



# ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій  
Кафедра цифрової економіки та системного аналізу

## СИЛАБУС (SYLLABUS) Дисципліна «Технології аналізу даних»/ «Data Analysis Technologies»

### ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА

Викладач	Роскладка Андрій Анатолійович
Науковий ступінь	Доктор економічних наук, кандидат фізико-математичних наук
Вчене звання	Професор
Посада	Завідувач кафедри цифрової економіки та системного аналізу
Адреса кафедри	м.Київ, вул. Кіото 19, каб. Б-113
E-mail	a.roskladka@knteu.edu.ua
Консультації	Відповідно до графіку індивідуальних консультацій на сайті кафедри

### ПОЛІТИКА АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ

<https://knteu.edu.ua/file/MzEyMQ==/c12a9f74e87d9154696ca0f761da2e5c.pdf>

#### Дотримання академічної доброчесності передбачає:

- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);
- посилання на джерела інформації у разі використання не авторських ідей, розробок, тверджень, відомостей і т.п.;
- дотримання норм законодавства про авторське право і суміжні права;
- надання достовірної інформації про результати власної наукової діяльності, використанні методики досліджень і джерела інформації.

#### Порушенням академічної доброчесності вважається:

- академічний плагіат – оприлюднення (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства;
- самоплагіат – оприлюднення (частково або повністю) власних раніше опублікованих наукових результатів як нових наукових результатів;
- фабрикація – вигадкування даних чи фактів, що використовуються в наукових дослідженнях;
- фальсифікація – свідомо зміна чи модифікація вже наявних даних, що стосуються наукових досліджень.

#### За порушення академічної доброчесності здобувачі освіти можуть бути притягнені до академічної відповідальності:

- повторне проходження оцінювання (модульний контроль, іспит, залік тощо);
- повторне проходження відповідного освітнього компонента освітньо-професійної програми;

- відрахування з Університету;
- позбавлення наданих університетом пільг;
- відмова у присудженні відповідного ступеня вищої освіти;

### ПОЛІТИКА ЩОДО ВІДВІДУВАННЯ ЗАНЯТЬ

- відвідування занять є обов'язковим;
- студент, який пропустив практичне заняття, самостійно вивчає матеріал (при виникненні питань може звертатися за консультацією згідно розкладу консультацій викладачів оприлюдненого на сайті кафедри) за наведеними джерелами, виконує завдання і дає його викладачу.
- за об'єктивних причин (наприклад, хвороба, міжнародне стажування та ін.) навчання може відбуватись в он-лайн формі за погодженням із викладачем дисципліни.

### ПОЛОЖЕННЯ ПРО АПЕЛЯЦІЮ РЕЗУЛЬТАТІВ ЕКЗАМЕНІВ У ДТЕУ

<https://knute.edu.ua/file/MjkwNQ==/cf2f392763bdbe0447eed3c254854ec5.pdf>

### ВРЕГУЛЮВАННЯ КОНФЛІКТНИХ СИТУАЦІЙ

Учасники освітнього процесу повинні дотримуватися принципів гідності, взаємоповаги, толерантності, доброчесності. Адміністрація ДТЕУ забезпечує попередження, запобігання, своєчасне виявлення та врегулювання конфліктних ситуацій, пов'язаних із цькуванням, дискримінацією, сексуальними домаганнями (див. Положення про врегулювання конфліктних ситуацій ДТЕУ (<https://knute.edu.ua/file/MjkwMjQ=/b91ca19cb0c629d8b9938ba46ccc41f5.pdf>)).

