

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М.Є. Жуковського
«Харківський авіаційний інститут»

Кафедра математичного моделювання та штучного інтелекту (№ 304)

ЗАТВЕРДЖУЮ

Гарант освітньої програми

 Дмитро ЧУМАЧЕНКО
(підпис) (ім'я та прізвище)

«30» 08 2024 р.

РОБОЧА ПРОГРАМА *ОБОВ'ЯЗКОВОЇ* НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Інтелектуальний аналіз даних

(назва навчальної дисципліни)

Галузі знань: 12 "Інформаційні технології"
(шифр і найменування галузі знань)

Спеціальність: 122 "Комп'ютерні науки"
(код та найменування напрямку підготовки)

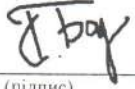
Освітня програма: "Інтелектуальні системи та технології"
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2024 рік

Розробник: Базілевич К.О., доцент каф.304, к.т.н., доцент
(прізвище та ініціали, посада, наукова ступінь та вчене звання)


(підпис)

Робочу програму навчальної дисципліни розглянуто на засіданні кафедри (№304)
математичного моделювання та штучного інтелекту
(назва кафедри)

Протокол № 1 від « 30 » серпня 2024 р.

Завідувач кафедри к.ф.-м.н., доц.
(науковий ступінь та вчене звання)


(підпис)

О. В. Карташов
(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показника	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів – 3,5	Галузі знань <u>12 "Інформаційні технології"</u> (код і найменування)	<i>Обов'язкова</i>
Кількість модулів – 1	Спеціальності <u>122 "Комп'ютерні науки"</u> (код і найменування) Освітні програми <u>"Інтелектуальні системи та технології"</u> (найменування) Рівень вищої освіти: перший (бакалаврський)	Навчальний рік
Кількість змістовних модулів – 2		2024/2025
Індивідуальне завдання Кореляційний аналіз ранжированих даних		Семестр
Загальна кількість годин – 64/105		5-й
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 2,56		Лекції*
		32 години
		Практичні, семінарські
	0 годин	
	Лабораторні	
	32 години	
Самостійна робота	41 година	
Вид контролю	модульний контроль, залік	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: для денної форми навчання: 1-й семестр – 64/41.

* Аудиторне навантаження може бути зменшене або збільшене на одну годину в залежності від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета: надання теоретичних та практичних аспектів аналізу даних та інтелектуального аналізу даних, спрямованих на пошук у необроблених даних раніше невідомих, практично корисних знань та закономірностей, необхідних для прийняття рішень, та розробки інформаційних технологій аналізу даних та інтелектуального аналізу даних.

Завдання: ознайомлення з теоретичними аспектами технології аналізу даних та інтелектуального аналізу даних, формування у студентів базових навичок застосування методів аналізу даних та інтелектуального аналізу даних з використанням інструментальних засобів.

Формування компетентностей:

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

СК11.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.

Результати навчання:

ПРН3. 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.

ПРН4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПРН12. 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 технології Data Mining, Text Mining, Web Mining.

Пререквізити: «Математичний аналіз», «Теорія ймовірності», «Математична статистика».

Кореквізити: «Випадкові процеси».

3. Програма навчальної дисципліни

Змістовий модуль 1. Перша група видів аналізу даних

Тема 1. Поняття вимірювання в статистичному дослідженні

Вимірювальні шкали номінативна шкала, рангова шкала, шкала інтервалів, шкала відносин. Правила ранжирування, перевірка правильності ранжирування, випадок однакових рангів. Форми обліку результатів спостереження.

Тема 2. Дисперсійний аналіз

Постановка задачі однофакторного дисперсійного аналізу. Основні положення дисперсійного аналізу. Статистичні оцінки факторної та остаткової дисперсій. Формулювання та перевірка гіпотези про рівність дисперсій. Постановка задачі двофакторного дисперсійного аналізу. Статистичні оцінки відповідних дисперсій. Формулювання та перевірка гіпотези про рівність дисперсій.

Тема 3. Кореляційний аналіз

Коефіцієнт детермінації як універсальна характеристика ступеню тісноти статистичного зв'язку. Кореляційне відношення. Дослідження лінійної залежності за допомогою парного коефіцієнта кореляції. Множинні та часткові коефіцієнти кореляції.

