



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Хмарні сховища і технології Big Data

Рівень вищої освіти	Перший (бакалаврський)
Код та назва спеціальності	F6 Інформаційні системи та технології
Галузь знань	F Інформаційні технології
Тип та назва освітньої програми	Освітньо-професійна програма «Прикладні інформаційні системи та технології»
Статус навчальної дисципліни	Обов'язкова компонента (ОК-25)
Курс, семестр викладання	4 курс, 1 (7) семестр
Трудомісткість навчальної дисципліни	6 кредитів ЄКТС (180 академічних годин), з них: денна (очна) форма навчання: лекц. – 30 год., практич. зан. – 18 год., лаб. зан. – 18 год., самот. роб. – 114 год.; заочна форма навчання: лекц. – 10 год., практич. зан. – 10 год., лаб. зан. – 10 год., самот. роб. – 150 год.
Мова викладання	Українська
Кафедра	Інформаційних та комп'ютерних систем
Факультет	Інформаційних технологій та кібербезпеки

Розробники / викладачі



Тихонова Олена Вікторівна,
старший викладач кафедри інформаційних та комп'ютерних систем, кандидат технічних наук

E-mail: elena.tykhonova@suitt.edu.ua

Консультації: згідно визначеного розкладу - ауд.
402 (головний корпус) або онлайн за посиланням
<https://us04web.zoom.us/j/3185149804?pwd=TmUybHZZYzBRK2dleUQrNVhPaG1wdz09>

Загальна інформація про дисципліну

Мета дисципліни	Формування у здобувачів фундаментальних знань та практичних навичок, необхідних для ефективного зберігання, управління та аналізу великих обсягів даних
Компетентності, формуванню яких сприяє дисципліна	<p>ЗК-5. Здатність вчитися і оволодівати сучасними знаннями.</p> <p>СК-4. Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).</p> <p>СК-6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.</p> <p>СК10. Здатність вибору, проектування, розгортання, інтегрування, управління, адміністрування та супроводжування інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.</p> <p>СК-12. Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).</p> <p>СК-15. Здатність використовувати технології віртуалізації та хмарних обчислень під час проектування, розробки, реконфігурації та реконструкції інформаційних систем.</p>
Програмні результати навчання	<p>ПРН-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.</p> <p>ПРН-7. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.</p> <p>ПРН13. Знати сучасні методи і технології проектування баз даних та знань та вміти застосовувати їх на практиці.</p> <p>ПРН-17. Застосовувати інформаційні технології та засоби для створення ІТ інфраструктури та її компонентів, вміти здійснювати їх технічне обслуговування.</p> <p>ПРН18. Застосовувати методи штучного інтелекту для рішення різноманітних задач у різних економічних сферах.</p>

Програма навчальної дисципліни

Тема 1.	Основні положення концепції Big Data. Визначення понять великих даних та причини їх появи. Сфери застосування Big Data та переваги їх використання для бізнесу. Порівняння «традиційних» даних та «великих» даних. Джерела надходження великих даних. Описова модель великих даних 3V: Volume (об'єм), Velocity (швидкість), Variety(різноманітність).
Тема 2.	Проблеми реляційних баз даних у контексті Big Data. Концепція NoSQL (NOSQL). Переваги та недоліки NOSQL підходу у порівнянні з реляційними базами даних. CAP-теорема. Основні види нереляційних баз даних (NOSQL БД) - документоорієнтовані, бази даних «ключ- значення», графові, колонкові. Бази даних NewSQL. Принципи онлайн-обробки

транзакцій (OLTP) та аналітичної онлайн-обробки даних (OLAP). Багатовимірна модель даних. Гіперкуб даних, зріз гіперкубу. Представлення гвперкубу за допомогою схем «Зірка» та «Сніжинка».

