

Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра комп'ютерних систем та технологій

Силабус курсу

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

Обсяг	Загальна кількість: кредитів - 3; годин - 90
Семестр, рік навчання	5 семестр, 3 рік
Дні, час, місце	За розкладом
Викладач (-і)	Коренкова Ганна Валентинівна
Контактний телефон	+380679515783
E-mail	av.korenkova@gmail.com
Робоче місце	Кафедра комп'ютерних систем та технологій ОНУ імені І.І. Мечникова, вул. Дворянська, 2
Консультації	Viber, Zoom, Google Class

КОМУНІКАЦІЯ

Комунікація зі студентами буде здійснюватися аудиторно, в месенджері Viber, Zoom, Google Classroom

АНОТАЦІЯ КУРСУ

Предмет вивчення курсу – методи та засоби аналізу даних та їх реалізація на мові Python.

Пререквізити курсу: Для вивчення курсу студенти потребують базових знань з дисциплін «Вища математика», «Дискретна математика», «Теорія ймовірності та математична статистика», «Програмування».

Метою є отримання навичок ефективної обробки, аналізу та візуалізації даних з використанням мови програмування Python.

Завдання дисципліни “Обробка та аналізу даних засобами Python” є сформуванню у студентів уявлення про:

- основи python та робота з даними;
- обробка даних засобами pandas;
- візуалізація даних;
- очищення та попередня обробка даних;
- сформувані уявлення про важливість чистоти та якості даних для успішного аналізу.

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.

ОПИС КУРСУ

Форми і методи навчання

Курс буде викладений у формі лекцій (17 год.) та лабораторних занять (17 год.), організації самостійної роботи студентів (56 год.).

Основна 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; практичні методи навчання - виконання лабораторних робіт, робота з літературними джерелами.

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.

Перелік рекомендованої літератури

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., Рєпка В. Б., Шатовська Т. Б. Інтелектуальний аналіз даних / Лесна Н. С., Рєпка 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://nbuv.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 1.1. Мечникова;
4. <http://odnb.odessa.ua/> - Сайт Одеської національної наукової бібліотеки.

ОЦІНЮВАННЯ

Методи поточного контролю: виконання завдань лабораторних робіт, контрольні роботи.

Форми і методи підсумкового контролю: іспит

Поточний та періодичний контроль					Індивідуальна 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

Самостійна робота студентів.

Самостійна робота представлена у формі підготовки до лекцій та лабораторних занять та індивідуальне самостійне завдання. Підготовка до лекцій перевіряється з використанням тестових завдань та поточного опитування на лекції. Звіт з лабораторної роботи студенти здають у письмовій формі на протязі 7 днів у Google Class. Кожна лабораторна робота оцінюється в 5 бал.

ПОЛІТИКА КУРСУ

Викладання дисципліни відбувається відповідно до Положення про організацію освітнього процесу в Одеському національному університеті імені І.І. Мечникова.

Політика щодо дедлайнів та перескладання: роботи мають здаватися вчасно. Деякі види робіт можуть здаватися із порушенням термінів з поважних причин. Перескладання модулів відбувається за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: дотримання академічної доброчесності здобувачами освіти передбачає дії відповідно до Положення про запобігання та виявлення академічного плагіату у освітній та науково-дослідній роботі учасників освітнього процесу та науковців Одеського національного університету імені І.І. Мечникова.

Політика щодо відвідування: відвідування лабораторних занять є обов'язковим та без запізень. За об'єктивних причин (наприклад, хвороба, міжнародне стажування, карантин тощо) навчання може відбуватись в онлайн формі (змішана форма навчання) на платформі із використанням інструментів Google WorkSpace та Classroom.

Мобільні пристрої дозволяється використовувати лише під час онлайн тестування та підготовки практичних завдань у процесі заняття.

Поведінка в аудиторії повинна відповідати загальним вимогам ділової та наукової етики: неприпустимо під час відповідей на занятті користуватися мобільними телефонами, порушувати дисципліну розмовами або в будь-який інший спосіб.