

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МІЖНАРОДНИЙ ГУМАНІТАРНИЙ УНІВЕРСИТЕТ
Кафедри комп'ютерних наук та інформаційних технологій

Соловська І.М., Григор'єва Т.І.

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Методичні рекомендації
для самостійної роботи здобувачів,
які навчаються за другим (магістерським) рівнем вищої освіти
зі спеціальності 112 «Статистика»

Одеса - 2024 рік

Рекомендовано навчально-методичною радою Міжнародного гуманітарного університету (Протокол № 13 від 16.08.2024 р.)

Соловська І.М., Григор'єва Т.І.

Інтелектуальний аналіз даних: методичні рекомендації для самостійної роботи здобувачів, які навчаються за другим (магістерським) рівнем вищої освіти зі спеціальності 112 «Статистика» [Електронне видання] / Соловська І.М., Григор'єва Т.І. Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 15 с.

Методичні рекомендації з дисципліни «Інтелектуальний аналіз даних» розроблено відповідно до навчального плану зі спеціальності 112 «Статистика», вони складаються з навчальної програми курсу, методичних рекомендацій з проведення практичних занять і виконання завдань для самостійної роботи здобувачів, списку рекомендованої літератури.

Матеріали призначено для студентів Міжнародного гуманітарного університету, які навчаються зі спеціальності 112 «Статистика», галузь знань – 11 Математика та статистика.

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни
		денна форма навчання
Кількість кредитів – 4 Загальна кількість годин – 120	Галузь - 11 «Математика і статистика» Спеціальність – 112 «Статистика»	обов'язкова
		Рік підготовки:
		1-й
		Семестр
		1-й
Мова навчання – українська	Рівень вищої освіти – другий (магістерський) рівень	Лекції
		14/6 год.
		Практичні, семінарські
		16/6 год.
		Лабораторні
		-
		Самостійна робота та індивідуальні завдання
		90/108 год.
		Вид контролю:
		залік

Дисципліна **«Інтелектуальний аналіз даних»** забезпечує формування системи теоретичних та практичних знань та навичок щодо застосування методів інтелектуального аналізу для вирішення прикладних задач обробки масивів великих даних. Завданнями дисципліни є вивчення принципів, методів та інтелектуального аналізу даних для вирішення економічних, соціальних та бізнес задач за допомогою концепцій виявлення знань в даних Knowledge Discovery in Data та «видобування» знань Data Mining, навчання ефективно використовувати методи інтелектуального аналізу даних для отримання практичних навичок при вирішенні прикладних задач та візуалізації отриманих результатів.

Мета викладання дисципліни – формування системи теоретичних знань щодо процесу виокремлення, дослідження, моделювання та інтелектуального аналізу даних із застосуванням статистичних та математичних методів, набуття практичних навичок роботи з сучасними пакетами прикладних програм для рішення задач аналізу та візуалізації даних Data Mining, що виникають в процесі професійної діяльності.

Передумови для вивчення дисципліни є знання і вміння, отримані студентом при вивченні навчальних дисциплін бакалаврської підготовки. Знання і вміння, отримані студентом при вивченні даної навчальної дисципліни, можуть бути використані при паралельному вивченні таких дисциплін: «Великі масиви даних», «Сучасні проблеми теорії ймовірностей та математичної статистики».

2. ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Студенти повинні знати:

- методи та засоби збору, аналізу, агрегації та обробки великих масивів даних Big Data різних видів,
- математичні моделі систем великих масивів даних та статистичні методи їхньої обробки,
- принципи формування хмарної інфраструктури різних технологій та сервісів, механізми управління хмарною інфраструктурою,
- статистичні методи обробки великих масивів даних.

Студенти повинні вміти:

- використовувати засоби інтеграції великих масивів даних, обирати та використовувати методи та алгоритми фільтрації, валідації та зберігання великих даних із застосовуванням ймовірносних розподілів,
- розробляти моделі та засоби обробки великих масивів даних із використанням математичних та статистичних методів,
- використовувати широкий спектр засобів аналізу великих масивів даних включно з використанням хмарних технологій, формуючи відповідний набір навичок та вмінь та відповідні висновки прийняття рішень при виконанні досліджень.

3. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Тема 1. Методи інтелектуального аналізу даних

Основні класи задач інтелектуального аналізу даних, задачі класифікації та кластеризації. Методи кластеризації. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

Методи класифікації. Постановка задачі класифікації та представлення результатів. Методи класифікації. Дерева рішень. Методи опорних векторів,

«найближчого сусіда», Байеса.

Задачі регресії. дисперсійний та регресійний аналіз. Аналіз часових рядів, кластерний аналіз, дерева рішень та генетичні алгоритми. Методи штучних нейронних мереж. Методи «k-найближчих сусідів».

Тема 2. Data Mining

Використання Data Mining в процесі прийняття рішень. Класифікація задач Data Mining (класифікації, кластеризації, асоціації, послідовності, прогнозування та візуалізація).

Задачі Data Mining. Прогнозування й візуалізація. Задачі прогнозування. Прогнозування і часові ряди. Тренд, сезонність і цикл. Види помилок та прогнозів. Візуалізація інструментів Data Mining. Методи візуалізації.

Основні методи Data Mining (регресія, нейронні мережі, генетичні алгоритми, дерева рішень). Рівні аналізу Data Mining.

Тема 3. Вирішення задач інтелектуального аналізу даних

Специфічні галузеві рішення Data Mining на практиці. Бізнес-завдання в умовах конкуренції. Застосування Data Mining при дослідженні економічних, соціальних та бізнес задач.

4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин							
	денна форма				Заочна форма			
	усього	у тому числі			усього	у тому числі		
		лекц.	прак	сам. роб.		лекц.	прак	сам. роб.
Тема 1. Методи інтелектуального аналізу даних	40	4	6	30	40	2	2	36
Тема 2. Data Mining	40	4	4	32	40	2	2	36
Тема 3. Вирішення задач інтелектуального аналізу даних	40	6	6	28	40	2	2	36

Усього годин	120	14	16	90	120	6	6	108
ПІДСУМКОВИЙ КОНТРОЛЬ - ЗАЛІК								

5. ПИТАННЯ ДО ПРАКТИЧНИХ ЗАНЯТЬ

Тема 1. Методи інтелектуального аналізу даних

Основні класи задач інтелектуального аналізу даних. Задачі класифікації та кластеризації.

Методи класифікації. Дерева рішень. Методи опорних векторів, «найближчого сусіда», Байеса.

Задачі регресії. Методи штучних нейронних мереж. Методи «k-найближчих сусідів».

Тема 2. Data Mining

Використання Data Mining в процесі прийняття рішень.

Задачі Data Mining. Прогнозування й візуалізація. Задачі прогнозування. Візуалізація інструментів Data Mining. Методи візуалізації.

Основні методи Data Mining. Рівні аналізу Data Mining.

Тема 3. Вирішення задач інтелектуального аналізу даних

Вирішення задач інтелектуального аналізу даних за допомогою програмного застосунку Weka.

Підготовка вихідних даних та вирішення задач класифікації та кластеризації за допомогою програмного застосунку Weka.

Підготовка вихідних даних та вирішення задач дерева рішень за допомогою програмного застосунку Weka

6. САМОСТІЙНА РОБОТА

До самостійної роботи студентів щодо вивчення дисципліни «Інтелектуальний аналіз даних» включаються:

1. Знайомство з науковою та навчальною літературою відповідно зазначених у програмі тем.
2. Опрацювання лекційного матеріалу.
3. Підготовка до практичних занять.
4. Консультації з викладачем протягом семестру.
5. Самостійне опрацювання окремих питань навчальної дисципліни.

6. Підготовка до підсумкового контролю.

ТЕМАТИКА ТА ПИТАННЯ ДО САМОСТІЙНОЇ ПІДГОТОВКИ ТА ІНДИВІДУАЛЬНИХ ЗАНЯТЬ

Тема 1. Методи інтелектуального аналізу даних

Основні класи задач інтелектуального аналізу даних, задачі класифікації та кластеризації. Методи кластеризації. Види кластерів. Міри близькості, засновані на відстанях. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

Методи класифікації. Постановка задачі класифікації та представлення результатів. Методи класифікації. Деревя рішень. Методи опорних векторів, «найближчого сусіда», Байеса.