- Тема 3.** Поняття Сховища Даних (Data Warehouse, DWH). Базова тришарова архітектура екосистеми DWH (шари джерел даних, сховища даних, користувачів). Вітрини даних. Додавання прошарків проміжних даних та вітрин даних до базової архітектури DWH. Процеси ETL та ELT. Традиційні підходи до організації сховища даних (Ральфа Кімбала, Біла Інмона). Сучасні методології побудови сховища даних (Data Vault, Anchor Modelling). Поняття Озера Даних (Data Lake). Особливості озер даних у порівнянні з DWH. Data lakehouse - гібридне рішення для зберігання даних
- Тема 4.** Обчислювальна парадигма MapReduce, фази відображення (Map) і згортки (Reduce), проміжний етап Shuffle. Переваги та недоліки моделі MapReduce. Актуальність технології MapReduce на поточний момент. Еволюція ролі MapReduce. Проект Apache Hadoop - відкрита реалізація концепції MapReduce. Основні компоненти Hadoop. HDFS - розподілена файлова система Hadoop. Концепція блоків в HDFS. Архітектура кластерної системи Hadoop 1.0.
- Тема 5.** Еволюція систем розподілених обчислень Hadoop. Обмеження Hadoop MapReduce v1. Версія Hadoop MapReduce 2.0, менеджер ресурсів YARN. Оновлення третьої версії Hadoop. Фреймворк Apache Spark для розподіленої обробки даних, його особливості. Компоненти Apache Spark. Apache Hive. Сучасне використання Apache Hive та Apache Spark.
- Тема 6.** Хмарні платформи даних. Приватна, публічна та гібридна хмара. Структура хмарних систем зберігання даних. Способи зберігання даних у хмарі: блоковий, об'єктний, файловий. Особливості використання, переваги та недоліки хмарного сховища для Big Data. Основні постачальники хмарних послуг (AWS, Azure, GCP) та їх служби обробки даних (S3, BigQuery, Snowflake, Redshift для сучасних сховищ даних/озер).
- Тема 7.** Інтелектуальний аналіз даних. Три класи задач аналізу: інформаційно-пошуковий, оперативно-аналітичний, інтелектуальний (Data Mining). Життєвий цикл аналітики даних. Методологія дослідження даних CRISP-DM. Основні етапи CRISP-DM: розуміння цілей і вимог проекту, розуміння даних, підготовка та попередня обробка даних, моделювання, оцінка, розгортання. Основні задачі Data Mining (описові, прогнозуючі). Дослідження науковців ДУІТЗ у сфері обробки великих даних, зокрема великих текстових масивів
- Тема 8.** Описові задачі Data Mining. Актуальність задачі пошуку асоціативних правил. Алгоритм Apriori для пошуку асоціативних правил. Ієрархічні та неієрархічні методи кластеризації. Англомеративний алгоритм кластеризації. Метод k-середніх для вирішення задач кластеризації. Правило «ліктя» визначення кількості кластерів для методу k-середніх
- Тема 9.** Прогнозуючі задачі Data Mining. Методи класифікації: наївний Байєсовський алгоритм, метод k-найближчих сусідів. Метрики для обчислення відстані між елементами даних. Нормалізація даних при обчисленні відстані між елементами. Метод опорних векторів та дерева рішень. Показник «забруднення» вузла дерева, інформаційний приріст способу розбиття, вибір найкращого розбиття вузла дерева рішень. Метод найменших квадратів вирішення задачі лінійної регресії.
- Тема 10.** Використання штучних нейронних мереж для вирішення задач інтелектуального аналізу даних. Поняття «штучний інтелект», «машинне навчання», «нейронні мережі». Основні компоненти нейронної мережі (вхідний, вихідний, прихований шари; ваги та зміщення; функція активації). Стратегія ініціалізації вагових коефіцієнтів та зміщень. Алгоритми нейронних мереж за принципом навчання: зі вчителем, без вчителя, з підкріпленням. Класифікація даних за допомогою нейронної мережі «зі вчителем». Кластеризація даних. Алгоритм нейронної мережі, що навчається «без вчителя». Автоенкодера. Нейронні мережі «з підкріпленням» в задачах аналізу даних. Процес навчання нейронної мережі; поняття функції втрат, градієнтного спуску та зворотного поширення помилки. Гіперпараметри алгоритмів нейронних мереж. Переваги нейронних мереж для вирішення задач аналізу даних.

Методи навчання

При вивченні навчальної дисципліни використовуються наступні методи навчання:

За джерелом знань

- Словесні методи: лекція, розповідь, пояснення, бесіда, дискусія
- Наочні методи: ілюстрація
- Практичні методи: лабораторні та практичні роботи

За характером пізнавальної діяльності

- Пояснювально-ілюстративний метод
- Частково-пошуковий метод

За рівнем активності

- Пасивні методи: лекції
- Активні методи: дискусії
- Інтерактивні методи: долучення студентів до розробки та проведення занять

Методи дистанційного навчання

- Відеоконференції в форматі лекцій або семінарів. Зв'язок здобувачів освіти з викладачем забезпечують різноманітні сучасні платформи, такі як: Zoom, Moodle, Google Meet та ін.
- Онлайн дискусії;
- Індивідуальне і групове консультування (викладач дає додаткові роз'яснення щодо виконання завдань через чати та e-mail- надсилання);

Стратегія оцінювання результатів навчання

Змістовий контент результатів навчання з дисципліни

Результати навчання з даної дисципліни, які здобувач освіти може продемонструвати та які можна ідентифікувати, оцінити і виміряти, розглядаються у вимірах 6 рівня Національної рамки кваліфікацій, що відповідає першому циклу вищої освіти Рамки кваліфікацій Європейського простору вищої освіти, а саме:

