

**Силабус курсу
Інтелектуальний
аналіз даних**

**Освітньо-професійна програма: 124 «Системний аналіз»
Ступінь вищої освіти – перший (бакалаврський)
Спеціальність: 124 «Системний аналіз»**

Рік навчання: I, Семестр: 2

Кількість кредитів: 5 Мова викладання: українська

Керівник курсу

ППП

к.е.н., доцент **Пасічник Роман Мирославович**

**Контактна
інформація**

roman.pasi@gmail.com

Опис дисципліни

У зв'язку з удосконаленням технологій запису та зберігання даних людство отримало колосальні потоки інформаційної “руди” в усіляких областях. Діяльність будь-якого підприємства (комерційного, виробничого, медичного, наукового тощо) тепер супроводжується реєстрацією та записом усіх подробиць його діяльності. Став очевидним той факт, що без продуктивної переробки, потоки сирих даних нікому непотрібні. Специфіка сучасних вимог до такої переробки є такою:

- дані мають необмежений обсяг;
- дані є різномірними (кількісними, якісними, текстовими);
- результати повинні бути конкретні та зрозумілі;
- інструменти для обробки сирих даних повинні бути прості у використанні.

Алгоритми традиційної математичної статистики тривалий час, як основні, підтримували концепцію усереднення з вибірки, що зводиться до операцій над фіктивними величинами (типу середньої температури аудиторій в усіх приміщеннях університету, середньої висоти будинку міста, що складається з палаців і халуп тощо). Методи математичної статистики виявилися корисними переважно для перевірки заздалегідь сформульованих гіпотез (verification-driven data mining) і для “грубого” розвідницького аналізу,

що становить основу оперативної аналітичної обробки даних (online analytical processing, OLAP).

В основу інтелектуального аналізу даних (discovery-driven data mining, Data Mining, ІАД) покладена концепція шаблонів (паттернів), що відбивають фрагменти багатоаспектних взаємин у даних. Ці шаблони являють собою закономірності, властиві підвибіркам даних, які можуть бути компактно виражені у зрозумілій людині формі. Пошук шаблонів проводиться методами, необмеженими апріорними припущеннями про структуру вибірки, та видами розподілів значень аналізованих показників. Важливе положення ІАД – нетривіальність розшукуваних шаблонів. Це означає, що знайдені шаблони повинні відбивати неочевидні, несподівані (unexpected) регулярності в даних, що становлять так звані приховані знання (hidden knowledge). До суспільства прийшло розуміння, що сирідані (raw data) містять глибинний шар знань, за грамотного “розкопування” якого можуть бути виявлені справжні “самородки”.

Отже, інтелектуальний аналіз даних – це процес виявлення в сирих даних раніше невідомих, нетривіальних, фактично корисних і доступних інтерпретації знань, необхідних для прийняття рішень у різних сферах людської діяльності.

Метою курсу «Інтелектуальний аналіз даних» є оволодіння методами сучасної обробки даних – інтелектуального аналізу даних (Data Mining, Knowledge Discovery in Data), аналітичного дослідження великих масивів інформації з метою виявлення нових раніше невідомих, практично корисних знань і закономірностей, необхідних для прийняття рішень; огляд методів, програмних продуктів і різних інструментальних засобів, які використовуються Data Mining; розгляд практичних прикладів застосування Data Mining; підготовка студентів до самостійної роботи з вирішення задач засобами Data Mining і розробки інтелектуальних систем.

Data Mining – мультидисциплінарна область, яка виникла і розвивається на базі таких наук як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та ін.

Результати навчання:

В результаті вивчення курсу "Інтелектуальний аналіз даних" студенти повинні:

- знати основні поняття лінійного та квадратичного аналізу вибірок, машин опорних векторів, стохастичного градієнтного спуску, методу найближчих сусідів, сумішей нормальних розподілів, наївних Байєсівських класифікаторів, дерев рішень, ансамблевих методів, напівконтрольованого навчання;

- вміти здійснювати лінійний та квадратичний аналіз вибірок, реалізовувати кластеризацію вибірок за допомогою машин опорних векторів, стохастичного градієнтного спуску, найближчих сусідів, наївних Байєсівських класифікаторів, дерев рішень, будувати прогнози на основі сумішей нормальних розподілів, ансамблевих методів, напівконтрольованого навчання та ізотонічної регресії.