Задачі регресії. дисперсійний та регресійний аналіз. Аналіз часових рядів, кластерний аналіз, дерева рішень та генетичні алгоритми. Методи штучних нейронних мереж. Методи «k-найближчих сусідів».

Тема 2. Data Mining

Використання Data Mining в процесі прийняття рішень. Класифікація задач Data Mining (класифікації, кластеризації, асоціації, послідовності, прогнозування та візуалізація).

Задачі Data Mining. Прогнозування й візуалізація. Задачі прогнозування. Прогнозування і часові ряди. Тренд, сезонність і цикл. Види помилок та прогнозів. Візуалізація інструментів Data Mining. Методи візуалізації.

Основні методи Data Mining (регресія, нейронні мережі, генетичні алгоритми, дерева рішень). Рівні аналізу Data Mining.

Тема 3. Вирішення задач інтелектуального аналізу даних

Специфічні галузеві рішення Data Mining на практиці. Бізнес-завдання в умовах конкуренції. Застосування Data Mining при дослідженні економічних, соціальних та бізнес задач.

7. ВИДИ ТА МЕТОДИ КОНТРОЛЮ

Робоча програма навчальної дисципліни «Big Data та хмарні технології» передбачає наступні види та методи контролю:

Види контролю	Складові оцінювання
поточний контроль , який здійснюється у ході: проведення практичних занять, виконання варіантів завдань, проведення консультацій та відпрацювань, підсумковий контроль.	50%
підсумковий контроль , який здійснюється у ході проведення іспиту.	50%

Методи діагностики знань (контролю)	фронтальне опитування, індивідуальне опитування, робота у групах, розв'язання задач і практичних завдань, іспит
--	---

8. ПИТАННЯ ДО ІСПИТУ

1. Дайте визначення поняттю «інтелектуальний аналіз даних».
2. Основні класи задач інтелектуального аналізу даних.
3. Приклади задач та даних. Основні типи даних. Принципи формалізації даних.
4. Задачі класифікації та кластеризації. Приклади та вихідні дані.
5. Поняття кластеризації. Основні цілі кластеризації.
6. Ієрархічна кластеризація.
7. Алгоритм кластеризації k-means. Приклади реалізації.
5. Основні принципи збору даних. Поняття якості даних. Найбільш критичні фактори якості даних.
6. Поняття, задачі і роль візуалізації даних. Візуалізація даних.
7. Задачі регресії. Приклади та умови використання.
8. Основні положення дисперсійного та регресійного аналізу.
9. Наведіть основні положення кластерного аналізу. Приклади реалізації.
10. Метод інтелектуального аналізу даних - дерева рішень. Приклади.
11. Поняття генетичного алгоритму. Структура генетичного алгоритму.
12. Використання Data Mining в процесі прийняття рішень. Наведіть приклади.
13. Класифікація задач Data Mining (класифікації, кластеризації, асоціації, послідовності, прогнозування та візуалізація).
14. Основні методи Data Mining (регресія, нейронні мережі, генетичні

алгоритми, дерева рішень).

15. Застосування Data Mining при дослідженні економічних, соціальних та бізнес задач.

16. Методи кластерного аналізу. Види кластерів. Міри близькості, засновані на відстанях.

17. Базові алгоритми кластеризації. Адаптивні методи кластеризації.

18. Методи класифікації. Постановка задачі класифікації та представлення результатів.

19. Методи класифікації. Дерева рішень.

20. Методи опорних векторів, «найближчого сусіда», Байеса.

21. Вирішення задач інтелектуального аналізу даних за допомогою програмного застосунку Weka.

22. Підготовка вихідних даних та вирішення задач класифікації за допомогою програмного застосунку Weka.

23. Підготовка вихідних даних та вирішення задач кластеризації за допомогою програмного застосунку Weka.