### УЗГОДЖЕННЯ ЗМІСТУ ДИСЦИПЛІНИ З ЦІЛЯМИ СТАЛОГО РОЗВИТКУ (ЦСР)

Зміст дисципліни «Технології аналізу даних» узгоджується з кількома ключовими Цілями сталого розвитку (ЦСР), зокрема:

1. **ЦСР 4. Забезпечення всеохоплюючої і справедливої якісної освіти та заохочення можливості навчання впродовж усього життя для всіх.** Дисципліна сприяє розвитку якісної освіти шляхом забезпечення студентів сучасними знаннями та навичками у сфері аналітики даних. Використання інноваційних технологій та інструментів навчання підвищує конкурентоспроможність студентів на ринку праці та готує їх до викликів цифрового майбутнього.
2. **ЦСР 8. Сприяння поступальному, всеохоплюючому та сталому економічному зростанню, повній і продуктивній зайнятості та гідній праці для всіх.** Інтелектуальний аналіз даних сприяє підвищенню продуктивності праці завдяки виявленню прихованих закономірностей та тенденцій у показниках діяльності, що стимулює інновації та сприяє економічному зростанню. Актуальні знання у сфері аналізу даних здатні забезпечити зайнятість та гідну працю у різних сферах людської діяльності.
3. **ЦСР 9. Створення стійкої інфраструктури, сприяння всеохоплюючій і сталій індустріалізації та інноваціям.** Прогресивні технології аналізу даних сприяють ефективному довгостроковому прогнозуванню показників діяльності, що дозволяє виявити слабкі місця бізнес-процесів, запровадити інноваційні методи управління і забезпечити сталій розвиток підприємств та організацій.
4. **ЦСР 11. Забезпечення відкритості, безпеки, життєстійкості й екологічної стійкості міст, інших населених пунктів.** Розумні міста передбачають використання розумних технологій, спрямованих на підвищення економічних, соціальних та екологічних стандартів. У свою чергу розумні технології ґрунтуються на обробці великих даних та їх інтелектуальному аналізі з використанням передових інструментів

аналітики, які безпосередньо становлять предмет вивчення дисципліни «Технології аналізу даних».

5. **ЦСР 12. Забезпечення переходу до раціональних моделей споживання і виробництва.** Технології аналізу даних спрямовані на підвищення якості процесів виробництва, оптимізацію розподілу ресурсів та мінімізацію витрат, забезпечуючи створення раціональних моделей споживання і виробництва у різних галузях сферах людської діяльності.

### ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назва дисципліни / тип дисципліни	Технології аналізу даних/ обов'язкова
Навчальний рік	2024-2025
Факультет	Факультет інформаційних технологій
Курс	1
Семестр	1
Освітній ступінь	Магістр
Галузь знань	12 «Інформаційні технології»
Спеціальність	122 «Комп'ютерні науки»
Загальна характеристика	Кількість годин –180 Кількість кредитів – 6 <b>Види занять:</b> лекції, практичні, самостійна робота. <b>Співвідношення аудиторних годин і годин самостійної роботи</b> - 60/120 <b>Мова викладання</b> – українська <b>Форма викладання</b> – очна
Підсумковий контроль	Екзамен
Програмне забезпечення	R, Tableau, Microsoft Power BI.
Обладнання	Проектор, комп'ютерна техніка із встановленим програмним забезпеченням та доступом до мережі Інтернет.
Необхідні попередні дисципліни	Впевнений користувач ПК, знання офісних програм, основ дискретної математики, математичної логіки, алгоритмізації та програмування, принципів роботи із система управління базами даних, • основ теорії ймовірностей та математичної статистики.
Методика вивчення	Методика вивчення дисципліни полягає у набутті студентами знань теоретичного та практично-прикладного характеру через інтерактивні лекції, практичні заняття, самостійну роботу, індивідуальні проекти. Студенти використовують сучасні мови програмування та аналітичні платформи для проведення інтелектуального аналізу даних. Особлива увага приділяється практичному застосуванню отриманих знань через виконання реальних кейсів.
Мета і завдання	<b>Метою</b> вивчення дисципліни «Технології аналізу даних» є надання фундаментальних теоретичних знань і набуття практичних навичок з питань формування, дослідження та всебічного аналізу даних у різних галузях сферах людської діяльності. <b>Завданням</b> дисципліни є надання студентам ґрунтовних знань в області аналітичних досліджень інформаційного простору, вивчення методів створення, передобробки, трансформації, аналізу та захисту даних.
<b>Місце дисципліни в освітньо-професійній програмі</b>	
Загальні компетентності	ЗК1. Здатність до абстрактного мислення, аналізу та синтезу. ЗК2. Здатність застосовувати знання у практичних ситуаціях.

