

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК

КАФЕДРА ІНТЕЛЕКТУАЛЬНИХ ІНФОРМАЦІЙНИХ СИСТЕМ

“ЗАТВЕРДЖУЮ”

Перший проректор

Котляр Ю.В.

“30” серпня 2024 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«МЕТОДИ ТА СИСТЕМИ МАШИННОГО НАВЧАННЯ»

Спеціальність:

122 Комп'ютерні науки

Розробники

Гожий О.П.

Калініна І.О.

Завідувач кафедри

Кондратенко Ю. П

Гарант освітньої програми

Гожий О.П.

Декан факультету

Бойко А.П.

Начальник НМВ

Шкірчак С.І.

1. Опис навчальної дисципліни

Найменування показника	Характеристика дисципліни	
Найменування дисципліни	Методи та системи машинного навчання	
Шифр курсу в освітній програмі	ПП 7	
Галузь знань	12 Інформаційні технології	
Спеціальність	122 Комп'ютерні науки	
Спеціалізація (якщо є)	-	
Освітня програма	Інтелектуальні інформаційні системи	
Рівень вищої освіти	Магістр	
Статус дисципліни	Нормативна	
Курс навчання	I	
Навчальний рік	2024-2025	
Номер семестрів:	Денна форма	Заочна форма
	2	-
Загальна кількість кредитів ЄКТС/годин	4,5 кредитів / 135 годин	
Структура курсу: – лекції – практичні – годин самостійної роботи студентів	Денна форма	Заочна форма
	18	-
	36	-
	81	-
Відсоток аудиторного навантаження	40%	-
Мова викладання	Українська	
Форма проміжного контролю	Тестування	
Форма підсумкового контролю	Іспит; Курсова робота	

2. Мета, завдання та результати вивчення дисципліни

Концепція викладання дисципліни:

Дисципліна «Методи та системи машинного навчання» присвячена розгляду актуальних питань використання сучасних інформаційних технологій для вирішення інтелектуальних задач і застосуванню інтелектуальних методів при вирішенні різноманітних проблем, зокрема проблем побудування систем машинного навчання, вирішенню попередньої обробки даних, задач нормалізації, задач класифікації та кластеризації, проблем побудови алгоритмів для вирішення завдань машинного навчання та прогнозування, розв'язання завдань розпізнавання образів та завдань глибинного навчання. Постійний розвиток технологій машинного навчання та інформаційних інтелектуальних систем, розвиток та оновлення програмного забезпечення у сфері інтелектуалізації обробки і представлення даних та знань створює передумови для застосування нових методів та інформаційних технологій вирішення інтелектуальних задач у різних галузях людської діяльності. У зв'язку із цим виникає необхідність викладати у програмі підготовки магістрів з навчального напрямку 122 – «Комп'ютерні науки» дисципліну «Методи та системи та машинного навчання».

Мета курсу: формування у студентів теоретичних знань з сучасних інформаційних технологій машинного навчання при вирішенні різноманітних задач, та ефективного використання інтелектуальних методів та їх реалізацій для розв'язування практичних завдань, та придбання навичок їх ефективного використання.

«Методи та системи машинного навчання» є *нормативно* дисципліною при підготовці магістрів з галузі знань «Інформаційні технології». Для студентів спеціальності 122 - "Комп'ютерні науки" викладається у 2 семестрі в обсязі 4,5 кредитів, в тому числі 54 годин аудиторних занять; з них 18 годин лекцій, 36 годин практичних занять; 81 години самостійної роботи. Згідно навчального плану по дисципліні «Методи та системи та машинного навчання» передбачено виконання *курсової роботи*.

Передумови для вивчення дисципліни.

Вивчення дисципліни передбачає наявність першого бакалаврського рівня. Таким чином

викладання курсу «Методи та системи машинного навчання» базується на знаннях, які студенти одержали при вивченні дисциплін «Вища математика», «Теорія алгоритмів», «Теорія ймовірностей, імовірнісні процеси і математична статистика», «Методи та системи штучного інтелекту» та інших. Студент повинен мати стійкі навички роботи на персональному комп'ютері, знати типові структури даних та основи програмування.

З дисципліни «Вища математика» використовуються знання і навички з лінійної алгебри та матричного обчислення.

З дисципліни «Теорія алгоритмів» використовуються знання і навички з розробки алгоритмів.

З дисципліни «Теорія ймовірностей, імовірнісні процеси і математична статистика» використовуються знання стосовно ймовірності, випадкової величини, випадкової функції, випадкового процесу, функції розподілу, щільності ймовірності, статистичних моментів їх оцінок.

З дисципліни «Методи та системи штучного інтелекту» використовуються знання і навички з розділу «Нейронні мережі».

В результаті вивчення дисципліни у студентів повинні бути сформовані **уявлення** про:

- Цілі і завдання досліджень в галузі машинного навчання, системах машинного навчання, принципи їх побудови та різних областях застосування.
- Проблеми і способи побудови нейронних мереж.
- Постановку задач класифікації, передбачення, розпізнавання образів і шляхи їх розв'язання.

Після вивчення дисципліни студенти повинні **знати**:

- Основні етапи та алгоритми, та процедури навчання та обробки даних.
- Основні алгоритми класифікації та кластеризації.
- Основні моделі нейронних мереж, методів і алгоритмів їх навчання.
- Основні поняття про навчання з підкріпленням.
- Структури та методи побудови систем МН та їх архітектурних особливостей залежно від особливостей розв'язуваної задачі. Етапи побудови систем МН.

В результаті вивчення дисципліни студенти повинні придбати **вміння** та **навички**:

- Будувати та застосовувати основні моделі нейронних мереж, класифікаторів та кластеризаторів для вирішення прикладних завдань МН.
- Ставити завдання побудови системи МН для вирішення різноманітних задач в різних предметних областях.
- Розробляти алгоритми для вирішення завдань класифікації, кластеризації та систем навчання за допомогою нейронних мереж в різних предметних областях.
- Розробляти компоненти систем машинного навчання, використовуючи сучасні програмні та інструментальні засоби.

