

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

Кафедра «Комп'ютерні науки»

ПОГОДЖЕНО

Гарант ОПП «Комп'ютерні науки»
доц. Холодняк Ю.В.
«_30_» __серпня____ 2023 р.

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри КН
доц. _____ Сергій ШАРОВ
«_31_» __серпня____ 2023 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Інтелектуальний аналіз даних»

для здобувачів ступеня вищої освіти «Бакалавр»
зі спеціальності 122 «Комп'ютерні науки»
за ОПП Комп'ютерні науки
(на основі повної загальної середньої освіти та ОКР «молодший спеціаліст»)

Факультет енергетики і комп'ютерних технологій

2023– 2024 н.р.


Робоча програма навчальної дисципліни «Інтелектуальний аналіз даних» для здобувачів ступеня вищої освіти «Бакалавр» зі спеціальності 122 «Комп'ютерні науки» за ОПП Комп'ютерні науки (на основі повної загальної середньої освіти), факультет енергетики і комп'ютерних технологій. Запоріжжя, ТДАТУ. 12 с.

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

Робочу програму затверджено на засіданні кафедри «Комп'ютерні науки»

Протокол № від « 30 » серпня 2023 року

В.о. зав. кафедри КН

 Сергій ШАРОВ

« 30 » серпня 2023 року

Схвалено методичною комісією факультет енергетики і комп'ютерних технологій зі спеціальності 122 «Комп'ютерні науки» за ОПП Комп'ютерні науки (на основі повної загальної середньої освіти)

Протокол № 1 від 31 серпня 2023 року

Голова, доц.  Олександр БОВК

«31» серпня 2023 року

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		<u>денна форма навчання</u> (денна або заочна)	
Кількість кредитів 4	Галузь знань 12 "Інформатика та обчислювальна техніка"	Обов'язкова	
Загальна кількість годин –120 годин	Спеціальність 122 "Комп'ютерні науки"	Курс	Семестр
Змістових модулів – 2		4-й, 2С	7-й, 3-й
Тижневе навантаження: - аудиторних занять 4 год. - самостійна робота студента 9 год.	Ступінь вищої освіти <u>«Бакалавр»</u>	Вид занять	Кількість годин
		Лекції	10 год.
		Лабораторні заняття	-
		Практичні заняття	20 год.
		Семінарські заняття	-
		Самостійна робота	90 год.
		Форма контролю: <u>диференційований залік</u>	

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета дисципліни “ Інтелектуальний аналіз даних ” (ІАД) - вивчення методів сучасної обробки даних – інтелектуального аналізу даних (Data Mining), пошуку у необроблених масивах даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів та різних інструментальних засобів, що використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з розв'язання різних економічних задач засобами Data Mining та розробки інтелектуальних систем. Розглядаються такі загальні поняття: статистичні пакети; нейронні мережі; еволюційні методи і алгоритми пошуку логічних закономірностей.

Завдання курсу - навчитися застосовувати сучасні інформаційні технології і програмні системи для обробки експериментальних і статистичних даних.

Вивчення навчальної дисципліни передбачає формування наступних компетентностей студента:

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

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК5. Здатність спілкуватися іноземною мовою.

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК12. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ЗК13. Здатність діяти на основі етичних міркувань.

ЗК14. Здатність реалізувати свої права і обов'язки як члена суспільства, усвідомлювати цінності громадянського (вільного демократичного) суспільства та не-

обхідність його сталого розвитку, верховенства права, прав і свобод людини і громадянина в Україні.

Спеціальні (фахові, предметні) компетентності.

ФК1. Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування.

ФК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

ФК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

ФК17. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики під час виконання функціональних завдань та обов'язків.

Програмні результати навчання:

РН3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.

РН8. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проектування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

РН12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.

РН18. Створювати концептуальні імітаційні моделі складних природних і економічних систем на основі їх дослідження та реалізувати їх за допомогою мов програмування і моделювання.

Soft skills:

- **комунікативні навички:** письмове, вербальне й невербальне спілкування; вміння 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

Міждисциплінарні зв'язки з урахуванням структурно-логічної схеми ОПП «Комп'ютерні науки».

Перелік навчальних дисциплін, знання з яких потрібні для вивчення освітньої компоненти ІАД: Дискретна математика, Теорія ймовірностей та математична статистика, Методи і системи штучного інтелекту.

