

ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ

Система забезпечення якості освітньої діяльності та якості вищої освіти
сертифікована на відповідність ДСТУ ISO 9001:2015 / ISO 9001:2015

Кафедра цифрової економіки та системного аналізу

ЗАТВЕРДЖЕНО

вченою радою ФІТ

(протокол № 14 від «26» 06 2024 р.)

Декан

 Олександр ХАРЧЕНКО

**ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ /
DATA ANALYSIS TECHNOLOGIES**

**РОБОЧА ПРОГРАМА /
COURSE OUTLINE**

освітній ступінь	магістр / master
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	122 Комп'ютерні науки / Computer Science 126 Інформаційні системи та технології / Information Systems and Technologies
освітня програма	Комп'ютерні науки / Computer Science Інформаційні системи та технології / Information Systems and Technologies

Київ 2024

Розповсюдження і тиражування без офіційного дозволу ДТЕУ заборонено

Автор: А. А. Роскладка, доктор економічних наук, професор

Робочу програму розглянуто і затверджено на засіданні кафедри цифрової економіки та системного аналізу 17.06.2024р., протокол № 19

Рецензенти: В. Ф. Гамалій, доктор фізико-математичних наук, професор,
О. І. Пурський, доктор фізико-математичних наук, професор

ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ / DATA ANALYSIS TECHNOLOGIES

РОБОЧА ПРОГРАМА / COURSE OUTLINE

освітній ступінь	магістр / master
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	122 Комп'ютерні науки / Computer Science 126 Інформаційні системи та технології / Information Systems and Technologies
освітня програма	Комп'ютерні науки / Computer Science Інформаційні системи та технології / Information Systems and Technologies

1. СТРУКТУРА ДИСЦИПЛІНИ ТА РОЗПОДІЛ ГОДИН ЗА ТЕМАМИ (ТЕМАТИЧНИЙ ПЛАН)

Для вивчення дисципліни відводиться 180 год / 6 кредитів ЄКТС (лекцій – 28 год., практичних занять – 28 год., самостійної роботи – 124 год.), підсумковий контроль – екзамен.

Назва теми	Кількість годин				Форми контролю*
	Усього годин/кредитів	з них:			
		лекції	практичні заняття	самостійна робота студентів	
<i>I семестр</i>					
Тема 1. Вступ до аналітики даних.	22	4	4	14	Т, ВПЗ, ВІЗ
Тема 2. Статистична обробка інформації в аналітичній системі R.	14	4	2	8	Т, ВПЗ, ВІЗ
Тема 3. Розвідувальна аналітика в системі R.	14	2	2	12	Т, ВПЗ, ВІЗ
Тема 4. Аналітична платформа Tableau.	22	4	4	12	Т, ВПЗ, ВІЗ
Тема 5. Аналітика в системі Microsoft Power BI.	26	4	10	12	Т, ВПЗ, ВІЗ
Тема 6. Консолідація та передобробка даних.	12	2	0	14	Т, ПР
Тема 7. Класифікація, регресія та прогнозування даних.	18	4	0	14	Т, ПР
Тема 8. Кластеризація даних	20	2	4	12	Т, ВПЗ, ВІЗ
Тема 9. Асоціативний аналіз даних	20	2	4	10	Т, ВПЗ, ВІЗ
Тема 10. Технології інтелектуальної обробки даних	12	2	0	10	Т, ПР
РАЗОМ	180/6	30	30	120	х
Підсумковий контроль – екзамен					

Примітка*:

Т – тестування; ВПЗ – виконання практичних завдань; ВІЗ – виконання індивідуальних завдань; ПР – підготовка презентації.

2. ТЕМАТИКА ТА ЗМІСТ ЛЕКЦІЙНИХ, ПРАКТИЧНИХ ЗАНЯТЬ, САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p>Знати: основні поняття Data Science, типи, форми і формати даних, основні задачі аналізу даних</p> <p>Вміти: визначати характеристики реальних наборів даних</p>	<p>Тема 1. Вступ до аналітики даних План лекції 1</p>	2
	<ol style="list-style-type: none"> 1. Глосарій <i>Data Science</i>. 2. Історія розвитку <i>Artificial Intelligence (AI)</i> і <i>Business Intelligence (BI)</i> 3. Дані, інформація та знання. 4. Типи та види даних. 	
	<p>План лекції 2</p> <ol style="list-style-type: none"> 1. Послідовність аналізу даних. 2. Методика <i>Knowledge Discovery in Databases</i>. 3. Технологія <i>Data Mining</i>. 4. Основні задачі аналізу даних. 	2
	<p>Список рекомендованих джерел:</p> <p>Основний: 1 [с. 14-39] Додатковий: 7 [с. 10-24]; 12 [с. 1-19, 169-195]; 13 [с. 29-112]; 15 [с. 695-710]; 19 [с. 45-56]; 20 [с. 1-7]; 22 [с. 1-29, 91-113]; 23 [с. 1-14]. Інтернет-ресурси: 29, 30, 33, 35, 37, 42-44.</p>	
	<p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Вступ до аналітики даних», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 1 [с. 14-39]; 7 [с. 10-24]; 12 [с. 1-19, 169-195]; 13 [с. 29-112]; 15 [с. 695-710]; 19 [с. 45-56]; 20 [с. 1-7]; 22 [с. 1-29, 91-113]; 23 [с. 1-14]; 29, 30, 33, 35, 37, 42-44.</p>	14
	<p>Практичні заняття 1-2 Створення набору даних з відкритих джерел</p> <p>Мета: формування набору даних для подальшого дослідження в різних аналітичних системах.</p>	4

