



ПРИКЛАДНІ ЗАДАЧІ МАШИННОГО НАВЧАННЯ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 Інформаційні технології
Спеціальність	122 Комп'ютерні науки
Освітня програма	Комп'ютерний моніторинг та геометричне моделювання процесів і систем
Статус дисципліни	Обов'язкова
Форма навчання	очна(денна)
Рік підготовки, семестр	4 курс осінній семестр
Обсяг дисципліни	4 кредити (120 год) 36 лек. 18 лаб. 66 СРС
Семестровий контроль/ контрольні заходи	Залік, МКР
Розклад занять	Науково-педагогічний працівник
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д.ф.-м.н., с.н.с., Матичин Іван Іванович, i.matychyn@kpi.ua, тел. 050-872-32-17 Практичні: д.ф.-м.н., с.н.с., Матичин Іван Іванович, i.matychyn@kpi.ua, тел. 050-872-32-17
Розміщення курсу	Кампус

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

В основі машинного навчання (англ. Machine Learning, ML) лежать алгоритми, що перетворюють інформацію в практично цінні знання. Саме тому машинне навчання є важливим в сучасну епоху великих даних. Без нього було б неможливо відслідковувати величезний потік інформації.

Курс поєднує практичні приклади з базовою теорією, що дає можливість застосовувати моделі та методи машинного навчання до прикладних задач аналізу даних.

Метою дисципліни є

- Формувати знання та навички щодо створення математичних моделей аналізу даних, використовуючи методи та алгоритми машинного навчання.
- Навчити студентів застосовувати математичні методи та програмні засоби для виявлення прихованих закономірностей у великих масивах даних.
- Сформувати базовий рівень математичних знань фахівців в галузі енергетики, для розв'язання різних задач у професійній діяльності: задач класифікації, кластеризації, регресії, виявлення аномалій.
- Розвивати інтелект, здібності до аналітичного та логічного мислення, навички самостійної роботи з науковою літературою.

Завдання. В результаті вивчення дисципліни у студентів повинні сформуватися наступні компетентності:

загальні:

- здатність до абстрактного мислення, аналізу та синтезу (ЗК 1),
- здатність проводити дослідження на відповідному рівні (ЗК 3),

фахові:

- 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 (ФК 2).

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі програмні результати навчання:

- 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. (ПР4).

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Пререквізити дисципліни. Знання та вміння, отримані на попередньому рівні освіти при вивченні дисциплін математичної підготовки, комп'ютерної дискретної математики, методів та систем штучного інтелекту.

Постреквізити дисципліни. Отримані при вивченні дисципліни «Прикладні задачі машинного навчання» знання формують базові знання для вивчення дисциплін, пов'язаних з моделюванням та проектуванням інформаційних систем.

3. Зміст навчальної дисципліни

Тема 1. Вступ до машинного навчання

Тема 2. Управління даними та їх інтерпретація

Тема 3. Класифікація за допомогою методу k найближчих сусідів

Тема 4. Класифікація за допомогою наївного баєсівського класифікатора

Тема 5. Класифікація за допомогою дерев рішень

Тема 6. Регресійні методи

Тема 7. Нейронні мережі та метод опорних векторів

Тема 8. Аналіз споживацького кошика за допомогою асоціативних правил

Тема 9. Кластеризація методом k середніх

Тема 10. Оцінка ефективності моделі

4. Навчальні матеріали та ресурси

Основна література

1. Кононова К. Ю. Машинне навчання: методи та моделі / К. Ю. Кононова. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с.
2. Lantz B. Machine learning with R / B. Lantz. – Birmingham: Packt Publishing, 2015. – 396 p.

Додаткова література

3. Raschka S. Python Machine Learning / S. Raschka, V. Mirjalili. – Birmingham: Packt Publishing, 2019. – 770 p.

5. Методика опанування навчальної дисципліни (освітнього компонента)

Тема 1. Вступ до машинного навчання

Лекція 1. Вступ до машинного навчання

Історія виникнення машинного навчання. Типи вхідних даних. Типи алгоритмів машинного навчання.

