

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

ЗАТВЕРДЖЕНО
на засіданні кафедри
інформатики та комп'ютерної техніки
Протокол № 21 від 29.08.2025 р.

ПОГОДЖЕНО
Проректор з навчально-методичної
роботи



Каріна НЕМАШКАЛО

**ТЕХНОЛОГІЇ РОБОТИ З BIG DATA
робоча програма навчальної дисципліни (РПНД)**

Галузь знань **F Інформаційні технології**
Спеціальність **F6 "Інформаційні системи і технології"**
Освітній рівень **перший (бакалаврський)**
Освітня програма **"Штучний інтелект"**

Статус дисципліни **обов'язкова**
Мова викладання, навчання та оцінювання **українська**

Розробник:
к.т.н., доц.

Ігор КОБЗЕВ

Завідувач кафедри
інформатики та комп'ютерної
техніки

Сергій УДОВЕНКО

Гарант програми

Сергій УДОВЕНКО

**Харків
2025**

ВСТУП

Дисципліна "Технології роботи з BIG DATA" формує у здобувачів необхідний обсяг теоретичних і практичних знань про великі дані Big Data, методи та засоби їхнього аналізу та обробки, основних принципів управління великими даними. Вивчення дисципліни забезпечує володіння знаннями щодо хмарних технологій та технологій віртуалізації, гіпервізорів, систем управління, балансування трафіку хмарних ресурсів та забезпечення гарантованої якості обслуговування. Знання, отримані при вивченні дисципліни "Технології роботи з BIG DATA" дозволять здобувачам виконувати інтелектуальний аналіз та обробку великих даних, формувати віртуальні хмарні середовища та володіти навичками контейнерної віртуалізації, володінню інструментами конфігурування віртуальних хмарних ресурсів і способами оцінки ефективності хмарної інфраструктури.

Навчальна дисципліна "Технології роботи з BIG DATA" є обов'язковою навчальною дисципліною та вивчається згідно з навчальним планом підготовки здобувачів за спеціальністю Ф6 "Штучний інтелект" першого (бакалаврського) рівня усіх форм навчання. Програму навчальної дисципліни розроблено у відповідності до вимог галузевого стандарту вищої освіти на базі освітньо-професійної програми підготовки бакалавра.

Метою викладання навчальної дисципліни є формування у майбутніх фахівців теоретичних і практичних знань з основ застосування методів інтелектуального аналізу даних для розв'язання спеціалізованих задач попередньої обробки, перетворення, класифікації даних та дослідження процесів у сфері використання інформаційних технологій.

Завданням навчальної дисципліни є формування у здобувачів теоретичних знань про основні поняття та методи аналізу і пошуку закономірностей у великих наборах даних та їх застосування при розв'язанні конкретних задач, розвиток математичного та логічного мислення здобувачів, навичок самостійної роботи над літературою.

Предметом вивчення навчальної дисципліни є методи, інструменти, алгоритми та архітектурні рішення, призначені для отримання, зберігання, обробки, аналізу та візуалізації великих обсягів структурованих і неструктурованих даних, які неможливо обробити традиційними системами.

Об'єктом вивчення навчальної дисципліни є процеси обробки та аналізу великих масивів даних (структурованих та неструктурованих), що характеризуються високим обсягом, швидкістю приросту та різноманітністю.

Навчальна дисципліна "Технології роботи з BIG DATA" надає здобувачам теоретичні знання та практичні навички вибору та застосування математичних моделей та методів опрацювання експериментальних даних та інших спеціально не підготовлених великих даних (Big Data), як основи задач математичного моделювання процесів у різних галузях людської діяльності.

Програма навчальної дисципліни передбачає навчання у формі лекцій, лабораторних занять та самостійної роботи здобувачів. Лекції, лабораторні заняття, індивідуальна робота та консультації проводяться з застосуванням персональних комп'ютерів, локальної мережі та мережі Інтернет у комп'ютерних класах (або з

використанням дистанційних технологій навчання в режимі он-лайн). Всі види занять забезпечуються необхідними електронними методичними матеріалами.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
ПР 6	ІК, КС 2, КС 12
ПР 8	ІК, КС 1, КС 9

де, ПР 6. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

ПР 8. Застосовувати правила оформлення проектних матеріалів інформаційних систем та технологій, знати склад та послідовність виконання проектних робіт з урахуванням вимог відповідних нормативно-правових документів для запровадження у професійній діяльності.

