

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ЗАТВЕРДЖУЮ  
В.о. декана факультету комп'ютерних  
інформаційних технологій

Ігор ЯКИМЕНКО

« 31 » 2023 р.

ЗАТВЕРДЖУЮ  
В.о. проєктора з науково-  
педагогічної роботи

Віктор ОСТРОВЕРХОВ

2023 р.

РОБОЧА ПРОГРАМА

з дисципліни «Виявлення та обробка аномальних даних»

Ступінь вищої освіти - бакалавр

Галузь знань – 12 Інформаційні технології

Спеціальність – 122 Комп'ютерні науки

Освітньо-професійна програма «Штучний інтелект»

Кафедра прикладної математики

Форма навчання	Курс	Семестр	Лекції, (год.)	Практичні заняття, (год.)	ІРС, (год.)	Тренінг, КІПЗ, (год.)	Самоост. робота студ., (год.)	Разом, (год.)	Залік, (сем.)
Денна	1	6	28	14	3	6	99	150	6

*Dr. M. P. B.*

Робочу програму склав доцент кафедри прикладної математики, канд. фіз.-мат. наук Андрій АЛІЛУЙКО

Робоча програма затверджена на засіданні кафедри прикладної математики, протокол № 1 від 28.08.2023 р.

Завідувач кафедри



Олеся МАРТИНЮК

Розглянуто та схвалено групою забезпечення спеціальності 122 Комп'ютерна наука, протокол № 1 від 30.08 2023 р.

Голова групи  
забезпечення спеціальності



Мирослав КОМАР

Гарант ОІП



Василь КОВАЛЬ

**СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
«Виявлення та обробка аномальних даних»**

**1. Опис дисципліни «Виявлення та обробка аномальних даних»**

<b>Дисципліна «Виявлення та обробка аномальних даних»</b>	<b>Галузь знань, спеціальність, освітньо-професійна програма, ступінь вищої освіти</b>	<b>Характеристика навчальної дисципліни</b>
Кількість кредитів ECTS – 5	Галузь знань – 12 Інформаційні технології	<b>Статус дисципліни:</b> вибіркова <b>Мова навчання:</b> українська
Кількість залікових модулів – 3	Спеціальність – 122 Комп’ютерні науки	Рік підготовки: 3 Семестр: <i>денна</i> – 6
Кількість змістових модулів – 2	Освітньо-професійна програма – Штучний інтелект	Лекції: <i>денна</i> – 28 год <i>заочна</i> – 8 год Лабораторні заняття: <i>денна</i> – 14 год.
Загальна кількість годин – 150	Ступінь вищої освіти – бакалавр	Самостійна робота: <i>денна</i> – 99 год Тренінг, КППЗ: <i>денна</i> – 6 год Індивідуальна робота: <i>денна</i> – 3 год
Тижневих годин – 10 год, з них аудиторних – 3 год.		Вид підсумкового контролю – залік

**2. Мета і завдання вивчення дисципліни «Виявлення та обробка аномальних даних»**

**2.1. Мета вивчення дисципліни**

Метою викладання дисципліни є формування системи теоретичних знань і практичних навичок застосування сучасних методів виявлення та обробки аномалій в наборах даних великих об’ємів.

**2.2. Завдання вивчення дисципліни**

Завдання вивчення дисципліни «Виявлення та обробка аномальних даних»: ознайомити студентів із сучасними методами пошуку та опрацювання аномалій даних; поглибити та розширити знання у сфері роботи з даними великих об’ємів; здобути практичні навички побудови та реалізації алгоритмів пошуку аномалій за допомогою середовища Python.

**2.3. Результати навчання**

В результаті вивчення навчальної дисципліни студент повинен:

**знати:**

- теоретичні основи пошуку аномалій;
- методи, алгоритми пошуку аномалій, обробки даних з використанням технологій машинного навчання;
- сучасні програмні засоби реалізації технологій обробки даних та пошуку аномалій в них.

**вміти:**

- обґрунтовано підбирати методи та алгоритми для пошуку аномалій при опрацюванні конкретних наборів даних;
- здійснювати оцінювання якості реалізованих алгоритмів та їх оптимізації;
- вирішувати прикладні задачі щодо оптимізації та аналізу даних, прогнозування із врахуванням виявлених аномалій даних;
- використовувати сучасні програмні засоби для реалізації технологій обробки даних та пошуку аномалій в них.

