



МАШИННЕ НАВЧАННЯ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>F Інформаційні технології</i>
Спеціальність	<i>F1 Прикладна математика</i>
Освітня програма	<i>Машинне навчання та математичне моделювання</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>1 курс, весняний семестр</i>
Обсяг дисципліни	<i>120 год. / 4 кредити, лекції 30 год., комп'ютерний практикум 16 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік, МКР(=колоквіум), РГР (=міні-проект)</i>
Розклад занять	<i>Лекція – 2 год. на тиждень, комп'ютерний практикум – 1 год. на тиждень</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: проф., д.т.н. Чертов Олег Романович, oleg.r.chertov@gmail.com Комп'ютерний практикум: Мальчиков Володимир Вікторович, mavr2k@gmail.com</i>
Розміщення курсу	

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Метою навчальної дисципліни є формування у студентів сучасного наукового світогляду в області методів машинного навчання, ґрунтовне ознайомлення із наявними методами розв'язування різноманітних задач з області машинного навчання та роз'яснення вагомих аспектів їх застосування.

Предметом навчальної дисципліни є методи контрольованого та неконтрольованого навчання і методи навчання з підкріпленням.

Під час вивчення даної дисципліни студенти набудуть такі **компетентності**:

- здатність до самонавчання, пошуку, оброблення та інтелектуального аналізу інформації з різних джерел, вміння виявляти, ставити та вирішувати проблеми;
- здатність формалізувати та розв'язувати складні задачі й проблеми, які потребують оновлення й інтеграції знань, часто в умовах неповної, неточної чи недостатньої інформації та суперечливих вимог;
- здатність будувати, навчати та оцінювати якість моделей машинного навчання, зокрема, основних класифікаторів, при розв'язанні задач;
- здатність виконувати обчислення, пов'язані з навчанням та роботою моделей машинного навчання та інтелектуального аналізу даних за допомогою мови Python та спеціалізованих бібліотек, правильно інтерпретувати отримані результати навчання моделі.

Програмними результатами навчання є такі:

- використовувати та адаптувати математичні теорії та моделі для забезпечення теоретичного підґрунтя розв'язання наукових та практичних задач;
- застосовувати існуючий математичний апарат, розробляти нові моделі, методи та алгоритми при вирішенні актуальних практичних задач широкого спектру;
- застосовувати методи здобуття знань із даних, методи оцінки та інтерпретації знайдених закономірностей;
- обирати набір ознак (факторів) для класифікації чи регресії та проводити попередню обробку даних, підбирати вид моделі машинного навчання у залежності від задачі, що розв'язується;
- Обґрунтовувати вибір метрик класифікації та регресії, принципів побудови векторних ознак, вирішуючих правил та класифікаторів, розуміти основні типи задач машинного навчання й інтелектуального аналізу великих обсягів даних.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Необхідно опанування університетських курсів алгебри, математичного аналізу, дискретної математики, теорії ймовірностей, математичної статистики, методів оптимізації, аналізу даних, програмування на мові Python. Також бажаним є знання основ машинного та глибинного навчання.

3. Зміст навчальної дисципліни

Розділ 1. Контрольоване навчання (навчання під наглядом)

- Тема 1.1. Класифікаційні метрики
- Тема 1.2. Дерева прийняття (ухвалення) рішень. Метричні методи класифікації
- Тема 1.3. Лінійні методи класифікації та регресії. Баєсівські методи. Ядрові методи. Метод опорних векторів

Розділ 2. Неконтрольоване навчання (навчання без нагляду)

- Тема 2.1. Кластеризація
- Тема 2.2. Асоціативні правила

Розділ 3. Ансамблі моделей. Навчання з підкріпленням. Окремі питання машинного навчання

- Тема 3.1. Комбінації моделей (ансамбль моделей)
- Тема 3.2. Відбір та синтез інформативних характеристик (зниження розмірності)
- Тема 3.3. Навчання з підкріпленням
- Тема 3.4. Основи статистичної теорії навчання розпізнаванню за Вапником та Червоненкісом
- Тема 3.5. Причинно-наслідкове навчання (Causal Learning)
- Тема 3.6. Навчання з гарантованою точністю (Conformal Learning)

