

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І.І.МЕЧНИКОВА
Кафедра комп'ютерних систем та технологій



“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної роботи

“ ”

_____ 2023 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**ВП 9.1. СПЕЦКУРС №2
ОБРОБКА ТА АНАЛІЗ ДАНИХ ЗАСОБАМИ Python**

Рівень вищої освіти	перший (бакалаврський)
Галузь знань	12 «Інформаційні технології»
Спеціальність	122- Комп'ютерні науки
Освітньо-професійна програма	Комп'ютерні науки

ОНУ
Одеса
2023

Робоча програма навчальної дисципліни «Спецкурс №2. Обробка та аналізу даних засобами Python». – Одеса: ОНУ, 2023. – 14с.

Розробники: кандидат фізико-математичних наук, доцент кафедри комп'ютерних систем та технологій Коренкова Ганна Валентинівна

Робоча програма затверджена на засіданні кафедри комп'ютерних систем та технологій

Протокол № 1 від. "30" серпня 2023 р.

Завідувач кафедри _____ (Юрій ГУНЧЕНКО)

Погоджено із гарантом ОПП «Комп'ютерні науки»

_____ (Алла КАМЄНЄВА)

Схвалено навчально-методичною комісією (НМК) ФМФІТ

Протокол № 1 від "31" серпня 2023 р.

Голова НМК _____ (Алла РАЧИНСЬКА)

Переглянуто та затверджено на засіданні кафедри комп'ютерних систем та технологій

Протокол № ____ від. "____" _____ 20__ р.

Завідувач кафедри _____ (_____)

Переглянуто та затверджено на засіданні кафедри комп'ютерних систем та технологій

Протокол № ____ від. "____" _____ 20__ р.

Завідувач кафедри _____ (_____)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, спеціалізація, рівень вищої освіти	Характеристика навчальної дисципліни
		<i>денна форма навчання</i>
Загальна кількість: кредитів – 3 годин –90 Змістових модулів -2	Галузь знань <u>12– інформаційні технології</u> Спеціальність <u>122 – комп’ютерні науки</u> Рівень вищої освіти: перший бакалаврський	вибіркова
		<i>Рік підготовки:</i>
		3
		<i>Семестр</i>
		5
		<i>Леції</i>
		17 год.
		<i>Практичні, семінарські</i>
		-
		<i>Лабораторні</i>
		17 год.
		<i>Самостійна робота</i>
		56 год.
Форма підсумкового контролю: залік		

2. Мета та завдання навчальної дисципліни

Мета: Отримання навичок ефективної обробки, аналізу та візуалізації даних з використанням мови програмування Python.

Завдання: Створення, очищення та аналіз наборів даних за допомогою мови програмування Python, а також вивчення методів візуалізації та висновків інсайтів з оброблених даних.

Процес вивчення дисципліни спрямований на формування елементів наступних **компетентностей**:

ІК. Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук або у процесі навчання, що передбачає застосування теорій та методів інформаційних технологій і характеризується комплексністю та невизначеністю умов.

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК8. Здатність генерувати нові ідеї (креативність).

У результаті вивчення навчальної дисципліни здобувач вищої освіти повинен

знати:

- Основні поняття аналізу даних та їх роль у різних сферах.
- Основи мови програмування Python, включаючи роботу зі змінними, умовними операторами, циклами та функціями.
- Основи роботи з бібліотекою Pandas для створення та маніпулювання об'єктами DataFrame.
- Методи очищення даних від пропущених значень, дублікатів та некоректних даних.
- Побудову графіків, діаграм та інших візуалізацій даних за допомогою бібліотеки Matplotlib.
- Основи статистичного аналізу даних, включаючи розрахунок середнього, медіани, стандартного відхилення та інших показників.
- Роботу з зовнішніми джерелами даних, такими як CSV файли, JSON та бази даних.

вміти:

- Створювати та редагувати таблиці (DataFrame) для обробки даних.

- Використовувати методи фільтрації, сортування та агрегації для виявлення залежностей в даних.
- Використовувати статистичні методи для аналізу розподілу та характеристик даних.
- Побудову візуалізацій для зручного сприйняття та представлення даних.
- Адаптувати набуті знання для вирішення конкретних практичних завдань та створення обґрунтованих висновків.

Що забезпечує наступні програмні результати навчання:

ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.

ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

ПР13. Володіти мовами системного програмування та методами розробки програм, що взаємодіють з компонентами комп'ютерних систем, знати мережні технології, архітектури комп'ютерних мереж, мати практичні навички технології адміністрування комп'ютерних мереж та їх програмного забезпечення.

