


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

Кафедра математичного моделювання та штучного інтелекту (№ 304)

ЗАТВЕРДЖУЮ

Керівник проектної групи

 Д. І. Чумаченко
(підпис) (ініціали та прізвище)

« 31 » 08 2021 р.

**РОБОЧА ПРОГРАМА ВИБІРКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Методи аналізу даних в соціально-економічних системах
(шифр і назва навчальної дисципліни)

Галузь 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" (4 роки), "Інтелектуальні системи та технології" (Скорочена форма навчання, 3 роки)
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2021 рік

Робоча програма навчальної дисципліни

«Методи аналізу даних в соціально-економічних системах»

(назва навчальної дисципліни)

для студентів за спеціальністю 122 "Комп'ютерні науки"

освітніми програмами "Інтелектуальні системи та технології"(4 роки), "Інтелектуальні системи та технології" (Скорочена форма навчання, 3 роки)

«4» серпня 2021 р. – 12 с.

РОЗРОБНИК ПРОГРАМИ:

Пічугіна О. С., професор кафедри 304, д. ф.-м. н., доцент
(прізвище та ініціали, посада, наукова ступень та вчене звання)



(підпис)

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

Протокол № 2 від «27» серпня 2021 р.

Завідувач
кафедри № 304

д.т.н., професор
(науковий ступінь і вчене звання)


(підпис)

А. Г. Чухрай
(ініціали та прізвище)

Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, рівень вищої освіти	Характеристика навчальної дисципліни
		Денна форма навчання
Кількість кредитів – 4.5	Галузь знань <u>12 "Інформаційні технології"</u> <small>(шифр та найменування)</small> Спеціальність <u>122 "Комп'ютерні науки"</u> <small>(код та найменування)</small> Освітня програма <u>"Інтелектуальні системи та технології"</u> <small>(найменування)</small> Рівень вищої освіти: <u>перший (бакалаврський)</u> <small>(найменування)</small>	Вибіркова
Модулів – 2		Рік підготовки:
Змістових модулів – 3		2021/2022
Індивідуальне науково-дослідне завдання: <u>розрахункова робота "Однокомпонентні економетричні моделі. Системи одночасних рівнянь"</u>		Семестр
Загальна кількість годин – 135 денна – 64/135		7-й
		Лекції
Тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 4.4		32 год.
		Практичні
		-
		Лабораторні
	32 год.	
	Самостійна робота	
	71 год.	
	Індивідуальна робота	
-		
Вид контролю		
модульний контроль, іспит		

Примітка

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить:

для денної форми навчання – 64/71.

2. Мета та завдання навчальної дисципліни

Мета навчальної дисципліни: надання теоретичних та практичних аспектів інтелектуального аналізу даних (Data Analysis, Data Mining), 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 навчальної дисципліни: ознайомлення з теоретичними аспектами технології Data Mining, формування у студентів базових навичок застосування методів інтелектуального аналізу даних з використанням інструментальних засобів DataMining.

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких **загальних компетентностей**:

- Здатність застосовувати знання у практичних ситуаціях (ЗК 2)
- Знання та розуміння предметної області та професійної діяльності (ЗК 3).
- Здатність вчитися і оволодівати сучасними знаннями (ЗК 6).

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких **фахових компетентностей**:

- Ґрунтовна математична підготовка, а також підготовка з теоретичних, методичних і алгоритмічних основ інформаційних технологій для використання математичного апарату під час вирішення прикладних і наукових завдань в області інформаційних систем і технологій (ФК 1)
- Базові знання в області системних досліджень і вміння застосовувати їх під час управління ІТ-проектами, здійснення моделювання систем, проведення системного аналізу об'єктів інформатизації, прийняття рішень, розробки методів і систем штучного інтелекту (ФК 4)

Програмні результати навчання.

Відповідно до освітньої програми студент повинен досягти наступних програмних результатів:

- ПРН 4. Знання сучасних методів побудови та аналізу ефективних алгоритмів і вміння їх реалізувати в конкретних застосуваннях (ПРН 4).
- ПРН 9. Знання принципів структурного програмування, сучасних процедурно-орієнтованих мов, основних структур даних і вміння їх застосовувати під час програмної реалізації алгоритмів професійних завдань (ПРН 9).

Міждисциплінарні зв'язки: для вивчення дисципліни «Методи аналізу даних в соціально-економічних системах» необхідно володіти запасом знань таких дисциплін, як «Алгебра та геометрія», «Математичний аналіз», «Методи обчислень», «Дискретна математика», «Теорія ймовірностей та математична статистика», «Аналіз даних», а в подальшому знання з дисципліни «Методи аналізу даних в соціально-економічних системах» стануть основою для вивчення таких дисциплін, як «Інтелектуальні системи підтримки прийняття рішень» та ін.

