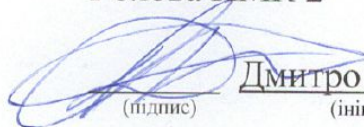


Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Голова НМК 2



Дмитро КРИЦЬКИЙ

(підпис)

(ініціали та прізвище)

« 31 » 08 2024 р.

**РОБОЧА ПРОГРАМА
ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Data Mining

(назва навчальної дисципліни)

Дисципліна індивідуального вибору 4

(назва вибіркового блоку)

Галузі знань: усі галузі, за якими відбувається підготовка здобувачів в університеті.

Спеціальності: усі спеціальності, за якими відбувається підготовка здобувачів в університеті.

Освітні програми: усі освітні програми відповідних спеціальностей, за якими відбувається підготовка здобувачів в університеті.

Форма навчання: денна

Рівень вищої освіти: другий (магістерський)

Харків 2024 рік

Розробник: Дмитро КРИЦЬКИЙ к.т.н., доцент каф. 105.

Марія ПИВОВАР асистент каф 105

(посада, науковий ступінь і вчене звання, ім'я та прізвище)


(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій
проектування

(назва кафедри)

Протокол № 1 від «30» 08 2024 р.

В.о. завідувача кафедри 105


(підпис)

Аліна АРТЬОМОВА

(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	Галузь знань усі галузі, за якими <u>відбувається підготовка здобувачів в університеті.</u> (шифр і найменування) Спеціальність <u>усі спеціальності наведених галузей</u> (код і найменування) Освітня програма <u>усі освітні програми наведених галузей</u> (найменування) Рівень вищої освіти: <u>другий (магістерський)</u>	<i>Вибіркова</i>
Кількість модулів – 2		Навчальний рік
Кількість змістовних модулів – 2		2024/2025
Індивідуальне завдання не передбачено		Семестр
Загальна кількість годин – 64 / 150		2-й
		Лекції*
Кількість тижневих годин для денної форми навчання: аудиторних – 4; самостійної роботи студента – 5,4.		32 години
		Практичні, семінарські*
		0 годин
		Лабораторні*
	32 години	
	Самостійна робота	
	86 годин	
Вид контролю		
	модульний контроль, іспит	

Співвідношення кількості годин ауд. занять до сам. роботи становить 0,74.

¹⁾ Аудиторне навантаження може бути зменшеним або збільшеним на одну годину в залежності від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета вивчення навчальної дисципліни полягає у наданні теоретичних та практичних аспектів інтелектуального аналізу даних (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.

Основними завданнями є ознайомлення студентів з теоретичними аспектами технології Data Mining, формування у студентів базових навичок застосування методів інтелектуального аналізу даних з використанням інструментальних засобів Data Mining.

У результаті вивчення навчальної дисципліни студент повинен:

знати :

- основні поняття та визначення інтелектуального аналізу даних;
- методи побудови моделей та аналізу залежностей у великих масивах даних;
- сучасні програмні засоби для проектування і розробки систем інтелектуального аналізу даних.

- класи систем Data Mining;
- стандарти Data Mining;
- методи візуального аналізу;

вміти:

- обґрунтовувати й аналізувати вибір конкретного типу моделі та методу інтелектуального аналізу даних при вирішенні відповідних практичних задач;
- використовувати сучасні програмні засоби для проектування та дослідження систем інтелектуального аналізу даних;
- аналізувати результати побудови та використання систем інтелектуального аналізу даних при вирішенні прикладних задач.

Компетентності, які набуваються:

1. Здатність застосовувати знання у практичних ситуаціях.
2. Знання та розуміння предметної області та розуміння професійної діяльності.
3. Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).
4. Здатність до пошуку, оброблення та аналізу інформації з різних джерел
5. Здатність забезпечити організацію обчислювальних процесів в інформаційних системах різного призначення з урахуванням архітектури, конфігурування, показників результативності функціонування операційних систем і системного програмного забезпечення

3. Програма навчальної дисципліни

МОДУЛЬ 1

Модуль 1.

