

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТАВРІЙСЬКИЙ ДЕРЖАВНИЙ АГРОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ДМИТРА МОТОРНОГО**

Кафедра «Комп'ютерні науки»

ЗАТВЕРДЖУЮ

В.о.зав. кафедрою КН

доцент Шу Сергій ШАРОВ

« 31 » серпня 2021 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Технології інтелектуального аналізу даних»

для здобувачів ступеня вищої освіти «Магістр»

зі спеціальності 122 "Комп'ютерні науки"

за ОПІ Комп'ютерні науки

(на основі ОС «Бакалавр»)

Факультет енергетики і комп'ютерних технологій

2021-2022 н.р.


Робоча програма навчальної дисципліни «Технології аналізу даних» для здобувачів ступеня вищої освіти «Магістр» зі спеціальності 122 «Комп'ютерні науки» (на основі ОС «Бакалавр»), факультет енергетики і комп'ютерних технологій. - Мелітополь, ТДАТУ, 2021 - 13 с.

Розробники: д.т.н., професор Малкіна В.М.

Робоча програма затверджена на засіданні кафедри «Комп'ютерні науки».

Протокол № 1 від 27 серпня 2021 року

В.о. завідувача кафедри КН

 Сергій ШАРОВ

« 27 » серпня 20 21 року

Схвалено методичною комісією факультет енергетики і комп'ютерних технологій

для здобувачів ступеня вищої освіти «Магістр» за спеціальністю 122 «Комп'ютерні науки» за ОПП Комп'ютерні науки (на основі ОС «Бакалавр»)

Протокол № 1 від « 31 » 08 2021 року

Голова  Діна НЕСТЕРЧУК

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	
Кількість кредитів 4	Галузь знань 12 "Інформаційні технології"	Обов'язкова	
Загальна кількість годин – 120 годин	Спеціальність 122 "Комп'ютерні науки"	Курс	Семестр
Змістових модулів – 2		1М-й	2-й
Тижневе навантаження: - аудиторних занять 4 год. - самостійна робота студента 4,5 год.	Ступінь вищої освіти «Магістр»	Вид занять	Кількість годин
		Лекції	26 год.
		Лабораторні заняття	26 год.
		Практичні заняття	-
		Семінарські заняття	-
		Самостійна робота	68 год.
		Форма контролю: <u>диф. залік</u>	

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни – сформувані фундаментальні теоретичні знання та практичні навички щодо суті та застосування засобів інтелектуального аналізу та організації даних в інформаційних системах; набути навички практичної роботи із програмними засобами для інтелектуального аналізу даних; навчити ефективного використання сучасних технологій та програмних засобів інтелектуального аналізу даних та обробки інформації.

Завдання навчальної дисципліни - надати інформацію про основні технології аналізу даних; основні концепції сховища даних; методи очистки та трансформації даних; методи побудови багатовимірних моделей даних; способи візуалізації даних; статистичні аспекти моделювання; методи побудови моделей та аналізу залежностей у великих масивах даних; сучасні програмні засоби та технології для інтелектуального аналізу даних; навчити студентів обґрунтовувати і аналізувати вибір конкретного типу моделі та методу аналізу даних при розв'язанні відповідних практичних задач; використовувати сучасні програмні засоби і інформаційні технології для інтелектуального аналізу даних; обирати методи та планувати експериментальні та теоретичні наукові дослідження; використовувати статистичні методи обробки та аналізу результатів досліджень; аналізувати результати побудови та використання систем інтелектуального аналізу даних при розв'язанні прикладних задач.

Результати навчання (з урахуванням soft skills):.

Інтегральна компетентність

Здатність розв'язувати задачі в галузі професійної та/або дослідницько-інноваційної діяльності у сфері комп'ютерних наук, що передбачає як вільне володіння наявними знаннями, так і спроможність їх застосування у професійній практиці.

Загальні компетенції

Здатність до абстрактного мислення, аналізу та синтезу.

Здатність застосовувати знання у практичних ситуаціях.

