



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

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Сучасні методи аналізу даних з використанням R

1. Загальна інформація про курс. Викладачі

Іванюк Віталій Анатолійович, доктор технічних наук, доцент, завідувач кафедри комп'ютерних наук
Телефон: +38 (068) 746 33 28

E-mail: wivanyuk@kpnpu.edu.ua

Профайл: <https://cs.kpnpu.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@kpnpu.edu.ua

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

Мова викладання: українська

2. Анотація до курсу

Цей курс спрямований на ознайомлення студентів з сучасними методами аналізу даних за допомогою мови програмування R. Студенти отримають практичні навички з обробки, візуалізації та моделювання даних, використовуючи потужні інструменти, які пропонує R. Курс охоплює широкий спектр тем, від основної статистики до складних методів машинного навчання.

3. Мета і завдання курсу

Метою курсу є підготовка фахівців, здатних ефективно працювати з даними, застосовуючи сучасні статистичні та машинні методи.

Завдання курсу

- Ознайомити студентів з основами мови програмування R.
- Навчити основним методам статистичного аналізу даних.
- Ознайомити з різними типами даних та їх особливостями.
- Розглянути методи візуалізації даних за допомогою R.
- Навчити будувати різноманітні статистичні моделі.
- Ознайомити з основами машинного навчання та його застосуванням в аналізі даних.
- Розвинути навички вирішення практичних задач аналізу даних.

4. Результати навчання

Проводити попередній аналіз даних для виявлення закономірностей та аномалій.

Використовувати різноманітні статистичні тести для перевірки гіпотез.

Будувати різні типи графіків для візуалізації даних.

Використовувати методи регресійного аналізу для прогнозування.

Застосовувати методи класифікації для розподілу даних на категорії.

Будувати кластерні моделі для виявлення груп об'єктів.

Оцінювати точність побудованих моделей.

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
Семинарські 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.

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, фабрикування 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

Тема 1. Вступ до R

- Інсталяція та налаштування R та RStudio.
- Основи синтаксису R.
- Структури даних: вектори, матриці, дата фрейми.
- Операції з даними.

Тема 2. Статистичний аналіз даних

- Описова статистика.
- Перевірка статистичних гіпотез.
- Кореляційний аналіз.
- Регресійний аналіз.

Тема 3. Візуалізація даних

- Пакет ggplot2.
- Створення різних типів графіків (гістограми, діаграми розсіювання, бокс-плоти тощо).
- Інтерактивна візуалізація.

Тема 4. Машинне навчання в R

- Лінійні моделі.
- Методи класифікації (логістична регресія, дерева рішень, випадкові ліси).
- Методи кластеризації (k-means, ієрархічна кластеризація).
- Оцінка моделей.

Тема 5. Практичні завдання

- Аналіз реальних наборів даних.
- Розробка власних проектів.
- Презентація результатів.

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.