



Силабус освітнього компонента Програма навчальної дисципліни



Обробка даних засобами Python

Шифр та назва спеціальності

122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Комп'ютерні науки

Кафедра

Системного аналізу та інформаційно-аналітичних технологій

Рівень освіти

Бакалавр

Тип дисципліни

Профільна підготовка, Вибіркова

Семестр

7

Мова викладання

Українська

Викладачі, розробники



Коваленко Сергій Володимирович

Serhii.Kovalenko@khai.edu.ua

Кандидат технічних наук, доцент, доцент кафедри системного аналізу та інформаційно-аналітичних технологій НТУ «ХПІ»

Досвід роботи – 27 років. Автор понад 100 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Проектування баз даних», «Проектування інформаційних систем», «Бази даних», «Комп'ютерна обробка зображень», «Обробка зображень та мультимедіа», «Основи комп'ютерної графіки», «Обробка даних засобами Python».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на ознайомлення студентів з основними поняттями, методами та засобами попередньої підготовки, статистичної обробки, візуалізації та аналізу даних, основами машинного навчання. На лабораторних заняттях студенти зможуть використовувати основні бібліотеки мови Python для роботи з даними (NumPy, SciPy, Matplotlib, Scikit-learn, Pandas тощо).

Мета та цілі дисципліни

Мета викладання дисципліни полягає у формуванні у студентів теоретичних знань і практичних здатностей застосовувати загальні методи і засоби підготовки, статистичної обробки, візуалізації та аналізу даних, використовуючи основні бібліотеки мови Python для роботи з даними, надання необхідного обсягу знань із практичного застосування мови Python та механізмів інтелектуального аналізу для вирішення прикладних задач обробки даних.

Формат занять

Лекції, лабораторні роботи, розрахункове завдання, самостійна робота, консультації. Підсумковий контроль – екзамен.

Компетентності

ВКП 2.2 Здатність використовувати прикладні математичні методи та програмні засоби для аналізу та синтезу інформаційно-аналітичних систем.

ВКП 2.3 Здатність використовувати математичні методи та програмні засоби в системах діагностики стану, розпізнавання образів, планування та прогнозування поведінки складних систем.

Результати навчання

РНП 2.2 Володіти методами графічного подання інформації, вміти розробляти та використовувати методи аналізу даних та розпізнавання образів при обробці статичної та динамічної графічної інформації.

РНП 2.3 Володіти навичками розробки програмного забезпечення для вирішення задач аналізу даних, методами та засобами оцінювання якості та тестування програмного забезпечення.

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5 кредитів ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з наступних дисциплін: "Аналіз даних", "Теорія ймовірностей та математична статистика".

Особливості дисципліни, методи та технології навчання

Лекції проводяться інтерактивно з використанням мультимедійних технологій. На лабораторних заняттях використовується проєктний підхід до навчання, акцентується увага на застосуванні інформаційних технологій. Навчальні матеріали доступні студентам.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Вступ. Введення до вивчення дисципліни. Предмет дисципліни. Мета викладання дисципліни. Завдання дисципліни Основні поняття обробки даних

Структура дисципліни. Поняття даних, основні завдання обробки даних, особливості обробки даних. Класифікація та загальний огляд етапів та методів обробки даних.

Тема 2. Елементи мови Python

Складові елементи python-програми. Структура файлу і кодування програми. Виконання Python-програм. Поняття складних типів даних. Класифікація типів даних.

Тема 3. Бібліотека NumPy. Створення масивів NumPy, арифметичні операції над масивами NumPy
Створення масивів NumPy. Арифметичні операції над масивами NumPy. Індексція та агрегування масивів в NumPy.

Тема 4. Бібліотека NumPy. Арифметичні операції над матрицями, індексція, агрегування, транспонування і зміна форми матриць

Створення матриць NumPy на прикладах. Арифметичні операції над матрицями, індексція, агрегування, транспонування і зміна форми матриць. Приклади обробки даних за допомогою NumPy. Часові ряди в NumPy.

Тема 5. Статистична обробка даних засобами Python

Виявлення статистичних характеристик масивів даних засобами NumPy. Дослідження та моделювання експериментальних даних.

Тема 6. Попередня обробка даних в бібліотеці Pandas. Інструменти завантаження даних в бібліотеці Pandas, засоби підготовки даних до обробки

Інструменти завантаження даних в бібліотеці Pandas. Засоби підготовки даних до обробки. Об'єднання та зміна форми даних.

Тема 7. Обробка даних в бібліотеці Pandas. Об'єкти Series, DataFrame, Index. Універсальні функції: збереження індексу, вирівнювання індексів, виконання операцій між об'єктами
Об'єкти Series, DataFrame, Index. Універсальні функції: збереження індексу, вирівнювання індексів, виконання операцій між об'єктами. Індексція та вибірка даних. Операції над даними в бібліотеці Pandas. Обробка відсутніх даних. Ієрархічна індексція.

Тема 8. Обробка даних в бібліотеці Pandas. Робота з часовими рядами

Об'єднання наборів даних: конкатенація і додавання в кінець. Агрегування і угруповання. Зведені таблиці. Векторизовані операції над рядками. Робота з часовими рядами.

Тема 9. Візуальне представлення даних. Основні види та способи представлення даних, їхні особливості, переваги і недоліки. Вибір засобів та інструментів представлення даних

Основні види та способи представлення даних, їхні особливості, переваги і недоліки. Вибір засобів та інструментів представлення даних.

Тема 10. Візуалізація за допомогою бібліотеки Matplotlib. Прості лінійні графіки. Налаштування графіки

Прості лінійні графіки. Налаштування графіки. Прості діаграми розсіювання. Візуалізація похибок. Графіки щільності і контурні графіки. Гістограми, розбиття по інтервалах і щільність.

