

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
(ініціали та прізвище)

31 серпня 2023 р.

**РОБОЧА ПРОГРАМА *ОБОВ'ЯЗКОВОЇ*
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

«Основи машинного навчання»
(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: РУБЕЛЬ Андрій, старший викладач, доктор філософії
(прізвище та ініціали, посада, науковий ступінь і вчене звання)


(підпис)

Робочу програму розглянуто на засіданні кафедри інформаційно-комунікаційних технологій ім. О. О. Зеленського №504

(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	<p>Галузь знань <u>12«Інформаційні технології»</u> (шифр і найменування)</p> <p>Спеціальність <u>126«Інформаційні системи та технології»</u> (код і найменування)</p> <p>Освітня програма <u>«Штучний інтелект та інформаційні системи»</u> (найменування)</p> <p>Рівень вищої освіти: перший (бакалаврський)</p>	<i>Обов'язкова</i>
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2023/2024
Індивідуальне завдання – <u>не передбачено навчальним планом</u>		Семестр
Загальна кількість годин – 64*/150		5/3**
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи здобувача – 5,4		Лекції*
		<u>32</u> години
		Практичні*
		<u>32</u> години
		Лабораторні*
	-	
Самостійна робота		
86 годин		
Вид контролю		
модульний контроль, залік		

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.

** група зі скороченим терміном навчання

2. Мета та завдання навчальної дисципліни

Мета вивчення: набуття студентами знань, основних принципів і практичних навичок в області аналізу даних і машинного навчання для вирішення прикладних задач у різних галузях в умовах широкого використання сучасних інформаційних технологій.

Завдання: вивчення основних алгоритмів машинного навчання і оволодіння практичними навичками побудови моделей машинного навчання з використанням мови програмування Python.

Компетентності, які набуваються:

загальні:

- здатність застосовувати знання у практичних ситуаціях;
- здатність вчитися і оволодівати сучасними знаннями;
- здатність розробляти та управляти проектами;
- здатність оцінювати та забезпечувати якість виконуваних робіт;

спеціальні (фахові):

- здатність аналізувати об'єкт проектування або функціонування та його предметну область;
- здатність до проектування, розробки, налагодження та вдосконалення системного, комунікаційного та програмно-апаратного забезпечення інформаційних систем та технологій, Інтернету речей (IoT), комп'ютерно-інтегрованих систем та системної мережної структури, управління ними;
- здатність проектувати, 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 функціональних завдань та обов'язків;
- здатність до аналізу, синтезу і оптимізації інформаційних систем та технологій з 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 проектах (стартапах).

Очікувані результати навчання:

- застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проектування і використання інформаційних систем та технологій;
- 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 проектних матеріалів інформаційних систем та технологій, знати склад та послідовність виконання проектних робіт з урахуванням вимог відповідних нормативно-правових документів для запровадження у професійній діяльності.

Пререквізити – “Основи програмування”, “Алгоритми і структури даних”, “Системний аналіз і математичне моделювання інформаційних систем”.

Кореквізити – “Машинне навчання і аналіз даних”, “Машинне навчання і аналіз даних (КП)”, “Методи оптимізації в машинному навчанні”.

3. Зміст навчальної дисципліни

Модуль 1.

Змістовний модуль 1. Основи теорії машинного навчання

Тема 1. Вступ до машинного навчання. Предмет вивчення і задачі дисципліни. Місце дисципліни в учбовому плані. Штучний інтелект і машинне навчання як його складова частина. Основні поняття машинного навчання: дані, ознаки, алгоритми. Основні позначення. Постановка завдання машинного навчання. Класифікація методів машинного навчання. Приклади прикладних завдань машинного навчання. Методологія машинного навчання CRISP-DM. Основні інструменти і бібліотеки.

Тема 2. Лінійна алгебра і бібліотеки Python. Лінійна алгебра у машинному навчанні. Вектори, векторний простір, матриці і операції над ними. Розв'язання систем лінійних рівнянь. Власні вектори і власні значення. Бібліотеки Python (NumPy, SciPy) і лінійна алгебра.

Тема 3. Обробка і аналіз даних за допомогою бібліотеки Pandas. Дані. Основи роботи з даними у бібліотеці Pandas. Основні структури даних бібліотеки Pandas. Основи попередньої обробки і аналізу даних. Завантаження, запис, очищення, фільтрація, обробка і групування даних у pandas.

Тема 4. Основи статистики і візуалізації даних. Основи статистики. Імовірнісні розподіли. Перевірка гіпотез. Оцінка найбільшої вірогідності. Візуалізація даних з використанням бібліотек Matplotlib, Seaborn, Pandas.