3. Програма навчальної дисципліни

Модуль 1

Змістовий модуль 1. Однокомпонентні економетричні моделі

Тема 1. Теоретичні основи статистичного аналізу економетричних моделей

- Основні задачі аналізу даних

Етапи аналізу даних. Класифікація типів змінних. Кількісні, ординальні та номінальні дані та робота з ними. Групування даних. Отримання вибіркового характеристик за згрупованими даними.

Багатовимірний статистичний аналіз (БСА) і його основні задачі та апарат дослідження. Класифікація показників статистичного аналізу. Розділи БСА та їх основні задачі.

- Попередня обробка даних

Визначення головних характеристик скалярних змінних: центру групування значень, розсіювання значень, скошеності та гостроверхості розподілу і т.д. Характеристики векторних спостережень. Перевірка основних гіпотез.

- Кореляційний аналіз

Характеристики парного та множинного статистичного зв'язку кількісних даних: індекс кореляції, коефіцієнт детермінації, кореляційне відношення, множинний коефіцієнт кореляції і т.д. Частинні коефіцієнти кореляції.

Аналіз рангових кореляцій. Ранг, об'єднаний ранг. Таблиці рангів. Парні рангові коефіцієнти кореляції. Критерії Спірмена, Кендала.

- Регресійний аналіз.

Класифікація регресійних моделей. Квазілінійні парні регресії. Коефіцієнт множинної детермінації R^2 , різні формули його обчислення, застосування його при перевірці адекватності моделей. Надійні зони базисних змінних та прогнозних значень для парних та множинних, лінійних нелінійних регресій (для множинної регресії – в матричній формі).

Тема 2. Дослідження реальних економетричних моделей

- Суттєво нелінійні регресійні моделі, що зводяться до лінійних.

Виробнича регресія Кобба-Дугласа (ВРКД). Гіпотези, яким має задовольняти функція залежності випуску продукції від затрат на нього. Перевірка їх виконання для ВРКД. Частинні та сумарні коефіцієнти еластичності та їх економічний зміст. Дослідження двофакторної ВРКД: ізокванти, темпи приросту показника. (Виробнича регресія Кобба-Дугласа-Тінбергена).

Логістична регресія: застосування в економіці та методи оцінювання її параметрів.

- Квазілінійні множинні регресії: економетрична модель аналізу індивідуального ринку. Основна гіпотеза для функції залежності попиту від ціни, перевірка її. Залежність товарообігу і прибутку від ціни, визначення прогнозних значень цін для максимального товарообігу і прибутку, прогнозні інтервали збільшення і зменшення товарообігу і прибутку при збільшенні ціни. Коефіцієнти еластичності попиту, товарообігу і прибутку та їх властивості. Зони еластичного попиту. Область практичного застосування квадратичної функції попиту, дослідження функції попиту для суспільства в залежності від ціни і категорій населення.
- Автокореляція збурень початкових даних. Автокореляція першого порядку, коефіцієнт автокореляції ρ та методи його оцінювання. Критерій Дарбіна-Уотсона, рекомендації по подальшому дослідженню моделі в разі наявної, відсутньої і невизначеної автокореляції. Матриця коваріацій початкових даних моделі з автокореляцією першого порядку. Узагальнений метод найменших квадратів.
- Авторегресивні і дистрибутивно-лагові моделі.

Лаги. Приклади економетричних моделей з лаговими змінними (функція споживання). Короткострокові та довгострокові мультиплікатори. Класифікація лагових моделей (скінченні та нескінченні). Труднощі у дослідженнях лагових моделях та методи їх усунення.

Метод Ширлі-Алмон дослідження дистрибутивно-лагових моделей: спрощена і загальна схема. Переваги застосування спрощеної і загальної схеми. Узагальнення методу на декілька незалежних лагових змінних.

Метод Койка дослідження дистрибутивно-лагових моделей. Основна гіпотеза методу. Застосування методу для однієї та декількох незалежних змінних, для випадку одного та кількох параметрів для різних змінних, виконання основної гіпотези методу, починаючи з деякої лагової змінної.

- Фіктивні змінні та їх застосування у регресійному аналізі.
Дослідження попиту-пропозиції в залежності від соціального статусу населення.