### **3. Програма навчальної дисципліни «Виявлення та обробка аномальних даних»**

#### *Змістовий модуль 1. Методи математичної статистики для виявлення аномалій*

##### **Тема 1. Вступ в виявлення аномальних даних.**

Поняття аномалій даних. Класифікація аномалій. Класифікація методів виявлення аномалій.

##### **Тема 2. Ймовірнісні та геометричні методи виявлення аномалій.**

Статистичні методи виявлення аномалій. Ймовірнісні моделі виявлення аномалій. Аналіз екстремальних значень. Виявлення аномалії на основі кута. Глибинні техніки виявлення аномалій

##### **Тема 3. Виявлення аномалій в часових рядах.**

Поняття аномалій в даних часових рядів. Статистичне управління процесом для виявлення аномалій. Авторегресійні моделі.

#### *Змістовий модуль 2. Машинне навчання в виявленні та обробці даних*

##### **Тема 4. Лінійні методи для виявлення аномалій.**

Лінійна регресія. Аналіз головних компонент (PCA). Однокласові опорні векторні машини (one-class SVMs).

##### **Тема 5. Виявлення аномалій методами на основі близькості.**

Метод k-найближчих сусідів (KNN) при виявленні аномалій. Метод кластеризації k-середніх при виявленні аномалій. Метод локального коефіцієнта викиду (LOF) при виявленні аномалій.

##### **Тема 6. Виявлення аномалій в даних великої розмірності.**

Особливості виявлення аномалії в даних великої розмірності. Метод підпростору з пакетуванням ознак для виявлення аномалій у багатовимірних наборах даних. Метод ізольованого лісу для виявлення аномалій у багатовимірних наборах даних (Isolation Forest).

##### **Тема 7. Контрольовані методи виявлення аномалій.**

Контрольоване виявлення аномалій. Економічне навчання. Адаптивна повторна вибірка.

##### **Тема 8. Оцінка методів виявлення аномалій.**

Метричний аналіз методів виявлення аномалій. Аналіз виявлених аномалій різних типів даних.

##### **Тема 9. Глибинне навчання для виявлення аномалій.**

Архітектура автокодувальника. Застосування автокодувальника для аналізу аномалій.

#### 4. Структура залікових кредитів з дисципліни «Виявлення та обробка аномальних даних»

##### Структура I залікового кредиту

Тема	Кількість годин					
	Лекції	Лабораторні заняття	ІРС	Тренінг, КПЗ	СРС	Контрольні заходи
<b>Змістовий модуль 1. Методи математичної статистики для виявлення аномалій</b>						
Тема 1. Вступ в виявлення аномальних даних	2				1	Поточне опитування
Тема 2. Ймовірнісні та геометричні методи виявлення аномалій	2	2	2	5	2	
Тема 3. Виявлення аномалій в часових рядах	4	2			3	
<b>Змістовий модуль 2. Машинне навчання в виявленні та обробці даних</b>						
Тема 4. Лінійні методи виявлення аномалій	4	2			1	Поточне опитування
Тема 5. Виявлення аномалій методами на основі близькості	4	2			4	
Тема 6. Виявлення аномалій в даних великої розмірності	4	2			1	
Тема 7. Контрольовані методи виявлення аномалій	4	2			7	
Тема 8. Оцінка методів виявлення аномалій	2	2			1	
Тема 9. Глибинне навчання для виявлення аномалій	2				1	
Разом	28	14	<b>3</b>	<b>6</b>	99	

## 5. Тематика лабораторних занять

### Лабораторна робота 1

**Тема.** Ймовірнісні та геометричні методи виявлення аномалій

**Мета:** Виробити навички виявлення аномалій даних статистичними, ймовірнісними та геометричними методами.

**Питання для обговорення:**

1. Статистичні тести: Z-оцінка, модифікована Z-оцінка, тест Грабса, відстань Махаланобіса.
2. Аналіз екстремальних значень: блокові максимуми (block maxima), пік над порогом (peak over threshold).
3. Виявлення аномалії на основі кута (Angle-Based Anomaly Detection).
4. Глибинна техніка виявлення аномалій (Depth-based techniques).

### Лабораторна робота 2

**Тема.** Виявлення аномалій в часових рядах

**Мета:** Навчити застосовувати числові характеристики числових рядів для виявлення аномалій. Виробити навички побудови ARIMA моделей для вивчення особливостей часових рядів.

**Питання для обговорення:**

1. Статистичне управління процесом для виявлення аномалій.
2. ARIMA моделювання для пошуку незвичної поведінки в даних часових рядів.

