

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АЕРОКОСМІЧНИЙ УНІВЕРСИТЕТ
ІМ. М. Є. ЖУКОВСЬКОГО
«ХАРКІВСЬКИЙ АВІАЦІЙНИЙ ІНСТИТУТ»**

**КАФЕДРА ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ
ТЕХНОЛОГІЙ
ІМ. О.О. ЗЕЛЕНСЬКОГО (№ 504)**

«ЗАТВЕРДЖУЮ»

Гарант освітньої програми



(підпис)

Віктор МАКАРЧУК

26 серпня 2024 р.

**РОБОЧА ПРОГРАМА ОБОВ'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

«Машинне навчання і аналіз даних»

(назва навчальної дисципліни)

Галузь знань: 17 «Електроніка, автоматизація та електронні комунікації»
(шифр і найменування галузі знань)

Спеціальність: 172 «Електронні комунікації та радіотехніка»
(код і найменування спеціальності)

Освітньо-наукова програма «Інформаційні мережі зв'язку»
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: другий (магістерський)

Харків 2024 рік

Розробник: доцент, к. ф.-м. н. Віктор МАКАРІЧЕВ

(посада, науковий ступінь і вчене звання, ім'я та прізвище)



(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційно-комунікаційних технологій ім. О.О. Зеленського

(назва кафедри)

Протокол № 1 від 26 серпня 2024 р.

Завідувач кафедри д.т.н., професор

(науковий ступінь і вчене звання)



(підпис)

Володимир ЛУКІН

(ім'я та прізвище)

1. Опис навчальної дисципліни

Найменування показника	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів – 5	<p>Галузь знань 17 «Електроніка, автоматизація та електронні комунікації» (шифр і найменування)</p> <p>Спеціальність 172 «Електронні комунікації та радіотехніка» (код і найменування)</p> <p>Освітньо-наукова програма «Інформаційні мережі зв'язку» (найменування)</p> <p>Рівень вищої освіти: другий (магістерський)</p>	Обов'язкова
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2024/2025
Індивідуальне завдання не передбачене (назва)		Семестр
Загальна кількість годин – 64/150		2-й
		Лекції*
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи здобувача – 5,4		32 години
		Практичні, семінарські
		32 години
		Лабораторні*
	-	
	Самостійна робота	
86 годин		
Вид контролю	модульний контроль, іспит	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: 64/86.

*Аудиторне навантаження може бути зменшене або збільшене на одну годину залежно від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета вивчення: формування у студентів теоретичних та практичних знань та навичок, що необхідні для розробки та застосування моделей машинного навчання й використання методів аналізу даних.

Завдання: оволодіння основними методами розробки моделей машинного навчання та аналізу даних.

Компетентності, які набуваються

Загальні компетентності:

- Здатність досліджувати проблеми з використанням системного аналізу, синтезу, комп'ютерного моделювання та методів оптимізації.

- Здатність аналізувати, верифікувати, оцінювати повноту інформації в ході професійної діяльності, за необхідності доповнювати й синтезувати відсутню інформацію й працювати в умовах невизначеності.

Фахові компетентності:

- Здатність застосовувати відповідні математичні, наукові і технічні методи, а також комп'ютерне програмне забезпечення для вирішення завдань в сфері розподілу і обробки інформації.

- Здатність організовувати і проводити експериментальні дослідження.

- Здатність продемонструвати знання і розуміння математичних принципів і методів, необхідних для підтримки спеціалізацій з телекомунікацій.

- Здатність застосовувати математичну теорію організації і планування експерименту, розробляти плани проведення досліджень, вибирати алгоритми опрацювання інформації, а також застосовувати необхідне програмне забезпечення для автоматизації обчислень.

- Здатність використовувати знання з математичних і логічних побудов, які є підґрунтям оптимізації інформаційних систем та мереж, їх окремих пристроїв, що проектуються, експлуатуються чи досліджуються.

Очікувані результати навчання:

- Знання і розуміння сучасних методів ведення науково-дослідних робіт, фізико-математичних методів, що застосовуються в інженерній і дослідницькій практиці, на рівні, необхідному для досягнення інших результатів освітньої програми.

- Знання сучасних методів і програмного забезпечення побудови адекватних теоретичних моделей і способів їх обґрунтування.

- Знання з математичних і логічних побудов, які є основою оптимізації інформаційних систем та мереж, їх окремих пристроїв, що проектуються, експлуатуються чи досліджуються.

- Здатність ефективно застосовувати роботу з комп'ютером, його технічним та програмним забезпеченнями (носіями інформації, базами даних тощо).

