



Силабус дисципліни «Інтелектуальні технології аналізу та попередньої обробки даних»



Викладач: *Калініна Ірина Олександрівна*
канд. техн. наук, доцент, доцент кафедри інтелектуальних
інформаційних систем

Очікувані результати навчання

В результаті вивчення дисципліни студент

має знати:

- концептуальні та методологічні знання у сфері аналізу даних для прийняття оптимальних рішень, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень в галузі аналізу даних на рівні останніх світових досягнень з комп'ютерної інженерії, IT-інфраструктур, інформаційних технологій;
- сучасні методи проведення досліджень у сфері прикладного аналізу даних для прийняття рішень в різних галузях науки шляхом вивчення теоретичних та практичних положень побудови поведінкових моделей функціональних залежностей та використання результатів аналізу даних для уточнення оптимальних параметрів прийняття рішень;
- наукові і математичні положення, що лежать в основі методів аналізу даних для отримання оптимальних параметрів прийняття рішень, методів побудови та дослідження математичних моделей та технологій адаптивних та інтелектуальних обчислень при аналізі даних.

має вміти:

- ефективно здійснювати пошук та критичний аналіз інформації з різних джерел щодо методів та технологій аналізу даних;

Обсяг дисципліни: 5,5 кредити ECTS (18 годин лекцій, 36 годин практичних робіт).

Мета: підготовка фахівців, здатних розв'язувати комплексні проблеми в галузі дослідницької діяльності у сфері аналізу даних для прийняття оптимальних рішень та використання результатів аналізу даних для уточнення наукових висновків та формування прогнозів щодо прийнятих рішень, що передбачає глибоке розуміння підходів до створення традиційних та створення нових поведінкових моделей.

Оригінальність навчальної дисципліни:

Авторський курс

Зміст дисципліни

Тема 1. Загальне поняття про аналіз даних та сфери застосування задач прикладного аналізу. Порівняння методів машинного навчання та інтелектуального аналізу даних. Класифікація задач машинного навчання. Задачі ймовірнісно-статистичного аналізу та попередньої обробки даних в задачах машинного навчання.

Тема 2. Загальне поняття про прикладні задачі аналізу даних. Задачі класифікації. Задача оцінювання позичальників. Задачі регресійної оцінки. Задача прогнозування споживчого попиту. Задачі кластеризації. Задача рубрикації текстів. Задача виділення термінів. Задача соціологічного опитування. Задачі пошуку асоціацій.

Тема 3. Постановка задачі аналізу даних. Основні визначення. Об'єкти та ознаки. Типи задач аналізу даних. Поняття про алгоритм аналізу даних. Означення методу навчання. Означення та способи формування функції втрат. Функціонал якості. Мінімізація емпіричного ризику. Cross-володація. Leave-one-out cross-володація.

Тема 4. Методи ідентифікації та заповнення пропусків в даних.

Тема 5. Методи і алгоритми виявлення та обробки екстремальних значень. Методи ідентифікації викидів.

- розв’язувати задачі синтезу та аналізу об’єктів дослідження при аналізі даних для прийняття оптимальних рішень;
- створювати та реалізовувати поведінкові математичні моделі опису функціональних залежностей при аналізі даних;
- застосовувати прикладні методи аналізу даних технології адаптивних та інтелектуальних обчислень для передбачення майбутніх станів досліджуваних об’єктів та процесів.
- застосовувати прикладні бібліотеки та програмні системи, які використовуються при машинному аналізі даних;
- володіти методами та технологіями програмування з використанням прикладних бібліотек та програмних систем, призначених для машинного аналізу реальних наборів даних.

Пререквізити

Відповідно до освітньої програми необхідно попередньо оволодіти знаннями з дисциплін: «Вища математика», «Основи програмування», «Об’єктно-орієнтовне програмування», «Теорія ймовірностей та математична статистика», «Алгоритми та структури даних», «Системний аналіз», «Методи та системи штучного інтелекту».

Пореквізити

Компетентності, знання та вміння, отримані в рамках вивчення даної дисципліни, можуть бути застосовані для отримання обґрунтованих результатів досліджень та підвищення наукового рівня кваліфікаційної роботи.

Семестровий контроль: іспит

Оцінювання:

За семестр: 60 балів

За іспит: 40 балів

Види робіт:

Практичні роботи

Технічне забезпечення

Тема 6. Методи і алгоритми згладжування даних.

Тема 7. Системне використання методів фільтрації даних в задачах машинного навчання. Цифрова фільтрація. Оптимальна фільтрація. Ймовірнісна фільтрація.

Тема 8. Системний підхід до нормалізації та стандартизації даних в завданнях машинного навчання. Лінійні методи нормалізації. Нелінійні методи нормалізації.

Тема 9. Інформаційна технологія ймовірнісно-статистичного аналізу та попередньої обробки даних.

Критерії оцінювання лабораторних/практичних/індивідуальних/робіт/ доповідей/проєктів

Максимальна кількість балів – студент з високою якістю самостійно виконав весь обсяг робіт, відповідає на всі питання, пов’язані з виконаними роботами, та робить додаткові розрахунки, які йому пропонує викладач. У викладача немає претензій щодо реалізації та вимог до виконання роботи.

70%-99% від максимальної кількості балів – студент з достатньою якістю виконав всі завдання, але в процесі роботи він робив деякі помилки, які, після вказування на них викладачем, самостійно виправляв. На деякі питання він відповідає з похибкою. Запропоновані викладачем додаткові розрахунки робить з деякою потугою. Не всі вимоги до виконання роботи дотримані.

40%-69% від максимальної кількості балів – студент самостійно виконав всі роботи, але якість реалізації недостатня (помилки при розрахунках, не всі вимоги до роботи дотримані). На питання щодо виконання робіт відповідає не зовсім чітко. Є помилки при відповідях.

1%-39% від максимальної кількості балів – студент самостійно виконав не всі роботи, при цьому якість реалізації недостатня (помилки при розрахунках, не дотримується вимог до оформлення роботи). На питання щодо виконання робіт відповідає не чітко. Є грубі помилки при відповідях.

0 балів – студент не виконав весь обсяг робіт, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам.

Для зарахування практичної роботи здобувач має змогу надати сертифікат (або інший документ) про проходження курсів на платформах *Prometheus*, *Coursera* тощо та за умови співпадіння направленості курсів блоку освітньої компоненти.

Заняття з дисципліни проводяться у дистанційному форматі (*GoogleMeet, Zoom* тощо) або у комп'ютерних класах з використанням мов програмування та програмних середовищ: R, Python, RStudio.

Політика щодо дедлайнів

Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.

Політика щодо академічної доброчесності

Передбачає самостійне виконання лабораторних робіт. Списування під час іспиту (в т. ч. із використанням мобільних пристроїв) заборонено. У разі виявлення плагіату або списування роботи не зараховуються.