

**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 ДТЕУ

(пост. П. 8 від «27» 10 2022 р.)

Ректор



**Анатолій МАЗАРАКІ**

**ІНСТРУМЕНТАЛЬНІ ЗАСОБИ  
БІЗНЕС-АНАЛІТИКИ /  
BUSINESS INTELLIGENCE TOOLS**

**ПРОГРАМА /  
COURSE SUMMARY**

**Київ 2022**

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

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

Програму розглянуто і затверджено на засіданні кафедри цифрової економіки та системного аналізу 02.03.2022 р., протокол № 7

Рецензенти: В. В. Кулаженко, кандидат економічних наук,  
М. Г. Шарафутдінов, бізнес-аналітик, директор з розвитку  
компанії «Center Research & Development».

**ІНСТРУМЕНТАЛЬНІ ЗАСОБИ  
БІЗНЕС-АНАЛІТИКИ /  
BUSINESS INTELLIGENCE TOOLS**

**ПРОГРАМА /  
COURSE SUMMARY**

## ВСТУП

Програма дисципліни «Інструментальні засоби бізнес-аналітики» призначена для студентів бакалаврів ДТЕУ галузі знань 12 «Інформаційні технології», спеціальностей 126 «Інформаційні системи та технології».

Програму підготовлено відповідно до Стандартів вищої освіти України, Стандарту вищої освіти та відповідних освітньо-професійних програм підготовки бакалаврів і магістрів ДТЕУ.

Програма складається з таких розділів:

1. Мета, завдання та предмет дисципліни.
2. Передумови вивчення дисципліни як вибіркової компоненти освітньої програми.
3. Результати вивчення дисципліни.
4. Зміст дисципліни.
5. Список рекомендованих джерел.

### ***1. МЕТА, ЗАВДАННЯ ТА ПРЕДМЕТ ДИСЦИПЛІНИ***

*Метою* вивчення дисципліни «Інструментальні засоби бізнес-аналітики» є формування у майбутніх фахівців необхідного рівня фундаментальних знань та набуття практичних навичок використання мови програмування *R* та її бібліотек для реалізації процесів збору, первинної обробки, трансформації, аналізу та візуалізації даних, виявлення нових властивостей та закономірностей у бізнес-інформації, створення моделей і програмних систем бізнес-аналітики.

*Завданням* вивчення дисципліни «Інструментальні засоби бізнес-аналітики» є надання студентам ґрунтовних знань в глибокого аналізу даних у процесі цифрової трансформації бізнесу.

*Предметом* вивчення дисципліни є технології та методика, що базуються на програмному середовищі *R* та використовують екосистему цієї мови для створення ефективних моделей глибокого аналізу даних бізнес-структур.

### ***2. ПЕРЕДУМОВИ ВИВЧЕННЯ ДИСЦИПЛІНИ ЯК ВИБІРКОВОЇ КОМПОНЕНТИ ОСВІТНЬОЇ ПРОГРАМИ***

*знання*

- основ інформаційних технологій (операційна система *Windows*, бази даних);
- основ дискретної математики, математичної логіки, алгоритмізації та програмування;
- основ теорії ймовірностей та математичної статистики (випадкові величини та їх числові характеристики, закони розподілу випадкових величин, статистичні гіпотези та методи їх перевірки);

*вміння*

- вільно працювати з офісними додатками *Microsoft Word, Microsoft Excel, Microsoft PowerPoint*.

### 3. РЕЗУЛЬТАТИ ВИВЧЕННЯ ДИСЦИПЛІНИ

Дисципліна «Інструментальні засоби бізнес-аналітики» забезпечує оволодіння студентами загальними та фаховими компетентностями і досягнення ними програмних результатів навчання за освітньо-професійними програмами:

#### «Інформаційні системи та технології» (ОС бакалавр)

Номер в освітній програмі	Зміст компетентності	Номер теми, що розкриває зміст компетентності
<i>Загальні компетентності за освітньою програмою</i>		
КЗ 1.	Здатність до абстрактного мислення, аналізу та синтезу.	1-6
КЗ 2.	Здатність застосовувати знання у практичних ситуаціях.	1-3,5,6
КЗ 6.	Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.	1-3,5,6
<i>Фахові компетентності за освітньою програмою</i>		
КС 1.	Здатність аналізувати об'єкт проектування або функціонування та його предметну область.	3,5,6
КС 4.	Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).	1-6
КС 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.	1-6
КС 13.	Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.	1-6

### 4. ЗМІСТ ДИСЦИПЛІНИ

#### Тема 1. Моделювання даних у бізнес-аналітиці

Основні поняття бізнес-аналітики. Аналітичні дані. Види аналітики. Актуальні бізнес-задачі аналітики даних.

