

КРИВОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Кафедра моделювання та програмного забезпечення



ЗАТВЕРДЖУЮ

Перший проректор

Владислав ЧУБАРОВ

2023 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Системи штучного інтелекту

спеціальність 121 – «Інженерія програмного забезпечення»
(шифр і назва напрямку підготовки)

факультет Інформаційних технологій
(назва інституту, факультету, відділення)

Форма навчання	Курс	Семестр	Всього годин за планом	Кількість національних кредитів	Всього аудиторних годин	Аудиторних годин, (у тому числі КЗ)		Самостійна робота (год.)	Контрольно-модульні роботи	Залік (сем.)	Екзамен (сем.)
						Лекції	Лабораторні				
Денна	3	6	135	4,5	72	36	36	63	2	*	—
	4	7	150	5	48	32	16	102	1	—	*
Всього			285	9,5	120	68	52	165			
Денна скорочена	2	4	135	4,5	72	36	36	63	2	*	—
	3	5	150	5	48	32	16	102	1	—	*
Всього			285	9,5	120	68	52	165			
Заочна	3	6	135	4,5	14	8	6	121	2	*	—
	4	7	150	5	10	6	4	140	1	—	*
Всього			285	9,5	24	14	10	261			
Заочна скорочена	2	4	135	4,5	14	8	6	121	2	*	—
	3	5	150	5	10	6	4	140	1	—	*
Всього			285	9,5	24	14	10	261			

Кривий Ріг – 2023 рік

Робочу програму навчальної дисципліни «Системи штучного інтелекту» для здобувачів першого (бакалаврського) рівня вищої освіти за освітньою програмою «Інженерія програмного забезпечення» розроблено згідно з ОПП галузі знань 12 «Інформаційні технології» зі спеціальності 121 «Інженерія програмного забезпечення».

Розробники: старший викладач кафедри МПЗ Шаповалова Н. Н.

Робоча програма затверджена на засіданні кафедри моделювання та програмного забезпечення

Протокол від "31" травня 2023 року № 7

Завідувач кафедри МПЗ, доцент, к.п.н.



Андрій СТРІЮК

Схвалено вченою радою факультету інформаційних технологій

Протокол від "27" 02 2023 року № 6

Голова вченої ради



Іван МУЗИКА

Схвалено групою забезпечення ОПП

Протокол від "31" травня 2023 року № 7

Гарант ОПП



Андрій СТРІЮК

ЗМІСТ

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ	4
2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ	5
3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ	7
4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ	8
5 ТЕМИ ПРАКТИЧНИХ ЗАНЯТЬ	9
6 ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ	9
7 САМОСТІЙНА РОБОТА	10
8 МЕТОДИ НАВЧАННЯ	12
9 МЕТОДИ КОНТРОЛЮ	13
10 РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ ЗДОБУВАЧІ	13
12 ПЕРЕЛІК ПИТАНЬ ДЛЯ ПІДСУМКОВОГО КОНТРОЛЮ ЗНАНЬ	18
11 ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ	
13 НАВЧАЛЬНО-МЕТОДИЧНІ МАТЕРІАЛИ З ДИСЦИПЛІНИ	20
14 ІНФОРМАЦІЙНІ РЕСУРСИ	21
15 ТЕРМІНОЛОГІЧНИЙ СЛОВНИК	21
16 ЗМІНИ ТА ДОПОВНЕННЯ ДО РОБОЧОЇ ПРОГРАМИ	26
Додаток до робочої програми. Робочий план	27

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, освітній рівень	Характеристика навчальної дисципліни			
		денна форма навчання	заочна форма навчання	денна скорочена форма навчання	заочна скорочена форма навчання
Кількість кредитів – 9,5	Галузь знань 12 <u>Інформаційні технології</u> (шифр і назва)	нормативна			
Модулів – 3	Спеціальність: <u>121 Інженерія програмного забезпечення</u> (код та найменування спеціальності)	Рік підготовки:			
Змістових модулів – 3		3,4-й	3,4-й	2,3-й	2,3-й
Загальна кількість годин – 285		Семестр			
		6,7-й	6,7-й	4,5-й	4,5-й
Тижневих годин для денної форми навчання: аудиторних – 4/3 самостійної роботи студента – 5	Ступінь вищої освіти: <u>бакалавр</u>	Лекції			
		68 год.	14 год.	68 год.	14 год.
		Практичні, семінарські			
		-	-	-	-
		Лабораторні			
		52 год.	10 год.	52 год.	10 год.
		Самостійна робота			
165 год.	261 год.	165 год.	261 год.		
Вид контролю: залік, екзамен					

Примітка

Співвідношення кількості годин аудиторних занять до самостійної роботи становить:
 для денної форми навчання – 0,72;
 для заочної форми навчання – 0,09;
 для скороченої форми навчання – 0,72
 для заочної скороченої форми навчання – 0,09;

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

2.1 **Метою** викладання навчальної дисципліни «Системи штучного інтелекту» є формування теоретичних знань та практичних навичок щодо використання понять штучного інтелекту, пошуку рішення у просторі станів, інтелектуальних агентів, машинного навчання та розгортання побудованих моделей у вигляді web-застосунку.

2.2 Основними **завданнями** вивчення навчальної дисципліни «Системи штучного інтелекту» є набуття теоретичних знань та практичних умінь з формування базового уявлення про галузі застосування систем штучного інтелекту; набуття вмінь і навичок розв'язання задач з використанням систем штучного інтелекту; опанування теоретичних і практичних питань створення та застосування систем штучного інтелекту; вивчення механізмів обробки і подання знань в інтелектуальних системах.

2.3 Відповідно до освітньої програми дисципліна забезпечує наступні **компетентності**:

Загальні компетентності

ЗК01 Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02 Здатність застосовувати знання у практичних ситуаціях.

ЗК05 Здатність вчитися і оволодівати сучасними знаннями.

ЗК06 Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

Фахові компетентності

СК01 Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення.

СК02 Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

СК03 Здатність розробляти архітектури, модулі та компоненти програмних систем.

СК04 Здатність формулювати та забезпечувати вимоги щодо якості програмного забезпечення у відповідності з вимогами замовника, технічним завданням та стандартами.

СК05 Здатність дотримуватися специфікацій, стандартів, правил і рекомендацій в професійній галузі при реалізації процесів життєвого циклу.

СК06 Здатність аналізувати, вибирати і застосовувати методи і засоби для забезпечення інформаційної безпеки (в тому числі кібербезпеки).

