

Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Кафедра теоретичної та прикладної системотехніки



Проректор
з науково-педагогічної роботи
Олександр ГОЛОВКО

“ЗАТВЕРДЖУЮ”

2022 р.

Робоча програма навчальної дисципліни

Методи машинного навчання

рівень вищої освіти другий (магістерський)

спеціальність 151 «Автоматизація та комп'ютерно-інтегровані технології»

освітня програма Комп'ютеризовані системи автоматизації та управління

вид дисципліни за вибором

факультет комп'ютерних наук

2022 / 2023 навчальний рік

Програму рекомендовано до затвердження вченою радою факультету комп'ютерних наук
«28» червня 2022 року, протокол №10

РОЗРОБНИКИ ПРОГРАМИ:

кандидат технічних наук, доцент кафедри теоретичної та прикладної системотехніки

Стрілець Вікторія Євгенівна

Програму схвалено на засіданні кафедри теоретичної та прикладної системотехніки

Протокол від «11» червня 2022 року, №12

Завідувач кафедри теоретичної та прикладної системотехніки

 Сергій ШМАТКОВ.

Програму погоджено з гарантом освітньої програми «Комп'ютеризовані системи управління
та автоматика»


Гарант освітньої програми «Комп'ютеризовані системи управління та
автоматика»

 Сергій ШМАТКОВ

Програму погоджено методичною комісією факультету комп'ютерних наук

Протокол від «24» червня 2022 року № 9

Голова методичної комісії факультету комп'ютерних наук

 Анатолій БЕРДНІКОВ

ВСТУП

Програма навчальної дисципліни «Методи машинного навчання» розроблена відповідно до освітньо-професійної програми підготовки другого (магістерського) рівня спеціальності 151 «Автоматизація та комп'ютерно-інтегровані технології» освітньої програми «Комп'ютеризовані системи автоматизації та управління».

1. Опис навчальної дисципліни

1.1. Метою викладання навчальної дисципліни є:

засвоєння студентами основ наукових і математичних положень, моделей і методів, що лежать в основі обробки даних і технології машинного навчання для дослідження складних систем та процесів, зокрема управління складними комп'ютерними системами; вироблення навичок з використання алгоритмів машинного навчання та програмних засобів, призначених для розв'язання задач машинного навчання.

1.2. Основними завданнями вивчення навчальної дисципліни є вивчення і набуття навичок застосування:

- методів попередньої обробки даних, отриманих під час проведення експериментів або вивчення процесів різної природи;
- методів розв'язання задач класифікації даних;
- методів ідентифікації математичних моделей систем та процесів;
- методів оцінювання інформативності (значущості) змінних;
- методів прогнозування даних.

В ході вивчення дисципліни у студента повинні формуватися такі компетентності.

Загальні компетентності (ЗК)

ЗК01. Здатність генерувати нові ідеї (креативність).

ЗК03. Здатність до абстрактного мислення, аналізу та синтезу.

Спеціальні (фахові, предметні) компетентності (ФК)

ФК01. Здатність здійснювати автоматизацію складних технологічних об'єктів та комплексів, створювати кіберфізичні системи на основі інтелектуальних методів управління та цифрових технологій з використанням баз даних, баз знань, методів штучного інтелекту, робототехнічних та інтелектуальних мехатронних пристроїв.

ФК04. Здатність аналізувати виробничо-технологічні системи і комплекси як об'єкти автоматизації, визначати способи та стратегії їх автоматизації та цифрової трансформації.

ФК07. Здатність застосовувати спеціалізоване програмне забезпечення та цифрові технології для розв'язання складних задач і проблем автоматизації та комп'ютерно-інтегрованих технологій.

ФК08. Здатність розробляти функціональну, технічну та інформаційну структуру комп'ютерно-інтегрованих систем управління організаційно-технологічними комплексами із застосуванням мережевих та інформаційних технологій, програмно-технічних керуючих комплексів, промислових контролерів, мехатронних компонентів, робототехнічних пристроїв та засобів людино-машинного інтерфейсу.

1.3. Кількість кредитів – 5

1.4. Загальна кількість годин – 150

1.5. Характеристика навчальної дисципліни
За вибором

Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
1-й	1-й
Лекції	
32 год.	год.
Практичні, семінарські заняття	
16 год.	год.
Лабораторні заняття	
0 год.	год.
Самостійна робота	
102 год.	год.
Індивідуальні завдання	
-год.	

