



Силабус курсу Прикладний аналіз даних

Освітньо-професійна програма «Системний аналіз»
Ступінь вищої освіти – бакалавр
Спеціальність: 124 «Системний аналіз»

Рік навчання: 3 Семестр: 5

Кількість кредитів: 6 Мова викладання: українська

Керівник курсу

ППП

к.е.н., доцент Данилюк Ірина Вадимівна

Контактна інформація

Irynadanylyuk9@gmail.com,

Опис

Мета навчальної дисципліни «Прикладний аналіз даних» — формування у студентів здатності аналізувати великі масиви даних, використовуючи сучасні інструменти та методи інтелектуального аналізу, зокрема Excel, MATLAB та SPSS. Курс спрямований на розвиток як теоретичних знань, так і практичних навичок для виявлення прихованих закономірностей, прогнозування та прийняття обґрунтованих рішень.

В курсі вивчаються основні методи інтелектуального аналізу даних, включаючи класифікацію, кластеризацію, регресійний аналіз, асоціативний аналіз, а також візуалізацію даних. Особлива увага приділяється методам попередньої обробки даних, аналізу часових рядів, а також зменшенню розмірності даних.

Оволодіння курсом повинне забезпечити студентам навички практичного використання методів інтелектуального аналізу даних для вирішення задач у різних сферах діяльності, зокрема аналізу великих даних, прогнозування та прийняття рішень на основі отриманих моделей.

Структура курсу

Години (лек. / сем.)	Тема	Результати навчання	Завдання
4/4	Тема 1. Встановлення пакету R та елементарні операції в оболонці	Володіти поняттями прикладного аналізу даних. Знати основні етапи процесу видобування знань із даних. Вміти встановити пакет R та здійснювати операції на його основі	Практичне завдання
4/4	Тема 2. Метод прогнозування часових рядів	Розрізняти типи даних: категорійні, числові, часові ряди тощо. Виявляти проблеми з даними: відсутні значення, шум, дублікати. Знати основні методи нормалізації та очищення даних.	Практичне завдання
4/4	Тема 3. Розклад часових рядів	Знати основні поняття часових рядів: тренд, сезонність, шум. Уміти проводити прогнозування на основі часових рядів.	Практичне завдання

4/4	Тема 4. Метод Холта – Вінтерса	Знати простий метод експоненційного згладжування Вміти створити часовий ряд, спрогнозувати дані за допомогою SES	Практичне завдання
4/4	Тема 5. Метод ARSMA	Знати методику, яка використовується для узгодження даних часових рядів і прогнозування. Вміти застосовувати моделі ARIMA для часових рядів	Практичне завдання
4/4	Тема 6. Побудова лінійної моделі	Уміти здійснити оцінку точності моделі. Знати процес додавання гнучкості до лінійних моделей	Практичне завдання
4/4	Тема 7. Теорія методу найближчого сусіда	Уміти використовувати моделі KNN в якості класифікатора Знати підхід найближчих сусідів для регресій	Практичне завдання
6/4	Тема 8. Регресійний аналіз.	Уміти будувати лінійну регресію та поліноміальну регресію. Володіти методом найменших квадратів.	Практичне завдання
6/6	Тема 9. Задачі класифікації	Володіти основними поняттями класифікації. Вміти будувати моделі класифікації: дерева рішень, логістична регресія. Здійснювати оцінку якості моделей класифікації.	Практичне завдання
6/6	Тема 10. Кластеризація	Володіти основними поняттями кластеризації. Знати методи кластеризації: метод k-середніх, ієрархічна кластеризація. Володіти метриками для оцінки якості кластеризації.	Практичне завдання

Політика оцінювання

Підсумковий бал (за 100-бальною шкалою) з дисципліни визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

Оцінювання

Модуль 1		Модуль 2		Модуль 3	Модуль 4	Модуль 5
Поточне опитування	Модульний контроль	Поточне опитування	Модульний контроль	Тренінги	Самостійна робота	Іспит
10%	10%	10%	10%	5%	15%	40%
Визначається як середнє арифметичне оцінок, отриманих на практичних заняттях (кожен здобувач має отримати не менше 5 оцінок)	Підсумкове тестування (20 тестів по 5 балів за кожен тест) за темами навчального курсу, теми 1 - 5	Оцінка визначається як середнє арифметичне за відповіді або проходження поточного тестування на кожному практичному занятті	Колоквіум за темами навчального курсу – оцінюється індивідуально за шкалою від 1 до 100 балів, охоплює теми 6-10	Оцінка, отримана за виконанні завдання	Оцінка, отримана за виконання завдання	Теоретичні запитання (2 по 20 балів); практичне завдання (60 балів) -макс. 100 балів

Шкала оцінювання:

За шкалою Університет	За національною шкалою	За шкалою ECTS
90-100	Відмінно	A (відмінно)
85-89	Добре	B (дуже добре)
75-84		C (добре)
65-74	Задовільно	D(задовільно)
60-64		E (достатньо)
35-59	Незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)

•

• **Політики курсу**

- **Політика щодо дедлайнів та перескладання:** Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (-20 балів). Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).
- **Політика щодо академічної доброчесності:** Усі письмові роботи перевіряються на наявність плагиату і допускаються до захисту із коректними текстовими запозиченнями не більше 20%. Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних девайсів). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування.
- **Політика щодо відвідування:** Відвідування занять є обов'язковим компонентом оцінювання, за яке нараховуються бали. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу.