Тема 11. Візуалізація за допомогою бібліотеки Matplotlib. Призначені для користувача настройки Matplotlib: конфігурації і таблиці стилів. Таблиці стилів

Множинні субграфіки. Стрілки і пояснювальні написи. Призначені для користувача настройки Matplotlib: конфігурації і таблиці стилів. Таблиці стилів. Побудова тривимірних графіків в бібліотеці Matplotlib.

Тема 12. Основи інтелектуального аналізу даних. Класифікація методів та алгоритмів інтелектуального аналізу даних

Класифікація методів та алгоритмів інтелектуального аналізу даних. OLAP, Data Mining, Text Mining, машинне навчання. Особливості їх застосування.

Тема 13. Елементи машинного навчання. Категорії машинного навчання. Приклади прикладних задач машинного навчання

Категорії машинного навчання. Приклади прикладних задач машинного навчання. Класифікація: передбачення дискретних міток. Знайомство з бібліотекою Scikit-Learn. Представлення даних в Scikit-Learn. Огляд популярних методів машинного навчання.

Тема 14. Методи машинного навчання без вчителя. Типи машинного навчання без вчителя.

Проблеми машинного навчання без вчителя

Типи машинного навчання без вчителя. Проблеми машинного навчання без вчителя. Зниження розмірності та виділення признаков. Кластеризація.

Тема 15. Методи машинного навчання з вчителем. Алгоритми машинного навчання з вчителем

Алгоритми машинного навчання з вчителем. Оцінки невизначеності для класифікаторів. Навчальна здатність, перенавчання та недонавчання.

Тема 16. Робота з текстовими даними

Строкові типи даних. Токенізація, стемінг та лематизація. Моделювання тем та кластеризація документів.

Теми практичних занять

Практичні роботи в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Основи IPython. Виконання оболонки IPython. Виконання Jupyter Notebook.

Тема 2. Огляд мови розмітки markdown. Експорт даних і звітів з Jupyter Notebook.

Тема 3. Робота з NumPy ndarray (об'єкт багатовимірного масиву).

Тема 4. Арифметичні операції з масивами NumPy.

Тема 5. Виявлення статистичних характеристик масивів даних засобами NumPy.

Тема 6. Інструменти завантаження даних в бібліотеці Pandas.

Тема 7. Робота з об'єктами Series та DataFrame.

Тема 8. Арифметичні операції та вирівнювання даних.

Тема 9. Візуальне представлення даних.

Тема 10. Побудова графіків за допомогою Pandas та Seaborn.

Тема 11. Побудова графіків за допомогою Matplotlib. Налаштування графіки.

- Тема 12. Методи класифікації даних. Методи кластерного аналізу.
Тема 13. Робота з бібліотекою Scikit-Learn. Представлення даних в Scikit-Learn.
Тема 14. Методи машинного навчання без вчителя.
Тема 15. Методи машинного навчання з вчителем.
Тема 16. Робота з текстовими даними.

Самостійна робота

Курс передбачає виконання розрахункового завдання. Результат моделювання та програмування оформлюється у письмовий звіт. Студентам також рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення та аналізу.

Література та навчальні матеріали

Основна література

1. J. Hunt: Advanced Guide to Python 3 Programming. // Springer, 2019. 494 p.
2. Кренивч А.П. Python у прикладах і задачах. Частина 2. Об'єктно-орієнтоване програмування. Навчальний посібник – К.: ВПЦ "Київський Університет", 2020. – 152 с.
3. Костюченко А. О. Основи програмування мовою Python: навчальний посібник. Ч.: ФОП Балакіна С.М., 2020. 180 с.
4. Копей В. Б. Мова програмування Python для інженерів і науковців : навчальний посібник / В. Б. Копей – ІваноФранківськ : ІФНТУНГ, 2019. – 272 с.
5. Ланде, Д. В. Оброблення надвеликих масивів даних (Big Data) [Електронний ресурс] : навчальний посібник для використання у навчальному процесі зі спеціальності 122 «Комп'ютерні науки» / Д. В. Ланде, І. Ю. Субач, А. Я. Гладун ; КПП ім. Ігоря Сікорського. – Електронні текстові дані. – Київ : КПП ім. Ігоря Сікорського, 2021. – 168 с.
6. 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.
7. Jesús Rogel-Salazar Advanced Data Science and Analytics with Python, CRC Press, 2020, 424 p.

Додаткова література

8. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою Pandas" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз / уклад.: С. М. Коваленко, С. В. Коваленко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2022. – 44 с.
9. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою Matplotlib" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 28 с.
10. Методичні вказівки до лабораторної роботи "Основи роботи з бібліотекою NumPy" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 48 с.
11. Методичні вказівки до лабораторної роботи "Основи роботи в середовищі Jupyter Notebook" з курсу "Обробка даних Python" [Електронний ресурс] : для студентів спец. 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 124 Системний аналіз, 126 Інформаційні системи і технології / уклад.: С. М. Коваленко, С. В. Коваленко, О. В. Шматко ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2021. – 28 с.
12. Путівник мовою програмування Python [Електронний ресурс]. – Режим доступу : <https://pythonguide.rozh2sch.org.ua>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді:
лабораторні роботи: 30% семестрової оцінки;
контрольні роботи: 20% семестрової оцінки;
розрахункове завдання: 30% семестрової оцінки;
екзамен: 20% семестрової оцінки.
Екзамен: письмове завдання (2 запитання з теорії + розв'язання 2 практичних завдань) та усна доповідь. Поточне оцінювання: 2 контрольні роботи та розрахункове завдання.

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

27.08.2023

Завідувач кафедри
Юрій ДОРОФЄЄВ

27.08.2023

Гарант ОП
Олена ЛОБАЧ