

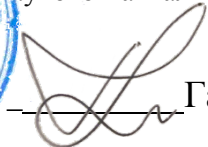
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

Інститут високих технологій

Кафедра теоретичних основ високих технологій



«ЗАТВЕРДЖУЮ»
Заступник директора
науково-навчальної роботи


Галина ГРАБЧУК

«22» березня 2021 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Технологія аналізу даних в природничих науках

для студентів

галузь знань	10 Природничі науки
спеціальність	102 Хімія
освітній рівень	другий (магістр)
освітньо-наукова програма	Високі технології (хімія та наноматеріали)
вид дисципліни	обов'язкова

Форма навчання	<u>денна</u>
Навчальний рік	2021/2022
Семестр	4
Кількість кредитів ECTS	3.0
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	залік

Викладач: Васильєв А.Г.

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

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

КИЇВ – 2021

Розробник: Васильєв Т.А., к.ф.-м.н., асистент кафедри теоретичних основ високих технологій

«ЗАТВЕРДЖЕНО»

Завідувач кафедри теоретичних основ високих технологій

 Валерій ЛОЗОВСЬКИЙ

Протокол № 11 від «03» 03 2021р.

Схвалено науково - методичною комісією

«Інституту високих технологій»

Київського національного університету імені Тараса Шевченка

Протокол від «⁰⁵» 03 2021 року № 3

Голова науково-методичної комісії  (Русінчук Н.М.)

1. Мета дисципліни – забезпечення майбутніх фахівців з високих технологій у прикладній фізиці та наноматеріалах 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. Володіння науково-теоретичним та практичним матеріалом навчальних дисциплін, які викладаються студентам освітнього рівня «Бакалавр».

2. Знання основних елементарних методів математичної статистики та програмування.

3. Володіння навичками елементарних операцій з дійсними числами та дійсними змінними.

3. Анотація навчальної дисципліни:

Предметом навчальної дисципліни "Технології аналізу даних в природничих науках" 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):

ЗК1. Знання та розуміння предметної області та розуміння професійної діяльності

ЗК2. Здатність вчитися та оволодівати сучасними знаннями.

ЗК3. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК4. Здатність застосовувати знання у практичних ситуаціях.

ЗК5. Здатність до адаптації та дії в новій ситуації.

ЗК6. Здатність генерувати нові ідеї (креативність), а також формулювати судження, маючи неповну або обмежену інформацію.

ЗК7. Навички використання інформаційних і-комунікаційних технологій в хімічних дослідженнях та професійній діяльності.

ЗК8. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ФК1. Здатність використовувати закони, теорії та концепції хімії у поєднанні із відповідними математичними інструментами для опису природних явищ.

ФК2. 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.

ФК5. 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)	Методи оцінювання та пороговий	Відсоток у підсумковій
---	------------------------------------	--------------------------------	------------------------

Код	Результат навчання	викладання і навчання	критерій оцінювання (за необхідності)	оцінці з дисципліни
1.1	Знати основні завдання технологій аналізу даних в природничих науках	Лекція	Модульна контрольна робота, залік	35
1.2	Знати основні методи підготовки даних	Лекція		
1.3	Знати основні методи створення класифікаційних моделей та методи машинного навчання у застосуванні до задач аналізу даних у природничих науках	Лекція	Модульна контрольна робота, залік	35
1.4	Знати основні методи підготовки, візуалізації та аналізу даних за допомогою пакетів прикладних програм та бібліотек мови програмування Python.	Лекція		
2.1	Вміти готувати дані для аналізу за допомогою пакетів прикладних програм та бібліотек мови програмування Python	Самостійна 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 (код) Програмні результати навчання (назва)	1.1	1.2	1.3	1.4	2.1	4.1
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.	+					+
8. Здатність формулювати нові гіпотези та наукові задачі в галузі хімії, вибирати напрями та відповідні методи для їх розв'язання на основі розуміння сучасної проблематики досліджень в галузі хімії та беручи до уваги наявні ресурси.	+	+	+	+	+	
ПРН09. Визначати напрями перспективних досліджень з урахуванням світових тенденцій розвитку науки, техніки й технологій.						+

