

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ

Декан факультету комп'ютерних
інформаційних технологій


Ігор ЯКИМЕНКО

«30» 08 2024 р.

ЗАТВЕРДЖУЮ

Проректор з
науково-педагогічної роботи


Віктор ОСТРОВЕРХОВ

2024 р.

РОБОЧА ПРОГРАМА

з дисципліни

„АНАЛІТИКА ВЕЛИКИХ ДАНИХ”

Ступінь вищої освіти – бакалавр

Галузі знань – 12 «Інформаційні технології»

Спеціальність – 124 «Системний аналіз»

Освітньо-професійна програма – «Системний аналіз»

Кафедра економічної кібернетики та інформатики

Форма навчання/ факультет	Курс	Семестр	Лекції	Лабораторні заняття	ІРС	СРС	Тренінг	Разом	Залік (семестр)
Денна	4	7	32	14	3	95	6	150	7

Тернопіль – 2024

Робочу програму склав старший викладач кафедри економічної кібернетики та інформатики Олег АДАМІВ

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики протокол № 1 від 28.08.2024 р.

Завідувач кафедри
д.е.н., професор



Леся БУЯК

Розглянуто та схвалено групою забезпечення спеціальності «Системний аналіз», протокол № 1 від 30.08.2024 р.

Голова ГЗС
д.т.н., професор



Роман ПАСІЧНИК

Гарант ОПП
д.т.н., професор



Роман ПАСІЧНИК

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Аналітика великих даних»

1. Опис дисципліни «Аналітика великих даних»

Дисципліна – Аналітика великих даних	Галузь знань, спеціальність, СВО	Характеристика навчальної дисципліни
Кількість кредитів ECTS – 5	Галузь знань – 12 «Інформаційні технології»	Вибіркова дисципліна Мова навчання - українська
Кількість залікових модулів – 3	Спеціальність – 124 «Системний аналіз»	Рік підготовки: Денна – 4 Семестр: Денна – 7
Кількість змістових модулів – 2	Ступінь вищої освіти – бакалавр	Лекції: Денна – 32 год. Лабораторні: Денна – 14 год.
Загальна кількість годин – 150 год.		Самостійна робота: Денна – 95 год. Тренінг – 6 год. Індивідуальна робота – 3 год.
Тижневих годин: Денна форма навчання – 10 год., з них аудиторних – 3 год.		Вид підсумкового контролю – залік

2. Мета і завдання дисципліни “Аналітика великих даних”

2.1. Мета вивчення дисципліни – 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.

2.2. Завдання вивчення дисципліни

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.

2.3. Передумови для вивчення дисципліни

Для успішного вивчення цієї дисципліни студенти повинні опанувати такі дисципліни: вища математика, теорія імовірностей та математична статистика.

2.4. Завдання лекційних занять

Мета проведення лекцій полягає у тому, щоб ознайомити студентів із головними питаннями курсу "Аналітика великих даних".

Завдання проведення лекцій полягає у:

– викладенні студентам у відповідності з програмою та робочим планом основних питань курсу "Аналітика великих даних";

– сформуванні у студентів цілісної системи теоретичних знань з курсу "Аналітика великих даних".

2.5. Завдання проведення практичних занять

Мета проведення практичних занять полягає у тому, щоб виробити у студентів практичні навички використання теоретичного матеріалу.

Завдання проведення практичних занять полягає у глибшому засвоєнні та закріпленні теоретичних знань, одержаних на лекціях.

3. Програма навчальної дисципліни “ Аналітика великих даних ”

Змістовий модуль 1. Основи Big Data

Тема 1. Введення в аналітику великих даних

Big Data: поняття, історія, актуальність. Концепція 5V. Сфери застосування великих даних: економіка, бізнес, наука. Професія аналітика великих даних. Етичні та правові аспекти роботи з великими даними. Приватність даних

Тема 2. Джерела великих даних

Інтернет речей (IoT) та зростання даних. Платформа Kaggle. Driven Data. Структуровані та неструктуровані дані. Дані в спокої. Дані в русі. Хмарні обчислення

Тема 3. Організація сховищ великих даних

Збереження великих даних. Технології зберігання: in-memory (Oracle Exadata, SAP HANA), NoSQL, Hadoop/MapReduce. Розгортання сховищ великих даних у хмарних середовищах

Змістовий модуль 2. Технології та інструменти Big Data

Тема 4. Основні технології та інструменти роботи з великими даними

Big Data Visualization. Методи візуалізації: текстів, кластерів, дерев рішень. Ландшафтна візуалізація. Багатовимірна візуалізація. Сіткова візуалізація. Бібліотеки для візуалізації. Класифікація. Gephi.

