

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЗАТВЕРДЖЕНО

на засіданні кафедри
інформаційних систем.
Протокол № 1 від 22.08.2023 р.

ПОГОДЖЕНО

Проректор з навчально-методичної роботи



Каріна НЕМАШКАЛО

ВСТУП ДО МАШИННОГО НАВЧАННЯ
робоча програма навчальної дисципліни (РПНД)

Галузь знань 12 "Інформаційні технології"
Спеціальність 121 "Інженерія програмного забезпечення"
Освітній рівень перший (бакалаврський)
Освітня програма "Інженерія програмного забезпечення"

Статус дисципліни вибіркова
Мова викладання, навчання та оцінювання українська

Розробник:
к.ф.-м.н., доцент

підписано КЕП

Віктор ЗАДАЧИН

Завідувач кафедри
інформаційних систем

Дмитро БОНДАРЕНКО

Гарант програми

Олег ФРОЛОВ

Харків
2024

ВСТУП

Сучасний розвиток інформаційних технологій характеризується все більш зростаючим рівнем використання інтелектуальних інформаційних систем. Навчальна дисципліна "Вступ до машинного навчання" є ключовим етапом в оволодінні здобувачами основами сучасної області інформаційних технологій, що вивчає теорії та методики машинного навчання.

Метою викладання даної навчальної дисципліни є надання здобувачам вищої освіти спеціальних знань щодо базових концепцій, методів та алгоритмів машинного навчання, а також розвиток їхньої здатності застосовувати ці знання на практиці, наприклад, для проектування інтелектуальних інформаційних систем.

Завданнями навчальної дисципліни є: вивчення основних алгоритмів машинного навчання, надбання навичок аналізу даних та роботи з інструментами машинного навчання.

Предметом вивчення дисципліни є теоретичні та прикладні аспекти машинного навчання, включаючи класичні та сучасні методи, техніки оптимізації, аналіз результатів та етичні аспекти використання алгоритмів машинного навчання.

Об'єктом вивчення дисципліни є процеси та методи, пов'язані із застосуванням комп'ютерних систем для навчання на основі даних. Розглядаються принципи вибору та підготовки даних, розроблення моделей, їхнє навчання та оцінювання ефективності.

У процесі навчання здобувачі вищої освіти отримують необхідні знання під час проведення аудиторних занять: лекційних та лабораторних. Також велике значення в процесі вивчення та закріплення знань має самостійна робота здобувачів.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
РН 01	ЗК 01, ЗК 02, ЗК 05, ЗК 06, СК 08
РН 05	ЗК 01, ЗК 02, СК 07, СК 08, СК 14
РН 18	ЗК01, ЗК02, ЗК05, СК07, СК14

де, РН 01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

РН 05. Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.

РН 18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.

ЗК 01. Здатність до абстрактного мислення, аналізу та синтезу.

- ЗК 02. Здатність застосовувати знання у практичних ситуаціях.
- ЗК 05. Здатність вчитися і оволодівати сучасними знаннями.
- ЗК 06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- СК 07. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.
- СК 08. Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення.
- СК 14. Здатність до алгоритмічного та логічного мислення.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст навчальної дисципліни

Змістовий модуль 1. Основні поняття і визначення машинного навчання.

Тема 1. Основи машинного навчання. Машинне навчання та штучний інтелект.

1.1. Data Mining, Big Data та математична статистика

1.2. Поняття штучного інтелекту, Machine Learning, Deep Learning та Data science.

Тема 2. Класи задач, які вирішують методами Machine Learning.

2.1. Регресія.

2.2. Класифікація та кластеризація.

2.3. Прогнозування часових рядів.

2.4. Пошук аномалій.

2.5. Розпізнавання образів.

Тема 3. Основні типи машинного навчання.

3.1. Навчання з вчителем.

3.2. Навчання без вчителем.

3.3. Напівконтрольоване навчання.

3.4. Навчання з підкріпленням.

Тема 4. Процес роботи з даними. Попередня обробка та візуалізація даних.

4.1. Статистичне оброблення даних.

4.2. Візуалізація даних.

4.3. Пропущені та аномальні дані.

4.4. Масштабування даних.

Тема 5. Класичні методи розв'язання задач регресійного аналізу та прогнозування.

5.1. Методи оцінки параметрів регресійної моделі.

5.2. Методи прогнозування часових рядів.

5.3. Оцінка якості моделі.

Тема 6. Класичні методи розв'язання задач класифікації та кластеризації.

6.1. Метрики.

6.2. Задача класифікації. Дерево рішень, наївний Байєс, метод k найближчих сусідів.

6.3. Задача кластеризації. Методи k -means, fuzzy c -means.

6.4. Оцінка якості моделі.

Змістовий модуль 2. Нейронні мережі

Тема 7. Апарат штучних нейронних мереж

7.1. Штучний нейрон. Види передаточних (активаційних) функцій. Архітектура нейромережі.

7.2. Навчання штучної нейронної мережі.