СК10 Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.

СК11 Здатність реалізовувати фази та ітерації життєвого циклу програмних систем та інформаційних технологій на основі відповідних моделей і підходів розробки програмного забезпечення.

СК12 Здатність здійснювати процес інтеграції системи, застосовувати стандарти і процедури управління змінами для підтримки цілісності, загальної функціональності і надійності програмного забезпечення.

СК13 Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.

СК14 Здатність до алгоритмічного та логічного мислення.

Програмні результати навчання освітньої програми, яким відповідає дисципліна:

ПР01 Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

ПР02 Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.

ПР05 Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.

ПР07 Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.

ПР11 Вибирати вихідні дані для проектування, керуючись формальними методами опису вимог та моделювання.

ПР13 Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.

ПР18 Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.

ПР19 Знати та вміти застосовувати методи верифікації та валідації програмного забезпечення.

В результаті вивчення дисципліни здобувачі повинні **знати:**

- типи задач машинного навчання;
- основні поняття машинного навчання;
- принципи побудови регресійних і класифікаційних моделей;
- методи навчання моделей машинного навчання;
- поняття перенавчання моделей та засоби боротьби з ним;
- принципи побудови вирішальних дерев та методів їх навчання;
- основні композиційні алгоритми;
- засади функціонування метричних алгоритмів.

вміти:

- будувати моделі машинного навчання;
- застосовувати спеціалізовані бібліотеки мови програмування Python для проектування моделей машинного навчання;
- застосовувати методи навчання моделей;
- використовувати засоби боротьби з перенавчанням моделей;
- виконувати валідацію моделей;
- застосовувати композиційні алгоритми і налаштовувати їх параметри;
- застосовувати високорівневі фреймворки машинного навчання;
- самостійно опановувати нові фреймворки та методи навчання моделей, використовувати документацію.

2.4 Міждисциплінарні зв'язки

При вивченні дисципліни використовуються знання здобувачів з дисциплін «Вища математика», «Теорія ймовірностей і математична статистика», «Основи програмування», «Алгоритми та структури даних», «Об'єктно-орієнтоване програмування», «Сучасні технології Internet-програмування», «Англійська мова (за професійним спрямуванням)».

Знання, одержані здобувачами при вивченні дисципліни, використовуються при вивченні дисциплін «Нейромережеві технології», «Технології інтелектуальних систем».

Вимоги до знань та умінь визначаються галузевими стандартами вищої освіти України.

3 ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Принципи побудови моделей машинного навчання

Тема 1. Загальні положення систем штучного інтелекту. Базові поняття штучного інтелекту. Визначення штучного інтелекту за М. Мінським. Тест Тюрінга. Галузі застосування штучного інтелекту. Сприйняття і розпізнавання образів. Логічні ігри. Розв'язання задач. Розуміння природної мови. Робототехніка. Класифікація систем штучного інтелекту.

Тема 2. Введення у машинне навчання. Типи задач машинного навчання. Навчання на розмічених даних (supervised learning). Навчання без учителя (unsupervised learning). Типи ознак в машинному навчанні. Поняття навчальної вибірки і відповіді у задачах навчання з учителем.

Тема 3. Лінійні моделі. Лінійні моделі в задачах регресії. Навчання моделей лінійної регресії. Градієнтний спуск для лінійної регресії. Стохастичний градієнтний спуск. Лінійна класифікація. Функція втрат в задачах класифікації.

Тема 4. Боротьба з перенавчанням моделей. Проблема перенавчання. Регуляризація. Оцінювання якості алгоритмів. Метод крос-валідації. Порівняння алгоритмів і підбір гіперпараметрів. Метрики якості в задачах регресії. Метрики якості в задачах класифікації. Якість оцінок приналежності до класу.

Змістовий модуль 2. Вирішальні дерева та їх композиції

Тема 5. Вирішальні дерева. Навчання вирішальних дерев. Критерії інформативності. Критерії зупину і стрижки дерев. Вирішальні дерева і категоріальні ознаки. Композиції дерев. Зміщення і розкид.

Тема 6. Градієнтний бустінг. Випадкові ліси. Композиції простих алгоритмів. Метод градієнтного бустінгу. Боротьба з перенавчанням в градієнтному бустінгу. Градієнтний бустінг для регресії і класифікації. Градієнтний бустінг над вирішальними деревами.

Змістовий модуль 3. Огляд методів та розгортання моделей штучного інтелекту

Тема 7. Байєсівський класифікатор. Теорема Баяса. Байєсівський підхід до теорії ймовірностей. Байєсівські моделі в задачах машинного навчання. Відновлення розподілів. Мінімізація ризиків. Мінімізація ризиків і аналіз функцій втрат.

Тема 8. Метричні алгоритми і метод опорних векторів. Метод k найближчих сусідів. Налаштування параметрів в kNN. Метрики в kNN. Прокляття розмірності. Рекомендуючі системи на основі kNN.

Тема 9. Метод опорних векторів (SVM). Ядра в методі опорних векторів. Розв'язання задач машинного навчання без вчителя. Задача кластеризації. SVM-регресія.

Тема 10. Розгортання моделі. Використання бібліотеки PyCaret для порівняння моделей. Налаштування моделей. Створення front-end частини застосунку. Фреймворк для створення web-застосунків Flask. Платформа для надання доступу до локальних серверів Ngrok.

4 СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин							
	денна / денна скорочена форма				заочна / заочна скорочена форма			
	усього	у тому числі			усього	у тому числі		
		л	лаб	с.р.		л	лаб	с.р.
1	2	3	4	5	6	7	8	9
Модуль 1								
Змістовий модуль 1. Принципи побудови моделей машинного навчання								
Тема 1. Загальні положення систем штучного інтелекту	11	4	6	1	11	2	–	9
Тема 2. Введення у машинне навчання	16	6	6	4	16	–	–	16
Тема 3. Лінійні моделі	27	6	6	15	27	2	2	23
Тема 4. Боротьба з перенавчанням моделей	25	6	6	13	25	–	–	25
Разом за змістовим модулем 1	79	22	24	33	79	4	2	73
Модуль 2								
Змістовий модуль 2. Вирішальні дерева та їх композиції								
Тема 5. Вирішальні дерева	29	8	6	15	29	2	2	25
Тема 6. Градієнтний бустінг	27	6	6	15	27	2	2	23
Разом за змістовим модулем 2	56	14	12	30	56	4	4	48
Разом за семестр	135	36	36	63	135	8	6	121
Модуль 3								
Змістовий модуль 3. Огляд методів та розгортання моделей штучного інтелекту								
Тема 7. Баєсівський класифікатор	24	6	4	14	24	2	2	20
Тема 8. Метричні алгоритми	24	8	4	12	24	–	–	24
Тема 9. Метод опорних векторів	24	8	4	12	24	2	–	22
Тема 10. Розгортання моделі	78	10	4	64	78	2	2	74
Разом за змістовим модулем 3	150	32	16	102	150	6	4	140
Разом за семестр	150	32	16	102	150	6	4	140
Усього годин	285	68	52	165	285	14	10	261

5 ТЕМИ ПРАКТИЧНИХ ЗАНЯТЬ

Не передбачено.