	ЗК5. Здатність вчитися й оволодівати сучасними знаннями. ЗК7. Здатність генерувати нові ідеї (креативність).
<b>Фахова компетентності</b>	СК4. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень. СК6. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.
<b>Програмні результати навчання</b>	РН2. Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур РН8. Розробляти математичні моделі та методи аналізу даних (включно з великими). РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими). РН11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

## ТЕМАТИКА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

### Тема 1. Вступ до аналітики даних.

Глосарій *Data Science*. Історія розвитку *Artificial Intelligence* і *Business Intelligence*. Зв'язок понять «дані», «інформація» та «знання». Підходи до аналізу даних, сутність та відмінності: *Data Science*, *Data Analysis*, *Data Engineering*, *Machine Learning* та *Deep Learning*.

Типи та види даних. Форми представлення даних. Вимірювання і шкали в аналізі даних. Реляційні та багатовимірні дані. Змінні, постійні та умовно-постійні дані. Довідкові, оперативні та архівні дані. Точкові дані та дані за період. Первинні і вторинні дані. Метадані. Формати зберігання даних. Типи наборів даних. Транзакційні дані.

Технологія *Knowledge Discovery in Databases*. Формування вибірки даних. Етапи аналізу даних. Технологія *Data Mining*. Задачі *Data Mining*: класифікація, регресія, кластеризація, асоціація даних.

Концепція *Big Data*. Масштаби великих даних. Сучасна архітектура *Big Data*. Методики аналізу великих даних. Аналітичний інструментарій для обробки великих даних. Етапи роботи з великими даними.

Концепція *Open Data*. Регулювання доступу до даних. Джерела публічної інформації у формі відкритих даних. Законодавство України щодо відкритих даних. Рейтинг відкритих даних *Global Open Data Index*. Світовий портал відкритих даних *Data.world*. Єдиний державний веб-портал відкритих даних *Data.gov.ua*. Вимоги до структури наборів відкритих даних.

### Тема 2. Статистична обробка інформації в аналітичній системі R.

Знайомство з R та *RStudio*. Імпорт аналітичних даних з текстового файлу, файлу *Excel*, *XML*-файлу. Добування даних із веб-сторінок. Імпорт даних із систем управління базами даних. Типи об'єктів бізнес-аналізу: вектори, матриці, масиви даних, таблиці даних, фактори, списки. Робота з *Data Frame*.

Перетворення даних. Робота із пропущеними даними. Діагностика пропущених даних. Сортування даних. Об'єднання наборів даних. Поділ множини даних на складові.

### Тема 3. Розвідувальна аналітика в системі R.

Створення графіки в R. Робота з діаграмами показників процесу. Прості методи візуалізації даних: точкові діаграми, стовпчикові діаграми, кругові та секторні діаграми, спінограми, гістограми, діаграми розмахів, скрипкові діаграми. Експорт результатів аналізу із середовища R.

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.

Створення причинно-наслідкових діаграм в R. Взаємозв'язки між числовими та категоріальними змінними. Діаграми розсіювання, бульбашкові діаграми, лінійні графіки, корелограми, мозаїчні діаграми, діаграми Клівленда. Категоризовані графіки.

#### **Тема 4. Аналітична платформа *Tableau*.**

Базові навички роботи у системі *Tableau*. Склад та призначення елементів екосистеми *Tableau*. Функціонал безкоштовної версії *Tableau Public*. Імпорт, обробка, візуалізація та експорт даних у системі *Tableau*. Інтерфейс *Tableau Public*.