Рекомендовані джерела інформації

1. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2020. 152 с
2. Болюбаш Н.М. Інтелектуальний аналіз даних : навч. посіб. / Н.М. Болюбаш. – Миколаїв : Вид-во ЧНУ ім. Петра Могили, 2023. – 320 с. // https://www.researchgate.net/publication/379994907_Bolubas_N_M_Intelektualnij_an_aliz_danih
3. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. 92 с.
4. Лупан І. В. Інтелектуальний аналіз даних Data Mining: навчально- методичний посібник / Ірина Володимирівна Лупан. – Кропивницький: ФОП Піскова М. А., 2022. – 112 с // <https://dspace.cusu.edu.ua/items/f2aabfff-9146-43cc-8525-761437210a00>
5. Інтелектуальний аналіз даних: Комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 122 «Комп'ютерні науки та інформаційні технології», спеціалізацій «Інформаційні системи та технології проектування», «Системне проектування сервісів» / О. О. Сергеев-Горчинський, Г. В. Іщенко ; КПП ім. Ігоря Сікорського.– Київ : КПП ім. Ігоря Сікорського, 2020. – 73 с.

Додаткова література:

6. Jiajun, Z., Zong, C., & Xia, R. (2022). Text Data Mining. Springer.
7. Li, B., Yue, L., Jiang, J., Chen, W., Li, X., Long, G., Fang, F., & Yu, H. (Ред.). (2022). Advanced Data Mining and Applications. Springer International Publishing. <https://doi.org/10.1007/978-3-030-95408-6>
8. Park, L. A. F., Gomes, H. M., Doborjeh, M., Boo, Y. L., Koh, Y. S., Zhao, Y., Williams, G., & Simoff, S. (Ред.). (2022). Data Mining. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8746-5>
9. Tan, Y., & Shi, Y. (Ред.). (2022). Data Mining and Big Data. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8991-9>
10. Ufuk Baytar, C. (Ред.). (2022). The Future of Data Mining. Nova Science Publishers. <https://doi.org/10.52305/kcin5931>
11. Cao, J. (2023). E-Commerce Big Data Mining and Analytics. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-3588-8>
12. Konys, A., & Nowak-Brzezińska, A. (2023). Knowledge Engineering and Data Mining. MDPI. <https://doi.org/10.3390/books978-3-0365-6789-1>
13. Mai, T. T., Crane, M., & Bezbradica, M. (2023). Educational Data Mining und Learning Analytics. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-39607-7>
14. Olson, D. L., & Lauhoff, G. (2023). Descriptives Data-Mining. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-21274-1>
15. Shah, K., Shah, N., Sawant, V., & Parolia, N. (2023). Practical Data Mining Techniques and

Applications. Auerbach Publications. <https://doi.org/10.1201/9781003390220>

16. Zhang, H. (2023a). Handbook of Mobility Data Mining, Volume 1: Data Preprocessing and Visualization. Elsevier.

17. Zhang, H. (2023b). Handbook of Mobility Data Mining, Volume 2: Mobility Analytics and Prediction. Elsevier.

18. Zhang, H. (2023c). Handbook of Mobility Data Mining, Volume 3: Mobility DataDriven Applications. Elsevier.

19. Kaixin Wang. Introduction to Gaussian process regression, Part 1: The basics // <https://medium.com/data-science-at-microsoft/introduction-to-gaussian-process-regression-part-1-the-basics-3cb79d9f155f>

20. Kaixin Wang. Introduction to Gaussian process regression, Part 1: Application to predicting concrete strength // <https://medium.com/data-science-at-microsoft/introduction-to-gaussian-process-regression-part-2-application-to-predicting-concrete-strength-3facef04f639>

21. Naive Bayes Classifiers // <https://www.geeksforgeeks.org/naive-bayes-classifiers/>

22. Scikit-learn. User Guide // https://scikit-learn.org/1.5/user_guide.html

23. Scikit-learn. Stochastic Gradient Descent // <https://scikit-learn.org/1.5/modules/sgd.html>

24. Scikit-learn. SGD: Maximum margin separating hyperplane // https://scikit-learn.org/dev/auto_examples/linear_model/plot_sgd_separating_hyperplane.html

25. Scikit-learn. SGD: Weighted samples // https://scikit-learn.org/dev/auto_examples/linear_model/plot_sgd_weighted_samples.html

26. Scikit-learn. NearestNeighbors. // <https://scikit-learn.org/1.5/modules/neighbors.html>