7.3. Перцептрон Розенблата. Мережа MLP.

7.4. Класи нейронних мереж. Мережі прямого поширення. Рекурентні нейромережі. Згорткові нейромережі.

Тема 8. Розв'язання задач регресії за допомогою нейронних мереж.

8.1. Етапи побудови моделі. Пошук архітектури, навчання та оцінка якості моделі. Застосування моделі.

8.2. Програмні інструменти для побудови регресійної моделі.

Тема 9. Розв'язання задач класифікації за допомогою нейронних мереж.

9.1. Етапи побудови моделі. Пошук архітектури, навчання та оцінка якості моделі. Застосування моделі.

9.2. Програмні інструменти для побудови моделі класифікації.

Тема 10. Розв'язання задач кластеризації за допомогою нейронних мереж.

10.1. Нейромережа Кохонена (SOM).

10.2. Етапи побудови моделі. Пошук архітектури, навчання та оцінка якості моделі. Застосування моделі.

10.3. Програмні інструменти для побудови моделі кластеризації.

Тема 11. Розв'язання задач прогнозування часових рядів за допомогою нейронних мереж.

11.1. Етапи побудови моделі. Пошук архітектури, навчання та оцінка якості моделі. Застосування моделі.

11.2. Програмні інструменти для побудови моделі прогнозування часових рядів.

Тема 12. Розв'язання задач розпізнавання образів за допомогою нейронних мереж.

12.1. Мережі Хопфілда та ВММ.

12.2. Етапи побудови моделі. Пошук архітектури, навчання та оцінка якості моделі. Застосування моделі.

12.3. Програмні інструменти для побудови моделі розпізнавання.

Перелік лабораторних занять за навчальною дисципліною наведено в табл. 2.

Перелік лабораторних занять

Назва теми	Зміст
Тема 2-5, 7, 8.	Розв'язання задач регресії методами Machine Learning.
Тема 2-4, 6, 7, 9.	Розв'язання задач класифікації методами Machine Learning.
Тема 2-4, 6, 7, 10.	Розв'язання задач кластеризації методами Machine Learning.
Тема 2, 5, 7, 11.	Розв'язання задач прогнозування часових рядів методами Machine Learning.
Тема 2, 7, 12.	Розв'язання задач розпізнавання образів за допомогою нейронних мереж.

Перелік самостійної роботи за навчальною дисципліною наведено в табл. 3.

Таблиця 3

Перелік самостійної роботи

Назва теми	Зміст
Тема 1 – 12	Вивчення лекційного матеріалу
Тема 2 – 12	Підготовка до лабораторних занять
Тема 1 – 12	Підготовка до екзамену

Кількість годин лекційних та лабораторних занять, а також годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

МЕТОДИ НАВЧАННЯ

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

Словесні (лекція (Тема 1, 2, 3, 4, 5, 6, 7, 8, 9), проблемна лекція (Тема 10, 12), лекція-візуалізація (Тема 11)).

Наочні (демонстрація (Тема 1-12)).

Лабораторна робота (Тема 4, 5, 6, 7, 8, 9, 10, 11, 12), кейс-метод (Тема 5).

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

Поточний контроль здійснюється під час проведення лекційних та лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів:

– для дисциплін з формою семестрового контролю екзамен (іспит): максимальна сума – 60 балів; мінімальна

сума, що дозволяє здобувачу вищої освіти скласти екзамен (іспит) – 35 балів.

Підсумковий контроль включає семестровий контроль та атестацію здобувача вищої освіти.

Семестровий контроль проводиться у формі семестрового екзамену (іспиту). Складання семестрового екзамену (іспиту) здійснюється під час екзаменаційної сесії.

Максимальна сума балів, яку може отримати здобувач вищої освіти під час екзамену (іспиту) – 40 балів. Мінімальна сума, за якою екзамен (іспит) вважається складеним – 25 балів.

Підсумкова оцінка за навчальною дисципліною визначається сумуванням балів за поточний та підсумковий контроль.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: захист лабораторних робіт (50 балів), письмові контрольні роботи (10 балів).

Семестровий контроль: Екзамен (40 балів).

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

Приклад екзаменаційного білета та критерії оцінювання для навчальної дисципліни.

Приклад екзаменаційного білета

Харківський національний економічний університет імені Семена Кузнеця
Перший (бакалаврський) рівень вищої освіти
Спеціальність «Інженерія програмного забезпечення»
Освітньо-професійна програма «Інженерія програмного забезпечення»
Семестр VII
Навчальна дисципліна «Вступ до машинного навчання»

ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № 1

Завдання (евристичне, 40 балів).

Провести кластеризацію багатовимірних об'єктів для даних «Adult» з Репозитарію «UCI Machine Learning repository» (<http://archive.ics.uci.edu/ml/index.php>), де вибірці даних відповідає папка з відповідною назвою, в якій знаходяться файл з описом предметної області та безпосередньо файл з даними.

