



Уманський національний
університет садівництва

Факультет економіки і
підприємництва

Кафедра інформаційних
технологій

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ «Технології BigData»

Рівень вищої освіти:	<u>Другий рівень вищої освіти</u>
Спеціальність:	<u>122 "Комп'ютерні науки"</u>
Освітня програма:	<u>"Комп'ютерні науки"</u>
Навчальний рік, семестр:	<u>2023-2024 н.р., семестр 1</u>
Курс (рік навчання)	<u>1 (2023)</u>
Форма навчання:	<u>Денна</u>
Кількість кредитів ЄКТС:	<u>5.5</u>
Мова викладання:	<u>українська</u>
Обов'язкова/вибіркова:	<u>обов'язкова</u>

Лектор курсу	Володимир Кучерук
Профайл лектора	https://ekis.udau.edu.ua/ua/pro-kafedru/vikladachi-ta-spivrobotniki/kucheruk-volodimir-yurijovich.html
Контактна інформація лектора (e-mail)	kaf_it@udau.edu.ua
Сторінка курсу в MOODLE	https://moodle.udau.edu.ua/course/view.php?id=1993

ОПИС ДИСЦИПЛІНИ

Мета курсу	оволодіння студентами навичок застосування різних методичних підходів до використання технологій та алгоритмів роботи з великими даними
Завдання курсу	набуття навичок проектування методів та засобів оброблення великих даних, навичок застосування різних технологій проектування систем оброблення великих даних
Компетентності	Загальні компетентності: ЗК01 Здатність до абстрактного мислення, аналізу та синтезу. ЗК03 Здатність спілкуватися державною мовою як усно, так і письмово. ЗК06 Здатність бути критичним і самокритичним. Спеціальні компетентності: СК03 Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області. СК04 Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень. СК09 Здатність розробляти та адмініструвати бази даних та знань.
Програмні результати навчання	РН7 Розробляти та застосовувати математичні методи для аналізу інформаційних моделей. РН8 Розробляти математичні моделі та методи аналізу даних (включно з великим). РН9 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

СТРУКТУРА КУРСУ

Тема	Години (лекції/ практичні (семинарські, лабораторні))	Зміст тем курсу	Завдання	Оцінювання (балів)
Змістовий модуль 1 Основні поняття інформаційних систем				
Тема 1. Вступ у великі дані.	4/4	Історія виникнення поняття BigData. Джерела великих даних. Інтернет речей. Основні поняття Hadoop, MapReduce. Проблема масштабування великих даних.	Вивчити термінологію та основні поняття великих даних	5
Тема 2. Загальний огляд інструментів та технологій великих даних.	4/6	Платформа Apache AirFlow. Delta Lake. Механізм запитів Apache Drill. Розподілене середовище Hadoop. Інфраструктура сховища даних HIVE. Платформа обробки великих даних HPCC Systems. Технології Hudi. Відкритий формат таблиці Iceberg. Платформа розподіленої потокової передачі подій Kafka. Розподілене сховище інформації та аналітична платформа Kylin. Система запитів SQL Presto. Система розподіленої обробки потоків Samza. Механізм обробки та аналізу даних у пам'яті Spark. Розподілена система обчислень у реальному часі Storm. База даних NoSQL.	Вивчити основні можливості існуючих платформ для оброблення великих даних та принципи їх функціонування	5
Тема 3. Розроблення програмного забезпечення для аналізу веб-сайтів.	4/4	Можливості інструментів аналізу даних. Роль Python в аналізі даних. Традиційна аналітика великих даних та аналітика нового покоління. Життєвий цикл аналізу даних. Відкриті дані, їх формати та засоби обробки. Веб-скрепінг. Витягування, перетворення та завантаження даних.	Вивчити базові елементи мови програмування Python, що використовуються при обробленні великих даних	5
Тема 4. Використання Python для аналізу великих даних.	4/6	Форматування даних про час та дату у Python. Читання та запис файлів в Python. Взаємодія із зовнішніми додатками. Програмування Python та SQLite. Призначення утиліти csvsql.	Вивчити базові елементи мови програмування Python, що використовуються при обробленні великих даних	5
Змістовий модуль 2. Основні принципи проектування інформаційного забезпечення				
Тема 5. Програмна бібліотека Pandas.	4/4	Процедура імпорту даних із файлів у Pandas. Імпорт даних з мережі Інтернет. Засоби для кореляційного аналізу в Pandas. Статистичні підходи до аналітики великих даних. Використання Pandas. Імпорт даних з файлів. Імпорт даних з мережі Інтернет. Описова статистика в Pandas. Засоби для кореляційного аналізу в Pandas.	Вивчити базові елементи Pandas, що використовуються при обробленні великих даних	10
Тема 6. Технології	4/4	Оброблення відсутніх даних. Перетворення типів даних та	Вивчити методи та алгоритми мови програмування Python, що	10

