

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

Кафедра інформаційних технологій проектування (№ 105)

ЗАТВЕРДЖУЮ

Голова НМК 2

(підпис)

Д.М. Крицький
(ініціали та прізвище)

«31» 05 2021 р.

РОБОЧА ПРОГРАМА ОBOB'ЯЗКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Системи і технології інтелектуального аналізу даних
(назва навчальної дисципліни)

Галузь знань 12 «Інформаційні технології»
(шифр і найменування галузі знань)

Спеціальність: 126 «Інформаційні системи та технології»

Освітні програми: «Інформаційні системи та технології підтримки віртуальних середовищ»

(найменування спеціалізації)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2021 рік

Робоча програма «Системи і технології інтелектуального аналізу даних»

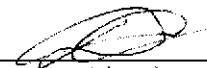
(назва дисципліни)

для студентів спеціальності 126 «Інформаційні системи та технології»,
спеціалізації «Інформаційні системи та технології підтримки віртуальних середовищ»

«28» травня 2021 р., – 10 с.

Розробник: Крицький Д.М., завідувач кафедри, к.т.н., доцент

(прізвище та ініціали, посада, науковий ступінь та вчене звання)


(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційних технологій проектування

(назва кафедри)

Протокол № 1 від «31» 08 2021 р.

Завідувач кафедри к.т.н., доцент

(науковий ступінь та вчене звання)


(підпис)

Д.М. Крицький

(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки (спеціальність, спеціалізація), рівень вищої освіти	Характеристика навчальної дисципліни	
		Денна форма навчання	Заочна форма навчання
Кількість кредитів – 4	Галузь знань <u>12 «Інформаційні технології»</u> (шифр та найменування) Спеціальність <u>126 «Інформаційні системи та технології»,</u> (код та найменування) Освітні програми <u>«Інформаційні системи та технології підтримки віртуальних середовищ»</u> (найменування) Рівень вищої освіти: перший (бакалаврський)	Цикл професійної підготовки (за вибором)	
Кількість модулів – 2		Навчальний рік	
Кількість змістових модулів – 6		2021/ 2022	
Індивідуальне завдання РГР <u>Проведення аналізу даних з використанням методів Data Mining</u> (назва)		Семестр	
Загальна кількість годин – 120		6-й	-
		Лекції¹⁾	
		28 годин	-
Кількість тижневих годин для денної форми навчання: аудиторних – 2,47 самостійної роботи студента – 3,84		Практичні, семінарські	
		0 годин	-
		Лабораторні¹⁾	
		19 годин	-
		Самостійна робота	
		73 годин	-
		Вид контролю	
		Іспит	-

Співвідношення кількості годин ауд. занять до сам. роботи становить для денної форми $47/73=0,64$.

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 бакалаврів 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.

3. Програма навчальної дисципліни

МОДУЛЬ 1

Змістовий модуль 1. Основи інтелектуального аналізу даних.

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.

Змістовний модуль 2. OLAP і Data Mining.

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.

Змістовний модуль 3. Методи пошуку асоціативних правил.

Тема 5. 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 6. Метод Apriori.

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

Змістовний модуль 4. Методи кластеризації.

Тема 7. Кластерний аналіз.

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

Тема 8. Сітка Кохонена.

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

МОДУЛЬ 2.

Змістовний модуль 5. Методи класифікації та регресії.

Тема 9. 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 10. 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.

Змістовний модуль 6. Ансамблі моделей та порівняння моделей.**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									
	Денна форма					Заочна форма				
	Усього	У тому числі				Усього	У тому числі			
		л	п	лаб.	с. р.		л	п	лаб.	с. р.
1	2	3	4	5	6	7	8	9	10	11
Модуль 1										
Змістовий модуль 1. Основи інтелектуального аналізу даних										
Тема 1. Задачі інтелектуального аналізу даних	12	2	-	2	2	-	-	-	-	-
Тема 2. Сховища даних.	14	2	-	2	4	-	-	-	-	-
Змістовний модуль 2. OLAP і Data Mining										
Тема 3. Упровадження Data Mining, OLAP і сховищ даних у СППР.	10	2	-	-	4	-	-	-	-	-
Тема 4. OLAP-системи.	12	2	-	2	2	-	-	-	-	-
Змістовний модуль 3. Методи пошуку асоціативних правил.										
Тема 5. Пошук асоціативних правил	8	2	-	-	2	-	-	-	-	-
Тема 6. Метод Apriori.	12	2	-	2	2	-	-	-	-	-
Змістовний модуль 4. Методи кластеризації.										
Тема 7. Кластерний аналіз.	8	2	-	2	2	-	-	-	-	-
Тема 8. Сітка Кохонена.	10	2	-	2	4	-	-	-	-	-
Разом за модулем 1.	84	16	-	12	18	-	-	-	-	-
Модуль 2										
Змістовний модуль 5. Методи класифікації та регресії.										
Тема 9. Методи класифікації	14	4	-	2	2	-	-	-	-	-
Тема 10. Статистичний аналіз.	8	4	-	2	2	-	-	-	-	-
Змістовний модуль 6. Ансамблі моделей та порівняння моделей..										
Тема 11. Види побудови та аналізу	10	2	-	3	2	-	-	-	-	-

ансамблів моделей										
Тема 12. Методи оцінки моделей.	6	2	-	-	4	-	-	-	-	-
Індивідуальне завдання	30	-	-	-	30					
Разом за модулем 2	78	12	-	7	40					
Усього годин	162	28	-	19	58	-	-	-	-	-

5. Теми семінарських занять

Семінарські заняття навчальним планом не передбачені.