Здатність спілкуватися державною мовою як усно, так і письмово.

Здатність вчитися й оволодівати сучасними знаннями.

Здатність бути критичним і самокритичним.

Здатність генерувати нові ідеї (креативність).

Здатність працювати в команді.

Здатність оцінювати та забезпечувати якість виконуваних робіт.

Спеціальні (фахові, предметні) компетентності.

Розуміння теоретичних засад комп'ютерних наук для об'єктивного оцінювання можливостей використання обчислювальної техніки в певних процесах людської діяльності і визначення перспективних інформаційних технологій.

Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття рішень.

Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук: алгоритми розв'язання обчислювальних та логічних задач, алгоритми паралельних та розподілених обчислень, алгоритми аналітичної обробки й інтелектуального аналізу великих даних з оцінкою їх ефективності та складності.

Професійне володіння сучасними комп'ютерними та інформаційними технологіями.

Soft skills:

- **комунікативні навички:** письмове, вербальне й невербальне спілкування; уміння грамотно спілкуватися по e-mail; вести суперечки і відстоювати свою позицію, спілкування в конфліктній ситуації; навички створення, керування й побудови відносин у команді;

- **уміння виступати привселюдно:** навички, необхідні для виступів на публіці; проводити презентації;

- **керування часом:** уміння справлятися із завданнями вчасно;

- **гнучкість і адаптивність:** гнучкість, адаптивність і здатність мінятися; уміння аналізувати ситуацію, орієнтування на вирішення проблем;

- **лідерські якості:** уміння спокійно працювати в напруженому середовищі; уміння ухвалювати рішення; уміння встановлювати мету, планувати;

- **особисті якості:** креативне й критичне мислення; етичність, чесність, терпіння, повага до колег

3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1.

Тема 1. Основні прийоми роботи в R. [1, с.15-22]

1. Запуск и початок роботи
2. Робота зі «скриптами»
3. Спеціальні змінні в R
5. Створення числових послідовностей в R
6. Класи даних (об'єктів) в R
7. Вектор
8. Матриці
9. Масиви
10. Список
11. Фактор
12. Фрейми (таблиці)
13. Читання даних з текстового файлу
14. Збереження текстових даних
15. Генерування випадкових послідовностей

Тема 2. Описова статистика в R. [1, с.35-42]

1. Що таке описова статистика.
2. Оцінка вибірових характеристик з використання спеціальних функцій.
3. Використання функцій summary () и додаткових пакетів.
4. Заповнення пропущених значень в таблицях даних.
5. Аналіз викидів.

Тема 3. Підбір параметрів і перевірка закону розподілу [1, с.45-54]

1. Поняття закон розподілу і параметри закону розподілу
2. Основні ймовірнісні розподілі
 - 2.1 Нормальний закон розподілу
 - 2.2 Показовий закон розподілу
- 2.4 Рівномірний закон розподілу
- 4 Перевірка на відповідність нормальному закону розподілу
 - 4.1 Графічні способи
 - 4.2 Формальні тести
- 5 Оцінка відповідності нормальному розподілу за допомогою критерію

Колмогорова та χ^2 - критерію Пірсона

Тема 4. Перевірка статистичних гіпотез в R. [1, с.55-65]

1. Параметричні і непараметричні статистичні гіпотези
2. Нульова й альтернативна гіпотези
3. Область прийняття гіпотези. Критична область. Критична точка
4. Алгоритм перевірки нульової гіпотези
5. Помилки першого та другого роду. Потужність критерію
6. Гіпотеза про рівність середніх двох генеральних сукупностей
7. Гіпотеза про однорідність дисперсій

Змістовий модуль 2

Тема 5. Дисперсійний аналіз [1, с.35-37]

1. Ідея методу дисперсійного аналізу
2. Однофакторний дисперсійний аналіз
3. Однофакторний дисперсійний аналіз в R
4. Двофакторний дисперсійний аналіз

Тема 6. Data Mining: Візуалізація даних [1, с.32-37]