4. Навчальні матеріали та ресурси

Базова література

1. Кононова К. Ю. Машинне навчання: методи та моделі. – Харків: ХНУ імені В. Н. Каразіна, 2020. – 301 с. (<https://github.com/katerynakononova/ML/blob/master/ML.pdf>)

2. Інтелектуальний аналіз даних та машинне навчання. Ч. 1 : Базові методи та засоби аналізу даних / Я. В. Іванчук та ін. / Вінниц. нац. техн. ун-т. - Вінниця : ВНТУ, 2021. - 68 с. (http://pdf.lib.vntu.edu.ua/books/2022/Ivanchuk_P1_2021_69.pdf).

Допоміжна література

1. Burkov, A. (2019) The Hundred-Page Machine Learning Book (<https://themlbook.com/>).
2. James, G.M., Witten, D.M., Hastie, T.J., & Tibshirani, R. (2021). An Introduction to Statistical Learning. Springer Texts in Statistics (https://hastie.su.domains/ISLR2/ISLRv2_website.pdf).

Інформаційні ресурси

1. Електронний кампус КПІ ім. Ігоря Сікорського. Матеріали з дисципліни «Машинне навчання». – Режим доступу : <http://login.kpi.ua>
2. Дистанційний курс “Машинне навчання” - Електронний ресурс — Режим доступу https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

5.1. Лекції

Лекція 1 Вступ до машинного навчання та науки про дані

Визначення науки про дані (через діаграми Венна). Визначення машинного навчання (за Томом Мітчелом). Типи навчання та класи задач машинного навчання. Базова схема машинного навчання.

Лекції 2-3 Класифікаційні метрики

Класифікаційні метрики: матриця помилок (Confusion matrix) і помилки першого та другого роду, доля правильних відповідей (Accuracy), точність (Precision) і повнота (Recall), F-міри (F-Scores), коефіцієнт кореляції Метьюса (Matthews Correlation Coefficient, MCC), Precision-Recall крива (curve), ROC («reciever operating characteristic»)-крива, Area under curve (AUC), коефіцієнт Джині (Gini coefficient). Log_Loss. Бінарний, багатомітковий та багатокласовий випадки.

Лекція 4 Метричні методи класифікації

Відмінність задач регресії та класифікації. Метрики, що базуються на відстані та подібності (distance-based vs similarity-based). Жадібні / лінійні алгоритми. Гіпотези компактності та неперервності. Алгоритм k найближчих сусідів. Яке значення k краще? Підбір вагів. Діаграма Вороного. Чутливість до шумів. Нормалізація характеристик. Приклади метрик (метрика Мінковського, косинусна міра, відстань Жаккарда, відстань Хеммінга). Наближені методи пошуку найближчих сусідів (алгоритм STOLP, метод LSH — locality sensitive hashing, композиція геш-функцій для міри Жаккарда).

Лекція 5 Баєсівські методи

Безумовна (априорна) ймовірність та умовна (апостеріорна) ймовірність. Теорема Баєса. Класифікатори на основі теореми Баєса. Наївний баєсівський алгоритм. Згладжування Лапласа.

Лекція 6 Кластеризація

Кластеризація як приклад навчання без нагляду. Типи кластерних алгоритмів. Характеристики кластерів. Критерії зупинки кластеризації. Метод K-середніх. Як визначити кількість кластерів? Ієрархічна агрегативна кластеризація. Як обчислити

близькість кластерів? Близькість кластерів. Моделі, засновані на щільності (DBSCAN). Кластеризаційні метрики

Лекції 7-8 Вибір та синтез інформативних характеристик (зниження розмірності)

Вибір інформативних характеристик (feature selection): виявлення неінформативних ознак за допомогою фільтрів; методи обгортки; методи, вбудовані в модель навчання. Синтез інформативних характеристик (feature extraction).

Лекції 9-10 Комбінації моделей (ансамбль моделей)

Загальна схема ансамблю моделей. За рахунок чого є перевага ансамблю над окремим класифікатором? Класичні моделі – Bagging, випадковий ліс, Boosting (підсилення)). Градієнтні методи (алгоритм LightGBM тощо).

