



МАШИННЕ НАВЧАННЯ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>05 Соціальні та поведінкові науки</i>
Спеціальність	<i>054 Соціологія</i>
Освітня програма	<i>Аналітика соціальних даних</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>Перший рік, 2 семестр</i>
Обсяг дисципліни	<i>4 кредити ЄКТС/120 годин: лекції 18 год., практичні заняття 36 год., самостійна робота 66 год.</i>
Семестровий контроль/ контрольні заходи	<i>Залік/МКР</i>
Розклад занять	<i>rozklad.kpi.ua 1 год. лекційних занять та 2 год. комп'ютерних практикумів 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'ютерні практикуми: Мальчиков Володимир Вікторович, mavr2k@gmail.com</i>
Розміщення курсу	<i>Електронний ресурс «Кампус КПІ» https://ecampus.kpi.ua/</i>

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, ознайомлення із рядом моделей машинного навчання та роз'яснення аспектів їх застосування.

Предметом навчальної дисципліни є методи контрольованого та неконтрольованого навчання, перехресної перевірки та відбору важливих характеристик.

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі результати навчання:

- 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;

- вміти формулювати та розв'язувати базові задачі машинного навчання.

Згідно з освітньо-науковою програмою, засвоєння дисципліни сприяє підсиленню таких компетенцій та програмних результатів навчання:

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

- *ФК 03 Здатність проектувати і виконувати соціологічні дослідження, розробляти й обґрунтовувати їхню методологію.*
- *ФК 04 Здатність збирати та аналізувати емпіричні дані з використанням сучасних методів соціологічних досліджень.*
- *ФК 11 Здатність здійснювати аналіз відкритих джерел даних (OSINT), аналізувати якісну інформацію, текстові дані, використовувати інтелектуальний аналіз для соціальних даних.*
- *ПРН 01 Аналізувати соціальні явища і процеси, використовуючи емпіричні дані та сучасні концепції і теорії соціології.*
- *ПРН 04 Застосовувати наукові знання, соціологічні та статистичні методи, цифрові технології, спеціалізоване програмне забезпечення для розв'язування складних задач соціології та суміжних галузей знань.*
- *ПРН 05 Здійснювати пошук, аналізувати та оцінювати необхідну інформацію в науковій літературі, банках даних та інших джерелах.*
- *ПРН 09 Планувати і виконувати наукові дослідження у сфері соціології, аналізувати результати, обґрунтовувати висновки.*
- *ПРН 12 Здійснювати аналіз відкритих джерел даних (OSINT), аналізувати якісну інформацію, текстові дані, використовувати інтелектуальний аналіз для соціальних даних.*

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

Пререквізити: необхідно опанування університетських курсів вищої математики (алгебри, математичного аналізу), теорії ймовірностей, математичної статистики, програмування.

Постреквізити: вивчення дисципліни сприяє поглибленню аналітичної та дослідницької підготовки студентів, формує навички побудови, оцінювання та інтерпретації моделей для аналізу даних, що може бути використано під час опанування фахових дисциплін, виконання практичних завдань і підготовки магістерської дисертації.

3. Зміст навчальної дисципліни

Розділ 1. Контрольоване навчання

- Тема 1.1. Вступ*
- Тема 1.2. Класифікаційні метрики*
- Тема 1.3. Древа прийняття рішень*
- Тема 1.4. Метричні методи класифікації та регресії*

Розділ 2. Неконтрольоване навчання. Складні моделі

- Тема 2.1. Кластеризація*
- Тема 2.2. Інженерія інформативних характеристик*
- Тема 2.3. Комбінація моделей*

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/katernakononova/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).*

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

5.1. Лекції

Лекція 1 Вступ до машинного навчання

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 («receiver operating characteristic»)-крива, Area under curve (AUC). Бінарний та багатокласовий випадки.

Лекції 4 Древа прийняття 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:

- *Information Gain (приріст інформації)*
- *Gini impurity (неоднорідність Джині)*
- *(Information) Gain ratio (нормалізований приріст інформації)*
- *Classification error (помилка класифікації).*

Переваги та недоліки дерев прийняття рішень.

Обрізання дерев (pre-pruning, post-pruning). Для чого воно застосовується?

Лекції 5 Метричні методи класифікації

Відмінність задач регресії та класифікації. Метрики, що базуються на відстані та подібності (distance-based vs similarity-based). Жадібні / ледачі 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).

Лекції 6 Кластеризація

Кластеризація як приклад неконтрольованого навчання. Типи кластерних алгоритмів. Характеристики кластерів. Критерії зупинки кластеризації. Метод K-середніх. Як визначити кількість кластерів? Ієрархічна агломеративна кластеризація.

Лекція 7 Інженерія інформативних характеристик

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

Лекції 8 Комбінації моделей

Загальна схема ансамблю моделей. За рахунок чого є перевага ансамблю над окремим класифікатором? *Bagging*. Приклади агрегуючих функцій. Просте і зважене голосування. Випадковий ліс. *Boosting* (підсилення).