1.6. Відповідно до вимог освітньо-кваліфікаційного рівня підготовки за результатами вивчення дисципліни студенти повинні –

знати:

- основні види машинного навчання;
- задачі, які відносяться до навчання без вчителя;
- задачі, які відносяться до навчання з вчителем
- методи машинного навчання для розв'язання задач класифікації, регресії та кластеризації даних;

уміти:

- здійснювати вибір методів машинного навчання для розв'язання задач класифікації, регресії та кластеризації даних;
- проводити верифікацію методів машинного навчання та оцінку їх якості застосування на основі існуючих критеріїв;
- розв'язувати задачі попереднього аналізу даних, класифікації, регресії та кластеризації даних із застосуванням спеціалізованих бібліотек та мов програмування;
- застосовувати підходи і методи машинного навчання для створення ефективних систем автоматизації складних процесів;
- пояснювати, кількісно та якісно оцінювати, корегувати отримані результати;

придбати навички:

- статистичної обробки даних;
- застосування методів машинного навчання для розв'язання задач класифікації, регресії та кластеризації даних;
- проведення верифікації методів, оцінки якості математичних методів на основі існуючих критеріїв;
- вирішення задач машинного навчання з застосуванням спеціалізованих бібліотек;

мати уявлення:

- про роль методів машинного навчання у створенні сучасних складних технічних систем; перспективах розвитку методів машинного навчання; про основні проблеми розробки сучасного програмного забезпечення для розв'язання задач аналізу, інтелектуальної обробки даних, управління складними комп'ютерними системами та ін.

В результаті вивчення дисципліни у студента повинні формуватися такі *програмні результати навчання (ПРН)*.

ПРН01. Створювати системи автоматизації, кіберфізичні виробництва на основі використання інтелектуальних методів управління, баз даних та баз знань, цифрових та мережових технологій, робототехнічних та інтелектуальних мехатронних пристроїв.

ПРН04. Застосовувати сучасні підходи і методи моделювання та оптимізації для дослідження та створення ефективних систем автоматизації складними технологічними та організаційно-технічними об'єктами.

ПРН05. Розробляти комп'ютерно-інтегровані системи управління складними технологічними та організаційно-технічними об'єктами, застосовуючи системний підхід із врахуванням нетехнічних складових оцінки об'єктів автоматизації.

ПРН07. Аналізувати виробничо-технічні системи у певній галузі діяльності як об'єкти автоматизації і визначати стратегію їх автоматизації та цифрової трансформації.

ПРН08. Застосовувати сучасні математичні методи, методи теорії автоматичного керування, теорії надійності та системного аналізу для дослідження та створення систем автоматизації складними технологічними та організаційно-технічними об'єктами, кіберфізичних виробництв.

ПРН10. Розробляти і використовувати спеціалізоване програмне забезпечення та цифрові технології для створення систем автоматизації складними організаційно-технічними об'єктами, професійно володіти спеціальними програмними засобами.

ПРН11. Дотримуватись норм академічної доброчесності, знати основні правові норми щодо захисту інтелектуальної власності, комерціалізації результатів науково-дослідної, винахідницької та проектної діяльності.

ПРН12. Збирати необхідну інформацію, використовуючи науково-технічну літературу, бази даних та інші джерела, аналізувати і оцінювати її.

ПРН13. Знати методи управління процесами різної природи, які побудовані на основі сучасних та перспективних методів математики, системного аналізу, штучного інтелекту.

2. Тематичний план навчальної дисципліни

Тема 1. Машинне навчання як напрям штучного інтелекту.

Етапи розвитку штучного інтелекту, виникнення і виокремлення машинного навчання. Визначення машинного навчання, його основні напрямки (з вчителем, без вчителя, з підкріпленням). Формальне поняття «навчання».

Тема 2. Задачі машинного навчання.

Задачі навчання з вчителем. Задачі навчання без вчителя. Способи навчання та оцінка його якості. Перенавчання і недонавчання, узагальнююча здатність.

Тема 3. Попередня обробка даних.

Статистичний аналіз даних. Виявлення та обробка відсутніх значень. Кодування нечислових ознак. Масштабування і стандартизація. Кодування.

Тема 4. Задача класифікації та методи її розв'язання.

Загальна постановка задачі класифікації. Наївний байєсівський класифікатор. Дерева рішень. Випадкові ліси. Логістична регресія. К-найближчих сусідів. Метод опорних векторів.