Знайомство з *R*. Доступ до інсталяції програмного середовища. Редагування та виконання коду. Принципи пакетної організації обчислень. Робота з великими масивами даних бізнесу. Графічний інтерфейс *GUI* для *R*.

*RStudio*. Робота з консоллю. Термінал *RStudio*. Редагування та виконання коду. Автоматичне створення функцій. Коментування коду. Історія команд. Робочий простір *Workspace*.

Введення даних з клавіатури. Імпорт аналітичних даних з текстового файлу, файлу *Excel*, *XML*-файлу. Добування бізнес-даних із веб-сторінок. Імпорт даних із систем управління базами даних.

#### **Список рекомендованих джерел**

*Основний:* 2-5.

*Додатковий:* 11,12,19,21,23,29,31.

*Інтернет-ресурси:* 32-35.

#### **Тема 2. Управління даними в R.**

Синтаксис мови R. Змінні. Типи об'єктів бізнес-аналізу: вектори, матриці, масиви даних, таблиці даних, фактори, списки. Робота з *Data Frame*

Перетворення даних (*Reshape, Split, Combine*). Робота із пропущеними даними. Перетворення типів. Сортування даних. Об'єднання наборів даних. Поділ множини даних на складові.

Функції R. Числові і текстові функції. Умовні оператори. Цикли. Функції користувача. Експорт результатів аналізу із середовища R.

#### **Список рекомендованих джерел**

*Основний:* 2-5.

*Додатковий:* 11,12,19,21,23,29,31.

*Інтернет-ресурси:* 32-35.

#### **Тема 3. Описова та індуктивна бізнес-аналітика.**

Описова статистика даних бізнес-процесів. Оцінка вибірових параметрів. Найбільш поширені помилки при виконанні статистичного аналізу. Таблиці частот та спряженості. Кореляції. Перевірка статистичної значущості кореляцій. Візуалізація кореляцій.

Закони розподілу випадкових даних: біноміальний, рівномірний, експонентний, нормальний розподіл, розподіл Пуасона. Підгонка розподілів. Аналіз викидів бізнес-даних. Заповнення пропущених значень у таблицях даних. Відновлюваність результатів спостережень з використанням генератора випадкових чисел.

Основні кроки побудови та верифікації моделей. Статистична перевірка гіпотез. Класичні методи і критерії статистики. Тест Шапіро-Уїлка. Рангові критерії Уїлкінсона-Манна-Уїтні. Тести Стюдента для залежних і незалежних вибірок даних. Непараметричні тести міжгрупових різниць. Рандомізація, бустреп та оцінка статистичної потужності вибірки. Оцінка кореляції. Тест  $\chi^2$ -квадрат (Пірсона). Тест Колмогорова-Смірнова. Критерій Фішера. Точний тест Фішера. Критерії Мак-Немара та Кохрана-Мантеля-Хензеля.

Дисперсійний аналіз (*ANOVA*). Однофакторний дисперсійний та коваріаційний аналіз. Лінійні моделі дисперсійного аналізу. Дисперсійний аналіз за Краскелом-Уолісом. Багатовимірний дисперсійний аналіз.

#### **Список рекомендованих джерел**

*Основний:* 2,3,5.

*Додатковий:* 7,12,13,17,19,21,23.

*Інтернет-ресурси:* 32-35.

#### **Тема 4. Візуалізація даних та результатів бізнес-аналізу**

Створення графіки в *R*. Робота з діаграмами показників бізнес-процесу. Прості методи візуалізації бізнес-даних: точкові діаграми, стовпчикові діаграми, кругові та секторні діаграми, спінограми, гістограми, діаграми розмахів, діаграми ядерної оцінки функції щільності, скрипкові діаграми.

