

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«МЕТОДИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ОБРОБКИ
ВЕЛИКИХ ДАНИХ (BIG DATA)»



Ступінь освіти
Освітня програма

бакалавр
Інформаційні системи та
технології
3 чверть

Тривалість
викладання

Заняття:

Лекції

II семестр 2025/2026 н.р.
2 години на тиждень

Лабораторні

2 години на тиждень

Мова викладання

українська

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/course/view.php?id=4107>

Кафедра, що викладає: Інформаційних технологій та комп'ютерної інженерії

Інформація про викладача:



Викладач:

Кожевников А.В., доцент.

Персональна
сторінка:

https://it.nmu.org.ua/ua/HR_staff/prepods/kozhevnykov.php

E-mail:

kozhevnykov.a.v@nmu.one

1. Анотація до курсу

Дані існують всюди. Важко уявити, скільки нових даних генерується кожен день. Дані можуть бути словами в книзі, вмістом електронної таблиці, зображеннями, відео, аудіо або потоком вимірювань, відправленим з пристрою моніторингу. Необроблені дані мають мало користі. Ми повинні обробити ці дані, а потім інтерпретувати вивід, щоб зробити їх корисним. Ці корисні дані тепер перетворилися в інформацію. Коли даних так багато, що традиційні способи їх обробки, зберігання та аналізу неможливо використовувати, це називається великими даними. Великі дані вимагають нових методів та інструментів, щоб зробити їх значущими. Цей курс знайомить здобувачів вищої освіти з цими методами та інструментами, які допоможуть використовувати всю міць великих даних.

2. Мета та завдання навчальної дисципліни

Мета дисципліни – формування умінь та компетенцій щодо методів та інформаційні технології обробки даних, розмір яких перевищує можливості звичайних програмних аналітичних платформ та баз даних по аналізу, зберіганню і управлінню великими даними (Big Data). Реалізація мети вимагає визначення дисциплінарних результатів навчання та адекватний відбір змісту навчальної дисципліни за цим критерієм.

Предметом вивчення дисципліни є технології, методи та засоби обробки і візуалізації великих даних.

Завдання вивчення дисципліни:

- оволодіння основними поняттями інтелектуального аналізу великих даних;
- ознайомлення з новітніми інформаційними технологіями аналітичної обробки великих даних;
- набуття практичних навичок використання методів і засобів інтелектуального аналізу великих даних.

3. Результати навчання

Здобувачі вищої освіти повинні:

Знати: основні поняття, методи, засоби, моделі та алгоритми інтелектуального аналізу великих даних.

Розуміти: принципи застосування технологій інтелектуального аналізу даних, перспективи і проблеми аналітики великих даних.

Уміти: вільно орієнтуватися на сучасному ринку аналітичних програмних продуктів; використовувати мову Python для створення конвеєра даних для збору, аналізу та візуалізації даних; застосовувати моделі машинного навчання для автоматизації завдань.

Набути компетенції. За результатами вивчення дисципліни здобувачі вищої освіти повинні бути здатні:

- обирати методи та розробляти на їх основі алгоритми аналізу даних;
- обирати та використовувати локальні та розподілені середовища обробки даних в тому числі з використанням штучного інтелекту;
- створювати програмні продукти аналізу даних мовою Python

4. Структура курсу

ЛЕКЦІЇ

Тема 1	1. Вступ 1.1 Мета і завдання дисципліни “Методи та техніки аналізу великих даних”. 1.2 Сфери застосування Big Data
Тема 2	2. Подання даних та їх попередня обробка 2.1 Визначення. 2.2 Шкали виміру ознак. 2.3 Життєвий цикл даних. 2.4 Попередня обробка даних. 2.5 Алгоритм ZET заповнення пробілів у таблицях даних. 2.6 Метадані. Життєвий цикл метаданих.
Тема 3	3. Прогнозування стохастичних залежностей. 3.1 Регресійний аналіз. Завдання регресійного аналізу. 3.2 Оцінка параметрів рівнянь парної регресії. 3.3 Система нормальних рівнянь. 3.4 Основні моделі парної регресії. Критерії оцінки якості моделей парної регресії. 3.5 Багатовимірний лінійний регресійний аналіз. Відбір факторних змінних. Мультиколінеарність. 3.6 Алгоритм Фаррара-Глобера. Оцінка параметрів рівняння множинної лінійної регресії. Стандартизоване рівняння множинної лінійної регресії. Покроковий відбір факторів множинної лінійної регресії 3.7 Аналіз та прогнозування часових рядів. Автокореляція часових рядів. Стаціонарність часових рядів. AR та MA процеси
Тема 4	4. Методи класифікації даних. Кластерний аналіз 4.1 Огляд методів класифікації даних. 4.2 Завдання кластерного аналізу. Міри відстані та збіжності. Основні метрики. 4.3 Основні методи кластерного аналізу: ієрархічні агломеративні та дивізійні методи, метод К-середніх Мак-Куїна. 4.4 Нечітка кластеризація, метод С-середніх
Тема 5	5. Методи класифікації даних. Дерева рішень 5.1 Структура дерева рішень і цільова функція. 5.2 Критерії обрання атрибута розбиття дерева. 5.3 Критерії припинення розбиття. Відсікання гілок. 5.4 Алгоритм ID3
Тема 6	6. Методи класифікації даних. Нейронні мережі 6.1 Елементи нейронної мережі. 6.2 Функції активації нейрона. 6.3 Основні архітектури мереж. 6.4 Помилка мережі. Навчання мереж. 6.5 Правила Хебба. Правило Відроу-Хоффа. 6.6 Реалізація бінарного класифікатора та предиктора лінійного часового тренду на основі нейронних мереж.
Тема 7	7. Програмні інструменти Big Data. 7.1 Мова Python 7.2 Статистичні функції мови Python. 7.3 Аналіз даних засобами мови: регресійний, кластерний. 7.4 Реалізація дерев рішень та нейронних мереж засобами мови.
Тема 8	8. Програмні інструменти Big Data. 8.1 Платформа Hadoop 8.2 Екосистема Hadoop. 8.3 Складові платформи: Hadoop Common, YARN, MapReduce, файлова система HDFS. Пісочниця Apache Hadoop