1. Основні поняття
2. Аналіз даних за допомогою графіків
3. Графічні функції високого рівня
 - 3.1 R - Кругові діаграми
 - 3.2 Стовпчикові діаграми
 - 3.3 Гістограми
 - 3.4 Лінійні графіки

Тема 7. Data Mining: Регресійний аналіз [1, с.64-78]

1. Основні поняття регресійного аналізу даних
 - 1.1 Первинний аналіз залежності
 - 1.2 Коефіцієнт кореляції Пірсона
2. Регресійний аналіз
 - 2.1 Обчислення коефіцієнтів рівняння регресії методом найменших квадратів.
 - 2.2 Гіпотези в регресійному аналізі
 - 2.3 Оцінка значимості регресійної моделі: t-критерій і F-критерій

Тема 8. Data Mining: Кластерний аналіз в R

1. Призначення кластерного аналізу
2. Короткі теоретичні відомості
 - 2.1 Особливості вихідних даних
 - 2.2 Відстань між класами
 - 2.3 Якість розбиття на класи
 - 2.4 Заходи відстані між об'єктами
3. Методи кластерного аналізу
 - 3.1 Різновиди агломеративного (об'єднують) методів
 - 3.2 Поняття про дендрограму
4. Реалізація методів кластеризації в R
 - 4.1 Реалізація методу k-means в R
 - 4.2 Реалізація ієрархічного алгоритму кластеризації в R

4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Номер тижня	Вид занять	Тема заняття або завдання на самостійну роботу	Кількість				
			годин				балів
			лк	лаб	сем. (пр.)	СРС	
Змістовий модуль 1							
1	Лекція 1	Основні прийоми роботи в R	2	-	-	-	-
	Лабораторна робота 1	Знайомство з мовою програмування R. Імпорт та збереження даних	-	2	-	-	3,5
	Самостійна робота 1	Технології KDD I Data Mining. Аналітичні платформи	-	-	-	6	3,5
2	Лекція 2	Основні прийоми роботи в R (продовження) Описова статистика	2	-	-	-	-
	Лабораторна робота 2	Знайомство з мовою програмування R. Імпорт та збереження даних (продовження)	-	2	-	-	3,5
	Самостійна робота 2	Технології KDD I Data Mining. Аналітичні платформи	-	-	-	5	3,5
3	Лекція 3	Описова статистика	2	-	-	-	-
	Лабораторна робота 3	Описова статистика і підгонка розподілів	-	2	-	-	3,5
	Самостійна робота 3	Очистка даних в ELT. Перетворення даних в ELT	-	-	-	6	3,5
4	Лекція 4	Підбір параметрів і перевірка закону розподілу	2	-	-	-	-
	Лабораторна	Визначення параметрів	-	2	-	-	3,5

	робота 4	рів і перевірка гіпотези про закон розподілу даних					
	Самостійна робота 4	Очистка даних в ELT. Перетворення даних в ELT	-	-	-	5	3,5
5	Лекція 5	Перевірка статистичних гіпотез	2	-	-	-	-
	Лабораторна робота 5	Перевірка статистичних гіпотез	-	2	-	-	3
	Самостійна робота 5	Злиття даних. Квантування. Нормалізація та кодування даних.	-	-	-	4	3
6	Лекція 6	Дисперсійний аналіз	2				
	Лабораторна робота 6	Дисперсійний аналіз в R	-	2	-	-	3
	Самостійна робота 6	Обробка дублікатів та протиріч. Виявлення аномальних значень. Відновлення пропущених значень.	-	-	-	4	3
7,8	ПМК 1	Підсумковий контроль за змістовий модуль 1					10
	Самостійна робота 7	Підготовка до ПМК 1				4	
Всього за змістовий модуль 1 - 58 год.			12	12	-	34	50
Змістовий модуль 2							
9	Лекція 7	Візуалізація даних в R	2	-	-	-	-
	Лабораторна робота 7	Візуалізація даних	-	2	-	-	3
	Самостійна робота 8	Зменшення кількості признаков	-	-	-	5	3
10	Лекція 8	Регресійний аналіз в R	2	-	-	-	-
	Лабораторна робота 8	Регресійний аналіз		2			3
	Самостійна робота 9	Зменшення кількості признаков.	-	-	-	4	3
11	Лекція 9	Регресійний аналіз в R (продовження)	2	-	-	-	-
	Лабораторна робота 9	Регресійний аналіз (продовження)	-	2	-	-	3
	Самостійна робота 10	Скорочення кількості значень при знаків та записів.	-	-	-	4	3

