

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



Затверджую

Декан факультету комп'ютерних
інформаційних технологій

Ігор ЯКИМЕНКО

2024 р.



Затверджую

Проректор з науково-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

30.08.2024

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.), зі змінами (протокол № 11 від 26.06.2024 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 124 «Системний аналіз», протокол № 1 від 30.08.2024 р.

Голова ГЗС



проф. Роман ПАСІЧНИК

Гарант ОПП



проф. Роман ПАСІЧНИК

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Машинне навчання»

1. Опис дисципліни «Машинне навчання»

Дисципліна – Машинне навчання	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів ECTS – 8	Галузь знань – 12 «Інформаційні технології»	Статус дисципліни – обов'язкова дисципліна циклу професійної підготовки, мова навчання – українська
Кількість залікових модулів – 5	Спеціальність – 124 «Системний аналіз»	Денна: Рік підготовки – 3 Семестр – 5
Кількість змістових модулів – 2	Ступінь вищої освіти – бакалавр	Денна: Лекції – 46 год. Лабораторні – 44 год.
Загальна кількість годин – 240		Денна: Самостійна робота – 133 год. Тренінг – 12 год. Індивідуальна робота – 5 год.
Тижневих годин: 16 год., з них аудиторних – 6 год.		Вид підсумкового контролю – екзамен

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:

К29. Здатність організовувати роботу з проектування та розроблення масштабованих систем аналізу даних.

К30. Здатність використовувати методи збору та структурування інформації в глобальних інформаційних мережах

К31. Здатність застосовувати методи аналізу даних в адаптивних інформаційних системах.

2.4. Передумови для вивчення дисципліни.

Вища математика, теорія імовірностей та математична статистика, основи програмування, системний аналіз.

2.5. Результати навчання:

ПР20. Застосовувати методи автоматизації вилучення даних із структурованих інформаційних систем в глобальних мережах.

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

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. Гребнева 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.

Змістовий модуль 2 – Класифікатори

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 бегінгу. Бегінг як метод пакетування із агрегацією бутстрапінгу. Моделі випадкового лісу, як удосконалений бегінг.

4. Структура залікового кредиту дисципліни "Машинне навчання"

	Кількість годин					
	Лекції	Лабораторні заняття	Індивідуальна робота	Тренінг	Самостійна робота	Контрольні заходи
Змістовий модуль 1 – Попередній аналіз вибірок						
Тема 1. Моделі машинного навчання	4	4	2	6	13	Поточне опитування
Тема 2. Аналіз лінійних моделей даних	4	4			13	
Тема 3. Гребенева регресія	4	4			13	
Тема 4. Реалізація лінійних моделей у статистичних пакетах	4	4			13	
Тема 5. Задачі класифікації	5	4			13	
Змістовий модуль 2 – Класифікатори						
Тема 6. Класифікація методом найближчих сусідів	5	4	3	6	13	Поточне опитування.
Тема 7. Машина опорних векторів	5	5			13	
Тема 8. Байєсівський класифікатор та дерево рішень	5	5			14	
Тема 9. Процеси Гауса	5	5			14	
Тема 10. Ансамблеві методи	5	5			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 гребенева мультилінійна регресійна модель.
3. Модуль гребеневої регресії.
4. Модуль вибору структури гребеневої 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'ягкі розділювачі.
3. Регуляризація.
4. Перехід просторів вищої розмірності.

Лабораторне заняття 8. Наївний баєсівський 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 використанням 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	Тренінг	Самостійна робота	Екзамен
Оцінка за даний модуль визначається як середнє арифметичне за роботу на лабораторних заняттях по темах №1-5	Підсумкова письмова робота за темами №1-5	Оцінка за даний модуль визначається як середнє арифметичне за роботу на лабораторних заняттях по темах №6-10	Підсумкова письмова робота за темами №6-10	Визначається як середнє арифметичне за виконання завдань за темами тренінгу	Оцінка за виконаний і представлений звіт із оброблення отриманих даних	1. Теоретична частина: 2 запитання по 20 балів (40 балів) 2. Практична частина: 2 задачі по 30 балів кожна (60 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)

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

Основна література:

