

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій

Затверджую
Декан факультету комп'ютерних
інформаційних технологій
Ігор ЯКИМЕНКО
" 2024р.



Затверджую
Проректор з інноваційно-педагогічної роботи
Віктор ОСТРОВЕРХОВ
" 2024 р..



РОБОЧА ПРОГРАМА
з дисципліни
«Машинне навчання»

Ступінь вищої освіти – перший(бакалаврський)
Галузь знань: 12 Інформаційні технології
Спеціальність: 124 Системний аналіз
Освітньо-професійна програма «Системний аналіз»

Кафедра економічної кібернетики та інформатики

Форма навчання	Курс	Семестр	Лекції	Лаб.	ІРС	Тре-нінг	СРС	Разом	Екзамен (семестр)
Денна	III	V	46	44	5	12	133	240	V

Тернопіль 2024

Робоча програма складена на основі освітньо-професійної програми підготовки бакалавра галузі знань 12 Інформаційні технології спеціальності 124 Системний аналіз, затвердженої на засіданні вченої ради ЗУНУ (протокол №9 від 15.06.2022 р.).

Робочу програму склав: професор кафедри економічної кібернетики та інформатики
Роман ПАСІЧНИК

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики, протокол № 1 від 28.08. 2024 р.

Завідувач кафедри

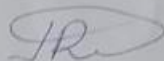


проф.

Леся БУЯК

Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол №__ від _____ 2024 р.

Голова ГЗС



проф. Роман ПАСІЧНИК

Гарант ОПП



проф. Роман ПАСІЧНИК

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Машинне навчання»

1. Опис дисципліни «Машинне навчання»

Дисципліна – Інтелектуальний аналіз даних	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів ECTS 8	Галузь знань – 12 «Інформаційні технології»	Статус дисципліни – обов'язкова, мова навчання українська
Кількість залікових модулів - 5	Спеціальність – 124 «Системний аналіз»,	<i>Денна:</i> Рік підготовки: 3 Семестр – 5
Кількість змістових модулів - 3	Ступінь вищої освіти – бакалавр	<i>Денна:</i> лекції – 46 год.; лаб.- 44 год
Загальна кількість годин - 240		Самостійна робота: 137 год., тренінг – 8 год Індивідуальна робота : 5 год.
Тижневих годин: 16 год., з них аудиторних – 6 год		Вид підсумкового контролю – <i>екзамен</i>

2. Мета й завдання вивчення дисципліни "Машинне навчання"

2.1. Мета вивчення дисципліни

Метою викладання дисципліни "Машинне навчання" є ознайомлення студентів з методологією підтримки прийняття рішень на основі методів машинного навчання із застосуванням їх для розв'язання прикладних задач.

2.2. Завдання вивчення дисципліни

В результаті вивчення курсу "Машинне навчання" 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;
- 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.

Предметом дисципліни є методи моделювання та дослідження процесів у складних системах та оцінки достовірності отриманих результатів.

2.3. Найменування та опис компетентностей, формування котрих забезпечує вивчення дисциплін:

К19. Здатність формалізувати проблеми, описані природною мовою, у тому числі за допомогою математичних методів, застосовувати загальні підходи до математичного моделювання конкретних процесів.

К27. Здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них.

2.4. Передумови для вивчення дисципліни.

Математичний аналіз, теорія імовірностей та математична статистика, програмування на Python.

2.5. Результати навчання:

ПР19. Проектувати, реалізовувати, супроводжувати програмні компоненти масштабовних систем роботи з даними і знаннями..

ПР21. Розуміти і застосовувати методи виявлення нових, нетривіальних та корисних знань в базах даних.

ПР22. Застосовувати методи і засоби штучного інтелекту у розробленні прикладних програмних засобів аналізу даних.

2.6. Завдання лекційних занять

Мета проведення лекцій полягає у тому, щоб ознайомити студентів із головними питаннями курсу "Машинне навчання".

Завдання проведення лекцій полягає у:

- викладенні студентам у відповідності з програмою та робочим планом основних питань курсу "Машинне навчання";
- сформуванні у студентів цілісної системи теоретичних знань з курсу "Машинне навчання".

