



**ЦЕНТРАЛЬНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

Кафедра кібербезпеки та програмного забезпечення



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ПОПЕРЕДНІЙ АНАЛІЗ ТА ПІДГОТОВКА ДАНИХ
першого рівня вищої освіти

м. Кропивницький

1. Загальна інформація

Назва дисципліни	Попередній аналіз та підготовка даних
Викладач	Лектор – Кислун Олег Андрійович, кандидат технічних наук, доцент кафедри кібербезпеки та програмного забезпечення
Контактний телефон	службовий: (0522)390-449 – робочі дні з 8 ³⁰ до 14 ²⁰
E-mail:	kyslun@gmail.com
Консультації	Очні консультації відповідно до затвердженого графіку консультацій Онлайн консультації засобами електронної пошти, месенджерів (Facebook-Messenger / Viber / Telegram) у робочі дні

2. Анотація дисципліни

Дисципліна «Попередній аналіз та підготовка даних» спрямована на формування у здобувачів вищої освіти теоретичних знань і практичних навичок роботи з даними, необхідних для подальшого статистичного аналізу, побудови моделей та прийняття обґрунтованих управлінських рішень. У межах курсу розглядаються методи збору, структурування, опису, форматування та попередньої обробки електронних масивів даних, а також принципи підготовки табличних даних до аналітичної обробки.

3. Мета і завдання дисципліни

Метою викладання дисципліни «Попередній аналіз та підготовка даних» є формування у здобувачів вищої освіти системних теоретичних знань і практичних умінь щодо попереднього аналізу та підготовки даних для подальшого застосування статистичних, математичних і аналітичних методів, а також використання інструментів обробки даних для отримання достовірної інформації та підтримки управлінських рішень.

Основними **завданнями** вивчення дисципліни є оволодіння здобувачами вищої освіти методами первинного аналізу даних і практиками підготовки електронних масивів до подальшої аналітичної обробки. Курс передбачає формування навичок роботи з табличними даними, їх структурування, очищення, форматування та опису, а також опанування методів виявлення аномальних значень, виправлення помилок і заповнення пропусків. Окремим завданням дисципліни є набуття практичного досвіду застосування сучасних програмних інструментів для аналізу та інтерпретації даних, підготовки аналітичних висновків і прогнозів у прикладних задачах.

4. Результати навчання

У результаті вивчення дисципліни здобувач вищої освіти повинен:

– знати методи та технології попереднього аналізу й підготовки даних для подальшого моделювання та дослідження об’єктів і процесів;

- знати основні підходи до пошуку в даних закономірностей, взаємозв'язків і тенденцій;
- знати засоби аналітичної обробки даних, принципи формування та перевірки гіпотез, а також вибору відповідних моделей аналізу.
- вміти виконувати попередній аналіз даних, оцінювати їх якість та придатність до подальшого використання;
- вміти здійснювати підготовку даних до аналізу (очищення, форматування, обробку пропусків і аномальних значень) для підтримки ухвалення управлінських рішень.

5. Обсяг дисципліни

Ознака дисципліни	
Кількість кредитів / годин	4/120
Нормативна / вибіркова	вибіркова
Вид підсумкового контролю	залік

6. Політика дисципліни

Академічна доброчесність:

Очікується, що студенти будуть дотримуватися принципів академічної доброчесності, усвідомлювати наслідки її порушення. Детальніше за посиланням URL : <http://www.kntu.kr.ua/doc/dobro.pdf>

Відвідування занять

Відвідування занять є важливою складовою навчання. Очікується, що всі студенти відвідають лекції і лабораторні заняття курсу.

Пропущені заняття повинні бути відпрацьовані не пізніше, ніж за тиждень до залікової сесії.

Поведінка на заняттях

Недопустимість: запізнь на заняття, списування та плагіат, несвоєчасне виконання поставленого завдання.

При організації освітнього процесу в Центральнотехнічному національному університеті студенти, викладачі та адміністрація діють відповідно до: Положення про організацію освітнього процесу; Положення про організацію вивчення навчальних дисциплін вільного вибору; Положення про рубіжний контроль успішності і сесійну атестацію студентів ЦНТУ, Кодексу академічної доброчесності ЦНТУ.

7. Програма навчальної дисципліни

Тема 1. Вступ до аналізу та підготовки даних

Роль даних у прийнятті рішень. Етапи роботи з даними: збір, підготовка, аналіз, інтерпретація.