6. Теми практичних занять

Практичні заняття навчальним планом не передбачені

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Багатовимірний аналіз даних. Робота з OLAP-кубами	2
2	Побудова та робота з фізичним сховищем даних.	2
3	Пошук асоціативних правил. Метод Apriori	2
4	Методи кластеризації: k-means, G-means.	2
5	Кластеризація за алгоритмом Кохонена	2
6	Класифікація за методом Байєса.	2
7	Побудова дерева рішень	2
8	Статистичний аналіз даних. Ансамбль моделей	2
9	Ансамбль моделей	3
	Разом	19

8. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		Денна форма навчання	Заочна форма навчання
1	Види СППР. Задачі інтелектуального аналізу даних	2	-
2	Види сховищ даних. Види даних в сховищах даних.	2	-
3	Побудова сховищ даних та проведення аналізу.	2	-
4	Робота з OLAP-системами. Переваги та недоліки OLAP-систем з позицій аналізу даних.	2	-
5	Задача пошуку асоціативних правил.	2	-
6	Метод Apriori та його різновиди.	2	-
7	Методи кластерного аналізу k-means та G-means, процедура Мак-Кіна, метод C-means.	2	-
8	Сіті Кохонена, що самоорганізуються. Карта Кохонена. Підготовка до модульної контрольної роботи №1	4	-
9	Методи класифікації. Класифікація за допомогою	2	

	дерева рішень.		
10	Методи статичного аналізу даних	2	-
11	Методи побудови ансамблю моделей	2	-
12	Методи оцінки моделей. Підготовка до модульної контрольної роботи №2.	4	
	Разом	28	-

9. Індивідуальне завдання

Зміст: проведення аналізу даних з використанням методів Data Mining. Форма звітності – подача пояснювальної записки та демонстрація роботи алгоритму аналізу даних з використанням інформаційних засобів.

Варіанти завдань – алгоритми пошуку асоціативних правил, побудови та аналізу OLAP-кубів, методи кластеризації, методи регресійного аналізу, методи класифікації, методи побудови часових рядів.

Обсяг звітних документів – 15-25 сторінок.

Тижні 8 – 14. Трудомісткість: 30 годин самостійної роботи.

План-графік виконання ІДЗ:

№	Найменування розділу	Обсяг, %	Тиждень здачі	Кількість сторінок ПЗ	Трудомісткість	
					аудиторн.	самостійн.
1	Постановка задачі	10	8	1 – 2	-	2
2	Формалізований опис задачі	10	8	1 – 2	-	2
3	Алгоритм виконання аналізу даних	25	9	3 – 6	-	12
4	Проведення аналізу даних	30	10	4 – 5	-	6
5	Аналіз алгоритму и методу аналізу даних	10	11	4 – 5	-	4
6	Виводи щодо проведеного дослідження та рекомендація щодо прийняття рішення по отриманим результатам	10	12	1 – 3	-	2
7	Перелік посилань	5	12	1-2		2
Разом		100		15 – 25	-	30

10. Методи навчання

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

Самостійна робота включає підготовку до лабораторних робіт, модульного контролю та заліку, виконання позааудиторної частини розрахунково-графічної роботи і вивчення вказаних вище тем за конспектом, літературними джерелами та програмною документацією.

11. Методи контролю

Контроль здійснюється згідно з “Положенням про модульно-рейтингову систему оцінювання знань студентів”.

Поточний контроль – відповідно до повноти, якості та своєчасності виконання лабораторних робіт; проміжний (модульний) контроль – письмові контрольні роботи на 8-му, та 17-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			
Виконання та захист лабораторних (практичних) робіт	3-5	4	10-18*
Модульний контроль	1-2	10	10-20
Модуль 2			
Робота на лекціях			
Виконання та захист лабораторних (практичних) робіт	3-5	5	14-22*
Модульний контроль	4-10	10	10-20
Виконання та захист РГР (РР, РК)	12-20	1	16 - 20
Всього за семестр			60...100

*Приведення балів за лабораторні роботи до загальної суми 40 балів за весь курс лабораторних робіт проводиться з використанням коефіцієнту приведення.

Семестровий контроль (іспит/залік) проводиться у разі відмови студента від балів поточного тестування та за наявності допуску до іспиту/заліку. При складанні семестрового іспиту/заліку 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 – 2,5 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

1. Завдання та сценарії виконання лабораторних робіт у електронній формі та допоміжні приклади надаються студентам на сервері кафедри.

14. Рекомендована література

14.1. Базова

1. Черняк О.І. Інтелектуальний аналіз даних: підручник. – К: Знання, 2014. – 599с.
2. Шумейко А. А. Интеллектуальный анализ данных (Введение в Data Mining): учеб. пособ. / А. А. Шумейко, С. Л. Сотник. – Днепропетровск: Белая Е.А., 2012. – 212 с.
3. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. – Запоріжжя: КПУ, 2011. – 268 с.

14.2. Допоміжна

1. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.

15. Інформаційні ресурси

1. BaseGroup Labs: Глоссарий. – [Электронный ресурс]. – Режим доступа: <https://basegroup.ru/community/glossary>.
2. BaseGroup Labs: Руководство по алгоритмам Deductor 5.2. – [Электронный ресурс]. – Режим доступа: <https://basegroup.ru/deductor/manual/guide-algorithm-520>.
3. BaseGroup Labs: Руководство аналитика. – [Электронный ресурс]. – Режим доступа: https://basegroup.ru/system/files/documentation/guide_analyst_5.2.0.pdf.