



**ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**
Факультет інформаційних технологій
Кафедра комп'ютерних наук та інформаційних систем

СИЛАБУС (SYLLABUS)
Дисципліна «Машинне навчання/Machine learning»

ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА

Викладач	Філімонова Тетяна Олегівна
Науковий ступінь	Кандидат фізико-математичних. наук
Вчене звання	Доцент
Посада	Доцент кафедри комп'ютерних наук та інформаційних систем
Адреса кафедри	м.Київ, вул. Кіото 19, каб. Б-507, Б-526
E-mail	compdep@knute.edu.ua
Консультації	Відповідно до графіку індивідуальних консультацій на сайті кафедри

ПОЛІТИКА АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ

<https://knute.edu.ua/file/MzEyMQ==/c12a9f74e87d9154696ca0f761da2e5c.pdf>

Дотримання академічної доброчесності передбачає:

- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);
- посилання на джерела інформації у разі використання не авторських ідей, розробок, тверджень, відомостей і т.п.;
- дотримання норм законодавства про авторське право і суміжні права;
- надання достовірної інформації про результати власної наукової діяльності, використанні методики досліджень і джерела інформації.

Порушенням академічної доброчесності вважається:

- академічний плагіат – оприлюднення (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства;
- самоплагіат – оприлюднення (частково або повністю) власних раніше опублікованих наукових результатів як нових наукових результатів;
- фабрикація – вигадкування даних чи фактів, що використовуються в наукових дослідженнях;
- фальсифікація – свідомо зміна чи модифікація вже наявних даних, що стосуються наукових досліджень.

За порушення академічної доброчесності здобувачі освіти можуть бути притягнені до академічної відповідальності:

- повторне проходження оцінювання (модульний контроль, іспит, залік тощо);
- повторне проходження відповідного освітнього компонента освітньо-професійної програми;
- відрахування з Університету;
- позбавлення наданих університетом пільг;
- відмова у присудженні відповідного ступеня вищої освіти;

ПОЛІТИКА ЩОДО ВІДВІДУВАННЯ ЗАНЯТЬ

- відвідування занять є обов'язковим;
- за об'єктивних причин (наприклад, хвороба, міжнародне стажування та ін.) навчання може відбутись в он-лайн формі за погодженням із викладачем дисципліни.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назва дисципліни / тип дисципліни	Машинне навчання/ обов'язкова
Навчальний рік	2023-2024
Факультет	Факультет інформаційних технологій
Курс	3
Семестр	6
Освітній ступінь	Бакалавр
Галузь знань	12 «Інформаційні технології»
Спеціальність	126 «Інформаційні системи та технології»
Загальна характеристика	Кількість годин –180 Кількість кредитів – 6 Види занять: лекції, лабораторні, самостійна робота. Співвідношення аудиторних годин і годин самостійної роботи - 78/102 Мова викладання – українська Форма викладання – очна
Підсумковий контроль	Екзамен
Програмне забезпечення	Python 3.12.1, Jupyter Notebook, Google Colab
Обладнання	Проектор, комп'ютерна техніка із встановленим програмним забезпеченням та доступом до мережі Інтернет.
Необхідні попередні дисципліни	«Чисельні методи програмування»; «Штучний інтелект»
Методика вивчення	Методика вивчення дисципліни полягає у набутті студентами знань теоретичного і практично-прикладного характеру під час лекцій, лабораторних занять, самостійної роботи та вивчення першоджерел і навчально-методичної літератури.
Мета і завдання	Метою вивчення дисципліни «Машинне навчання» є засвоєння основ методології машинного навчання і надання студентами теоретичних знань та формування практичних навичок побудови систем машинного навчання з метою аналізу складних соціально-економічних систем. Завданням вивчення дисципліни «Машинне навчання» є теоретична та практична підготовка з питань: сутності понять і категорій методології машинного навчання та сфер його застосування; організації процесу машинного навчання; застосування алгоритмів машинного навчання; застосування сучасних бібліотек машинного навчання; побудови етапів розробки системи машинного навчання; застосування різних типів процесу машинного навчання; методів вирішення основних типів задач машинного навчання - регресія, кластеризація, класифікація, зниження розмірності; застосування різних механізмів навчання нейронних мереж; визначення адекватності побудованих моделей машинного навчання.
Місце дисципліни в освітньо-професійній програмі	
Загальні	КЗ 5. Здатність вчитися і оволодівати сучасними знаннями.

компетентності	КЗ 6. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.
Фахові компетентності (результати навчання)	КС 4 Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші). КС 6 Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.
Програмні результати навчання	ПР 6. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності. ПР 7. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій.

ТЕМАТИКА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Тема 1. Загальні поняття дисципліни «Машинне навчання» (Machine learning).

Теорія обчислювального навчання. Типи навчання. Індуктивне навчання на основі емпіричних даних (навчальній вибірці). Дедуктивне навчання на основі формалізації знань експертів. Типи машинного навчання: навчання з вчителем, навчання без вчителя, навчання з підкріпленням. Задачі машинного навчання: кластеризація, регресія, зниження розмірності.