ПР17. Виконувати організацію комп'ютерних паралельних та розподілених обчислень, застосовувати чисельні методи та алгоритми для паралельних структур, мови паралельного програмування при розробці та експлуатації паралельного та розподіленого програмного забезпечення.

3. Зміст навчальної дисципліни ЗМІСТОВИЙ МОДУЛЬ 1.

ОСНОВИ ОБРОБКИ ТА АНАЛІЗУ ДАНИХ З Python

Тема 1. Введення до аналізу даних

Роль аналізу даних у сучасному світі. Поняття даних, основні завдання обробки даних, особливості обробки даних. Класифікація та загальний огляд етапів та методів обробки даних.

Тема 2. Огляд бібліотек мови Python для аналізу даних

Змінні та типи даних. Умовні оператори та цикли. Функції та обробка винятків. Робота зі списками, кортежами, словниками та множинами. Маніпулювання даними: фільтрація, перетворення, агрегація. Основні елементи та налаштування бібліотек мови Python. Структури даних: Series та DataFrame. Читання та запис даних з файлів.

Тема 3. Очищення та підготовка даних

Робота з пропущеними значеннями. Основні кроки підготовки даних перед аналізом. Виявлення та видалення дублікатів. Методи обробки пропущених значень: видалення, заповнення, інтерполяція. Вплив пропущених значень на аналіз та як їх керувати. Причини виникнення дублікатів у даних. Методи виявлення дублікатів. Підходи до рішення проблеми дублікатів: видалення, об'єднання. Робота з зовнішніми джерелами даних.

ЗМІСТОВИЙ МОДУЛЬ 2.

ВІЗУАЛІЗАЦІЯ, СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ

Тема 4. Візуалізація даних

Значення візуалізації в аналізі даних. Основні типи графіків та діаграм. Основи створення графіків за допомогою Matplotlib. Налаштування заголовків, підписів, легенд та осей графіків. Створення мультиплотів та субплотів для порівняння даних. Додаткові елементи візуалізацій.

Тема 5. Статистичний аналіз даних

Використання бібліотеки NumPy для обчислення статистичних показників. Обчислення основних показників описової статистики: середнє, медіана, мода, дисперсія. Візуалізація розподілу даних за допомогою гістограм, діаграм розсіювання тощо.

4. Структура навчальної дисципліни

Назви тем	Кількість годин				
	Денна форма				
	Усього	у тому числі			
		л	п/с	лаб	ср
1	2	3	4	5	6
ЗМІСТОВИЙ МОДУЛЬ 1. ОСНОВИ ОБРОБКИ ТА АНАЛІЗУ ДАНИХ З Python					
Тема 1. Введення до аналізу даних	16	2		4	10
Тема 2. Огляд бібліотек мови Python для аналізу даних	20	5		3	12
Тема 3. Очищення та підготовка даних	16	2		2	12
Разом за змістовим модулем 1	52	9		9	34
ЗМІСТОВИЙ МОДУЛЬ 2. ВІЗУАЛІЗАЦІЯ, СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ					
Тема 4. Візуалізація даних	19	4		4	11
Тема 5. Статистичний аналіз даних	19	4		4	11
Разом за змістовим модулем 2	38	8		8	22
Усього годин	90	17		17	56

5. Теми семінарських занять

Семінарські заняття не передбачені.

6. Теми практичних занять

Практичні заняття не передбачені.

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Знайомство з середовищем Python та Jupyter Notebook.	4

	Встановлення Python та Jupyter Notebook. Основи роботи з Jupyter Notebook: створення, виконання, збереження.	
2	Основи роботи з бібліотекою Pandas. Зчитування даних з CSV файлу. Огляд та фільтрація даних. Робота з пропущеними значеннями.	3
3	Очищення та підготовка даних Видалення дублікатів. Обрізання та об'єднання даних. Використання функцій для обробки текстових даних.	2
4	Візуалізація даних з бібліотеками Matplotlib та Seaborn. Побудова різних типів графіків. Додавання осей, легенди, назв графіків. Використання Seaborn для стилізації графіків.	4
5	Статистичний аналіз даних. Обчислення середнього, медіани, стандартного відхилення. Кореляція між змінними.	4
	Разом	17

8. Самостійна робота

№ з/п	Назва теми/питання для підготовки, завдання	Кількість годин
1	Введення до аналізу даних	10
2	Огляд бібліотек мови Python для аналізу даних	12
3	Очищення та підготовка даних	12
4	Візуалізація даних Індивідуальна розрахункова робота	11
5	Статистичний аналіз даних	11
	Разом	56

Критерії оцінювання індивідуальної розрахункової роботи:

1. Критерії для оцінювання:

- своєчасність виконання;
- самостійність виконання (у разі доведеного плагіату бали за роботу анулюються);
- повнота дослідження даних;
- відповідність формальним критеріям (структура, послідовність, логічність, якість оформлення тощо).
- вміння застосовувати теоретичні знання для рішення практичних завдань.