### Лабораторна робота 3

**Тема.** Лінійні методи виявлення аномалій

**Мета:** Набути практичний досвід роботи з моделями лінійної регресії, з PCA, з однокласними SVM при виявленні аномалій.

**Питання для обговорення:**

1. Лінійна регресія. Аналіз головних компонент (PCA).
2. Однокласові опорні векторні машини (one-class SVMs).
3. Порівняння результативності застосування методів.

### Лабораторна робота 4

**Тема.** Виявлення аномалій методами на основі близькості

**Мета:** Навчитися застосовувати методи KNN, K-Means, LOF для виявлення аномалій.

**Питання для обговорення:**

1. Метод k-найближчих сусідів (KNN) при виявленні аномалій.
2. Метод кластеризації k-середніх (K-Means) при виявленні аномалій.
3. Метод локального коефіцієнта викиду (LOF) при виявленні аномалій.

### Лабораторна робота 5

**Тема.** Виявлення аномалій в даних великої розмірності

**Мета:** Розуміти проблеми виявлення аномалій в даних великої розмірності. Володіти метод підпростору з пакетуванням ознак та методом ізольований ліс (Isolation Forest) для виявлення аномальних даних.

**Питання для обговорення:**

1. Метод підпростору з пакетуванням ознак для виявлення аномалій у багатовимірних наборах даних.
2. Метод ізольованого лісу для виявлення аномалій у багатовимірних наборах даних (Isolation Forest).

## **Лабораторна робота 6**

### **Тема.** Контрольовані методи виявлення аномалій

**Мета:** Начитися застосовувати методи контрольованого виявлення аномалій. Отримати практичний досвід проведення економічного навчання, адаптивної повторної вибірки та методів посилення при виявленні аномалій.

#### **Питання для обговорення:**

1. Контрольоване виявлення аномалій.
2. Економічне навчання (Cost-sensitive learning).
3. Адаптивна повторна вибірка (adaptive resampling).
4. Методи підсилення (boosting methods).

## **Лабораторна робота 7**

### **Тема.** Оцінка методів виявлення аномалій

**Мета:** Навчити оцінювати методи виявлення аномалій.

#### **Питання для обговорення:**

1. Оцінка виявлених аномалій.
2. Показники точності виявлення аномалій.

## 6. Комплексне практичне індивідуальне завдання

Для набуття практичних навичок у студентів особливе значення має виконання комплексного практичного індивідуального завдання (КПЗ) з дисципліни «Виявлення та обробка аномальних даних» виконується самостійно кожним студентом. Студент самостійно вибирає набори даних (одновимірні, багатовимірні, часовий ряд). Набори даних можна завантажити із сайтів: <https://www.kaggle.com/>; <https://www.timeseriesclassification.com/dataset.php>, або інших сайтів. Здійснює підготовку даних і проводить пошук аномалій та їх опрацювання. КПЗ оформлюється згідно з встановленими вимогами.

## 7. Тренінг з дисципліни

Мета тренінгу з дисципліни «Виявлення та обробка аномальних даних» – оволодіння навичками застосування методів обробки та пошуку аномальних даних різних галузей.

Проведення тренінгу дозволяє: забезпечити практичне засвоєння теоретичних знань, отриманих у процесі вивчення дисципліни «Виявлення та обробка аномальних даних»;

№ з/п	Вид роботи	Порядок проведення тренінгу
1	Вступна частина	Ознайомлення студентів з темою тренінгового заняття і видача завдання
2	Практична частина	Виконання завдань студентами згідно з індивідуальним завданням; оформлення короткого звіту
3	Підведення підсумків	Презентація та оцінювання виконаних завдань

## 8. Самостійна робота студентів

З метою засвоєння дисципліни «Виявлення та обробка аномальних даних» студенти повинні володіти значним обсягом інформації, частину якої вони отримують і опрацьовують шляхом самостійної роботи.

№ п/п	Тематика	Кількість годин	
		ДФН	ЗФН
1	Вступ в виявлення аномальних даних	11	14
2	Ймовірнісні та геометричні методи виявлення аномалій	11	15
3	Виявлення аномалій в часових рядах	11	15
4	Лінійні методи виявлення аномалій	11	15
5	Виявлення аномалій методами на основі близькості	11	15
6	Виявлення аномалій в даних великої розмірності	11	16
7	Контрольовані методи виявлення аномалій	11	16
8	Оцінка методів виявлення аномалій	11	16
9	Глибинне навчання для виявлення аномалій	11	16
<b>Разом:</b>		99	138

## 9. Методи навчання

У навчальному процесі застосовуються: лекції, практичні та індивідуальні заняття, консультації, самостійна робота, метод опитування, тестування, виконання КПЗ.