6 ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин	
		денна/денна скорочена	заочна/заочна скорочена
1.	Початок роботи з пакетом обчислювальної математики NumPy і пакетом для наукових і інженерних розрахунків SciPy	4	–
2.	Препроцесінг даних	4	–
3.	Візуалізація даних в Pandas	4	–
4.	Лінійна регресія і основні бібліотеки Python для аналізу даних і наукових обчислень	4	–
5.	Лінійна регресія і стохастичний градієнтний спуск	2	2
6.	Лінійна регресія: перенавчання і регуляризація	4	–
7.	Перенавчання моделей і боротьба з ним	2	–
8.	Вирішальні дерева	6	2
9.	Метод градієнтного бустінгу своїми руками	6	2
РАЗОМ ЗА СЕМЕСТР:		36	6
10.	Використання акселераторів обчислень у Google Colaboratory	2	–
11.	Розробка спам-фільтру на основі баєсівського класифікатора	2	1
12.	Розробка рекомендаційної системи на основі метричних алгоритмів	2	1
13.	Використання алгоритмів колаборативної фільтрації для рекомендаційних систем	2	–
14.	Вирішення задачі кластеризації. Ядерні трюки методу SVM	2	–
15.	Реалізація задачі регресії методом опорних векторів	2	–
16.	Робота з бібліотекою PyCaret. Оптимізація моделей	2	–
17.	Розгортання моделі у web-застосунок. Фреймворк Flask	2	2
РАЗОМ ЗА СЕМЕСТР:		16	4
Усього годин		52	10

7 САМОСТІЙНА РОБОТА

На самостійну роботу студентам денної і скороченої форми навчання відведено 183 години, заочної – 261 година.

Самостійна робота студентів при вивченні дисципліни «Системи штучного інтелекту» залучає такі складові:

- опрацювання лекційного матеріалу з кожної теми;
- опрацювання літератури по темі;
- вивчення окремих тем або питань, що передбачені для самостійного опрацювання;
- підготовка до виконання, а також до захисту лабораторних робіт;
- підготовка до проведення контрольних заходів.

Розподіл годин самостійної роботи

№ з/п	Назва теми	Кількість годин			
		денна	заочна	денна скорочена	заочна скорочена
1.	Основні поняття в області штучного інтелекту.	1	4	1	4
2.	Огляд стану проблем штучного інтелекту.	2	4	2	4
3.	Основні напрямки досліджень в області штучного інтелекту.	2	5	2	5
4.	Постановка завдань навчання по прецедентах. Об'єкти і ознаки. Типи шкал: бінарні, номінальні, порядкові, кількісні.	2	5	2	5
5.	Типи задач: класифікація, регресія, прогнозування, кластеризація. Приклади прикладних задач.	2	5	2	5
6.	Модель алгоритмів, метод навчання, функція втрат і функціонал якості, принцип мінімізації емпіричного ризику, узагальнююча здатність.	2	6	2	6
7.	Методика експериментального дослідження та порівняння алгоритмів на модельних і реальних даних.	2	6	2	6
8.	CRISP-DM – міжгалузевий стандарт ведення проектів інтелектуального аналізу даних.	2	5	2	5
9.	Лінійний класифікатор, безперервні апроксимації порогової функції втрат. Зв'язок з методом максимуму правдоподібності.	2	5	2	5
10.	Метод стохастичного градієнта.	2	6	2	6
11.	Теорема Новікова про збіжність.	2	4	2	4
12.	Евристики: ініціалізація ваг, порядок пред'явлення об'єктів, вибір величини градієнтного кроку, «вибивання» з локальних мінімумів.	2	5	2	5
13.	Метод стохастичного середнього градієнта SAG.	1	6	1	6
14.	Проблема мультиколінеарності і перенавчання.	1	7	1	7
	Підготовка до контрольних заходів (модуль 1)	8	–	8	–
Разом за змістовим модулем 1:		33	73	33	73
15.	Принцип максимуму спільного правдоподібності даних і моделі. Квадратичний (гаусівський) і лапласівський регуляризатор.	2	4	2	4

№ з/п	Назва теми	Кількість годин			
		денна	заочна	денна скорочена	заочна скорочена
16.	Налаштування порога вирішального правила за критерієм кількості помилок I і II роду	2	6	2	6
17.	Крива помилок (ROC curve).	2	4	2	4
18.	Алгоритм ефективного побудови ROC-кривої.	2	6	2	6
19.	Градієнтний метод максимізації AUC.	2	4	2	4
20.	Обмеження лінійних методів (приклад: XOR).	2	6	2	6
21.	Вирішальні дерева.	4	6	4	6
22.	Ансамблі. Бутстреп. Бегінг. Випадковий ліс.	2	6	2	6
23.	AdaBoost, градієнтний бустінг. XGBoost	4	6	4	6
	Підготовка до контрольних заходів (модуль 2)	8	-	8	-
Разом за змістовим модулем 2:		30	48	30	48
24.	Бассівська регуляризація.	16	18	16	18
25.	Метод найближчих сусідів і його узагальнення.	14	18	14	18
26.	Метод найближчих сусідів (kNN).	16	20	16	20
27.	Підбір числа сусідів за критерієм змінного контролю.	16	16	16	16
28.	Узагальнений метричний класифікатор, поняття відступу.	16	20	16	20
29.	Метод потенційних функцій, градієнтний алгоритм.	16	18	16	18
	Підготовка до контрольних заходів (модуль 3)	8	30	8	30
Разом за змістовим модулем 3:		102	140	102	140
Разом:		165	261	165	261

8 МЕТОДИ НАВЧАННЯ

Використовуються наступні методи навчання: лекції, лабораторні заняття, самостійна робота.

