



Силабус курсу

АНАЛІТИКА ВЕЛИКИХ ДАНИХ

Ступінь вищої освіти – бакалавр

Галузь знань – 01 Освіта / Педагогіка

Спеціальність – 015 Професійна освіта (за спеціалізаціями)

Спеціалізація – 015.39 Цифрові технології

Освітньо-професійна програма:

“Професійна освіта (Цифрові технології)”

Рік навчання: IV Семестр: VIII

Кількість кредитів: 4 Мова викладання: українська

Керівник курсу

ПП

Адамів Олег Петрович

Контактна інформація

E-mail: oleh.adamiv@wunu.edu.ua

Опис дисципліни

Передбачено вивчення принципів, методів та технологій обробки великих обсягів даних, включно з побудовою сховищ, застосуванням алгоритмів аналітики та інструментів (Hadoop, Spark) для ефективного отримання, аналізу та візуалізації даних у різних галузях.

В результаті вивчення курсу "Аналітика великих даних" студенти повинні:

- знати концепти та джерела Big Data: ознайомитися з характеристиками великих даних (5V), особливостями роботи з структурованими й неструктурованими даними, а також з поняттям IoT (Інтернет Речей) як головним рушієм зростання обсягів даних;
- опанувати технології та інструменти обробки великих даних: знати принципи та інфраструктуру Hadoop (HDFS, YARN, MapReduce), а також NoSQL-системи для зберігання розподілених даних; навчитися застосовувати Apache Spark (Spark Core, Spark Streaming, Spark SQL, MLlib) для обчислень у пам'яті, потокової обробки та машинного навчання; оволодіти базовими підходами до візуалізації (Big Data Visualization, Gephi) та технологіями хмарних обчислень у контексті сховищ великих даних;
- вміти використовувати результати аналітики для прийняття рішень, дотримуючись етичних і правових норм: застосовувати методи візуалізації, класифікації й аналізу великих даних для формування обґрунтованих висновків та стратегічних рішень у бізнесі та науці; урахувати етичні вимоги, законодавчі обмеження й принципи збереження приватності під час роботи з конфіденційною інформацією.

Структура курсу

№	Тема	Результати навчання	Завдання
1	Введення в аналітику великих даних	Знати базові поняття Big Data, їх ключові характеристики (5V: обсяг, швидкість, різноманітність, достовірність, цінність) та відмінність роботи з великими даними від традиційної аналітики. Вміти наводити приклади реальних задач, де застосовується аналітика великих даних (маркетинг, фінанси, охорона здоров'я, освіта, IoT) і формулювати очікувані вигоди від їх використання. Розуміти місце аналітики великих даних у прийнятті управлінських і наукових рішень та її зв'язок з цифровою трансформацією організацій.	Тести, питання

2	Джерела великих даних	Знати основні джерела Big Data: транзакційні системи, сенсори й пристрої Інтернету речей (IoT), лог-файли, соціальні мережі, мобільні додатки, відкриті державні дані, мультимедійні потоки. Вміти класифікувати дані на структуровані, напівструктуровані та неструктуровані, наводити приклади кожного типу та оцінювати потенційну корисність цих даних для аналітики. Розуміти, що саме стрімке зростання кількості пристроїв IoT та онлайн-сервісів є ключовим рушієм вибухового збільшення обсягів даних.	Тести, питання
3	Організація сховищ великих даних	Знати основні підходи до побудови сховищ великих даних: класичні сховища (Data Warehouse, Data Mart), озера даних (Data Lake), гібридні архітектури, а також базові принципи розподіленого зберігання. Вміти пояснити, чому для Big Data застосовують масштабовані розподілені файлові системи й NoSQL-сховища, та чим вони відрізняються від традиційних реляційних БД. Розуміти роль архітектури сховища у забезпеченні доступності, надійності та продуктивності систем аналітики великих даних.	Тести, питання
4	Основні технології та інструменти роботи з великими даними	Знати ключові технології Big Data-аналітики: Hadoop-екосистему (HDFS, YARN, MapReduce), NoSQL-системи (key-value, документні, колоночні, графові бази), інструменти потокової обробки та візуалізації даних. Вміти орієнтуватися в типових сценаріях використання цих технологій: коли доцільно обрати Hadoop, коли – in-memory платформи на зразок Spark, коли – NoSQL-сховища. Розуміти, як вибір інструментів пов'язаний з вимогами до обсягу, швидкості надходження даних, типу аналітики та обмежень інфраструктури.	Тести, питання
5	Технології Hadoop Big Data. Розподілена обробка MapReduce	Знати архітектуру Hadoop: роль HDFS як розподіленої файлової системи, YARN як менеджера ресурсів та MapReduce як моделі розподілених обчислень. Вміти на інтуїтивному рівні пояснити, як парадигма MapReduce розбиває обробку великих наборів даних на дві фази (Map і Reduce), розподіляє завдання між вузлами кластера та забезпечує масштабованість. Розуміти переваги Hadoop для пакетної обробки дуже великих обсягів даних, а також його обмеження (затримки, орієнтація на batch-аналітику) у порівнянні з сучасними in-memory рішеннями.	Тести, питання
6	Apache Spark	Знати основні компоненти Apache Spark: Spark Core, Spark SQL, Spark Streaming, MLlib, GraphX та його ключову ідею — обробку даних у пам'яті для пришвидшення обчислень. Вміти описувати типові сценарії використання Spark: інтерактивний аналіз, потокова обробка, машинне навчання на великих наборах даних, інтеграція з Hadoop та хмарними платформами. Розуміти переваги Spark над суто MapReduce-підходом (швидкість, гнучкість, зручні API для аналітики й ML) та його роль як одного з базових інструментів сучасного аналітика великих даних.	Тести, питання