Графічні параметри візуалізації даних: розташування діаграм на сторінці, системи координат, колір та прозорість, налаштування відображення графіка, легенда, заголовки графіка та осей, розмітка осей, підписи та анотації, фасети, графічні теми.

Розширена графіка для просунутої бізнес-аналітики. Чотири графічні системи *R*: *Base*, *Grid*, *Lattice*, *ggplot2*. Пакет *lattice*. Умовні змінні. Функції для зміни формату даних. Пакет *ggplot2*. Інтерактивна графіка. Взаємодія з діаграмами: ідентифікація точок. Пакет *playwith*. Пакет *lattice*. Створення інтерактивної графіки за допомогою пакета *iplots*. Пакет *rggobi*.

Діаграми розсіювання, бульбашкові діаграми, лінійні графіки, корелограми, мозаїчні діаграми, діаграми Клівленда. Категоризовані графіки.

#### **Список рекомендованих джерел**

*Основний:* 2,3,5.

*Додатковий:* 8,12,16,19,21,23,25,29.

*Інтернет-ресурси:* 33-35.

#### **Тема 5. Прогностична бізнес-аналітика.**

Регресійний бізнес-аналіз. Моделі регресії при різних видах функції втрат. МНК-регресія: проста лінійна, поліноміальна, множинна лінійна регресія. Діагностика та відбір регресійних моделей. Мультиколінеарність.

Нестандартні спостереження за реальними даними: викиди, точки високої напруженості, впливові спостереження. Способи коригування нестандартних спостережень: видалення спостережень, перетворення змінних, додавання або видалення змінних.

Узагальнені лінійні моделі і функція *glm()*. Логістична регресія в *R*. Пуасонівська регресія. Моделі згладжування. Узагальнені моделі регресії. Моделі пробіт-та логіт-регресії.

Вибір оптимальної регресійної моделі бізнес-даних. Порівняння моделей. Критерії вибору моделей оптимальної складності. Крос-валідація. Відносна важливість. Прогнозування даних в *R*. Порівняння ефективності різних моделей при прогнозуванні.

#### **Список рекомендованих джерел**

*Основний:* 2,3,5.

*Додатковий:* 7,12,21,23.

*Інтернет-ресурси:* 33-35.

## Тема 6. Розширена бізнес-аналітика

Задачі *Data Mining* в *R*. Бінарні матриці та асоціативні правила. Пошук логічних закономірностей у даних. Бінарні класифікатори. Дискримінантний аналіз. Моделі класифікації. Іриси Фішера і метод *k-neighbors*. Класифікаційні дерева, основні принципи їх роботи та імплементація в *R*. Наївний класифікатор Байеса.

Кластерний аналіз. Бібліотека *NbClust*. Багатовимірне шкалювання для візуалізації кластерів. Ієрархічна кластеризація. Алгоритм побудови дендрограми. Самоорганізовані карти Кохонена в задачах кластеризації та сегментації бізнес-даних.

Макет призначеного для користувача інтерфейсу. Додавання керуючих віджетів. Пакети *gWidgets* та *fgui*. Основні підходи до розробки веб-програм. Пакет *shiny*. Поширення створених веб-додатків.

Способи сканування бізнес-даних. *CSS*-селектори. *Web scraping* з пакетом *rvest*. *PhantomJS* і збереження вмісту динамічних веб-сторінок. Збір інформації за допомогою *API*.

Просторовий аналіз даних і створення картограм. Моделі просторових даних. Векторні та растрові дані. Проста карта: використання растрового рисунка і підрахунок відстаней. Аналіз просторового розміщення даних бізнес-аналізу. Особливості аналізу та візуалізації просторових даних. Використання сервісів картографічної системи *Google Maps*. Побудова картограм з використанням *R*.

### Список рекомендованих джерел

*Основний*: 1,3.

*Додатковий*: 6,8,10,15,16,20,24,26-28.

*Інтернет-ресурси*: 33-35.

