

АНАЛІТИЧНІ СИСТЕМИ BIG-DATA



КОНТАКТИ

Факультет менеджменту
Кафедра економічної
кібернетики, комп'ютерних наук
та інформаційних технологій

[https://www.mnau.edu.ua/
faculty-men/kaf-econ-kiber/](https://www.mnau.edu.ua/faculty-men/kaf-econ-kiber/)

РОЗРОБНИК

Жебко Олександр Олегович
викладач

[https://www.mnau.edu.ua/
faculty-men/kaf-econ-kiber/
zhebko/](https://www.mnau.edu.ua/faculty-men/kaf-econ-kiber/zhebko/)

ОПИС ДИСЦИПЛІНИ

- Семестр – 8
Кількість кредитів – 4
Кількість змістових модулів – 2
Загальна кількість годин - 120
- Лекції – 26 годин
 - Практичні заняття – 26 годин
 - Самостійна робота – 68 годин

ПРАКТИЧНА КОРИСТЬ

- База для роботи з сучасною екосистемою Big Data
- Підготовка до масштабного статистичного аналізу мовою R
- Розуміння та розробка розподілених пайплайнів мовою Python
- Вміння масштабувати алгоритми машинного навчання (Machine Learning)
- Перші навички роботи на стику Data Science та Data Engineering



АНОТАЦІЯ

Курс призначений для засвоєння студентами методологічних та концептуальних теоретичних відомостей про основні поняття аналітики великих даних та основні технології, що застосовуються для зберігання і пошуку в великих даних, а також підготовка кадрів вищої кваліфікації, які мають практичний досвід застосування методів аналізу великих даних та реалізації програм для аналітики великих даних, отримання досвіду у застосуванні сучасних методів роботи з Big-Data у вибраній галузі досліджень, а також у професійній діяльності. Зміст дисципліни охоплює коло питань, пов'язаних з технологіями зберігання і обробки великих даних, статистичними методами аналізу даних, сучасними програмні засоби аналізу великих обсягів інформації, збором та зберіганням великих даних, методами обробки і аналізу великих даних та візуалізацією вихідної інформації і аналітичних даних.



МЕТА

Формування у здобувачів вищої освіти теоретичних знань і практичних навичок щодо роботи з великими даними, проектування аналітичних систем та застосування сучасних інструментів Big Data для зберігання, обробки та аналізу масштабних інформаційних потоків. Вивчення дисципліни спрямоване на опанування архітектури розподілених систем, принципів паралельних обчислень, методів потокової і пакетної обробки, а також інструментів аналізу великих масивів даних у реальному часі.



ЗМІСТ ДИСЦИПЛІНИ

Модуль 1

2 - кредити

Технології та системи обробки великих даних мовою програмування R

- Джерела великих даних. Інтернет Речей. Визначення Big Data
- Аналіз великих даних в R. Фактори, списки, фрейми та дії над ними
- Експорт, імпорт та оброблення великих даних в R
- Основні інструменти аналізу та візуалізації великих даних в R
- Технології Hadoop Big Data. Розподілена обробка MapReduce
- Платформа Apache Spark

Модуль 2

2 - кредити

Технології та системи обробки великих даних мовою програмування Python

- Роль Python в аналізі великих даних. Відкриті дані, їх формати та засоби оброблення
- Процедура імпорту даних із файлів у Pandas. Засоби для кореляційного аналізу в Pandas
- Оброблення відсутніх даних. Перетворення типів даних та маніпулювання дата фреймами у Python
- Регресійний аналіз великих даних в Python
- Помилки в аналізі великих даних та прогнозній аналітиці. Оцінка помилок регресії засобами Python.
- Алгоритми класифікації великих даних. Застосування та проблеми класифікацій. Модель класифікатора дерева рішень
- Модуль Pyplot. Інструмент Plotly. Типи візуалізації даних. Візуалізація аномалій



ФОРМА ПІДСУМКОВОГО КОНТРОЛЮ

Залік