ІК. Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в області інформаційних систем та технологій, або в процесі навчання, що характеризуються комплексністю та невизначеністю умов, які потребують застосування теорій та методів інформаційних технологій.

КС 1. Здатність аналізувати об'єкт проектування або функціонування та його предметну область.

КС 2. Здатність застосовувати стандарти в області інформаційних систем та технологій при розробці функціональних профілів, побудові та інтеграції систем, продуктів, сервісів і елементів інфраструктури організації.

КС 9. Здатність розробляти бізнес-рішення та оцінювати нові технологічні пропозиції.

КС 12. Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст навчальної дисципліни

Змістовий модуль 1. Основи Big Data

Тема 1. Головні поняття, функції і задачі великих даних

Основні поняття та визначення. Історія розвитку. Джерела великих даних. Застосування в економіці, бізнесі, сільському господарстві, промисловості. Приклади використання. Великі дані в наукових сферах. Особливості застосування. Вимоги до професії аналітика великих даних.

Тема 2. Методи аналізу великих даних

Основні етапи життєвого циклу. Збір, консолідація і очищення даних. Побудова моделей, роль машинного навчання.

Тема 3. Технології роботи з великими даними

Збір та консолідація даних, "аналітична пісочниця", "озеро даних", пакетна аналітика, аналітика реального часу, гібридна аналітика, робота з СУБД.

Тема 4. Розподілена обробка великих даних та метод Map-Reduce

Збереження великих даних. Технології: in-memory (Oracle Exadata, SAP HANA), NoSQL, Hadoop/MapReduce. Алгоритм MapReduce. Apache Hadoop, HDFS, HBase, YARN, Hive, Pig, Storm як система потокової обробки.

Змістовий модуль 2. Технології та інструменти роботи з Big Data

Тема 5. Використання неструктурованих баз даних в Big Data

Огляд декількох СУБД, побудованих на основі нереляційних моделей баз даних. Загальні правила використання і звернення до них при розробці інформаційних систем. Розбір застосування програмних засобів реалізації нереляційних моделей баз даних.

Тема 6. Інструменти аналітичної обробки даних у сховищі великих даних

Apache Spark – інфраструктура кластерних обчислень. Компоненти Spark – MapReduce, HDFS, YARN, Hive, Pig, Zookeeper, Flume та інші. Алгоритм Apache Hive. Мова Pig Latin (Apache Pig). Планування завдань Apache Oozie.

Тема 7. Аналіз даних в R.

Історія розвитку мови R. Можливості мови R. Об'єкти, пакети, функції. Вектори, матриці та операції над ними в R. Фактори, списки, фрейми та дії над ними.

Тема 8. Основні типи віртуалізації

Програмна та апаратна віртуалізація. Контейнерна віртуалізація. Огляд платформ віртуалізації.

Перелік лабораторних занять за навчальною дисципліною наведено в табл. 2

Таблиця 2

Перелік лабораторних занять

Назва теми та / або завдання	Зміст
Тема 1-3. Завдання 1	Встановлення і використання Linux Mint на Oracle VM VirtualBox
Тема 4. Завдання 2	Основи роботи з Apache Hadoop
Тема 5. Завдання 3	Основи роботи з MapReduce
Тема 5. Завдання 4	Основи роботи з MongoDB

Тема 6. Завдання 5	Використання мови програмування Python для обробки великих даних
Тема 7-8. Завдання 6.	Обробка статистичних великих даних за допомогою мови програмування R

Перелік самостійної роботи за навчальною дисципліною наведено в табл. 3

Таблиця 3

Перелік самостійної роботи

Назва теми	Зміст
Тема 1 - 8	Вивчення лекційного матеріалу
Тема 1 - 8	Підготовка до лабораторних занять

Кількість годин лекційних та лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

МЕТОДИ НАВЧАННЯ

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

- словесні (лекції за темами 1-8), елементи проблемних лекції (за темами 1-8);
- наочні (демонстрації (теми 1-8);
- практичні (лабораторні заняття за темами 1-8 навчальної дисципліни).