12	Лекція 10	Data Mining: кластеризація	2				
	Лабораторна робота 10	Задача кластеризації. Ієрархічні алгоритми (алгоритм) найближ- чого сусіда.	-	2	-	-	3
	Самостійна робота 11	Технології Data Mining.	-	-	-	4	3
13	Лекція 11	Data Mining: кластеризація (продо- вження)	2	-	-	-	-
	Лабораторна робота 11	Задача кластеризації. Метод k-means	-	2	-	-	3
	Самостійна робота 12	Скорочення кількості значень при знаків та записів.	-	-	-	4	3
14	Лекція 12	Data Mining: класифікації	2				
	Лабораторна робота 12	Класифікація. Алго- ритм CART		2			3
	Самостійна робота 13	Скорочення кількості значень при знаків та записів.				4	3
15	Лекція 13	Data Mining: класифікації	2				
	Лабораторна робота 13	Класифікація. Алго- ритм CART (продов- ження)		2			2
	Самостійна робота 14	Скорочення кількості значень при знаків та записів.				4	2
16,17	ПМК 2	Підсумковий контроль за змістовий модуль 2					10
	Самостійна робота 15	Підготовка до ПМК 2				5	
Всього за змістовий модуль 2 - 62 год.			14	14	-	34	50
Всього з навчальної дисципліни - 120 год.							100

5 ПЕРЕЛІК ПИТАНЬ, ЩО ВІНОСЯТЬСЯ НА ПІДСУМКОВИЙ КОНТРОЛЬ

ПМК-1

1. Принципи аналізу даних.
2. Структуровані дані.
3. Підготовка даних до аналізу.
4. Технології KDD і Data Mining.
5. Аналітичні платформи.
6. Алгоритми Data Mining
7. Задача консолідації.
8. Сховища даних.
9. Основні концепції сховища даних.
10. Багатовимірні сховища даних.
11. Реляційні сховища даних.
12. Гібридні сховища даних.
13. Віртуальні сховища даних.
14. Нечіткі зрізи.
15. ETL. Видобування даних в ETL.
16. Очистка даних в ETL.
17. Перетворення даних в ETL.
18. Завантаження даних в ETL.
19. Завантаження даних із локальних джерел.
20. Збагачення даних.
21. Вступ до трансформації даних. Трансформація упорядкованих даних.
22. Групування даних.
23. Злиття даних.
24. Квантування.
25. Нормалізація та кодування даних.
26. Вступ до візуалізації. Візуалізатори загального значення.
27. OLAP – аналіз.
28. Візуалізатори для оцінювання якості моделей.
29. Візуалізатори, що застосовуються для інтерпретації результатів аналізу.
30. Описова статистика(графічна візуалізація)
31. Оцінка параметрів і перевірка статистичних гіпотез.
32. Непараметричні тести

ПМК 2

1. Оцінка якості даних.
2. Технології та методи оцінки якості даних.
3. Очистка та первинна обробка даних.

4. Фільтрація даних.
5. Обробка дублікатів та протиріч.
6. Виявлення аномальних значень.
7. Відновлення пропущених значень.
8. Скорочення розмірності.
9. Зменшення кількості признаков.
10. Скорочення кількості значень признаков та записів.
11. Класифікація і регресія. Основні поняття.
12. Проста лінійна регресія.
13. Оцінка відповідності простої лінійної регресії реальним даним.
14. Проста регресійна модель.
15. Множино – лінійна регресія.
16. Модель множино – лінійної регресії.
17. Регресія з категоріальними вхідними змінними.
18. Методи відбору змінних в регресійні моделі.
19. Обмеження застосування регресійних моделей.
20. Основи логістичної регресії.
21. Інтерпретація моделі логістичної регресії.
22. Множино - логістична регресія.

