

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Інститут високих технологій

Кафедра теоретичних основ високих технологій



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Технології аналізу даних

для студентів

галузь знань 10 Природничі науки
спеціальність **102 Хімія**
освітній рівень **магістр**
освітня програма **Високі технології (Хемоінформатика)**
вид дисципліни **вибіркова**

Форма навчання	<u>денна</u>
Навчальний рік	<u>2021/2022</u>
Семестр	<u>3</u>
Кількість кредитів ECTS	<u>3</u>
Мова викладання, навчання та оцінювання	<u>українська</u>
Форма заключного контролю	<u>екзамен</u>

Викладач: кан.фіз.-мат.наук. Васильєв Т.А.

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

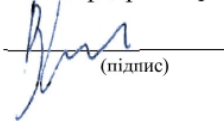
на 20__/20__ н.р. _____ (_____) «__» 20__ р.
(підпис, ПІБ, дата)

КИЇВ – 2021

Розробник: кан.фіз.-мат.наук. Васильєв Т.А.

ЗАТВЕРДЖЕНО

Зав. кафедри теоретичних основ високих технологій


(підпис)

Валерій ЛОЗОВСЬКИЙ

Протокол № _____ від «_____» _____ 2022 року

Схвалено науково - методичною комісією Інституту високих технологій

Протокол від «05» березня 2021 року №3

Голова науково-методичної комісії


(підпис)

(Русінчук Н.М.)

(прізвище та ініціали)

ВСТУП

1. Мета дисципліни :

забезпечення майбутніх фахівців з високих технологій необхідним апаратом аналізу даних, формування у них базових знань для розв'язування практичних задач зі сфери їх наукової діяльності; розвиток умінь формулювання прикладних задач аналізу великого обсягу даних з орієнтацією на проблеми фахової діяльності.

2. Попередні вимоги до опанування або вибору навчальної дисципліни :

1. Знання основ вищої математики та теорії ймовірності.
2. Знання основних елементарних методів математичної статистики.
3. Володіння навичками елементарних операцій з дійсними числами та дійсними змінними.

3. Анотація навчальної дисципліни

Предметом навчальної дисципліни "Технології аналізу даних" є дані, що виникають у теоретичних і прикладних дослідженнях природничих наук. У курсі вивчається і систематизується набір сучасних методів статистичної та аналітичної обробки даних, що дозволяють робити оцінку отриманих даних, їх обробку, виявлення особливостей в даних, та інтерпретацію результатів міждисциплінарних досліджень. Курс включає в себе приклади застосування запропонованих методів дослідження у природничих науках. Приведені розгорнуті способи візуального представлення даних за допомогою бібліотек мови програмування Python.

4. Завдання (навчальні цілі)

Навчання дисципліні має на меті розвивати у студентів такі компетентності:

Загальні компетентності (ЗК)

1. Знання та розуміння предметної області та розуміння професійної діяльності.
6. Здатність генерувати нові ідеї (креативність).
8. Здатність оцінювати та забезпечувати якість виконуваних робіт.
9. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності).
12. Здатність працювати автономно.
14. Здатність до пошуку, критичного аналізу та обробки інформації з різних джерел.

Фахові компетентності (ФК)

1. Здатність використовувати закони, теорії та концепції хімії у поєднанні із відповідними математичними інструментами для опису природних явищ.
2. Здатність будувати адекватні моделі хімічних явищ, досліджувати їх для отримання нових висновків та поглиблення розуміння природи, в тому числі з використанням методів молекулярного, математичного і комп'ютерного моделювання.
5. Здатність застосовувати методи комп'ютерного моделювання для вирішення наукових, хіміко-технологічних проблем та проблем хімічного матеріалознавства.

5. Результати навчання за дисципліною:

Результат навчання (1, знати; 2, вміти; 3, комунікація; 4, автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
1.1	Знати основні завдання технології аналізу даних з застосуванням пакетів комп'ютерних прикладних програм.	Лекція	Модульна контрольна робота, іспит	35
1.2	Знати основні контрольовані і неконтрольовані методи аналізу даних	Лекція		

1.3	Знати основні поняття і методи машинного навчання у застосуванні до задач аналізу даних у природничих науках.	Лекція	Модульна контрольна робота, іспит	35
1.4	Знати основні методи візуалізації даних за допомогою пакетів прикладних програм Origin, Microsoft Excel тощо та бібліотек мови програмування Python.	Лекція, семінарська робота		
2.1	Вміти готувати дані для аналізу за допомогою пакетів прикладних програм Origin, Microsoft Excel тощо та бібліотек мови програмування Python.	Самостійна робота, семінарська робота	Підготовка реферату	30
4.1	Вміти знаходити необхідну інформацію у літературі та електронних базах, оцінювати її релевантність та достовірність	Самостійна робота		

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни	1.1	1.2	1.3	1.4	2.1	4.1
Програмні результати навчання						
P2. Глибоко розуміти основні факти, концепції, принципи і теорії, що стосуються предметної області, опанованої у ході магістерської програми, використовувати їх для розв'язання складних задач і проблем, а також проведення досліджень з відповідного напрямку хімії.	+	+	+	+	+	+
P3. Застосовувати отримані знання і розуміння для вирішення нових якісних та кількісних задач хімії.	+	+	+	+	+	+
P5. Володіти методами комп'ютерного моделювання структури, параметрів і динаміки хімічних систем.	+	+	+	+	+	+
P6. Знати методологію та організації наукового дослідження.			+	+	+	+
P8. Вміти ясно і однозначно донести результати власного дослідження до фахової аудиторії та/або нефхівців.				+	+	+
P9. Збирати, оцінювати та аналізувати дані, необхідні для розв'язання складних задач хімії, використовуючи відповідні методи та інструменти роботи з даними.	+	+	+	+	+	+
P13. Аналізувати наукові проблеми та пропонувати їх вирішення на абстрактному рівні шляхом декомпозиції їх на складові, які можна дослідити окремо.	+	+	+	+	+	+

7. Схема формування оцінки.