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Здійснити аналіз основних світових та вітчизняних джерел відкритої інформації: <i>World Data Lab, Open Data Barometer, Open Datasets World Bank Open Data</i>, Єдиний державний веб-портал відкритих даних https://data.gov.ua/, Державна служба статистики України http://www.ukrstat.gov.ua/, Національні інформаційні системи https://nais.gov.ua/ тощо. 2. Сформувавши з відкритих баз даних, проаналізованих у п. 1, або іншого офіційного джерела відкритої інформації набір даних, який містить: <ul style="list-style-type: none"> - географічні дані (назви країн, міст, регіонів); - дані інших різних типів (дата/час, текстові, цілочислові, дійсні, логічні); - не менше 10 факторів (стовпців); - не менше 100 записів (рядків). 3. Зберегти сформований набір даних у форматах: <i>*.xlsx; *.csv; *.txt</i>. 4. У звіті до практичного завдання навести: <ul style="list-style-type: none"> - посилання на джерело інформації, звідки було сформовано набір даних; - короткий загальний опис набору даних; - перелік факторів (полів, стовпців) набору даних із вказанням типу даних у них. 	

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p>Знати: базові принципи роботи в аналітичній системі R</p> <p>Вміти: завантажувати дані в систему R, проводити аналітичну обробку масивів даних, представлених у вигляді <i>Data Frame</i></p>	<p>Тема 2. Статистична обробка інформації в аналітичній системі R План лекції 3</p> <ol style="list-style-type: none"> 1. Десктопні та хмарні рішення R (<i>RStudio, Posit Cloud</i>). 2. Типи даних у R. 3. Імпорт даних в середовище R. 4. Первинна обробка даних. <p>Список рекомендованих джерел: Основний: 4 [с. 3-13]. Додатковий: 15 [с. 29-72]; 20 [с. 61-78]. Інтернет-ресурси: 39-41.</p> <p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Статистична обробка інформації в аналітичній системі R», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 4 [с. 3-13]; 15 [с. 29-72]; 20 [с. 61-78]; 39-41.</p> <p><i>На основі набору даних, сформованого при вивченні теми 1, провести первинну обробку даних у середовищі R</i></p> <ol style="list-style-type: none"> 1. Завантажте файл у форматі <i>csv</i> з власного датасету, імпортуйте його в R з використанням функції <i>read.csv()</i>. 2. Переконайтеся в коректності процесу імпорту. Виведіть за допомогою функції <i>View</i> повний імпортований датасет. Виведіть у консоль перші та останні 50 записів датасету. 3. Сформууйте датасет <i>ds100</i>, який містить перші 100 записів і обрані самостійно 5 його полів. 4. Оберіть одне із символічних полів та впорядкуйте датасет <i>ds100</i> за цим полем в алфавітному порядку. 5. Оберіть одне із числових полів та впорядкуйте датасет <i>ds100</i> за цим полем у порядку спадання його значень. 6. Оберіть одне із символічних полів датасету <i>ds100</i> і виведіть у консоль всі унікальні значення цього поля. 7. Оберіть одне із значень символічного поля і відфільтруйте з датасету <i>ds100</i> рядки, які містять це текстове значення. 	<p>4</p> <p>8</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>8. Застосуйте на свій розсуд декілька фільтрів датасету <i>ds100</i> одночасно. Комплексний фільтр повинен містити принаймні одну умову з текстовим значенням і одну умову з числовим значенням та логічними операціями порівняння.</p> <p>9. Визначте середнє значення за одним із числових стовпців датасету <i>ds100</i>.</p> <p style="text-align: center;">Практичне заняття 3 Базова аналітика у системі R</p> <p>Мета: вивчити основні структури даних в R, засоби введення, імпорту та первинної обробки наборів даних.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Завантажити дані із заданого файлу. 2. Вивести імпортований датасет на екран і додатково вивести його в окремому вікні з використанням функції перегляду <i>View</i>. 3. Вивести 8 перших рядків датасету. 4. Вивести 10 останніх рядків в окремому вікні з використанням функції <i>View</i>. 5. Визначити вимірність набору даних (кількість рядків і стовпців) за допомогою функції <i>dim</i>. 6. Визначити окремо кількість рядків та стовпців, використавши функції <i>ncol()</i> та <i>nrow()</i>. 7. Отримати повний технічний опис датасету (функція <i>str</i>). 8. Перевірити, чи містить датасет пропущені значення. Заповнити пропущені значення параметрами <i>NA</i>. 9. Підрахувати кількість рядків датасету, що містять значення <i>NA</i> з використанням функцій <i>complete.cases (dataset)</i> та <i>is.na()</i>. 10. Підключити бібліотеки <i>VIM</i> та <i>misc</i>. Застосувати функції <i>aggr()</i> бібліотеки <i>VIM</i> та <i>matrixplot</i> бібліотеки <i>misc</i> для візуалізації пропущених значень датасету. 	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p>Знати: принципи візуалізації даних, графічні бібліотеки R</p> <p>Вміти: застосовувати графічні можливості системи R, керувати параметрами візуалізації</p>	<p style="text-align: center;">Тема 3. Розвідувальна аналітика в системі R План лекції 4</p> <ol style="list-style-type: none"> 1. Огляд графічних систем R. 2. Базові методи візуалізації даних у R (графіки, гістограми, стовпчикові та кругові діаграми). 3. Розвинуті графічні методи (корелограми, спінограми, скрипкові та мозаїчні діаграми, діаграми розсіювання, мультиграфіки). 4. Вибір оптимальної форми візуалізації даних. <p style="text-align: center;">Список рекомендованих джерел:</p> <p>Основний: 4 [с. 14-230]. Додатковий: 15 [с. 115-240]. Інтернет-ресурси: 39-41.</p> <p>Самостійна робота студентів. Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Розвідувальна аналітика в системі R», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 4 [с. 14-230]; 15 [с. 115-240]; 39-41.</p> <p><i>На основі набору даних, сформованого при вивченні теми 1, провести первинну обробку даних у середовищі R</i></p> <ol style="list-style-type: none"> 1. Завантажте файл у форматі csv з власного датасету, імпортуйте його в R з використанням функції read.csv(). 2. Завантажте файл у форматі csv з власного датасету, сформованого при виконанні практичного завдання 1, та імпортуйте його в R з використанням функції read.csv(). 3. Точковий графік. Оберіть два числових поля, які мають залежні значення і побудуйте точковий графік залежності одного параметру від іншого у пакеті base. 4. Лінійний графік. Оберіть декілька числових полів, які мають залежні значення і побудуйте в одній системі координат для кожної пари факторів лінійні графіки залежності одного параметру від іншого у пакеті base. 5. Гістограма. Оберіть текстове поле (категорію). Побудуйте гістограму, в якій горизонтальна вісь містить назви категорій, а вертикальна містить частоту значень кожної категорії у відсотках. 	<p style="text-align: center;">2</p> <p style="text-align: center;">12</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p>Вміти: створювати та форматовувати звіти у системі <i>Tableau Desktop</i>, створювати хмарні додатки у середовищі <i>Tableau Public</i></p> <p>інтерактивні візуальні компоненти <i>Power BI Desktop</i>, публікувати аналітичні звіти на платформі <i>Power BI Service</i>, створювати вбудований код звіту на веб-ресурсах</p>	<p>2. Візуалізатори системи <i>Tableau</i>. 3. Дашборди в <i>Tableau</i>. 4. Практичне застосування аналітичної платформи <i>Tableau</i>.</p> <p>Список рекомендованих джерел: Основний: 5 [с. 2-80] Додатковий: 10 [с. 83-133] Інтернет-ресурси: 31, 32, 36.</p> <p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Аналітична платформа <i>Tableau</i>», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 5 [с. 2-80]; 10 [с. 83-133]; 31, 32, 36. Обрати на веб-ресурсі порталу відкритих даних https://tradingeconomics.com набір економічних показників України: https://tradingeconomics.com/ukraine/indicators. Завантажити набір даних до платформи <i>Tableau</i> за прямим доступом до відповідної веб-сторінки. Побудувати звіт з використанням різних типів стовпчикових та лінійних діаграм, секторних, кільцевих та бульбашкових діаграм, карток та багаторядкових карток, карт. Здійснити візуальну фільтрацію даних за трьома обраними показниками.</p> <p>Практичні заняття 5-6 Аналіз даних у системі <i>Tableau</i> Мета: імпорт, моделювання, передобробка, аналіз та візуалізація даних у системі <i>Tableau</i>.</p> <p>Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Ознайомтеся із екосистемою <i>Tableau</i> на сайті додатку https://www.tableau.com 2. Встановіть додаток <i>Tableau Desktop</i> та активуйте безкоштовну <i>trial-версію</i> програми: https://www.tableau.com/products/desktop та ознайомтеся з інтерфейсом <i>Tableau Desktop</i>. 3. Завантажте базу даних із заданого файлу. 4. Створіть модель даних. Налаштуйте зв'язки між таблицями бази даних. 5. Перевірте коректність розподілу завантажених даних на категорії <i>Dimensions</i> та <i>Measures</i>. 	<p>12</p> <p>4</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>6. Обчисліть загальну кількість значень вказаного параметру і побудуйте лінійчасту діаграму за вказаний період.</p> <p>7. Обчисліть загальну кількість записів бази даних за вказаним параметром і побудуйте лінійчасту діаграму, впорядкувавши її за спаданням значення параметру.</p> <p>8. Побудуйте бульбашкову діаграму за вказаним параметром.</p> <p>9. Створіть карту і позначте на ній області, які відповідають географічним даним датасету.</p> <p>10. Створіть дерево карти (treemap) за вказаним параметром. Значення параметру потрібно вказати у центрі кожної прямокутної області treemap.</p> <p>11. Створіть дашборд з чотирма візуалізаціями, створеними у завданнях 7-11 (на вибір)</p> <p>12. Створіть таблицю з розподілом значень вказаного у завданні параметру і відповідного відсотку серед загальної кількості значень.</p> <p>13. Створіть дашборд, поєднавши таблицю з однією зі створених візуалізацій (на вибір).</p> <p>14. Зареєструйтеся в системі Tableau Public (https://public.tableau.com/authoring)</p> <p>15. Виконайте завдання 5-13 у системі Tableau Public.</p>	
<p>Знати: основні інструменти візуалізації аналітичних звітів, засоби захисту й публікації звітів, оновлення даних у побудованих звітах.</p> <p>Вміти: створювати та формувати інтерактивні візуальні компоненти <i>Power BI Desktop</i>, публікувати аналітичні звіти на платформі</p>	<p>Тема 5. Аналітика в системі Microsoft Power BI</p> <p>План лекції 7</p> <ol style="list-style-type: none"> 1. Інтерфейс та налаштування середовища <i>Power BI Desktop</i>. 2. Завантаження даних <i>Power BI</i> з різних інформаційних джерел. 3. Редактор запитів <i>Power Query</i>. 4. Трансформація, очищення та фільтрування даних у <i>Power BI</i>. <p>План лекції 8</p> <ol style="list-style-type: none"> 1. Принципи створення інтерактивних візуалізацій в системі <i>Power BI Desktop</i> 2. Базові візуальні елементи для побудови дашбордів аналітичних даних. 3. Публікація звітів та створення інформаційних панелей на порталі <i>Power BI</i>. 4. Просунута аналітика даних у <i>Power BI Desktop</i>. <p>Список рекомендованих джерел:</p>	<p>2</p> <p>2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><i>Power BI Service</i>, створювати вбудований код звіту на веб-ресурсах</p>	<p>Основний: 3 [с. 1-46, 173-209]. Додатковий: 19 [с. 70-153]. Інтернет-ресурси: 28, 34.</p> <p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Аналітика в системі Microsoft Power BI», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 3 [с. 1-46, 173-209]; 19 [с. 70-153]; 28, 34. Знайти серед веб-ресурсів порталу відкритих даних https://data.gov.ua набір даних за однією з груп: будівництво, економіка та бізнес, освіта і культура, сільське господарство, транспорт, держава, земля, охорона здоров'я, соціальний захист, фінанси, екологія, молодь і спорт, податки, стандарти, юстиція. Завантажити набір даних до платформи <i>Power BI</i> за прямим доступом до відповідної веб-сторінки. Провести трансформацію даних, залишивши тільки стовпці, які будуть використані для дослідження. Побудувати модель завантажених даних. Додати до існуючих таблиць в моделі принаймні один додатковий обчислювальний стовпець і одну міру. Побудувати звіт з використанням базових елементів візуалізації даних.</p> <p style="text-align: center;">Практичне заняття 7 Створення та аналіз бази даних в системі Power BI Desktop</p> <p>Мета: набуття базових навичок роботи в аналітичній платформі Power BI.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. За даними Державної служби статистики України http://www.ukrstat.gov.ua сформувати базу даних щодо цін за оренду однокімнатної квартири у м. Києві. 2. Здійснити реєстрацію у програмі Microsoft Power BI Desktop, використовуючи корпоративний акаунт @knu.edu.ua. Безпосередньо в аналітичній платформі сформувати інформаційну базу, використовуючи дані за останні 24 місяці. Налаштувати необхідний формат даних таблиці (ціле число, дата, текстове поле, грошові одиниці). 3. Провести візуалізацію даних, використовуючи такі інструменти: <ul style="list-style-type: none"> - графік; - стовпчикова діаграма; - кругова діаграма. 	<p style="text-align: center;">12</p> <p style="text-align: center;">2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>4. Змінити кольори елементів візуалізацій, шрифт і стиль надписів, встановлених за замовченням. Продемонструвати інтерактивність візуалізацій, по черзі виділивши їх окремі компоненти. Побудувати агреговані таблиці, здійснивши ієрархію дат до рівня кварталу та року. Таблиці розташувати на окремих аркушах в Power BI. Використавши панель аналітики, за даними таблиці.</p> <p>5. Встановити на графіку лінії максимуму, мінімуму та середнього значення.</p> <p>6. Побудувати прогноз ціни оренди однокімнатної квартири на наступні 6 місяців.</p> <p style="text-align: center;">Практичне заняття 8 Аналіз даних із веб-ресурсів</p> <p>Мета: отримання практичних навиків обробки даних із веб-ресурсів з використанням карт.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Завантажити дані про переможців Чемпіонату Європи з футболу УЄФА, які можна знайти в таблиці результатів на сторінці Вікіпедії за адресою http://en.wikipedia.org/wiki/UEFA_European_Football_Championship 2. Зробити попередню очистку даних перед завантаженням їх у платформу. Відкрити редактор Power Query та видалити всі непотрібні стовпці, залишивши тільки Year (Рік) і Final Winners (Переможці фіналу). З використанням меню Replace виконайте заміну значень, видаливши слово «Details» з усіх рядків стовпця Year. За допомогою пункту Rename залишити у стовпці Final Winner тільки переможців фіналів, яким можна присвоїти ім'я Country (Країна). 3. Створити візуалізацію звіту з використанням інструменту Map. 4. Змінити величину маркера країни на карті, щоб розмір маркера відображав кількість перемог для кожної країни. 5. Розфарбувати точку даних Germany в різні кольори, щоб розділити переможців з ФРН та об'єднаної Німеччини. Змінити колір та стиль заголовку карти. 6. Побудувати кругову діаграму, що відображає внесок кожної країни у загальну кількість перемог. 7. Провести сортування секторів діаграми так, щоб країни найбільшою кількістю перемог відображались на круговій діаграмі поруч. 	4

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p style="text-align: center;">Практичні заняття 9-10</p> <p style="text-align: center;">Створення багатосторінкових звітів у системі Power BI</p> <p>Мета: створення системи навігації багатосторінкового звіту, стандартних візуальних зображень, візуалізацій на основі карти, побудова аналітичного прогнозу.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Побудувати аналітичний звіт за даними датасету про кількість туристів, що перебували у колективних засобах розміщення (КЗР) в різних регіонах України. 2. Звіт має містити 4 сторінки з наступним змістом: <ol style="list-style-type: none"> 1) перша сторінка має містити головне меню із гіперпосиланнями-переходами на інші сторінки звіту; 2) на другій сторінці потрібно відобразити карту України з можливістю обрати один або декілька регіонів і відображенням інформації про кількість туристів у КЗР у відповідних регіонах України; 3) на третій сторінці потрібно побудувати графіки залежності кількості туристів у КЗР у регіонах за роками та кругову діаграму частки туристів у регіонах у загальній кількості туристів, що перебували у КЗР України; 4) четверта сторінка має містити прогноз кількості туристів у КЗР обраного регіону на наступні три роки. 3. Змінити параметри візуалізацій, які пропонуються за замовченням у Power BI. 	4
<p style="text-align: center;">Знати:</p> <p>принципи та критерії оптимальності консолідації даних, методи оцінювання якості даних, алгоритми очищення та збагачення даних.</p>	<p style="text-align: center;">Тема 6. Консолідація та передобробка даних</p> <p style="text-align: center;">План лекції 9</p> <ol style="list-style-type: none"> 1. Поняття консолідації. Схема процесу консолідації. 2. Технології та методи оцінки якості даних. 3. Очищення та збагачення даних. 4. Основні методи трансформації даних. <p style="text-align: center;">Список рекомендованих джерел:</p> <p>Основний: 4 [с. 5-13]. Додатковий: 20 [с. 109-119]. Інтернет-ресурси: 37, 38.</p>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Консолідація та передобробка даних», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 4 [с. 5-13]; 20 [с. 109-119]; 37, 38.</p> <p>Підготувати наукове повідомлення (на 5-7 хв.) або презентацію на одну із запропонованих тем:</p> <ol style="list-style-type: none"> 1. Архітектури сховищ даних. Спеціалізовані сховища даних. 2. ETL-процес. 3. OLAP-системи. Схеми побудови сховищ даних. 4. Очищення від шумів і згладжування рядів даних. 5. Фільтрація даних. Відновлення пропущених значень. 6. Редагування аномальних значень. Методи корегування аномальних значень. 7. Проблема репрезентативності. Обробка дублікатів і протиріч. 8. Семплінг. Усунення незначущих факторів 	14
<p>Знати: основні методи класифікації даних, регресії та прогнозування даних, логістичної регресії</p> <p>Вміти: ідентифікувати задачі, які можуть бути розв’язані методами класифікації даних, методами регресійного аналізу та прогнозування даних з використанням лінійної, нелінійної, динамічної та логістичної регресії.</p>	<p style="text-align: center;">Тема 7. Класифікація, регресія та прогнозування даних</p> <p style="text-align: center;">План лекції 10</p> <ol style="list-style-type: none"> 1. Огляд методів класифікації. 2. Точність класифікації. 3. Оцінювання класифікаційних методів. 4. Байєсівська класифікація. <p style="text-align: center;">План лекції 11</p> <ol style="list-style-type: none"> 1. Діагностика та відбір регресійних моделей даних. 2. Логістична регресія. 3. Методи прогнозування даних. 4. Дерева рішень. <p style="text-align: center;">Список рекомендованих джерел:</p> <p>Основний: 2 [с. 135-204]. Додатковий: 10 [с. 393-449, 571-611]; 13 [с. 311-382]; 23 [с. 516-524].</p>	2 3

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>яким можливо провести кластеризацію його значень. Обґрунтувати вибір міри відстані при кластеризації значень показника та кількість кластерів. Провести кластеризацію значень показника на основі алгоритму <i>k-means</i>.</p> <p style="text-align: center;">Практичні заняття 11-12 Принципи кластеризації даних. Алгоритм <i>k-means</i>.</p> <p>Мета: набуття практичних навичок роботи з алгоритмом кластеризації <i>k-means</i>.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Дано 10 об'єктів (A, B, C, D, E, F, G, H, I, K), які представлено координатами точок на площині. 2. Використовуючи алгоритм <i>k-means</i>, провести кластерний аналіз даних, розбивши початкову множину точок на 2 кластери. За початкові центри кластерів прийняти точки E і J. 3. Використовуючи алгоритм <i>k-means</i>, провести кластерний аналіз даних, розбивши початкову множину точок на 3 кластери. За початкові центри кластерів прийняти точки A, F і J. 4. Порівняти отримані результати і обрати оптимальну кількість кластерів. 5. На кожному етапі кластерного аналізу провести візуалізацію результатів розбиття на кластери. 6. Для визначення належності об'єкта до певного кластеру використати евклідову відстань та відстань Манхеттена. 	4
<p>Знати: основи афінитивного аналізу, принципи генерації популярних наборів даних та асоціативних правил.</p> <p>Вміти: використовувати алгоритм <i>a priori</i> для побудови та аналізу асоціативних правил</p>	<p style="text-align: center;">Тема 9. Асоціативний аналіз даних План лекції 13</p> <ol style="list-style-type: none"> 1. Афінитивний аналіз. 2. Основні поняття <i>Rules Mining</i>. 3. Значущість асоціативних правил. 4. Програмі засоби пошуку асоціативних правил <p style="text-align: center;">Список рекомендованих джерел:</p> <p>Основний: 2 [с. 71-89] Додатковий: 6 [с. 38-41]; 10 [284-287]; 20 [с. 205-216]; 23 [с. 535-538] Інтернет-ресурси: 37, 38.</p>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Асоціативний аналіз даних», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 2 [с. 71-89]; 6 [с. 38-41]; 10 [284-287]; 20 [с. 205-216]; 23 [с. 535-538]; 37, 38.</p> <p><i>На основі набору даних, сформованого при вивченні теми 1, згенерувати популярні набори значень показника, побудувати систему асоціативних правил та визначити їх кількісні характеристики (підтримку, достовірність, ліфт, леввередж, покращення). Виокремити множину сильних правил за показником добутку підтримки та достовірності правила.</i></p> <p style="text-align: center;">Практичні заняття 13-14</p> <p style="text-align: center;">Створення асоціативних правил. Алгоритм <i>Apriori</i></p> <p>Мета: набуття практичних навичок формування популярних наборів та генерації асоціативних правил та на основі алгоритму <i>Apriori</i>.</p> <p style="text-align: center;">Завдання до практичного заняття</p> <ol style="list-style-type: none"> 1. Роздрібна мережа по продажу товарів побутової хімії поставила задачу аналізу купівельних кошиків для оптимізації розміщення товарів на вітринах і проведення крос-продажів. Відділ маркетингу надав 5000 чеків, у яких відображені покупки, зроблені клієнтами магазинів. 2. Передбачити, які товари покупці можуть вибрати в залежності від того, що вже є в їхніх кошиках; виявити найбільш популярні товарні набори, які складаються з більше ніж одного предмету; запропонувати рекламні акції типу: «Кожному, хто купив A і B – товар C в подарунок». 3. Використовуючи алгоритм <i>apriori</i>, побудувати асоціативні правила і зробити їх інтерпретацію. 4. Сформувані власну базу даних для подібної задачі, провести генерацію популярних наборів і визначити систему асоціативних правил. Серед початкових даних повинно бути: не менше 20 різних видів товару; кожен чек повинен містити від 2 до 4 різних видів товару; кількість чеків не менше 50. 	<p style="text-align: center;">10</p> <p style="text-align: center;">4</p>
<p>Знати: принципи візуального аналізу</p>	<p style="text-align: center;">Тема 10. Технології інтелектуальної обробки даних План лекції 14</p>	<p style="text-align: center;">2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p>даних (<i>Visual Mining</i>), аналізу текстової інформації (<i>Text Mining</i>), аналізу даних в режимі реального часу (<i>Real-Time Mining</i>), аналізу даних з мережі Інтернет (<i>Web Mining</i>).</p>	<ol style="list-style-type: none"> 1. Візуальний аналіз даних – <i>Visual Mining</i>. 2. Задача та етапи аналізу текстових даних – <i>Text Mining</i>. 3. Ідея <i>Data Mining</i> у реальному часі – <i>Real-Time Mining</i>. 4. Аналіз даних з мережі Інтернет – <i>Web Mining</i> <p>Список рекомендованих джерел: Основний: 1 [с. 42-66, 446-460]. Додатковий: 7 [с. 84-95]; 10 [с. 287-294]; 13 [с. 383-404]; 15 [с. 483-495]; 20 [с. 121-133]; 22 [с. 197-235]. Інтернет-ресурси: 38.</p> <p>Самостійна робота студентів. Опрацювання матеріалу лекції з теми «Технології інтелектуальної обробки даних», підготовка до тестування, ознайомлення з основними та додатковими джерелами, інтернет-ресурсами: 1 [с. 42-66, 446-460]; 7 [с. 84-95]; 10 [с. 287-294]; 13 [с. 383-404]; 15 [с. 483-495]; 20 [с. 121-133]; 22 [с. 197-235]; 38.</p> <p>Підготувати наукове повідомлення (на 5-7 хв.) або презентацію на одну із запропонованих тем:</p> <ol style="list-style-type: none"> 1. Деревовидні карти та канонічні дерева у процесі <i>Visual Mining</i>. 2. Рекомендаційні системи зі штучним інтелектом. 3. Веб-системи аналітики даних у режимі реального часу. 4. Інтелектуальний аналіз текстових даних. 5. Інструментальні засоби добування інформації з Інтернет-джерел. 	10
<i>Підсумковий контроль – екзамен</i>		
Разом		180

Примітка:

курсивом позначено інтерактивні методи навчання

3. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основний

1. Талах М.В. Технології обробки Big Data. Навчальний посібник/ М.В. Талах – Чернівці: Чернівецький нац. ун-т, 2024. – 454 с.
2. Черноус Г., Фаренюк Я., Діденко І. Дата майнінг для економістів : навч. посіб (англ. мовою). Київ: Видавництво «Ліра-К», 2023. 290 с.
3. Clark D. *Beginning Microsoft Power BI. A Practical Guide to Self-Service Data Analytics.* – Apress, 2020. – 417 p.
4. Kabacoff R. *Modern Data Visualization with R.* – CRC Press, 2024. – 272 p.
5. Meier M., Baldwin D. *Mastering Tableau 2021. Implement advanced business intelligence techniques and analytics with Tableau Packt Publishing, 2021.* – 794 p.

Додатковий

6. Гороховатський В.О., Творошенко І.С. *Методи інтелектуального аналізу та оброблення даних: навч. посібник.* – Харків: ХНУРЕ, 2021. 92 с.
7. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань: Навч. посібник. Львів: «Магнолія-2006», 2021. 276 с.
8. Bell R. *Big data in travel, consumption and marketing: big data for business.* Independently published, 2021. – 142 p.
9. Buisson F. *Behavioral Data Analysis with R and Python. Customer-Driven Data for Real Business Results.* – O'Reilly, 2021. – 360 p.
10. Camm J. D., Cochran J. J., Fry M. J., Ohlmann J. W. *Business Analytics.* – Cengage AU, 2024. – 1037 p.
11. Chinnici M., Pop F., Negru C. *Data Science and Big Data Analytics in Smart Environments.* CRC Press, 2021. – 292 p.
12. Consoli S., Recupero D. R., Saisana M. *Data Science for Economics and Finance: Methodologies and Applications.* Springer, 2021. – 369 p.
13. Evans J. R. *Business Analytics: Methods, Models, and Decisions: Pearson, 2021.* 705 p.
14. Ferrar J., Green D. *Excellence in People Analytics: How to Use Workforce Data to Create Business Value.* Kogan Page, 2021. – 384 p.
15. Irizarry A. *Introduction to Data Science. Data Analysis and Prediction Algorithms with R: Chapman and Hall/CRC, 2020.* 743 p.
16. Kakulapati V. *Open Data: ITeXLi, 2022.* 77 p.
17. Kempson S. *Data Analysis And Business Performance: The Basics Of Data Analysis And What You Should Know: How Does Data Analysis Help Business.* Kindle Edition, 2021. – 360 p.
18. Landkamer T. *Utilizing Data For Growth Of Your Business: An Essential Guide To Data Analysis Everyone Must Know: How Can Data Improve A Business?* Kindle Edition, 2021. – 360 p.
19. Larson B. *Data Analysis with Microsoft Power BI: Mc Graw Hill, 2020.* 741 p.
20. Mariani M. C., Tweneboah O. K., Beccar-Varela M. P. *Data Science in Theory and Practice: Techniques for Big Data Analytics and Complex Data Sets: John Wiley & Sons, Inc., 2022.* 403 p.
21. Matloff N. *Probability and Statistics for Data Science: Math + R + Data.* – London: Chapman & Hall, 2019. – 376 p.
22. Mishra B.K., Kumar V., Panda S.K., Tiwari P. *Handbook of Research for Big Data. Concepts and Techniques: CRC-Press, 2022.* 389 p.
23. Ragsdale C. T. *Spreadsheet Modeling and Decision Analysis. A Practical Introduction to Business Analytics: Cengage, 2022.* 908 p.
24. Shmueli G., Bruce P., Yahav I., Patel N., Lichtendahl K. *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R.* – New Jersey: Willey, 2018. – 544 p.
25. Stephenson D., Elder J. *Business Skills for Data Scientists: Practical Guidance in Six Key Topics.* Data Science Innovation, 2021. – 304 p.

26. Wade R. Advanced Analytics in Power BI with R and Python: Ingesting, Transforming, Visualizing. Apress, 2020. – 440 p.
27. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020. – 298 p.

Internet-ресурси

28. Основи роботи з Power BI // URL: <https://docs.microsoft.com/ru-ru/power-bi/desktop-data-sources>
29. Портал відкритих даних. URL: <https://data.gov.ua>
30. Система аналітики відкритих даних Clarity Project. URL: <https://clarity-project.info>
31. Старовойтенко К. Побачити та зрозуміти дані. Гайд по роботі з Tableau. URL: <https://dou.ua/forums/topic/39355/>
32. Clustering in Tableau – Learn the Steps to Perform it Easily. URL: <https://data-flair.training/blogs/clustering-in-tableau/>
33. Earth Data. Open Access for Open Science. URL: <https://www.earthdata.nasa.gov>
34. Microsoft Power BI Guided Learning [Електронний ресурс]. – Режим доступу: <https://docs.microsoft.com/uk-ua/power-bi/guided-learning>
35. Open Data Soft Platform. URL: <https://www.opendatasoft.com/>
36. Pandey P. Cluster Analysis in Tableau. Learn how to cluster your data in Tableau easily. URL: <https://towardsdatascience.com/cluster-analysis-in-tableau-1f19acd0c647>
37. Rancher A. An Introduction to Big Data Concepts URL: https://www.suse.com/c/rancher_blog/an-introduction-to-big-data-concepts/
38. RapidMiner Academy: learn data science and RapidMiner from leading industry experts. URL: <https://academy.rapidminer.com/>
39. RStudio-education. URL: <https://github.com/rstudio-education>
40. RStudio. URL: <https://www.rstudio.com/products/rstudio/download/#download>
41. The Comprehensive R Archive Network. URL: <https://cran.r-project.org/>
42. The official portal for European data. URL: <https://data.europa.eu/>
43. The Open Data World Portal. URL: <https://data.world>
44. World Data Lab. URL: <https://worlddata.io/>

* Курсивом виділені джерела, що є у бібліотеці ДТЕУ, або наявні повнотекстові електронні версії джерел