ЛАБОРАТОРНІ ЗАНЯТТЯ

Лабораторна робота 1	Імовірнісні розподіли дискретних випадкових величин і точкові оцінки їх параметрів
Лабораторна робота 2	Імовірнісні розподіли неперервних випадкових величин і точкові оцінки їх параметрів
Лабораторна робота 3	Спеціальні функції математичної статистики, інтервальні оцінки параметрів випадкових величин і тести для перевірки статистичних гіпотез
Лабораторна робота 4	Кореляційний аналіз

Лабораторна робота 5	Регресійний аналіз
Лабораторна робота 6	Кластерний аналіз
Лабораторна робота 7	Знайомство з пісочницею Apache Hadoop

5. Технічне обладнання та/або програмне забезпечення

Використовуються лабораторна та інструментальна бази випускової кафедри інформаційних технологій та комп'ютерної інженерії, а також комп'ютерне та мультимедійне обладнання:

1. Персональний комп'ютер або ноутбук зі сталим доступом до мережі Інтернет
2. Активованний акаунт університетської пошти (student.i.p.@nmu.one) на Ofic365.
3. Активний обліковий запис у системі дистанційної освіти Moodle.
4. Програмне забезпечення:
 - 4.1. Операційна система Windows 10 або Windows 11.
 - 4.2. Office 365.
 - 4.3. Вільно розповсюджене середовище математичних розрахунків Scilab.
 - 4.4. Середовище програмування Python на базі оболонки Anaconda.
 - 4.5. Віртуальні машини Oracle VM VirtualBox та HDP Sandbox.

6. Система оцінювання та вимоги

6.1. Шкали оцінювання

Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
75 – 89	добре
60 – 74	задовільно
0 – 59	незадовільно

6.2. Підсумкова оцінка

Здобувачі вищої освіти можуть отримати підсумкову оцінку з навчальної дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів складатиме не менше як 60 балів.

Максимальне оцінювання становить:

Теоретична частина	Лабораторні роботи	Самостійна робота з підготовки до лабораторних робіт	Разом
51	42	7	100

Теоретична частина оцінюється за результатами online тесту, який містить 27 запитань.

6.3. Критерії оцінювання теоретичної частини курсу

Для кожного запитання тесту потрібно вибрати єдину правильну відповідь. Кількість варіантів відповіді в залежності від характеру запитання може складати від 5-

ти до 10-ти. Кількість балів, які нараховуються за правильну відповідь залежності від складності запитання може складати від 1-го до 4-х. У разі неправильної відповіді на запитання, здобувач вищої освіти отримує за нього 0 балів. Максимальна кількість балів, яку можна набрати за виконання тесту становить 51.

6.4. Критерії оцінювання лабораторних робіт

За кожну лабораторну роботу здобувач вищої освіти може отримати до 6 балів (усього 42 бали), а саме:

6 балів: програма обробки даних правильно функціонує, супроводжується достатньою кількістю коментарів, звіт з роботи оформлений згідно з методичними рекомендаціями і містить повні, інформативні та обґрунтовані висновки.

4 бали: програма обробки даних правильно функціонує, але містить незначні помилки, які суттєво не впливають на отримувані результати, або кількість коментарів недостатня, або звіт з роботи оформлений із незначними відхиленнями від методичних рекомендацій або висновки не повністю задовольняють вимогам повноти, інформативності та обґрунтованості.

2 бали: програма обробки даних функціонує, але містить помилки, які суттєво впливають на отримувані результати, або коментарі відсутні, або звіт з роботи оформлений із суттєвими відхиленнями від методичних рекомендацій, або висновки не задовольняють вимогам повноти, інформативності та обґрунтованості.