Тема 4. Множинний регресійний аналіз

Регресійний аналіз. Основні ідеї регресійної моделі. Модель лінійної регресії. Оцінка якості моделі, інтерпретація та оцінки коефіцієнтів регресії, рівень значущості коефіцієнтів. Обмеження регресійної моделі - мультиколінеарність, гомоскедастичність. Нелінійні залежності. Причини виникнення, приклади різних форм нелінійних залежностей в регресійних моделях. Модель логістичної регресії. Інтерпретація коефіцієнтів логістичної регресії.

Змістовий модуль 2. Друга група видів аналізу даних

Тема 5. Дискримінантний аналіз

Поняття класу як генеральної сукупності та базова ідея ймовірностно-статистичних методів класифікації. Функції збитків та ймовірність неправильної класифікації. Параметричний дискримінантний аналіз з повною інформацією. Параметричний дискримінантний аналіз у випадку нормальних класів.

Тема 6. Елементи Data-Mining.

Знайомство з Data Mining. Класифікація, кластеризація, проста та складна класифікація, багатовимірна класифікація, штучна та природна класифікація. Постановка задачі автоматичної класифікації, відстані між окремими об'єктами та класами об'єктів. Функціонали якості розбиття на класи та екстремальна постановка задачі кластер-аналіза. Паралельні кластер-процедури. Послідовні кластер-процедури.

Тема 7. Аналіз часових рядів

Стаціонарність, автоковаріації і автокореляції. Основні описові статистики для часових рядів. Використання лінійної регресії з детермінованими чинниками для моделювання часового ряду. Прогнози по регресії з

детермінованими чинниками. Лаговий оператор. Оптимальне в середньоквадратичних сенсі прогнозування. Згладжування часового ряду.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	денна форма					
	усього	у тому числі				
л		лаб	пр	інд	с.р.	
1	2	3	4	5	6	7
Модуль 1						
Змістовий модуль 1. Перша група видів аналізу даних						
Тема 1. <i>Поняття вимірювання в статистичному дослідженні</i>	6	2	2	–	–	2
Тема 2. <i>Дисперсійний аналіз</i>	20	6	6	–	–	8
Тема 3. <i>Кореляційний аналіз</i>	12	4	4	–	–	4
Тема 4. <i>Регресійний аналіз</i>	20	6	8	–	–	6
Модульний контроль	2	2	–	–	–	–
Разом за змістовим модулем 1	60	20	20	0	0	20
Змістовий модуль 2. Друга група видів аналізу даних						
Тема 5. <i>Дискримінантний аналіз</i>	14	4	4	–	–	6
Тема 6. <i>Елементи Data-Mining</i>	12	4	4	–	–	4
Тема 7. <i>Аналіз часових рядів</i>	12	2	4	–	–	6
Модульний контроль	2	2	–	–	–	–
Індивідуальне завдання – розрахункова	5	–	–	–	–	5
Разом за змістовим модулем 2	45	12	12	0	0	21
Усього годин	105	32	32	0	0	41

5. Теми семінарських занять

№ з/п	Назва теми	Кількість годин
1		

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1		

7. Теми лабораторних занять

№ з/п	Назва теми	Кількі
1	Перевірка гіпотези про вид розподілу ознаки	2
2	Однофакторний дисперсійний аналіз	2
3	Двофакторний дисперсійний аналіз	2
4	Кореляційний аналіз	4
5	Лінійна множинна регресія	4
6	Нелінійна множинна регресія	4
7	Дискримінантний аналіз	4
8	Елементи Data-Mining	6
9	Аналіз часових рядів	4
	Разом	32

8. Самостійна робота

№ з/п	Назва теми	Кількіс ть
1	Поняття рангової кореляції	2
2	Основні задачі статистичного аналізу зв'язків між ранжировками за лач	8
3	Ранговий коефіцієнт кореляції Спірмена, перевірка його	4
4	Ранговий коефіцієнт кореляції Кендалла, перевірка його	6
5	Коефіцієнт конкордації Кендалла, перевірка його	6
6	Кореляційний аналіз категоризованих даних	4
7	Метод головних компонент	6
8	Виконання розрахункової роботи	5
	Разом	41

9. Індивідуальні завдання

№ з/п	Назва теми
1	Кореляційний аналіз ранжованих даних

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами, опублікованими кафедрою (методичні посібники), проведення олімпіад.

