

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

Кафедра математичного моделювання та штучного інтелекту (№ 304)
(назва кафедри)

ЗАТВЕРДЖУЮ

Керівник проектної групи


(підпис)

Дмитро ЧУМАЧЕНКО
(ініціали та прізвище)

« 31 » 08 2021 р.

РОБОЧА ПРОГРАМА ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Аналіз даних

(назва навчальної дисципліни)

Галузі знань:

12 "Інформаційні технології"

(шифр і найменування галузі знань)

Спеціальність:

122 "Комп'ютерні науки"

(код та найменування напряму підготовки)

Освітня програма:

"Інтелектуальні системи та технології"

(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2021 рік

Робоча програма Аналіз даних
(назва дисципліни)
для студентів за спеціальностями 122 "Комп'ютерні науки",
освітньою програмою "Інтелектуальні системи та технології"
« 27 » серпня 2021 р., – 9 с.

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

К.О.Баз
(підпис)

Робочу програму розглянуто на засіданні кафедри математичного
моделювання та штучного інтелекту
(назва кафедри)

Протокол № 2 від « 27 » серпня 2021 р.

Завідувач кафедри Д.Т.Н., доц.
(наукова ступінь та вчене звання)

ДТ
(підпис)

Андрій ЧУХРАЙ
(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 – 5	Галузі знань <u>12 «Інформаційні технології»</u> (код і найменування) Спеціальності <u>122 «Комп’ютерні науки»,</u> (код і найменування) Освітні програми <u>«Інтелектуальні системи та технології»</u> (найменування) Рівень вищої освіти: перший (бакалаврський)	Цикл за вибором
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2021/2022
Індивідуальне завдання		Семестр
Кореляційний аналіз ранжированих даних		5-й
Загальна кількість годин – 64/150		Лекції*
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента –5,6		32 години
		Практичні, семінарські
		0 годин
		Лабораторні
	32 години	
	Самостійна робота	
	86 годин	
	Вид контролю	модульний контроль, залік

Співвідношення кількості годин аудиторних занять до самостійної роботи становить 64/86.

2. Мета та завдання навчальної дисципліни

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

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

ЗК2 – Здатність застосовувати знання у практичних ситуаціях

ЗК3 – Знання та розуміння предметної області та професійної діяльності

ЗК6 – Здатність вчитися і оволодівати сучасними знаннями

3. Програма навчальної дисципліни

Змістовий модуль 1

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

Тема 5. Дискримінантний аналіз

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

Тема 6. Кластер-аналіз

Класифікація без навчання. 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.
	120	30	30	-	-	60
Модуль 1						
Змістовий модуль 1.						
Тема 1. Поняття вимірювання в статистичному дослідженні	6	2	2	–	–	2
Тема 2. Дисперсійний аналіз	24	6	6	–	–	12
Тема 3. Кореляційний аналіз	20	4	4	–	–	12
Тема 4. Регресійний аналіз	28	8	8	–	–	12
Разом за змістовим модулем 1	78	20	20	0	0	38
Модуль 2						
Змістовий модуль 2.						
Тема 5. Дискримінантний аналіз	22	4	4	–	–	14
Тема 6. Кластер-аналіз	22	4	4	–	–	14
Тема 7. Аналіз часових рядів	28	4	4	–	–	20
Разом за змістовим модулем 2	72	12	12	0	0	48
Усього годин	150	32	32	0	0	86

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	Кількість годин
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	Кластер-аналіз	6
9	Аналіз часових рядів	4
	Разом	32

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	<i>Поняття рангової кореляції</i>	10
2	Основні задачі статистичного аналізу зв'язків між ранжировками задач	8
3	Ранговий коефіцієнт кореляції Спірмена, перевірка його значущості	10
4	Ранговий коефіцієнт кореляції Кендалла, перевірка його значущості	10
5	Коефіцієнт конкордації Кендалла, перевірка його значущості	10
6	<i>Кореляційний аналіз категоризованих даних</i>	10
7	<i>Метод головних компонент</i>	10
8	Основні чисельні характеристики головних компонент	10
9	Оптимальні властивості головних компонент	8
	Разом	86

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 роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Виконання і захист лабораторних робіт	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 (практичних) робіт	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 іспиту 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

13. Методичне забезпечення

Навчально-методичний комплекс дисципліни, який включає в себе:

- робоча програма дисципліни;
- конспект лекцій, підручники (навчальні посібники), в тому числі в електронному вигляді, які за змістом повністю відповідають робочій програмі дисципліни;
- лабораторних та практичних робіт, а також рекомендації для самостійної підготовки;
- тематики індивідуальних завдань;
- приклади розв'язування типових задач чи виконання типових завдань;

- питання, тести для контрольних заходів;
- каталоги інформаційних ресурсів.

Електронний ресурс, на якому розміщено навчально-методичний комплекс дисципліни – http://library.khai.edu/library/fulltexts/doc/002_Anali_Dani.pdf

14. Рекомендована література

Базова

1. Анализ данных/ Н.С. Бакуменко, О.С. Радивоненко. – Учеб. пособие по лабораторному практикуму. – Харьков: Нац. аэрокосм. ун-т «Харьк. авиац. ин-т.», 2007 – 87с.
2. Решение экономических задач с использованием статистических пакетов обработки данных: учеб. Пособие по лаб. практикуму / М.С. Мазорчук, Н.С. Бакуменко. – Харьков: Нац. аэрокосм. ун-т «Харьк. авиац. ин-т», 2008. – 95с.

14. Рекомендована література

Базова

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрии. – М.: ЮНИТИ, 1998.
2. Анализ данных: учебник для академического бакалавриата / под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2016. — 490 с
3. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. – М.: ИНФРА, 1998, - 528 с.
4. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистика в науке и бизнесе. – К.: МОРИОН, 2002. – 640 с.

Допоміжна

1. Осипов Г.В. Андреев Э.П. Методы измерения в социологии. – М.: Наука, 1977. – 108 с.
2. Адлер Ю.П. Маркова Е.В. Планирование эксперимента при поиске оптимальных условий. – М. : Наука, 1976. – 280 с.
3. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 755 с.
4. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя: КПУ, 2011. – 268 с.

15. Інформаційні ресурси

Сайт кафедри <http://k304.khai.edu/>