оброблення великих даних в Python.		маніпулювання дата фреймами у Python. Оброблення відсутніх даних. Перетворення типів даних. Маніпулювання дата фреймами. Регресійний аналіз даних в Python. Методи та типи аналізу машинного навчання. Регресійний аналіз. Типи регресійного аналізу. Застосування регресійного аналізу. Помилки в аналізі даних та прогнозній аналітиці. Оцінка помилок регресії засобами Python. Призначення бібліотеки scikit-learn.	використовуються при обробленні великих даних	
Тема 7. Класифікація та візуалізація великих даних.	4/4	Проблеми класифікації. Алгоритми класифікації. Візуалізація класифікацій. Застосування та валідація класифікацій. Модуль Pyplot. Інструмент Plotly. Типи візуалізації даних. Візуалізація аномалій.	Вивчити базові елементи бібліотеки Python Pyplot, що використовуються при обробленні великих даних	10
Topic 8. Big data analysis using the R programming language.	4/4	History of R language development. R language capabilities. Objects, packages, functions. Vectors, matrices and operations on them in R. Factors, lists, frames and operations on them. Exporting, importing, and processing data in R. Exporting and importing data in R. Using R for time series analysis. Data processing in R. Basic tools for data analysis and visualization in R. The plot() function and its parameters. Management of general parameters - arguments of graphic functions. Graph types in R.	Вивчити базові елементи мови програмування R, що використовуються при обробленні великих даних	10
Модульний контроль				10
Всього за 1 семестр	32/42			70
Підсумковий контроль				30
Всього за курс				100

ПОЛІТИКИ КУРСУ

Політика оцінювання	В основу рейтингового оцінювання знань закладена 100-бальна шкала оцінювання (максимально можлива сума балів, яку може набрати здобувач за всіма видами контролю знань з дисципліни з урахуванням поточної успішності, самостійної роботи, науково-дослідної роботи, модульного контролю, підсумкового контролю тощо). Встановлюється, що при вивченні дисципліни до моменту підсумкового контролю (іспиту) здобувач може набрати максимально 70 балів. На підсумковому контролі (іспит) здобувач може набрати максимально 30 балів, що в сумі і дає 100 балів.
Політика щодо академічної доброчесності	Під час підготовки рефератів (есе) та індивідуальних науково-дослідних завдань, проведення контрольних заходів здобувачі повинні дотримуватися правил академічної доброчесності, які визначено Кодексом доброчесності Уманського НУС. Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Жодні форми порушення академічної доброчесності не толеруються. Виявлення ознак академічної недоброчесності в письмовій роботі здобувача є підставою для її незарахування викладачем, незалежно від масштабів плагіату

Політика щодо відвідування	Відвідування занять є обов'язковим. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись індивідуально (за погодженням із деканом факультету)
-----------------------------------	--

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
		для екзамену
90 – 100	A	Відмінно
82-89	B	Добре
74-81	C	
64-73	D	
60-63	E	Задовільно
35-59	FX	незадовільно з можливістю повторного складання
0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни