

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ 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  
(пост. п. від 2021 р.)  
Ректор



А. А. Мазаракі

**АНАЛІТИКА BIG DATA /  
BIG DATA ANALYTICS**

**РОБОЧА ПРОГРАМА /  
COURSE OUTLINE**

освітній ступінь	Магістр / Master
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	124 Системний аналіз / System Analysis
спеціалізація	Інформаційні технології та бізнес-аналітика (Data Science) / Information Technologies and Business Analytics (Data Science)

**Київ 2021**

**Розповсюдження і тиражування без офіційного дозволу КНТЕУ заборонено**

**Автор:** А.А. РОСКЛАДКА, докт. екон. наук, проф.

Розглянуто і схвалено на засіданні кафедри цифрової економіки та системного аналізу «25» травня 2021 р., протокол №10.

**Рецензент:** В. В. Кулаженко, кандидат економічних наук, доцент, доцент кафедри цифрової економіки та системного аналізу

**АНАЛІТИКА BIG DATA /  
BIG DATA ANALYTICS**

**РОБОЧА ПРОГРАМА /  
COURSE OUTLINE**

<b>освітній ступінь</b>	<b>магістр / master</b>
<b>галузь знань</b>	<b>12 Інформаційні технології / Information Technologies</b>
<b>спеціальність</b>	<b>124 Системний аналіз / System Analysis</b>
<b>освітня програма</b>	<b>Інформаційні технології та бізнес-аналітика (Data Science) / Information Technologies and Business Analytics (Data Science)</b>

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

Для вивчення дисципліни відводиться 180 год / 6 кредитів ЄКТС (лекцій – 20 год., лабораторних занять – 42 год., самостійної роботи – 118 год.), підсумковий контроль – письмовий екзамен.

Назва теми	Кількість годин				Форми контролю*
	Усього годин/ кредитів	із них:			
		лекції	лабора- торні заняття	самостій- на робота студентів	
<i>III семестр</i>					
Тема 1. Концепція <i>Big Data</i>	12	2	0	10	Т, ПР
Тема 2. Концепція <i>Open Data</i>	16	2	4	10	Т, ВПЗ, ВІЗ, ПР
Тема 3. Первинна обробка великих даних у середовищі <i>R</i>	20	2	6	12	Т, ВПЗ, ВІЗ
Тема 4. Графічний аналіз великих даних	20	2	6	12	Т, ВПЗ, ВІЗ
Тема 5. Створення інтерактивних веб-додатків для аналізу великих даних у <i>R Shiny</i>	24	2	8	14	Т, ВПЗ, ВІЗ
Тема 6. Реактивне програмування процедур обробки великих даних у <i>Shiny</i>	24	2	8	14	Т, ВПЗ, ВІЗ
Тема 7. Аналіз великих даних у <i>Power BI</i>	22	2	6	14	Т, ВПЗ, ВІЗ
Тема 8. Веб-скрапінг та парсінг великих даних	18	2	4	12	Т, ВПЗ, ВІЗ
Тема 9. Технології <i>MapReduce, Hadoop</i> і <i>Spark</i> в аналітиці великих даних	12	2	0	10	Т, ПР
Тема 10. Методи машинного навчання для надвеликих даних	12	2	0	10	Т, ПР
<i>Підсумковий контроль – письмовий екзамен</i>					
<b>РАЗОМ</b>	<b>180/6</b>	<b>20</b>	<b>42</b>	<b>118</b>	<b>x</b>

### Примітка\*:

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

## 2. ТЕМАТИКА ТА ЗМІСТ ЛЕКЦІЙНИХ, ЛАБОРАТОРНИХ ЗАНЯТЬ І САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Знати:</b> основні терміни та поняття аналітики великих даних</p> <p><b>Вміти:</b> виявляти характеристики та проблеми в обробці великих даних</p>	<p align="center"><b>Тема 1. Концепція <i>Big Data</i></b> <b>План лекції 1</b></p> <p>1. Історія розвитку <i>Big Data</i>. 2. Визначення та архітектура <i>Big Data</i>. 3. Методики аналізу великих даних. 4. Джерела великих даних.</p> <p align="center"><b>Список рекомендованих джерел:</b></p> <p>Основний: 1 [с. 1-22]; 4 [с. 7-24]. Додатковий: 9 [с. 7-17]; 12 [с. 29-59]; 18 [с. 1-7]; 20 [с. 157-197, 297-318]; 21 [с. 3-18]; 22; 25 [с. 9-20]. Інтернет-ресурси: 32; 34; 36; 37.</p> <p><b>Самостійна робота студентів.</b> Вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, підготовка презентації або наукового повідомлення 1 [с. 1-22]; 4 [с. 7-24]; 9 [с. 7-17]; 12 [с. 29-59]; 18 [с. 1-7]; 20 [с. 157-197, 297-318]; 21 [с. 3-18]; 22; 25 [с. 9-20]; 32; 34; 36; 37.</p> <p>Підготувати наукове повідомлення (на 5-7 хв.) або презентацію про концепцію <i>Big Data</i>:</p> <ul style="list-style-type: none"> <li>- Кліффорд Лінч – основоположник терміну <i>Big Data</i>;</li> <li>- розвиток поняття <i>Big Data</i> у сучасній науці про дані;</li> <li>- феноменальне прискорення нагромадження даних та їх ускладнення;</li> <li>- структуровані та неструктуровані великі дані;</li> <li>- проблеми надшвидкого оновлення великих даних;</li> <li>- приклади застосування великих даних у різних областях людської діяльності.</li> </ul>	<p align="center">2</p> <p align="center">10</p>



Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p style="text-align: center;"><b>Лабораторні заняття 1-2</b></p> <p style="text-align: center;"><b>Аналіз інформаційного забезпечення галузевої аналітики</b></p> <p><b>Мета:</b> дослідження джерел публічної інформації в різних галузях людської діяльності, організація автоматизованого доступу до джерел відкритих великих даних.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання:</i></p> <ul style="list-style-type: none"> <li>- здійснити аналіз основних світових джерел відкритої інформації на основі міжнародних рейтингів <i>Global Open Data Index</i> та <i>Open Data Barometer Growth from Knowledge (GFK)</i>;</li> <li>- проаналізувати складові частини світового порталу відкритих даних <i>Data.world</i>, завантажити набори великих даних у різних форматах представлення;</li> <li>- дослідити принципи організації доступу до відкритих даних в Україні (Державна служба статистики України, Єдиний державний веб-портал відкритих даних <i>Data.gov.ua</i>, <i>YouControl</i>, <i>SMIDA</i>, <i>ProZorro</i>).</li> </ul>	4
<p style="text-align: center;"><b>Знати:</b></p> <p>засоби представлення, основні типи та методи управління великими даними в середовищі <i>R</i></p>	<p style="text-align: center;"><b>Тема 3. Первинна обробка великих даних у середовищі <i>R</i></b></p> <p style="text-align: center;"><b>План лекції 3</b></p> <ol style="list-style-type: none"> <li>1. Імпорт великих даних у середовище <i>R</i>.</li> <li>2. Вирішення проблем з даними: пропущені значення, аномальні значення, дублікати, протиріччя, невірний формат представлення даних.</li> <li>3. Первинна обробка числових та нечислових (категоріальних, факторних, бінарних, логічних, рядкових) даних.</li> <li>4. Робота з каналами великих даних.</li> <li>5. Функції <i>R</i> для аналізу складних даних.</li> </ol>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Вміти:</b> використовувати пакети та функції <i>R</i> для опису структури набору великих даних; проводити описову аналітику та здійснювати перевірку статистичних гіпотез в середовищі <i>R</i></p>	<p><b>Список рекомендованих джерел:</b> Основний: 1 [с. 63-114]; 2 [с. 14-18]; 4 [с. 25-126]. Додатковий: 7 [с. 42-52]; 13 [с. 6-14, 38-70]; 14 [с. 29-72]; 15 [с. 30-80]; 18 [с. 61-78]; 24 [с. 3-13]; 25 [с. 39-80]. Інтернет-ресурси: 28; 29; 31.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 1 [с. 63-114]; 2 [с. 14-18]; 4 [с. 25-126]; 7 [с. 42-52]; 13 [с. 6-14, 38-70]; 14 [с. 29-72]; 15 [с. 30-80]; 18 [с. 61-78]; 24 [с. 3-13]; 25 [с. 39-80]; 28; 29; 31. <i>На основі набору великих даних, сформованого при вивченні теми 2, провести первинну обробку даних у середовищі RStudio з метою подальшого якісного аналізу.</i></p> <p><b>Лабораторне заняття 3</b> <b>Імпорт та попередній аналіз великих даних</b> <b>Мета:</b> вивчити основні структури даних в <i>R</i>, засоби введення, імпорту, анотування та акуратизації наборів великих даних. <i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i> <i>2. Практичні завдання: 1) організувати ведення даних в середовище <i>R</i> з використанням різних структур (vector, matrix, array, data frame, factor, list); 2) провести імпорт даних у середовище <i>R</i> з різних джерел (текстовий файл з роздільниками, файл Excel, XML-файл, файли імпорту із систем SPSS, SAS, Stata, netCDF, HDF5, файли СУБД ); 3) здійснити анотування імпортованих даних шляхом вказування заголовків факторів та пояснення</i></p>	<p>12</p> <p>2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>значень змінних; 4) провести фільтрацію рядків отриманих наборів даних з використанням функції <code>filter()</code>; 5) здійснити базові операції з імпортованим набором даних: переставлення рядків – функція <code>arrange()</code>, вибір стовпців – функція <code>select()</code>, додавання нових змінних – функція <code>mutate()</code>; 6) отримати підсумкові дані для введених та імпортованих наборів даних з використанням функції <code>summarize()</code>; 7) провести акуратизацію даних, звівши їх формату, придатному для подальшого аналізу, використовуючи функції пакету <code>tidyr</code>: <code>gather()</code> – зведення стовпців, <code>spread()</code> – розосередження стовпців, <code>separate()</code> – поділ стовпців, <code>unite()</code> – об'єднання стовпців.</p> <p style="text-align: center;"><b>Лабораторне заняття 4</b>  <b>Методи аналізу різних типів великих даних у R</b></p> <p><b>Мета:</b> отримання досвіду роботи з різними типами великих даних: реляційними, рядковими, факторними, даними дати і часу.</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичне завдання: 1) виконати основні операції з реляційними даними із застосуванням пакету <code>dplyr</code> (встановлення первинного, зовнішнього та сурогатного ключів, мінливих та фільтруючих з'єднань між таблицями); 2) здійснити операції над рядковими даними за допомогою пакету <code>stringr</code> (визначення довжини рядка – <code>str_length()</code>, об'єднання рядків – <code>str_c()</code>, витяг підрядка – <code>str_sub()</code>, перетворення регістрів – <code>str_to_lower()</code>, <code>str_to_upper()</code>, <code>str_to_title()</code>, сортування рядків – <code>str_sort()</code>, <code>str_order()</code>, пошук шаблонів в рядку – <code>str_detect()</code>); 3) виконати роботу з факторами у пакеті <code>forcast</code> (створення факторів – <code>factor()</code>, зміна порядку слідування факторів – <code>fct_reorder()</code>, зміна рівнів представлення факторів – <code>fct_relevel()</code>),</p>	2



Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p><i>впорядкування рівнів факторів – <code>ltr_infreq()</code>); 4) здійснити аналіз даних у форматі дати і часу за допомогою пакету <code>lubridate</code> (отримання поточної дати або дати/часу – <code>today()</code> і <code>now()</code>, створення значень дати/часу на основі рядків – <code>ymd()</code>, <code>mdy()</code>, <code>dmy()</code>, створення значень дати/часу на основі окремих компонентів – <code>make_date()</code>, <code>make_datetime()</code>, витяг окремих компонентів дати/часу – <code>year()</code>, <code>month()</code>, <code>mday()</code>, <code>yday()</code>, <code>wday()</code>, <code>hour()</code>, <code>minute()</code>, <code>second()</code>).</i></p> <p style="text-align: center;"><b>Лабораторне заняття 5</b></p> <p style="text-align: center;"><b>Методи роботи з пропущеними даними</b></p> <p><b>Мета:</b> виявлення, аналіз закономірностей та відновлення пропущених даних.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання: 1) використати функції <code>is.na()</code>, <code>is.nan()</code>, <code>is.infinite()</code> для виявлення пропущених, неприпустимих та нескінчених значень у даних; 2) визначення записів у наборі даних, що містять пропущені значення – <code>complete.cases()</code>; 3) провести візуальний аналіз пропущених даних за допомогою пакету <code>VIM</code> (функції <code>aggr()</code>, <code>matrixplot()</code>, <code>scattMiss()</code>); 4) виявити джерела пропущених даних; 5) застосувати метод множинного відновлення пропущених даних за допомогою пакету <code>misc</code>; 6) використати функції пакетів <code>Hmisc</code>, <code>Mvnmle</code>, <code>Cat</code>, <code>kmi</code>, <code>mix</code>, <code>rap</code> для реалізації попарного видалення та нестохастичного відновлення даних.</i></p>	2
<p><b>Знати:</b> методи візуалізації великих даних, засоби роботи з графічними пакетами <i>R</i></p>	<p style="text-align: center;"><b>Тема 4. Графічний аналіз великих даних</b></p> <p style="text-align: center;"><b>План лекції 4</b></p> <p>1. Огляд графічних систем <i>R</i>. 2. Інтерактивна візуалізація великих даних. 3. Базові методи візуалізації даних у <i>R</i>.</p>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Вміти:</b> використовувати пакети <i>Base, Grid, Lattice, ggplot2</i> та інші для проведення процедур візуалізації великих даних</p>	<p>4. Розвинуті графічні методи (корелограми, спінограми, скрипкові та мозаїчні діаграми, діаграми розсіювання, мультиграфіки).</p> <p>5. Вибір оптимальної форми візуалізації великих даних.</p> <p><b>Список рекомендованих джерел:</b> Основний: 1 [с. 377-393]; 3 [с. 42-155]. Додатковий: 13 [с. 172-195, 324-348]; 14 [с. 113-232]; 15 [с. 45-71, 119-140, 373-399]; 18 [с. 121-133]. Інтернет-ресурси: 28; 29; 31.</p> <p><b>Самостійна робота студентів.</b> Вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, підготовка презентації або наукового повідомлення 1 [с. 377-393]; 3 [с. 42-155]; 13 [с. 172-195, 324-348]; 14 [с. 113-232]; 15 [с. 45-71, 119-140, 373-399]; 18 [с. 121-133]; 28; 29; 31. <i>На основі набору великих даних, сформованого при вивченні теми 2, провести візуальний аналіз великих даних, застосувавши чотири графічних системи R: base, lattice, ggplot, grid.</i></p> <p><b>Лабораторне заняття 6</b> <b>Базові технології візуалізації даних</b> <b>Мета:</b> створення основних діаграм та налаштування їх параметрів. <i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i> <i>2. Практичні завдання: 1) завантажити вбудований датасет <i>mpg</i> і вивести інформацію з описом його полів і загальної структури; визначити неперервні та категоріальні змінні датасету; 2) створити точкову діаграму, що описує залежність змінної <i>hwy</i> від змінної <i>cy</i>; 3) змінити тип</i></p>	<p>12</p> <p>2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>символів та ліній; 4) налаштувати кольори складових частин діаграми; 5) встановити заголовки діаграми та осей; 6) додати текстову інформацію (легенду, анотацію) на діаграму; 7) здійснити об'єднання декількох діаграм в одну структуру.</p> <p style="text-align: center;"><b>Лабораторне заняття 7</b> <b>Візуальний аналіз взаємозв'язку великих даних</b></p> <p><b>Мета:</b> вивчення способів відображення взаємозв'язку між двома змінними (двовимірні взаємозв'язки) та декількома змінними (багатовимірні взаємозв'язки).</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичні завдання: 1) на основі даних датасету <i>mtcars</i> побудувати діаграму розсіювання з лініями найкращої апроксимації з використанням функцій <i>abline()</i> та <i>scatterplot()</i>; 2) створити матриці діаграм розсіювання з використанням функцій <i>pairs()</i>, <i>scatterplotMatrix()</i>, <i>cpairs()</i>; 3) для набору великих даних створити діаграму розсіювання високої щільності за допомогою функції <i>smoothScatter()</i>; 4) здійснити об'єднання даних в шестикутні комірки за допомогою функції <i>hexbin()</i>; 5) побудувати тривимірну діаграму розсіювання з використанням функції <i>scatterplot3d()</i>; 6) створити інтерактивні діаграми розсіювання з можливістю обертання (функція <i>plot3d()</i> пакету <i>rgl</i>, функція <i>scatter3d()</i> пакету <i>Rcmdr</i>); 7) створити корелограму для набору даних <i>mtcars</i> та дати інтерпретацію використання функції <i>corrgram()</i>; створити мозаїчні діаграми (функція <i>mosaic()</i> з пакету <i>vcd</i>).</p>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p style="text-align: center;"><b>Лабораторне заняття 8</b></p> <p style="text-align: center;"><b>Передові методи візуалізації великих даних</b></p> <p><b>Мета:</b> ознайомитися з передовими методами візуалізації великих даних з використанням систем <i>base, grid, lattice, ggplot2</i>.</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичні завдання: 1) на основі даних <i>singer</i> пакету <i>lattice</i> про зріст співаків та їх вокальних партій побудувати категоризовану діаграму для зросту виконавців різних вокальних партій; 2) використовуючи умовні змінні для <i>mtcars</i> та категоризовану діаграму для залежності витрат пального і маси автомобіля, візуалізувати залежність витрат пального від маси автомобіля з різним об'ємом двигуна; 3) додати до категоризованої діаграми регресійні лінії, «лінії-щітки» та координатну сітку; 4) створити діаграму ядерної оцінки функції щільності значень витрат пального для автомобілів з різними типами коробки передач; 5) побудувати діаграми розмахів для значень витрат пального автомобілями з різною кількістю циліндрів; 6) застосувати функції пакетів <i>plywith, latticist, iplots, rggobi</i> для створення інтерактивних діаграм з можливістю зміни параметрів в режимі редагування.</p>	2
<p style="text-align: center;"><b>Знати:</b></p> <p style="text-align: center;">базові принципи розробки веб-додатків з використанням пакету <i>Shiny</i></p>	<p style="text-align: center;"><b>Тема 5. Створення інтерактивних веб-додатків для аналізу великих даних у R Shiny</b></p> <p style="text-align: center;"><b>План лекції 5</b></p> <p>1. Огляд пакету <i>Shiny</i>.</p> <p>2. Базові операції з великими даних в межах односторінкового макету.</p> <p>3. Класифікація методів обробки великих даних для багатосторінкових макетів.</p> <p>4. Ровинутий інтерфейс користувача веб-додатку.</p>	2

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Вміти:</b> створювати веб-додатки для обробки великих даних у середовищі R</p>	<p>5. Програмні засоби створення інтерфейсу користувача. <b>Список рекомендованих джерел:</b> Основний: 5 [с. 24-49, 91-226]. Інтернет-ресурси: 27-31.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 5 [с. 24-49, 91-226]; 27-31.</p> <p><i>На основі набору великих даних, сформованого при вивченні теми 2, створити інтерактивний веб-додаток у середовищі R Shiny, завантажити у нього набір даних та здійснити аналіз текстової, графічної та табличної інформації про завантажений набір даних з використанням багатосторінкового шаблону веб-додатку.</i></p> <p><b>Лабораторні заняття 9-10</b> <b>Створення односторінкових макетів обробки великих даних</b> <b>Мета:</b> набуття навичок у створенні стандартних односторінкових макетів обробки великих даних у пакеті Shiny.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання: 1) Створити додаток Shiny, який відображає набір даних Orange у таблиці даних зі значеннями за замовченням; 2) Видалити назви рядків у наборі даних; 3) показати лише 7 рядків і дозволити змінити кількість відображених рядків на 14, 21, 28 або 35; 4) вирівняти текст стовпців (як значення, так і заголовки) по центру; 5) видалити поле пошуку (верхнє праворуч) і текст інформації про таблицю (нижнє ліворуч); 6)</i></p>	<p>14</p> <p>4</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>додати кнопки «Копіювати» та «csv», щоб дозволити зберегти таблицю в буфер обміну та завантажити її як csv-файл відповідно; 7) додати поле фільтра для кожного стовпця внизу таблиці; 8) дозволити вибір лише одного рядка, а не кількох рядків; 9) видалити опцію сортування таблиці; 10) відобразити в інтерфейсі користувача (<i>textOutput</i>) вибрані значення <i>tree</i> та <i>age</i>.</p> <p style="text-align: center;"><b>Лабораторні заняття 11-12</b></p> <p style="text-align: center;"><b>Створення багатопанельних макетів обробки великих даних</b></p> <p><b>Мета:</b> ознайомитися з основними методами створення головної та бічних панелей управління у додатку <i>Shiny</i>, панелей для виведення текстової, табличної та графічної інформації для розширеної аналітики великих даних.</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичні завдання: 1) створити сторінку <i>Navbar</i> із трьома панелями з використанням функції <i>navbarPage()</i>, яка створює програму зі стандартною панеллю <i>Navbar</i> у верхній частині. Назвіть додаток «Cars», а панелі «Plot», «Summary» та «Other» відповідно; 2) додати бічну панель у вкладці «Plot»; 3) додати на бічній панелі перемикач, за допомогою якого користувач зможе вибрати точкову діаграму або лінійну діаграму; 4) додати <i>radioButtons()</i> із двома кнопками всередині бічної панелі для перемикання між точковою та лінійною діаграмою; 5) у <i>mainPanel()</i> створити два графіки з використанням функції <i>plotOutput()</i>; 6) у <i>tabPanel()</i> «Summary» за допомогою функцій <i>verbatimTextOutput()</i> та <i>renderPrint()</i> вивести підсумкову статистику набору даних автомобілів; 7) у першій додатковій вкладці <i>tabPanel()</i> «Table» розмістити таблицю даних вашого набору даних, використовуючи функції <i>datatableOutput()</i> і <i>renderDataTable()</i>;</p>	4

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>8) на панелі «Description» описати заданий набір даних, який містить швидкість автомобілів і відстань, пройдену до зупинки; 9) перетворити програму за допомогою <code>navlistPanel()</code> замість <code>navbarPage()</code>; 10) створити чотири <code>headerPanel()</code> по одній на кожену <code>tabPanel</code> та назвати їх як відповідні <code>tabPanel</code>. Використати <code>position</code> у своєму <code>sidebarLayout</code>, щоб розмістити бічну панель зліва направо.</p>	
<p><b>Знати:</b> Основні концепції реактивного програмування, принципи створення та реалізації реактивних виразів у <i>Shiny</i></p> <p><b>Вміти:</b> використовувати реактивні вирази для створення процедур обробки великих даних у <i>R Shiny</i></p>	<p><b>Тема 6. Реактивне програмування процедур обробки великих даних у <i>Shiny</i></b> <b>План лекції 6</b></p> <ol style="list-style-type: none"> <li>1. Реакція на події у веб-додатку. Основи реактивного програмування.</li> <li>2. Реактивні вирази та графіки.</li> <li>3. Основи модульного програмування в <i>Shiny</i>.</li> <li>4. Ефективні прийоми реактивного програмування в аналізі великих даних.</li> <li>5. Оптимізація роботи веб-додатку <i>Shiny</i>.</li> </ol> <p><b>Список рекомендованих джерел:</b> Основний: 5 [с. 50-74, 227-368]. Інтернет-ресурси: 27-31.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 5 [с. 50-74, 227-368]; 27-31.</p> <p><i>На основі набору великих даних, сформованого при вивченні теми 2, створити інтерактивний веб-додаток у середовищі R Shiny, завантажити у нього набір даних, розробити декілька реактивних процедур та функцій для аналізу текстової, графічної та табличної інформації. Здійснити</i></p>	<p>2</p> <p>14</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p><i>аналіз геоданих із завантаженого набору з використанням геосистеми Leaflet.</i></p> <p style="text-align: center;"><b>Лабораторні заняття 13-14</b></p> <p style="text-align: center;"><b>Реактивні функції в аналітиці великих даних у Shiny</b></p> <p><b>Мета:</b> засвоїти практичні прийоми роботи з реактивними <i>render</i>-функціями у додатку Shiny.</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичне завдання: 1) додати <i>tabsetPanel</i> до <i>mainPanel</i> свого додатка Shiny; 2) розмістити <i>tabPanel</i> у <i>tabsetPanel</i>, назвати панель «Data Table»; 3) поставити другу <i>tabPanel</i> біля першої та назвати її «Summary»; 4) завантажити набір даних, який ви хочете проаналізувати («something.txt»), зі свого комп'ютера у свій каталог за допомогою кнопок RStudio; 5) використати різноманітні функції, які перетворюють R-об'єкти на вихід для вашого інтерфейсу користувача:</p> <ul style="list-style-type: none"> <li>- <i>htmlOutput</i>: необроблений HTML-код;</li> <li>- <i>imageOutput</i>: зображення;</li> <li>- <i>plotOutput</i>: графік;</li> <li>- <i>tableOutput</i>: таблиця;</li> <li>- <i>textOutput</i>: текст;</li> <li>- <i>uiOutput</i>: необроблений HTML-код;</li> <li>- <i>verbatimTextOutput</i>: текст;</li> </ul> <p>6) розмістити <i>output</i> функцію всередині <i>sidebarPanel</i> або <i>mainPanel</i> в сценарії <i>ui.R</i>; 7) додати <i>datatableOutput</i> до «Data Table», назвати аргумент «Table»; 8) створити запис, визначивши новий елемент для виведення у функції. Ім'я елемента має відповідати імені реактивного</p>	4



Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p><i>елемента, який ви створили в ui.R.</i></p> <p><i>Кожен запис має містити вихід однієї з функцій Shiny render. Кожна render функція відповідає певному типу реактивного об'єкта:</i></p> <ul style="list-style-type: none"> <li>- <i>renderImage: зображення, збережені як посилання на вихідний файл;</i></li> <li>- <i>renderPlot: графіки;</i></li> <li>- <i>renderPrint: будь-який друкований результат;</i></li> <li>- <i>renderTable: фрейм даних, матриця, інші структури, подібні до таблиці даних;</i></li> <li>- <i>renderText: рядки символів;</i></li> <li>- <i>renderUI: об'єкт meзу Shiny або HTML</i></li> </ul> <p><i>9) додати відповідну render функцію до server.R, щоб створити таблицю даних набору даних; 10) деактивувати текстовий фільтр, встановивши searching значення «FALSE»; 11) вимкнути розбиття на сторінки таблиці даних; 12) встановити точну кількість рядків, які відобразатимуться, рівною 10; 13) створити меню довжини, щоб повністю контролювати вибір кількості рядків, які ми хочемо відобразити, призначити кожному номеру мітку меню: 5 – «5», 10 – «10», 15 – «15», -1 – «Всі»; 14) створити меню довжини зі значеннями (10,20,30,-1) і призначити кожне зі значень відповідній мітці меню.</i></p> <p style="text-align: center;"><b>Лабораторні заняття 15-16</b>  <b>Обробка геоданих з використанням R Shiny</b></p> <p><b>Мета:</b> навчитися опрацьовувати, аналізувати та формувати географічні дані з використанням системи Leaflet у Shiny.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання: 1) використовуючи бібліотеку Leaflet створити</i></p>	4

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>карту з географічною сіткою, обрати для неї довільну форму сітки користувача, але не ту, що пропонується за замовченням. Доступних постачальників можна переглянути за лінком <a href="https://leaflet-extras.github.io/leaflet-providers/preview/index.html">https://leaflet-extras.github.io/leaflet-providers/preview/index.html</a>; 2) за замовченням установити координати центру Швейцарії з рівнем масштабування 7 (скористайтеся Google, щоб знайти відповідні координати); 3) обмежити параметри масштабування мінімальним рівнем 5 і максимальним рівнем 9; 4) використати набір даних <code>leaflet::gadmCHN</code> для зображення кордони кантонів; 5) змінити кольори рамки та заливки на кольори державного прапора Швейцарії (червоний і білий відповідно), змінити непрозорість і товщину ліній кордону відповідно до ваших естетичних уподобань; 6) додати маркер, що вказує на центр кантону «Берн», де знаходиться однойменна столиця Швейцарії; 7) навівши курсор на кантон, змінити межі цього кантону на темно-червоні; 8) навівши курсор на кантон, відобразити мітку з його назвою; 9) додати на карту функцію вимірювання відстані/площі (одиницями вимірювання повинні бути кілометри (для відстаней) і квадратні метри (для площ)); 10) додати до своєї карти функцію міні-карти.</p>	
<p><b>Знати:</b> базові принципи роботи зі скриптами R в аналітичній платформі <i>Microsoft Power BI</i></p>	<p align="center"><b>Тема 7. Аналіз великих даних у Power BI</b></p> <p align="center"><b>План лекції 7</b></p> <ol style="list-style-type: none"> <li>1. Завантаження великих даних з різних джерел у платформу <i>Power BI</i>.</li> <li>2. Імпорт даних з R <i>Power BI</i> засобами <i>API</i>.</li> <li>3. Інтеграція скриптів з обробкою великих даних у R до системи <i>Power BI</i>.</li> <li>4. Створення візуалізації в <i>Power BI</i> за допомогою R</li> </ol> <p align="center"><b>Список рекомендованих джерел:</b></p> <p>Основний: 3 [с. 157-249].</p>	<p align="center">2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Вміти:</b> будувати аналітичні звіти в <i>Microsoft Power BI</i> з використанням аналітики великих даних, проведеної у системі <i>R</i></p>	<p>Додатковий: 12 [с. 113-136]; 17 [с. 70-153] Інтернет-ресурси: 26.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 3 [с. 157-249]; 12 [с. 113-136]; 17 [с. 70-153]; 26.</p> <p><i>На основі набору великих даних, сформованого при вивченні теми 2:</i></p> <ul style="list-style-type: none"> <li>- імпортувати дані до платформи <i>Microsoft Power BI</i> з використанням скрипта <i>R</i>;</li> <li>- здійснити налаштування типів та трансформацію даних, встановити потрібні типи даних, налаштувати заголовки полів, здійснити фільтрацію непотрібних рядків та стовпців, здійснити злиття та розбиття даних, групування, приведення даних до необхідної форми з використанням функцій <i>R</i> для обробки даних;</li> <li>- розрахувати щонайменше один додатковий обчислювальний стовпець і міру засобами <i>R</i>.</li> </ul> <p style="text-align: center;"><b>Лабораторне заняття 17</b></p> <p style="text-align: center;"><b>Візуалізація великих даних у <i>Power BI</i> за допомогою <i>R</i></b></p> <p><b>Мета:</b> створення візуалізацій користувача в аналітичній платформі <i>Power BI</i> з використанням засобів <i>R</i>.</p> <p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичне завдання: 1) створіть стовпчикову діаграму в <i>R</i> на основі аналізу набору даних за допомогою функцій пакету <i>ggplot2</i>; 2) використовуючи результати завдання 1, створіть діаграму з анотацією,</p>	<p style="text-align: center;">14</p> <p style="text-align: center;">2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p><i>бульбашкову діаграму та інші засоби візуалізації/, які характеризують заданий набір даних; 3) перенесіть код скрипта R у Power BI; 4) налаштуйте параметри візуалізації у середовищі Power BI; 5) побудуйте прогноз даних у Power BI з використанням пакетів та функцій прогностичної аналітики R; 6) додайте лінію регресії на прогнозну модель та налаштуйте її зовнішній вигляд.</i></p> <p style="text-align: center;"><b>Лабораторне заняття 18</b></p> <p style="text-align: center;"><b>Створення інформаційної моделі даних Power BI за допомогою R</b></p> <p><b>Мета:</b> отримання практичних навиків завантаження великих даних до аналітичної платформи Power BI засобами R.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання: 1) створіть датафрейм на основі динамічного об'єднання файлів у форматі csv в середовищі R; 2) імпортуйте пакети tidyverse(), readxl() для зчитування файлів Excel за допомогою R і створіть оболонку функції combine_sheets; 3) об'єднайте інформацію з аркушів Excel в один датафрейм за допомогою функції map_dfr (); 4) перенесіть створений код у редактор скриптів R на платформі Power BI; 5) реалізуйте зчитування даних з SQL Server у Power BI за допомогою R; 6) здійсніть зчитування та завантаження даних у Power BI з API за допомогою R.</i></p> <p style="text-align: center;"><b>Лабораторне заняття 19</b></p> <p style="text-align: center;"><b>Перетворення великих даних у Power BI за допомогою методів R</b></p> <p><b>Мета:</b> застосування технології R для перетворення та форматування даних, завантажених до аналітичної платформи Power BI.</p>	<p style="text-align: center;">2</p> <p style="text-align: center;">2</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</p> <p>2. Практичне завдання: 1) організувати захист конфіденційних даних у Power BI за допомогою пакетів tidyverse та stringr; 2) змінити формат завантажених даних на основі розробленого шаблону з правильним форматуванням; 3) здійснити пошук шаблонів у записах на основі умовної логіки за допомогою R; 4) активізувати функцію геокодування адрес в R за допомогою Google Geocoding API; 5) створити обчислювальний стовпець в Power BI з відстанями між заданими географічними точками з використанням функцій R.</p>	
<p><b>Знати:</b> принципи обробки даних з URL адрес в системі R</p> <p><b>Вміти:</b> здійснювати основні операції веб-скрапінгу та парсінгу з використанням пакету rvest</p>	<p style="text-align: center;"><b>Тема 8. Веб-скрапінг та парсінг великих даних</b> <b>План лекції 8</b></p> <ol style="list-style-type: none"> <li>1. Принципи організації доступу до Інтернет із системи R.</li> <li>2. Вирішення проблем з кодуванням веб-сторінок.</li> <li>3. Організація веб-скрапінгу із застосуванням пакету rvest.</li> <li>4. Парсінг веб-сторінок за допомогою RSelenium.</li> <li>5. Веб-скрапінг та парсінг соціальним мереж.</li> </ol> <p style="text-align: center;"><b>Список рекомендованих джерел:</b></p> <p>Основний: 1 [с. 327-356]; 2 [с. 192-204]; 4 [с. 240-369]. Додатковий: 9 [с. 178-220]; 10 [с. 33-98]; 14 [с. 435-441, 477-481]. Інтернет-ресурси: 30; 33.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 1 [с. 327-356]; 2 [с. 192-204]; 4 [с. 240-369]; 9 [с. 178-220]; 10 [с. 33-98]; 14 [с. 435-441, 477-481]; 30; 33.</p>	<p style="text-align: center;">2</p> <p style="text-align: center;">12</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p><i>Здійснити веб-скрапінг набору великих даних, який було ідентифіковано у темі 2 без попереднього завантаження датасету на комп'ютер. Провести парсинг веб-сторінки з набором даних, виявити потрібні фактори та записи і завантажити набір даних потрібного формату до системи R з використанням бібліотеки rvest.</i></p> <p style="text-align: center;"><b>Лабораторні заняття 20-21</b></p> <p><b>Скрапінг та парсинг веб-сторінок з використанням пакету rvest</b>  <b>Мета:</b> застосування пакету rvest для проведення парсингу та веб-скрапінгу, створення аналізатора вмісту веб-сторінок.</p> <p><i>1. Обговорення питань за проблематикою теми. Презентація індивідуальних завдань.</i></p> <p><i>2. Практичне завдання: 1) завантажити розширення Selector Gadget із сайту selectorgadget.com, додати його до свого веб-браузера та закріпити на панелі розширень; 2) проаналізувати вміст сайту IMDb.com – найбільшого в світі джерела інформації в галузі кінематографу, звернути увагу на необхідні поля (назва фільму, рік випуску, тривалість фільму в хвилинах, жанр фільму, короткий опис сюжету, рейтинг, кількість набраних голосів); 3) з використанням пакетів rvest і dplyr створити аналізатор назв фільмів, починаючи з 2020 року шляхом визначення CSS-селекторів HTML-елементів на сторінці</i>  <a href="https://www.imdb.com/search/title/?year=2020&amp;amp">https://www.imdb.com/search/title/?year=2020&amp;amp</a> визначити</p> <p><i>4) подібним чином провести парсинг реєстру елементів необхідних даних; 5) записати Отриманий кадр даних movie_list у файл csv за допомогою функції write.csv(movie_list, file_name); 6) провести аналіз глибинний веб-скрапінг, здійснивши аналіз вкладених посилань з інформацією про кадровий склад акторів кожного фільму; 7) здійснити веб-скрапінг даних з декількох сторінок, використавши цикл for.</i></p>	4

Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Знати:</b> ключові аспекти технологій <i>MapReduce</i>, <i>Hadoop</i> та <i>Spark</i></p> <p><b>Вміти:</b> ідентифікувати неструктуровані великі дані, які потребують використання технологій <i>MapReduce</i>, <i>Hadoop</i> і <i>Spark</i></p>	<p><b>Тема 9. Технології <i>MapReduce</i>, <i>Hadoop</i> і <i>Spark</i> в аналітиці великих даних</b> <b>План лекції 9</b></p> <ol style="list-style-type: none"> <li>1. Неструктуровані великі дані.</li> <li>2. Проблеми застосування базових методів для аналізу неструктурованих великих даних.</li> <li>3. Алгоритми обробки великих даних із використанням технології <i>MapReduce</i>.</li> <li>4. Поточкова обробка великих даних за технологією <i>Hadoop</i>.</li> <li>5. Технологія <i>Spark</i> в аналітиці великих даних.</li> </ol> <p><b>Список рекомендованих джерел:</b> Основний: 1 [с. 295-323]; 2 [с. 37-141]; 4 [127-239, 370-408]. Додатковий: 9 [с. 242-259, 291-305]. Інтернет-ресурси: 30; 31; 33; 35.</p> <p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 1 [с. 295-323]; 2 [с. 37-141]; 4 [127-239, 370-408]; 9 [с. 242-259, 291-305]; 30; 31; 33; 35.</p> <p>Підготувати наукове повідомлення (на 5-7 хв.) або презентацію про сучасні технології в аналітиці <i>Big Data</i>:</p> <ul style="list-style-type: none"> <li>- розгортання хмарних віртуальних машин за допомогою <i>Hadoop</i>;</li> <li>- розподілена файлова система <i>Hadoop</i> (HDFS) і фреймворки <i>MapReduce</i>;</li> <li>- налаштування віртуальної машини для включення основних</li> </ul>	<p>2</p> <p>10</p>

Результати навчання	Навчальна діяльність	Робочий час студента, год.
	<p>бібліотек <i>HDFS</i>;</p> <ul style="list-style-type: none"> <li>- корисні додаткові інструменти для керування даними в <i>HDFS</i>;</li> <li>- інтеграція статистичного середовища <i>R</i> із <i>Hadoop</i> в одновузловий кластер;</li> <li>- керування файлами в <i>HDFS</i> і виконання простих завдань <i>MapReduce</i> за допомогою пакету <i>rhadoop</i>;</li> <li>- аналіз наборів даних у багатовузловому кластері <i>HDInsight</i> у <i>Microsoft Azure</i>;</li> <li>- оцінка характеристик та функцій <i>Spark</i>;</li> <li>- підключення <i>RStudio Server</i> до ресурсів <i>HDFS</i>, включаючи <i>Hadoop</i> і <i>Spark</i>;</li> <li>- перетворення даних <i>Spark</i>, їх агрегація та аналіз за допомогою відкритих даних <i>Bay Area Bike Share</i>.</li> </ul>	
<p><b>Знати:</b> основні засоби аналізу надвеликих даних з використанням нейронних мереж, алгоритмів машинного навчання та систем штучного інтелекту</p>	<p style="text-align: center;"><b>Тема 10. Методи машинного навчання для надвеликих даних</b> <b>План лекції 10</b></p> <ol style="list-style-type: none"> <li>1. Поняття та характеристичні ознаки надвеликих даних.</li> <li>2. Передумови застосування методів машинного навчання в аналітиці <i>Big Data</i>.</li> <li>3. Інтеграція моделей машинного навчання до платформи <i>Power BI</i>.</li> <li>4. Розгортання моделей штучного інтелекту <i>Azure Cognitive Services</i>.</li> <li>5. Система <i>IBM Watson</i>.</li> </ol> <p style="text-align: center;"><b>Список рекомендованих джерел:</b></p> <p>Основний: 1 [с. 359-393]; 2 [с. 14-178]; 3 [с. 268-333]; 4 [409-465]. Додатковий: 7 [с. 373-413]; 14 [с. 497-672]; 20 [с. 235-277, 319-364]. Інтернет-ресурси: 30; 31; 33; 35.</p>	2



Результати навчання	Навчальна діяльність	Робочий час студента, год.
<p><b>Вміти:</b> ідентифікувати структури надвеликих даних, що потребують використання методів та моделей машинного навчання</p>	<p><b>Самостійна робота студентів.</b> Підготовка до лабораторного заняття за матеріалами теми, вивчення основних та додаткових джерел, Інтернет-ресурсів, підготовка до тестування, виконання індивідуальних завдань 1 [с. 359-393]; 2 [с. 14-178]; 3 [с. 268-333]; 4 [409-465]; 7 [с. 373-413]; 14 [с. 497-672]; 20 [с. 235-277, 319-364]; 30; 31; 33; 35.</p> <p>Підготувати наукове повідомлення (на 5-7 хв.) або презентацію про прогресивні засоби аналізу надвеликих даних:</p> <ul style="list-style-type: none"> <li>- основні технології великих даних, доступні на ринку, та їх інтеграція з мовою <i>R</i>;</li> <li>- статистичні інструменти для аналізу надвеликих даних;</li> <li>- створення чутливих спам-фільтрів та сканування електронної пошти на віруси;</li> <li>- надвеликі дані в технології комп'ютерного зору, включаючи розпізнавання обличчя, поведінки та об'єктів у динаміці;</li> <li>- аналіз надвеликих даних у системах протидії та запобігання кіберзлочинам;</li> <li>- методи машинного навчання у політтехнологіях;</li> <li>- конфігурація засобів аналізу та налаштування тексту в <i>IBM Watson</i>;</li> <li>- рекомендаційні алгоритми на службі аналітики <i>Big Data</i>;</li> <li>- обробка надвеликих наборів даних у <i>translate</i>-системах зі штучним інтелектом.</li> </ul>	10
<i>Підсумковий контроль – письмовий екзамен</i>		
<b>Разом</b>		<b>180</b>

**Примітка:**

*курсивом позначено інтерактивні методи навчання*

### 3. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

#### Основний

1. *Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data /EMC Education Services. Indianapolis : John Wiley & Sons, Inc, 2015. 432 p.*
2. *Prajapati V. Big Data Analytics with R and Hadoop: Packt Publishing, 2013. 238 p.*
3. *Wade R. Advanced Analytics in Power BI with R and Python: Ingesting, Transforming, Visualizing. Apress, 2020. – 440 p.*
4. *Walkowiak S. Big Data Analytics with R: Packt Publishing, 2016. 491 p.*
5. *Wickham H. Mastering Shiny: O'Reilly Media, 2021 352 p.*

#### Додатковий

6. *Bell R. Big data in travel, consumption and marketing: big data for business. Independently published, 2021. – 142 p.*
7. *Casella G., Fienberg S., Olkinand I. An Introduction to statistical learning with application in R: Springer, 2017. 440 p.*
8. *Chinnici M., Pop F., Negru C. Data Science and Big Data Analytics in Smart Environments. CRC Press, 2021. – 292 p.*
9. *Cuesta H., Kumar S. Practical Data Analysis: Packt Publishing, 2016. 330 p.*
10. *Danneman N., Heimann R. Social Media Mining with R: Packt Publishing, 2014. 122 p.*
11. *Etaati L. Advanced Analytics with Power BI and R. - Auckland: RADACAD Systems Limited, 2017. - 179 p.*
12. *Evans J. R. Business Analytics: Methods, Models, and Decisions: Pearson, 2021. 705 p.*
13. *Hafner S. An Introduction to R for Beginners. - Hafner Consulting LLC, 2019. - 360 p.*
14. *Irizarry A. Introduction to Data Science. Data Analysis and Prediction Algorithms with R: Chapman and Hall/CRC, 2020. 743 p.*
15. *Kabacoff R. R in Action. Data analysis and graphics with R. – Manning: Shelter island, 2015. 608 p*
16. *Kakulapati V. Open Data: ITeXLi, 2022. 77 p.*
17. *Larson B. Data Analysis with Microsoft Power BI: Mc Graw Hill, 2020. 741 p.*
18. *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.*
19. *Matloff N. Probability and Statistics for Data Science: Math + R + Data. – London: Chapman & Hall, 2019. – 376 p.*
20. *Mishra B.K., Kumar V., Panda S.K., Tiwari P. Handbook of Research for Big Data. Concepts and Techniques: CRC-Press, 2022. 389 p.*
21. *Moreira J. M., de Carvalho A.C.P.L.F, Horvath T., A General Introduction to Data Analytics: John Wiley & Sons, Inc., 2019. 328 p.*
22. *Rancher A. An Introduction to Big Data Concepts URL: [https://www.suse.com/c/rancher\\_blog/an-introduction-to-big-data-concepts/](https://www.suse.com/c/rancher_blog/an-introduction-to-big-data-concepts/)*

23. Sharda R., Delen D., Turban E. Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support. Pearson; 11th edition, 2019. – 832 p.
24. Майборода Р. Є., Сугакова О. В. Аналіз даних за допомогою пакета R: навчальний посібник. – К.: ВПЦ «Київський університет», 2015. – 65 с.
25. Негрей М., Гнот Т. Аналітика з R. – Київ: Компрінт, 2020. – 236 с.

*Інтернет-ресурси*

26. Microsoft Power BI Guided Learning. URL: <https://docs.microsoft.com/uk-ua/power-bi/guided-learning>
27. Official portal of R Shiny. URL: <https://shiny.rstudio.com/>
28. RStudio. URL: <https://www.rstudio.com/products/rstudio/download/#download>
29. RStudio-education. URL: <https://github.com/rstudio-education>
30. Scraping the Web in R. URL: <https://www.ssc.wisc.edu/sscc/pubs/web scraping-r/scraping-the-web.html>
31. The Comprehensive R Archive Network. URL: <https://cran.r-project.org>
32. The Open Data World Portal. URL: <https://data.world>
33. Web Scraping in R with rvest. URL: <https://www.dataquest.io/blog/web-scraping-in-r-rvest>
34. Аналітична система *YouControl*. URL: <https://youcontrol.com.ua>
35. Парсинг даних у R. URL: [http://rstudio-pubs-static.s3.amazonaws.com/6955\\_2c6760795680448f8ab9f47c08669ca0.html](http://rstudio-pubs-static.s3.amazonaws.com/6955_2c6760795680448f8ab9f47c08669ca0.html)
36. Портал відкритих даних. URL: <https://data.gov.ua>
37. Система аналітики відкритих даних *Clarity Project*. URL: <https://clarity-project.info>

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