

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Інститут високих технологій

Кафедра теоретичних основ високих технологій

«ЗАТВЕРДЖУЮ»

Заступник директора
з науково-педагогічної
роботи

Галина ГРАБЧУК

« 25 » серпень 2020 року



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ
для студентів

галузь знань	09	<u>«Біологія»</u>
спеціальність	091	<u>«Біологія»</u>
освітній рівень		<u>Магістр</u>
освітня програма		<u>«Біоінформатика та структурна біологія»</u>
вид дисципліни		<u>обов'язкова</u>

Форма навчання	<u>денна</u>
Навчальний рік	2021/2022
Семестр	<u>1</u>
Кількість кредитів ECTS	5
Мова викладання, навчання та оцінювання	українська
Форма заключного контролю	екзамен

Викладач: Васильєв Т.А.

Пролонговано: на 2021/2022 н.р. (підпис, ПІБ, дата) «05» 05 2021 р. проф. 03

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

КИЇВ – 2020

ИНО 03

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

«ЗАТВЕРДЖЕНО»

Завідувач кафедри теоретичних основ високих технологій


_____ Валерій ЛОЗОВСЬКИЙ

Протокол № 16 від «10» серпня 2020 р.

Схвалено науково - методичною комісією

«Інституту високих технологій»

Київського національного університету імені Тараса Шевченка

Протокол від «10» серпня 2020 року № 3

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

1. Мета дисципліни – забезпечення майбутніх фахівців з високих технологій у біотехнології 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 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):

Дисципліна забезпечує набуття студентами таких компетентностей:

ЗК02. Здатність використовувати інформаційні та комунікаційні технології.

СК01. Здатність користуватися новітніми досягненнями біології, необхідними для професійної, дослідницької та/або інноваційної діяльності. Здатність застосовувати знання у професійній діяльності з урахуванням новітніх досягнень, у т.ч. для дослідницької роботи.

СК02. Здатність формулювати задачі моделювання, створювати моделі об'єктів і процесів на прикладі різних рівнів організації живого із використанням математичних методів й інформаційних технологій.

СК03. Здатність користуватися сучасними інформаційними технологіями та аналізувати інформацію в галузі біології і на межі предметних галузей.

СК11. Вміння розробляти програмне забезпечення для обробки біомолекулярних даних.

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, лабораторна робота		
2.1	Вміти готувати дані для аналізу за допомогою пакетів прикладних програм Origin, Microsoft Excel тощо та бібліотек мови програмування Python.	Самостійна робота, лабораторна 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
ПР2. Використовувати бібліотеки, інформаційні бази даних, інтернет ресурси для пошуку необхідної інформації.	+	+	+	+	+	+
ПР4. Розв'язувати складні задачі в галузі біології, генерувати та оцінювати ідеї.	+	+	+	+	+	+
ПР11. Проводити статистичну обробку, аналіз та узагальнення отриманих експериментальних даних із використанням програмних засобів та сучасних інформаційних технологій.					+	+
ПР16. Моделювати об'єкти і процеси у живих організмах та їхніх компонентах із використанням математичних методів й інформаційних технологій.					+	+
ПР17. Розробляти програмне забезпечення для обробки біомолекулярних даних.					+	+

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 і практичних занять**

№ п/п	Назва теми*	Кількість годин		
		лекції	семінари/ практичні/ лабораторні вибрати необхідне	Самостій на 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	10
2	<i>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</i>	4	0	10
3	<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	4	20
4	<i>Розбиття на класи (групи). Розбиття на класи об'єктів, які описуються даними. Класифікаційне дерево. Область зразка. Підрахунок вірогідності групи</i>	4	2	20
	<i>Модульна контрольна робота № 1</i>	1	0	
	<i>Всього</i>	13	6	60
<u>ЗМ2: Основні завдання аналізу даних</u>				
5	<i>Типові завдання обробки даних. Класифікація та підрахування вірогідності належності до класу; регресія. Визначення подібностей і кластеризація. групування за збігами; профілювання або опис поведінки. Прогнозування зв'язків і семплювання. причинно-наслідкове моделювання</i>	6	0	8
6	<i>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 К-найближчих сусідів</i>	4	4	20
7	<i>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.</i>	4	4	20

	<i>Модульна контрольна робота № 2</i>	1		
	<i>Всього</i>	15	8	48
	ВСЬОГО	28	14	108

Загальний обсяг 150 год., в тому числі:

Лекцій – **28** год.

Лабораторні заняття –**14** год.

Самостійна робота – **108** 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.

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.

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