Змістовий модуль 1.

Тема 1. Задачі інтелектуального аналізу даних.

Інтелектуальний аналіз даних як міждисциплінарна область. Основні поняття та цілі інтелектуального аналізу. Задачі систем підтримки прийняття рішень (СППР). Типи підсистем аналізу в складі СППР. Вимоги до систем інтелектуального аналізу.

Тема 2. Сховища даних.

Концепція сховища даних (СД). Методи первісної обробки даних. Методи дослідження структури даних: візуалізація та автоматичне групування даних. Представлення даних у вигляді гіперкубу.

Тема 3. Упровадження Data Mining, OLAP і сховищ даних у СППР.

Категорії даних в СД: детальні дані, агреговані, метадані. Детальні дані: вимір та факт. Типи фактів. Види агрегованих даних: адитивні, напівадитивні, неадитивні. Загальна архітектура ХД. Вітрина даних. Структура СППР з фізичним СД. Структура СППР з віртуальним СД. Структура СППР з вітриною даних. Структура СППР з СД і вітриною даних.

Тема 4. OLAP-системи.

Характеристики та призначення OLAP-систем. Операції з гіперкубом в OLAP-системах: зріз, обертання, деталізація, консолідація. Задачі Data Mining: класифікація, регресія, пошук асоціативних правил, кластеризація. Стандарти Data Mining. Інструменти Data Mining.

Тема 5. Методи класифікації.

Правила класифікації. Етапи класифікації: побудова моделі, класифікація нових даних. Види методів класифікації: статистичні та методи машинного навчання. Алгоритми побудови дерева рішень. Повне дерево рішень. Міри ефективності дерев рішень. Критерії вибору кращих атрибутів. Індекс Джині. Зменшення ентропії або приріст інформації. Тест хи-квадрат. Зменшення дисперсії. F-критерій Фішера. Алгоритми ID3. Алгоритм C4.5. Алгоритм CART. Методи опорних векторів, «найближчого сусіда», Naïve Bayes.

Тема 6. Кластерний аналіз.

Цілі кластеризації. Ієрархічна та секційна кластеризація. Методи кластеризації: процедура Мак-Кіна, метод k-means, G-means.

МОДУЛЬ 2.

Змістовний модуль 2.

Тема 7. Метод Apriori.

Послідовне відображення шаблонів даних. Метод Apriori. Різновиди методу Apriori.

Тема 8. Статистичний аналіз.

Аналіз багатовимірних угруповань. Статистична обробка даних. Статистична обробка часових рядів і прогнозування. Кореляційний і регресійний аналіз даних. Лінійна і логістична регресія. Множинний регресійний аналіз. Лінійна множинна регресійна модель. Перевірка адекватності моделі. Нелінійне оцінювання параметрів.

Тема 9. Пошук асоціативних правил.

Основні поняття методів пошуку асоціативних правил: асоціаційні правила, транзакція, предметний набір. Структура асоціативного правила. Показники зв'язку між наборами предметів: підтримка, достовірність, значимість, ліфт, леверідж.

Тема 10. Сітка Кохонена.

Алгоритм роботи з сіткою Кохонена. Алгоритм Fuzzy C-means. Кластеризація по Гюстафсону-Кесселю. Адаптивні методи кластеризації. Проблеми методів кластеризації.

Тема 11. Види побудови та аналізу ансамблів моделей.

Комбінування рішень. Види ансамблів моделей. Метод бэггінга для аналізу моделей. Метод бустінга для аналізу моделей. Альтернативні методи побудови ансамблів моделей: аддитивна регресія, адитивна логістична регресія.

Тема 12. Методи оцінки моделей.