При записі результатів виконання завдання необхідно дотримуватися наступних вимог: дати формулювання загальної постановки задачі до якої відноситься завдання; пояснити сенс основних атрибутів; провести підготовку даних; побудувати модель; провести оцінку якості моделі на тестових даних; застосувати модель для розв'язання завдання; дати інтерпретацію отриманих результатів; побудувати графіки, які необхідні для інтерпретації, як вхідних даних, так і отриманих результатів.

Затверджено на засіданні кафедри інформаційних систем протокол № ____ від «__» _____ 20__ р.

Екзаменатор

к.ф.-м.н., доц. Задачин В.М.

Критерії оцінювання

Екзаменаційний білет складається з одного евристичного завдання. Підсумкова оцінка за іспит - максимум 40 балів.

В результаті виконання завдання повинно бути два файли: перший в форматі DOC (пояснення та скриншоти результатів роботи програми), другий – скрипт програми (Python або R), що реалізує виконання завдання. Ці файли потрібно прикріпити, як відповідь на завдання на курсі ПНС.

При записі результатів виконання завдання (в першому файлі) необхідно дотримуватися наступних вимог: дати формулювання загальної постановки задачі до якої відноситься завдання; пояснити сенс основних атрибутів; провести підготовку даних; побудувати модель; провести оцінку якості моделі на тестових даних; застосувати модель для розв'язання завдання; дати інтерпретацію отриманих результатів; побудувати графіки, які необхідні для інтерпретації, як вхідних даних, так і отриманих результатів.

Завдання оцінюється згідно шкали представленої в табл. 4.

Таблиця 4

Критерії оцінювання евристичного завдання

40 балів	Завдання виконане в повному обсязі. Представлено скрипт програми, отримані результати, проведений аналіз отриманих результатів з прив'язкою до сутності задачі, дані відповіді на поставлені в умові задачі питання. Зроблені висновки. Дотримані всі вимоги до запису результатів виконання завдання.
35 балів	Завдання виконане в повному обсязі. Представлено скрипт програми, отримані результати, проведений аналіз отриманих результатів з прив'язкою до сутності задачі, дані відповіді на поставлені в умові задачі питання. Зроблені висновки. Дотримані не всі вимоги до запису результатів виконання.
30 балів	Завдання виконане в повному обсязі. Представлено скрипт програми, отримані результати, проведений аналіз отриманих результатів з прив'язкою до сутності задачі, дані відповіді не на всі поставлені в умові задачі питання. Зроблені висновки. Дотримані не всі вимоги до запису результатів виконання.
25 балів	Завдання виконане. Є скрипт програми, програма працює, але є незначні невідповідності умові поставленої задачі. Отримані результати, але не проведений аналіз отриманих результатів, немає відповідей на поставлені в умові задачі питання. Дотримані не всі вимоги до запису результатів виконання.
20 балів	Є скрипт програми, але вона не працює. Дотримані не всі вимоги до запису результатів виконання завдання, але приведена логіка моделювання.
10 балів	Завдання не виконане. Є фрагмент скрипту програми і приведена логіка моделювання. Вимоги до запису результатів виконання завдання не дотримані.
0 балів	Завдання не виконане.

Завдання може бути оцінено і в проміжні бали (наприклад, 24-29, 31-34, 36-39) в залежності від наявності та ступеня детальності пояснень до процесу виконання завдання.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Нікольський Ю. В. Системи штучного інтелекту: Навчальний посібник / Ю. В. Нікольський, В.В. Пасічник, Ю. М. Щербина. – Київ: Магнолія, 2021. – 280 с.
2. Троцько В.В. Методи штучного інтелекту: навчально-методичний і практичний посібник / В.В. Троцько. – Київ: Університет "КРОК", 2020. – 86 с.
3. Попередня обробка та аналіз даних: лабораторний практикум для студ. спеціальності 113 «Прикладна математика» /Уклад.: Н. Е. Кондрук. Ужгород: УжНУ, 2023. - 41 с.

Додаткова

4. Stephan S. Jones, Frank M. Groom Artificial Intelligence and Machine Learning for Business for Non-Engineers. – CRC Press, 2019. – 148 p.
5. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2nd Edition – O’Reilly Media, 2019. –856 p.
6. Burkov A. The Hundred-Page Machine Learning Book – Andriy Burkov, 2019. –141 p.
7. Задачин В.М. Моделювання систем та методи оптимізацій: методичні рекомендації до лабораторних робіт для студентів галузі знань 12 "Інформаційні технології" першого (бакалаврського) рівня: [Електронне видання] – Харків: Вид. ХНЕУ ім. С. Кузнеця, 2019. – 210 с. <http://repository.hneu.edu.ua/handle/123456789/22458>

Інформаційні ресурси

8. Introduction to Mathematical Optimization : with Python – <https://indrag49.github.io/Numerical-Optimization/>
9. Colaboratory – <https://colab.research.google.com/>
10. Numerical Programming with Python – <https://python-course.eu/numerical-programming/>