Лекція 11 Асоціативні правила

Модель ринкового кошику. Інші приклади (пов'язані поняття, антиплагіат, біомаркери). Властивість антимонотонності підтримки. Підтримка (support). Рівень довіри (confidence). Асоціативні правила. Алгоритм A-Priori. Алгоритм FP-Growth. Підвищення ефективності обробки популярних наборів (метод локальних предметних наборів; семплінг).

Лекції 12-13 Основи статистичної теорії навчання розпізнаванню за Вапником та Червоненкисом

Лекція 14 Навчання з підкріпленням

Агент і його дії. Стан, траєкторія, стратегія. Задача про багаторукого бандита (multi-armed bandit). Жадібна стратегія агента. Вибір дії з максимальною верхньою оцінкою цінності. Марковський процес. Рівняння Беллмана. Метод Q-learning.

Лекції 15-16 Колоквіуми

Лекції 17-18 Лекції запрошених фахівців (Causal Learning, Conformal Learning тощо).

5.2. Комп'ютерний практикум

Протягом семестру студенти повинні виконати цикл робіт комп'ютерного практикуму, які у сукупності утворюють окремий міні-проект, який здається як розрахункова робота (РР).

Всього повинно бути виконано три роботи комп'ютерного практикуму (=три етапи розробки міні-проекту):

- 1) підготовка та затвердження бізнес-ідеї міні-проекту: мета, цільова аудиторія, функціонал, конкуренти, SWOT-аналіз, пошук/збір наборів даних;
- 2) розробка та затвердження прототипу системи;
- 3) розробка та захист мінімально життєздатного продукту (Minimal Viable Product, MVP).

6. Самостійна робота студента

1 Ядрові методи. Метод опорних векторів

Лінійний бінарний класифікатор з геометричної точки зору. Максимізація відступу (margin). Підхід з опуклою оболонкою. Дуальні (двоїсті) задачі для лінійного та квадратичного програмування. Метод опорних векторів: постановка задачі, kernel trick, ядро як узагальнення скалярного добутку. Теорема Мерсера. Переваги і недоліки SVM.

5 год.

2 Древа прийняття (ухвалення) рішень

Використання дерев прийняття рішень для класифікації. Експоненційний розмір простору гіпотез. Класифікація неперервних даних за допомогою дерев. Жадібний алгоритм

	побудови дерева. Проблеми при наявності атрибутів з багатьма значеннями. Побудова по дереву правил прийняття рішень. Алгоритм ID3. Порівняння критеріїв розщеплення. Переваги та недоліки дерев прийняття рішень. Обрізання дерев (pre-pruning, post-pruning). Для чого воно застосовується?	5 год.
3	Лінійні методи класифікації та регресії. Логістична регресія Задачі регресії та класифікації як оптимізаційні задачі. Метод мінімізації емпіричного ризику. Поняття відступу (margin). Регуляризація (параметр регуляризації, ступінь регуляризатора). Логістична регресія. Коефіцієнт детермінації.	5 год.
4	Виконання робіт комп'ютерного практикуму	30 год.
5	Підготовка до колоквиуму та заліку	21 год.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

- Відвідування лекцій є дуже бажаним.
- На лекційних заняттях студенти повинні слухати уважно, не створювати шуму. Дозволяється (і рекомендується!) задавати «гарні» питання для з'ясування незрозумілих моментів та відповідати на запитання викладача. Мобільні телефони, планшети, смарт-годинники та інші гаджети мають бути вимкнені або встановлені на беззвучний режим, якщо вони не використовуються на занятті.
- Роботи комп'ютерного практикуму виконуються за допомогою мови Python, можливість використання сторонніх бібліотек визначається для кожної роботи окремо. Захист робіт комп'ютерного практикуму відбувається в усній формі, очно або в режимі конференції (при дистанційній формі навчання). На захисті студент має розповісти хід виконання роботи, сформулювати висновки, відповісти на питання викладача. Роботи потрібно здавати та захищати вчасно, за запізнення оцінка знижується.
- Колоквиум здається під час одного чи декількох із лекційних занять.
- Під час вивчення дисципліни студенти повинні дотримуватися правил академічної доброчесності, що передбачає зокрема неприпустимість плагіату, списування та інших способів видавання чужого доробку за свій. За недотримання цих правил передбачається покарання, що включає в себе вилучення порушника/порушників з відповідного контрольного заходу та виставлення за нього оцінки «0».