Методи порівняння моделей. Оцінка похибки моделей. Похибки помилкової класифікації. Оцінка моделей з Lift-кривими. Оцінка моделей за Gain-діаграмами. Profit-криві. ROC-аналіз.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин				
	Усього	У тому числі			
		л	п	лаб.	с. р.
Модуль 1					
Тема 1. Задачі інтелектуального аналізу даних	12	2	2	-	6
Тема 2. Сховища даних.	14	2	2	-	6
Тема 3. Упровадження Data Mining, OLAP і сховищ даних у СППР.	10	3	-	-	7
Тема 4. OLAP-системи.	12	3	2	-	7
Тема 5. Методи класифікації	8	3	6	-	7
Тема 6. Кластерний аналіз.	12	3	4	-	7
Модульний контроль 1	2	-	-	-	2
Разом за модулем 1.	74	16	16	-	42
Модуль 2					
Тема 7. Метод Apriori.	8	2	3	-	7
Тема 8. Статистичний аналіз.	10	2	3	-	7
Тема 9. Пошук асоціативних правил	14	3	5	-	7
Тема 10. Сітка Кохонена.	8	3	-	-	7
Тема 11. Види побудови та аналізу ансамблів моделей	10	3	5	-	7
Тема 12. Методи оцінки моделей.	6	3	-	-	7
Модульний контроль 2	2	-	-	-	2
Разом за модулем 2	76	16	16	-	44
Усього годин	150	32	32	-	86

5. Теми семінарських занять

Семінарські заняття навчальним планом не передбачені.

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Знайомство з програмною бібліотекою Weka	4
2	Класифікація даних з використанням програмного пакету Weka	4
3	Кластеризація даних з використанням програмного пакету Weka	4
4	Регресійний аналіз з використанням програмного пакету Weka	5
5	Побудова асоціативних правил з використанням програмного пакету Weka	5
6	Побудова дерева рішень	5
7	Ансамбль моделей	5
	Разом	32

7. Теми лабораторних занять

Лабораторні заняття навчальним планом не передбачені

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Види СППР. Задачі інтелектуального аналізу даних	6
2	Види сховищ даних. Види даних в сховищах даних.	6
3	Побудова сховищ даних та проведення аналізу.	7
4	Робота з OLAP-системами. Переваги та недоліки OLAP-систем з позицій аналізу даних.	7
5	Задача пошуку асоціативних правил.	7
6	Метод Apriori та його різновиди.	7
7	Модульний контроль	2
8	Методи кластерного аналізу k-means та G-means, процедура Мак-Кіна, метод C-means.	7
9	Сіті Кохонена, що самоорганізуються. Карта Кохонена. Підготовка до модульної контрольної роботи №1	7
10	Методи класифікації. Класифікація за допомогою дерева рішень.	7
11	Методи статичного аналізу даних	7
12	Методи побудови ансамблю моделей	7
13	Методи оцінки моделей. Підготовка до модульної контрольної роботи №2.	7
14	Модульний контроль	2
	Разом	86

9. Індивідуальне завдання

Індивідуальне завдання навчальним планом не передбачено.

10. Методи навчання

При проведенні лекцій, практичних робіт та самостійної роботи використовуються такі методи навчання як словесні (пояснення, розповідь, бесіда, навчальна дискусія та ін.); наочні (ілюстрування, демонстрування, самостійне спостереження) та практичні, а саме лекції проводяться з використанням основних розділів конспекту лекцій в електронній формі, демонстрацій окремих прийомів роботи з інструментальним середовищем, практичні роботи виконуються з використанням навчальних та ліцензованих робочих версій середовищ програмування та аналізу даних.

Самостійна робота включає підготовку до практичних робіт, модульного контролю та іспиту, вивчення вказаних вище тем за конспектом, літературними джерелами та програмною документацією.

11. Методи контролю

Контроль здійснюється згідно з «Положення про рейтингове оцінювання досягнень студентів».

Поточний контроль – відповідно до повноти, якості та своєчасності виконання лабораторних робіт та розділів домашнього завдання; проміжний (модульний) контроль – письмові контрольні роботи на 8-му та 16-му тижнях; підсумковий контроль – письмовий іспит.

