

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
БУДІВНИЦТВА І АРХІТЕКТУРИ

**БАКАЛАВР**

Кафедра інформаційних технологій

**«ЗАТВЕРДЖУЮ»**

Декан факультету автоматизації і  
інформаційних технологій



/ І.В. Русан /

« 04 » червня 2021 року

**НАВЧАЛЬНА РОБОЧА ПРОГРАМА ДИСЦИПЛІНИ**

«Інтелектуальний аналіз даних»

(назва навчальної дисципліни)

шифр	назва спеціальності
122	Комп'ютерні науки

Розробник:

Єрукаєв А.В., кандидат технічних наук

(прізвище та ініціали, науковий ступінь, звання)

(підпис)

Робоча програма затверджена на засіданні кафедри інформаційних технологій

протокол № 17 від " 18 " травня 2021 року

Завідувач кафедри

інформаційних технологій

(підпис)

/ Цюцюра С.В. /

(прізвище та ініціали)

Схвалено навчально-методичною комісією спеціальності (НМКС):  
"Комп'ютерні науки"

протокол № 6 від " 4 " червня 2021 року

Голова НМКС

(підпис)

/ Терентьев О.О. /

(прізвище та ініціали)

**ВИТЯГ З НАВЧАЛЬНОГО ПЛАНУ 2021-2022 рр.**

Шифр	ОР, бакалавр	Форма навчання:										Форма контролю	Семестр	Відмітка про погодження
	Назва спеціальності (спеціалізації)	Кредитів на сем.	Обсяг годин					Кількість індивідуальних робіт						
			Всього	Аудиторних										
				Разом	У тому числі			КП	КР	РГР	Роб			
Л	Лр	Пз												
122	Комп'ютерні науки	<b>4,0</b>	<b>120</b>	<b>40</b>	<b>20</b>	<b>20</b>				<b>1</b>		<b>Зал</b>	<b>5</b>	

### Мета та завдання навчальної дисципліни

Мета дисципліни – активно закріпити, узагальнити, поглибити й розширити знання, отримані при застосуванні моделей та методів інтелектуального аналізу даних при проектуванні та розробці інформаційних систем. Завдання дисципліни полягають у формуванні навичок автоматичного пошуку закономірностей в масивах інформації.

#### Компетенції студентів, що формуються в результаті засвоєння дисципліни

Код	Зміст	Програмні результати навчання
<b>Інтегральна компетентність</b>		
<b>ІК</b>	Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі комп'ютерних наук або у процесі навчання, що передбачає застосування теорій та методів інформаційних технологій і характеризується комплексністю та невизначеністю умов.	
<b>Загальні компетентності</b>		
<b>ЗК1</b>	Здатність до абстрактного мислення, аналізу та синтезу.	<p><b>ПР4.</b> Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.</p> <p><b>ПР12.</b> Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.</p>
<b>ЗК2</b>	Здатність застосовувати знання у практичних ситуаціях.	<p><b>ПР4.</b> Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.</p> <p><b>ПР12.</b> Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.</p>
<b>ЗК7</b>	Здатність до пошуку, оброблення та аналізу інформації з різних джерел.	<b>ПР1.</b> Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.
<b>Спеціальні (фахові) компетентності</b>		
<b>СК2</b>	Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.	<p><b>ПР1.</b> Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.</p> <p><b>ПР3.</b> Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.</p> <p><b>ПР4.</b> Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.</p>
<b>СК11</b>	СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.	<p><b>ПР1.</b> Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.</p> <p><b>ПР12.</b> Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining.</p>

**Програма навчальної дисципліни  
Змістовний модуль 1. Аналіз даних.**

**Тема лекційного заняття 1. Вступ.**

"Корисні" дані. Ознаки, прецеденти. Типи даних. Бізнес-дані. Data Mining.

**Тема лекційного заняття 2. Дослідження даних.**

Звіт про якість даних. Неперервні, категоріальні ознаки. Потужність, процентиль, частота, доля.

**Тема лекційного заняття 3. Візуалізація даних.**

Гістограма. Групування даних. Типи гістограм. Проблеми з якістю даних. IQR.

**Тема лекційного заняття 4. Поглиблене дослідження даних.**

Діаграма розсіювання. Маленькі діаграми. Стекові діаграми.

**Тема лекційного заняття 5. Обчислення коваріації та кореляції.**

Вибіркова коваріація. Кореляція. Матриці візуалізації.

**Змістовний модуль 2. Класифікація методом ID3.**

**Тема лекційного заняття 6. Підготування даних.**

Нормалізація. Групування. Вибіркові методи.

**Тема лекційного заняття 7. Гра "Вгадай, хто".**

Стратегії обробки. Послідовності питань. Інформативні ознаки.

**Тема лекційного заняття 8. Дерева рішень.**

Генерування передбачень. Склад та використання дерев рішень.

**Тема лекційного заняття 9. Модель ентропії Шенона.**

Ентропія множини. Двійковий логарифм. Результати ймовірностей.

**Тема лекційного заняття 10. Приріст інформації.**

Міра інформативності. Процес обчислення. Значення інформативності.

**Індивідуальна робота**

**Теми лабораторних занять**

№ з/п	Назва теми
1	Формування звіту про якість даних за обраною темою.
2	Опис даних на основі гістограм, стовпчикових діаграм та блочних діаграм.
3	Дослідження даних за матрицею діаграм розсіювання.
4	Побудова ID3 за категоріальними та неперервними ознаками.
5	Усунення шуму та вибіркової дисперсії в ID3.

**Самостійна робота**

№ з/п	Назва теми
1	Вступ.
2	Дослідження даних.
3	Візуалізація даних.
4	Поглиблене дослідження даних.
5	Обчислення коваріації та кореляції.
6	Підготування даних.
7	Гра "Вгадай, хто".
8	Дерева рішень.
9	Модель ентропії Шенона.
10	Приріст інформації.
11	Алгоритм ітеративного дихотомізатора 3.
12	Неперервні ознаки в ID3.

**Методи контролю та оцінювання знань студентів**

Поточне оцінювання (кількість балів)		Сума
Змістовий модуль	Змістовий модуль	

№1	№2	
45	55	100

### Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	<b>A</b>	відмінно	зараховано
82-89	<b>B</b>	добре	
74-81	<b>C</b>		
64-73	<b>D</b>	задовільно	
60-63	<b>E</b>		
35-59	<b>FX</b>	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	<b>F</b>	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

### Рекомендована література

1. Інтелектуальний аналіз даних: практикум [ Комплект] / Фісун М.Т., Кравець І.О., Казмірчук П.П., Ніколенко С.Г. – Л.: «Новий Світ – 2000», 2020. – 162 с.
2. Data Mining - пошук знань в даних / Анатолій Гладун. - К.: ТОВ "ВД "АДЕФ\_Україна"", 2016. - С. 452

### Інформаційні ресурси

<http://library.knuba.edu.ua>