Тема 5. Технології Hadoop Big Data. Розподілена обробка MapReduce

Hadoop Big Data. Компоненти Hadoop Big Data (HDFS, YARN). MapReduce. Обробка даних у розподілених файлових системах. Масштабованість.

Тема 6. Apache Spark

Архітектура Apache Spark. Spark Core. Spark SQL. Поточкова обробка Spark Streaming. Машинне навчання MLlib. Обчислення в пам'яті

4. Структура залікового кредиту дисципліни

«Аналітика великих даних»

Денна форма навчання

Назва теми	Кількість годин					
	Лекції	Лабораторні заняття	Самостійна робота	Індивідуальна робота	Тренінг	Контрольні заходи
Змістовий модуль 1.						
Тема 1. Введення в аналітику великих даних	4	2	14		3	поточне опит.
Тема 2. Джерела великих даних	4	2	14			поточне опит.

Тема 3. Організація сховищ великих даних	4	2	15			ПОТОЧНЕ ОПИТ.
Змістовий модуль 2.						
Тема 4. Основні технології та інструменти роботи з великими даними	8	4	18	3	3	ПОТОЧНЕ ОПИТ.
Тема 5. Технології Hadoop Big Data. Розподілена обробка MapReduce	6	2	18			ПОТОЧНЕ ОПИТ.
Тема 6. Apache Spark	6	2	16			ПОТОЧНЕ ОПИТ.
Всього	32	14	95	3	6	

5. Тематика лабораторних занять

Лабораторна робота № 1.

Тема: Дослідження базових характеристик великих даних

Мета: Ознайомитися з поняттям Big Data та концепцією 5V, навчитися визначати типи даних (структуровані/неструктуровані) й розглянути основні сфери практичного застосування великих обсягів даних

Лабораторна робота № 2

Тема: Джерела великих даних

Мета: Освоїти методи пошуку та імпорту великих даних із різних джерел (IoT, Kaggle, DrivenData), врахувати особливості «даних у спокої» та «даних у русі» й ознайомитися з можливостями хмарних обчислень для обробки інформації.

Лабораторна робота № 3

Тема: Організація сховищ великих даних

Мета: Навчитися розгортати та налаштовувати середовище зберігання (NoSQL або Hadoop), ознайомитися з in-memory технологіями (Oracle Exadata, SAP HANA) та відпрацювати операції читання/запису у розподіленій файловій системі (HDFS).

Лабораторна робота № 4.

Тема: Основні технології та інструменти роботи з великими даними

Мета: Ознайомитися з налаштуванням базових інструментів для візуалізації великих даних (Gephi або аналогічні платформи), навчитися відображати кластери і деревоподібні структури та первинно аналізувати отримані графічні результати.

Лабораторна робота №5.

Тема: Основні технології та інструменти роботи з великими даними

Мета: Освоїти поглиблені методи Big Data Visualization (ландшафтна, сіткова, багатовимірні візуалізація), оцінити їхню ефективність для різних типів даних і сформулювати рекомендації з вибору оптимального підходу.

Лабораторна робота № 6.

Тема: Технології Hadoop Big Data. Розподілена обробка MapReduce

Мета: Освоїти роботу в середовищі Hadoop (HDFS, YARN), ознайомитися з розподіленою обробкою даних за допомогою MapReduce та оцінити масштабованість системи під час аналізу великих обсягів інформації.

Лабораторна робота № 7.

Тема: Apache Spark

Мета: Освоїти застосування Apache Spark (Spark Core, Spark SQL, Spark Streaming, MLlib) для обчислень у пам'яті, вивчити його можливості потокової обробки й машинного навчання та порівняти продуктивність зі звичайною моделлю MapReduce.