Навчальна лекція – це логічне, послідовне викладання змісту навчання, яке характеризується судженнями, висновками, підсумком. Вона охоплює основний теоретичний матеріал однієї або кількох тем навчальної дисципліни. Призначенням лекції є формування у здобувачів фундаментальних знань з дисципліни, а також визначає основний зміст і характер усіх інших навчальних занять та самостійної роботи здобувачів із цієї дисципліни.

Лабораторне заняття – форма організації навчання, яку проводять за завданням і під керівництвом НПП. Основні дидактичні цілі – експериментальне підтвердження вивчених теоретичних положень навчальної дисципліни та формування вмінь й навичок їх практичного застосування. Проведення лабораторного заняття ґрунтується на попередньо підготовлених наборах завдань різної складності для розв’язання на занятті. Лабораторне

заняття проводиться у навчальних лабораторіях з використанням пристосованого до умов навчального процесу устаткування.

Самостійна робота здобувача є основним способом оволодіння навчальним матеріалом у час, вільний від обов'язкових аудиторних занять. Мета виконання самостійної роботи – поглиблення, узагальнення й закріплення теоретичних знань і практичних умінь здобувачів із дисципліни шляхом вироблення вміння самостійної роботи з навчальною і фаховою літературою та інформацією в мережі Інтернет.

Самостійна робота здобувачів здійснюється у формі: підготовки до лекцій і лабораторних занять, опрацюванні тем, винесених на самостійне опрацювання. Самостійну роботу здобувач може виконувати у бібліотеці, комп'ютерних класах, а також у домашніх умовах.

Підготовка до лекцій передбачає самостійне опрацювання теоретичного матеріалу. При цьому необхідно звернути увагу на необхідність чіткого засвоєння основних термінів та визначень, розуміння їх змісту, обов'язкового аналізу використання теоретичних положень для розв'язання наданих прикладів.

Самоперевірку засвоєння навчального матеріалу здобувач здійснює за контрольними запитаннями, що надано після кожної теми у конспекті лекцій та іншій літературі, та після кожного лабораторного заняття у відповідних методичних вказівках. Якщо на деякі запитання здобувач не може надати відповіді, то необхідно повторити вивчення навчального матеріалу, або визначити правильну відповідь з викладачем на консультації.

Під час вивчення даної дисципліни використовуються:

– мультимедійні освітні технології: інтерактивні лекції (презентації) із використанням програм MS Power Point, Google Slides у поєднанні з анімацією та звуковим супроводом; перегляд відеороликів за окремими пунктами тем занять, використання електронних посібників;

– діалогові технології: організація групових обговорень, використання «мозкового штурму».

Лекції проводяться з використанням технічних засобів навчання й супроводжуються демонстрацією презентацій за допомогою проектора.

У разі виникнення необхідності забезпечення навчального процесу в дистанційному режимі супровід та контроль знань реалізовується за допомогою дистанційного курсу, розробленого в Google Classroom. Онлайн лекції, консультації та усні відповіді на питання, захист лабораторних робіт проводиться за допомогою Google Meet або Zoom.

9 МЕТОДИ КОНТРОЛЮ

Основними завданнями контролю знань здобувачів вищої освіти з дисципліни є оцінювання засвоєння теоретичних знань і практичних навичок, отриманих під час навчання.

Контрольні заходи мають виконувати наступні функції:

- стимулювати систематичну самостійну роботу над навчальним матеріалом;
- забезпечувати закріплення та реалізацію набутих теоретичних знань при підготовці до практичних занять;
- прищеплювати навички відповідального ставлення до своїх обов'язків, самостійного цілеспрямованого пошуку потрібної інформації, чіткої організації свого робочого дня.

Оцінювання знань здобувачів складається з поточного та підсумкового контролю.

Поточний контроль знань здобувачів вищої освіти передбачає оцінювання за наступними основними напрямками:

- перевірка теоретичних знань;
- перевірка підготовки до лабораторних занять.

З даних компонентів складаються загальні бали, які фіксуються в журналі викладача.

Оцінювання рівня засвоєння теоретичних знань здобувачів вищої освіти проводиться під час усної співбесіди зі здобувачами по теоретичним матеріалам, за результатами захисту лабораторних робіт й виконання самостійної роботи. Підсумковим контролем є залік у 6 семестрі (4 семестрі для скороченої форми навчання) і екзамен у 7 семестрі (5 семестрі для скороченої форми навчання).

10 РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ ЗДОБУВАЧІ

Використовується модульно-рейтингова система оцінювання, яка передбачає розподіл балів за виконання всіх запланованих видів робіт.

10.1 Денна форма навчання

Максимальна кількість балів за модуль при умові його бездоганного виконання для студентів *денної* форми навчання

№ з/п	Назва змістового модуля	Максимальна кількість балів
1	Принципи побудови моделей машинного навчання:	
	Лабораторні роботи	25
	КМР за 1 змістовий модуль	25
2	Вирішальні дерева та їх композиції:	
	Лабораторні роботи	25
	КМР за 2 змістовий модуль	25
Всього за 6 (4 для скороченої форми) семестр		100
3	Огляд методів та розгортання моделей штучного інтелекту:	
	Лабораторні роботи	75
	КМР за 3 змістовий модуль	25
Всього за 7 (5 для скороченої форми) семестр		100

Сума складається з балів, що накопичив студент у ході поточного контролю.

Лабораторні роботи у модулі відображують оволодіння навичками та вміння застосовувати знання на практиці і сукупно відповідають 50-ти відсоткам ваги у 6 (4 для скороченої форми) семестрі та 75-ти відсоткам ваги у 7 (5 для скороченої форми) семестрі. Бали розподіляються поміж лабораторними роботами модуля у відповідності до їх відносної складності. При зниженні якості виконання тієї чи іншої лабораторної роботи, знижується і кількість балів, якою вона оцінюється.

Бали за контрольно-модульні роботи складають 50 відсотків та 25 відсотків відповідно.

Контрольно-модульна робота (КМР) може дати максимально 25 балів при найвищій якості виконання. При зниженні якості КМР знижується і сума балів відповідно до шкали, що наводиться у таблиці:

Шкала оцінювання контрольних робіт

Відсоток вірних компонентів КМР	0 – 30	31 – 60	61 – 75	76 – 85	86 – 94	95 – 100
Сума балів за КМР	0	5	10	15	20	25

10.2 Заочна форма навчання

Максимальна кількість балів за модуль при умові його бездоганного виконання для студентів заочної форми навчання

№ з/п	Назва змістового модуля	Максимальна кількість балів
1	Принципи побудови моделей машинного навчання:	
	Лабораторні роботи	5
2	Вирішальні дерева та їх композиції:	
	Лабораторні роботи	15
	КР	80
	Всього за 6 (4 для скороченої форми) семестр	100
3	Огляд методів та розгортання моделей штучного інтелекту:	
	Лабораторні роботи	20
	КР	80
	Всього за 7 (5 для скороченої форми) семестр	100

Сума складається з балів, що накопичив студент у ході поточного контролю.