2.7. Завдання проведення практичних занять

Мета проведення практичних занять полягає у тому, щоб виробити у студентів практичні навички використання теоретичного матеріалу.

Завдання проведення практичних занять полягає у глибшому засвоєнні та закріпленні теоретичних знань, одержаних на лекціях.

3. Програма дисципліни "Машинне навчання"

Змістовий модуль 1 – Попередній аналіз вибірок

Тема 1. Моделі машинного навчання

Підходи до трактування поняття машинне навчання. Статистична група методів у машинному навчанні. Категорії машинного навчання. Навчання з учителем та без учителя. Навчання з підкріпленням. Глибоке навчання. Глибоке навчання із підкріпленням.

Тема 2. Аналіз лінійних моделей даних

Лінійна модель даних із розбиттям на підвибірки. Реалізація мультилінійної регресії за допомогою пакету Scikit-learn. Відбір пояснюючих змінних регресійної моделі. Фільтрація змінних за кореляціями.

Тема 3. Гребенева регресія

Загальна гребнева мультилінійна регресійна модель. Мультиколінеарність. Регуляризація співвідношення ідентифікації моделі. Вибір структури гребневої регресійної моделі.

Тема 4. Реалізація лінійних моделей у статистичних пакетах

Середовище R. Операції в R. Побудова лінійних моделей в R. Порівняння точності лінійних моделей в R. Побудова багатовимірної лінійної моделі методом виключень. Вибір головних лінійно-незалежних компонент.

Тема 5. Задачі класифікації

Простий класифікатор. Логістична класифікація. Логістична класифікація на реальних даних. Лінійна дискримінантна класифікація. Квадратичний дискримінантний аналіз. Класифікація методом головних компонент.

Тема 6. Класифікація методом найближчих сусідів

Аналіз методу найближчих сусідів. Класифікація методом найближчих сусідів. Коефіцієнти кореляції для класифікації. Регресія методом найближчих сусідів.

Тема 7. Машина опорних векторів

Теорія методу машини опорних векторів. Площина максимального розділення. Нелінійна SVM класифікація. Підтримка візуалізації SVM моделі.

Тема 8. Байєсівський класифікатор та дерево рішень

Теорія наївних байєсівських класифікаторів. Приклад використання Байєсівських класифікаторів. Основні поняття теорії дерев рішень. Ключові елементи теорії дерев прийняття рішень. Приклади застосування дерев прийняття рішень.

Тема 9. Процеси Гауса

Вступ до Гаусівських процесів. Регресія на гаусівських процесах. Процеси перехресного розкладу. Порівняння методів перехресного розкладу. Приклад на зменшення розмірності задачі.

Тема 10. Ансамблеві методи

Бутстрапінг як метод випадкового формування вибірок. Основи бутстрапінгу та беггінгу. Беггінг як метод пакетування із агрегацією бутстрапінгу. Моделі випадкового лісу, як удосконалений беггінг

4. Структура залікового кредиту дисципліни "Машинне навчання"

	Кількість годин					
	Лекції	Лабораторні заняття	Індивідуальна робота	Тренінг	Самотійна робота	Контр. заходи
Змістовий модуль 1 – Попередній аналіз вибірок						
Тема 1. Моделі машинного навчання	4	3	2	6	13	поточне опитування
Тема 2. Аналіз лінійних моделей даних	4	3			13	
Тема 3. Гребенева регресія	4	4			13	
Тема 4. Реалізація лінійних	4	4			13	

моделей у статистичних пакетах						
Тема 5. Задачі класифікації	4	4			13	
Змістовий модуль 2 – Класифікатори						
Тема 6. Класифікація методом найближчих сусідів	4	4			13	По- точне опит.
Тема 7. Машина опорних векторів	4	4			13	
Тема 8. Байєсівський класифікатор та дерево рішень	4	4	3	6	14	
Тема 9. Процеси Гауса	4	4			14	
Тема 10. Ансамблеві методи	4	4			14	
Разом	46	44	5	12	133	

5. Тематика лабораторних занять

Лабораторне заняття 1. Моделі машинного навчання.