## 5. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

### Основний

1. Etaati L. *Advanced Analytics with Power BI and R*. - Auckland: RADACAD Systems Limited, 2017. - 179 p.
2. Hafner S. *An Introduction to R for Beginners*. - Hafner Consulting LLC, 2019. - 360 p.
3. Kabacoff R. *R in Action. Data analysis and graphics with R*. – Manning: Shelter island, 2015. – 608 p
4. Roskladka A., Ivanova O., Kulazhenko V. *Data Scientist: a glance into the future // Зовнішня торгівля: економіка, фінанси, право*. 2019. № 3. С. 109-120
5. Негрей М., Гнот Т. Аналітика з *R*: навч. посібник / М. Негрей, Т. Гнот. – Київ: Компринт, 2020. – 236 с.

### Додатковий

6. Burger S. *Introduction to Machine Learning with R: Rigorous Mathematical Analysis*. - Sebastopol: O'Reilly Media, 2018. - 226 p.
7. Casella G., Flienbergs S., Olkin I. *An Introduction to statistical learning with application in R*. - Springer, 2017. - 440 p.
8. Cbang W. *R Graphics Cookbook* - O'Reilly Media, 2013. - 413 p.

9. Dietrich D., HelleR B., Yang B. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. – John Wiley & Sons, Inc., 2015. – 420 p.
10. Gago M. Creation of Graphical User Interfaces for R. - Universitat Autònoma de Barcelona, 2013. - 89 p.
11. Golemund G. HandsOn Programming with R. - O'Reilly Media, 2014. - 225 p.
12. Horton N., Kleinman K. Using R for Data Management, Statistical Analysis and Graphics. - CRC Press, 2011. - 296 p.
13. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning with Applications in R. - Springer, 2013. - 441 p.
14. Kumar V. Data Mining with R. - CRC Press, 2011. - 289 p.
15. Ledolter J. Data Mining and Business Analytics with R. - New Jersey: Willey, 2013. - 368 p.
16. Maindonald & J. Brown J. Data analysis and graphics using R. - Cambridge University Press, 2010. - 565 p.
17. Matloff N. Probability and Statistics for Data Science: Math + R + Data. – London: Chapman & Hall, 2019. – 376 p.
18. Nolan D., Lang D.T. Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving. London: Chapman and Hall, 2015. - 539 p.
19. Paradis E. R for beginners. - Montpellier, 2005 - 76 p.
20. 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.
21. Teetor P. R Cookbook. - O'Reilly Media, 2011. - 436 p.
22. Torgo L. Data Mining with R: Learning with Case Studies. – London: Chapman and Hall, 2017. – 446 p.
23. Viswanathan V. Data Analytics with R: A hands-on approach. - New Jersey: Infivista Inc, 2015. - 420 p.
24. Walkowiak S. Big Data Analytics with R: Leverage R Programming to uncover hidden patterns in your Big Data. - Birmingham: Packt Publishing, 2016. - 506 p.
25. Wickham H. ggplot2: Elegant graphics for data analysis. - Springer, 2017. - 217 p.
26. Wickham H., Golemund G. R for Data Science. – Sebastopol: O'Reilly Media, 2017. – 522 p.
27. Zhao Y., Cen Y. Data Mining Applications with R. - Elsevier, 2014. - 470 p.
28. Zumel N., Mount J. Practical Data Science with R. - Manning Publications, 2014. - 417 p.
29. Гнатюк В. Вступ до R на прикладах. - Харків: ХНЕУ, 2010. - 107 с.
30. Дроздов В. С. И др. Платформа Rstudio для анализа больших данных // Материалы 4 Международной научно-практической конференции "BIG DATA and Advanced Analytics" (Минск, Республика Беларусь, 3-4 мая 2018 г.). - С. 110-122
31. Майборода Р. Є., Сугакова О. В. Аналіз даних за допомогою пакета R: навчальний посібник. – К.: ВПЦ «Київський університет», 2015. – 65 с.

### *Інтернет-ресурси*

32. RStudio - <https://www.rstudio.com/products/rstudio/download/#download> (дата звернення 02.11.2020).
33. RStudio-education - <https://github.com/rstudio-education> (дата звернення 02.11.2020).
34. The Comprehensive R Archive Network - <https://cran.r-project.org/> (дата звернення 02.11.2020).
35. The R Project for Statistical Computing - <https://www.r-project.org/> (дата звернення 02.11.2020).

\* Курсивом виділені джерела, що є у бібліотеці ДТЕУ