Тема 2. Введення в машинне навчання на мові програмування Python.

Машинне навчання та Python. Переваги використання Python в процедурах машинного навчання. Інсталяція Python. Пакет прикладних програм Anaconda X.X. Загальні поняття мови програмування Python. Вбудовані бібліотеки Python - Jupyter Notebook, NumPy, SciPy, Matplotlib, Pandas, Mllearn. Бібліотека Scikit-learn – призначення та інсталяція. Бібліотека TensorFlow – призначення та інсталяція. Бібліотека Theano – призначення та інсталяція. Бібліотека CUDA – призначення та інсталяція. Бібліотека Keras – призначення та інсталяція. Бібліотека PyTorch – призначення та інсталяція.

Тема 3. Етапи розробки моделі машинного навчання.

Етап підготовки (представлення) даних. Ознакове описання об'єкту. Вибір ключових ознак. Етап конструювання алгоритму (вибір типу задачі). Етап тренування та тестування моделі. Перенавчання. Тренувальна та тестувальна вибірки на прикладі нейронної мережі. Типи ознак (бінарні, номінальні, порядкові, кількісні). Етап валідації алгоритму на тестових даних. Функціонал якості, квадратична та абсолютна похибки. Мінімізація помилки. Зміна сили зв'язків.

Тема 4. Методи вирішення типових задач в машинному навчанні.

Візуалізація даних (використання бібліотеки Matplotlib). Лінійна алгебра (вектори, матриці, тензори). Статистика (Опис одиночного набору даних. Показники центру розподілу. Показники варіації. Кореляція. Парадокс Сімпсона). Теорія ймовірностей (Залежність і незалежність. Умовна ймовірність. Теорема Байеса. Випадкові величини. Неперервні розподіли. Нормальний розподіл. Центральна гранична теорема). Гіпотеза і висновок. Перевірка статистичних гіпотез. Р-значення. Довірчі інтервали. Підгонка Р-значення. Проведення А/В-тестування. Байєсівський статистичний висновок. Лінійна регресія. Градієнтний спуск. Логістична регресія. Навчання логістичної регресії. Софт-макс регресія.

Тема 5. Машинне навчання з учителем.

Класифікація і регресія. Узагальнююча здатність, перенавчання та недонавчання. Взаємозв'язок між складністю моделі і розміром набору даних. Алгоритми машинного навчання з учителем: набори даних, метод k найближчих сусідів, лінійні моделі, наївні байєсовські класифікатори, дерева рішень. Ансамблі дерев рішень, метод опорних векторів, нейронні мережі. Оцінки

невизначеності для класифікаторів. Функція рішень. Прогнозування ймовірностей. Невизначеність в мультикласовій класифікації.

Тема 6. Машинне навчання без учителя.

Типи машинного навчання без учителя. Проблеми машинного навчання без учителя. Попередня обробка даних і масштабування. Різні види попередньої обробки даних. Застосування перетворень даних. Масштабування навчального і тестового наборів. Вплив попередньої обробки даних на машинне навчання без учителя. Зниження розмірності, виділення ознак і множинне навчання. Аналіз головних компонент (PCA). Факторизація невід'ємних матриць (NMF). Множинне навчання за допомогою алгоритму t-SNE. Кластеризація. Кластеризація k-середніх. Агломеративна кластеризація. Кластеризація DBSCAN. Порівняння і оцінка якості алгоритмів кластеризації.

Тема 7. Машинне навчання з підкріпленням.

Навчання з підкріпленням. Етапи та основні проблеми. Функція підкріплення. Середовища та ресурси. Кластеризація та зменшення розмірності в машинному навчанні з підкріпленням. Автокодувальники.

Тема 8. Нейронні мережі, як найпоширеніший метод машинного навчання.

Нейронні мережі. Перцептрони. Нейронні мережі прямого розповсюдження. Нейронні мережі зі згортою (конволютивні). Мах-пулінг. Процес навчання нейронної мережі методом зворотного поширення помилки. Оцінка помилки. Зворотне поширення помилки. Посилення зв'язків. Зміна величини кроку. Адаптивна зміна величини кроку. Локальні мінімуми. VDG-архітектура нейронної мережі. Нейронна мережа Елмана. Архітектура рекурентної нейронної мережі Елмана, приклад роботи. LSTM-архітектура нейронних мереж, приклад роботи.

Тема 9. Типи даних і конструювання ознак.

Категоріальні змінні. Пряме кодування (dummy-змінні). Біннінг, дискретизація, лінійні моделі і дерева. Взаємодії і поліноми. Одномірні нелінійні перетворення. Автоматичний відбір ознак. Одномірні статистики. Відбір ознак на основі моделі. Ітеративний відбір ознак. Застосування експертних знань.

Тема 10. Оцінка і поліпшення якості моделі.

