

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

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

ЗАТВЕРДЖУЮ

Керівник проектної групи

 А. Г. Чухрай
(підпис) (ініціали та прізвище)

« 30 » 08 2023 р.

РОБОЧА ПРОГРАМА ОBOB'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: **другий (магістерський)**

Харків 2023 рік

Робоча програма Інтелектуальний аналіз даних
(назва дисципліни)
для студентів за спеціальністю 122 "Комп'ютерні науки"
освітньою програмою "Інтелектуальні системи та технології"

«27» серпня 2023 р. – 8 с.

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


(підпис)

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

Протокол № 1 від « 30 » серпня 2023 р.

Завідувач кафедри д.т.н., проф.
(науковий ступінь та вчене звання)


(підпис)

А. Г. 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)	Цикл професійної підготовки (обов'язковий)
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: другий (магістерський)	Навчальний рік
Кількість змістовних модулів – 2		2023/2024
Індивідуальне завдання Використання методів Data Mining для аналізу		Семестр
Загальна кількість годин – 64/150		10-й
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 5,375		Лекції*
	32 години	
	Практичні, семінарські	
	0 годин	
	Лабораторні	
	32 години	
Самостійна робота		
86 годин		
Вид контролю		
модульний контроль, іспит		

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: для денної форми навчання: 1-й семестр – 64/86.

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

2. Мета та завдання навчальної дисципліни

Мета: поглиблення знань та ефективне використання сучасних методів інтелектуального аналізу даних та обробки інформації з використанням програмного забезпечення.

Завдання: оволодіння сучасною методологією та особливостями застосування інтелектуальної обробки даних; засвоєння майбутніми фахівцями теоретичних основ інформаційних систем, орієнтованих на застосування стандартів інтелектуального аналізу даних; отримання практичних навичок використання і адаптації деяких найбільш відомих систем та бібліотек для реалізації інтелектуального аналізу даних.

ЗК1 – Здатність до абстрактного мислення, аналізу та синтезу

ЗК2 – Здатність застосовувати знання у практичних ситуаціях.

ЗК5 – Здатність вчитися й оволодівати сучасними знаннями.

СК3.Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК4.Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

СК6.Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

3. Програма навчальної дисципліни

Змістовий модуль 1

Тема 1. Визначення, основні задачі інтелектуального аналізу даних (ІАД).

Інтелектуальний аналіз даних. Визначення Data Mining. Методи і задачі Data Mining. Зв'язки з іншими областями знань. Класифікація стадій Data Mining. Поняття Data Science. Аналіз даних та інтелектуальний аналіз даних. Приклади рішень, продуктів які створені за допомогою Data Science. Основні області застосування Data Science. Використання Machine learning.

Тема 2. Препроцесінг інформації

Цикл одержання попередньої обробки аналізу даних, інтерпретації результатів та їхнього використання. Ентропія і кількість інформації. Нормалізація і стандартизація вихідних значень. Алгоритм «вибілювання» входів. Виключення неінформативних факторів. Аналітико-евристичні алгоритми визначення вагомих інформативних ознак. Метод Шенона. Метод Кульбака-Лейблера.

Змістовий модуль 2

Тема 3. Алгоритми Data Mining: задача класифікації

Завдання Data Mining: класифікація задач Data Mining, завдання класифікації і регресії, завдання пошуку асоціативних правил. Постановка задачі класифікації. Представлення результатів: правила класифікації, дерева рішень, математичні функції. Методи опорних векторів, «найближчого сусіда», алгоритм побудови 1-правил, метод Naive Bayes. Аналіз багатомірних угруповань. Класифікація об'єктів у випадку невідомих розподілень даних. Методи оцінювання помилок класифікації. Оцінка параметрів логістичної регресії на основі методу оцінки шансів і ймовірностей. Оцінка ймовірності за допомогою байєсовського класифікатора (Naive Bayes).

Тема 4. Алгоритми Data Mining: задача кластеризації

Базові алгоритми кластеризації. Адаптивні методи кластеризації. Постановка завдання кластеризації: формальна постановка задачі, заходи близькості, засновані на відстанях, які використовуються в алгоритмах кластеризації. Представлення результатів кластеризації. Види кластерів. Базові алгоритми кластеризації: класифікація алгоритмів, ієрархічні алгоритми, неієрархічні алгоритми. Адаптивні методи кластеризації: вибір найкращого рішення і якість кластеризації, використання формальних критеріїв якості в адаптивній кластеризації, приклад адаптивної кластеризації.