1. Категорії машинного навчання
2. Використання пакету ScikitLearn.
3. Приклади задач навчання з учителем та без учителя.
4. Приклади задач глибокого навчання та навчання із підкріпленням.

Лабораторне заняття 2. Аналіз лінійних моделей даних

1. Модуль мультилінійної регресії.
2. Відбір пояснюючих змінних регресійної моделі
3. Фільтрація змінних за кореляціями
4. Мультилінійна регресійна модель на Scikit-learn

Лабораторне заняття 3. Гребенева регресія

1. Регуляризація.
2. Загальна гребнева мультилінійна регресійна модель.
3. Модуль гребневої регресії
4. Модуль вибору структури гребневої регресії.

Лабораторне заняття 4. Реалізація лінійних моделей у статистичних пакетах

1. Середовище R.
2. Побудова лінійних моделей в R.
3. Порівняння точності лінійних моделей в R.
4. Спрощена лінійна модель.

Лабораторне заняття 5. Задачі класифікації

1. Простий класифікатор.
2. Відбір змінних для класифікації.
3. Логістична класифікація.
4. Лінійний дискримінантний аналіз

Лабораторне заняття 6. Класифікація методом найближчих сусідів

1. Пояснення методу найближчих сусідів.
2. Класифікація методом найближчих сусідів.
3. Коефіцієнти кореляції для класифікації.
4. Регресія методом найближчих сусідів

Лабораторне заняття 7. Машини опорних векторів

1. Мета алгоритму опорних векторів.
2. Розділяюча гіперплощина. М'яккі розділювачі.
3. Регуляризація.
4. Перехід просторів вищої розмірності.

Лабораторне заняття 8. Наївний байєсівський класифікатор

1. Теорема Байєса. Докази. Апостеріорна імовірність.
2. Класифікаційна імовірність. Наївне припущення.
3. Класова імовірність. Умовні імовірності.
4. Види класифікаторів: нормальний, мультиноміальний, Бернуллі.

Лабораторне заняття 9. Процеси Гауса

1. Вступ до Гаусівських процесів
2. Регресія на гаусівських процесах.
3. Модуль перехресного розкладу.
4. Зменшення розмірності задач.

Лабораторне заняття 10. Ансамблеві методи

1. Призначення ансамблевих методів.
2. Бутстрапінг як метод випадкового створення вибірок.
3. Беггінг як метод пакетування із агрегацією бутстрапінгу.
4. Моделі випадкового лісу, як удосконалений беггінг.

6. Самостійна робота

Самостійна робота "Розробка регресійних та класифікаційних моделей бізнес процесів"

Мета роботи: Розробити моделі для формування рекомендацій на основі статистичних масивів

Завдання:

1. Кожен студент отримує реальні статистичні дані.
2. На основі специфіки обраного статистичного матеріалу, студент проводить дослідження, яке включає:

- Аналіз природи та ролі атрибутів статистичного масиву
- Вибір регресійних та класифікаційних моделей
- Навчання та тестування моделей
- Формування висновків та рекомендацій для користувачів

3. Студент повинен підготувати звіт, який містить:

- Опис особливостей отриманих статистичних масивів
- Детальний опис структури регресійних та класифікаційних моделей
- Опис результатів навчання моделей
- Опис результатів тестування моделей
- Опис висновків і рекомендацій отриманих в результаті моделювання

Роботу необхідно здати викладачу у вигляді письмового звіту та презентації розробленого додатку. Захист роботи відбувається у формі усної презентації з демонстрацією функціоналу додатку.

7. Організація та проведення тренінгу з дисципліни "Машинне навчання"

Цей тренінг охоплює ключові аспекти моделювання із використанням нових неklasичних підходів. Учасники отримають досвід роботи з різноманітними методологіями ансамблевих методів.

Мета тренінгу:

Забезпечити учасників теоретичними знаннями та практичними навичками в галузі побудови та аналізу моделей у складних ситуаціях.