0 балів: програма обробки даних не функціонує, або звіт з роботи відсутній, або не містить висновків, або висновки протирічають фактично отриманим результатам.

У разі отримання позитивної оцінки здобувач вищої освіти може підвищити її шляхом очної співбесіди з викладачем. При цьому за кожну правильну відповідь на запропоновані запитання нараховується 1 бал.

6.5. Критерії оцінювання самостійної роботи

За підготовку вхідних даних згідно з варіантом завдання до кожної лабораторної роботи здобувач вищої освіти отримує:

1 бал (разом до 7 балів).

0 балів: підготовку вхідних даних не виконано, дані розраховані невірно або не відповідають варіанту завдання.

7. Політика курсу

7.1. Політика щодо академічної доброчесності

Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи

фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка". http://www.nmu.org.ua/ua/content/activity/us_documents/System_of_prevention_and_detection_of_plagiarism.pdf.

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікаційна політика

Здобувачі вищої освіти повинні мати активовану університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання

Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4. Відвідування занять

Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в он-лайн формі за погодженням з керівником курсу.

7.5. Політика щодо оскарження оцінювання

Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

7.6. Студентоцентризований підхід

Для врахування інтересів та потреб здобувачів вищої освіти на початку вивчення курсу здобувачам вищої освіти пропонується відповісти у системі Moodle на низку питань щодо інформаційного наповнення курсу. Відповідно до результатів опитування формується траєкторія навчання з урахуванням потреб здобувачів вищої освіти.

Під час навчання здобувачі вищої освіти реалізують своє право вибору індивідуальних завдань лабораторних робіт.

Наприкінці вивчення курсу та перед початком сесії здобувачам вищої освіти пропонується анонімно заповнити у системі Moodle електронні анкети для оцінки рівня задоволеності методами навчання і викладання та врахування пропозицій стосовно покращення змісту навчальної дисципліни. За результатами опитування вносяться відповідні корективи у робочу програму та силабус.

8. Рекомендовані джерела інформації

Основні:

1. Zgurovsky M.Z. Big Data: Conceptual Analysis and Applications. [Текст] / M.Z. Zgurovsky, Y.P. Zaychenko // Springer, 2021, 298 p.
2. Кожевников А.В. Методи та інформаційні технології обробки великих даних (Big Data) [Електронний ресурс] : методичні рекомендації до виконання лабораторних робіт для здобувачів ступеня бакалавра освітньо-професійної програми «Інформаційні системи та технології» зі спеціальності 126 Інформаційні системи та технології / А. В. Кожевников ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2025. – 89 с.
3. Akerkar R. Models of Computation for Big Data [Текст] / R. Akerkar // Springer, 2018, 110 p.
4. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics [Текст] / P. Ghavami // CreateSpace Independent Publishing Platform, 2016, 204 p.
5. Feeney K. Engineering Agile Big-Data Systems [Текст] / K. Feeney, J. Davies, J. Welch, S. Hellmann, C. Dirschl, A. Koller, P. Francois, A. Marciniak // River Publishers, 2018, 436 p.
6. Big Data Fundamentals courses [Електрон. ресурс]. Режим доступа: <https://cognitiveclass.ai/learn/big-data> (дата звернення: 20.08.2024).
7. Big Data Analytics [Електрон. ресурс]. Режим доступа: <https://cognitiveclass.ai/learn/analytics/> (дата звернення: 20.08.2024).

Додаткові:

1. Aziukovskyi O., Udovik I., Kozhevnykov A., Powroźnik T. Creating using the mathcad system of laboratory experimentation on the subject «intelligent data analysis» [Текст] / Матеріали XVI міжнародної конференції "Проблеми використання інформаційних технологій у сфері освіти, науки та промисловості" – Дніпро: НТУ «Дніпровська політехніка», 2021. – С. 21–26.
2. Кожевников А.В., Удовик І.М. Створення відкритої нейронної мережі бінарного класифікатора засобами системи Mathcad [Текст] / Матеріали XVI міжнародної конференції "Проблеми використання інформаційних технологій у сфері освіти, науки та промисловості" – Дніпро: НТУ «Дніпровська політехніка», 2021. – С. 27–32.
3. Кожевников А.В., Удовик І.М. Створення відкритої нейронної мережі предиктора лінійного часового ряду засобами системи Mathcad [Текст] / Матеріали XVI міжнародної конференції "Проблеми використання інформаційних технологій у сфері освіти, науки та промисловості" – Дніпро: НТУ «Дніпровська політехніка», 2021. – С. 45–49.
4. UCI Machine Learning Repository [Електрон. ресурс]. Режим доступа: <http://archive.ics.uci.edu/ml/index.php/> (дата звернення: 21.10.2024).
5. HDFS Commands Guide – Apache Hadoop 3.4.1 [Електрон. ресурс]. Режим доступа: <https://hadoop.apache.org/docs/r3.4.1/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html> (дата звернення: 21.10.2024).