



Силабус курсу

Інтелектуальний аналіз даних

Ступінь вищої освіти-бакалавр

Спеціальність 122 «Комп'ютерні науки»

Освітньо-професійна програма: «Комп'ютерні науки»

Рік навчання: III, Семестр: I

Кредитів: 5 Мова викладання: українська

Керівник курсу

ППП

к.т.н., доцент Лип'яніна-Гончаренко Христина Володимирівна

Контактна інформація

xrustya.com@gmail.com, +380968000965

Опис дисципліни

Даний курс знайомить з предметною областю взаємозв'язків з наукою про дані (data science) та аналізом даних (data analytics). Оволодіння навичками оперування великими даними шляхом використання спеціалізованих програмно-апаратних засобів, зокрема хмарних сервісів, спеціалізованих систем зберігання даних, розподілених файлових систем тощо. Розвиток уміння працювати з великими даними з урахуванням їх ключових характеристик: обсягу, різноманітності, мінливості й забезпечуючи відповідний рівень швидкості їх опрацювання. Розвиток навичок ефективного використання математичного, алгоритмічного та програмного забезпечення для розв'язання основних задач предметної області великих даних.

Завдання навчальної дисципліни «Інтелектуальний аналіз даних» – опанувати базові принципи побудови моделей даних, ознайомлення з концепцією Knowledge Discovery in Data (виявлення знань в даних) і Data Mining («видобування» знань), навчитися ефективно використовувати методи здобуття знань з великих масивів даних, ознайомитися з основними типами задач, що можуть бути вирішені за допомогою методів інтелектуального аналізу даних та отримати практичні навички з використання інструментальних засобів інтелектуального аналізу даних при вирішенні прикладних задач і навчитися інтерпретувати отримані результати.

Структура курсу

Години (лек./лаб.)	Тема	Результати навчання	Завдання
2/-	Тема 1. Концепції інтелектуального аналізу даних та DataMining	Ознайомлення з необхідністю аналізу великих, складних, багатих інформацією наборів даних, цілями та основними завданнями процесу інтелектуального аналізу даних. Знати технології інтелектуального аналізу даних та Data Mining, ітераційний характер процесу інтелектуального аналізу даних та його основних кроків, вплив якості даних на процес інтелектуального аналізу даних та концепцію великих даних та науки про дані. Вміти застосовувати алгоритми Data Mining для вирішення реальних завдань аналізу даних, визначення корисних шаблонів та тенденцій у даних.	Питання
2/4	Тема 2. Попередня	Знати методи попередньої обробки даних, такі як нормалізація, трансформація та	Лабораторна робота

	обробка та аналіз даних	очищення даних. Вміти проводити експлоративний аналіз даних, виявляти аномалії та готувати дані для подальшого моделювання.	
2/4	Тема 3. Стратегії обробки даних з відсутністю інформації та залежністю від часу	Знати методи обробки пропущених значень та часових рядів. Вміти використовувати відповідні техніки для аналізу та прогнозування даних, що змінюються в часі.	Лабораторна робота
2/4	Тема 4. Аналіз великих наборів даних та методи зменшення розмірності	Знати принципи аналізу великих наборів даних та основні методи зменшення розмірності, такі як аналіз головних компонентів та лінійна дискримінантна аналіз. Вміти застосовувати ці методи для вирішення реальних завдань аналізу даних.	Лабораторна робота
2/4	Тема 5. Методи візуалізації	Знати основні методи візуалізації даних, такі як гістограми, точкові діаграми та теплові карти. Вміти використовувати візуалізацію для інтерпретації результатів аналізу та комунікації з іншими учасниками проекту.	Лабораторна робота
2/4	Тема 6. Статистичний аналіз	Знати основні статистичні поняття та методи аналізу даних, включаючи гіпотезу перевірки, кореляційний аналіз та регресійний аналіз. Вміти проводити статистичний аналіз для виявлення залежностей у даних.	Лабораторна робота
2/-	Тема 7. Методи оцінки результатів	Знати методи оцінки результатів моделей машинного навчання, такі як крос-валідація, ROC-аналіз та матриця плутанини. Вміти використовувати ці методи для оцінки та порівняння різних моделей.	Питання
4/4	Тема 8. Лінійний дискримінантний аналіз	Знати принципи лінійного дискримінантного аналізу та його застосування для задач класифікації. Вміти застосовувати лінійний дискримінантний аналіз для вирішення реальних завдань аналізу даних.	Лабораторна робота
6/6	Тема 9. Регресійні моделі	Знати основні типи регресійних моделей, методи їхнього обрання та оцінювання параметрів. Розуміти принципи відбору змінних, переваги та недоліки лінійної та логістичної регресії, методи регуляризації та їх вплив на якість моделей. Вміти інтерпретувати результати регресійного аналізу, застосовувати регресійні моделі для прогнозування значень цільової змінної та аналізу взаємозв'язку між змінними.	Лабораторна робота
2/6	Тема 10. Прогнозування часових рядів	Знати основні методи та моделі прогнозування часових рядів, такі як ARIMA, експоненціальне згладжування та інші. Розуміти принципи сезонної декомпозиції, трендів та циклічних компонент часових рядів. Вміти використовувати спеціалізоване програмне забезпечення для аналізу та	Лабораторна робота