Лекція 2. Види машинного навчання

Навчання під наглядом. Навчання без нагляду. Навчання з підкріпленням. Різниця між контрольованим та неконтрольованим навчанням.

Тема 2. Управління даними та їх інтерпретація

Лекція 3. Машинне навчання з використанням R

Основи мови програмування R. Встановлення і завантаження R-пакетів. Встановлення та робота з RStudio.

Лекція 4. Структури даних в R

Вектори. Фактори. Списки. Фрейми даних. Матриці і масиви. Імпорт та збереження даних.

Тема 3. Класифікація за допомогою методу k найближчих сусідів

Лекція 5. Метод k-NN.

Алгоритм методу k-NN. Збір та підготовка даних. Навчання моделі на даних. Оцінка ефективності моделі. Підвищення ефективності моделі.

Тема 4. Класифікація за допомогою наївного баєсівського класифікатора

Лекція 6. Наївний баєсівський класифікатор.

Теоретичні основи та теорема Баєса. Основні поняття баєсівських методів. Наївний баєсівський алгоритм. Приклад.

Тема 5. Класифікація за допомогою дерев рішень

Лекція 7. Дерева рішень.

Алгоритм дерева рішень C5.0. Приклад. Алгоритм 1R. Алгоритм RIPPER. Правила, побудовані на основі дерев рішень. Приклад.

Лекція 8. Випадкові ліси

Ансамблі. Бутстреп. Беггінг. Похибка Out-of-bag. Випадковий ліс. Порівняння з деревом рішень і беггінгом.

Тема 6. Регресійні методи

Лекція 9. Лінійна регресія

Поняття регресії. Метод найменших квадратів. Теорема Гауса-Маркова. Декомпозиція похибки. Коефіцієнт детермінації.

Лекція 10. Логістична регресія

Лінійний класифікатор. Логістична регресія як лінійний класифікатор. Принцип максимальної вірогідності і логістична регресія. L2-регуляризація логістичної функції втрат.

Тема 7. Нейронні мережі та метод опорних векторів

Лекція 11. Основи нейронних мереж

Біологічний нейрон та перцептрон. Багатошарові мережі прямого розповсюдження. Навчання методом оберненого розповсюдження (backpropagation). Функції активації. Функції втрат. Гіперпараметри.

Лекція 12. Глибокі мережі

Глибоке навчання (deep learning). Загальні архітектурні принципи глибоких мереж. «Будівельні блоки» глибоких мереж.

Лекція 13. Основні архітектури глибоких мереж

Згорткові нейронні мережі (CNN). Рекурентні нейронні мережі. Рекурсивні нейронні мережі.

Лекція 14. Вибір глибокої мережі для вирішення проблеми

Табличні дані і багатошарові перцептрони. Зображення і згорткові нейронні мережі. Послідовності, часові ряди і рекурентні нейронні мережі.

Тема 8. Аналіз споживацького кошика за допомогою асоціативних правил

Лекція 15. Метод асоціативних правил

Алгоритм Apriori для пошуку асоціативних правил. Вимірювання цікавості правила: підтримка і довіра. Побудова набору правил за принципом Apriori. Приклад: аналіз споживацького кошика.

Тема 9. Кластеризація методом k середніх

Лекція 16. Кластерний аналіз

Типи вхідних даних. Цілі кластеризації. Функції відстані. Методи кластеризації.

Лекція 17. Метод k середніх

Алгоритм методу k середніх. Вибір значення k. Переваги та недоліки методу k середніх.

Приклад.

Тема 10. Оцінка ефективності моделі

Лекція 18. Оцінка ефективності моделі.

Вимірювання ефективності класифікації. Прогнози класифікатора. Аналіз матриць невідповідностей. Візуалізація компромісів ефективності за допомогою ROC-кривих. Метод відкладених даних.

