**Силабус курсу**

# Інтелектуальний аналіз даних

**Ступінь вищої освіти-бакалавр**
**Спеціальність 122 «Комп'ютерні науки»**
**Освітньо-професійна програма: «Штучний інтелект»**

**Рік навчання:** III, **Семестр:** V

**Кредитів:** 5 **Мова викладання:** українська

## Керівник курсу


| | |
|---|---|
| **ПІП** | к.т.н., доцент Ліп'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 знать 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 данихта 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 |
| 2/2 | Тема 2. Попередня обробка та аналіз | Знати методи попередньої обробки даних, такі як нормалізація, трансформація та очищення | Практична робота |

| | | | |
|---|---|---|---|
| | даних | даних. Вміти проводити експлоративний аналіз даних, виявляти аномалії та готувати дані для подальшого моделювання. | |
| 2/2 | 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. | Практична 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. | Практична робота |
| 2/2 | 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. | Практична робота |
| 2/2 | 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. | Практична робота |
| 2/2 | 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. | Практична робота |
| 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. | Практична робота |
| 4/4 | 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. | Практична робота |
| 4/4 | 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 прогнозування часових рядів, оцінювати точність прогнозів та вибирати оптимальні моделі на основі діагностичних перевірок та критеріїв якості. | Практична робота |

| | | | |
|---|---|---|---|
| 4/4 | 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. | Практична робота |

## Літературні джерела

### Основні джерела інформації:


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. Електронний курс з дисципліни «Інтелектуальний аналіз даних» для студентів галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки» на платформі Moodle ЗУНУ /Ліп’яніна-Гончаренко Х.В., Тернопіль, 2021.
3. Методичні вказівки до виконання лабораторних робіт з дисципліни «Інтелектуальний аналіз даних» для студентів освітньо-професійної програми «Штучний інтелект» спеціальності 122 «Комп’ютерні науки» першого (бакалаврського) рівня вищої освіти / Укладач: Ліп’яніна-Гончаренко Х.В. ¬– Тернопіль, ЗУНУ, 2024, – с.101
4. Методичні вказівки та завдання для самостійної роботи з дисципліни «Інтелектуальний аналіз даних» для студентів освітньо-кваліфікаційного рівня «бакалавр» спеціальності 122 «Комп’ютерні науки» / Укладач: Ліп’яніна-Гончаренко Х.В. ¬ Тернопіль, ЗУНУ, 2024, – с.25.
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. Сергеєв-Горчинський, Г. В. Іщенко ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2020. – 73 с.
6. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченко, 2020. 152 с.
7. Ліп’яніна-Гончаренко, Х. (2024). Узагальнений принцип синтезу інформаційної технології інтелектуального аналізу соціально-економічних даних ТГ. *Measuring and Computing Devices in Technological Processes*, (1), 359–367. https://doi.org/10.31891/2219–9365–2024–77–48
8. Ліп’яніна-Гончаренко, Х. (2024). Методологія аналізу соціально-економічних даних територіальних громад. *Herald of Khmelnytskyi National University. Technical Sciences*, 337(3(2)), 446–450. https://doi.org/10.31891/2307–5732–2024–337–3–67
9. Ліп’яніна-Гончаренко, Х. (2024). Методи аналізу соціально–економічних даних територіальних громад для адаптивного управління ресурсами. *Computer Systems and Information Technologies*, (3), 92–97. https://doi.org/10.31891/csit–2024–3–12
10. Ліп'яніна-Гончаренко, Х. В. (2023). Метод формування навчальної вибірки для масивів даних на основі машинного навчання. *Наукові записки НаУКМА. Комп'ютерні науки*, (6), 30–35. https://doi.org/10.18523/2617–3808.2023.6.30–35
11. Ліп’яніна-Гончаренко, Х., & Юрків, Х. (2023). Методи бустингового машинного навчання для нестаціонарних часових рядів. *Measuring and Computing Devices in Technological Processes*, (3), 19–30. https://doi.org/10.31891/2219–9365–2023–75–2
12. Ліп’яніна-Гончаренко, Х., Комар, М., Юрків, Х., & Лук’янчук, В. (2023). Концептуальна модель інтелектуальної оцінки наслідків техногенних катастроф. *Вісник Хмельницького національного університету. Серія: Технічні науки*, 329(6), 230–237. https://doi.org/10.31891/2307–5732–2023–329–6–230–237
13. Комар, М., Ліп’яніна-Гончаренко, Х., Кіт, І., Мадараш, Р., & Юрків, Х. (2023). Інтелектуальний метод виявлення джерел мультилінгвальної дезінформації. *Measuring and Computing Devices in Technological Processes*, (2), 221–230. https://doi.org/10.31891/2219–9365–2023–74–31
14. Lipianina-Honcharenko, K., Wolff, C., Sachenko, A., Kit, I., & Zahorodnia, D. (2023).