Літературні джерела

1. Ланде Д. В., Субач І. Ю., Гладун А. Я. Оброблення надвеликих масивів даних (BIG DATA) : навчальний посібник. Київ 2021. 168 с.
2. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань: навчальний посібник. Львів: Магнолія, 2021. 276 с.
3. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник для студентів спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Л. М. Олещенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с. // <https://ela.kpi.ua/items/4ad484d3-aa21-4f7e-be09-a6d8e1d1e8bc>
4. Тарн В.І., Гордієнко Ю. Г., Стіренко С.Г. / Технології BIG DATA Практикум // Київ: КПІ ім. Ігоря Сікорського, 2022. 56 с.
5. Чаплінський В. Р., Кушнір О. К., Свідер О. П. Аналіз великих даних та їх візуалізація для потреб бізнесу. Ефективна економіка. 2021. № 6. – URL: <http://www.economy.nauka.com.ua/?op=1&z=8979>. DOI: 10.32702/2307-2105-2021.6.89
6. Benjamin S. Baumer, Daniel T. Kaplan , Nicholas J. Horton. Modern Data Science with R. 2nd Edition. Chapman and Hall/CRC, 2021. 632 p.
7. Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model. O'Reilly Media, Inc., 2023. 576 p.

Політика оцінювання

Політика щодо дедлайнів і перескладання. Для виконання усіх видів завдань студентами і проведення контрольних заходів встановлюються конкретні терміни. Перескладання модулів відбувається з дозволу дирекції факультету за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності. Списування під час проведення контрольних заходів заборонені. Під час контрольного заходу студент може користуватися лише дозволеними допоміжними матеріалами або засобами, йому забороняється в будь-якій формі обмінюватися інформацією з іншими студентами, використовувати, розповсюджувати, збирати варіанти контрольних завдань.

Політика щодо відвідування. За об'єктивних причин (наприклад, карантин, воєнний стан, хвороба, закордонне стажування тощо) навчання може відбуватись в дистанційній формі за погодженням із керівником курсу з дозволу дирекції факультету.

Політика щодо визнання результатів навчання

Відповідно до «Положення про визнання в Західноукраїнському національному університеті результатів попереднього навчання»

(https://www.wunu.edu.ua/pdf/pologenya/Polozhennya_ruzult_poper_navch.pdf) здобувачам вищої освіти може бути зараховано результати навчання (неформальної/інформальної освіти, академічної мобільності тощо) на підставі підтвердних документів (сертифікати, довідки, документи про підвищення кваліфікації тощо). Рішення про зарахування здобувачу результатів (певного освітнього компонента в цілому, або ж окремого виду навчальної роботи за таким освітнім компонентом) приймається уповноваженою Комісією з визнання результатів навчання за процедурою, визначеною вищезазначеним положенням.

Оцінювання

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Аналітика великих даних» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

Модуль 1		Модуль 2		Модуль 3	Модуль 4
20 %	20%	20 %	20%	5 %	15 %
Поточне оцінювання	Модульний контроль 1	Поточне оцінювання	Модульний контроль 2	Тренінг	Самостійна робота
Оцінка визначається як середнє арифметичне з отриманих оцінок за перший змістовий модуль	Виконання модульного завдання	Оцінка визначається як середнє арифметичне з отриманих оцінок за другий змістовий модуль	Виконання модульного завдання	Оцінка за виконання завдання	Оцінка за виконання самостійного завдання

Шкала оцінювання

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90–100	відмінно	A (відмінно)
85–89	добре	B (дуже добре)
75–84		C (добре)
65–74	задовільно	D (задовільно)
60–64		E (достатньо)
35–59	незадовільно	FX (незадовільно з можливістю повторного складання)
1–34		F (незадовільно з обов'язковим повторним курсом)