Знання, які студенти набудуть при вивченні курсу «Методи машинного навчання» будуть необхідними при подальшому навчанні та освоєнні фахових та спеціальних дисциплін, а також у виробничій діяльності зі спеціальності.

Викладання дисципліни направлене на формування у студента наступних загальних та спеціальних (фахових, предметних) компетентностей (згідно зі стандартом вищої освіти України для другого (магістерського) рівня від 28.04.2022 № 393):

Загальні компетентності:

- ЗК 5. Здатність вчитися й оволодівати сучасними знаннями.
- ЗК 7. Здатність генерувати нові ідеї (креативність).

Спеціальні (фахові, предметні) компетентності:

- СК 08. Здатність розробляти і реалізовувати проекти зі створення програм-ного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності

застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.

– СК 12. Здатність розробляти елементи програмного забезпечення для реалізації інтелектуальних технологій, а саме нейронних мереж, методів нечіткої логіки, еволюційних методів обчислень, біоінспірованих методів в інтелектуальних системах.

– СК 13. Здатність розробляти та реалізовувати проекти з інтелектуальних систем в тому числі за допомогою методів машинного і глибокого навчання, імітаційних моделей, застосовуючи нові інструментальні засоби розробки, сучасні бібліотеки мов програмування та сучасні візуальні технології.

Програмні результати навчання:

– РН 1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.

– РН 10. Проектувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

– РН 11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

– РН 20. Розробляти та застосовувати елементи програмного забезпечення для реалізації інтелектуальних технологій, а саме нейронних мереж, методів нечіткої логіки, еволюційних методів обчислень, біоінспірованих методів в інтелектуальних системах.

– РН 21. Розробляти та реалізовувати проекти на основі інтелектуальних компонентів в тому числі за допомогою методів машинного і глибокого навчання, імітаційних моделей, застосовуючи нові інструментальні засоби розробки, сучасні бібліотеки мов програмування та сучасні візуальні технології.

3. Програма навчальної дисципліни

Денна форма:

№	Теми	Лекції	Практичні (групові)	Самостійна робота
<i>РОЗДІЛ 1. ВСТУП. Предметна область та основні поняття машинного навчання.</i>				
1	Тема 1. Вступ в машинне навчання. Головні розділи. Основні поняття.	1	2	7
2	Тема 2. Типи систем машинного навчання. Головні завдання та проблеми машинного навчання.	1	4	7
<i>РОЗДІЛ 2. Етапи розробки проекту машинного навчання.</i>				
3	Тема 3. Етапи проекту МН.	1	6	7
4	Тема 4. Вибір та навчання моделі даних. Налаштування моделей даних.	1	6	7
<i>РОЗДІЛ 3. Методи класифікації та кластеризації</i>				
5	Тема 5. Методи класифікації. Одновимірний та багатовимірні методи. Методи нормалізації даних.	2	4	8
6	Тема 6. Навчання моделей. Зменшення розмірності. (Узагальнення).	2	2	7
7	Тема 7. Методи кластеризації.	2	6	7
8	Тема 8. Дерева прийняття рішень. Ансамблеві методи МН.	2	6	8

РОЗДІЛ 4. Нейромережеві технології МН.				
9	Тема 9. Нейромережеві технології МН.	2	-	7
10	Тема 10. Глибоке навчання та згорткові нейронні мережі.	2	-	8
11	Тема 11. Рекурентні мережі (RNN). Типи рекурентних нейронних мереж. Бібліотеки <i>TensorFlow</i> , <i>Keras</i> , <i>Scikit-Learn</i> .	2	-	8
РАЗОМ		18	36	81

4. Зміст навчальної дисципліни

4.1. План лекцій

№	Тема заняття / план
1	Тема 1: Вступ в машинне навчання. Головні розділи. Основні поняття. 1) Предмет та зміст дисципліни. 2) Два типи навчання: навчання з вчителем, навчання без вчителя. Навчання з підкріпленням. Тема 2: Типи систем машинного навчання. Головні завдання та проблеми машинного навчання. 1) Головна схема МН. 2) Етапи МН. Особливість кожного етапу. 3) Приклади завдань МН.
2	Тема 3: Етапи проекту МН. 1) Робота з реальними даними. 2) Отримання даних. 3) Підготовка даних для алгоритмів МН (нормалізація та агрегація даних). Тема 4: Вибір та навчання моделі даних. Налаштування моделей даних. 1) Методи попередньої обробки даних. 2) Налаштування моделей даних. 3) Приклади вирішення прикладних завдань.
3	Тема 5: Вступ до методів класифікації. 1) Двійковий класифікатор. 2) Байєсівський класифікатор. 3) Показники якості класифікації. 4) Багатокласова та багатозначна класифікація.
4	Тема 6: Навчання моделей. 1) Матриця неточностей. Методи побудови МН. 2) Лінійна регресія. 3) поліноміальна регресія. 4) логістична регресія.
5	Тема 7: Методи кластеризації. 1) Методи кластеризації. 2) Методи опорних векторів. 3) Лінійна класифікація SVM. 4) Регресія SVM.
6	Тема 8: Древа прийняття рішень. Ансамблеві методи МН. 1) Древа прийняття рішень. 2) Ансамблеві методи МН. 3) Бустінг, Беггінг.
7	Тема 9: Основи нейронних мереж. 1) Біологічний нейрон та модель штучного нейрону.

	2) Класифікація НМ. 3) Перцептрон. 4) Багатошаровий перцептрон. 5) Алгоритми навчання. Приклад побудови системи розпізнавання символів.
8	Тема 10: Глибоке навчання та згорткові нейронні мережі. 1) Завдання глибокого навчання. 2) Архітектура CNN. 3) Типи шарів CNN. Створення класифікаторів зображень на основі CNN.
9	Тема 11: Рекурентні мережі (RNN). Типи рекурентних нейронних мереж . Бібліотеки <i>TensorFlow, Keras, Scikit-Learn</i>. 1) Базові рекурентні нейронні мережі в бібліотеках <i>TensorFlow</i> . 2) Навчання рекурентних мереж. 3) Навчання RNN для прогнозування часових рядів.