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 заліку, якщо під час семестру набрав менше ніж 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 і практичних занять

№ п/п	Назва теми*	Кількість годин		
		лекції	семінари/ практичні/ лабораторні і виврати необхідне	Самостійна робота
<u>ЗМ1: Основні методи підготовки даних до аналізу</u>				
1	<i>Основні поняття обробки даних.</i> Контрольовані і неконтрольовані методи аналізу даних. Міжгалузевий стандартний процес обробки даних CRISP-DM. Розуміння і підготовка даних. Моделювання, оцінка, запуск.	2	0	5
2	<i>Основні методи підготовки даних для аналізу.</i> Доповнення даних. Перетворення символічних та текстових даних на числові. Нормалізація і стандартизація даних. Використання бібліотек Pandas, Numpy, Matplotlib для підготовки та візуалізації даних	2	0	5
3	<i>Методи стиснення даних.</i> Коваріаційна матриця. Метод головних компонент. Лінійний дискримінантний аналіз. Ядерний аналіз головних компонент.	4	0	10
4	<i>Поняття ентропії та приросту інформації при аналізі даних.</i> Вибір інформативних параметрів, що описують об'єкти. Розрахунок ентропії та приросту інформації при створенні класифікаційного бінарного дерева	2	0	5
5	<i>Задачі семплювання.</i> Створення зразків випадкових величин методом Монте-Карло. Статистичні характеристики, оцінка ймовірності та довірчий інтервал. Використання критеріїв Стюдента, Фішера, Фостера-Стюарта. Довірчі інтервали оцінювання вибіркового середнього та вибіркової дисперсії. Кластеризація і створення вибірки.	2	0	5
	<i>Модульна контрольна робота № 1</i>	1	0	
	<i>Всього</i>	15	0	30
<u>ЗМ2: Основні завдання аналізу даних</u>				
6	<i>Задачі класифікації.</i> Класифікація та підрахування вірогідності належності до класу. Дерева прийняття рішень. Дерева класифікації і регресії. Випадкові ліси.	2	0	5
7	<i>Методи машинного навчання.</i> Поняття машинного навчання, перенавчання і генералізації. Адаптивний лінійний нейрон. Правило порогового навчання. Логістична регресія. Метод опорних векторів. Нелінійна класифікація. Метод К-найближчих сусідів	6	0	10
8	<i>Використання можливостей мови Python для аналізу даних.</i> Бібліотеки для статистичної обробки даних (Scipy, sumpy). Бібліотека Scikit-learn для аналізу даних	6	0	15

	<i>Модульна контрольна робота № 2</i>	1		
	<i>Всього</i>	15	0	30
	ВСЬОГО	30	0	60

Загальний обсяг 90 год., в тому числі:

Лекцій – **30** год.

Практичні заняття – **0** год.

Самостійна робота – **60** год.

9. Рекомендовані джерела:

Основна: (Базова)

1. S. Raschka, V. Mirjalili. Python machine learning, Packt Publishing, 2019.
2. J. Unpingco, Python for probability, statistics, and machine learning, Springer, 2019.
3. П. І. Бідюк, О. М. Терентьев, Т. І. Просянкіна-Жарова, Прикладна статистика Навчальний посібник, Вінниця ПП «ТД» Едельвейс і К», 2013.

Додаткова:

1. J. Walkenbach. Excel 2013 Formulas. Wiley, 2018.
2. О. М. Васильєв, Програмування мовою Python. Тернопіль: Богдан, 2019.

Інтернет-ресурси:

1. Origin user guide, OriginLab Corp. 2020.
https://d2mvzyuse3lwjc.cloudfront.net/pdfs/Origin2020b_Documentation/English/Origin_User_Guide_2020b_E.pdf#zoom=100
2. UCI Machine Learning Repository
<https://archive.ics.uci.edu/ml/index.php>