Тема 2. Типи даних та джерела формування наборів даних

Структуровані й неструктуровані дані. Формати файлів і табличні набори даних. Первинні вимоги до якості даних.

Тема 3. Описова статистика та первинний аналіз даних

Основні статистичні характеристики. Розподіли, кореляція. Виявлення первинних тенденцій і відхилень.

Тема 4. Очищення даних та усунення помилок

Пошук некоректних значень, дублювань, логічних помилок. Правила очищення та виправлення даних.

Тема 5. Обробка пропущених значень та аномалій

Типи пропусків. Методи заповнення пропущених значень. Виявлення та обробка викидів (outliers).

Тема 6. Перетворення, нормалізація та узгодження даних

Масштабування, нормалізація, стандартизація. Перетворення ознак та приведення даних до єдиного формату.

Тема 7. Дослідницький аналіз даних (EDA) та візуалізація

Візуалізація даних. Пошук закономірностей, взаємозв'язків і тенденцій на основі EDA.

Тема 8. Формування і перевірка гіпотез на основі даних

Побудова гіпотез. Перевірка статистичних припущень. Інтерпретація результатів аналізу.

Тема 9. Підготовка даних для моделювання та прогнозування

Відбір ознак, підготовка навчальних вибірок. Балансування даних. Підготовка даних до побудови моделей.

Тема 10. Інструментальні засоби аналізу та підготовки даних

Огляд програмних засобів для обробки даних. Практичні сценарії підготовки наборів даних для прикладних задач.

8. Система оцінювання та вимоги

Види контролю: поточний, підсумковий.

Методи контролю: спостереження за навчальною діяльністю, усне опитування, письмовий контроль, тестовий контроль.

Форма підсумкового контролю: залік.

Поточний контроль передбачає оцінювання виконання лабораторних робіт з урахуванням правильності виконання завдань, якості захисту та дотримання термінів подання, а також рівня засвоєння теоретичного матеріалу.

Підсумковий контроль здійснюється у формі заліку та спрямований на перевірку рівня опанування дисципліни і здатності застосовувати знання на практиці. Підсумкова оцінка формується за результатами роботи протягом семестру; у разі набрання понад 60 балів залік може бути зарахований за поточним рейтингом.

9. Рекомендована література

Базова

1. Інтелектуальний аналіз даних. Частина 1 / М.В. Талах, В.В. Дворжак. Чернівці: Технодрук, 2022. 367 с.
2. Глибинне навчання для комп'ютерного зору. Частина 1 / В.В. Дворжак, М.В. Талах. Чернівці: Технодрук, 2022 р. 271 с.
3. Lee W.-M. Python Machine Learning Wiley, 2019. 307 p. ISBN: 978-1-119-54567-5.
4. Chinnamgari Sunil Kumar. R Machine Learning Projects: Implement supervised, unsupervised, and reinforcement learning techniques using R 3.5 Packt Publishing, 2019. 325 p. ISBN 978-1-78980-7943.
5. Lesmeister C. Mastering Machine Learning with R 2nd ed. Packt Publishing, 2017. 420 p. ASIN B01N5XJ3O0.

Допоміжна

6. Saiz A.Z., Gonzalez C.Q., Gil L.H., Ruiz D.M. An Introduction to Data Analysis in R: Hands- on Coding, Data Mining, Visualization and Statistics from Scratch Springer, 2020. 291 p. ISBN 978-3-030-489977 (eBook).
7. George F. Luger. Artificial Intelligence: Structures and Strategies for Complex Problem Solving, 5e Paperback. 2004. – 928 p.
8. Ranga Suri N.N.R., Murty N., Athithan M.G. Outlier Detection: Techniques and Applications. A Data Mining Perspective / N.N.R. Ranga Suri, N. Murty, M.G. Athithan.,- Cham: Springer International Publishing, 2019, 214 p.

Інформаційні ресурси

9. Онлайн-курси Prometheus. – URL: <https://prometheus.org.ua/>
10. Онлайн-курси Coursera. – URL: <https://www.coursera.org>
11. Академія Cisco. – URL: <https://www.netacad.com>
12. Он-лайн ресурс з інформаційних технологій. – URL:<https://habr.com>
13. Он-лайн ресурс з інформаційних технологій. – URL:<https://dou.ua/>
14. Пошукова система. – URL:<https://www.google.com/>
15. Он-лайн ресурс перегляду відеоуроків.– URL:<https://www.youtube.com>