4.2 План практичних (семінарських, лабораторних, півгрупових) занять

№ ПР	Тема заняття / план	Кількість годин
1	Практична робота №1. Розвідувальний аналіз та попередня обробка даних.	6
2	Практична робота №2. Регресія (частина 1). Моделі лінійної регресії	6
3	Практична робота №3. Регресія (частина 2). Регресійні моделі на деревах.	6
4	Практична робота №4. Методи класифікації (частина 1).	6
5	Практична робота №5. Методи класифікації (частина 2).	6
6	Практична робота №6. Методи кластеризації.	6

Методичні рекомендації щодо виконання практичних робіт знаходяться в Moodle3.

Для зарахування практичної роботи здобувач має змогу надати сертифікат (або інший документ) про проходження курсів на платформах *Prometheus, Coursera* тощо та за умови співпадіння напрямленості курсів блоку освітньої компоненти.

4.3. Завдання для самостійної роботи

4.3.1. Загальні положення

Одним з основних напрямів успішного засвоєння матеріалів навчальної дисципліни є самостійна робота студентів над основною й додатковою літературою з вивчення й використання сучасних комп'ютерних технологій при рішенні вимірювальних задач.

Основними видами самостійної роботи є:

1. Вивчення лекційного матеріалу.
2. Вивчення рекомендованої літератури.
3. Вивчення термінів і основних понять з тем навчальної дисципліни.
4. Підготовка до лабораторних занять і розробка ескізів документів з кожної лабораторної роботи.
5. Підготовка до тестового контролю з навчальної дисципліни.
6. Робота з опрацювання та вивчення рекомендованої літератури.
7. Систематизація вивченого матеріалу перед іспитом.

4.3.2. Обов'язкові види самостійної роботи

Фіксований перелік тем для виконання індивідуальних з дисципліни у семестрі студентам не пропонується. Теми обираються студентами самостійно та є засобом поглиблення знань про

алгоритми та засоби їх побудови, які розглядаються в межах дисципліни. Крім того, можуть бути розглянутими деякі специфічні використання методів та алгоритмів.

Теми індивідуальних занять узгоджуються з викладачем протягом семестру, до початку залікового тижня.

Теми інформаційних повідомлень співпадають з темами та основними питаннями, які розглядаються на лекціях. В інформаційних повідомленнях також можуть розглядатись новітні засоби та методи в сфері обчислювальних алгоритмів.

4.3.3. Додаткові теми для самостійної роботи

I. Поглиблені алгоритми та архітектури

1. **Стекінг та блендинг:** методи комбінування моделей для підвищення точності.
2. **Аналіз роботи та оптимізацій в XGBoost, LightGBM та CatBoost:** порівняння провідних реалізацій градієнтного бустингу.
3. **Архітектура Трансформерів (Transformers)** та механізм уваги (Attention Mechanism).
4. **BERT та його похідні (RoBERTa, ALBERT):** застосування трансформерів для задач обробки природної мови (NLP).
5. **Графові нейронні мережі (Graph Neural Networks, GNN):** аналіз даних у графових структурах.
6. **Архітектури U-Net та Mask R-CNN** для семантичної та екземплярної сегментації зображень.
7. **Навчання з одним шансом, декількома шансами та нульовим шансом (One-shot, Few-shot, Zero-shot learning).**

II. Генеративні моделі та представлення даних

8. **Генеративно-змагальні мережі (Generative Adversarial Networks, GANs):** принципи роботи, архітектури (DCGAN, StyleGAN) та проблеми навчання.
9. **Варіаційні автокодувальники (Variational Autoencoders, VAEs)** як інструмент для генерації даних.
10. **Дифузійні моделі (Diffusion Models):** сучасний підхід до генерації високоякісних зображень.
11. **Навчання представлень (Representation Learning):** від Word2Vec до сучасних контекстуалізованих ембедингів.
12. **Контрастивне навчання (Contrastive Learning)** для самонавчання на нерозмічених даних (напр., SimCLR).
13. **Векторні бази даних (Vector Databases)** та їх застосування для швидкого пошуку схожості.

III. Пояснюваність, етика та надійність III (XAI & Trustworthy AI)

14. **Методи інтерпретації моделей "чорної скриньки":** LIME та SHAP.
15. **Візуалізація та пояснення роботи згорткових нейронних мереж** (напр., Grad-CAM).
16. **Алгоритмічна справедливість (Fairness):** виявлення та пом'якшення упереджень (bias) у моделях.
17. **Атаки на змагання (Adversarial Attacks)** та методи захисту нейронних мереж.
18. **Причинно-наслідковий висновок (Causal Inference)** та його роль у машинному навчанні.
19. **Квантифікація невизначеності (Uncertainty Quantification)** в прогнозах моделі.

IV. Робота зі специфічними даними та задачами

20. **Прогнозування часових рядів** за допомогою рекурентних мереж (LSTM/GRU) та трансформерів.
21. **Методи виявлення аномалій (Anomaly Detection)** в даних.
22. **Байєсівські методи в машинному навчанні:** основи та практичне застосування.

23. **Навчання з підкріпленням (Reinforcement Learning):** алгоритми Deep Q-Networks (DQN) та Policy Gradients.
24. **Багатозадачне навчання (Multi-task Learning)** та його переваги.
25. **Мета-навчання (Meta-Learning):** концепція "навчання навчатися".