Пререквізити – сучасні методи обробки сигналів.

Кореквізити – інтернет речей, інтелектуальна обробка даних дистанційного зондування.

Постреквізити - глибинне навчання в інфокомунікаціях.

3. Зміст навчальної дисципліни

Модуль 1.

Змістовний модуль №1. Базові засоби машинного навчання і аналізу даних.

Тема 1. Вступ до дисципліни. Предмет вивчення і задачі дисципліни. Місце дисципліни в навчальному плані. Місце у штучному інтелекті. Сфера застосувань.

Тема 2. Оцінка моделей машинного навчання. Індикатори ефективності, оцінювання передбачень моделей, оцінювання класифікаторів та моделей кластеризації, перехресна перевірка моделей, відбір моделей.

Тема 3. Зниження розмірності. Метод головних компонент. Алгоритми t-SNE та UMAP й налаштування їх гіперпараметрів. Метод частинної регресії найменших квадратів. Застосування.

Модульний контроль

Змістовний модуль №2. Передові засоби машинного навчання.

Тема 1. Метод опорних векторів. Метод опорних векторів (SVM): побудова, лінійно розділені та нерозділені множини, використання ядер, метод SVM-LS, обмеження та застосування.

Тема 2. Дерева та ліси. Тренування класифікаційного та регресійного дерева прийняття рішень, візуалізація, ідентифікація важливих ознак у випадкових лісах, відбір важливих ознак у випадкових лісах, обробка незбалансованих класів, управління розміром дерева, покращення результативності за допомогою бустингу, оцінювання випадкових лісів за допомогою помилок позапакетних спостережень.

Тема 3. Наївний баєсів класифікатор. Тренування класифікатора для безперервних ознак, тренування класифікатора для дискретних та лічильних ознак, тренування наївного байєсового класифікатора для бінарних ознак, калібрування передбачуваних ймовірностей.

Модульний контроль

4. Структура навчальної дисципліни

Назва змістовного модуля і тем	Кількість годин				
	Усього	У тому числі			
		л	п	лаб.	с. р.
1	2	3	4	5	6
Модуль 1					
Змістовний модуль №1. Базові засоби машинного навчання і аналізу даних					
Тема 1. Вступ до дисципліни. Предмет вивчення і задачі дисципліни. Місце дисципліни в навчальному плані. Місце у штучному інтелекті. Сфера застосувань.	12	2	2	-	8
Тема 2. Оцінка моделей машинного навчання. Індикатори ефективності, оцінювання передбачень моделей, оцінювання класифікаторів та моделей кластеризації, перехресна перевірка моделей, відбір моделей.	20	4	4	-	10
Тема 3. Зниження розмірності. Метод головних компонент. Алгоритми t-SNE та UMAP й налаштування їх гіперпараметрів. Метод частинної регресії найменших квадратів. Застосування.	33	6	10	-	18
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 1	67	14	16	-	36
Змістовний модуль №2. Передові засоби машинного навчання.					
Тема 1. Метод опорних векторів. Метод опорних векторів (SVM): побудова, лінійно розділені та нерозділені множини, використання ядер, метод SVM-LS, обмеження та застосування.	38	6	6	-	20
Тема 2. Дерева та ліси. Тренування класифікаційного та регресійного дерева прийняття рішень, візуалізація, ідентифікація важливих ознак у випадкових лісах, відбір важливих ознак у випадкових лісах, обробка незбалансованих класів, управління розміром дерева, покращення результативності за допомогою бустингу, оцінювання випадкових лісів за допомогою помилок	34	6	6	-	18

позапакетних спостережень. Застосування.					
Тема 3. Наївний бассів класифікатор. Тренування класифікатора для безперервних ознак, тренування класифікатора для дискретних та лічильних ознак, тренування наївного байєсового класифікатора для бінарних ознак, калібрування передбачуваних ймовірностей. Застосування.	24	4	4	-	12
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 2	98	18	16	-	50
Усього годин	150	32	32	-	86

