



Силабус освітнього компонента Програма навчальної дисципліни



Збір, обробка та аналіз даних на Python та R

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерного моделювання, прикладної
фізики та математики

Освітня програма
Прикладна математика. Комп'ютерне та
математичне моделювання

Кафедра
Математичне моделювання та інтелектуальні
обчислення в інженерії (161)

Рівень освіти
Бакалавр

Тип дисципліни
Вибіркова (профільна)

Семестр
5

Мова викладання
Українська

Викладачі, розробники



Розова Людмила Вікторівна

Lyudmyla.Rozova@khi.edu.ua

Кандидат технічних наук, доцент, досвід роботи – 20 років. Автор понад 40 наукових та методичних праць. Провідний лектор з дисциплін: «Збір, обробка та аналіз даних на Python та R», «Об'єктно-орієнтоване програмування», «Object Oriented Programming», викладач практики
Orcid: <https://orcid.org/0000-0002-0781-7473>

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на вивчення підходів для роботи з даними за допомогою бібліотек мови програмування Python та R. Дисципліна викладається у 5му семестрі та передбачає: 32 годин лекцій, 16 години лабораторних занять, 72 годин самостійної роботи. Підсумковий контроль – залік.

Мета та цілі дисципліни

Дисципліна спрямована на засвоєння майбутніми фахівцями знань та набуття умінь та навичок для підготовки, відбору, обробки, виявлення закономірностей, аналізу та інтерпретації різного типу даних з застосуванням алгоритмів, реалізованих в функціях бібліотеках мов програмування Python та R та сформуванню навички практичного їх застосування при роботі з даними.

Формат занять

Лекції, лабораторні роботи. Підсумковий контроль - залік

Компетентності

ЗК10. Навички у використанні інформаційних і комунікаційних технологій.
ФК05. Здатність проектувати бази даних, інформаційні системи та ресурси.

ФК09. Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.

ФК14. Здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату.

Результати навчання

РН06. Володіти основними методами розробки дискретних і неперервних математичних моделей об'єктів та процесів, аналітичного дослідження цих моделей на предмет існування та єдиності їх розв'язку.

РН10. Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних.

РН11. Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.

РН13. Використовувати в практичній роботі спеціалізовані програмні продукти та програмні системи комп'ютерної математики.

РН19. Збирати та інтерпретувати відповідні дані й аналізувати складності в межах своєї спеціалізації для донесення суджень, які відбивають відповідні соціальні та етичні проблеми.

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредита ECTS): лекції – 32 год., лабораторні роботи – 16 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Вступ до спеціальності. Ознайомча практика, Алгоритмізація та програмування, Об'єктно-орієнтоване програмування, Організація баз даних, Математична логіка, теорія алгоритмів та структура даних.

Особливості дисципліни, методи та технології навчання

Заняття проводяться інтерактивно, з використанням мультимедійних технологій. Для опанування практичних навичок застосовується наступні вільні програмні середовища: Anaconda Jupyter Notebook для мови Python та R-studio для мови R. Студент зобов'язаний відвідувати всі заняття згідно розкладу, виконувати лабораторні роботи. Дотримуватися етики поведінки. З метою оволодіння необхідною якістю освіти з дисципліни потрібно відвідуваність і регулярна підготовленість до занять.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Введення у аналіз даних

Для чого необхідно обробляти дані для аналізу? Поняття науки о даних. Основні етапи аналізу даних. Задачі спеціаліста дата-аналітика. Існуючі мови програмування для аналізу даних. Їх порівняння.

Тема 2. Мова програмування Python, як одна із поширених мов, що застосовується для аналізу даних.

Мультипарадигмальна мова програмування Python. Введення, повторення. Особливості мови програмування Python: синтаксис, основні типи даних, введення-виведення даних, оператори, умови, цикли, робота з послідовностями даних, кортежи, списки, множини, словники.

Тема 3. Інформаційна грамотність. Використання мови Python для дослідження даних.

Поняття інформативної грамотності. Застосування знань з математики та статистики для обробки даних. Виявлення закономірностей в даних. Бібліотеки мови Python для дослідження даних. Особливості бібліотеки NumPy для роботи з масивами даних.

Тема 4. Робота з наборами даних. Бібліотека Python Pandas .

Датасети та датафрейми. Особливості роботи з функціями бібліотеки Pandas по створенню та обробці структур даних. Підготовка даних для аналізу.