Тема 5. Задача регресії та методи її розв'язання.

Загальна постановка задачі регресії. Базова регресійна модель. Лінійна регресія. Методи нелінійної регресії (дерева рішень, випадкові ліси, метод опорних векторів).

Тема 6. Задача кластеризації та методи її розв'язання.

Загальна постановка задачі кластеризації. Метод k-середніх. Метод c-середніх. Агломеративна (ієрархічна) кластеризація.

3. Структура навчальної дисципліни

Назви розділів і тем	Кількість годин	
	Денна форма	
	Всього	у тому числі:

		Л	ПЗ	Лаб. роб.	Інд.	СР
1	2	3	4	5	6	7
Тема 1. Машинне навчання як напрям штучного інтелекту.	14	2	2			10
Тема 2. Задачі машинного навчання.	25	6	2			17
Тема 3. Попередня обробка даних.	25	6	2			17
Тема 4. Задача класифікації та методи її розв'язання.	31	6	6			21
Тема 5. Задача регресії та методи її розв'язання.	25	6	2			17
Тема 6. Задача кластеризації та методи її розв'язання.	30	6	2			20
Усього годин	150	32	16			102

4. Теми практичних занять

№ п/п	Назва теми	Кількість годин
1	Програмні засоби машинного навчання. Мова програмування Python. Бібліотеки numpy, pandas.	2
2	Пошук та завантаження наборів даних для аналізу. Статистичний аналіз.	2
3	Попередня обробка даних. Робота з відсутніми даними.	2
4	Задача класифікації. Наївний байєсівський класифікатор.	2
5	Задача класифікації. Дерева рішень і випадкові ліси.	2
6	Задача класифікації. К-найближчих сусідів. Метод опорних векторів.	2
7	Задача регресії. Лінійна регресія. Методи нелінійної регресії.	2
8	Задача кластеризації. Метод k-середніх. Ієрархічна кластеризація.	2
	Разом	16

5. Завдання для самостійної роботи

№ п/п	Зміст	Кількість годин
1	Ознайомитися з мовою програмування Python, бібліотеками машинного навчання, вбудованими наборами даних.	17
2	Розглянути існуючі підходи до опрацювання відсутніх даних у наборах даних.	17
3	Розглянути метрики оцінювання якості роботи методів машинного навчання.	21
4	Провести порівняльний аналіз методів класифікації даних.	17
5	Провести порівняльний аналіз методів регресії.	20
6	Підготовка до підсумкової контрольної роботи	10
	Разом	102

6. Індивідуальні завдання

7. Методи контролю

Контроль роботи студентів при вивченні дисципліни і засвоєння ними навчального матеріалу здійснюється на практичному занятті шляхом проведення контрольних опитувань

і захисту звітів з практичних завдань. Підсумковий контроль здійснюється при виконанні 1 контрольної роботи і відповіді на залікове завдання.

Студенти, що не виконали впродовж семестру 1 контрольну роботу, а також не представили і не захистили звіти з практичних завдань, до заліку не допускаються.

Заліковий білет містить одне теоретичне і одне практичне питання. Максимальна кількість балів за відповіді на кожне питання складає по 20 балів.

При дистанційному навчанні видача практичних завдань та контроль їх виконання здійснюється за допомогою сервісу дистанційного навчання Google Classroom. Лекційні заняття проводяться із використанням сервісу відео-конференцій Google Meet.

Підсумковий контроль у вигляді заліку проводиться шляхом розгорнутої письмової відповіді на білет.

8. Схема нарахування балів

Бали за поточний контроль знань впродовж семестру (по темах)					Контрольна робота, передбачена навчальним планом	Разом	Залік	Сума
T1, 2	T3	T4	T5	T6				
8	8	16	8	8	12	60	40	100

T1, T2 ... – теми розділів.

За темами T1, 2 студент отримує 8 балів за виконання практичної роботи 1.

За темою T3 студент отримує 8 балів за виконання практичної роботи 2.

За темою T4 студент отримує 16 балів за виконання практичних робіт 3 і 4.

За темою T5 студент отримує 8 балів за виконання практичної роботи 5.

За темою T6 студент отримує 8 балів за виконання практичної роботи 6.