Модульний контроль

Модуль 2

Змістовий модуль 2. Багатовимірний статистичний аналіз. Системи одночасних рівнянь

Тема 3. Дослідження систем одночасних рівнянь

- Системи одночасних рівнянь (СОР).

Економетричні моделі з СОР і тотожностями. Матрична форма СОР. Методи оцінювання параметрів одночасних рівнянь, вибір методу дослідження в залежності від міри ідентифікованості рівнянь СОР. Необхідна умова і критерій ідентифікованості економетричної моделі. Практична перевірка ідентифікованості, методи боротьби з неідентифікованістю рівнянь.

Системи одночасних рівнянь (СОР). Форми СОР – прогнозна і структурна. Перехід від однієї до іншої і необхідні умови можливості цього переходу.

- Рекурсивні моделі одночасних рівнянь.

Визначення міри ідентифікованості рівнянь у рекурсивній моделі з 5-ти рівнянь. Оцінювання параметрів рекурсивної моделі. Приклад реальної економетричної рекурсивної моделі.

– Дисперсійний аналіз. Модифікації МНК.

Непрямий метод найменших квадратів (НМНК). Реалізація НМНК для економетричної моделі відкритої економіки України, яка є точно ідентифікованою. НМНК для системи двох регресій з центрованими величинами. Порівняння оцінок коефіцієнтів структурної форми, одержаних обома способами.

Двокроковий метод найменших квадратів (ДМНК). Етапи методу, реалізація їх в матричній формі. Модифікований двокроковий метод найменших квадратів (МДМНК). Етапи методу, реалізація їх в матричній формі. Реалізація методів для моделі з двома одночасними рівняннями. Порівняння ДМНК і МДМНК. Економетрична модель відкритої економіки України, яка є неідентифікованою, та застосування ДМНК і МДМНК в її дослідженні.

– Причинний аналіз.

Графічне представлення причинних відношень. Структурні коефіцієнти. Складання систем одночасних рівнянь по причинним діаграмам. Графічне представлення лінійних і нелінійних причинних відношень, способи переходу від нелінійних до лінійних причинних відношень та їх графічне представлення.

Правила редукції причинних діаграм.

Змістовий модуль 3. Аналіз даних та машинне навчання

Тема 4. Задачі аналізу даних та машинного навчання та статистичні методи їх розв'язання

– Задачі класифікації.

Простори об'єктів, спостережень, ознак. Основні постановки задач класифікації. Класифікація з учителем у випадку нормальних спостережень. Підстановочне розв'язуюче правило. Лінійний дискримінантний аналіз

– Кластерний аналіз. Огляд методів кластерного аналізу. Метод к-середніх

– Метод головних компонент

Дослідження вибірових спостережень: метод Фаррара-Глобера визначення мультиколінеарності вибірових даних і метод головних компонент дослідження міри розсіювання вибірових даних. Застосування методу множників Лагранжа у визначенні перших двох головних компонент. Практична реалізація методу головних компонент: пошук базису власних векторів і характеристичних значень симетричної матриці.

Модульний контроль

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин				
	денна форма				
	усього	у тому числі			
л		п	лаб	с.р.	
1	2	3	4	5	6
Модуль 1					
Змістовий модуль 1. Однокомпонентні економетричні моделі					
Тема 1. Теоретичні основи статистичного аналізу економетричних моделей	25	6		4	15
Тема 2. Дослідження реальних економетричних моделей	42	13		14	15
Модульний контроль 1	1	1			
Разом за змістовим модулем 1	68	20	0	18	30
Модуль 2					
Змістовий модуль 2. Багатовимірний статистичний аналіз. Системи одночасних рівнянь					
Тема 3. Дослідження систем одночасних рівнянь	39	8		10	21
Разом за змістовим модулем 2	39	8	0	10	21
Змістовий модуль 3. Аналіз даних та машинне навчання					
Тема 4. Задачі аналізу даних та машинного навчання та статистичні методи їх розв'язання	27	3		4	20
Модульний контроль 2	1	1			
Разом за змістовим модулем 3	28	4	0	4	20
Усього годин	135	32	0	32	71