9. Методи навчання

Підготовка студентів здійснюється на лекційних та лабораторних заняттях, але у значній мірі покладається на самостійне вивчення матеріалу студентами денної форми навчання протягом семестру.

Під час викладання дисципліни використовуються словесні та наочні методи навчання:

лекції, бесіда, пояснення, робота з літературними джерелами.

Під час проведення лекцій використовуються наступні методи навчання: пояснювально-ілюстративний метод, інформаційно-рецептивний; репродуктивний метод (репродукція - відтворення); метод проблемного викладу; частково-пошуковий метод.

Під час лабораторних та практичних занять використовуються наступні методи навчання частково-пошуковий, або евристичний метод, дослідницький; при захисті лабораторних робіт та індивідуальних завдань використовується дискусійний метод. Під час самостійної роботи використовується дослідницький метод (студент опановує літературу за вказаною темою, виконує індивідуальне розрахункове завдання та доповідь).

10. Методи контролю

Поточний та підсумковий контроль здійснюється в результаті виконання лабораторних робіт. Поточний контроль: опитування, виконання лабораторних робіт; тестові завдання. Форми оцінювання: усне опитування, перевірка лабораторної роботи, тестування.

11. Питання для підсумкового контролю

1. Основні завдання обробки даних.
2. Бібліотеки Python для аналізу даних.
3. DataFrame в бібліотеці Pandas.
4. Методи зчитування даних з CSV файлу в Pandas.
5. Обробка пропущених значень в Pandas.
6. Об'єднання двох DataFrame в Pandas.
7. Методи Python для статистичного аналізу даних.
8. Графік розподілу в Seaborn.
9. Кількісний та якісний аналіз даних.
10. Методи для видалення дублікатів у Pandas.
11. Підходи для очищення даних від шуму.
12. Які функції допомагають здійснювати фільтрацію даних в Pandas?
13. Бібліотека Plotly для створення інтерактивних графіків.
14. Jupyter Notebook для аналізу даних.
15. Бібліотека SciPy для статистичного аналізу даних.
16. Функції для обчислення статистичних показників в Python.
17. Типи графіків, які підтримує бібліотека Seaborn.
18. Розрахунок середнього, медіани та стандартного відхилення в Pandas.

19. Побудова кругової діаграми за допомогою бібліотеки Matplotlib.
20. Функції apply та map для обробки даних в Pandas.
21. Бібліотека seaborn для побудови стилізованих графіків.
22. Обробка категоріальних даних в Pandas.
23. Бібліотека SciPy для проведення статистичних тестів.
24. Методи в Pandas, які дозволяють вибирати підмножини даних.
25. Бібліотеку Plotly для створення 3D графіків.

12. Розподіл балів, які отримують студенти

Поточний та періодичний контроль					Індивідуальн а розрахункова робота	Сума балів
Змістовий модуль 1			Змістовий модуль 2			
T1	T2	T3	T4	T5	25	100
5	5	5	5	5		
Контрольна робота за змістовим модулем 1 - 25			Контрольна робота за змістовим модулем 2 - 25			

ФОРМУВАЛЬНЕ ОЦІНЮВАННЯ

Види навчальної роботи	Бали за одне заняття (завдання)	Кількість завдань	Сумарна кількість балів
ЗМІСТОВИЙ МОДУЛЬ 1. ФІЗИЧНІ ОСНОВИ КВАНТОВОЇ ІНФОРМАТИКИ			
Виконання і захист лабораторних робіт	5	3	0-15
Контрольна робота за змістовим модулем			0-25
Усього за змістовим модулем 1			0-40
ЗМІСТОВИЙ МОДУЛЬ 2. КВАНТОВІ РОЗРАХУНКИ			
Виконання і захист лабораторних робіт	5	2	0-10
Контрольна робота за змістовим модулем			0-25
Усього за змістовим модулем 2			0-35

Індивідуальна розрахункова робота			0-25
Підсумкова сума балів			0-100

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проєкту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
85-89	B	добре	
75-84	C		
70-74	D	задовільно	
60-69	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

При оцінюванні в балах рівня засвоєння матеріалу використовуються загальні критерії оцінювання навчальних досягнень здобувачів вищої освіти:

