

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

**Інститут високих технологій**

Кафедра теоретичних основ високих технологій



**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ В ПРИРОДНИЧИХ НАУКАХ**

для студентів

галузь знань	№16	«Хімічна та біоінженерія»
спеціальність	№162	«Біотехнології та біоінженерія»
освітній рівень	Магістр	
освітня програма	«Високі технології (Біотехнологія)»	
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.

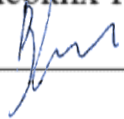
Пролонговано: на 20\_\_/20\_\_ н.р. \_\_\_\_\_ (\_\_\_\_\_) «\_\_» 20\_\_р.  
(підпис, ПІБ, дата)

на 20\_\_/20\_\_ н.р. \_\_\_\_\_ (\_\_\_\_\_) «\_\_» 20\_\_р.  
(підпис, ПІБ, дата)

Розробник: Васильєв Тарас Анатолійович, кандидат фізико-математичних наук, асистент кафедри теоретичних основ високих технологій

ЗАТВЕРДЖЕНО

Завідувач кафедри теоретичних основ високих технологій

  
\_\_\_\_\_ Валерій ЛОЗОВСЬКИЙ

Протокол № 11 від «03» березня 2021р.

Схвалено науково - методичною комісією Інституту високих технологій Київського національного університету імені Тараса Шевченка

Протокол від «05» березня 2021 року №3

Голова науково-методичної комісії   
\_\_\_\_\_ Наталя РУСІНЧУК

«05» березня 2021 року

**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:

K01. Здатність проведення досліджень на відповідному рівні.

K05. Здатність виявляти ініціативу та підприємливість.

K08. Здатність здійснювати пошук необхідної інформації в науковій і технічній літературі, базах даних та інших джерелах.

K09. Здатність відбирати та аналізувати релевантні дані, у тому числі за допомогою сучасних методів аналізу даних і спеціалізованого програмного забезпечення.

K11. Здатність розробляти нові біотехнологічні об’єкти і технології та підвищувати ефективність існуючих технологій на основі експериментальних та/або теоретичних досліджень та/або комп’ютерного моделювання, в тому числі методів біоінформатики.

K12. Здатність планувати і виконувати експериментальні роботи в галузі біотехнології з використанням сучасних обладнання та методів, інтерпретувати отримані дані на основі сукупності сучасних знань та уявлень про об’єкт і предмет дослідження, робити обґрунтовані висновки.

K23. Здатність розробляти/застосовувати комбіновані біотехнології за допомогою міждисциплінарних підходів, зокрема, з використанням технологій матеріалознавства та хімічних технологій.

## 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, залік	35
1.2	Знати основні 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, залік	35
1.4	Знати основні методи візуалізації даних за допомогою пакетів прикладних програм Origin, Microsoft Excel тощо та бібліотек мови програмування Python.	Лекція		
2.1	Вміти готувати дані для аналізу за допомогою пакетів прикладних програм Origin, Microsoft Excel тощо 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
ПР18. Знаходити необхідну інформацію у науковій та довідниковій літературі, електронних базах, інших джерелах інформації, оцінювати її релевантність та достовірність	+	+	+	+	+	+
ПР19. Оцінювати актуальність досліджуваних наукових проблем, придатність відомих наукових методів для їх дослідження на основі аналізу наявних даних та публікацій у провідних виданнях	+	+	+	+	+	+
ПР21. Мати навички планування та виконання експериментальних досліджень як особисто, так і у колективі, критичного аналізу отриманих результатів; оформлення результатів досліджень у вигляді звіту, наукової публікації, презентації на наукових та інших заходах.					+	+
ПР26. Застосовувати методи біоінформатики та обчислювальної структурної біології для раціонального дизайну (біо)молекул та матеріалів з заданою біологічною активністю.					+	+

## 7. Схема формування оцінки.

### 7.1 Форми оцінювання студентів:

#### - семестрове оцінювання:

1. Модульна контрольна робота 1 – РН 1.1; 1.2. – 35 балів/ 20 балів
2. Модульна контрольна робота 2 – РН 1.3; 1.4 – 35 балів/ 20 балів
3. Оцінювання самостійного проєктного завдання РН 2.1, 4.1 – 30 балів

#### - підсумкове оцінювання: немає.

### 7.2 Організація оцінювання:

Модульні контрольні роботи 1 і 2 проводяться після завершення лекцій зі змістовних модулів 1 і 2, відповідно. Самостійне проєктне завдання оцінюється протягом семестру.

### 7.3 Шкала відповідності оцінок

Зараховано / Passed	60-100
Не зараховано / Fail	0-59

## 8. Структура навчальної дисципліни.

### Тематичний план лекцій

№ п/п	Назва теми*	Кількість годин		
		лекції	-	Самостійна робота
<b><u>ЗМ1: Основні методи підготовки даних до аналізу</u></b>				
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 прикладних програмних пакетів Origin та Microsoft Office для обробки даних</i>	6	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	0	5
4	<i>Розбиття на класи (групи). Розбиття на класи об'єктів, які описуються даними. Класифікаційне дерево. Область зразка. Підрахунок вірогідності групи</i>	4	0	10
	<i>Модульна контрольна робота № 1</i>	1	0	

	<i>Всього</i>	15	0	30
<b>ЗМ2: Основні завдання аналізу даних</b>				
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	10
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	0	10
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	0	10
	<i>Модульна контрольна робота № 2</i>	1		
	<i>Всього</i>	15	0	30
	<b>ВСЬОГО</b>	<b>30</b>	<b>0</b>	<b>60</b>

**Загальний обсяг 90 год.**, в тому числі:

Лекцій – **30 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.

**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](https://d2mvzyuse3lwjc.cloudfront.net/pdfs/Origin2020b_Documentation/English/Origin_User_Guide_2020b_E.pdf#zoom=100)