Лабораторні роботи у модулі відображують оволодіння навичками та вміння застосовувати знання на практиці і сукупно відповідають 20-ти відсоткам ваги. Бали розподіляються поміж лабораторними роботами модуля у відповідності до їх відносної складності

Бали за контрольні роботи складають 80 відсотків відповідно. При зниженні якості виконання тієї чи іншої лабораторної роботи, знижується і кількість балів, якою вона оцінюється.

Контрольна робота (КР) може дати максимально 80 балів при найвищій якості виконання. При зниженні якості КР знижується і сума балів відповідно до шкали, що наводиться у таблиці:

Шкала оцінювання контрольних робіт

Відсоток вірних компонентів КР	0 – 30	31 – 60	61 – 75	76 – 85	86 – 94	95 – 100
Сума балів за КР	30	40	50	60	70	80

Шкала оцінювання лабораторних робіт

№ мод уля	№ зан.	Вид роботи	Тема	Максимальна кількість балів			
				денна	заочна	денна скорочена	скорочена заочна
1	1,2	Лабораторна робота № 1	Початок роботи з пакетом обчислювальної математики NumPy і пакетом для наукових і інженерних розрахунків SciPy	5	–	5	–
	3,4	Лабораторна робота № 2	Препроцесінг даних	5	–	5	–
	5,6	Лабораторна робота № 3	Візуалізація даних в Pandas	5	–	5	–
	7,8	Лабораторна робота № 4	Лінійна регресія і основні бібліотеки Python для аналізу даних і наукових обчислень	5	–	5	–
	9	Лабораторна робота № 5	Лінійна регресія і стохастичний градієнтний спуск	5	5	5	5
Контрольно-модульна робота				25	–	25	–
Разом по модулю				50	-	50	-
2	10	Лабораторна робота № 6	Лінійна регресія: перенавчання і регуляризація	5	–	5	–
	11	Лабораторна робота № 7	Перенавчання моделей і боротьба з ним	5	–	5	–
	12- 14	Лабораторна робота № 8	Вирішальні дерева	5	5	5	5
	15- 17	Лабораторна робота № 9	Метод градієнтного бустінгу своїми руками	10	10	10	10
Контрольно-модульна робота / Контрольна робота				25	80	25	80
Разом по модулю				50	-	50	-
Разом за семестр				100			
3	1	Лабораторна робота № 10	Використання акселераторів обчислень у Google Colaboratory	5	–	5	–
	2	Лабораторна робота № 11	Розробка спам-фільтру на основі байєсівського класифікатора	10	5	10	5
	3	Лабораторна робота № 12	Розробка рекомендаційної системи на основі метричних алгоритмів	10	5	10	5
	4	Лабораторна робота № 13	Використання алгоритмів колаборативної фільтрації для рекомендаційних систем	10	–	10	–
	5	Лабораторна робота № 14	Вирішення задачі кластеризації. Ядерні трюки методу SVM	10	–	10	–
	6	Лабораторна робота № 15	Реалізація задачі регресії методом опорних векторів	10	–	10	–
	7	Лабораторна робота № 16	Робота з бібліотекою PyCaret. Оптимізація моделей	10	–	10	–
	8	Лабораторна робота № 17	Розгортання моделі у web-застосунок. Фреймворк Flask	10	10	10	10
Контрольно-модульна робота / Контрольна робота				25	80	25	80

№ модуля	№ зан.	Вид роботи	Тема	Максимальна кількість балів			
				денна	заочна	денна скорочена	скорочена заочна
Разом по модулю				100			
Разом за семестр				100			

Оцінювання кожної лабораторної роботи ведеться за показниками, наведеними в таблиці:

№ з/п /Критерій оцінювання	Максимальна кількість балів (денна форма/заочна форма)			
	Своєчасність виконання	Правильність виконання	Захист роботи	Всього за роботу
Лабораторна робота № 1	1/-	1/-	3/-	5/-
Лабораторна робота № 2	1/-	1/-	3/-	5/-
Лабораторна робота № 3	1/-	1/-	3/-	5/-
Лабораторна робота № 4	1/-	1/-	3/-	5/-
Лабораторна робота № 5	1/-	1/5	3/-	5/5
Лабораторна робота № 6	1/-	1/-	3/-	5/-
Лабораторна робота № 7	1/-	1/-	3/-	5/-
Лабораторна робота № 8	1/-	1/5	3/-	5/5
Лабораторна робота № 9	1/-	4/-	5/-	10/-
Всього за семестр	9/-	12/10	29/-	50/10
Лабораторна робота № 10	1/-	1/-	3/-	5/
Лабораторна робота № 11	1/-	4/10	5/-	10/10
Лабораторна робота № 12	1/-	4/10	5/-	10/10
Лабораторна робота № 13	1/-	4/-	5/-	10/
Лабораторна робота № 14	1/-	4/-	5/-	10/
Лабораторна робота № 15	1/-	4/-	5/-	10/
Лабораторна робота № 16	1/-	4/-	5/-	10/
Лабораторна робота № 17	1/-	4/10	5/-	10/10
Всього за семестр	8/-	29/30	38/-	75/30

Під своєчасністю практичного виконання та своєчасністю захисту лабораторної роботи розуміється виконання та захист у тиждень згідно із графіком робіт.

Правильність виконання роботи оцінюється наступним чином:

- робота виконана без зауважень - максимальний бал;
- робота виконана достатньо повно з деякими зауваженнями – дві третини від максимального балу;
- робота виконана не повністю – одна третина від максимального балу;
- при перевірці роботи виявлені грубі помилки – 0 балів.

Захист лабораторної роботи передбачає відповіді на контрольні питання, які представлені у методичних вказівках до виконання лабораторних робіт відповідно до кожної теми.

Для допуску до підсумкового контролю студент повинен виконати графік навчального процесу, усі види запланованих завдань і протягом семестру отримати в сумі не менше 50 балів.