Оцінка за національно ю шкалою та відсоток від максималь ної кількості балів	Теоретична підготовка	Практична підготовка
	Здобувач освіти	
	зараховано (90-100% від максимальн ої кількості балів)	у повному обсязі володіє навчальним матеріалом, вільно, самостійно та аргументовано його викладає під час усних виступів та письмових відповідей; глибоко та всебічно розкриває зміст теоретичних питань, використовуючи при цьому нормативну, обов'язкову та додаткову літературу; робить самостійні висновки, виявляє причинно-наслідкові зв'язки; самостійно знаходить додаткову інформацію та використовує її для

	реалізації поставлених перед ним завдань. Здобувач здатен виділяти суттєві ознаки вивченого за допомогою операцій синтезу, аналізу, виявляти причинно-наслідкові зв'язки, формувати висновки і узагальнення, вільно оперувати фактами та відомостями.	оцінити результати власної практичної діяльності; виконує творчі завдання та ініціює нові шляхи їх виконання; вільно використовує набуті теоретичні знання при аналізі практичного матеріалу; проявляє творчий підхід до виконання індивідуальних та колективних завдань при самостійній роботі.
зараховано (75-89% від максимальної кількості балів)	достатньо повно володіє навчальним матеріалом, обґрунтовано його викладає під час усних виступів та письмових відповідей, використовуючи при цьому нормативну та обов'язкову літературу; при представленні деяких питань не вистачає достатньої глибини та аргументації, застосовує знання для розв'язання стандартних ситуацій; самостійно аналізує, узагальнює і систематизує навчальну інформацію, але допускаються при цьому окремі несуттєві неточності та незначні помилки.	правильно вирішив більшість розрахункових /тестових завдань за зразком; має стійкі навички виконання завдання
зараховано (60-74% від максимальної кількості балів)	володіє навчальним матеріалом на репродуктивному рівні або відтворює певну частину навчального матеріалу з елементами логічних зв'язків, знає основні поняття навчального матеріалу; має ускладнення під час виділення суттєвих ознак вивченого; під час виявлення причинно-наслідкових зв'язків і формулювання висновків.	може використовувати знання в стандартних ситуаціях, має елементарні, нестійкі навички виконання завдання. Правильно вирішив половину розрахункових/тестових завдань. Здобувач має ускладнення під час виділення суттєвих ознак вивченого; під час виявлення причинно-наслідкових зв'язків і формулювання висновків.
не зараховано (35-59% від максимальної кількості балів)	володіє навчальним матеріалом поверхово й фрагментарно (без аргументації та обґрунтування); безсистемно виокремлює випадкові ознаки вивченого; не вмє робити найпростіші операції аналізу і синтезу; робити узагальнення, висновки; під час відповіді допускаються суттєві помилки	недостатньо розкриває сутність практичних завдань, допускаючи при цьому суттєві неточності, правильно вирішив окремі розрахункові/тестові завдання за допомогою викладача, відсутні

		сформовані уміння та навички.
не - зараховано (0-34% від максимальної кількості балів)	не володіє навчальним матеріалом	виконує лише елементи завдання, потребує постійної допомоги викладача

13. Методичне забезпечення

Робоча програма навчальної дисципліни; силабус, мультимедійні презентації, конспект лекцій

14. Рекомендована література

Основна

1. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. – Київ. – 2017. – 150 с.
2. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. – К.: ІСЗІ КПІ ім. Ігоря Сікорського», 2018. — 297 с.
3. Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. - O'Reilly Media – 2022. – 579p.
4. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних / Я. В. Іванчук, В. І. Месюра, А. А. Яровий, О. Д. Манжілевський – Вінниця : ВНТУ, 2021. – 69 с.

Додаткова

5. Лесна Н. С., Репка В. Б., Шатовська Т. Б. Інтелектуальний аналіз даних / Лесна Н. С., Репка В. Б., Шатовська Т. Б.. -Харків.: ХНУРЕ, 2003. - 110 с.
6. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних -К...: КНЕУ, 2007. -375 с.
7. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник. К., 2014. 599 с
8. Гладун Ф.Я., Рогушина Ю.В. Data Mining: пошук знань в даних: підручник. Київ: ТОВ «ВД «АДЕФ-Україна», 2016. 452 с.
9. Путівник мовою програмування Python. <https://pythonguide.rozh2sch.org.ua/>

15. Електронні інформаційні ресурси

1. <http://nbuy.gov.ua/> - Сайт Національної бібліотеки України імені

В. І. Вернадського;

2. <http://www.dnpb.gov.ua/> - Сайт Державної науково-педагогічної бібліотеки України імені В.О. Сухомлинського;

3. <http://onu.edu.ua/>- Сайт бібліотеки ОНУ імені І.І. Мечникова;

4. <http://odnb.odessa.ua/> - Сайт Одеської національної наукової бібліотеки.