12. Розподіл балів, що отримують студенти

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			
Робота на лекціях	0...1	5	0...5
Виконання та захист практичних робіт	0.5	3	0...15
Модульний контроль	0...25	1	0...25
Модуль 2			
Робота на лекціях	0...1	5	0...5
Виконання та захист практичних робіт	0.5	4	0...20
Модульний контроль	0...30	1	0...30
Всього за семестр			0...100

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування та за наявності допуску до іспиту/залику. При складанні семестрового іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з 20 теоретичних та практичних питань у формі тесту с закритими та відкритими питаннями. Максимальна кількість балів за кожне питання – 5 балів (сума – 100 балів).

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

- мати уявлення про весь теоретичний матеріал;
- знати у повному обсязі не менш половини тем теоретичного матеріалу;
- знати додатковий матеріал за двома темами.

Необхідний обсяг умінь для одержання позитивної оцінки:

- володіти технологією застосування теоретичних знань на практиці;
- вміти самостійно знаходити довідкові матеріали;
- вміти застосовувати отриманні знання практично.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Знати основні положення теоретичного матеріалу. Знати базові положення щодо призначення та способів використання моделей інтелектуального аналізу даних в різних сферах життя.

Добре (75-89). Знати основний теоретичний матеріал в повному обсязі. Володіти технологією пошуку довідкової літератури. Вміти вирішувати базові задачі інтелектуального аналізу даних.

Відмінно (90-100). Знати основний і додатковий теоретичний матеріал в повному обсязі. Орієнтуватися в довідковій літературі. Вміти вирішувати задачі вибору методів інтелектуального аналізу даних для рішення практичних задач, вміти будувати власні сценарії

проведення інтелектуального аналізу даних, проводити повноцінний аналіз отриманих результатів та проводити оцінку ступеню достовірності отриманих результатів.

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

Увесь науково методичний комплект з дисципліни розміщено на освітньому порталі Classroom <https://classroom.google.com/u/0/c/NzM4ODM5MDQxNTA5>

14. Рекомендована література

1. Бойко І.В., М.Р. Петрик, Г.Б. Цуприк. Інформаційні технології видобутку даних (Data mining, високопродуктивні обчислення у складних системах): навчальний посібник. Тернопіль: : ТНТУ 2020 – 62 с
2. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.
3. Лупан І.В. Інтелектуальний аналіз даних Data Mining: навчально-методичний посібник. – Кропивницький, ФОП Піскова М. А., 2022. – 112 с.
4. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
5. Чанишев Р. І. Інформаційні технології та Data mining в соціальних науках: навчально-методичний посібник для підготовки здобувачів вищої освіти галузі знань 05 «Соціальні та поведінкові науки» [Електронний ресурс]. / Р. І. Чанишев, М. А. Яценко ; Нац. ун-т «Одеська юрид. академія». – Одеса: Фенікс, 2024. – 706 с. – Режим доступу: <https://doi.org/10.32837/11300.27676>

15. Інформаційні ресурси

1. Eibe Frank, Mark A. Hall, and Ian H. Witten. The WEKA Workbench – [Електронний ресурс]. – Режим доступу: https://ml.cms.waikato.ac.nz/weka/Witten_et_al_2016_appendix.pdf
2. Гавриленко, О. В. Навчальний посібник з дисциплін “Аналіз даних” та “Аналіз даних в управляючих системах” для студентів спеціальності 126 - Інформаційні системи та технології / О. В. Гавриленко ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2020. – 85 с. [Електронний ресурс]. – Режим доступу: <https://ela.kpi.ua/handle/123456789/38757>

ДОДАТОК

Перелік галузей знань, спеціальностей та освітніх програм, за якими відбувається підготовка здобувачів в університеті