Перелік задач для тренінгу:

1. Збір та аналіз даних із Веб-середовища
2. Використання бутстрапінгу для формування представницьких вибірок.
3. Використання беггінгу як методу пакетування із агрегацією бутстрапінгу.
4. Використання моделей випадкового лісу із удосконалений методики беггінгу
5. Порівняння ефективності методик

Ці завдання дозволять студентам отримати практичний досвід формування моделей у складних

ситуаціях оброблення інформації з використанням ансамблевих методів.

Порядок проведення тренінгу:

Вступна частина проводиться з метою ознайомлення студентів із запропонованими завданнями тренінгу.

Організаційна частина полягає у створенні робочого настрою у колективі студентів.

Практична частина реалізується шляхом виконання одного вибраного завдання тренінгу.

Підведення підсумків. Обговорення результатів виконаних завдань. Обмін думками з питань, що виносились на тренінг.

8. Методи навчання.

У навчальному процесі застосовуються: лекції, в тому числі з використання мультимедійного проєктора та інших ТЗН; лабораторні роботи, індивідуальні заняття; самостійна робота студентів, робота в Інтернет.

9. Засоби оцінювання та методи демонстрування результатів навчання

- У процесі вивчення дисципліни "Машинне навчання" використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне опитування;
- модульне тестування та опитування;
- оцінювання лабораторних робіт;
- оцінювання тренінгів;
- оцінювання результатів самостійної роботи.

10. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни "Машинне навчання" визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

Модуль 1		Модуль 2		Модуль 3	Модуль 4	Модуль 5
10%	10%	10%	10%	5%	15%	40%
Поточне оцінювання	Модульний контроль 1	Поточне оцінювання	Модульний контроль 2	Тренінги	Самостійна робота	Екзамен
Оцінка за даний модуль визначається як середнє арифметичне за роботу на практичних заняттях №1-5	Підсумкова письмова робота за темами №1-5	Оцінка за даний модуль визначається як середнє арифметичне за роботу на практичних заняттях №6-10.	Підсумкова письмова робота за темами №6-10	Визначається як середнє арифметичне за виконання завдань за темами №1-5 тренінгу..	Оцінка за виконаний і представлений звіт із оброблених отриманих даних	1. Теоретична частина: 2 запитання по 20 балів (40 балів) 2. Практична частина 2 задачі по 30 балів кожна (60 балів).

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

№	Найменування	Номер теми
1.	Персональний комп'ютер	1-10

2.	Програмне середовище Python	1-10
3	Прикладні пакети scikit-learn, numpy, scipy	1-10

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Scikit-learn. User Guide. https://scikit-learn.org/stable/user_guide.html
2. О. І. Шеремет, О. В. Садовой. Метод опорних векторів (SVM). <https://www.dstu.dp.ua/Portal/Data/74/72/3st13-17.pdf>
3. Scikit-learn. Stochastic Gradient Descent. <https://scikit-learn.org/stable/modules/sgd.html>
4. Scikit-learn. SGD: Maximum margin separating hyperplane. https://scikit-learn.org/stable/auto_examples/linear_model/plot_sgd_separating_hyperplane.html#sphx-glr-auto-examples-linear-model-plot-sgd-separating-hyperplane-py
5. SGD: Weighted samples. https://scikit-learn.org/stable/auto_examples/linear_model/plot_sgd_weighted_samples.html#sphx-glr-auto-examples-linear-model-plot-sgd-weighted-samples-py
6. Nearest Neighbors regression. https://scikit-learn.org/stable/auto_examples/neighbors/plot_regression.html#sphx-glr-auto-examples-neighbors-plot-regression-py
7. Nearest Neighbors. <https://scikit-learn.org/stable/modules/neighbors.html>
8. Kaixin Wang. Introduction to Gaussian process regression. <https://medium.com/data-science-at-microsoft/introduction-to-gaussian-process-regression-part-1-the-basics-3cb79d9f155f>
9. Naive Bayes Classifiers. <https://www.geeksforgeeks.org/naive-bayes-classifiers/>.
10. Decision Tree. <https://www.geeksforgeeks.org/decision-tree/>