Тема 5. Лінійна регресія. Лінійні моделі машинного навчання. Регресія. Формулювання завдання лінійної регресії. Метод найменших квадратів (МНК). Зведення до задачі оптимізації. Функція втрат. Аналітичний розв'язок МНК. Наближений чисельний метод розв'язання МНК. Алгоритм градієнтного спуску. Стохастичний градієнтний спуск. Теорема Гаусса-Маркова. Проблеми аналітичного методу МНК. Проблема мультиколінеарності. Розкладання помилки прогнозу лінійної регресії на зміщення (bias) і розкид (variance). Проблема недонавчання і перенавчання моделі. Регуляризація лінійної регресії. Регресія LASSO. Гребенева (Ridge) регресія. ElasticNet. Критерії якості моделі. Метрики регресії: середньоквадратична похибка (MSE, RMSE), коефіцієнт детермінації, середня абсолютна похибка (MAE), середня абсолютна відсоткова помилка (MAPE). Вибір метрики.

Тема 6. Поліноміальна регресія. Модель поліноміальної регресії. Масштабування ознак. Стандартизація ознак. Нормалізація ознак. Крос-валідація моделі. Розбиття вибірки на навчальну і тестову вибірки.

Модульний контроль

Змістовний модуль 2. Методи класифікації і метричні методи навчання

Тема 1. Логістична регресія. Лінійна класифікація. Постановка задачі лінійної класифікації. Бінарна класифікація. Логістична регресія як лінійний класифікатор. Логістична функція. Оцінка максимальної правдоподібності і логістична регресія. Регуляризація логістичної регресії. Багатокласова класифікація. Багатокласова логістична регресія. Метрики класифікації: метрика Accuracy, матриця помилок (confusion matrix), точність (precision), повнота (recall), міра F1, TPR (true positive rate), FPR (false positive rate), AUC, крива ROC.

Тема 2. Наївний баєсів класифікатор. Імовірнісна модель класифікатора. Теорема Байєса. Наївний класифікатор Байєса.

Тема 3. Дерева рішень. Визначення вирішального дерева. Критерії розгалуження. Критерій інформативності у задачі регресії. Критерій інформативності у задачі класифікації. Критерій Джині (Gini index). Методи регуляризації дерев рішень.

Тема 4. Метод найближчого сусіда. Метричні методи навчання. Постановка задачі класифікації у випадку використання метода найближчих сусідів. Метод кнайближчих сусідів

(k-nearest neighbors (KNN)). Метрики відстані. Пошук найближчих сусідів. Постановка задачі регресії у випадку використання метода найближчих сусідів. Ядерна регресія. Формула Надарая–Ватсона.

Тема 5. Кластеризація. Метод k–середніх. Навчання без учителя. Задача кластеризації. Кластеризація методом k-середніх. Вибір початкового наближення і метрики відстані.

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. Основи теорії машинного навчання					
Тема 1. Вступ до машинного навчання	6	2	1	-	3
Тема 2. Лінійна алгебра і бібліотеки Python	16	4	4	-	8
Тема 3. Обробка і аналіз даних за допомогою бібліотеки Pandas	16	4	4	-	8
Тема 4. Основи статистики і візуалізації даних	9	2	1	-	6
Тема 5. Лінійна регресія	19	6	5	-	8
Тема 6. Поліноміальна регресія	6	1	1	-	4
Модульний контроль	4	1	-	-	3
Разом за змістовним модулем 1	76	20	16	-	40
Змістовний модуль 2. Методи класифікації і метричні методи навчання					
Тема 1. Логістична регресія	16	4	4	-	8
Тема 2. Наївний баєсів класифікатор	12	2	2	-	8
Тема 3. Дерева рішень	16	2	4	-	10
Тема 4. Метод найближчого сусіда	14	2	4	-	8
Тема 5. Кластеризація. Метод k–середніх	12	1	2	-	9
Модульний контроль	4	1	-	-	3
Разом за змістовним модулем 2	74	12	16		46
Усього годин	150	32	32	-	86
Модуль 2					
Індивідуальне завдання	-	-	-	-	-
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	Бібліотеки Python лінійна алгебра	4
2	Обробка даних за допомогою бібліотеки Pandas	6
3	Візуалізація даних	2
4	Лінійна регресія	6
5	Поліноміальна регресія	2
6	Логістична регресія	2
7	Наївний класифікатор Байеса	2
8	Дерева рішень	4
9	Метод найближчого сусіда	2
10	Кластеризація методом k-середніх	2
	Разом	32

7. Теми лабораторних занять

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

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1	Опрацювання матеріалу лекцій	46
2	Підготовка до практичних занять	25
3	Опрацювання матеріалів та результатів отриманих на практичних заняттях	15
	Разом	86

9. Індивідуальні завдання

Не передбачене навчальним планом.

