностичне, 10 балів).

Здійснити імітацію реалізації операцій відбору батьків та кросоверу еволюційного алгоритму (з формуванням нової популяції) для розв'язку діофантового рівняння $a + 2b + 3c + 4d = 30$, з обмеженнями $1 \leq a, b, c, d \leq 30$, де a, b, c і d – деякі додатні цілі числа.

Для реалізації завдання необхідно обрати: функцію фітнесу F_{fit} та тип кросоверу. Початкова популяція з 5 хромосом (комбінації значень коефіцієнтів a, b, c і d відповідно):

хромосома1: 11, 6, 9, 23; **хромосома2:** 10, 16, 19, 21; **хромосома3:** 14, 22, 19, 23; **хромосома4:** 17, 8, 9, 20; **хромосома 5:** 21, 6, 9, 13.

Критерії оцінювання

Підсумкові бали за екзамен складаються із суми балів за виконання всіх завдань, що округлені до цілого числа за правилами математики.

Алгоритм вирішення кожного завдання включає окремі етапи, які відрізняються за складністю, трудомісткістю та значенням для розв'язання завдання. Тому окремі завдання та етапи їх розв'язання оцінюються відокремлено один від одного таким чином:

Завдання 1 (стереотипне, від 15 до 0 балів)

Дане завдання оцінюється за 15-бальною шкалою.

15 балів виставляється у випадку, якщо поставлене завдання розкрито повністю: наведено вірну та розгорнуту відповідь;

14-13 балів – якщо поставлене завдання виконано в основному правильно, але є відповідь є недостатньо розгорнутою;

12-11 балів – якщо поставлене завдання виконано в основному правильно, але є незначні помилки у формулюваннях чи формальному описі;

10-9 балів – якщо поставлене завдання виконано в основному правильно, але присутні суттєві помилки у деяких формулюваннях чи формальному описі;

8-7 балів – якщо поставлене завдання виконано в основному правильно, але деякі важливі аспекти завдання висвітлено неповністю;

6-5 балів – якщо поставлене завдання виконано, але деякі частини відповіді надані з помилками та в неповному вигляді;

4-3 балів – якщо поставлене завдання виконано частково або зі значними помилками;

2-1 балів – якщо поставлене завдання майже не виконано, або виконано неякісно;

0 балів – якщо відсутня відповідь на поставлене завдання .

Завдання 2. (діагностичне, від 15 до 0 балів).

Дане завдання оцінюється за 15-бальною шкалою.

15 балів виставляється у випадку, якщо рішення поставленого завдання характеризується чіткістю, обґрунтованістю, правильними необхідними поясненнями, обчисленнями та перетвореннями. Продемонстровано вміння розв'язувати задачу за допомогою методів обчислювального інтелекту;

14-13 балів – якщо поставлене завдання виконано в основному правильно, але є відповідь є недостатньо розгорнутою;

12-11 балів – якщо поставлене завдання виконано в основному правильно, але не наведено всіх пояснень, недостатньо обґрунтовано методику розв'язання або не зроблено необхідні висновки;

10-9 балів – якщо поставлене завдання виконано без пояснень, але правильно використано методи, факти, формули і залежності. Можливі 1-2 незначні помилки в обчисленнях, які не впливають на правильність подальшого розв'язання;

8-7 балів – якщо поставлене завдання виконано в основному правильно, але деякі важливі аспекти завдання висвітлено неповністю;

6-5 балів – якщо поставлене завдання виконано, але деякі частини відповіді надані з помилками та в неповному вигляді; завдання розв'язано лише частково та є помилки, що суттєво вплинули на процес правильного розв'язання задачі;

4-3 балів – якщо поставлене завдання виконано частково або зі значними помилками;

2-1 балів – якщо поставлене завдання майже не виконано, або виконано неякісно;

0 балів – якщо відсутня відповідь на поставлене завдання.

Завдання 3. (евристичне, від 10 до 0 балів).

Дане завдання оцінюється за 10-бальною шкалою.