Знання – мати системні знання про концепцію Big Data, її ключові характеристики (модель 3V), основні сфери застосування та джерела даних; розуміти відмінності між реляційними та нереляційними базами даних (NoSQL), знати основні види NoSQL-баз, принципи роботи OLAP/OLTP та моделі сховищ даних ("Зірка", "Сніжинка"); знати основні архітектурні підходи до побудови сховищ даних (DWH, Data Lake, Data Lakehouse) та хмарні платформи для зберігання даних; розуміти принципи роботи розподілених файлових систем (HDFS) і обчислювальних парадигм (MapReduce, Apache Spark); розуміти принципи роботи ключових алгоритмів інтелектуального аналізу даних (Data Mining);

Уміння/навички – вміти застосовувати принципи багатовимірного моделювання для організації даних у сховищах; мати практичні навички виконання базових кроків обробки даних в Apache Spark; мати практичні навички застосування алгоритмів кластерного аналізу, класифікації, регресії, нейронних мереж для вирішення простих задач аналізу даних, вміти інтерпретувати результати роботи алгоритмів;

Комунікація – вміти чітко та послідовно пояснювати принципи роботи основних технологій Big Data та алгоритмів Data Mining; бути здатним представляти результати виконання практичних робіт, обґрунтовуючи вибір методів та зрозуміло пояснюючи.

(практичні, лабораторні, семінарські) заняття та завдання, що виконуються під час самостійної роботи.
До підсумкового контролю допускаються здобувачі, які за результатами поточного оцінювання набрали не менше 60 балів.
Підсумковий контроль проводиться у формі **екзамену**, який передбачає перевірку рівня теоретичних знань, практичних умінь і навичок, а також здатності їх застосовувати у професійній діяльності.

Політика навчальної дисципліни

Відвідування

Здобувачі вищої освіти самостійно планують відвідування лекційних занять, що проводяться в межах дисципліни згідно академічного розкладу. Присутність на практичних заняттях та контрольних заходах (екзамен) є обов'язковою. Важливим є своєчасне виконання індивідуальних завдань в межах самостійної роботи, передбачених програмою дисципліни.

Дотримання принципів академічної доброчесності

Підготовка усіх завдань, письмових робіт і т. ін., що виконуються в межах дисципліни, здійснюється здобувачем вищої освіти самостійно, на засадах академічної доброчесності. Викладач має право для перевірки робіт застосовувати різні програмні засоби.

Умови зарахування пропущених занять

Відпрацювання академічної заборгованості з дисципліни можливо до початку екзаменаційної сесії. Процедура узгоджується з викладачем, згідно його розкладу консультацій.

Інші умови

Навчально-методичні матеріали з дисципліни розміщені на гугл-диску НМКД та платформі Moodle, доступ до них здобувачу надається на початку вивчення дисципліни ОК-25, згідно розкладу академічних занять та робочого навчального плану ОПП «Прикладні інформаційні системи та технології».

Рекомендовані джерела інформації

Базові підручники та навчальні посібники

- Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020, 298 p. Режим доступу: <https://vdoc.pub/download/big-data-conceptual-analysis-and-applications-40crds1qhta0>
- Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. 105 p.
- Victor Tikhonov, Eduard Siemens, Yevhen Vasiliu, Valery Sitnikov, Abdullah Taher, Olena Tykhonova, Kateryna Shulakova and
- Serhii Tikhonov. “Context-Defined Model of Open Systems Interaction for IoT Cybersecurity Issues Study” // Proc. of 12-th Int. Conf. on Applied Innovation in IT, 2024/11/30, Vol. 12, Issue 2, pp.35-44. 2024, ISSN: 2199-8876. (Scopus).

Методичні рекомендації та розробки викладачів дисципліни

- Методичні вказівки до практичних занять з дисципліни «Технології Big Data» / Уклад.: В.І. Тіхонов, О.В. Тихонова. Одеса: ДУІТЗ, 2025. 84 с.
- Методичні рекомендації до лабораторних робіт з дисципліни «Технології Big Data» / Уклад.: В.І. Тіхонов, О.В. Тихонова. Одеса: ДУІТЗ, 2025. 84 с.

Інформаційні ресурси

- An introduction to big data: <https://opensource.com/resources/big-data>.
- Big Data Resource: <https://www.sciencedirect.com/topics/computer-science/big-data-resource>.
- Best Resources to learn Big Data: <https://www.analyticsvidhya.com/resources-big-data/>.

Рік введення силябусу – 2025 р. Затверджено рішенням кафедри публічного управління та цифрової економіки
(Протокол від 26 серпня 2025 р. № 1)

В. о. завідувача кафедри



Роман ЦАРЬОВ

Гарант освітньої програми



Роман ЦАРЬОВ

Викладач:



Олена ТИХОНОВА