5. Теми семінарських занять

№ з/п	Назва теми	Кількість годин
		Денна форма навчання
	—	

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин
	—	

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1.	Нелінійна регресія 1 (підбір найкращої регресії зі списку заданих за допомогою функції “ЛИНЕЙН”).	2
2.	Нелінійна регресія 2 (для двох регресійних моделей визначити тип замін, знайти точкові і інтервальні оцінки прогнозу і перевірити адекватність). <ul style="list-style-type: none"> • Ч.1. • Ч.2 	2
3.	Логістична регресія.	2
4.	Виробнича регресія Кобба-Дугласа (двофакторна і трифакторна).	2
5.	Аналіз індивідуального ринку (розрахунок коефіцієнтів еластичності, прогнозів і надійних зон базисних і прогнозних змінних для попиту, товарообігу і прибутку).	2
6.	Фіктивні змінні	2
7.	Гетероскедастичність. УМНК	2
8.	Авторегресивні і дистрибутивно-лагові моделі. Дослідження дистрибутивно-лагової моделі методом Ширлі-Алмон за допомогою спрощеної і загальної схеми.	2
9.	Дистрибутивно-лагові моделі. Метод Койка	2
10.	Автокореляція. Узагальнений метод найменших квадратів (УМНК).	2
11.	Системи одночасних рівнянь. Рекурсивні моделі	2
12.	Системи одночасних незалежних рівнянь, дослідження моделі “попит-пропозиція”.	2
13.	Непрямий МНК.	2
14.	Двокроковий МНК.	2
15.	Дискримінантний аналіз.	2
16.	Визначення головних компонент та виявлення мультиколінеарності експериментальних даних.	2
	<i>Разом</i>	32

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1.	Теоретичні основи статистичного аналізу економетричних моделей	15
2.	Дослідження реальних економетричних моделей	15
3.	Дослідження систем одночасних рівнянь	21
4.	Задачі аналізу даних та машинного навчання та статистичні методи їх розв'язання	20
	<i>Разом</i>	71

9. Індивідуальні завдання

1. Виконання розрахункової роботи “Однокомпонентні економетричні моделі. Системи одночасних рівнянь”

У РР входить 6 з 16-ти лабораторних робіт на вибір.

Структура роботи:

- 1) Титульний лист із зазначенням номеру варіанту на додаток до стандартної інформації про студента і предмет;
- 2) По кожній роботі
 - завдання;
 - хід виконання роботи (детальний опис з математичними моделями, використаними позначеннями, підстановкою у формули та перетвореннями).

Зауваження. Пункт 3) виконується лише по тих лабораторних роботах, на яких студент був відсутнім і не захистив роботу на момент здачі РР;

- Розрахункові листи xls;
- Загальні висновки – на одну сторінку максимум;
- Джерела інформації - не обов'язково.

Структура роботи може бути довільною – все підряд по кожній роботі або спочатку всі завдання, потім усі звіти, далі розрахункові листи

Обсяг роботи – 20 сторінок.

10. Методи навчання

1. Пояснювально-ілюстративний (інформаційно-рецептивний) метод: та метод проблемного виконання (лекція).
2. Репродуктивний (лабораторні роботи).
3. Частково-пошуковий (евристичний) та дослідницький: (самостійна робота та виконання РР).

Дисципліна «Інтелектуальні системи підтримки прийняття рішень» передбачає лекційні (в т. ч. із використанням мультимедійного обладнання), лабораторні заняття під керівництвом викладача та самостійну роботу студента за підручниками і методичними матеріалами (методичні посібники і мережеві ресурси), що забезпечує закріплення теоретичних знань, сприяє набуттю практичних навичок і розвитку самостійного наукового мислення. Передбачено регулярні консультації.

11. Методи контролю

Визначення рівня засвоєння студентом навчального матеріалу дисципліни здійснюється шляхом проведення поточних і підсумкових контролів. У завдання поточного контролю входить систематична перевірка розуміння й засвоєння студентом програмного матеріалу, виконання завдань на лабораторних роботах, уміння самостійно проробляти тексти складання конспектів, написання звітів, рефератів, здатності усно або письмово представляти певний матеріал. Перед підсумковим контролем ставиться завдання перевірки глибини засвоєння студентом програмного матеріалу дисципліни, логіки й взаємозв'язки між її окремими розділами, здатності творчо використати придбані знання, уміння сформулювати своє відношення до проблеми, що впливає зі змісту дисципліни.