24. Підготовка вихідних даних та вирішення задач дерева рішень за допомогою програмного застосунку Weka.

25. Візуалізація рішень прикладних задач за допомогою програмного застосунку Weka.

9. КРИТЕРІЇ ПІДСУМКОВОЇ ОЦІНКИ ЗНАНЬ СТУДЕНТІВ (для іспиту / заліку)

Рівень знань оцінюється:

– «відмінно» / «зараховано» А - від 90 до 100 балів. Студент виявляє особливі творчі здібності, вміє самостійно знаходити та опрацьовувати необхідну інформацію, демонструє знання матеріалу, проводить узагальнення і висновки. Був присутній на лекціях та практичних заняттях, під час яких давав вичерпні, обґрунтовані, теоретично і практично правильні відповіді, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

– «добре» / «зараховано» В - від 82 до 89 балів. Студент володіє знаннями матеріалу, але допускає незначні помилки у формуванні термінів, категорій, проте за допомогою викладача швидко орієнтується і знаходить правильні відповіді. Був присутній на лекціях та практичних заняттях, має конспект з виконаними

завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

– «добре» / «зараховано» C - від 74 до 81 балів. Студент відтворює значну частину теоретичного матеріалу, виявляє знання і розуміння основних положень, з допомогою викладача може аналізувати навчальний матеріал, але дає недостатньо обґрунтовані, невичерпні відповіді, допускає помилки. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи та активність у науково-дослідній роботі;

– «задовільно» / «зараховано» D - від 64 до 73 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на середньому рівні, допускає помилки, серед яких є значна кількість суттєвих. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи;

– «задовільно» / «зараховано» E - від 60 до 63 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на рівні, вищому за початковий, значну частину його відтворює на репродуктивному рівні, на всі запитання дає необґрунтовані, невичерпні відповіді, допускає помилки, має неповний конспект з завданнями до самостійної роботи.

– «незадовільно з можливістю повторного складання» / «не зараховано» FX – від 35 до 59 балів. Студент володіє матеріалом на рівні окремих фрагментів, що становлять незначну частину навчального матеріалу.

– «незадовільно з обов'язковим повторним вивченням дисципліни» / «не зараховано» F – від 0 до 34 балів. Студент не володіє навчальним матеріалом.

Таблиця відповідності результатів контролю знань за різними шкалами

100-бальною шкалою	Шкала за ECTS	За національною шкалою	
		іспит	залік
90-100 (10-12)	A	Відмінно	зараховано
82-89 (8-9)	B	Добре	
74-81 (6-7)	C		
64-73 (5)	D	Задовільно	
60-63 (4)	E		
35-59 (3)	FX	незадовільно	не зараховано
1-34 (2)	F		

10. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Recommendation ITU-T Y.3500. Information technology. Cloud technology. Overview and vocabulary. 08/2014. URL: <https://www.itu.int/rec/T-REC-Y.3500-201408-I>.

Recommendation ITU-T Y.3501. Information technology. Cloud technology. Framework and high-level requirements. 06/2016. URL: <https://www.itu.int/rec/T-REC-Y.3500-201408-I>.

2. ETSI GS NFV 002 V1.1.1 Network Functions Virtualisation (NFV). Architectural Framework. 2013-10. URL: https://www.etsi.org/deliver/etsi_gs/nfv/001_099/002/01.01.01_60/gs_nfv002_v010101p.pdf

3. Литвин В. В., Пасічник В. В., Нікольський Ю. В. Аналіз даних та знань: навчальний посібник. Львів: Магнолія, 2021. 276 с.

4. Іванчук Я.В., Месюра В.І., Яровий А.А., Манжілевський О.Д. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних. Вінниця: ВНТУ, 2021. 69 с.

5. Олещенко Л.М. Технології оброблення великих даних. Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с. [Електронний ресурс] (istu.edu.ua).

6. 2. Тарн В.І., Гордієнко Ю. Г., Стіренко С.Г. / Технології BIG DATA Практикум // Київ: КПІ ім. Ігоря Сікорського, 2022. – 56 с.

7. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC, CRC Press, 2019.