		прогнозування часових рядів, оцінювати точність прогнозів та вибирати оптимальні моделі на основі діагностичних перевірок та критеріїв якості.	
2/6	Тема 11. TextMining і WebMining	Знати основні методи та інструменти, що використовуються для аналізу тексту та веб-даних. Вміти застосовувати методи TextMining для аналізу великих обсягів текстових даних, використовувати WebMining для збору та аналізу веб-даних, визначати ключові слова та теми в текстах, вирішувати завдання сентимент-аналізу та ідентифікації авторства.	Лабораторна робота

Літературні джерела

1. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.
2. Бахрушин В. Є. Методи аналізу даних: Навч. посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2019. – 268 с.
3. 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 ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2020. – 73 с.
4. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2020. 152 с.
5. Jiajun, Z., Zong, C., & Xia, R. (2022). Text Data Mining. Springer.
6. 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>
7. 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>
8. Tan, Y., & Shi, Y. (Ред.). (2022). Data Mining and Big Data. Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-8991-9>
9. Ufuk Baytar, C. (Ред.). (2022). The Future of Data Mining. Nova Science Publishers. <https://doi.org/10.52305/kcin5931>
10. Cao, J. (2023). E-Commerce Big Data Mining and Analytics. Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-3588-8>
11. Konys, A., & Nowak-Brzezińska, A. (Ред.). (2023). Knowledge Engineering and Data Mining. MDPI. <https://doi.org/10.3390/books978-3-0365-6789-1>
12. Mai, T. T., Crane, M., & Bezbradica, M. (Ред.). (2023). Educational Data Mining und Learning Analytics. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-39607-7>
13. Olson, D. L., & Lauhoff, G. (2023). Deskriptives Data-Mining. Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-21274-1>
14. Shah, K., Shah, N., Sawant, V., & Parolia, N. (2023). Practical Data Mining Techniques and Applications. Auerbach Publications. <https://doi.org/10.1201/9781003390220>
15. Zhang, H. (2023a). Handbook of Mobility Data Mining, Volume 1: Data Preprocessing and Visualization. Elsevier.
16. Zhang, H. (2023b). Handbook of Mobility Data Mining, Volume 2: Mobility Analytics and Prediction. Elsevier.
17. Zhang, H. (2023c). Handbook of Mobility Data Mining, Volume 3: Mobility Data-Driven Applications. Elsevier.

Політика оцінювання

Політика щодо дедлайнів та перескладання: Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

Політика щодо академічної доброчесності: Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних пристроїв). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування.

Політика щодо відвідування: Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбутись в он-лайн формі за погодженням із керівником курсу.

Оцінювання

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен
20 %	20 %	20 %	40 %
1. Виконання та захист лабораторних робіт (5 робіт по 5 балів) – 25 балів 2. Модульна контрольна робота – 75 балів	1. Виконання та захист лабораторних робіт (5 робіт по 5 балів) – 25 балів 2. Ректорська контрольна робота – 75 балів	1. Виконання завдань під час тренінгу – 20 балів 2. Написання та захист КПЗ – 80 балів	1. Тестові завдання (25 тестів по 2 бали) – 50 балів 2. Завдання 1 – 25 балів 3. Завдання 2 – 25 балів

Шкала оцінювання:

За шкалою ЗУНУ	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)