Тема 5. Візуалізація даних за допомогою Python Matplotlib.

Особливості бібліотеки Matplotlib для візуалізації відібраних даних. Основні функції бібліотеки, їх особливості застосування. Інтерпретація візуального представлення даних.

Тема 6. Мова програмування R для аналізу даних.

Особливості мови R, її застосування в науці о даних. Знайомство з синтаксисом мови, операторами. Середовище програмування R studio.

Тема 7. Робота зі структурами даних за допомогою мови R.

Маніпулювання даними в R. Робота з масивами даних, Застосування матричних операцій. Перетворення масивів. Робота з датафреймами, перетворення, відбір даних. Інтерпретування відібраних даних.

Тема 8. Візуалізація даних за допомогою мови R.

Можливості візуалізації даних з використанням мови R. Побудова різних типів діаграм, гістограмм, Інтерпретація візуалізації відібраних даних.

Теми лабораторних робіт

Тема 1. Основи програмування за допомогою мови Python.

Завантаження, установка та налагодження необхідного програмного середовища. Створення простих програм на мові Python, повторення роботи з алгоритмами та даними.

Тема 2. Робота з масивами в Python NumPy.

Бібліотека NumPy. Створення, заповнення та робота з масивами даних різного розміру. Матричні операції.

Тема 3. Робота зі структурами даних в Python Pandas.

Створення датафреймів. Зчитування та запис даних у файли типу .txt, .csv, та excel. Функції відбору, обробки та аналізу даних з застосуванням статистичних величин. Створення зведених таблиць даних.

Тема 4. Візуалізація даних за допомогою бібліотеки Python Matplotlib.

Підключення бібліотеки Matplotlib, практичне застосування функцій для побудови діаграм, графіків, їх параметрів. Створення декількох візуалізацій одночасно. Збереження та виведення графічного результату. Інтерпретація даних.

Тема 5. Приклад роботи над проектом з використанням мови Python.

Приклад застосування вивченого матеріалу для аналізу даних на прикладі існуючих датасетів з наборами даних. Датасет Титанік. Виконання індивідуальних завдань.

Тема 6. Мова програмування R для аналізу даних.

Завантаження, установка та налагодження роботи програмного середовища R studio . Знайомство з синтаксисом, типами даних, операторами мови R . Написання простих програм. Завантаження датасетів. Основні функції по обробці структур даних в R. Застосування функцій по відбору, перетворенню та аналізу наборів даних.

Тема 7. Візуалізація даних за допомогою мови R.

Функції побудови діаграм, гістограм, графіків, скатерплотів даних. Їх особливості, параметри.

Тема 8. Приклад роботи над проектом з використанням мови R.

Приклад застосування вивчених функцій і підходів на прикладі розбору датасетів з наборами даних. Виконання індивідуальних завдань.

Самостійна робота складається з наступних компонентів

Опрацювання лекційного матеріалу. Підготовка до лабораторних занять. Самостійне вивчення тем та питань, які не викладаються на лекційних заняттях. Виконання індивідуальних завдань.

Література та навчальні матеріали

- 1.°Бен С. Python. Збірник вправ. Введення в мову Python із завданнями та рішеннями. – Харків, 2021. – 23 8 с.
- 2.°Беррі Пол. Head First.- К. “Фабула”, 2021. – 624 с.
- 3.°Висоцька В., Оборська О. Python. Алгоритмізація та програмування.- Н.Світ 2000. – 2021.-514 с.
- 4.°Маттес Е. Пришвидшений курс Python. – Львів.: Видавництво Старого Лева, 2021. – 600 с.
- 5.°Яковенко А. В. Основи програмування. Python. Частина 1. – Київ: КІШ ІМ. Ігоря Сікорського, 2018.- 195 с.
- 6.°Gutttag John V. Introduction to Computation and Programming Using Python with Application to Understanding Data.- 2016.- 591 p.
- 7.°Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
8. Jenine K. Harris, Statistics With R. Washington University in St.Louis, USA, 2020.
9. Sarah Stowell. Using R for Statistics. Apress, 2014.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Бали за залік нараховуються за рейтингом:
Виконання лабораторних робіт – 40 балів
Виконання індивідуальних завдань – 50 балів
Теоретичне опитування – 10 балів

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

Дата погодження, підпис

Завідувач кафедри
Олексій ВОДКА

Дата погодження, підпис

Гарант ОП
Геннадій ЛЬВОВ