Семестровий контроль здійснюється у формі заліку в шостому семестрі для денної та заочної форм навчання, у четвертому семестрі для денної скороченої та заочної скороченої форми навчання; екзамену – відповідно у сьомому і п'ятому семестрах.

У разі виконання студентом усіх видів поточних контрольних заходів залік виставляється студенту на підставі зарахованих балів протягом семестру. Результати заліку оцінюються за 100-бальною шкалою. У відомість оцінка проставляється як у балах національної шкали, так і за шкалою ECTS.

У семестрі, в якому семестровий контроль здійснюється у формі екзамену, 100 балів, набрані студентом за результатами поточного контролю складають 60 відсотків, а екзаменаційне завдання – 40 відсотків. Результати екзамену оцінюються за 100-бальною шкалою. У відомість оцінка проставляється як у балах національної шкали, так і за шкалою ECTS:

Шкала оцінювання

Національна шкала успішності	Оцінка ECTS	Визначення ECTS	100-бальна система оцінювання
відмінно/ зараховано	A	ВІДМІННО - відмінне виконання лише з незначними помилками	90...100
добре/ зараховано	B	ДУЖЕ ДОБРЕ - вище середнього рівня з кількома помилками	80...89
	C	ДОБРЕ - у цілому правильно робота з певною кількістю помилок і недоліків	71...79
задовільно/ зараховано	D	ЗАДОВІЛЬНО - непогано, але зі значною кількістю грубих помилок	61...70
	E	ДОСТАТНЬО - виконання задовольняє мінімальні потреби	50...60
незадовільно/ не зараховано	FX	НЕЗАДОВІЛЬНО - із можливістю повторного складання	30...49
	F	НЕЗАДОВІЛЬНО - з обов'язковим повторним вивчення дисципліни	0...29

При наявності у здобувачів результатів **неформального навчання** за освітнім компонентом «Системи штучного інтелекту» у повному обсязі, визнання та оцінювання результатів здійснюється відповідно до «Положення про порядок визнання у Криворізькому національному університеті результатів навчання, отриманих в умовах неформальної освіти». У випадку, якщо за підсумками визнання результатів неформального навчання визнається тільки частина результатів навчання, заявнику зараховуються окремі види навчальної роботи за освітнім компонентом «Системи штучного інтелекту».

Нижче наведені окремі види навчальної роботи, які можуть бути зараховані здобувачеві при наявності сертифікату про успішне проходження рекомендованих онлайн курсів.

Тема	Посилання на рекомендовані курси
Введення у машинне навчання	https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about https://www.coursera.org/learn/machine-learning-basics
Лінійні моделі	https://www.coursera.org/learn/machine-learning-basics https://www.coursera.org/learn/machine-learning-with-python
Боротьба з перенавчанням моделей	https://www.coursera.org/learn/machine-learning-basics
Вирішальні дерева	https://www.coursera.org/specializations/machine-learning-introduction
Градiєнтний бустiнг	https://www.coursera.org/learn/machine-learning
Метричні алгоритми	https://www.coursera.org/learn/machine-learning-basics
Метод опорних векторів	https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about

11 ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

- Аудиторія персональних комп'ютерів класу Celeron, AMD, Pentium, Core 2 Duo, Core i5, i7 (або вище) з операційною системою типу Windows 7, 8 або 10. Забезпеченість комп'ютерами – 12 шт. на 25 студентів.
- Підключення до мережі Інтернет.
- Програмне забезпечення: Python 3, Anaconda, PyCharm, Jupyter notebook, Google Colaboratory, бібліотеки NumPy, SciPy, Pandas, Matplotlib, TensorFlow, PyBrain, Keras.

12 ПЕРЕЛІК ПИТАНЬ ДЛЯ ПІДСУМКОВОГО КОНТРОЛЮ ЗНАНЬ

Модуль 1

1. Які підвибірки можна зробити при використанні бутстрапа?
2. Поняття об'єкту в машинному навчанні.
3. Типи задач машинного навчання.
4. Обчислення прогнозування в лінійній моделі.
5. Навчання з учителем.
6. Постановка задачі навчання по прецедентах.
7. Поняття градієнта в машинному навчанні.
8. Етапи розробки моделі машинного навчання.
9. Особливості задач машинного навчання без учителя.
10. Крок в градієнтному спуску.
11. Навчання без учителя.
12. Середньоквадратична помилка в задачах регресії.
13. Поняття ознак в машинному навчанні.
14. Лінійна модель навчання.
15. Порівняння стохастичного градієнтного спуску зі звичайним градієнтним методом.

16. Поняття алгоритма (моделі) в машинному навчанні.
17. Навчання лінійної моделі.
18. Як можна позбутися вільного члена в лінійних моделях?
19. Особливості задач регресії, класифікації, кластеризації.
20. Метод градієнтного спуску.

Модуль 2

21. Перенавчання моделей машинного навчання.
22. Типи ознак.
23. Стохастичний градієнтний спуск.
24. Чому дерева великої глибини мають високе зміщення?
25. Дійсні ознаки.
26. Проблема перенавчання в задачах лінійної регресії.
27. Що таке "кількість об'єктів в вершині" вирішального дерева?
28. Категоріальні ознаки.
29. Проблема перенавчання і недонавчання моделей.
30. Як влаштований процес побудови вирішального дерева?
31. Задачі класифікації.
32. Методи боротьби з перенавчанням.
33. Яку проблему вирішують критерії зупину і стрижка дерев?
34. Поняття регуляризації.
35. Розв'язання задачі регресії за допомогою вирішальних дерев.
36. Мінімізація функціоналу помилок.
37. Оцінка якості моделей. Крос-валідація.
38. Як можна використовувати категоріальні ознаки у вирішальному дереві в підході, який будує n -арні дерева?
39. Поняття вирішальних дерев.
40. Яку форму матиме розділяюча поверхня, побудована деревом з умовами виду $[x_j < t]$ в вершинах? Вважайте, що у вибірці дві ознаки.

Модуль 3

41. Геометричний сенс градієнта.
42. Критерій зупинення і стрижка дерев.
43. Композиції дерев.
44. Види регуляризаторів.
45. Рандомізація в навчанні композицій дерев.
46. Недоліки оцінки якості алгоритму по відкладеній вибірці.
47. Випадкові ліси.
48. Поняття відступу в оцінки якості класифікатора.
49. Функціонал якості моделі лінійної регресії.
50. Алгоритм побудови випадкового лісу.
51. Верхні оцінки на порогову функцію втрат при навчанні лінійних класифікаторів.
52. Недоліки випадкових лісів.
53. Компонента розкиду в розкладанні помилки.
54. Поняття градієнтного бустінгу.
55. У якому просторі градієнтний бустінг здійснює градієнтний спуск?
56. Проблема перенавчання градієнтного бустінгу
57. Чим градієнтний бустінг відрізняється від випадкового лісу?
58. Стохастичний градієнтний бустінг.