10. Методи навчання

При викладанні курсу використовуються наступні навчальні методи:

- демонстрація;
- ілюстрація;
- розповідь;
- спостереження;
- дослідження;
- практичне заняття;
- виконання завдань.

11. Методи контролю

- 1) поточний контроль (оцінювання роботи студентів на практичних заняттях);
- 2) модульний контроль за змістовними модулями;
- 3) семестровий контроль у вигляді заліку.

12. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Виконання і захист практичних робіт	0...4	5	0...20
Модульний контроль	0...30	1	0...30
Змістовний модуль 2			
Виконання і захист практичних робіт	0...4	5	0...20
Модульний контроль	0...30	1	0...30
Усього за семестр			0...100

Семестровий контроль (залік) проводиться у разі відмови студента від балів поточного контролю і за наявності допуску до заліку. Під час складання семестрового заліку студент має можливість отримати максимум 100 балів.

Критерії оцінювання роботи здобувача протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити всі практичні завдання і здати тестування. Знати і вміти побудувати просту лінійну модель регресії і класифікації за допомогою бібліотеки scikit-learn.

Добре (75-89). Твердо знати мінімум, захистити всі практичні завдання, здати тестування та поза аудиторну самостійну роботу. Вміти все, що вказано у попередньому пункті і вміти обирати метод машинного навчання відповідно до поставлених вимог та особливостей задачі, що вирішується. Вміти самостійно реалізовувати лінійні алгоритми машинного навчання.

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

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. Навчально-методичне забезпечення дисципліни "Прикладний аналіз даних на Python" для бакалаврів / М-во освіти і науки України, Нац. аерокосм. ун-т ім. М. Є. Жуковського "Харків. авіац. ін-т", Каф. вищ. математики та систем. аналізу (№ 405) ; розроб. В. О. Макарічев. - Харків. - Нац. аерокосм. ун-т ім. М. Є. Жуковського "Харків. авіац. ін-т", 2020. - 177 с. - <http://library.khai.edu/library/fulltexts/2021/complex/>

2. Маттес Е. Пришвидшений курс Python: практичний, проєктно-орієнтований вступ до програмування / Е. Маттес. - Львів. - Вид-во Старого Лева, 2021. - 600 с.

3. Навчально-методичне забезпечення дисципліни "Програмування на PYTHON" для бакалаврів / М-во освіти і науки України, Нац. аерокосм. ун-т ім. М. Є. Жуковського "Харків. авіац. ін-т", Каф. вищ. математики та систем. аналізу (№ 405) ; розроб. В. О. Макарічев. - Харків. - Нац. аерокосм. ун-т ім. М. Є. Жуковського "Харків. авіац. ін-т", 2020. - 133 с. - <http://library.khai.edu/library/fulltexts/2021/complex/>

14. Рекомендована література

Базова

1. Басюк, Т. М. Машинне навчання : навчальний посібник призначений для студентів, що навчаються за першим (бакалаврським) рівнем вищої освіти за спеціальностями галузі знань 12 "Інформаційні технології" /Т.М. Басюк, В.В. Литвин, Л.М. Захарія, Н.Е. Кунанець; Національний університет "Львівська політехніка", Інститут комп'ютерних наук та інформаційних технологій, Кафедра інформаційних систем та мереж. – Львів :Видавництво "Новий світ-2000", 2021. – 329 с.

Допоміжна

1. Hastie T., Friedman J., Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2017.
URL:<https://web.stanford.edu/~hastie/ElemStatLearn>.
2. Bishop C. Pattern recognition and machine learning. Springer, 2006. URL:
<https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/>

15. Інформаційні ресурси

1. <https://www.python.org/doc/>
2. <https://jupyter.org/>
3. Google Colab, <https://colab.research.google.com/notebooks/welcome.ipynb?hl=uk>
4. <https://numpy.org/doc/stable/>
5. <https://matplotlib.org/stable/users/index.html>
6. Matplotlib tutorials, <https://matplotlib.org/stable/tutorials/index.html>
7. Pandas documentation, <https://pandas.pydata.org/docs/>
8. Seaborn user guide and tutorial, <https://seaborn.pydata.org/tutorial.html>
9. Бібліотека scikit-learn, <https://scikit-learn.org/stable/>