4. Перелік тем

1. Основи інтелектуального аналізу даних..
2. Процес виявлення знань.
3. Вирішення задачі класифікації.
4. Вирішення задачі регресії.
5. Методи аналізу часових рядів.
6. Вирішення задачі кластеризації.
7. Вирішення задачі пошуку асоціативних правил.
8. Сховища даних.
9. Оперативний аналіз даних.
10. Генетичні алгоритми.

6. Рекомендовані джерела інформації

1. Актуальні проблеми Data Mining : навчальний посібник для студентів факультету комп'ютерних наук та кібернетики / О. О. Марченко, Т. В. Россада. К. : КНУ ім. Т. Шевченка, 2018. – 150 с.
2. Інтелектуальний аналіз даних : Комп'ютерний практикум : навч. посібник / О. О. Сергеев-Горчинський, Г. В. Іщенко. – К. : КПІ ім. Ігоря Сікорського, 2018. – 73 с.
3. Kantarzic M. Data Mining. Concepts, Models, Methods and Algorithms / M. Kantarzic, 3rd Ed. – Publisher : Wiley, 2019. – 672 p.
4. A. Agresti. Statistical methods for social sciences. Boston: Pearson, 2018.
5. Аналіз даних та знань : навчальний посібник / Литвин В. В., Пасічник В. В., Нікольський Ю. В. – Львів : Магнолія-2006 , 2021. – 276 с.

Політика оцінювання

- **Політика щодо дедлайнів та перескладання:** Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку (-20 балів). Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).
- **Політика щодо академічної доброчесності:** Усі письмові роботи перевіряються на наявність плагіату і допускаються до захисту із коректними текстовими запозиченнями не більше 20%. Списування під

час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних девайсів). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування (наприклад, програма Kahoot).

- **Політика щодо відвідування:** Відвідування занять є обов'язковим компонентом оцінювання, за яке нараховуються бали. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу.

Оцінювання Система оцінювання та вимоги

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен	Разом
20%	20%	20%	40%	100%
1. Усне опитування під час заняття (5 теми по 10 балів = 50 балів) 2. Письмова робота = 50 балів	1. Усне опитування під час заняття (5 тем по 10 балів = 50 балів) 2. Письмова робота = 50 балів	1. Написання та захист КПЗ = 60 балів. 3. Виконання завдань під час тренінгу = 40 балів	1. 3 запитання по 20 балів = 60 балів 2. Задача = 40 балів	

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)

9. Політики курсу

Академічна доброчесність. Дотримання академічної доброчесності студентами передбачає:

- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);

- посилення на джерела інформації у разі використання ідей, розробок, тверджень, відомостей;

- дотримання норм законодавства про авторське право і суміжні права;

- надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використанні методики досліджень і джерела інформації.

Порушенням академічної доброчесності вважається:

академічний плагіат - оприлюднення (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства;

самоплагіат - оприлюднення (частково або повністю) власних раніше опублікованих наукових результатів як нових наукових результатів;

фабрикація - вигадкування даних чи фактів, що використовуються в освітньому процесі або наукових дослідженнях;

фальсифікація - свідомо зміна чи модифікація вже наявних даних, що стосуються освітнього процесу чи наукових досліджень;

списування - виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання, зокрема під час оцінювання результатів навчання.

За порушення академічної доброчесності здобувачі освіти можуть бути притягнені до такої академічної відповідальності:

- повторне проходження оцінювання (контрольна робота, іспит, залік тощо);
- повторне проходження відповідного освітнього компонента освітньої програми.

Політика запізнення. За несвоєчасно виконані завдання буде накладено штраф 10 відсотків від загальної кількості балів за це завдання. Примітка. Виключення можуть бути зроблені до невчасно зданих завдань з поважних причин.

Політика щодо відвідування. Відвідування занять є обов'язковим. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватися в он-лайн формі за погодженням із керівником курсу.