Передобробка даних в системі *Tableau Prep*. Обробка даних в аналітичній платформі *Tableau*. Візуалізатори системи *Tableau*: діаграми, теплові карти, спарклайни. Дашборди в *Tableau*. Практичне застосування аналітичної платформи *Tableau*.

#### **Тема 5. Аналітика в системі *Microsoft Power BI*.**

Інтерфейс *Power BI Desktop*. Завантаження даних *Power BI* з різних інформаційних джерел. Імпорт даних із реляційних баз даних, текстового файлу, вхідного каналу даних та сервісів аналізу даних.

Формування та аналіз датасетів. Трансформація даних у *Power Query*. Підготовка даних до аналізу. Зміна даних у *Power Query*. Трансформація, очищення та фільтрування даних у *Power BI*. Приведення даних до необхідної форми.

Візуалізація даних у *Power BI Desktop*. Створення таблиць та матриць візуалізації даних. Побудова стрічкових, кругових діаграм та гістограм. Побудова лінійних та точкових діаграм. Створення візуалізацій на основі карт. Інтерактивні візуалізації у *Power BI*.

Публікація звітів та створення інформаційних панелей на порталі *Power BI*. Оновлення даних в опублікованих звітах.

#### **Тема 6. Консолідація та передобробка даних**

Поняття консолідації. Основні критерії оптимальності консолідації даних. Джерела даних. Основні задачі консолідації даних. Схема процесу консолідації. Очищення даних. Збагачення даних.

Рівні якості даних: технічний, аналітичний та концептуальний. Оцінка придатності даних до аналізу. Технології та методи оцінки якості даних. Профайлінг. Візуальна оцінка якості даних. Причини надходження в систему «брудних даних».

Трансформація даних. Основні методи трансформації даних. Перетворення часових рядів. Квантування даних. Сортуння даних. Злиття даних. Об'єднання даних. Налаштування набору даних. Нормалізація даних. Трансформація впорядкованих даних. Групування та розгрупуння даних. Внутрішнє та зовнішнє з'єднання даних. Нормалізація та кодування даних.

#### **Тема 7. Класифікація, регресія та прогнозування даних**

Огляд методів класифікації. Точність класифікації. Оцінювання класифікаційних методів. Статистичні методи аналізу. Байєсівська класифікація.

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

Логістична регресія. Оцінки максимальної правдоподібності. Значущість вхідних змінних. Використання логістичної регресії для розв'язування задач класифікації у фінансовій сфері. Тест Чоу. ROC-аналіз. Множинна логістична регресія. Пуассонівська регресія. Моделі згладжування. Узагальнені моделі регресії. Моделі пробіт-та логіт-регресії.

Методи прогнозування даних. Часовий ряд та його компоненти. Моделі прогнозування часових рядів. Порівняння ефективності різних моделей при прогнозуванні.

Дерева рішень. Структура дерева рішень. Алгоритми побудови дерев рішень. Міри ефективності дерев рішень. Критерії вибору найкращих атрибутів розгалуження. Регресійне дерево рішень. Спрощення дерев рішень. Критерії оптимізації дерев рішень.

## Тема 8. Кластеризація даних

Формальна постановка задачі кластеризації. Задачі кластеризації: вивчення даних, полегшення аналізу, стиснення даних, прогнозування, виявлення аномалій. Приклади кластеризації в різних областях знань. Представлення результатів кластеризації даних.

Базові алгоритми кластеризації. Алгоритм кластеризації *k-means*. Критерій збіжності алгоритму. Міри відстаней у кластеризації. Міри Евкліда і Манхеттена. Алгоритм *g-means*. Кластеризація за Гюстафсоном-Кеселем.

Програмні засоби кластеризації та сегментації даних. Мережі Кохонена. Самоорганізуючі карти Кохонена. Методика побудови карти Кохонена. Вибір кількості нейронів карти. Алгоритм навчання мережі Кохонена.

Ієрархічні алгоритми кластерного аналізу. Міри схожості. Методи об'єднання та зв'язку. Ітеративні алгоритми кластерного аналізу.