6. Самостійна робота студента

Розв'язування всіх лабораторних робіт для власних наборів даних та аналіз отриманих результатів.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Відвідування лекційних та практичних занять є обов'язковим за винятком поважних причин (хвороби, форс-мажорних обставин).

В разі пропущення занять з поважних причин викладач надає можливість студенту виконати усі або деякі завдання практичних занять (винятком є виконання деяких завдань у зв'язку із закінченням навчального процесу).

В разі пропущення занять без поважних причин, а також через порушення граничного терміну виконання завдання (deadline) студент може отримати зменшену кількість балів від максимальної оцінки за відповідне завдання.

Протягом семестру студенти:

- виконують та захищають завдання практичних занять (комп'ютерні практикуми) у відповідні терміни,

- пишуть модульну контрольну роботу,

- повинні позитивно закрити дві атестації (в кінці березня та в середині травня),

- по закінченні навчального процесу складають екзамен.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Система рейтингових (вагових) балів та критерії оцінювання

Максимальна кількість балів з кредитного модуля дорівнює 100.

Рейтинг студента з дисципліни складається з балів, що він отримує за:

- виконання та захист лабораторних робіт,
- дві модульні контрольні роботи (МКР) тривалістю 1 акад. година кожна.

1. Виконання завдань практичних занять (комп'ютерні практикуми)

Завдання практичного заняття являє собою індивідуальне виконання лабораторних робіт, що пов'язані з рішенням на ЕОМ заданої задачі комп'ютерного моделювання.

Вагові бали завдань наведено у таблиці.

<i>Види завдань</i>	<i>Внесок до семестрового рейтингу балів</i>
Лабораторна робота №0. Знайомство з середовищем R	-
Лабораторна робота №1. Регресійний аналіз	10
Лабораторна робота №2. Машина опорних векторів	5
Лабораторна робота №3. Метод k-NN	5
Лабораторна робота №4. Баєсівська класифікація	5
МКР №1	20
Лабораторна робота №5. Дерева рішень	5
Лабораторна робота №6. Кластеризація k-means	5
Лабораторна робота №7. Асоціативні правила	5
Лабораторна робота №8. Нейронні мережі та глибоке навчання	20
МКР №2	20

Максимальна кількість балів за всі завдання дорівнює 100 балів.

Критерії оцінювання

Підготовка до роботи (у відсотках від максимальної кількості балів за відповідну роботу):

- протокол відповідає вимогам, охайний – 20 %;
- протокол відповідає вимогам, але є чисельні виправлення – 10 %;

Виконання завдання:

- робота виконана повністю і вірно протягом відведеного часу – 50 %;
- робота виконана пізніше зазначеного терміну – 20 %;

Якість захисту роботи:

- студент вірно і повністю відповів на запитання – 30 %;
- студент при відповіді допустив несуттєві неточності – 20 %;
- студент при відповіді на запитання допустив суттєві неточності, але самостійно виправив їх – 10 %.

2. Модульний контроль

Ваговий бал – 10×2 .

Контрольна робота складається з 10 тестових завдань. За кожен вірну відповідь на запитання надається 2 бали.

Сума вагових балів контрольних заходів протягом семестру складає:

$$R = 60 + 20 + 20 = 100 \text{ балів.}$$

Необхідною умовою допуску до екзамену є зарахування усіх завдань практичних занять, а також стартовий рейтинг (r_c) не менше 40% від R, тобто 40 балів.

Сума балів переводиться до екзаменаційної оцінки згідно з таблицею:

Бали (RD)	Традиційна оцінка
95..100	Відмінно
85...94	Дуже добре
75...84	Добре
65...74	Задовільно
60...64	Достатньо
RD<=60	Незадовільно
RD < 40 або не виконані інші умови допуску до екзамену	Не допущений

Робочу програму навчальної дисципліни (силабус):

Складено професор, д.ф.-м.н., с.н.с., Матичин Іван Іванович

Ухвалено кафедрою ЦТЕ (протокол № 1 від 01.07.2022р)

Погоджено Методичною комісією факультету (протокол № 10 від 04.07.2022 р.)