59. Градієнт якої функції обчислюється на кожній ітерації градієнтного бустінга?
60. Відбір ознак за допомогою регуляризатора.
61. Поняття вибірки, алгоритм навчання.
62. Для чого потрібно скорочення кроку в градієнтному бустінгу?
63. Задача бінарної класифікації.
64. Навчання випадкового лісу.
65. Аналітичний метод підбору параметрів лінійної регресії.
66. Глибина вирішальних дерев в випадкових лісах.
67. Чому дерева великої глибини мають високий розкид?
68. Недоліки оцінки якості алгоритму по відкладеній вибірці.
69. Недоліки градієнтного методу.
70. Чому дерева великої глибини мають високе зміщення?
71. Збіжність стохастичного градієнтного методу.
72. Функції втрат в задачах класифікації.

13 НАВЧАЛЬНО-МЕТОДИЧНІ МАТЕРІАЛИ З ДИСЦИПЛІНИ

13.1 Навчальна та довідкова література

1. Шаховська Н. Б. Системи штучного інтелекту: навч. посібник / Н. Б. Шаховська, Р. М. Камінський, О. Б. Вовк. Львів: Видавництво Львівської політехніки, 2018. 392 с.
2. Ertel W. Introduction to Artificial Intelligence. Springer International Publishing 2017. 356 p.
3. Шаров С.В. Інтелектуальні інформаційні системи: навч. посіб. / С.В. Шаров, Д.В. Лубко, В.В. Осадчий. Мелітополь: Вид-во МДПУ ім. Б. Хмельницького, 2015. 144 с.
4. Willi Richert, Luis Pedro Coelho. Building Machine Learning Systems with Python 2-nd edition / Packt Publishing, 2013. – 290 p.
5. Оптимізаційні методи та моделі. Підручник. / [Л. В. Забуранна, Н. В. Попрозман, Н. А. Клименко, О. І. Попрозман, С. В. Забуранний]. – К.: 2014. – 372 с.
6. Wolfgang Ertel Introduction to Artificial Intelligence 2nd edition: Springer International Publishing AG, 2017, 356 p.
7. Charu C. Aggarwal Neural Networks and Deep Learning / Springer International Publishing AG, part of Springer Nature, 2018, 497 p.
8. Ткаченко Р. О. Нейромережеві засоби штучного інтелекту: навч. посібник / Р. О. Ткаченко, П. Р. Ткаченко, І. В. Ізонін. Львів: Видавництво Львівської політехніки, 2017. 208 с.
9. Deng L.y Yu D. Deep Learning: Methods and Applications // Foundations and Trends in Signal Processing, 2014, vol. 7, no. 3-4. — P. 197-387.
10. Richard S. Sutton, Andrew G. Barto. Reinforcement Learning : An Introduction MIT Press, Adaptive Computation and Machine Learning Ser.: 2018. - 552 p.

13.2. Методична література

1. Методичні вказівки до виконання лабораторних робіт з дисципліни «Системи штучного інтелекту» розділ «Методи і моделі машинного навчання» для студентів всіх форм навчання зі спеціальності 121 «Інженерія програмного забезпечення». – 67 с. Укладачі: Саїтгарєєв Н. Х., к.т.н., доцент, Доценко І. О., ст. викладач, Шаповалова Н. Н., ст. викладач, 2021 р.
2. Методичні вказівки до виконання лабораторних робіт з дисципліни «Системи штучного інтелекту» розділ «Штучні нейронні мережі» для студентів всіх форм навчання зі спеціальності 121 «Інженерія програмного забезпечення». – 52 с. Укладачі: Саїтгарєєв Н. Х., к.т.н., доцент, Доценко І. О., ст. викладач, Шаповалова Н. Н., ст. викладач, 2021 р.

3. Методичні вказівки до виконання лабораторних робіт з дисципліни «Системи штучного інтелекту» розділ «Розпізнавання образів і системи комп'ютерного зору» для студентів всіх форм навчання зі спеціальності 121 «Інженерія програмного забезпечення». – 50 с. Укладачі: Саїтгарєєв Н. Х., к.т.н., доцент, Доценко І. О., ст. викладач, Шаповалова Н. Н., ст. викладач, 2018 р.

14 ІНФОРМАЦІЙНІ РЕСУРСИ

До складу інформаційних ресурсів навчальної дисципліни входять:

1. Бібліотека Криворізького національного університету (м. Кривий Ріг, вул. Пушкіна, 37). – Режим доступу: <http://lib.knu.edu.ua/>

Internet-ресурси:

2. Introduction to Machine Learning. [Електронний ресурс] – Режим доступу: <https://developers.google.com/machine-learning/crash-course/ml-intro>
3. Документація NumPy. [Електронний ресурс] – Режим доступу: <https://numpy.org/doc/stable/index.html#numpy-docs-mainpage>
4. Документація SciPy. [Електронний ресурс] – Режим доступу: <https://docs.scipy.org/doc/scipy/>
5. Документація Pandas. [Електронний ресурс] – Режим доступу: <https://pandas.pydata.org/>
6. Документація Matplotlib. [Електронний ресурс] – Режим доступу: <https://matplotlib.org/>
7. Документація Flask. [Електронний ресурс] – Режим доступу: <https://flask.palletsprojects.com/en/2.3.x/>
8. Документація ngrok. [Електронний ресурс] – Режим доступу: <https://ngrok.com/>
9. Документація rucaret. [Електронний ресурс] – Режим доступу: <https://rucaret.org/>

15 ТЕРМІНОЛОГІЧНИЙ СЛОВНИК

Accuracy (точність). The number of correct classification predictions divided by the total number of predictions. That is: $Accuracy = \frac{\text{correct predictions}}{\text{correct predictions} + \text{incorrect predictions}}$.

Activation function (активаційна функція). A function that enables neural networks to learn nonlinear (complex) relationships between features and the label. Popular activation functions include: ReLU, Sigmoid.