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.						
Тема 1. <i>Визначення, основні задачі ІАД</i>	22	2	–	–	–	20
Тема 2. <i>Препроцесінг інформації</i>	46	8	12	–	–	24
Модульний контроль	2	2	–	–	–	-
Разом за змістовим модулем 1	70	12	12	0	0	44
Модуль 2						
Змістовий модуль 2.						
Тема 3. <i>Алгоритми Data Mining: задача класифікації</i>	40	10	10	–	–	20
Тема 4. <i>Алгоритми Data Mining: задача кластеризації</i>	36	8	10	–	–	20
Модульний контроль	2	2	–	–	–	–
Індивідуальне завдання – розрахункова робота	2	–	–	–	–	2
Разом за змістовим модулем 2	80	20	20	0	0	42
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	Використання методу Шенона для оцінки інформативності ознак	4
2	Використання методу Кульбака-Лейблера для оцінки інформативності ознак	4
3	Скорочення розмірності ознакового простору	4
4	Задача класифікації. Оцінка параметрів логістичної регресії на основі методу оцінки шансів і ймовірностей	4
5	Задача класифікації. Оцінка ймовірності за допомогою байєсовського класифікатора (Naive Bayes)	4
6	Базові алгоритми кластеризації	4
	Разом	32

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Використання Machine learning	20
2	Алгоритм «вибілювання» входів. Аналітико-евристичні алгоритми визначення вагомих інформативних ознак. Методи опорних векторів	24
3	Методи опорних векторів. Аналіз багатомірних угруповань. Методи оцінювання помилок класифікації.	20
4	Адаптивні методи кластеризації	20
5	Виконання розрахункової роботи	2
	Разом	86

9. Індивідуальні завдання

№ з/п	Назва теми
1	Використання методів Data Mining для аналізу даних заданої предметної галузі

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами.

11. Методи контролю

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

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	3	12..15
Модульний контроль	10..25	1	10..25
Змістовний модуль 2			
Виконання і захист лабораторних робіт	3..5	3	12..15
Виконання і захист РР	16..20	1	16..20
Модульний контроль	10..25	1	10..25
Усього за семестр			60...100

Під час складання іспиту 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 методів інтелектуального аналізу даних; методи препроцесінгу інформації; основні поняття вибіркового методу; методи перевірки статистичних гіпотез; основні області застосування Data Science та Data Mining; цикл одержання попередньої обробки аналізу даних, інтерпретації результатів та їхнього використання; методи оцінки інформативності; алгоритми класифікації даних; алгоритми кластеризації даних.

Необхідний обсяг вмінь для одержання позитивної оцінки:

Використовувати методи Шенона та Кульбака-Лейблера для оцінки інформативності ознак; знижувати розмірність ознакового простору; проводити оцінку параметрів логістичної регресії на основі методу оцінки шансів і ймовірностей; вирішувати задачу класифікації за допомогою байєсовського класифікатора; вирішувати задачу кластеризації; представляти змістовну інтерпретацію результатів інтелектуального аналізу даних.

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.

Добре (75-89). Твердо знати мінімум, захистити всі індивідуальні завдання, виконати всі модульні контрольні роботи.

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

Методичне забезпечення:

- робоча програма дисципліни;
- конспект лекцій;
- методичні вказівки та рекомендації для виконання робіт та проектів, розрахункових та розрахунково-графічних робіт, лабораторних та практичних робіт, а також рекомендації для самостійної підготовки.

14. Рекомендована література

Базова

1. Інтелектуальний аналіз даних : навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя : ЗНТУ, 2012. – 278 с.
2. Ian H. Witten. Data Mining: Practical Machine Learning Tools and Techniques / Ian H. Witten, Eibe Frank and Mark A. Hall — [3rd Edition] — Morgan Kaufmann, 2011. — P. 664. — ISBN 9780123748560.
3. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.
4. Encyclopedia of artificial intelligence / Eds.: J. R. Dopico, J. D. de la Calle, A. P. Sierra. – New York : Information Science Reference, 2009. – Vol. 1–3. – 1677 p.
5. Ситник В.Ф. Засоби дейтамайнінгу для аналізу бізнесових рішень. Науково-практичний журнал "Науково-технічна інформація", №3, 2002. - с. 60-64.
6. Ситник В.Ф. Системи підтримки прийняття рішень: Навч.посібник / В. Ф. Ситник - К.:КНЕУ, 2004. – 614 с.

Допоміжна

- 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.

15. Інформаційні ресурси

1. Сайт кафедри математичного моделювання та обчислювального інтелекту.
2. Язык програмування з документацією, Python 3.x., [Електронний ресурс]. URL: www.python.org