V. Системи машинного навчання та MLOps

26. **Активне навчання (Active Learning):** стратегії для мінімізації обсягу розмітки даних.
27. **Слабке керування (Weak Supervision):** створення навчальних наборів за допомогою функцій розмітки (напр., Snorkel).
28. **Федеративне навчання (Federated Learning)** для тренування моделей на децентралізованих даних із збереженням приватності.
29. **Моніторинг моделей у продакшені:** виявлення дрейфу даних (data drift) та дрейфу концепції (concept drift).
30. **Стратегії розгортання (deployment) моделей:** Canary, Blue/Green, Shadow.
31. **Автоматизоване машинне навчання (AutoML):** від підбору гіперпараметрів до пошуку архітектур нейронних мереж (NAS).
32. **Оптимізація моделей для мобільних пристроїв та Edge AI** (квантизація, дистиляція).
33. **Інструменти для версіонування даних та моделей (DVC, MLflow).**
34. **Оркестрація робочих процесів МН** за допомогою інструментів як Apache Airflow або Kubeflow Pipelines.
35. **Безсерверні (Serverless) архітектури** для розгортання та масштабування ML-сервісів.

4.3.4. Вибіркові види самостійної роботи

Студентам пропонується виконання творчих завдань для самостійного опрацювання (одне – за вибором студента). Виконання творчих завдань не є обов’язковим, але може бути зараховане як залікове завдання, що надає шанс набрати бажану кількість балів до сесії.

Кожне творче завдання оцінюється в 10 балів та являє собою практичне завдання з виконання певних алгоритмів або окремих програмних модулів.

За результатами виконання творчого завдання студент повинен оформити звіт, в якому будуть задокументовані на скрін-шотах послідовно всі дії з виконання завдання, до отримання кінцевого результату.

План курсової роботи (КР)

Планом навчального процесу передбачено виконання студентами курсової роботи (КР). Навчальний посібник з рекомендаціями щодо виконання КР: Гожий О.П., Калініна І.О. Методи та системи та машинного навчання: Навчальний посібник. – Миколаїв: Вид-во МДГУ ім. Петра Могили, 2024 знаходиться в Moodle3.

Приклад завдання на курсову роботу ЗАВДАННЯ на курсову роботу з дисципліни «Методи та системи машинного навчання» студенту _____ групи _____ Варіант № _____																															
1. Тема роботи Розробка моделі класифікації на основі структурованих даних.	3. Етапи обробки даних Збір даних, очистка даних, створення признаков, вибірка даних, заміщення даних, нормалізація даних, рефакторинг, створення моделі, оцінка моделі, оптимізація моделі, матриця неточностей, крива ROC, крива навчання, розгортання моделі. Зуваження! Для розв'язку задачі необхідно використання мови <i>Python</i> та бібліотеки <i>pandas</i> .																														
2. Вхідні дані для роботи Значення признаков та векторів задається в залежності від структурованої вибірки для кожного варіанту окремо. а) визначення формату структурованих даних, признаки $N_1, N_2, \dots, N_p$; б) вектори (об'єкти): $M_1, M_2, \dots, M_s$; в) значення признаку по кожному вектору. Приклад вхідних даних для завдання класифікації <i>Passenger</i>	4. Зміст пояснювальної записки • Анотація. • Завдання на курсову роботу. • Постановка задачі. • Розв'язання завдання класифікації на основі методів машинного навчання (задача класифікації об'єктів при певній ситуації). • Результат класифікації, його обґрунтування. • Висновки. • Список використаних джерел літератури (15-20 найменувань). 5. Перелік графічного матеріалу Пояснювальна записка містить рисунки, що забезпечують ілюстрацію результатів роботи, а також Power-Point презентацію (10-12 слайдів). Зуваження! Для захисту курсової роботи перед комісією кожен студент готує текст доповіді (до 5-7 хвилин).																														
<table border="1"><thead><tr><th>Позначення</th><th>Роз'яснення</th></tr></thead><tbody><tr><td><i>pclass</i></td><td>клас пасажирів (1 – перший, 2 – другий, 3 – третій)</td></tr><tr><td><i>survival</i></td><td>виживання (0 – ні, 1 – так)</td></tr><tr><td><i>name</i></td><td>ім'я</td></tr><tr><td><i>sex</i></td><td>стать</td></tr><tr><td><i>age</i></td><td>вік</td></tr><tr><td><i>sibsp</i></td><td>кількість братів, сестер або подружжя на борту</td></tr><tr><td><i>parch</i></td><td>кількість батьків або дітей на борту</td></tr><tr><td><i>ticket</i></td><td>номер квитка</td></tr><tr><td><i>fare</i></td><td>тариф пасажирів</td></tr><tr><td><i>cabin</i></td><td>каюта</td></tr></tbody></table>	Позначення	Роз'яснення	<i>pclass</i>	клас пасажирів (1 – перший, 2 – другий, 3 – третій)	<i>survival</i>	виживання (0 – ні, 1 – так)	<i>name</i>	ім'я	<i>sex</i>	стать	<i>age</i>	вік	<i>sibsp</i>	кількість братів, сестер або подружжя на борту	<i>parch</i>	кількість батьків або дітей на борту	<i>ticket</i>	номер квитка	<i>fare</i>	тариф пасажирів	<i>cabin</i>	каюта	<table border="1"><tbody><tr><td><i>embarked</i></td><td>точка посадки (C – Шербур, Q – Квінстаун, S – Саутгемптон)</td></tr><tr><td><i>boat</i></td><td>рятувальна шлюпка</td></tr><tr><td><i>body</i></td><td>ідентифікаційний номер тіла</td></tr><tr><td><i>home.dest</i></td><td>дім або місце призначення</td></tr></tbody></table>	<i>embarked</i>	точка посадки (C – Шербур, Q – Квінстаун, S – Саутгемптон)	<i>boat</i>	рятувальна шлюпка	<i>body</i>	ідентифікаційний номер тіла	<i>home.dest</i>	дім або місце призначення
Позначення	Роз'яснення																														
<i>pclass</i>	клас пасажирів (1 – перший, 2 – другий, 3 – третій)																														
<i>survival</i>	виживання (0 – ні, 1 – так)																														
<i>name</i>	ім'я																														
<i>sex</i>	стать																														
<i>age</i>	вік																														
<i>sibsp</i>	кількість братів, сестер або подружжя на борту																														
<i>parch</i>	кількість батьків або дітей на борту																														
<i>ticket</i>	номер квитка																														
<i>fare</i>	тариф пасажирів																														
<i>cabin</i>	каюта																														
<i>embarked</i>	точка посадки (C – Шербур, Q – Квінстаун, S – Саутгемптон)																														
<i>boat</i>	рятувальна шлюпка																														
<i>body</i>	ідентифікаційний номер тіла																														
<i>home.dest</i>	дім або місце призначення																														