10 балів виставляється у випадку, якщо поставлене завдання розкрито повністю: використано один з можливих методів його розв'язання;

9 балів виставляється у випадку, якщо поставлене завдання виконано в основному правильно, але при його розв'язанні допущені незначні помилки у формулюваннях;

8 балів – якщо поставлене завдання виконано в основному правильно, але при його розв'язанні допущені помилки, що вплинули на остаточний результат;

7-6 бали – якщо поставлене завдання виконано в основному правильно, але деякі важливі аспекти завдання висвітлено неповністю;

5 бали – якщо поставлене завдання виконано, але підрозділи/процеси сформовані з помилками та в неповному вигляді

4-3 бали – якщо поставлене завдання виконано частково або зі значними помилками;

2-1 бал – якщо поставлене завдання майже не виконано, або виконано неякісно;

0 балів – якщо відсутня відповідь на поставлене завдання.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Інтелектуальний аналіз даних : навчальний посібник / О.В. Харкянен, О.М. М'якшило, С.В. Грибков. – Київ : НУХТ, 2019. – 170 с. Режим доступу: <https://library.nuft.edu.ua/>

2. Шабельник Т. В. Математичні методи інтелектуального аналізу даних : навч. посібник для здобувачів першого рівня вищої освіти спеціальності Системний аналіз / Т. В. Шабельник, О. Ф. Дяченко. – Маріуполь : МДУ, 2021. – 163 с. URL: <http://repository.hneu.edu.ua/handle/123456789/28088>

3. Акіменко В. В. Прикладні задачі інтелектуального аналізу даних (Data Mining) / В. Акіменко. – Київ : КНУ ім. Тараса Шевченка, 2018 р. – 162 с. [Електронний документ]/ Режим доступу: <https://www.researchgate.net> > DATA-MINING

Додаткова

4. Полянський В. О. Методи та моделі Data Science в системі управління безпекою / В. О. Полянський // Big data analytics: моделювання та інформаційні технології : міжнар. наук. симпозіум 20 бер. 2019 р. тези доп. – Київ : 2019. – С. 160–163. URL: <http://repository.hneu.edu.ua/handle/123456789/22431>

5. Аналіз та обробка потоків даних засобами обчислюваного інтелекту: монографія / Є. В. Бодяньський, Д. Д. Пелешко, О.А. Винокурова, С.В. Машталір, Ю.С. Іванов. – Львів : Видавництво Львівської політехніки, 2016. – 236 с.

6. Троцько В.В. Методи штучного інтелекту: навчально-методичний і практичний посібник. – Київ : Університет економіки та права "КРОК", 2020. – 86 с.

7. Сучасні інформаційні технології та системи: монографія / за заг. ред . В. С. Пономаренка / С.Г. Удовенко, Л.Е. Чала. Глава 6. Дослідження посилованих методів ранжування документів за результатами інформаційного пошуку (С. 91-117) // – Харків : Видавництво "Стиль", 2021. – 182 с .

8. Дудінова О.Б. Інтелектуальна обробка просторових даних в ГІС ландшафтно-екологічного моніторингу / О.Б. Дудінова, С.Г. Удовенко, Л.Е. Чала // Біоніка інтелекту. – 2020. – Вип. 2 (95). – С. 43-50.

9. Удовенко С.Г. Метод порівняння текстово-графічних фрагментів в електронних документах за гібридним критерієм / С.Г. Удовенко, Л.Е. Чала, Є.С. Кушвід // Біоніка інтелекту. – 2019. – Вип. 1 (92). – С. 71-76. Режим доступу : <http://repository.hneu.edu.ua/handle/123456789/22700>.

10. Удовенко С.Г. Комбінований метод нечіткої кластеризації даних в системах технічної діагностики / С.Г. Удовенко, Д.В. Келембет, О.В. Тесленко // Системи обробки інформації. – 2020. – Вип. 1 (160). – С. 7-17. Режим доступу: DOI:10.30748/soi.2020.160.01.0

Інформаційні ресурси

11. Інтелектуальний аналіз даних. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця [Електронний ресурс]. – Режим доступу : <https://pns.hneu.edu.ua/course/view.php?id=8049>

12. Моделювання систем у програмному середовищі Scilab&Xcos 5.5.1 [Електронний документ] – Режим доступу : <https://www.kv.by/node/10948>