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Рейтинг студента з кредитного модуля складається з:

- 1) балів за виконання та захист циклу робіт комп'ютерного практикуму (міні-проєкту);
- 2) балів за відповіді на колоквиумі (як виконання модульної контрольної роботи, МКР);
- 3) додаткових балів за активну роботу лекціях чи оригінальне виконання робіт комп'ютерного практикуму.

СИСТЕМА РЕЙТИНГОВИХ БАЛІВ

1. Бали за виконання та захист циклу робіт комп'ютерного практикуму (міні-проєкту)

Міні-проєкт виконується тільки колективно – бригадою студентів (переважно із трьох осіб), але оцінюється вклад кожного студента окремо.

Протягом семестру студенти повинні виконати три етапи розробки міні-проєкту:

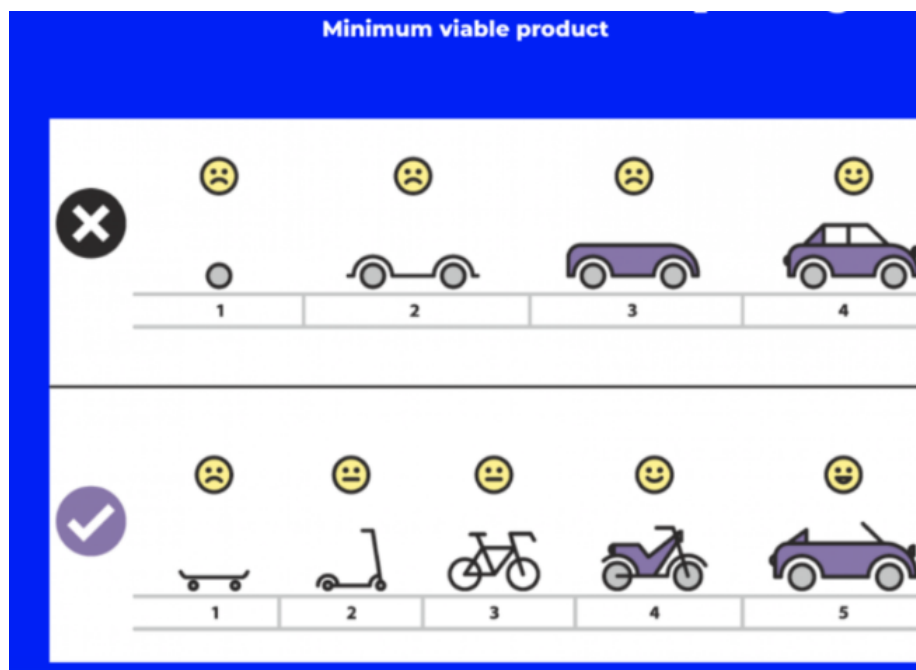
- 1) підготовка та затвердження бізнес-ідеї міні-проєкту: мета, цільова аудиторія, функціонал, конкуренти, SWOT-аналіз, пошук/збір наборів даних (максимум – 14 балів);

SWOT-аналіз



[Ілюстрація із Вікіпедії]

- 2) розробка та затвердження прототипу системи (максимум – 20 балів);
- 3) розробка та захист мінімально життєздатного продукту (Minimal Viable Product, MVP) (максимум – 30 балів).



[Ілюстрація із <https://habr.com/ru/companies/productstar/articles/508892/>]

За невчасно зданий без поважної причини етап розробки міні-проєкту оцінка знижується на 2 бали за кожен пропущений тиждень від кінцевого терміну здачі етапу:

етап		бізнес-ідея, 14 б.			прототип, 20 б.		MVP, 30 б.	
номер заняття	1	2	3	4	5	6	7	8

Максимальна кількість балів за розробку та захист міні-проєкту:

14 балів + 20 балів + 30 балів = **64 бали.**

2. Бали за відповіді на колоквіумі

Колоквіум проводиться як усна співбесіда, на якій кожен студент повинен відповісти на три основні запитання від викладача. Для уточнення відповіді студента чи перевірки – наскільки глибоко він розуміє матеріал, викладач може задавати запитання, додаткові до основного.