Перехресна перевірка. Перехресна перевірка в scikit-learn. Переваги перехресної перевірки. Стратифікована k-блокова перехресна перевірка та інші стратегії. Решітчастий пошук. Небезпека перенавчання параметрів і перевірочний набір даних. Решітчастий пошук з перехресної перевіркою. Метрики якості моделі і їх обчислення. Метрики для бінарної класифікації. Метрики для мультикласової класифікації. Метрики регресії. Використання метрик оцінки для відбору моделі.

Тема 11. Об'єднання алгоритмів в ланцюги та конвеєри.

Відбір параметрів з використанням попередньої обробки. Побудова конвеєрів. Використання конвеєра, поміщеного в об'єкт GridSearchCV. Загальний інтерфейс конвеєра. Метод побудови конвеєрів за допомогою функції make_pipeline. Робота з атрибутами етапів. Оптимальні параметри етапів конвеєра. Вибір оптимальної моделі конвеєра.

Тема 12. Робота з текстовими і графічними даними.

Строкові типи даних. Приклад застосування: аналіз тональності кіно відгуків. Подання текстових даних у вигляді «кошика слів». Застосування моделі «кошика слів» до синтетичного набору даних. Модель «мішка слів» для кіноотзивов. Стоп-слова. Масштабування даних за допомогою TensorFlow-idf. Дослідження коефіцієнтів моделі. Модель «кошика слів» для послідовностей з декількох слів (n-грам). Просунута токенизація, стемінг і лематизація. Моделювання тем і кластеризація документів. Латентне розміщення Діріхле. Введення в обробку зображень. Завантаження та показ зображень. Бінаризація. Гаусове розмивання. Проста класифікація зображень. Обчислення ознак по зображенню. Створення власних ознак. Використання ознак для пошуку подібних зображень. Локальні представлення ознак. Прикладна задача: конвеєр розпізнавання облич. Ознаки в методі HOG. Метод HOG - простий детектор обличь.

**Перелік навчальних робіт студентів та оцінки їх у балах з дисципліни
«Машинне навчання»**

Види робіт	К-сть балів
Лабораторне заняття №1. Тема: «Загальні поняття дисципліни «Машинне навчання» (Machine learning)»	4
Лабораторне заняття №2. Тема: «Введення в машинне навчання на мові програмування Python»	4
Лабораторне заняття №3. Тема: «Етапи розробки моделі машинного навчання»	4
Лабораторне заняття №4. Тема: «Методи вирішення типових задач в машинному навчанні»	4
Лабораторне заняття №5. Тема: «Машинне навчання з учителем»	4
Лабораторне заняття №6. Тема: «Машинне навчання без учителя»	4
Лабораторне заняття №7. Тема: «Машинне навчання з підкріпленням»	4
Лабораторне заняття №8. Тема: «Нейронні мережі, як найпоширеніший метод машинного навчання»	4
Лабораторне заняття №9. Тема: «Типи даних і конструювання ознак»	4
Лабораторне заняття №10. Тема: «Оцінка якості моделі».	2
Лабораторне заняття №11. Тема: «Реалізація крос-валідації».	2
Лабораторне заняття №12. Тема: «Об'єднання алгоритмів в ланцюги та конвеєри»	4
Лабораторне заняття №13. Тема: «Робота з текстовими даними»	3
Лабораторне заняття №14. Тема: «Робота з графічними даними».	3
Модульний контроль	20
Виконання індивідуального завдання (СР)	30
Разом: Аудиторна робота	70
Самостійна робота (СР)	30
Всього:	100

КОНТРОЛЬ ТА КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАТЬ СТУДЕНТІВ

При вивченні дисципліни використовуються наступні форми контролю знань студентів: поточний; модульний; підсумковий.

Поточний контроль передбачає перевірку теоретичних питань, самостійної роботи, практичних робіт та усне опитування по кожній практичній роботі. По даному виду контролю оцінювання знань здійснюється у відповідності до бального розподілу наведеного в попередній таблиці.

Модульний контроль передбачає виконання модульної контрольної роботи. Всі завдання оцінюються в 20 балів. Перше завдання (теоретичне) – 4 бали, друге завдання (практичне) – 8 балів, третє завдання (практичне) – 8 балів.

Формою підсумкового контролю є екзамен. Екзаменаційна оцінка (100 балів) є результатом виконання двох теоретичних питань (2 x 20 балів = 40 балів) та практичного завдання (60 балів).

Результуюча оцінка з дисципліни визначається як середня від балів набраних протягом семестру та отриманих на іспиті.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основний:

1. Chollet F. Deep Learning with Python / F. Chollet – Manning Publications Co., 2021 – 504 p.
2. Galea A. Beginning Data Science with Python and Jupyter / A. Galea. – Packt Publishing Ltd, 2018. – 194 p.
3. Muller A.C. Introduction to Machine Learning with Python / A.C. Muller, S. Guido. – Published by O'Reilly Media, Inc., 2017. – 376 p.
4. Falk K. Practical Recommender Systems / K. Falk .– Manning, 2019. – 432 p.