7.1 Форми оцінювання студентів: - семестрове оцінювання:

1. Модульна контрольна робота 1 – РН 1.1; 1.2. – 15 балів/ 20 балів
2. Модульна контрольна робота 2 – РН 1.3; 1.4 – 15 балів/ 20 балів
3. Оцінювання реферату РН 2.1 – 30 балів

- підсумкове оцінювання: у формі іспиту

Підсумкова оцінка з освітнього компоненту в цілому: підсумковою формою контролю за яким встановлено іспит визначається як сума оцінок (балів) за всіма успішно оціненими результатами навчання під час семестру (оцінки нижче мінімального порогового рівня до підсумкової оцінки не додаються) та оцінки, отриманої під час іспиту.

Формою проведення іспиту є тестова контрольна робота. Результатами навчання, які оцінюються в тестовій контрольній роботі, є РН 1.1-1.4. Максимальна кількість балів, які можуть бути отримані студентом, становить 40 балів за 100 бальною шкалою. Перескладання семестрового контролю з метою покращення позитивної оцінки не допускається.

- умови допуску до підсумкового іспиту:

Обов'язковим для заліку є успішне написання 2 модульних контрольних робіт (по кожній не менше 50% правильних відповідей), реферату. Студент не допускається до іспиту, якщо під час семестру набрав менше ніж 20 балів (рекомендований мінімум 36 балів).

7.2 Організація оцінювання: Модульні контрольні роботи 1 і 2 проводяться після завершення лекцій зі змістовних модулів 1 і 2, відповідно. Реферат оцінюється протягом семестру.

7.3 Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

8. Структура навчальної дисципліни.

Тематичний план лекцій та лабораторних занять

Номер і назва теми*	Кількість годин		
	лекції	Семінарські, практичні та лаб. заняття	Самостійна робота
Основні поняття обробки даних. Контрольовані і неконтрольовані методи аналізу даних. Міжгалузевий стандартний процес обробки даних CRISP-DM. Розуміння і підготовка даних. Моделювання, оцінка, запуск.	2		6
Основні методи підготовки даних для аналізу. Доповнення даних. Перетворення символічних та текстових даних на числові. Нормалізація і стандартизація даних. Використання прикладних програмних пакетів Origin та Microsoft Office для обробки даних	2		7
Поняття ентропії та приросту інформації при аналізі даних. Вибір інформативних параметрів, що описують об'єкти. Розрахунок ентропії та приросту інформації при створенні класифікаційного бінарного дерева	2	4	10
Розбиття на класи (групи). Розбиття на класи об'єктів, які описуються даними. Класифікаційне дерево. Область зразка. Підрахунок вірогідності групи	2	3	10
Типові завдання обробки даних. Класифікація та підрахування вірогідності належності до класу; регресія. Визначення подібностей і кластеризація. групування за збігами;	2		7

профілювання або опис поведінки. Прогнозування зв'язків і семплювання. причинно-наслідкове моделювання			
Елементарні поняття методів машинного навчання. Поняття машинного навчання, перенавчання і генералізації. Адаптивний лінійний нейрон. Бінарна класифікація. Правило порогового навчання. Логістична регресія. Метод опорних векторів. Нелінійна класифікація. Метод Кнайближчих сусідів	2	4	10
Використання можливостей мови Python для аналізу даних. Огляд синтаксису мови Python. Встановлення бібліотек Python і використання їх для обробки даних. Використання пакетів мови Python для редагування та візуалізації даних, розміщених у файлі Microsoft Excel	2	4	10
Модульна контрольна робота	1		
ВСЬОГО	15	15	60

Загальний обсяг 90 год., в тому числі:

Лекції – 15 год

Практичні – 15 год.

Самостійна робота - 60 год.

9. Рекомендовані літературні джерела:

Основна:

1. S. Raschka, V. Mirjalili. Python machine learning, Packt Publishing, 2019.
2. J. Unpingco, Python for probability, statistics, and machine learning, Springer, 2019.

Додаткова:

1. Ф. Провост, Т. Фоусетт, Data science для бізнесу, К.: Наш формат, 2019.
2. J. Walkenbach. Excel 2013 Formulas. Wiley, 2018.
3. О. М. Васильєв, Програмування мовою Python. Тернопіль: Богдан, 2019

10. Додаткові джерела та інформація:

1. Origin user guide, OriginLab Corp. 2020.
https://d2mvzyuse3lwjc.cloudfront.net/pdfs/Origin2020b_Documentation/English/Origin_User_Guide_2020b_E.pdf#zoom=100