5. Теми семінарських занять

№ п/п	Назва теми	Кількість годин
1	Не передбачено навчальним планом	
	Разом	

6. Теми практичних занять

№ п/п	Назва теми	Кількість годин
1	Програмні засоби дисципліни	2
2	Засоби аналізу ефективності та оцінювання передбачень моделей. Оцінювання класифікаторів.	2
3	Оцінювання моделей кластеризації. Відпрацювання методів перехресної перевірки моделей. Відпрацювання методів відбору моделей	2
4	Метод головних компонент та його застосування	2
5	Алгоритм t-SNE, налаштування гіперпараметрів, застосування	2
6	Алгоритм UMAP, налаштування гіперпараметрів, застосування	2
7	Порівняльний аналіз PCA, t-SNE та UMAP на прикладі обробки наборів MNIST та Fashion MNIST	2
8	Метод частинної регресії найменших квадратів, інтерпретація результатів та застосування.	2
9	Класичний варіант методу опорних векторів: програмна реалізація, аналіз, обмеження, застосування	2
10	Метод опорних векторів (SVM): використання ядерних функцій, порівняльний аналіз	2
11	Метод опорних векторів (SVM): обмеження, застосування.	2
12	Метод LS-SVM: випадки лінійно розділених та нерозділених множин, використання ядерних функцій, порівняльний аналіз, обмеження, застосування	2
13	Тренування класифікаційного дерева прийняття рішень та класифікаційного випадкового лісу. Тренування регресійного дерева прийняття рішень та регресійного випадкового лісу	2
14	Візуалізація моделей дерева прийняття рішень, Ідентифікація важливих ознак у випадкових лісах. Відбір важливих ознак у випадкових лісах. Обробка незбалансованих класів	2
15	Керування розміром дерева. Покращення результативності за допомогою бустингу, оцінювання випадкових лісів за допомогою помилок позапакетних спостережень	2
16	Тренування класифікаторів для безперервних ознак. Тренування	2

	класифікатора для дискретних та лічильних ознак	
	Разом	32

7. Теми лабораторних занять

№ п/п	Назва теми	Кількість годин
1	Не передбачено навчальним планом	

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1	Сфера застосувань методів машинного навчання і аналізу даних	8
2	Розширені програмні засоби аналізу та відбору моделей	10
3	Реальні кейси використання алгоритмів PCA, t-SNE, UMAP, PLS	18
4	Реальні кейси використання метод опорних векторів та обмеження	20
5	Реальні кейси використання методів дерев та лісів, обмеження, порівняльний аналіз	18
6	Реальні кейси використання наївних баєсівських класифікаторів	12
	Разом	86

9. Індивідуальні завдання

Не передбачене.

10. Методи навчання

При викладанні курсу використовуються наступні навчальні методи:

- наочний метод (ілюстрація, демонстрація);
- створення ситуації зацікавленості;
- словесний метод (розповідь, лекція, бесіда, пояснення);
- спостереження;
- дослідження;
- пояснювально-ілюстративний метод;
- практичний (вправи).

11. Методи контролю

Для контролю успішності в даному курсі використано:

- поточний контроль (на практичних заняттях);
- модульний контроль за змістовними модулями;
- семестровий контроль у вигляді іспиту.

12. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Модуль 1			
Робота на лекціях	0...1	6	0...6
Виконання і захист практичних робіт	0...2	9	0...18
Модульний контроль	0...16	1	0...16
Модуль 2			
Робота на лекціях	0...1	8	0...8
Виконання і захист практичних робіт	0... 2	15	0...30
Модульний контроль	0...22	1	0...22
Усього за семестр			0...100

Білет для іспиту складається з 2 теоретичних завдань. Максимальна кількість балів за одне завдання – 50

Критерії оцінювання роботи здобувача протягом семестру

Зараховано (60-74). Показати мінімум знань та умінь, відпрацювати практичні роботи, здати тестування. Знати основні методи машинного навчання і аналізу даних.

Зараховано (75-89). Твердо знати мінімум, відпрацювати практичні роботи, здати тестування та позааудиторну самостійну роботу. Окрім базових знань знати особливості методів, межу використання, вміти проводити порівняльний аналіз.

Зараховано (90-100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та вміти застосовувати їх.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

13. Рекомендована література

Базова

1. Кононова К.Ю. Машинне навчання: методи та моделі, 2020.
<https://github.com/katerynakononova/ML>
2. Albon C., Machine Learning with Python Cookbook, 2018.
3. Charu C. Aggarwal, Data Mining: The Textbook, 2015.
4. Kuhn M., Johnson K., Applied Predictive Modeling, 2013.

Допоміжна

5. Документація NumPy, <https://numpy.org/doc/stable/>.
6. Документація scikit-learn, https://scikit-learn.org/stable/user_guide.html.

14. Інформаційні ресурси

Науково-технічна бібліотека ХАІ - Режим доступу: <http://library.khai.edu>.

Youtube-канал спільноти DeepLearning.AI - <https://www.youtube.com/@Deeplearningai>