Відповідь на кожне запитання оцінюється максимум у 12 балів.

Критерії оцінювання:

- 12 балів — відповідь повна і правильна;
- 8-11 балів — відповідь правильна, але неповна або містить незначні неточності;
- 1-7 балів — відповідь неповна і неточна;
- 0 балів — відповідь неправильна або відсутня.

Максимальна кількість балів за колоквіум:

12 балів × 3 запитання = **36 балів.**

3. Заохочувальні бали за:

- виступ на лекції (за попередньою домовленістю з лектором) із доповіддю по матеріалам навчального курсу (або близьким до них) чи з доповіддю про особисто отримані наукові результати з машинного навчання або його застосувань, надається до 6-х заохочувальних балів;
- активність на лекціях, тобто за
 - відповіді на запитання лектора до загальної аудиторії,
 - знаходження описок/помилки у лекціях,
 - задавання «правильних запитань», тобто запитань, які свідчать про вдумливу роботу студента з навчальним матеріалом
 нараховується загалом до 4-х заохочувальних балів.

4. Визнання результатів навчання, отриманих у неформальній освіті, тощо

За проходження певного елементу неформальної освіти (наприклад, онлайн-курсів Coursera або аналогічних), за перемогу або участь в тематичних хакатонах або інших змаганнях, студенту можуть бути нараховані додаткові бали, зараховано виконання лабораторних робіт за всю дисципліну з максимально можливою оцінкою, або, взагалі, виставлено 100 балів.

Конкретний розмір заохочення студента визначається викладачем, що читає лекції, виходячи із повноти, важливості та результатів проходження цим студентом відповідних елементів неформальної освіти. Але в будь-якому випадку визнання результатів навчання, отриманих у неформальній освіті, може бути зараховано лише у випадку попереднього узгодження з викладачем.

5. Розрахунок шкали (R) рейтингу

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_c = 64 \text{ бали} + 36 \text{ балів} = 100 \text{ балів.}$$

Неодмінною умовою допуску до заліку є набрання студентом 30 балів за семестр. Інакше студент у першу відомість отримує оцінку «не допущено», а потім повинен виконати ці умови (набирати бали для допуску). Виконавши умови допуску до заліку, студент під час першого та/або другого перескладання пише залікову контрольну роботу.

Якщо наприкінці семестру студент набрав не менше 60 рейтингових балів, а також виконав умови допуску до заліку, він отримує залік «автоматом» відповідно до таблиці.

Студент може спробувати підвищити свою оцінку шляхом написання залікової контрольної роботи, при цьому його бали, отримані за семестр, анулюються.

У випадку, якщо сума рейтингових балів менша від 60, але виконано умови допуску до семестрового контролю, студент виконує залікову контрольну роботу, при цьому його рейтинг анулюють, після чого бали нараховують за результатами виконання залікової контрольної роботи.

Залікова контрольна робота проводиться як колоквиум з 5 письмовими та усними завданнями. Відповідь на кожне завдання оцінюють у 20 балів.

Критерії оцінювання відповідей на кожне завдання:

20 балів — правильна та змістовна відповідь;

15–19 балів — відповідь правильна, змістовна, але має незначні вади;

10–14 балів — відповідь містить незначні помилки чи є неповною;

1–9 балів — відповідь містить суттєві помилки та є неповною;

0 балів — немає відповіді.

Для отримання студентом відповідних оцінок його рейтингова оцінка R_D переводиться згідно з такою таблицею:

Рейтингові бали, R_D	Оцінка
95–100	відмінно
85–94	дуже добре
75–84	добре
65–74	задовільно
60–64	достатньо
Сума балів < 60	незадовільно
Рейтинг за семестр < 30	не допущено

Робочу програму навчальної дисципліни (силабус):

Складено професором кафедри ПМА, д.т.н., проф. ЧЕРТОВИМ О. Р.

Ухвалено кафедрою ПМА (протокол № 14 від 14.05.2025 р.).

Погоджено Методичною комісією факультету прикладної математики (протокол № 11 від 23.05.2025 р.).