



Кам'янець-Подільський національний університет
імені Івана Огієнка
Фізико-математичний факультет
Кафедра комп'ютерних наук

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ **R для науки про дані**

1. Загальна інформація про курс. Викладачі

Іванюк Віталій Анатолійович, доктор технічних наук, доцент, завідувач кафедри комп'ютерних наук

E-mail: wivanyuk@kpmu.edu.ua

Профайл: <https://cs.kpmu.edu.ua/2019/11/04/ivaniuk-vitalij-anatolijovych/>

Консультації: щочетверга – 16.00-18.00 (очно та онлайн)

Пилипюк Тетяна Михайлівна, доцент, доцент кафедри комп'ютерних наук

E-mail: pylypyuk.tetiana@kpmu.edu.ua

Профайл: <https://cs.kpmu.edu.ua/2019/10/30/pylypiuk-tetiana-mykhajlivna/#more-649>

Мова викладання: українська

2. Анотація до курсу

Цей курс є практичним посібником, що навчає студентів повному циклу аналізу даних мовою R. Фокусуючись на сучасній екосистемі Tidyverse (яка є індустріальним стандартом), студенти навчатимуться не просто використовувати R, а мислити як аналітики даних.

Курс охоплює весь шлях: від імпорту "сирих" даних (з файлів та баз даних) та їх очищення до побудови просунутих візуалізацій (ggplot2) та застосування статистичних моделей для отримання дієвих інсайтів.

3. Мета і завдання курсу

Мета: Надати студентам практичні навички для проходження повного циклу роботи з даними — від збору та очищення до моделювання та візуалізації — з використанням мови R та її сучасної екосистеми Tidyverse.

Завдання:

Сформулювати ґрунтовне розуміння синтаксису R та ключових структур даних (вектори, data frames).

Навчити ефективно імпортувати, очищувати та трансформувати "сирі" дані, використовуючи інструменти Tidyverse (dplyr) та підключатися до баз даних.

Навчити створювати інформативні, багатoshарові та готові до публікації візуалізації даних за допомогою "граматики графіки" (ggplot2).

Надати навички застосування ключових статистичних методів та побудови базових моделей машинного навчання (напр., регресія, класифікація) в середовищі R.

Розвинути здатність "розповідати історію" даними (data storytelling), перетворюючи результати аналізу на зрозумілі висновки та дієві інсайти для вирішення практичних кейсів.

4. Результати навчання

Писати відтворювані R-скрипти, впевнено використовуючи базовий синтаксис, функції та ключові структури даних

Імпортувати, очищувати та маніпулювати "сирими" даними з різноманітних джерел, готуючи їх для аналізу.

Інтегрувати R з реляційними базами даних: програмно підключатися до БД та виконувати SQL-запити для отримання, фільтрації та зміни даних.

"Розповідати історію" за допомогою даних, створюючи інформативні та готові до публікації візуалізації для ефективної комунікації результатів аналізу.

Розробляти повний аналітичний пайплайн: від підключення до джерела даних (файлу чи БД) до фінального звіту зі скриптами та візуалізаціями, які дають чіткі відповіді на поставлені запитання.

5. Формат курсу

Очний курс з елементами дистанційного навчання в системі Moodle.

6. Обсяг і ознаки курсу

Найменування показників	Характеристика навчальної дисципліни
Кількість кредитів ЄКТС	3/4
Загальний обсяг годин	90/120
Кількість годин навчальних занять	30/40
Лекційні заняття	12
Практичні заняття	0
Семінарські заняття	0
Лабораторні заняття	18/28
Самостійна та індивідуальна робота	60/80
Форма підсумкового контролю	Залік

7. Пререквізити і кореквізити курсу

Перелік дисциплін, які бажано, але не обов'язково, вивчити раніше: Програмування, Об'єктно-орієнтоване програмування, Організація баз даних та знань, Теорія ймовірностей та математична статистика, Інтелектуальний аналіз даних.

8. Технічне й програмне забезпечення, обладнання

Для проведення лекцій використовується мультимедійне забезпечення. Проведення лабораторних робіт здійснюється в лабораторії із необхідним апаратним та програмним забезпеченням.

Програмне забезпечення: R, RStudio, бібліотеки R.

9. Політика курсу

Увесь навчальний контент розміщено в модульному середовищі навчання K-ПНУ імені Івана Огієнка – moodle. Підготовка та виконання завдань лабораторних робіт є обов'язковими для кожного студента.

Академічна доброчесність. Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикавання джерел списування, втручання в роботу інших студентів становлять приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.

Відвідання занять. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати всі заняття. Виконання усіх лабораторних робіт є обов'язковим. Студенти зобов'язані дотримуватися термінів виконання усіх видів робіт, передбачених курсом.

Форми поточного та підсумкового контролю. Поточний контроль реалізується на лабораторних заняттях. Підсумковий контроль не здійснюється, а виставляється у формі заліку на основі поточного контролю.

10. Програма курсу

Модуль 1. Основи R та Інструментарій

- Тема 1. Вступ до R та RStudio.
- Тема 2. Ключові структури даних в R.
- Тема 3. Data Frames — "таблиці" в R.
- Тема 4. Програмування в R.

Модуль 2. Отримання та "Очищення" даних

- Тема 5. Імпорт даних з файлів.
- Тема 6. Інтеграція R з Базами Даних.
- Тема 7. Трансформація даних (Tidyverse: dplyr).
- Тема 8. Агрегація та очищення даних.

Модуль 3. Дослідження та Візуалізація

- Тема 9. Дослідницький аналіз даних (EDA) та "ГраMATика графіки".
- Тема 10. Побудова просунутих візуалізацій (ggplot2).
- Тема 11. Data Storytelling: Комунікація результатів.

Модуль 4. Інтеграція та фінальний проєкт

- Тема 12. Фінальний проєкт: Наскрізний аналітичний кейс.

11. Система оцінювання та вимоги

Курс складається з 1-го навчального (змістовного) модуля.

Поточний контроль полягає в перевірці теоретичних знань та практичних умінь і навичок під час лабораторних занять.

Результати перевірки самостійної роботи, в тому числі виконання домашніх завдань, студентів також входять до поточного контролю. Максимальний бал оцінки поточної успішності студентів на практичних заняттях приймається рівним 12.

Підсумковий контроль проводиться у формі заліку.

Рейтингова оцінка студентам виставляється відповідно до Положення про рейтингову систему оцінювання навчальних досягнень здобувачів вищої освіти Кам'янець-Подільського національного університету імені Івана Огієнка після проведення всіх навчальних занять та контрольних заходів з навчальної дисципліни.

Студенти, які мають академічну заборгованість за результатами підсумкового контролю, зобов'язані ліквідувати її в терміни, визначені графіком ліквідації академічної заборгованості.

12. Рекомендована література

Основна література та інформаційні ресурси:

1. Banachewicz K., Massaron L. The Kaggle book: Data analysis and machine learning for competitive data science. Birmingham, UK: Packt Publishing Ltd., 2022. 531 p.
2. Google Cloud Data Analytics Certificate. URL: <https://www.cloudskillsboost.google/paths/420>
3. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of massive data sets. Cambridge university press.
4. Ланде Д.В., Субач І.Ю., Гладун А.Я. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. Київ 2021. – 168 с.
5. Фред Нванганга, Майк Чеппелл. Practical Machine Learning in R. Wiley. John Wiley & Sons, LTD. 2020.
6. Елінор Джонс, Майкл Дж. Кроулі, Саймон Харден. The R Book. Wiley. John Wiley & Sons, LTD. 2022.