Перелік навчальних дисциплін, вивчення яких у подальшому базується на матеріалі освітньої компоненти ІАД: Теорія прийняття рішень

3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Аналіз даних на основі регресійних методів

Тема 1 Технології аналізу даних [1, 3, 6]

- 1 Вступ до аналізу даних
- 2 Принципи аналізу даних
- 3 Методика отримання знань
- 4 Поняття алгоритмів data mining

Тема 2 Інтелектуальний аналіз даних на основі регресії [1; 2, 3, 6, 7]

- 1 Регресія, основні поняття
- 2 Проста лінійна регресія

Тема 3 Регресійний аналіз [1, 3, 4, 5, 6, 7]

- 1 Проста лінійна регресійна модель
- 2 Гіпотези в регресійному аналізі
- 3 Оцінка значимості регресійної моделі: t-критерій і f-критерій
- 4 Множинна лінійна регресія
- 5 Побудова та аналіз моделі множинної лінійної регресії
- 6 Оцінка значимості множинної регресійної моделі
- 7 Обмеження при застосуванні регресійної моделі

Тема 4 Аналіз і прогнозування часових рядів [1, 2, 3, 5, 6]

- 1 Поняття прогнозування
- 2 Поняття часового ряду
- 3 Моделі часових рядів
- 4 Автокореляція при дослідженні часових рядів
- 5 Моделі прогнозування часових рядів
- 6 Прогнозування на основі методу декомпозиції часового ряду

Змістовий модуль 2 Аналіз даних методами машинного навчання

Тема 5 Класифікація. машинне навчання. Дерева рішень [1, 2, 3]

- 1 Основні поняття дерев рішень
- 2 Алгоритми побудови дерев рішень

Тема 6 Кластеризація [1, 2]

- 1 Основні поняття
- 2 Методи кластерного аналізу
- 3 Алгоритм кластеризації k-means
- 4 Приклад роботи алгоритму k-means
- 5 Оцінка якості багатовимірної класифікації

Тема 7 Задача асоціації [1, 2, 3]

Асоціативні правила. Пошук асоціативних правил. Генерація асоціативних правил. Послідовні шаблони. Алгоритм Apriori. Алгоритм FPG.

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. Аналіз даних на основі регресійних методів							
1	Лекція 1	Технології аналізу даних. Основні поняття статистичного аналізу даних	1	-	-	-	-
	Практична робота 1	Первинна обробка статистичних даних. Описова статистика	-	2	-	-	6
	Самостійна робота 1	Можливості системи Statistica. Перевірка статистичних гіпотез Приклади застосування Data Mining.	-	-	-	9	2
2	Лекція 2	Класифікація і регресія. Регресійний аналіз	2	-	-	-	-
	Практична робота 2	Проста лінійна регресія	-	2	-	-	6
	Самостійна робота 2	Архітектура, класифікація та стандарти програмних систем дейтамайнінгу	-	-	-	9	2
3	Лекція 3	Аналіз і прогнозування часових рядів. Класифікація, машинне навчання, дерево рішень.	2	-	-	-	-
	Практична робота 3	Нелінійна регресія.	-	2	-	-	6
	Самостійна робота 3	Характеристика сучасного стану ринку систем дейтамайнінгу та напрями розвитку	-	-	-	9	2
	Практична робота 4	Множинна регресія	-	2	-	-	6

	Самостійна робота 4	Характеристики процесів і активностей та методів дейтамайнінгу	-	-	-	9	2
	Практична робота 5	Аналіз часових рядів	-	2	-	-	6
	Самостійна робота 5	Розроблення програмного забезпечення штучних нейромереж	-	-	-	9	2
4,5	ПМК 1	Підсумковий контроль за змістовий модуль 1					10
Всього за змістовий модуль 1 - 60 год.			5	10	-	45	50
Змістовий модуль 2. Аналіз даних методами машинного навчання							
6	Лекція 4	Класифікація. Алгоритм покриття. Data Mining: кластеризація.	3	-	-	-	-
	Практична робота 4	Класифікація за допомогою дерева рішень. Алгоритм CART	-	2	-	-	5
	Самостійна робота 4	Програмне забезпечення нейромереж	-	-	-	9	3
7	Лекція 5	Задачі асоціації. Консолідація даних	2				
	Практична робота 5	Класифікація. Алгоритм покриття.		2			5
	Самостійна робота 5	Підготовка до практичної роботи				9	3
	Практична робота 8	Дерево рішень. Алгоритм ID3	-	2	-	-	5
	Самостійна робота 7	Програмне забезпечення нейромереж	-	-	-	9	3
	Практична робота 9	Задача кластеризації	-	2	-	-	5
	Самостійна робота 8	Загальна схема генетичних алгоритмів	-	-	-	9	3
8	Практична робота 10	Пошук асоціативних правил. Алгоритм Apriori		2			5
	Самостійна робота 10	Підготовка до практичної роботи				9	3
9,10	ПМК 2	Підсумковий контроль за змістовий модуль 2					10
Всього за змістовий модуль 2 - 60 год.			5	10	-	45	50