6 РЕКОМЕНДОВАНА ЛІТЕРАТУРА

ОСНОВНА

1. Ситник В.Ф., Краснюк М.Т. Интеллектуальный анализ данных (дейтамайнінг): Навч. посібник. - К.: КНЕУ, 2007. – 376 с.
2. Барсегян А.А. и др. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.
3. Дюк В., Самойленко А. Data Mining: учебный курс. - СПб: Питер, 2001. -368 с.
4. Корнеев В.В., Гареев А.Ф., Васютин С.В., Райх В.В. Базы данных. Интеллектуальная обработка информации. – М.: «Нолидж», 2000. – 352 с., ил
5. Андрейчиков А.В., Андрейчикова О.Н. Интеллектуальные информационные системы: Учебник. – М.: Финансы и статистика, 2004. – 424 с.
6. Дубровин В.И., Субботин С.А., Богуслаев А.В., Яценко В.К. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей: Монография. – Запорожье: ОАО "Мотор-Сич", 2003. – 279 с.
7. Круглов В.В., Борисов В.В. Искусственные нейронные сети: теория и практика.-М.: Горячая линия-Телеком, 2001. – 382 с.
8. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского. – М.: Горячая линия-Телеком, 2004. – 452 с.

ДОПОМІЖНА

9. Бондарев В.Н., Аде Ф.Г. Искусственный интеллект. – Севастополь: СевНТУ, 2002. – 615 с.
10. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности.-Харьков: Основа, 1997. – 112 с.
11. Дубровін В.І., Субботін С.О. Методи оптимізації та їх застосування в задачах навчання нейронних мереж: Навчальний посібник. – Запоріжжя: ЗНТУ, 2003. – 136 с.
12. Дьяконов В. МАТЛАБ 6: учебный курс. – СПб.: Питер, 2001. – 592 с.
13. Зайченко Ю.П. Основы проектирования интеллектуальных систем. Навчальний посібник. – К.: Слово, 2004. – 352 с.
14. Кричевский М.Л. Интеллектуальные методы в менеджменте. – СПб.: Питер, 2005. – 304 с.
15. Люгер Дж.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем / Пер. с англ. – М.: Вильямс, 2005. – 864 с.
16. Митюшкин Ю.И., Мокин Б.И., Ротштейн А.П. Soft Computing: идентификация закономерностей нечеткими базами знаний. – Винница: УНИВЕРСУМ-Винница, 2002. – 145 с.

17. Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2004. – 344 с.
18. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.: Пер с англ.. – М.: Вильямс. – 2006. – 1408 с.
19. Рідкокаша А.А., Голдер К.К. Основи систем штучного інтелекту. Навчальний посібник. – Черкаси: "ВІДЛУННЯ-ПЛЮС", 2002. – 240 с.
20. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Винница: УНІВЕРСУМ-Вінниця, 1999. – 320 с.
21. Усков А.А., Кузьмин А.В. Интеллектуальные технологии управления. Искусственные нейронные сети и нечеткая логика. – М.: Горячая линия-Телеком, 2004. – 143 с.
22. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. – М.: Финансы и статистика, 2004. – 320 с.

7 ІНФОРМАЦІЙНІ РЕСУРСИ

- 1 Наукова бібліотека ТДАТУ <http://www.tsatu.edu.ua/biblioteka/>
- 2 Сайт кафедри КН <http://www.tsatu.edu.ua/kn/course/technologii-analizy-danyh-ta-jih-orhanizacija/>
- 3 Навчально-інформаційний портал ТДАТУ <http://nip.tsatu.edu.ua/course/view.php?id=1501#section-11>