11. Методи контролю

Проведення поточного контролю, письмового модульного контролю, фінальний контроль у вигляді заліку.

12. Критерії оцінювання та розподіл балів, які отримують студенти

Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної	Бали за одне заняття	Кількість занять	Сумарна кількість
Змістовний модуль 1			
Виконання і захист	3..5	5	15..25
Виконання і захист РР	13..15	1	13..15
Модульний контроль	10..20	1	10..20
Змістовний модуль 2			
Виконання і захист лабораторних	3..5	4	12..20
Модульний контроль	10..20	1	10..20
Усього за семестр			60...100

Семестровий контроль (залік) проводиться у разі відмови студента від балів поточного тестування й за наявності допуску до заліку. Під час складання семестрового заліку студент має можливість отримати максимум 100 балів.

Білет для заліку складається з двох теоретичних та двох практичних завдань, кожне завдання оцінюється 25 балами.

12.2 Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

Основні цілі та вихідні передумови застосування статистичних методів; методи попередньої обробки даних; основні поняття вибіркового методу; методи перевірки статистичних гіпотез; характеристики статистичного зв'язку кількісних даних; моделі дисперсійного аналізу статистичних даних; методи кореляційного аналізу статистичних даних; методи регресійного аналізу; методи дискримінантного аналізу; методи аналізу часових рядів.

Необхідний обсяг вмінь для одержання позитивної оцінки:

Дослідження та характером статистичних даних; знаходити чисельні характеристики статистичних розподілів вибірок даних; перевіряти основні гіпотези щодо параметрів розподілення даних; видаляти аномальні спостереження у скалярних та векторних даних; робити одно факторний та двофакторний дисперсійний аналіз даних; знаходити рівняння лінійної регресії; перевіряти значущість коефіцієнтів лінійної регресії; знаходити довірчі інтервали для коефіцієнтів лінійної регресії; будувати лінійну множинну регресію; знаходити рівняння нелінійної регресії; визначати статистичні оцінки для параметрів нелінійної регресії; знаходити сукупності у даних методами дискримінантного аналізу; представляти змістовну

інтерпретацію результатів статистичного аналізу.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити всі індивідуальні завдання та здати модульні контрольні роботи. Знати основи статистичної обробки даних за допомогою параметричних статистичних методів, основи аналізу часових рядів.

Добре (75-89). Твердо знати мінімум, захистити всі індивідуальні завдання, виконати всі модульні контрольні роботи. Уміти аналізувати дані різних типів, правильно обирати метод обробки за типом даних (різні шкали, залежні чи незалежні вибірки), здійснювати прогнозування за допомогою алгоритмічних та аналітичних моделей.

Відмінно (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та уміти застосовувати їх.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

Під час складання семестрового *заліку* здобувач має можливість отримати максимум 100 балів.

13. Методичне забезпечення

Методичне забезпечення:

- робоча програма дисципліни;
- конспект лекцій, підручники (навчальні посібники);
- методичні вказівки та рекомендації для виконання лабораторних та практичних робіт, а також рекомендації для самостійної підготовки;
- приклади розв'язування типових задач чи виконання типових завдань;
- питання для контрольних заходів.

Посилання курс у системі дистанційного навчання Ментор

<https://mentor.khai.edu/course/view.php?id=3156>

14. Рекомендована література

Базова

1. Бахрушин В.Є. Аналіз даних. – Запоріжжя: ГУ "ЗІДМУ", 2006. – 170 с.
2. Гирко В.Л. Багатомірний статистичний аналіз. – К.: Вища школа, 1988.

– 320 с.

3. Королюк В.С., Боровских Ю.В. Асимптотичний аналіз розподілу статистик. – К.: Наукова думка, 1984. – 301 с.

4. Паніотто В.І., Максименко В.С., Харченко Н.М. Статистичний аналіз соціологічних даних. – К.: Вид. дім "КМ Академія", 2004. – 270с.

Допоміжна

1. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика. – К.:КНЕУ, 2001.-336 с.

2. Іващенко П.О., Семеняк І.В., Іванов В.В. Багатовимірний статистичний аналіз. – Харків: Основа, 1992. – 144 с.

15. Інформаційні ресурси

Сайт кафедри математичного моделювання та обчислювального інтелекту.