<i>Диференційований залік</i>	-
<i>Всього з навчальної дисципліни – 60+60 = 120 год.</i>	<i>100</i>

5. ПЕРЕЛІК ПИТАНЬ, ЩО ВІНОСЯТЬСЯ НА ПІДСУМКОВИЙ МОДУЛЬНИЙ КОНТРОЛЬ №1

1. Сутність аналітичних технологій.
2. Поняття інтелектуального аналізу даних.
3. Етапи та методи пошуку нових знань.
4. Основні моделі інтелектуальних обчислень.
5. Засоби програмної підтримки інтелектуального аналізу даних.
6. Новітні напрями застосування Data Mining.
7. Концепція сховищ даних.
8. Технології побудови сховищ даних.
9. Вітрини та кіоски даних .
10. OLAP-технологія.
11. Основні архітектури OLAP-систем.
12. OLAP-системи та інтернет-технології.
13. Поняття та можливості нейрокомп'ютерних технологій.
14. Архітектура нейронних мереж.
15. Нейронні мережі Хопфілда та Кохонена.
16. Програмні засоби реалізації нейромережових технологій.
17. Сучасна практика та перспективні напрями застосування нейротехнологій.

ПЕРЕЛІК ПИТАНЬ, ЩО ВІНОСЯТЬСЯ НА ПІДСУМКОВИЙ МОДУЛЬНИЙ КОНТРОЛЬ №2

1. Основні поняття теорії асоціативних правил.
2. Програмні засоби пошуку асоціативних правил.
3. Практичний аспект застосування технології асоціативних правил.
4. Дерева рішень — загальні принципи технології.
5. Комп'ютерні системи та напрямки застосування дерев рішень.
6. Концептуальні засади еволюційної теорії. Основні положення теорії генетичних алгоритмів.
7. Моделі генетичних алгоритмів.
8. Програмне забезпечення та сфери застосування генетичних алгоритмів.
9. Концепція нечітких обчислень.
10. Нечітка логіка в системах Data Mining.
11. Програмне забезпечення нечітких методів.
12. Сучасна практика застосування нечітких методів.

6 МЕТОДИ НАВЧАННЯ

Методи навчання, які використовуються в процесі проведення лекційних та практичних занять з навчальної дисципліни ІАД: лекції з використанням мультимедійних презентацій та проблемного викладання, робота в групах, розв'язування ситуаційних завдань, кейсів

6. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг): Навч. Посібник/ В. Ф. Ситник, М.Т. Краснюк - К: КНЕУ, 2007. - 376 с.
2. Акіменко В.В. Прикладні задачі інтелектуального аналізу даних (DATA MINING). К.: КНУ ім. Тараса Шевченка, 2018. 152 с.
3. Черняк О.І. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко ; Київ. нац. ун-т ім. Т. Шевченка. — К. : Знання, 2014. — 599 с.
4. Ситник В.Ф. Засоби дейтамайнінгу для аналізу бізнесових рішень. Науково-практичний журнал "Науково-технічна інформація", №3, 2002. - с. 60-64.
5. McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. O'Reilly Media. September 20, 2022. - 621 p.

Допоміжна

6. Малкіна В. М., Зінов'єва О.Г. Інтелектуальний аналіз даних: Лабораторний практикум. Частина I – Мелітополь: Люкс, 2021. – 150 с.
7. Єремєєв В.С. Теорія ймовірностей та математична статистика/В.С. Єремєєв, Д.О. Сосновських, О.В. Тітова. - Навчальний посібник: Мелітополь: ТОВ «Видавничий будинок ММД», 2009. – 188 с.
8. Ivanova, I., Serdyuk, M., Malkina, V., Tymoshchuk, T., & Shkinder-Barmina, A. (2022). Assessment of the influence of weather factors on the quantitative indicators of sweet cherry fruits by Ridge regression. Scientific Horizons, 25(5), 60-73.

7. ІНФОРМАЦІЙНІ РЕСУРСИ

- 1 Наукова бібліотека ТДАТУ
- 2 Сайти кафедри КН <http://www.tsatu.edu.ua/kn/course/intelektualnyj-analiz-danyh/>
- 3 Освітній портал ТДАТУ <http://op.tsatu.edu.ua/course/view.php?id=888>
- 4 Джерела Internet