



Силабус курсу

Виявлення та обробка аномальних даних

Ступінь вищої освіти – бакалавр
Галузь знань 12 «Інформаційні технології»
Спеціальність 122 «Комп'ютерні науки»
Освітньо-професійна програма: «Штучний інтелект»

Рік навчання: III, Семестр: VI

Кредитів: 5 Мова викладання: українська

Керівник курсу

ППП

к.фіз.-мат.н., доцент Андрій Алілуйко

Контактна інформація

aliluyko82@gmail.com, +38 0972784899

Опис дисципліни

Мета викладання даної дисципліни – сформувати систему знань студентів в області виявлення та опрацювання аномалій в наборах даних великих об'ємів. В дисципліні основний акцент робиться на розумінні вибору та застосування окремих моделей виявлення аномалій залежно від особливостей наборів даних та їх аномалій.

В дисципліні основна увага приділяється задачам виявлення аномалій із використанням різних алгоритмів.

Структура курсу

Години (лек./лабор.)	Тема	Результати навчання	Завдання
2/-	1. Вступ в виявлення аномальних даних	Розуміти поняттям аномалій та їх видів. Знати основні статистики та математичні поняття для виявлення та опрацювання аномальних даних. Класифікувати методи дослідження аномалій даних та їх опрацювання	Питання
2/4	2. Ймовірнісні та геометричні методи виявлення аномалій	Володіти статистичними та імовірнісними методами виявлення аномалій. Здійснювати аналіз екстремальних значень. Застосовувати методи виявлення аномалії на основі кута та глибинні техніки виявлення аномалій	Питання, лабораторна робота
4/4	3. Виявлення аномалій в часових рядах	Вміти виявляти аномалії часових рядів. Здійснювати статистичне управління процесом для виявлення аномалій. Застосовувати авторегресійні моделі при аналізі аномалій	Питання, лабораторна робота
4/4	4. Лінійні методи для виявлення аномалій	Вміти застосовувати лінійну регресію, метод головних компонент (PCA), метод однокласових опорних векторів (one-class SVMs) для виявлення аномалій	Питання, лабораторна робота
4/2	5. Виявлення аномалій	Володіти теоретичними знаннями про	Питання,

	методами на основі близькості	методи близькості даних. Вміти застосовувати метод k-найближчих сусідів (KNN), метод кластеризації k-середніх, метод локального коефіцієнта викиду (LOF) при виявленні аномалій	лабораторна робота
4/4	6. Виявлення аномалій в даних великої розмірності	Розуміти проблеми виявлення аномалій за допомогою даних великої розмірності. Володіти методом ізольований ліс (Isolation Forest) та методом підпростору з пакетуванням ознак для виявлення аномалій у багатовимірних наборах даних	Питання, лабораторна робота
2/4	7. Контрольовані методи виявлення аномалій	Вміти здійснювати контрольоване виявлення аномалій методами економічного навчання та адаптивної повторна вибірки	Питання, лабораторна робота
4/4	8. Оцінка методів виявлення аномалій	Вміти здійснювати метричний аналіз методів виявлення аномалій. Здійснювати аналіз виявлених аномалій різних типів даних	Питання, лабораторна робота
4/4	9. Глибинне навчання при виявленні аномалій	Застосовувати штучні нейронні мережі при виявленні аномалії. Розуміти архітектуру автокодувальника. Знати принципи роботи автокодувальника для аналізу аномалій	Питання, лабораторна робота

Літературні джерела

Основна література

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. Julien Lesouple, Cédric Baudoin, Marc Spigai, Jean-Yves Tourneret. Generalized isolation forest for anomaly detection. *Pattern Recognition Letters*. 2021. Vol. 149. P. 109-119.
3. 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.
4. Документація по пакету PYOD – URL: <https://pyod.readthedocs.io/en/latest>
5. 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.
6. 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.

Додаткова література

7. 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.
8. Anomaly Detection. Intel course – URL: <https://www.intel.com/content/www/us/en/developer/learn/course-anomaly-detection.html>

Політика оцінювання

Політика щодо дедлайнів і перескладання. Для виконання завдань і проведення контрольних заходів встановлюються конкретні терміни. Перескладання модулів є обов'язковим.

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

Політика щодо відвідування. Оцінювання студентів здійснюється обов'язково з усіх тем дисципліни. Відвідування занять, при необхідності консультацій, робота на дистанційній платформі Moodle є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, карантин, військовий стан, хвороба, закордонне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу з дозволу дирекції факультету.

Оцінювання

Модуль 1		Модуль 2		Модуль 3	Модуль 4
20%	20%	20%	20%	5%	15%
Поточне оцінювання	Модульний контроль 1	Поточне оцінювання	Модульний контроль 2	Тренінги	Самостійна робота
Визначається як середнє арифметичне з оцінок, отриманих за виконання та захист лабораторних робіт. Опитування проводиться з тем 1-5	Тестові завдання (10 тестів по 2 бали за тест) – макс. 20 балів Задача 1 – макс. 40 балів Задача 2 – макс. 40 балів	Визначається як середнє арифметичне з оцінок, отриманих під час занять. Опитування проводиться з тем 6-9	Тестові завдання (10 тестів по 2 бали за тест) – макс. 20 балів Задача 1 – макс. 40 балів Задача 2 – макс. 40 балів	Визначається як середнє арифметичне з оцінок, отриманих за виконання завдань на тренінгу	Визначається як середнє арифметичне з оцінок, отриманих за виконання завдань самостійної роботи

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)