12. Розподіл балів, які отримують студенти (екзамен)
12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття завдання)	Кількість занять (завдань)	Сумарна кількість балів
Модуль 1			
Змістовий модуль 1			
Робота на лекціях	0...2	6.5	0...13
Виконання і захист лабораторних робіт	0...3	7	0...21
Модульний контроль	0...8	1	0...8
Модуль 2			
Змістовий модуль 2			
Робота на лекціях	0...2	4	0...8
Виконання і захист лабораторних робіт	0...3	5	0...15
Змістовий модуль 3			
Робота на лекціях	0...2	1.5	0...3
Виконання і захист лабораторних робіт	0...3	2	0...6
Модульний контроль	0...8	1	0...8
Виконання і захист РР	0...6	1	0...6
Усього за семестр			0...100

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

Білет для іспиту/заліку складається з двох теоретичних і двох практичних запитань, вартість кожного з яких і результативній оцінці складає 25 балів зі 100.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

- предмет та основні поняття багатовимірного аналізу даних (БАД) та економетрики;
- класифікацію задач БАД та підходи до розв'язання кожного класу;
- можливості Excel для розв'язання різних класів задач БАД та економетричного аналізу.

Необхідний обсяг вмінь для одержання позитивної оцінки:

- використовувати основні методи і процедури статистичного аналізу даних;
- проводити попередню обробку вхідних даних;
- готувати вхідні дані для практичних розрахунків на основі реальних ситуацій;
- проводити математичне моделювання поставленої задачі;
- обґрунтовувати вибір статистичного методу розв'язування задач БАД та способів їх реалізації на ЕОМ;
- проводити необхідні обчислення з отримання розв'язку задачі БАД
- здійснювати аналіз отриманих результатів;
- коректно і зрозуміло оформляти розв'язок задач, давати економічну чи фізичну інтерпретацію результатів.

12.3 Критерії оцінювання роботи студента протягом семестру.

Задовільно (60-74). Мати мінімум знань та умінь, виконати та захистити розрахункову роботу, здати модульний контроль. Відвідати не менше половини лекцій, виконати та захистити на бал 60%-74% від максимальної оцінки всі лабораторні роботи (або виконати та захистити на вищий бал 60%-74% лабораторних робіт).

Добре (75 - 89). Твердо оволодіти основною частиною необхідних знань і умінь, виконати та захистити розрахункову роботу, здати модульний контроль. Відвідати не менше 3/4 лекцій, виконати та захистити на бал 75%-89% від максимальної оцінки всі лабораторні роботи (або виконати та захистити на вищий бал 75%-89% лабораторних робіт).

Відмінно (90 - 100). У повному обсязі оволодіти основним та додатковим матеріалом, що стосується знань і умінь. Виконати та захистити розрахункову роботу, здати модульний контроль. Відвідати усі лекції, виконати та захистити на бал щонайменше 90% від максимальної оцінки всі лабораторні роботи (або виконати та захистити на вищий бал щонайменше 90% лабораторних робіт).

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

1. Електронний ресурс, на якому розміщено навчально-методичний комплекс дисципліни «Методи аналізу даних в соціально-економічних системах»

14. Рекомендована література

Базова

1. Андрейчиков, А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике: М.: Финансы и статистика, 2000. - 368 с.; 2-е изд. - 2004. - 470 с.
2. Толбатов Ю.А. Математична статистика та задачі оптимізації в алгоритмах і програмах: Навч. Посібник.- К.: Вища шк. - 1994. - 399 с.
3. Толбатов Ю.А. Эконометрика: Підручник для студентів екон. Спец. Вузів.- К.: Четверта хвиля. - 1997. - 320 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. Пособие для вузов.- М.: Высш.школа. - 1977. - 479 с.
5. Джонстон Дж. Эконометрические методы.- М.: Статистика. - 1980г. - 444 с.

Допоміжна

1. Вероятность и математическая статистика: Энциклопедия/ под ред. Ю.В.Прохорова.- М.: Большая Российская энциклопедия. - 1999. - 910 с.
2. Хейс. Д. Причинный анализ в статистических исследованиях.- М.: Финансы и статистика. - 1981. - 225 с.
3. Колемаев В.А и др. Теория вероятностей и математическая статистика: Учеб пособие для экономических спец. Вузов.- М.: Высш. Шк. - 1991. - 400 с.

15. Інформаційні ресурси

1. сайт кафедри №304 - <http://k304.khai.edu>
2. Greene, William (2012). Econometric Analysis (7th ed.). Pearson Education. pp. 34, 41–42. ISBN 9780273753568.
3. Wooldridge, Jeffrey (2013). Introductory Econometrics, A modern approach. South-Western, Cengage learning. ISBN 978-1-111-53104-1.
4. Pearl, Judea (2000). Causality: Model, Reasoning, and Inference. Cambridge University Press. ISBN 978-0521773621.
5. Н. Р. Pesaran (1990), "Econometrics," Econometrics: The New Palgrave, p. 2, citing Ragnar Frisch (1936), "A Note on the Term 'Econometrics'," Econometrica, 4(1), p. 95.