Адаптивні методи кластеризації. Проблеми алгоритмів кластеризації. Невизначеність у виборі критерія якості кластеризації. Проблеми машинних ресурсів при кластеризації великих даних. Задача вибору кількості кластерів.

## Тема 9. Асоціативний аналіз даних

Афінитивний аналіз. Поняття типової транзакції. Предметний набір. Основні поняття *Rules Mining*. Асоціативні правила. Умова та наслідок асоціативного правила. Підтримка та достовірність правил. Значущість асоціативних правил. Міри корисності правил. Ліфт, левередж та покращення асоціативних правил. Алгоритм *apriori*. Пошук предметних наборів. Генерація асоціативних правил.

Програмні засоби пошуку асоціативних правил. Секвенціальний аналіз. Практичний аспект застосування технології асоціативних правил.

## Тема 10. Технології інтелектуальної обробки даних

Звітність та стандартні схеми потоків аналітичної інформації. Етапи та методи пошуку нових знань. Організація бізнес-аналітики рівня *Business Intelligence (BI)*.

Інструментальні технології інтелектуальної обробки даних. Візуальний аналіз даних *Visual Mining*. Характеристики засобів візуалізації даних. Методи візуалізації.

Задача та етапи аналізу текстів *Text Mining*. Методи класифікації текстових документів. Видалення стоп-слів. Стеммінг. *N*-грами. Методи кластеризації текстових документів: ієрархічні, бінарні. Задача анотування текстів. Пошук асоціацій. Первинний витяг ключових понять.

Ідея *Data Mining* у реальному часі *Real-Time Mining*. Адаптивне добування даних. Інструменти *Data Mining* у реальному часі. Адаптація системи до загальної концепції. Рекомендаційні машини. Класифікація рекомендаційних машин. Агентне навчання.

Проблеми аналізу даних з мережі Інтернет. Етапи *Web Mining*. Категорії *Web Mining*. Аналіз використання веб-ресурсів. Використання веб-структур та веб-контенту. Аналіз структури сегмента мережі. Персоналізація інформації.

### Перелік навчальних робіт студентів та оцінки їх у балах з дисципліни «Технології аналізу даних»

Види робіт	К-сть балів
Практичні заняття №1, 2 «Створення набору даних з відкритих джерел»	10
Практичне заняття №3 «Базова аналітика у системі R»	5
Практичне заняття №4 «Візуальний аналіз даних у системі R»	5
Практичні заняття №5, 6 «Аналіз даних у системі Tableau»	10
Практичне заняття №7 «Створення та аналіз бази даних в системі Power BI Desktop»	5

Практичне заняття №8 «Аналіз даних із веб-ресурсів»	5
Практичні заняття №9, 10 «Створення багатосторінкових звітів у системі Power BI»	10
Практичні заняття №11, 12 «Принципи кластеризації даних. Алгоритм k-means»	10
Практичні заняття №13, 14 «Створення асоціативних правил. Алгоритм Apriori»	10
<b>Разом: Аудиторна робота</b>	<b>70</b>
<b>Самостійна робота (СР)</b>	<b>30</b>
<b>Всього:</b>	<b>100</b>

### **КОНТРОЛЬ ТА КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАТЬ СТУДЕНТІВ**

При вивченні дисципліни використовуються наступні форми контролю знань студентів: поточний та підсумковий.

**Поточний контроль** передбачає тестову перевірку теоретичних питань, самостійної роботи, практичних робіт та усне опитування щодо результатів виконання практичних завдань. За даним видом контролю оцінювання знань здійснюється у відповідності до бального розподілу наведеного у попередній таблиці.

**Формою підсумкового контролю** є екзамен. Екзаменаційна оцінка (100 балів) є результатом виконання тестових завдань (50 питань = 50 балів) та практичного завдання (50 балів).

**Результуюча оцінка з дисципліни** визначається як середня від балів набраних протягом семестру та отриманих на іспиті.

### **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. Чорноус 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.