Допоміжна

8. Зінченко О.В., Іщеряков С.М., Прокопов С.В., Сєрих С.О., Василенко В.В. Хмарні технології. – Навчальний посібник. – К: ФОП Гуляєва В.М., 2020.

9. Григор'єва Т.І., Соловська І.М. Сучасні проблеми теорії ймовірностей та математичної статистики: методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Григор'єва Т.І., Соловська І.М. Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 47 с.

10. 12. Григор'єва Т.І., Соловська І.М. Аналіз даних та комп'ютерна статистика: методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Григор'єва Т.І., Соловська І.М. Кафедри комп'ютерних

наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 50 с.

11. 13. Соловська І.М., Григор'єва Т.І. Великі масиви даних: : методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Соловська І.М., Григор'єва Т.І., Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 15 с.

12. 14. Соловська І.М., Григор'єва Т.І. Статистичне моделювання та програмування: : методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Соловська І.М., Григор'єва Т.І. Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 15 с.

13. Олексюк В. Основи хмарних технологій / В. Олексюк. – Тернопіль: Тернопільський обласний комунальний інститут післядипломної педагогічної освіти. – 156 с.

14.

15. Strelkovskaya I., Solovskaya I., Makoganiuk A. Different extrapolation methods in Problems of Forecasting. Advances in Information and Communication Technology and Systems. Lecture Notes in Networks and Systems. 2020. Vol. 152. Springer. P. 217-228. https://doi.org/10.1007/978-3-030-58359-0_12

16. Стрелковська І.В., Соловська І.М., Стрелковська Ю.О. Застосування дійсних та комплексних сплайнів в задачах інфокомунікацій. Проблеми телекомунікацій. – 2021. – № 1(28). – С. 3-19.

17. Стрелковська І.В. Використання сплайн-апроксимації та сплайн-екстраполяції в задачах інфокомунікацій / І.В. Стрелковська, І.М. Соловська // VII Всеукраїнська науково-практична конференція «Інформаційне суспільство: проблеми та перспективи»: матеріали конф., Одеса, 20 травня 2022 р.: тези доп. – Одеса: НУ ОЮА, 2022. – С. 56-61. <http://dspace.onua.edu.ua/handle/11300/18196>

18. Стрелковська І. В., Соловська І. М., Стрелковська Ю. О., Використання методів сплайн-функцій в телекомунікаційних та інформаційних технологіях Міжнародна конференція «Передові технології в інформаційно-комунікаційній інженерії»: матеріали конф., 17-20 липня 2023р.: тези доп. – Одеса: МГУ, 2023. – С. 70-79.

Інформаційні ресурси

19. Електронний каталог Національної парламентської бібліотеки України [політемат. база даних містить відом. про вітчизн. та зарубіж. кн., брош., що надходять у фонд НПБ України]. Київ: Нац. парлам. б-ка України, 2002-2023. URL: catalogue.nplu.org .

20. Національна бібліотека України імені В. І. Вернадського Київ: НБУВ, 2013-2023. URL: www.nbuv.gov.ua

21. Відкритий програмний застосунок для аналізу даних та машинного навчання. URL: www.weka.io

ЗМІСТ

1. ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	3
2. ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ	4
3. ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	4
4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ.....	5
5. ПИТАННЯ ДО ПРАКТИЧНИХ ЗАНЯТЬ	6
6. САМОСТІЙНА РОБОТА.....	6
7. ВИДИ ТА МЕТОДИ КОНТРОЛЮ	7
8. ПИТАННЯ ДО ІСПИТУ.....	9
9. КРИТЕРІЇ ПІДСУМКОВОЇ ОЦІНКИ ЗНАНЬ СТУДЕНТІВ.....	10
10. РЕКОМЕНДОВАНА ЛІТЕРАТУРА.....	11
ЗМІСТ.....	14

Навчальне видання

Соловська Ірина Миколаївна

Григор'єва Тетяна Ігорівна

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ

Методичні рекомендації
для самостійної роботи здобувачів,
які навчаються за другим (магістерським) рівнем вищої освіти
зі спеціальності 112 «Статистика»

Українською мовою