

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

Затверджую  
Декан факультету комп'ютерних  
інформаційних технологій  
  
Гор Якименко  
" " 2024р.



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



Затверджую  
Директор ННІОТ  
  
Святослав Питель  
" " 2024р.



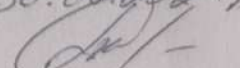
## РОБОЧА ПРОГРАМА

з дисципліни  
«Аналіз даних»

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

Кафедра економічної кібернетики та інформатики

Форма навчання	Курс	Семестр	Лекції	Практ.	ІРС	Тре-нінг	СРС	Разом	Залік (семестр)
Денна	I	II	32	14	5	6	93	150	II
Заочна	I	II	8	4	-	-	138	150	II

30.08.2024  


Робоча програма складена на основі освітньо-професійної програми підготовки бакалавра галузі знань освітньо-професійної програми підготовки бакалавра галузі знань 12 Інформаційні технології спеціальності 124 Системний аналіз, затвердженої на засіданні вченої ради ЗУНУ (протокол № 11 від 26.06.24)

Робочу програму склав: професор кафедри економічної кібернетики та інформатики ПАСІЧНИК Роман Мирославович, Данилюк Ірина Вадимівна

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики, протокол № 1 від 28.08.2024 р.

Завідувач кафедри



проф. БУЯК Леся Михайлівна

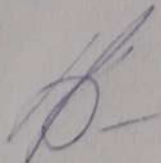
Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол №1 від 30.08. 2024 р.

Голова ГЗС



проф. ПАСІЧНИК Роман Мирославович

Гарант ОПП



проф. БОДНАР Дмитро Ількович

# СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

## «Аналіз даних»

### 1. Опис дисципліни «Аналіз даних»

Дисципліна – Аналіз даних	Галузь знань, спеціальність, ступінь вищої освіти	Характеристика навчальної дисципліни
Кількість кредитів ECTS 5	<b>Галузь знань – 12</b> «Інформаційні технології»	Нормативна дисципліна циклу професійної підготовки, мова навчання - <i>українська</i>
Кількість залікових модулів - 4	<b>Спеціальність – 124</b> «Системний аналіз»,	<i>Денна:</i> Рік підготовки: 1 Семестр – 1 <i>Заочна:</i> Рік підготовки: 1 Семестр – 2
Кількість змістових модулів - 2	Ступінь вищої освіти – магістр	<i>Денна:</i> лекції – 32 год.; практ.- 14 год <i>Заочна:</i> лекції – 8 год.; практ.- 4 год
Загальна кількість годин - 150 год.		Самостійна робота: 93 год., Тренінг – 6 год. Індивідуальна робота : 5 год.
Тижневих годин: 10 год., з них аудиторних – 3 год		Вид підсумкового контролю – залік

## **2. Мета й завдання вивчення дисципліни "Аналіз даних"**

### **2.1. Мета вивчення дисципліни**

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

### **2.2. Завдання вивчення дисципліни**

В результаті вивчення курсу "Аналіз даних" студенти повинні володіти основними поняттями 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.

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

В результаті вивчення курсу "Аналіз даних" студенти повинні володіти такими навичками:

- здатність до абстрактного мислення, аналізу та синтезу;
- здатність до пошуку оброблення та аналізу інформації з різних джерел;
- здатність моделювати, прогнозувати та проектувати складні системи і процеси на основі методів та інструментальних засобів системного аналізу;
- здатність застосовувати теорію і методи Data Science для здійснення інтелектуального аналізу даних з метою виявлення нових властивостей та генерації нових знань про складні системи;
- здатність управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів;
- здатність управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів;
- здатність розробляти і реалізовувати наукові та прикладні проекти в галузі інформаційних технологій та дотичні до неї міждисциплінарні проекти.

**2.4. Результати навчання.** Результатом вивчення дисципліни «Системи прийняття рішень в умовах невизначеності» є:

- спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері системного аналізу та інформаційних технологій і є основою для оригінального мислення та проведення досліджень;
- застосовувати методи машинного навчання та інтелектуального аналізу даних, математичний апарат нечіткої логіки, теорії ігор та розподіленого штучного інтелекту для розв'язання складних задач системного аналізу;
- застосовувати методи машинного навчання та скрапінгу для видобування та аналізу структурованих даних із глобальних мереж;
- вільно презентувати та обговорювати усно і письмово результати досліджень та інновацій, інші питання професійної діяльності державною та англійською мовами.