4.4. Забезпечення освітнього процесу

Заняття з дисципліни проводяться у дистанційному форматі (GoogleMeet, Zoom тощо) або у комп'ютерних класах з використанням:

- PC зі швидкісним доступом до мережі Internet;
- ОС Windows 10;
- Браузери: Chrome / Opera / Mozilla Firefox / MS Edge.
- Ноутбук, проектор, екран.
- Комплект слайд-презентацій по курсу.
- Програмне забезпечення для демонстрацій слайд-презентацій.
- Мови та середовища програмування Python, R, RStudio

5. Підсумковий контроль

Кожне залікове завдання складається з теоретичної та практичної частини. Перелік теоретичних питань наведений нижче:

I. Основи машинного навчання

1. Що таке машинне навчання та чим воно відрізняється від традиційного програмування?
2. Назвіть три основні категорії машинного навчання та коротко опишіть кожну з них.
3. Що таке "ознака" (feature) та "мітка" (label) у контексті машинного навчання?
4. Поясніть різницю між завданнями класифікації та регресії. Наведіть по одному прикладу для кожного.
5. Для чого використовуються навчальний, валідаційний та тестовий набори даних?
6. Що таке функція втрат (loss function) і яка її роль у процесі навчання моделі?
7. Що таке параметричні та непараметричні моделі машинного навчання?
8. Поясніть концепцію "прокляття розмірності" (curse of dimensionality).
9. Опишіть процес інжинірингу ознак (feature engineering). Чому він важливий?
10. Що таке нормалізація та стандартизація даних? Коли їх слід застосовувати?

II. Навчання з учителем (*Supervised Learning*)

11. Опишіть принцип роботи алгоритму лінійної регресії.
12. У чому полягає основна ідея логістичної регресії та для яких завдань вона використовується?
13. Як працює метод k-найближчих сусідів (k-Nearest Neighbors, k-NN)? Які його переваги та недоліки?
14. Поясніть концепцію машини опорних векторів (Support Vector Machine, SVM). Що таке гіперплощина, що розділяє?
15. Що таке "трюк з ядром" (kernel trick) в алгоритмі SVM?
16. Опишіть процес побудови дерева рішень (Decision Tree).
17. Які критерії використовуються для розщеплення вузлів у деревах рішень (напр., індекс Джині, ентропія)?
18. Як працює алгоритм Наївного Баєса (Naive Bayes)? У чому полягає його "наївність"?
19. Що таке перенавчання (overfitting) та недонавчання (underfitting) моделі?
20. Поясніть компроміс між зміщенням та дисперсією (bias-variance tradeoff).
21. Що таке регуляризація? Опишіть L1 (Lasso) та L2 (Ridge) регуляризацію.
22. Що таке ансамблеві методи? Назвіть основні підходи до їх створення.
23. Поясніть різницю між беггінгом (bagging) та бустингом (boosting).
24. Як працює алгоритм випадкового лісу (Random Forest)?
25. Опишіть основну ідею градієнтного бустингу (Gradient Boosting).

III. Навчання без учителя (*Unsupervised Learning*)

26. Які основні завдання вирішує навчання без учителя?
27. Опишіть алгоритм кластеризації K-середніх (K-Means).
28. Як можна визначити оптимальну кількість кластерів (k) для алгоритму K-Means?
29. Поясніть принцип роботи ієрархічної кластеризації. Чим відрізняються агломеративний та дивізійний підходи?
30. Як працює алгоритм кластеризації DBSCAN? Які його ключові параметри?
31. Яка мета методів зниження розмірності?
32. Опишіть метод головних компонент (Principal Component Analysis, PCA).
33. Для чого використовується алгоритм t-SNE (t-distributed Stochastic Neighbor Embedding)?
34. Що таке пошук асоціативних правил? Наведіть приклад.
35. Поясніть метрики "підтримка" (support), "достовірність" (confidence) та "підйом" (lift).

IV. Нейронні мережі та глибоке навчання

36. Що таке перцептрон і як він працює?
37. Яка роль функцій активації в нейронних мережах? Наведіть приклади.
38. Опишіть архітектуру багатoshарового перцептрона (Multi-Layer Perceptron, MLP).
39. Поясніть принцип роботи алгоритму зворотного поширення помилки (Backpropagation).

40. Що таке згорткові нейронні мережі (Convolutional Neural Networks, CNN) і для яких завдань вони є ефективними?
41. Які основні шари використовуються в CNN (згортковий, пулінг)?
42. Що таке рекурентні нейронні мережі (Recurrent Neural Networks, RNN)? У чому їхня особливість?
43. Поясніть проблему зникаючого та вибухаючого градієнта в RNN.
44. Що таке архітектури LSTM та GRU? Яку проблему вони вирішують?
45. Поясніть механізм уваги (attention mechanism) у глибокому навчанні.
46. Що таке трансферне навчання (transfer learning)?