Галузі знань: 02 Культура і мистецтво, 03 Гуманітарні науки, 05 Соціальні та поведінкові науки, 07 Управління та адміністрування, 08 Право, 10 Природничі науки, 11 Математика та статистика, 12 Інформаційні технології, 13 Механічна інженерія, 14 Електрична інженерія, 15 Автоматизація та приладобудування, 16 Хімічна та біоінженерія, 17 Електроніка та телекомунікації, 19 Архітектура та будівництво, 26 Цивільна безпека, 27 Транспорт, 28 Публічне управління та адміністрування, 29 Міжнародні відносини

Спеціальності: 029 Інформаційна, бібліотечна та архівна справа, 035 Філологія, 051 Економіка, 053 Психологія, 071 Облік і оподаткування, 072 Фінанси, банківська справа та страхування, 073 Менеджмент, 075 Маркетинг, 076 Підприємництво, торгівля та біржова діяльність, 081 Право, 101 Екологія, 103 Науки про Землю, 113 Прикладна математика, 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія, 124 Системний аналіз, 125 Кібербезпека, 126 Інформаційні системи та технології, 131 Прикладна механіка, 133 Галузеве машинобудування, 134 Авіаційна та ракетно-космічна техніка, 141 Електроенергетика, електротехніка та електромеханіка, 142 Енергетичне машинобудування, 144 Теплоенергетика, 151 Автоматизація та комп'ютерно-інтегровані технології, 152 Метрологія та інформаційно-вимірвальна техніка, 153 Мікро- та наносистемна техніка, 163 Біомедична інженерія, 172 Телекомунікації та радіотехніка, 173 Авіоніка, 193 Геодезія та землеустрій, 272 Авіаційний транспорт, 262 Правоохоронна діяльність, 274 Автомобільний транспорт, 281 Публічне управління та адміністрування, 292 Міжнародні економічні відносини

Освітні програми: Інформаційна, бібліотечна та архівна справа, Прикладна лінгвістика, Економіка підприємства, Психологія, Облік і оподаткування, Фінанси, банківська справа та страхування, Менеджмент, Логістика, Управління проектами, Маркетинг, Підприємництво, торгівля та біржова діяльність, Право, Екологія та охорона навколишнього середовища, Космічний моніторинг Землі, Обчислювальний інтелект, Інженерія програмного забезпечення, Інформаційні технології проектування, Комп'ютеризація обробки інформації та управління, Інтелектуальні системи та технології, Комп'ютерні технології в біології та медицині, Комп'ютерні системи та мережі, Системне програмування, Системний аналіз і управління, Безпека інформаційних і комунікаційних систем, Інформаційні системи та технології підтримки віртуальних середовищ, Розподілені інформаційні системи, Штучний інтелект та інформаційні системи, Динаміка і міцність машин, Роботомеханічні системи і логістичні комплекси, Комп'ютерний інжиніринг, Проектування, виробництво та сертифікація авіаційної техніки, Авіаційні двигуни та енергетичні установки, Ракетно-космічна техніка, Інтелектуальні безпілотні транспортні засоби, Комп'ютерно-інтегроване управління в енергетиці, Нетрадиційні та відновлювані джерела енергії, Газотурбінні установки і компресорні станції, Енергетичний менеджмент, Інженерія мобільних додатків, Комп'ютерні системи технічного зору, Комп'ютерні технології проектування та виробництва, Комп'ютерно-інтегровані технологічні процеси і виробництва, Інтелектуальні інформаційні вимірвальні системи, Якість, стандартизація та сертифікація, Мікро- та наносистемна техніка, Біомедична інженерія, Інформаційні мережі зв'язку, Радіоелектронні пристрої, системи та комплекси, Системи автономної навігації та адаптивного управління літальних апаратів, Геоінформаційні системи і технології, Правоохоронна діяльність, Інтелектуальні транспортні системи, Технічне обслуговування та ремонт повітряних суден і авіадвигунів, Автомобілі та автомобільне господарство, Публічне управління та адміністрування, Міжнародна економіка