



Силабус освітнього компонента

Програма навчальної дисципліни



Обробка та аналіз даних на Python та R

Шифр та назва спеціальності

122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерного моделювання, прикладної фізики та математики

Освітня програма

Комп'ютерні науки. Моделювання, проектування та комп'ютерна графіка

Кафедра

Математичне моделювання та інтелектуальні обчислення в інженерії (161)

Рівень освіти

Бакалавр

Тип дисципліни

Профільований пакет 2, Вибіркова

Семестр

5

Мова викладання

Українська

Викладачі, розробники



Розова Людмила Вікторівна

Lyudmyla.Rozova@khpi.edu.ua

Кандидат технічних наук, доцент

Фахівець в галузі розробки наукоємного програмного забезпечення, математичного моделювання, обробки та аналізу даних, організації проектного навчання. Автор понад 40 наукових та методичних праць.

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна зосереджується на вивченні ключових інструментів та методів обробки даних, включаючи використання Python та R для виконання розширеного аналізу даних, візуалізації та моделювання, що надає студентам необхідні навички для розв'язання складних завдань у сфері аналітики даних. Ця дисципліна допомагає освоїти практичні навички роботи з великими обсягами даних, розроблювати та застосовувати алгоритми машинного навчання та статистичні методи для вивчення та використання інформації, закладаючи основи для успішної кар'єри в області аналітики.

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 навичок практичного їх застосування при роботі з даними.

Формат занять

Лекції, лабораторні роботи, самостійна робота. Підсумковий контроль - залік

Компетентності

ЗК2: Здатність застосовувати знання у практичних ситуаціях.

ЗК9: Здатність працювати в команді.

ЗК12: Здатність оцінювати та забезпечувати якість виконуваних робіт.

СК3: Здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем

СК7: Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів

СК9: Здатність реалізувати багаторівневу обчислювальну модель на основі архітектури клієнт-сервер, включаючи бази даних, знань і сховища даних, виконувати розподілену обробку великих наборів даних на кластерах стандартних серверів для забезпечення обчислювальних потреб користувачів, у тому числі на хмарних сервісах

СК11: Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач

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

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

Обсяг дисципліни

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

Передумови вивчення дисципліни (пререквізити)

Вступ до спеціальності. Ознайомча практика, Алгоритмізація та програмування, Архітектура обчислювальних систем, 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, кортежі, 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, скатерплових 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

Самостійна робота включає опрацювання лекційного матеріалу; підготовку до лабораторних занять; самостійне вивчення тем та питань, які не викладаються на лекційних заняттях; виконання індивідуального завдання.

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.°Guttag 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

28.08.2023

Завідувач кафедри
Олексій ВОДКА

28.08.2023

Гарант ОП
Оксана ТАТАРІНОВА