V. Оцінка моделей та системи машинного навчання

47. Поясніть метрики для оцінки класифікаторів: точність (accuracy), влучність (precision), повнота (recall), F1-міра.
48. У яких випадках точність (accuracy) може бути поганою метрикою для оцінки моделі?
49. Що таке матриця помилок (confusion matrix)?
50. Поясніть, що таке ROC-крива (Receiver Operating Characteristic) та метрика AUC (Area Under the Curve).
51. Що таке перехресна валідація (cross-validation) і навіщо вона потрібна?
52. Опишіть процедуру k-блокової перехресної валідації (k-fold cross-validation).
53. Що таке гіперпараметри моделі? Чим вони відрізняються від параметрів?
54. Опишіть методи пошуку оптимальних гіперпараметрів: пошук по сітці (Grid Search) та випадковий пошук (Randomized Search).
55. Що таке навчання з підкріпленням (Reinforcement Learning)? Назвіть його ключові елементи.
56. Поясніть різницю між розвідуванням (exploration) та експлуатацією (exploitation) у навчанні з підкріпленням.
57. Що таке MLOps? Назвіть основні етапи життєвого циклу моделі в MLOps.
58. Для чого потрібні системи контролю версій даних (напр., DVC)?
59. Яка роль контейнеризації (напр., Docker) у розгортанні моделей машинного навчання?
60. Що таке AutoML і які завдання він покликаний вирішити?

Типові тестові завдання

1. Який тип машинного навчання використовується, коли алгоритм навчається на даних, що не мають попередньо визначених міток або категорій?
А) Навчання з учителем (Supervised Learning);
Б) Навчання без учителя (Unsupervised Learning);
В) Напівконтрольоване навчання (Semi-supervised Learning);
Г) Навчання з підкріпленням (Reinforcement Learning).
2. Яке завдання вирішує модель, що прогнозує ціну будинку на основі його характеристик (площа, кількість кімнат тощо)?
А) Асоціація;
Б) Кластеризація;
В) Класифікація;
Г) Регресія.
3. Що таке 'перенавчання' (overfitting) в машинному навчанні?
А) Модель ідеально працює як на навчальних, так і на тестових даних;
Б) Процес навчання моделі займає занадто багато часу;
В) Модель занадто складна, вона 'запам'ятовує' навчальні дані, включаючи шум, і погано узагальнює на нових даних;
Г) Модель занадто проста і не може вловити основні закономірності в даних.

4. В основі якого алгоритму лежить побудова ієрархічної структури з правил 'якщо-тоді'?
- А) Нейронна мережа;
 - Б) Дерево рішень (Decision Tree);
 - В) Метод k-найближчих сусідів (k-NN);
 - Г) Машина опорних векторів (SVM)
5. Який елемент є базовим будівельним блоком нейронної мережі?
- А) Кластер;
 - Б) Нейрон (або перцептрон);
 - В) Піксель;
 - Г) Вектор.
6. Алгоритм К-середніх (K-Means) є прикладом якого типу завдань?
- А) Навчання з підкріпленням;
 - Б) Регресія;
 - В) Класифікація;
 - Г) Кластеризація.
7. Який метод навчання моделі передбачає взаємодію 'агента' з 'середовищем' для максимізації 'винагороди'?
- А) Генетичні алгоритми;
 - Б) Навчання без учителя;
 - В) Навчання з підкріпленням;
 - Г) Навчання з учителем.
8. Для чого в основному використовується метод градієнтного спуску?
- А) Для розділення даних на навчальну та тестову вибірки;
 - Б) Для вибору найбільш значущих ознак у наборі даних;
 - В) Для візуалізації багатовимірних даних;
 - Г) Для пошуку мінімуму функції втрат шляхом ітеративного коригування параметрів моделі.
9. Що таке крос-валідація (cross-validation)?
- А) Метод для прискорення навчання нейронних мереж;
 - Б) Спосіб створення нових ознак з існуючих;
 - В) Алгоритм для боротьби з перенавчанням шляхом спрощення моделі;
 - Г) Процедура оцінки ефективності моделі, яка ділить дані на кілька частин (фолдів) і послідовно використовує їх для навчання та тестування.
10. Процес створення нових вхідних змінних (ознак) з наявних даних для покращення продуктивності моделі називається:
- А) Інжиніринг ознак (Feature Engineering);
 - Б) Вибір ознак (Feature Selection);
 - В) Нормалізація даних;
 - Г) Зменшення розмірності

Типові задачі для розв'язання

Практична робота № 0: Алгоритми класифікації. Алгоритм створення байесівського класифікатора

Вступ. Наївний байесівський: класифікатор (*Naive Bayes classifier*) - це простий класифікатор, заснований на використанні теореми Байеса, яка описує ймовірність події з урахуванням пов'язаних з ним умов. Такий класифікатор створюється за допомогою присвоювання

міток класів екземплярів задачі. Останні подаються у вигляді векторів значень ознак. При цьому вважається, що значення будь-якого заданої ознаки не залежить від значень інших ознак. Його припущення про незалежність розглянутих ознак і становить наївну частину байєсівського класифікатора.

Послідовність виконання практичної роботи.

Для створення наївного байєсівського класифікатора необхідно:

1. Створити новий файл *Python* і імпортувати наступні пакети.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.naive_bayes import GaussianNB
from sklearn import cross_validation

from utilities import visualize_classifier
```

2. Завантажити вхідні данні з файлу `data_multivar_nb(N).txt` За відповідним варіантом (N-номер варіанта за журналом).

3. Створити байєсівський класифікатор. В даному випадку використовується гауссовський наївний байєсовський класифікатор, в якому передбачається, що значення, асоційовані з кожним класом, діють за законом розподілу Гаусса.

```
# Створення наївного байєсівського класифікатора
classifier = GaussianNB()
```

4. Провести навчання класифікатора, використовуючи тренувальні дані.

Запустимо класифікатор на тренувальних даних і спрогнозуємо результати.

```
# Прогнозування значень для тренувальних даних
y_pred = classifier.predict(X)
```

5. Обчислити якість (accuracy) 1 класифікатора, порівнявши передбачені значення з істинними мітками, а потім візуалізувати результат.