Analysis syntax (синтаксичний аналіз). Checking the expression for its occurrence in the set of constructed expressions. 2. When processing natural language under A.S. is understood as the construction of the syntactic structure of a sentence in some natural language. In intelligent systems that implement the functions of understanding texts in a limited natural language, A.S. is carried out in two stages. On the first one, a surface syntactic structure is built, in which parts of speech and relations between them participate. At the second stage, there is a transition to a deep syntactic structure, which is essentially a cognitive structure associated with the reflection of reality in a text in natural language. For A.S. created software systems - parsers.

AdaGrad. A sophisticated gradient descent algorithm that rescales the gradients of each parameter, effectively giving each parameter an independent learning rate.

Agent (агент). In reinforcement learning, the entity that uses a policy to maximize the expected return gained from transitioning between states of the environment.

Anomaly detection (виявлення аномалій). The process of identifying outliers. For example, if the mean for a certain feature is 100 with a standard deviation of 10, then anomaly detection should flag a value of 200 as suspicious.

Artificial intelligence (штучний інтелект). A non-human program or model that can solve sophisticated tasks. For example, a program or model that translates text or a program or model that identifies diseases from radiologic images both exhibit artificial intelligence. Formally, machine learning is a sub-field of artificial intelligence. However, in recent years, some organizations have begun using the terms artificial intelligence and machine learning interchangeably.

Bagging (бегінг). A method to train an ensemble where each constituent model trains on a random subset of training examples sampled with replacement. For example, a random forest is a collection of decision trees trained with bagging. The term bagging is short for bootstrap aggregating.

Bag of words (мішок слів). A representation of the words in a phrase or passage, irrespective of order. For example, bag of words represents the following three phrases identically:

the dog jumps

jumps the dog

dog jumps the

Each word is mapped to an index in a sparse vector, where the vector has an index for every word in the vocabulary. For example, the phrase the dog jumps is mapped into a feature vector with non-zero values at the three indices corresponding to the words the, dog, and jumps. The non-zero value can be any of the following:

A 1 to indicate the presence of a word. A count of the number of times a word appears in the bag. For example, if the phrase were the maroon dog is a dog with maroon fur, then both maroon and dog would be represented as 2, while the other words would be represented as 1. Some other value, such as the logarithm of the count of the number of times a word appears in the bag.

Bayesian optimization (басівська оптимізація). A probabilistic regression model technique for optimizing computationally expensive objective functions by instead optimizing a surrogate that quantifies the uncertainty via a Bayesian learning technique. Since Bayesian optimization is itself very expensive, it is usually used to optimize expensive-to-evaluate tasks that have a small number of parameters, such as selecting hyperparameters.

Centroid (центроїд). The center of a cluster as determined by a k-means or k-median algorithm. For instance, if k is 3, then the k-means or k-median algorithm finds 3 centroids.

Centroid-based clustering (центроїдна кластеризація). A category of clustering algorithms that organizes data into nonhierarchical clusters. k-means is the most widely used centroid-based clustering algorithm. Contrast with hierarchical clustering algorithms.

Classification model (класифікаційна модель). A model whose prediction is a class. For example, the following are all classification models:

A model that predicts an input sentence's language (French? Spanish? Italian?).

A model that predicts tree species (Maple? Oak? Baobab?).

A model that predicts the positive or negative class for a particular medical condition.

In contrast, regression models predict numbers rather than classes.

Two common types of classification models are: binary classification, multi-class classification.

Data analysis (аналіз даних). Obtaining an understanding of data by considering samples, measurement, and visualization. Data analysis can be particularly useful when a dataset is first received, before one builds the first model. It is also crucial in understanding experiments and debugging problems with the system.

DataFrame. A popular pandas datatype for representing datasets in memory. A DataFrame is analogous to a table or a spreadsheet. Each column of a DataFrame has a name (a header), and each row is identified by a unique number. Each column in a DataFrame is structured like a 2D array, except that each column can be assigned its own data type.

Ensemble (ансамбль). A collection of models trained independently whose predictions are averaged or aggregated. In many cases, an ensemble produces better predictions than a single model. For example, a random forest is an ensemble built from multiple decision trees. Note that not all decision forests are ensembles.

False negative (FN) (хибний пропуск). An example in which the model mistakenly predicts the negative class. For example, the model predicts that a particular email message is not spam (the negative class), but that email message actually is spam.

Gradient (градієнт). The vector of partial derivatives with respect to all of the independent variables. In machine learning, the gradient is the vector of partial derivatives of the model function. The gradient points in the direction of steepest ascent.

Hyperparameter (гіперпараметр). The variables that you or a hyperparameter tuning service adjust during successive runs of training a model. For example, learning rate is a hyperparameter. You could set the learning rate to 0.01 before one training session. If you determine that 0.01 is too high, you could perhaps set the learning rate to 0.003 for the next training session. In contrast, parameters are the various weights and bias that the model learns during training.

Label (мітка). In supervised machine learning, the "answer" or "result" portion of an example. Each labeled example consists of one or more features and a label. For instance, in a spam detection dataset, the label would probably be either "spam" or "not spam." In a rainfall dataset, the label might be the amount of rain that fell during a certain period. A program or system that trains a model from input data. The trained model can make useful predictions from new (never-before-seen) data drawn from the same distribution as the one used to train the model.

Noise (шум). Broadly speaking, anything that obscures the signal in a dataset. Noise can be introduced into data in a variety of ways. For example:

Human raters make mistakes in labeling.

Humans and instruments mis-record or omit feature values.

Overfitting (перенавчання). Creating a model that matches the training data so closely that the model fails to make correct predictions on new data.

Regularization can reduce overfitting. Training on a large and diverse training set can also reduce overfitting.

Pandas. A column-oriented data analysis API built on top of numpy. Many machine learning frameworks, including TensorFlow, support pandas data structures as inputs. See the pandas documentation for details.

Random forest (випадковий ліс). An ensemble of decision trees in which each decision tree is trained with a specific random noise, such as bagging. Random forests are a type of decision forest

Scoring (рахунок). The part of a recommendation system that provides a value or ranking for each item produced by the candidate generation phase.

Training sample (навчальна вибірка). A set of examples and counterexamples for the formation of decision rules. Included in the learning table.

Tree binary (бінарне дерево). Representation of the search process in the form of a tree, each vertex of which is associated with the value of the search key in such a way that all smaller keys are concentrated in its left subtree, and all large ones - in the right one.

Weight (вага). A value that a model multiplies by another value. Training is the process of determining a model's ideal weights; inference is the process of using those learned weights to make predictions.