Критерії оцінювання знань студентів за практичні роботи

Вимоги	Кількість балів
<ul style="list-style-type: none"> ▪ Завдання відзначається повнотою виконання без допомоги викладача. ▪ Визначає рівень поінформованості, потрібний для прийняття рішень. Вибирає інформаційні джерела. ▪ Робить висновки і приймає рішення у ситуації невизначеності. Володіє уміннями творчо-пошукової діяльності. 	8
<ul style="list-style-type: none"> ▪ Завдання – повні, з деякими огріхами, виконані без допомоги викладача. ▪ Планує інформаційний пошук; володіє способами систематизації інформації; ▪ Робить висновки і приймає рішення у ситуації невизначеності. Володіє уміннями творчо-пошукової діяльності. 	6-7
<ul style="list-style-type: none"> ▪ Завдання відзначається неповнотою виконання без допомоги викладача. ▪ Студент може зіставити, узагальнити, систематизувати інформацію під керівництвом викладача; вільно застосовує вивчений матеріал у стандартних ситуаціях. 	4-5
<ul style="list-style-type: none"> ▪ Завдання відзначається неповнотою виконання за консультацією викладача. 	3

<ul style="list-style-type: none"> ▪ Застосовує запропонований вчителем спосіб отримання інформації, має фрагментарні навички в роботі з підручником, науковими джерелами; ▪ Вибирає відомі способи дій для виконання фахових методичних завдань. 	
Завдання відзначається фрагментарністю виконання за консультацією викладача або під його керівництвом.	1-2

Критерії оцінювання знань студентів за контрольну роботу

Вимоги	Кількість балів
Повнота виконання завдання повна, студент здатен формулювати закони та закономірності, структурувати судження, умовиводи, доводи, описи.	9-12
Повнота виконання завдання повна, студент здатен формулювати операції, правила, алгоритми, правила визначення понять.	5-8
Повнота виконання завдання елементарна, студент здатен вибирати відомі способи дій для виконання фахових завдань.	3-5
Повнота виконання завдання фрагментарна.	1-2

Критерії оцінювання залікових робіт студентів

Вимоги	Кількість балів
Показані всебічні систематичні знання та розуміння навчального матеріалу; безпомилково виконані завдання.	35-40
Показані повні знання навчального матеріалу; помилки, якщо вони є, не носять принципового характеру.	30-35
Показано повне знання необхідного навчального матеріалу, але допущені помилки.	20-30
Показано повне знання необхідного навчального матеріалу, але допущені суттєві помилки	10-20
Показано недосконале знання навчального матеріалу, допущені суттєві помилки.	5-10
Показано недосконале знання навчального матеріалу, допущені суттєві помилки, які носять принциповий характер; обсяг знань не дозволяє засвоїти предмет.	1-5

Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка	
	для чотирирівневої шкали оцінювання	для дворівневої шкали оцінювання
90 – 100	відмінно	зараховано
70-89	добре	
50-69	задовільно	
1-49	незадовільно	не зараховано

9. Рекомендована література

Основна література

1. MacKay David J.C. Information Theory, Inference and Learning Algorithms / David J.C. MacKay. – Cambridge University Press, 2003. – 628 p.

2. Bishop Christopher M. Pattern Recognition and Machine Learning / Christopher M. Bishop. – New York: Springer, 2006. – 738 p.
3. VanderPlas J. Python Data Science HandBook. Essential tools for working with data / Jake VanderPlas. – O'Reilly, 2018. – 576 p.
4. Andrew Ng. Machine Learning Yearning. – 2018. – 118 p.
5. Rasmussen C.E. Gaussian Processes for Machine Learning / C.E. Rasmussen, C.K.I. Williams. – Cambridge, Massachusetts: MIT Press, 2006. – 248 p.
6. Merphy K. Machine Learning: A Probabilistic Perspective / Kevin Merphy. (<https://mitpress.mit.edu/books/machine-learning-0>).

Допоміжна

1. Довбиш А.С. Основи теорії розпізнавання образів: навч. посіб / А.С. Довбиш, І.В. Шелехов. – Суми: Сумський державний університет, 2015. – 109 с.
2. Sutton Richard S., Barto Andrew G. Reinforcement Learning: An Introduction. – Cambridge, Massachusetts: MIT Press, 2018. – 426 p.
3. Hastie Trevor, Tibshirani Robert, Friedman Jerome. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. – New York: Springer, 2009. – 745 p.

10. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. <https://www.datacamp.com/courses/deep-learning-in-python>
2. <https://www.kaggle.com/learn/machine-learning>
3. Datasets. – <https://www.kaggle.com>