```
# Обчислення якості класифікатора
accuracy = 100.0 * (y == y_pred).sum() / X.shape[0]
print("Accuracy of Naive Bayes classifier =", round(accuracy, 2), "%")

# Візуалізація результатів роботи класифікатора
visualize_classifier(classifier, X, y)
```

6. Підготувати звіт.

«0» варіант іспитового білету з зазначенням максимальної кількості балів за кожне виконане завдання:

№	Питання білету	Максимальна кількість балів
1	Головні завдання та проблеми машинного навчання.	10
2	Дерева прийняття рішень. Ансамблеві методи МН. Наведіть приклади.	15
3	Байєсівський класифікатор. Приклад використання.	15
	Всього	40

6. Критерії оцінювання та засоби діагностики результатів навчання

а) для денної форми навчання:

№	Вид діяльності (завдання)	Максимальна кількість балів
1	Практична робота №1	5
2	Практична робота №2	5
3	Практична робота №3	5
4	Практична робота №4	5
5	Практична робота №5	5

6	Практична робота №6	5
10	Виконання контрольного тестового завдання	5+5
11	Самостійна робота студента	20
12	Разом за семестр	60
13	Іспит	40
14	Всього	100

Критерії оцінювання завдань для досягнення максимальної кількості балів при виконанні практичних робіт

Максимальна кількість балів (відповідно до попередньої таблиці) – студент з високою якістю самостійно виконав весь обсяг робіт, відповідає на всі питання, пов’язані з виконаними роботами, та робить додаткові розрахунки, які йому пропонує викладач. У викладача немає претензій щодо реалізації та вимог до виконання роботи.

4 бали – студент з достатньою якістю виконав всі завдання, але в процесі роботи він робив деякі помилки, які, після вказування на них викладачем, самостійно виправляв. На деякі питання він відповідає з похибкою. Запропоновані викладачем додаткові розрахунки робить з деякою потугою. Не всі вимоги до виконання роботи дотримані.

3 бали – студент самостійно виконав всі роботи, але якість реалізації недостатня (помилки при розрахунках, не всі вимоги до роботи дотримані). На питання щодо виконання робіт відповідає не зовсім чітко. Є помилки при відповідях.

2-1 бали – студент самостійно виконав не всі роботи, при цьому якість реалізації недостатня (помилки при розрахунках, не дотримується вимог до оформлення роботи). На питання щодо виконання робіт відповідає не чітко. Є грубі помилки при відповідях.

0 балів – студент не виконав весь обсяг робіт, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам.

При отриманні незадовільної оцінки студент має право виправити всі помилки або виконати нові варіанти завдань, якщо викладач невпевнений, що студент виконав їх самостійно. Такий варіант пропонується, коли студент має багато пропусків занять.

Критерії оцінювання для досягнення максимальної кількості балів при виконанні контрольного тестового завдання

Оцінювання контрольного тестового завдання студентів по дисципліні «Методи та системи машинного навчання» проводиться по 5-бальній шкалі.

Тести допомагають отримати об'єктивніші оцінки рівня знань, умінь, навиків, перевірити відповідність вимог до підготовки студентів заданим стандартам, виявити пропуски в підготовці студентів.

Виходячи з технологічності процедури тестування відповіді кодуються двійковим кодом: **1** – істинно і **0** – помилково, і у такому вигляді можуть поступати в сучасні автоматизовані системи обробки інформації. Також може використовуватись відсоткове кодування відповіді (згідно з кількістю правильних відповідей у разі можливості надання декількох відповідей) відповідно до вагового коефіцієнту кожної відповіді з усіх можливих.

Тестування з дисципліни «Методи та системи машинного навчання» студенти можуть проходити як одноразово, так і необмежену кількість разів, поки не досягнуть результату, який відповідає їх уяві про власні знання. Чим буде більше наполегливим та зацікавленим студент - тим скоріше він досягне найвищого рівня професійних знань та вмінь. Така наполегливість також є складовою мети дисципліни.

Тестові бали переводяться в традиційну систему оцінок. Наприклад, якщо випробовуваний виконав більше 90 % завдань, то він отримує оцінку “відмінно” (5 балів), що вирішив від 75 до 90 % завдань “добре” (4 бали), від 50 до 75 % – “задовільно” (3 бали). Якщо студент виконав менш ніж 50% завдань, то він обов’язково повинен перескласти тестові завдання для отримання допуску до заліку.

Критерії оцінювання для досягнення максимальної

кількості балів при виконанні курсової роботи

Оцінювання курсової роботи студентів по дисципліні «Методи та системи машинного навчання» проводиться на основі рейтингової системи оцінювання знань по 100-бальній системі.

Таблиця. Шкала оцінювання курсової роботи з дисципліни «Методи та системи машинного навчання»

Оцінка в балах	Оцінка за національною шкалою	Оцінка за шкалою ECTS	
		Оцінка	Пояснення
90-100	Відмінно	A	<i>Відмінно</i> студент з високою якістю самостійно виконав весь обсяг завдань курсової роботи, відповідає на всі питання, пов'язані з виконаними роботами. При захисті курсової роботи допускається лише незначна кількість помилок не принципового характеру. У викладача (та членів комісії) немає претензій щодо реалізації, оформленню та представленню курсової роботи.
82-89	Добре	B	<i>Дуже добре</i> студент виконав всі завдання курсової роботи, але в процесі роботи він робив деякі помилки, які, після вказування на них викладачем, самостійно виправляв. На деякі питання комісії він відповідає з похибкою або не дуже впевнено. Не всі вимоги до оформлення курсової роботи дотримані.