6. Самостійна робота

№ п/п	ЗМІСТ
1	NoSQL-рішення у сфері Big Data
2	Безпека та приватність великих даних
3	Хмарна інфраструктура для Big Data
4	Еволюція Hadoop-екосистеми
5	Машинне навчання на платформі Spark
6	Порівняння пакетної та потокової обробки даних
7	Роль IoT у формуванні великих даних
8	Дашборди та візуалізація Big Data
9	Когнітивні системи в контексті великих даних
10	Big Data в аграрному секторі
11	Застосування великих даних у фінансах
12	Обробка неструктурованих даних
13	Етичні та правові аспекти аналітики великих даних
14	AutoML: перспективи автоматизації аналітики Big Data

Індивідуальні завдання з дисципліни “Аналітика великих даних” студенти виконують самостійно на основі сформованого випадковим чином вибіркового завдання. Вони охоплюють усі основні теми дисципліни “Аналітики великих даних”. Завдання оформлюються у відповідності із встановленими вимогами; при його виконанні та оформленні завдань студент використовує комп'ютерну техніку. Виконання завдань є одним із обов'язкових складових модулів залікового кредиту з дисципліни “Аналітика великих даних”.

7. Тренінг з дисципліни

Тематика. Дослідження основних можливостей збору, обробки та візуалізації великих даних

Тренінг передбачає виокремлення наступних етапів:

- 1) Обрати джерело великих даних.
- 2) Налаштувати середовище для зберігання даних, завантажити вибраний масив і забезпечити належну структуру сховища.
- 3) Провести первинне очищення й трансформацію даних, зокрема застосувати MapReduce або Spark Core/SQL для агрегування та підготовки інформації до аналізу.
- 4) За потреби використати бібліотеки машинного навчання (наприклад, Spark MLlib) для пошуку закономірностей і побудови моделей прогнозування.
- 5) Задіяти інструменти візуалізації з метою наочного відображення ключових результатів і зв'язків у даних.

8. Методи навчання

У навчальному процесі застосовуються: лекції, лабораторні та індивідуальні заняття, консультації, метод опитування, тестування, виконання завдань для самостійної роботи та тренінгу.

9. Засоби оцінювання та методи демонстрування результатів навчання

У процесі навчання дисципліни «Аналітика великих даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- стандартизовані тести;
- поточне опитування;
- залікове модульне тестування та опитування;
- презентації результатів виконаних завдань та досліджень;
- оцінювання результатів завдань тренінгу та самостійної роботи;
- студентські презентації та виступи на наукових заходах;
- завдання на лабораторному обладнанні, тренажерах, реальних об'єктах тощо;
- залік.

10. Критерії, форми поточного та підсумкового контролю

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

Модуль 1		Модуль 2	Модуль 3
40%	40%	5%	15%
Поточне оцінювання	Модульний контроль	Тренінг	Самостійна робота
Оцінка за даний модуль визначається як середнє арифметичне за роботу на лабораторних заняттях	Підсумкова письмова робота	Визначається як середнє арифметичне за виконання завдань тренінгу	Оцінка за виконання самостійної роботи

11. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№ з/п	Найменування	Номер теми
1.	Персональний комп'ютер	1-6
2.	Apache Hadoop, Apache Spark	1-6

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

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. Тарн В.І., Гордієнко Ю. Г., Стіренко С.Г. / Технології BIG DATA Практикум // Київ: КПІ ім. Ігоря Сікорського, 2022. 56 с.

Додаткова література:

1. 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>
2. 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.nayka.com.ua/?op=1&z=8979>.
DOI: [10.32702/2307-2105-2021.6.89](https://doi.org/10.32702/2307-2105-2021.6.89)
3. Benjamin S. Baumer, Daniel T. Kaplan , Nicholas J. Horton. Modern Data Science with R. 2nd Edition. Chapman and Hall/CRC, 2021. 632 p.
4. Hadley Wickham, Mine Çetinkaya-Rundel, Garrett Grolemund. R for Data Science: Import, Tidy, Transform, Visualize, and Model. O'Reilly Media, Inc., 2023. 576 p.