1. К. Ю. Кононова. Машинне навчання: методи та моделі. Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с. // https://moodle.znu.edu.ua/pluginfile.php/593075/mod_folder/intro/%D0%91%D0%B0%D0%B7%D0%BE%D0%B2%D0%B8%D0%B9%20%D0%BF%D1%96%D0%B4%D1%80%D1%83%D1%87%D0%BD%D0%B8%D0%BA_2%20%28%D0%9A%D0%BE%D0%BD%D0%BE%D0%BD%D0%BE%D0%B2%D0%B0%20%D0%9A.%20%D0%AE.%20%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%B5%20%D0%BD%D0%B0%D0%B2%D1%87%D0%B0%D0%BD%D0%BD%D1%8F%20-%20%D0%BC%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%20%D1%82%D0%B0%20%D0%BC%D0%BE%D0%B4%D0%B5%D0%BB%D1%96%29.pdf
2. Іван Пінцак. Використання машинного навчання та аналізу даних для прогнозування тенденцій у електронній комерції // <https://journals.politehnica.dp.ua/index.php/it/article/view/555>
3. Владислав Пилипенко. Дослідження точності методів машинного навчання при прогнозуванні успішності студентів. // <https://heraldts.khmnua.edu.ua/index.php/heraldts/article/view/244>
4. Пасічник Р.М. Курс лекцій з дисципліни «Машинне навчання» для студентів спеціальності 124 «Системний аналіз», освітньо – професійної програми «Системний аналіз», ступінь вищої освіти «бакалавр» всіх форм навчання. – Тернопіль: Західноукраїнський національний університет, 2022. – 51 с.
5. Пасічник Р.М. Методичні вказівки до практичних занять з курсу «Машинне навчання» для студентів спеціальності 124 «Системний аналіз», освітньо – професійної програми «Системний аналіз», ступінь вищої освіти «бакалавр» всіх форм навчання. – Тернопіль: Західноукраїнський національний університет, 2022. – 50 с.
6. Пасічник Р.М. Методичні вказівки для самостійної роботи з курсу «Машинне навчання» для студентів спеціальності 124 «Системний аналіз», освітньо – професійної програми «Системний аналіз», ступінь вищої освіти «бакалавр» всіх форм навчання. – Тернопіль: Західноукраїнський національний університет, 2022. – 32 с.

Додаткова література:

1. Зайченко Ю.П. Основи проектування інтелектуальних систем. – К.: “Слово”, 2004. – 352с.
2. Scikit-learn. User Guide. https://scikit-learn.org/stable/user_guide.html
3. О.І. Шеремет, О.В. Садовой. Метод опорних векторів (SVM). <https://www.dstu.dp.ua/Portal/Data/74/72/3st13-17.pdf>
4. Scikit-learn. Stochastic Gradient Descent. <https://scikit-learn.org/stable/modules/sgd.html>
5. 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
6. 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
7. Nearest Neighbors regression. https://scikit-learn.org/stable/auto_examples/neighbors/plot_regression.html#sphx-glr-auto-examples-neighbors-plot-regression-py
8. Nearest Neighbors. <https://scikit-learn.org/stable/modules/neighbors.html>
9. Kaixin Wang. Introduction to Gaussian process regression. <https://medium.com/data-science-at-microsoft/introduction-to-gaussian-process-regression-part-1-the-basics-3cb79d9f155f>
10. Naive Bayes Classifiers. <https://www.geeksforgeeks.org/naive-bayes-classifiers/>.
11. Decision Tree. <https://www.geeksforgeeks.org/decision-tree/>
12. Роман Пасічник, Франк Отоо. Метод побудови математичної моделі процесу у формі диференціального рівняння на підставі інтервальних різницевих рівнянь. Measuring and

Computing Devices in Technological Processes. № 2 (2023). Pp 134-145. DOI: <https://doi.org/10.31891/2219-9365-2023-74-17> ISSN: 2219-9365.

https://www.researchgate.net/publication/372512381_METOD_POBUDOWI_MATEMATICNOI_MODELI_PROCESU_U_FORMI_DIFERENCIALNOGO_RIVNANNA_NA_PIDSTAVI_INTERVALNIH_RIZNICEVIH_RIVNAN

13. Mykola Dyvak, Andriy Melnyk, Andriy Dyvak, Frank Otoo. Transformation of Mathematical Model for Complex Object in Form of Interval Difference Equations to a Differential Equation. International Journal of Computing, 22(2), 219-224.

<https://computingonline.net/computing/article/view/3091>

14. Андрій Миколайович Мельник, Микола Петрович Дивак, Роман Мирославович Пасічник. Метод виявлення неактуальної інформації в сервісно-орієнтованих корпоративних системах на прикладі систем оцінювання якості ґрунтів. ІТКІ, Вінниця, вип. 50, вип. 1, с. 45–54, Квіт 2021. <https://itce.vntu.edu.ua/index.php/itce/article/view/785>

15. Mykola Dyvak, Artur Rot, Roman Pasichnyk, Vasyl Tymchyshyn, Nazar Huliiev, Yurii Maslyiak. Monitoring and Mathematical Modeling of Soil and Groundwater Contamination by Harmful Emissions of Nitrogen Dioxide from Motor Vehicles. Sustainability 2021, 13(5), 2768; 2 quartiles. <https://www.mdpi.com/2071-1050/13/5/2768>

16. Mykola Dyvak, Roman Pasichnyk. Inductive and Deductive Approaches to Modeling the Daily Cycle of Carbon Monoxide Concentrations Due to Air Pollution by Motor Vehicles. Advanced Computer Information Technologies ACIT'2024. Pp.77-71.

<https://ieeexplore.ieee.org/document/10712476>