

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Інститут високих технологій

Кафедра теоретичних основ високих технологій

«ЗАТВЕРДЖУЮ»

Заступник директора
з науково-педагогічної роботи

Галина ГРАБЧУК

« 22 » 05 2021 року

Протокол № 3

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ В ПРИРОДНИЧИХ НАУКАХ
для студентів**

галузь знань 10 « Природничі науки »
спеціальність 105 « Прикладна фізика та наноматеріали »
освітній рівень Магістр
освітня програма « Високі технології (Прикладна фізика та наноматеріали) »
вид дисципліни обов'язкова

Форма навчання	<u>денна</u>
Навчальний рік	2021/2022
Семестр	4
Кількість кредитів ECTS	3
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	залік

Викладач: 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: 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):

Згідно з вимогами Стандарту вищої освіти України (другий (магістерський) рівень вищої освіти (восьмий рівень НРК України), галузь знань 10 Природничі науки, спеціальність 105 Прикладна фізика та наноматеріали дисципліна забезпечує набуття студентами таких компетентностей:

інтегральної:

здатність ставити і розв’язувати задачі аналізу великого обсягу даних, які виникають у дослідженнях природничих наук.

загальних:

ЗК04. Навички використання інформаційних і комунікаційних технологій.

ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК08. Навички міжособистісної взаємодії.

ЗК10. Навички здійснення безпечної діяльності.

ЗК13. Здатність розв’язувати складні задачі і проблеми, що потребує оновлення та інтеграції знань, часто в умовах неповної/недостатньої інформації та суперечливих вимог.

ЗК16. Здатність генерувати нові ідеї.

ЗК18. Здатність провадження дослідницької та інноваційної діяльності на відповідному рівні.

спеціальних (фахових, предметних):

ФК04. Здатність встановлювати взаємозв'язок внутрішньої структури елементів та компонентів сучасного обладнання з їх електричними і електрофізичними характеристиками та параметрами.

ФК06. Здатність встановлювати області застосування виробів електронної техніки.

ФК09. Здатність відслідковувати найновіші досягнення в області прикладної фізики та високих технологій, вивчаючи наукову літературу та взаємокорисно спілкуючись із колегами.

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 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)	1.1	1.2	1.3	1.4	2.1	4.1
Програмні результати навчання (назва)						

ПРН01. Володіти поглибленим рівнем знань у прикладній фізиці, наноматеріалознавстві, високих технологіях та споріднених областях, включаючи методики проведення експериментів і технології отримання наноматеріалів, рівень цих знань повинен бути достатнім для проведення наукових досліджень на рівні останніх світових досягнень і направленим на їх розширення та поглиблення.			+		+	
ПРН02. Використовувати знання в галузі прикладної фізики, математики, електроніки та інформаційних технологій для виконання наукових досліджень та розв'язання виробничих задач.						+
ПРН05. Обговорювати та знаходити прогресивні та інноваційні рішення проблем і завдань при виконанні науково-технічних та виробничих проектів.	+					+
ПРН06. Встановлювати та аргументувати нові залежності між параметрами та характеристиками фізичних систем.	+	+	+	+	+	
ПРН09. Визначати напрямки 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 вибрати необхідне	Самостійна робота
<u>ЗМ1: Основні методи підготовки даних до аналізу</u>				
1	<i>Основні поняття обробки даних. Контрольовані і неконтрольовані методи аналізу даних. Міжгалузевий стандартний процес обробки даних CRISP-DM. Розуміння і підготовка даних. Моделювання, оцінка, запуск.</i>	2	0	5
2	<i>Основні методи підготовки даних для аналізу. Доповнення даних. Перетворення символічних та текстових даних на числові. Нормалізація і стандартизація даних. Використання бібліотек Pandas, Numpy, Matplotlib для підготовки та візуалізації даних</i>	2	0	5
3	<i>Методи стиснення даних. Коваріаційна матриця. Метод головних компонент. Лінійний дискримінантний аналіз. Ядерний аналіз головних компонент.</i>	4	0	10
4	<i>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</i>	2	0	5
5	<i>Задачі семплювання. 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.</i>	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. Класифікація та підрахування вірогідності належності до класу. Дерева прийняття рішень. Дерева класифікації і регресії. Випадкові ліси.</i>	2	0	5
7	<i>Методи машинного навчання. Поняття машинного навчання, перенавчання і генералізації. Адаптивний лінійний нейрон. Правило порогового навчання. Логістична регресія. Метод опорних векторів. Нелінійна класифікація. Метод К-найближчих сусідів</i>	6	0	10
8	<i>Використання можливостей мови Python для</i>	6	0	15

	<i>аналізу даних. Бібліотеки для статистичної обробки даних (Scipy, sumpy). Бібліотека Scikit-learn для аналізу даних</i>			
	<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. Просянкін-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>