В умовах змішаної форми навчання подання лекційного матеріалу та/або проведення лабораторних занять та групових та індивідуальних консультацій відбувається з використанням платформи Zoom, в умовах звичайної аудиторної форми заняття проводяться очно, в аудиторіях та комп'ютерних залах.

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

Поточний контроль здійснюється під час проведення лекційних та лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів:

– для дисциплін з формою семестрового контролю залік: максимальна сума – 100 балів; мінімальна сума – 60 балів.

Підсумковий контроль включає семестровий контроль та атестацію здобувача вищої освіти.

Семестровий контроль проводиться у формах диференційованого заліку або заліку.

Підсумкова оцінка за навчальною дисципліною визначається:

– для дисциплін з формою семестрового контролю залік – сумуванням всіх балів, отриманих під час поточного контролю.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: захист лабораторних робіт (50 балів), тести (20 балів), індивідуальне завдання (30 балів).

Семестровий контроль: Залік

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Wiktorski T. Data-intensive Systems Principles and Fundamentals using Hadoop and Spark / Tomasz Wiktorski. – Springer Nature Switzerland AG. : Springer International Publishing, 2019. – 524 p.

2. О'Ніл К. BIG DATA. Зброя математичного знищення (MIM) / Кейт О'Ніл. – Київ : Форс, 2019. – 336 с.

3. Mieko Tanaka-Yamawaki. Principal Component Analysis and Randomness Test for Big/ Data Analysis. Practical Applications of RMT-Based Technique – Springer International Publishing. – 2022. – 152 p.

4. Акіменко В. В. Прикладні задачі інтелектуального аналізу даних (Data Mining) / В. Акіменко. – Київ : КНУ ім. Тараса Шевченка, 2018 р. – 162 с.

Додаткова

5. Полянський В. О. Методи та моделі Data Science в системі управління безпекою / В. О. Полянський // Big data analytics: моделювання та інформаційні технології : міжнар. наук. симпозіум 20 бер. 2019 р. тези доп. – Київ : 2019. – С. 160–163. URL: <http://repository.hneu.edu.ua/handle/123456789/22431>

6. Minukhin S. Performance study of the DTU model for relational databases on the Azure platform / S. Minukhin // Сучасний стан наукових досліджень та технологій в промисловості. – 2022. – № 1 (19). – С. 27-39. URL: <http://repository.hneu.edu.ua/handle/123456789/27739>

7. Гілдер Д. Життя після Google. Занепад великих даних і становлення блокчейн-економіки / Джордж Гілдер. – Київ: Форс, 2021. – 320 с.

8. Базилевич О. В. / Застосування аналізу великих даних для оцінки ефективності рекламної кампанії підприємства / О. В. Базилевич, Ю. Е. Парфьонов // Інформаційні технології та системи : матеріали міжнар. наук.-практ. конф., 8 - 9 квіт. 2021 р.: тези допов. – Харків: ХНЕУ ім. Семена Кузнеця, 2021. – С. 26. URL: <http://repository.hneu.edu.ua/handle/123456789/25305>

9. Лістьєв З. С. Моделювання та моніторинг роботи з великими даними на хмарній платформі Azure: дипломна робота на здобуття освітнього ступеня магістра: спец. 122 "Комп'ютерні науки" / З. С. Лістьєв. – Харків : ХНЕУ ім. С. Кузнеця, 2021. URL: <http://repository.hneu.edu.ua/handle/123456789/26959>

Інформаційні ресурси

10. Технології роботи з BIG DATA. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця. – URL : <https://pns.hneu.edu.ua/course/view.php?id=8358>
11. Introduction to Big Data. Coursera. - URL: <https://www.coursera.org/learn/big-data-introduction/home/welcome>
12. Big Data, Hadoop, and Spark Basics. EDX - URL: <https://learning.edx.org/course/course-v1:IBM+BD0225EN+2T2021/home>