Intelligent method for classifying the level of anthropogenic disasters. *Big Data and Cognitive Computing*, 7(3), 157. https://doi.org/10.3390/bdcc7030157 (Індексована в наукометричній базі Web of Science та Scopus, що відноситься до Q2 відповідно до SCImago Journal & Country Rank)

15. Lipianina-Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S., & Kit, I. (2023). Intelligent information system for product promotion in internet market. *Applied Sciences*, 13(17), 9585. https://doi.org/10.3390/app13179585 (Індексована в наукометричній базі Web of Science та Scopus, що відноситься до Q2 відповідно до SCImago Journal & Country Rank)

16. Lipianina-Honcharenko, K., Komar, M., Osolinskyi, O., Shymanskyi, V., Havryliuk, M., & Semaniuk, V. (2024). Intelligent waste–volume management method in the smart city concept. *Smart Cities*, 7(1), 78–98. https://doi.org/10.3390/ smartcities7010004 (Індексована в наукометричній базі Web of Science та Scopus, що відноситься до Q1 відповідно до SCImago Journal & Country Rank)

17. Lipianina-Honcharenko, K., Komar, M., Melnyk, N., & Komarnytsky, R. (2024). Sustainable information system for enhancing virtual company resilience through machine learning in smart city socio-economic scenarios. *Economics*, 12(2). https://doi.org/10.2478/eoik–2024–0022 (Індексована в наукометричній базі Web of Science та Scopus, що відноситься до Q2 відповідно до SCImago Journal & Country Rank)

18. Lipianina-Honcharenko, K., Bodyanskiy, Y., Kustra, N., & Ivasechko, A. (2024). OLTW–TEC: Online learning with sliding windows for text classifier ensembles. *Frontiers in Artificial Intelligence*, 7, 1401126. https://doi.org/10.3389/frai.2024.1401126 (Індексована в наукометричній базі Scopus та відноситься до Q2 відповідно до SCImago Journal & Country Rank)

19. Bodyanskiy, Y. V., Lipianina-Honcharenko, K. V., & Sachenko, A. O. (2022). Ensemble of adaptive predictors for multivariate nonstationary sequences and its online learning. *Radio Electronics, Computer Science, Control*, (4(67)), 91–97, https://doi.org/10.15588/1607–3274–2023–4–9 (Індексована в наукометричній базі Web of Science)

20. Lipianina-Honcharenko, K., Bodyanskiy, Y., & Sachenko, A. (2023). Intelligent information system of the city's socio-economic infrastructure. *System Research and Information Technologies*, (3), 108–120. https://doi.org/10.20535/srit.2308–8893.2023.3.08 (Індексована в наукометричній базі Scopus та відноситься до Q4 відповідно до SCImago Journal & Country Rank)

**Додаткові джерела інформації:**

21. Jiajun, Z., Zong, C., & Xia, R. (2022). Text Data Mining. Springer.

22. 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

23. 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

24. Tan, Y., & Shi, Y. (Ред.). (2022). Data Mining and Big Data. Springer Nature Singapore. https://doi.org/10.1007/978-981-19-8991-9

25. Ufuk Baytar, C. (Ред.). (2022). The Future of Data Mining. Nova Science Publishers. https://doi.org/10.52305/kcin5931

26. Cao, J. (2023). E-Commerce Big Data Mining and Analytics. Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3588-8

27. Konys, A., & Nowak-Brzezińska, A. (2023). Knowledge Engineering and Data Mining. MDPI. https://doi.org/10.3390/books978-3-0365-6789-1

28. 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

29. Olson, D. L., & Lauhoff, G. (2023). Deskriptives Data-Mining. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-21274-1

30. Shah, K., Shah, N., Sawant, V., & Parolia, N. (2023). Practical Data Mining Techniques and Applications. Auerbach Publications. https://doi.org/10.1201/9781003390220

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

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

Prediction. Elsevier.
33. Zhang, H. (2023c). Handbook of Mobility Data Mining, Volume 3: Mobility Data-Driven Applications. Elsevier.

**Перелік баз даних, які можуть бути використані для реалізації практичних задач:**

1. World Bank Open Data - https://data.worldbank.org/
2. Kaggle Datasets - https://www.kaggle.com/datasets
3. UCI Machine Learning Repository - https://archive.ics.uci.edu/ml/
4. European Social Survey (ESS) - https://www.europeansocialsurvey.org/
5. National Climatic Data Center (NCDC) - https://www.ncei.noaa.gov/
6. Global Terrorism Database (GTD) - https://www.start.umd.edu/gtd/
7. Google Trends - https://trends.google.com/trends/
8. OECD Data - https://data.oecd.org/
9. IMDb Datasets - https://www.imdb.com/interfaces/
10. HealthData.gov - https://www.healthdata.gov/
11. OpenStreetMap - https://www.openstreetmap.org/
12. US Census Bureau - https://www.census.gov/data.html
13. Yelp Open Dataset - https://www.yelp.com/dataset
14. Eurostat - https://ec.europa.eu/eurostat
15. Financial Data from Yahoo Finance - https://finance.yahoo.com/
16. CDC Wonder - https://wonder.cdc.gov/
17. GDELT Project - https://www.gdeltproject.org/
18. FRED Economic Data - https://fred.stlouisfed.org/
19. Airbnb Data - https://www.airbnb.com/data
20. NASA Earth Observing System Data and Information System (EOSDIS) - https://earthdata.nasa.gov/

## 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 тощо) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу.

## Оцінювання

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Інтелектуальний аналіз даних» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

| Модуль 1 | | Модуль 2 | | Модуль 3 | Модуль 4 | Модуль 5 |
|---|---|---|---|---|---|---|
| 10 % | 10 % | 10 % | 10 % | 5% | 15% | 40% |
| Поточне оцінювання | Модульний контроль 1 | Поточне оцінювання | Модульний контроль 2 | Тренінг. | Самостійна робота. | Екзамен. |

| | | | | | | |
|---|---|---|---|---|---|---|
| Оцінка визначається як середнє арифметичне з оцінок, отриманих за виконання і захист 4 лабораторних робіт | Модульна контрольна робота (25 тестових питань) | Оцінка визначається як середнє арифметичне з оцінок, отриманих за виконання і захист 4 лабораторних робіт | Модульна контрольна робота (одне теоретичне питання та одна задача) | Виконання завдань під час тренінгу | Виконання завдання для самостійної роботи | - 10 тестових питань.<br>- Одне теоретичне питання.<br>- Одна задача. |