## 10. Засоби оцінювання та методи демонстрування результатів навчання

У процесі вивчення дисципліни «Виявлення та обробка аномальних даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне опитування та тестування;
- оцінювання результатів модульної контрольної роботи;
- оцінювання результатів ректорської контрольної роботи;
- оцінювання результатів КПЗ;
- залік.



## 11. Політика оцінювання

*Політика щодо дедлайнів і перескладання.* Для виконання індивідуальних завдань і проведення контрольних заходів встановлюються конкретні терміни. Перескладання модулів відбувається з дозволу дирекції факультету за наявності поважних причин (наприклад, лікарняний).

*Політика щодо академічної доброчесності.* Використання друкованих і електронних джерел інформації під час контрольних заходів та екзаменів заборонено.

*Політика щодо відвідування.* Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, карантин, воєнний стан, хвороба, закордонне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу з дозволу дирекції факультету.

## 12. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Виявлення та обробка аномальних даних» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

### Для заліку

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Разом (%)
30%	40%	30%	100
Виконання та захист лабораторних робіт (4 роботи по 10 балів) – макс. 40 балів Модульна контрольна робота – макс. 60 балів	Виконання та захист лабораторних робіт (3 роботи по 10 балів) – макс. 30 балів Модульна контрольна робота – макс. 70 балів	Виконання завдань під час тренінгу – макс. 20 балів Виконання та захист КПІЗ – макс. 80 балів	100

### Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90–100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75–84		C (добре)
65–74	задовільно	D (задовільно)
60-64		E (достатньо)
35–59	незадовільно	FX (незадовільно з можливістю повторного складання)
1–34		F (незадовільно з обов'язковим повторним курсом)

## 13. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

№	Найменування	Номер теми
1.	Комунікаційне програмне забезпечення (Zoom) для проведення занять у режимі онлайн (за необхідності)	1–9
2.	Комунікаційна навчальна платформа (Moodle) для організації дистанційного навчання (за необхідності)	1–9
3.	Python	1–9
4.	Мультимедійне обладнання	1–9

## РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Ane Blázquez-García, Angel Conde, Usue Mori, and Jose A Lozano. A review on outlier/anomaly detection in time series data. *ACM Computing Surveys (CSUR)* 54, 2021. 3. P. 1–33.
2. Anomaly Detection. Intel course – URL: <https://www.intel.com/content/www/us/en/developer/learn/course-anomaly-detection.html>
3. Charu C. Aggarwal, *Outlier Analysis*, Second Edition, Springer International Publishing AG 2017. P. 465.
4. Julien Lesouple, Cédric Baudoin, Marc Spigai, Jean-Yves Tournet. Generalized isolation forest for anomaly detection. *Pattern Recognition Letters*. 2021. Vol. 149. P. 109-119.
5. Mohammad Braei and Sebastian Wagner. Anomaly detection in univariate time-series: A survey on the state-of-the-art. arXiv preprint arXiv:2004.00433, 2020.
6. Paul Boniol, Michele Linardi, Federico Roncallo, Themis Palpanas, Mohammed Meftah, and Emmanuel Remy. Unsupervised and scalable subsequence anomaly detection in large data series. *The VLDB Journal*. 2021. 30, 6, P. 909–931.
7. Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407, 2019.
8. Xiaodan Xu, Huawen Liu, and Minghai Yao. Recent Progress of Anomaly Detection, *Hindawi Complexity*. 2019. Article ID 2686378, 11 pages – URL: <https://doi.org/10.1155/2019/2686378>
9. Гавриленко С.Ю., Зозуля В.Д. Дослідження методів виявлення аномалій на етапі попередньої обробки даних. *Системи управління, навігації та зв'язку*. 2022. Вип. 1 (67). с. 52-56.
10. Завгородній В.В. Завгородня Г.А. Валявська Н.О. Герасименко О.О. Калюжний О.В. Степовий А.В. Пошук аномалій у даних за допомогою машинного навчання. *Вчені записки ТНУ імені В.І. Вернадського*. Серія: Технічні науки. 2022. Т. 33 (72). № 3 с. 39-43.