



Силабус освітнього компонента Програма навчальної дисципліни



Обробка даних засобами Python

Шифр та назва спеціальності

124 – Системний аналіз

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Системний аналіз і управління

Кафедра

Системного аналізу та інформаційно-аналітичних технологій

Рівень освіти

Магістр

Тип дисципліни

Спеціальна (фахова), Обов'язкова

Семестр

1

Мова викладання

Українська

Викладачі, розробники



Коваленко Сергій Володимирович

Serhii.Kovalenko@khp.edu.ua

Кандидат технічних наук, доцент, професор 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», «Обробка текстової та графічної інформації», «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).

Мета та цілі дисципліни

Мета викладання дисципліни полягає в 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

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність спілкуватися іноземною мовою.

ЗК3. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК4. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності).

ЗК5. Здатність розробляти проєкти та управляти ними.

СК1. Здатність інтегрувати знання та здійснювати системні дослідження, застосовувати методи математичного та інформаційного моделювання складних систем та процесів різної природи.

СК3. Здатність розробляти системи підтримки прийняття рішень та рекомендаційні системи.

СК6. Здатність застосовувати теорію і методи Data Science для здійснення інтелектуального аналізу даних з метою виявлення нових властивостей та генерації нових знань про складні системи.

СК7. Здатність управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів.

СК8. Здатність розробляти і реалізовувати наукові та прикладні проєкти в галузі інформаційних технологій та дотичні до неї міждисциплінарні проєкти.

СК11. Вміння використовувати моделі та методи Data Mining для розв'язання задач інтелектуального аналізу даних.

Результати навчання

РН1. Спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері системного аналізу та інформаційних технологій і є основою для оригінального мислення та проведення досліджень.

РН2. Будувати та досліджувати моделі складних систем і процесів застосовуючи методи системного аналізу, математичного, комп'ютерного та інформаційного моделювання.

РН3. Застосовувати методи розкриття невизначеностей в задачах системного аналізу, розкривати ситуаційні невизначеності та невизначеності в задачах взаємодії, протидії та конфлікту стратегій, знаходити компроміс при розкритті концептуальної невизначеності.

РН4. Розробляти та застосовувати методи, алгоритми та інструменти прогнозування розвитку складних систем і процесів різної природи.

РН5. Використовувати міри оцінювання ризиків та застосовувати їх при аналізі багатofакторних ризиків в складних системах.

РН6. Застосовувати методи машинного навчання та інтелектуального аналізу даних, математичний апарат нечіткої логіки, теорії ігор та розподіленого штучного інтелекту для розв'язання складних задач системного аналізу.

РН7. Розробляти інтелектуальні системи в умовах слабо структурованих даних різної природи.

РН9. Розробляти та застосовувати моделі, методи та алгоритми прийняття рішень в умовах конфлікту, нечіткої інформації, невизначеності та ризиків.

РН10. Зрозуміло і недвозначно доносити власні знання, висновки та аргументацію до фахівців і нефахівців, зокрема до осіб, які навчаються.

РН11. Вільно презентувати та обговорювати усно і письмово результати досліджень та інновацій, інші питання професійної діяльності державною та англійською мовами.

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)

Передумови навчання – немає.

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. Florent Buisson. (2021). Behavioral Data Analysis with R and Python. USA: O'Reilly, 336 p.

(<https://dokumen.pub/qdownload/behavioral-data-analysis-with-r-and-python-customer-driven-data-for-real-business-results-9781492061373-1492061379.html>)

2. Jake VanderPlas. (2023). Python Data Science Handbook. 2-nd Edition USA: O'Reilly Media, 564 p.

(<https://dokumen.pub/qdownload/python-data-science-handbook-essential-tools-for-working-with-data-2nbsped-1098121228-9781098121228.html>)

3. 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.

(<https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/36454/120997.pdf?sequence=2&isAllowed=y>)

4. 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. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/61382>
5. 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. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55469>
6. 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. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55468>
7. 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. – URI: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/55466>
8. Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python. Частина 1 / Б. І. Мокін, В. Б. Мокін, О. Б. Мокін. – Вінниця : ВНТУ, 2022. – 124 с. (<https://iq.vntu.edu.ua/repository/getfile.php/5097.pdf>)
9. Костюченко А. О. Основи програмування мовою Python: навчальний посібник. Ч.: ФОП Балакіна С. М., 2020. 180 с.
(<https://epub.chnpu.edu.ua/jspui/bitstream/123456789/5584/1/%D0%9E%D1%81%D0%BD%D0%BE%D0%B2%D0%B8%20%D0%BF%D1%80%D0%BE%D0%B3%D1%80%D0%B0%D0%BC%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F%20%D0%BC%D0%BE%D0%B2%D0%BE%D1%8E%20Python.pdf>)

Додаткова література

10. Попередня обробка та аналіз даних: лабораторний практикум для студ. спеціальності 113 «Прикладна математика» /Уклад.: Н. Е. Кондрук. Ужгород: УжНУ, 2023. – 41 с.
(<https://dspace.uzhnu.edu.ua/jspui/bitstream/lib/53297/1/%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%D1%87%D0%BA%D0%B0%20%283%29.pdf>)
11. NumPy. [Electronic resource]. Access mode: <https://numpy.org/>
12. The Jupyter Notebook. [Electronic resource]. Access mode: <https://jupyter.org/>
13. Pandas. [Electronic resource]. Access mode: <https://pandas.pydata.org/>
14. Matplotlib: Visualization with Python. [Electronic resource]. Access mode: <https://matplotlib.org/>
15. Seaborn. [Electronic resource]. Access mode: <https://seaborn.pydata.org/>
16. Scikit-learn Machine Learning in Python. [Electronic resource]. Access mode: <https://scikitlearn.org/stable/>

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

26.08.2024

Завідувач кафедри
Юрій ДОРОФЄЄВ

26.08.2024

Гарант ОП
Валерій СЕВЕРИН