75-81	Добре	C	<i>Добре</i> студент виконував завдання курсової роботи вище середнього рівня з кількома суттєвими помилками, але після вказування на них викладачем, самостійно виправляв їх. На деякі питання комісії він відповідає з похибкою або з потугою. Не всі вимоги до виконання або оформлення курсової роботи дотримані.
67-74	Задовільно	D	<i>Задовільно</i> студент самостійно виконав всі завдання курсової роботи, але якість реалізації недостатня (помилки при розрахунках, не всі вимоги до роботи дотримані). На питання щодо виконання завдань відповідає не зовсім чітко. Є помилки при відповідях.
60-66	Задовільно	E	<i>Достатньо</i> студент самостійно виконав не всі завдання курсової роботи, при цьому якість задовольняє мінімальним критеріям (помилки при розрахунках, не дотримується вимог до оформлення роботи). На питання щодо виконання завдань відповідає не чітко. Є грубі помилки при відповідях.
35-59	Незадовільно	FX	<i>Незадовільно</i> (з можливістю обрання іншого варіанта завдання, його виконання та процедури захисту) студент виконав менш ніж 50% обсягу завдань курсової роботи, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам.

Оцінка в балах	Оцінка за національною шкалою	Оцінка за шкалою ECTS	
		Оцінка	Пояснення
1-34	Незадовільно	F	<i>Незадовільно</i> (з можливістю обрання іншого варіанта завдання, його виконання та процедури захисту) студент виконав менш ніж 25% обсягу завдань курсової роботи, або виконав з грубими помилками. Він має проблеми з розрахунками, не знає теоретичного матеріалу, програмна реалізація не відповідає поставленим вимогам, або вона відсутня.

Студент, який за виконання та захист курсової роботи набрав менше, ніж 60 балів, обирає інший варіант завдання для курсової роботи та виконує його.

7. Рекомендовані джерела інформації

7.1. Основні:

1. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2019.
2. Esposito D., Esposito F. Introducing Machine Learning. Pearson Education, EBook, 2020. 400 p.
3. Geron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. O'Reilly Media, Inc., 2019. 688 p.
4. Hastie T., Tibshirani R, Friedman J. The Elements of Statistical Learning (2nd edition). Springer, 2019.
5. Kelleher J. D., Mac Namee B., D'Arcy A. Fundamentals of Machine Learning for Predictive Data Analytics. Algorithms, Worked Examples, and Case Studies. Second Edition. The MIT Press, Cambridge, Massachusetts, London, England, 2020. 798 p.
6. Lakshmanan V., Robinson S., Munn M. Machine Learning Design Patterns: Solutions to Common Challenges in Data Preparation, Model Building, and MLOps 1st Edition. O'Reilly Media, Inc., 2020. 448 p.
7. Lantz B. Machine Learning with R. Expert techniques for predictive modeling, 3rd Edition, Packt Publishing, 2019. 458 p.
8. Murphy K. Machine Learning: A Probabilistic Perspective. MIT Press, 2019.
9. Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. Cambridge University Press, 2019.
10. Кононова К. Ю. Машинне навчання. – Харків: ХНУ ім. В. Н. Каразіна, 2020. – 301 с.
https://moodle.znu.edu.ua/pluginfile.php/593075/mod_folder/intro/Базовий%20підручник_2%20%208Кононова%20К.%20Ю.%20Машинне%20навчання%20-%20методи%20та%20моделі%29.pdf
- 11.

7.2 Додаткові:

1. Gozhyj A., Nechakhin V., Kalinina I. Solar Power Control System based on Machine Learning Methods. 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT). 23-26 September 2020. Zbarazh, Ukraine. Vol. 1, pp. 24 – 27. Publisher: IEEE 21 January 2021. DOI: 10.1109/CSIT49958.2020.9321953.
2. Kalinina I., Gozhyj A. Methodology for Solving Forecasting Problems Based on Machine Learning Methods. Lecture Notes on Data Engineering and Communications Technologies (Switzerland). 2023, Vol. 149, pp. 105-125. (ISSN: 2367 – 4512).
3. Bidyuk P., Kalinina I., Gozhyj A. An Approach to Identifying and Filling Data Gaps in Machine Learning Procedures. Lecture Notes on Data Engineering and Communications Technologies (Switzerland). 2022. Vol. 77, pp. 164-176. (ISSN: 2367- 4512).
4. Bidyuk, P., Kalinina, I., Zhebko, O., Gozhyj, A., Hannichenko, T. Classification System Based on Ensemble Methods for Solving Machine Learning Tasks. CEUR Workshop Proceedings, 2023, 3426, pp. 1–11. CEUR-WS.org/Vol-3426/paper5.pdf.

5. Kalinina, I., Bidyuk, P., Gozhyj, A., Malchenko, P. Combining Forecasts Based on Time Series Models in Machine Learning Tasks. CEUR Workshop Proceedings, 2023, 3426, pp. 25–35. CEUR-WS.org/Vol-3426/paper2.pdf.
6. Kalinina I., Gozhyj A. Methodology for Solving Forecasting Problems Based on Machine Learning Methods. Lecture Notes on Data Engineering and Communications Technologies (Switzerland). 2023. Vol. 149, pp. 105-125. (ISSN: 2367 – 4512). (bookchapter, Scopus).

7.3 Інтернет-ресурси

1. Interactive course. Deep Learning in Python [Електронний ресурс]. – Режим доступу: <https://www.datacamp.com/courses/deep-learning-in-python>.
2. Machine Learning [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/learn/machine-learning>.
3. Datasets [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/datasets>.
4. Машинне навчання і нейронні мережі: бібліотека PHP-ML [Електронний ресурс]. – Режим доступу: <http://echo.lviv.ua/dev/5469>.

7.4 Курси

1. Machine Learning, Data Science and Generative AI with Python - <https://www.udemy.com/course/data-science-and-machine-learning-with-python-hands-on/?couponCode=ST14MT32124>
2. Python for Data Science and Machine Learning Bootcamp - <https://www.udemy.com/course/python-for-data-science-and-machine-learning-bootcamp/?couponCode=ST14MT32124>
3. Професійний сертифікат IBM Machine Learning - https://www.coursera.org/professional-certificates/ibm-machine-learning?irclickid=wk7RwhV-LxyPUoxz6aSAq3DPUkHyn2TIPVKSUg0&irgwc=1&utm_medium=partners&utm_source=impact&utm_campaign=3310965&utm_content=b2c