

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)

Олексій РУБЕЛЬ

31 серпня 2023 р.

РОБОЧА ПРОГРАМА ОBOB'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Машинне навчання і аналіз даних»
(назва навчальної дисципліни)

Галузь знань: 12 «Інформаційні технології»
(шифр і найменування галузі знань)

Спеціальність: 126 «Інформаційні системи та технології»
(код і найменування спеціальності)


Освітня програма «Штучний інтелект та інформаційні системи»
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2023 рік

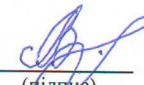
Розробник: МАКАРІЧЕВ Віктор, доцент каф. 504, к.ф.-м. н.
(прізвище та ініціали, посада, науковий ступінь та вчене звання)


(підпис)

Робочу програму розглянуто на засіданні кафедри _____
інформаційно-комунікаційних технологій ім. О.О. Зеленського
(назва кафедри)

Протокол № 1 від 31 серпня 2023 р.

Завідувач кафедри д.т.н., професор
(науковий ступінь і вчене звання)


(підпис)

Володимир ЛУКІН
(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показника	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів – 5,5	<p>Галузь знань 12 «Інформаційні технології» (шифр і найменування)</p> <p>Спеціальність 126 «Інформаційні системи та технології» (код і найменування)</p> <p>Освітня програма «Штучний інтелект та інформаційні системи» (найменування)</p> <p>Рівень вищої освіти: перший (бакалаврський)</p>	Обов'язкова
Кількість модулів – 2		Навчальний рік
Кількість змістовних модулів – 2		2023/2024
Індивідуальне завдання <u>не передбачене</u> (назва)		Семестр
Загальна кількість годин – 80*/165		6/4** -й
Кількість тижневих годин для денної форми навчання: аудиторних – 5 самостійної роботи здобувача – 5,3		Лекції*
		32 години
		Практичні, семінарські
		48 годин
		Лабораторні*
	-	
	Самостійна робота	
	85 годин	
	Вид контролю	
	модульний контроль, іспит	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: 80/85.

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

** група за скороченою формою навчання

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 застосовувати знання у практичних ситуаціях.
- Здатність вчитися і оволодівати сучасними знаннями.
- Здатність розробляти та управляти проєктами.
- Здатність оцінювати та забезпечувати якість виконаних робіт.

Спеціальні компетентності:

- Здатність аналізувати об'єкт проєктування або функціонування та його предметну область.
- Здатність до проєктування, розробки, налагодження та удосконалення системного, комунікаційного та програмно-апаратного забезпечення інформаційних систем та технологій, інтернету речей (IoT), комп'ютерно-інтегрованих систем та системної мережної структури, управління ними.
- Здатність проєктувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).
- Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання завдань та обов'язків.
- Здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій за використанням математичних моделей і методів.
- Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет).
- Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.
- Здатність формувати нові конкурентоспроможні ідеї й реалізовувати їх у проєктах (стартапах).

Очікувані результати навчання:

- Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.
- Застосовувати правила оформлення проєктних матеріалів інформаційних систем та технологій, знати склад та послідовність виконання проєктних робіт з урахуванням вимог відповідних нормативно-правових документів для запровадження у професійній діяльності.

Пререквізити – основи машинного навчання.

Кореквізити – машинне навчання та аналіз даних (курсний проєкт), глибинне навчання (курсний проєкт).

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

Модуль 2.

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 байєсів 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	6	-	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	-	17
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 1	67	14	18	-	35
Модуль 2					
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	12	-	20
Тема 2. Дерева та ліси. Тренування класифікаційного та регресійного дерева прийняття рішень, візуалізація,	34	6	10	-	18

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. Застосування.					
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. Застосування.	24	4	8	-	12
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 2	98	18	30	-	50
Усього годин	165	32	48	-	85

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	2
3	Оцінювання класифікаторів. Оцінювання моделей кластеризації	2
4	Відпрацювання методів перехресної перевірки моделей. Відпрацювання методів відбору моделей	2
5	Метод головних компонент та його застосування	2
6	Алгоритм t-SNE, налаштування гіперпараметрів, застосування	2
7	Алгоритм UMAP, налаштування гіперпараметрів, застосування	2
8	Порівняльний аналіз PCA, t-SNE та UMAP на прикладі обробки наборів MNIST та Fashion MNIST	2
9	Метод частинної регресії найменших квадратів, інтерпретація результатів та застосування.	2
10	Класичний варіант методу опорних векторів: програмна реалізація, аналіз, обмеження, застосування	2
11	Метод опорних векторів (SVM): використання ядерних функцій, порівняльний аналіз, обмеження, застосування.	6
12	Метод LS-SVM: випадки лінійно розділених та нерозділених множин, використання ядерних функцій, порівняльний аналіз, обмеження, застосування	4
13	Тренування класифікаційного дерева прийняття рішень та класифікаційного випадкового лісу	2
14	Тренування регресійного дерева прийняття рішень та регресійного випадкового лісу	2
15	Візуалізація моделей дерева прийняття рішень, Ідентифікація важливих ознак у випадкових лісах. Відбір важливих ознак у випадкових лісах	2

16	Обробка незбалансованих класів. Керування розміром дерева	2
17	Покращення результативності за допомогою бустингу, оцінювання випадкових лісів за допомогою помилок позапакетних спостережень	2
18	Тренування класифікаторів для безперервних ознак	2
19	Тренування класифікатора для дискретних та лічильних ознак	2
20	Тренування наївного байєсового класифікатора для бінарних ознак	2
21	Калібрування передбачуваних ймовірностей	2
	Разом	48

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	17
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	85

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

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

Добре (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.

Відмінно (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. Методичне забезпечення

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.
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>