### **3. Програма дисципліни "Аналіз даних"**

#### ***Змістовий модуль 1 – Попередній аналіз вибірок***

##### **○ Тема 1. Лінійний і квадратичний дискримінантний аналіз**

Лінійний дискримінантний аналіз (LDA) із застосуванням генеративного підходу для класифікації. Моделі класів та їх коваріацій. Випадок двійкової класифікації. Гребенева регресія.

Квадратний дискримінантний аналіз (QDA) із індивідуальною коваріаційною матрицею. Межова квадратична поверхня.

##### **Тема 2. Машини опорних векторів**

Мета алгоритму опорних векторів. Розділяюча гіперплощина. М'яккі розділювачі. Регуляризація. Перехід просторів вищої розмірності. Розділяючі гіперплощини.

##### **Тема 3. Стохастичний градієнтний спуск**

Рух за антиградієнтом. Обмеження на крок спуску. Стохастична точка спуску. Множина точок пошуку. Використання стохастичного градієнтного спуску для у лінійних моделях машинного навчання.

##### **Тема 4. Метод найближчих сусідів**

Класифікація точок на основі класифікації найближчих сусідів. Підбір кількості сусідів. Підбір кількості сусідів на невеликій навчальній вибірці. Вибір метрики у відстанях між сусідами. Переваги та недоліки методу найближчих сусідів.

##### **Тема 5. Суміші нормальних розподілів**

Гаусівські процеси. Постійне ядро. Ядро білого шуму. Квадратично-експоненціальне ядро. Базисна функція (RBF). Періодичне ядро. Розподіл імовірностей для нових спостережень. Прогнозування Гаусівського процесу. Якість інтерполяції та екстраполяції.

#### ***Змістовий модуль 2 – Класифікатори***

##### **Тема 6. Наївний баєсівський класифікатор**

Теорема Байєса. Докази. Апостеріорна імовірність. Класифікаційна імовірність. Наївне припущення. Класова імовірність. Умовні імовірності. Припущення щодо умовних імовірностей. Види класифікаторів: нормальний, мультиноміальний, Бернуллі.

##### **Тема 7. Деревя рішень**

Структура дерева рішень: вузли, гілки, листки, функції, правила, результати. Алгоритм дерева класифікації та регресії. Сильні та слабкі сторони підходу дерева рішень

## **Тема 8. Ансамблеві методи**

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

## **Тема 9. Напівконтрольоване навчання**

Основна відмінність між контрольованим та неконтрольованим навчанням. Використання невеликої кількості розмічених та великої кількості нерозмічених даних. Приклади напівконтрольованого навчання. Маркування аудіофайлів, веб-контенту, білкових структур.

## **Тема 10. Ізотонічна регресія**

Проблема знаходження монотонної функції, що мінімізує похибку на експериментальних даних. Покрокове використання динамічного програмування. Застосування ізотонічної регресії. Корекція прогнозованих даних. Моделювання порядкових змінних. Врахування відсутніх значень. Виявлення викидів.

#### 4. Структура залікового кредиту дисципліни "Аналіз даних"

Денна форма

	Кількість годин					
	Лекції	Практичні заняття	Самостійна робота	Індивідуальна робота	Тренінг	Контроль заходів
Змістовий модуль 1 – Попередній аналіз вибірок						
Тема 1. Лінійний і квадратичний дискримінантний аналіз	2	1	9	2	3	поточне опит.
Тема 2. Машини опорних векторів	4	1	9			поточне опит.
Тема 3. Стохастичний градієнтний спуск	4	1	9			поточне опит.
Тема 4. Метод найближчих сусідів	4	1	9			поточне опит.
Тема 5. Суміші нормальних розподілів	4	2	9			модульн контр
Змістовий модуль 2 – Класифікатори						
Тема 6. Наївний баєсівський класифікатор	2	1	9	3	3	поточне опит.
Тема 7. Дерева рішень	2	1	9			поточне опит.
Тема 8. Ансамблеві методи	4	2	10			поточне опит.
Тема 9. Напівконтрольоване навчання	4	2	10			поточне опит.
Тема 10. Ізотонічна регресія	4	2	10			модул. контр.
Разом	32	14	93	5	6	

Заочна форма

	Кількість годин				
	Лекції	Практичні заняття	Самостійна робота	Індивідуальна робота	Контроль заходів
Змістовий модуль 1 – Попередній аналіз вибірок					
Тема 1. Лінійний і квадратичний дискримінантний аналіз	1	-	-	13	поточне опит.
Тема 2. Машини опорних векторів	1	-	-	13	поточне опит.
Тема 3. Стохастичний градієнтний спуск	-	1	-	14	поточне опит.
Тема 4. Метод найближчих сусідів	-	1	-	14	поточне опит.
Тема 5. Суміші нормальних розподілів	1	-	-	14	модульн контр
Змістовий модуль 2 – Класифікатори					
Тема 6. Наївний баєсівський класифікатор	1		-	14	поточне опит.
Тема 7. Дерева рішень	1	1	-	14	поточне опит.
Тема 8. Ансамблеві методи	1	1	-	14	поточне опит.

Тема 9. Напівконтрольоване навчання	1		-	14	поточне опит.
Тема 10. Ізотонічна регресія	1		-	14	Контр.
Разом	8	4	-	138	

## 5. Тематика практичних занять

### Практичне заняття 1

Тема 1. Лінійний і квадратичний дискримінантний аналіз.

1. Лінійний дискримінантний аналіз (LDA)
2. Гребенева регресія.
3. Квадратний дискримінантний аналіз (QDA) із індивідуальною коваріаційною матрицею.
4. Межова квадратична поверхня.

Тема 2. Машини опорних векторів

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

### Практичне заняття 2

Тема 3. Стохастичний градієнтний спуск

1. Організація методу спуску
2. Обмеження на крок спуску.
3. Множина точок пошуку.
4. Використання стохастичного градієнтного спуску у лінійних моделях машинного навчання.

Тема 4. Метод найближчих сусідів

1. Класифікація точок на основі класифікації найближчих сусідів.
2. Підбір кількості сусідів.
3. Вибір метрики у відстанях між сусідами.
4. Переваги та недоліки методу найближчих сусідів.

### Практичне заняття 3

Тема 5. Суміші нормальних розподілів

1. Гаусівські процеси. Постійне ядро.
2. Ядро білого шуму.
3. Квадратично-експоненціальне ядро.
4. Прогнозування Гаусівського процесу.

### Практичне заняття 4

Тема 6. 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.



## Тема 7. Дерева рішень

1. Структура дерева рішень.
2. Вузли, гілки, листки, функції, правила, результати.
3. Алгоритм дерева класифікації та регресії.
4. Сильні та слабкі сторони підходу дерева рішень.

### Практичне заняття 5

## Тема 8. 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.

### Практичне заняття 6

## Тема 9. Напівконтрольоване навчання

1. Основна відмінність між контрольованим та неконтрольованим навчанням.
2. Використання невеликої кількості розмічених та великої кількості нерозмічених даних.
3. Приклади напівконтрольованого навчання.
4. Маркування аудіофайлів, веб-контенту, білкових структур.

### Практичне заняття 7

## Тема 10. Ізотонічна регресія

1. Проблема знаходження монотонної функції, що мінімізує похибку на експериментальних даних.
2. Покрокове використання динамічного програмування.
3. Корекція прогнозованих даних.
4. Моделювання порядкових змінних.

## 6. Самостійна робота

1. Вступна частина: ознайомлення студентів з темою тренінгового заняття і видача завдання.
2. Практична частина: виконання завдань студентами згідно з індивідуальним завданням:
  - 2.1. Попередній лінійний та квадратичний аналіз вибірок
  - 2.2. Байєсівська класифікація
  - 2.3. Класифікація деревом рішень
  - 2.4. Застосування ансамблевих методів
3. Перевірка результатів з допомогою програмного середовища Python (прикладні пакети scikit-learn, numpy, scipy); оформлення короткого звіту.
4. Підведення підсумків: обговорення результатів виконаних завдань.

## 7. Тренінг з дисципліни «Аналіз даних»

Тематика: Ансамблеві методи.

Порядок проведення:

1. Обрати індивідуальні дані
2. Використати бутстрапінг для випадкового створення вибірок.
3. Використати беггінг як метод пакетування із агрегацією бутстрапінгу.
4. Використати моделі випадкового лісу, як удосконалений беггінг.
5. Порівняти ефективність методів
6. Обговорення та презентація результатів

### 8. Засоби оцінювання та методи демонстрування результатів навчання

- У процесі вивчення дисципліни «Аналіз даних» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне опитування;
- залікове модульне тестування та опитування;
- аналітичні звіти, реферати, есе;
- оцінювання результатів СРС;
- розрахункові роботи;
- тренінг.

### 9. Критерії, форми поточного та підсумкового контролю

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

Модуль 1		Модуль 2	Модуль 3
40%	40%	5%	15%
Поточне оцінювання	Модульний контроль	Тренінги	Самостійна робота
Визначається як середнє арифметичне оцінок, отриманих на практичних заняттях (кожен здобувач має отримати не менше 4 оцінок)	Підсумкова контрольна робота, яка охоплює всі теми	Визначається як оцінка, отримана за виконанні завдання під час тренінгу	Визначається як оцінка, отримана за виконання завдання самостійної роботи

### 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)

## 10. 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](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](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](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](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/>