Лекція 9 Типовий підхід до розв'язання задач машинного навчання

Демонстрація типових підходів до збору даних та їх очищення, заповненню пропущених значень, проведення розвідувального аналізу даних (*Exploratory data analysis*), відбору інформативних характеристик, вибору моделей машинного навчання та оптимізації їх гіперпараметрів на прикладі *Jupyter Notebook* з прогнозування пасажирів, що врятувалися після того, як затонув «Титанік».

5.2. Практичні заняття (комп'ютерні практикуми)

1. Практичне заняття 1-3. Аналіз даних із використанням *Python*.
2. Практичне заняття 4-6. Застосування метрик класифікації.
3. Практичне заняття 7-9. Дерева прийняття рішень.
4. Практичне заняття 10-13. Метричні методи класифікації та регресії. Алгоритм к найближчих сусідів.
5. Практичне заняття 14-16. Кластеризація.
6. Практичне заняття 17. МКР.
7. Практичне заняття 18. Залік.

6. Самостійна робота студента

- Підготовка до захисту завдань на комп'ютерні практикуми 56 год.
- Підготовка до модульної контрольної роботи 4 год.
- Підготовка до заліку 6 год.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

- Відвідування лекцій та комп'ютерних практикумів 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, користуватися ними не дозволяється, якщо вони не застосовуються для організації навчання.
- Комп'ютерні практикуми виконуються за допомогою однієї з мов програмування (рекомендується *Python*), можливість використання сторонніх бібліотек визначається для кожного практикума окремо. Захист комп'ютерних практикумів 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.
- Модульна контрольна робота виконується під час одного із практичних 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». Якщо при перевірці

комп'ютерних практикумів або модульної контрольної роботи виявляються ознаки списування, оцінка «0» виставляється і тому, хто списав, і тому, хто дав списати.

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

Рейтинг студента з дисципліни складається з балів, що він отримує:

- 1) за виконання, здачу та захист комп'ютерних практикумів;
- 2) за виконання модульної контрольної роботи (МКР);

СИСТЕМА РЕЙТИНГОВИХ БАЛІВ

1. Виконання, здача та захист комп'ютерних практикумів

За семестр студент повинен підготувати, здати і захистити викладачу 5 комп'ютерних практикумів. При оцінюванні комп'ютерних практикумів, як і при розв'язанні реальних практичних задач для конкретного Замовника, враховується працездатність програми, своєчасність її надання та вміння пояснити роботу програми й інтерпретувати отримані результати.

Окремо перевіряється академічна доброчесність – студент повинен довести викладачу, що він розуміє, що саме і як саме виконується в програмі та може в ній пояснити особливості коду і міркування, що лягли в основу прийнятих рішень при програмній реалізації.

За виконання та здачу одного комп'ютерного практикуму студент може отримати бали за кожну із наступних трьох складових:

1) (своєчасність): якщо комп'ютерний практикум надіслано викладачу не пізніше обумовленого терміну і при цьому безпомилково проходить базове тестування (запускається без помилок та дає коректний результат хоча б для одного стандартного тестового прикладу), то за це нараховується 4 бали;

2) (працездатність):

- якщо програма працює безпомилково, то за це нараховується 3 бали;
- якщо програма має несуттєві недоліки, інколи дає неправильний результат, то за це нараховується від 1 до 2 балів;
- якщо програма у більшості випадків працює некоректно, то за це нараховується 0 балів;

3) (презентабельність):

- якщо студент правильно чи майже правильно відповідає на питання стосовно функціональності програми та особливостей її реалізації, то за це нараховується від 6 до 7 балів;
- якщо студент частково плутається при відповіді на питання стосовно функціональності програми та особливостей її реалізації, то за це нараховується від 3 до 5 балів;
- якщо студент частково плутається при відповіді на питання стосовно функціональності програми та не може обґрунтувати вибір інструментальних засобів для її реалізації, то за це нараховується від 1 до 2 балів;
- якщо студент не може пояснити основну функціональність програми та особливості її реалізації, тобто очевидно має місце порушення академічної

добросовісності, то всі бали за відповідний комп'ютерний практикум обнуляються і, загалом, за нього студент отримує 0 балів;

- щоб зрозуміти, наскільки студент розібрався з теоретичним матеріалом, який відповідає комп'ютерному практикуму, викладач може задати студенту декілька питань по відповідній теорії, і в залежності від правильності відповідей на них збільшує чи зменшує оцінку за «презентабельність».

Таким чином, за один комп'ютерний практикум один студент максимально може отримати:

4 бали + 3 бали + 7 балів = **14 балів**.

Загалом, за цикл комп'ютерного практикуму студент максимально може отримати:

14 балів × 5 = **70 балів (R_{Lab})**.

Викладач має право заохотити студента певною кількістю балів (максимально – 5 балів за семестр) за надання цікавої і оригінальної алгоритмічної чи програмної реалізації завдання комп'ютерного практикуму.

2. Модульний контроль

Ваговий бал – **30 балів (R_{сw})** за одногодинну МКР.

Контрольна робота включає в себе завдання двох типів:

- 1) виконати певні розрахунки (побудувати PR- чи ROC-криву тощо);
- 2) вибрати правильну чи правильні відповіді із наданих і обґрунтувати свій вибір.

Кожне завдання контрольної роботи оцінюється (у залежності від його складності) в два, три чи п'ять балів.

Максимальну кількість балів студент отримує у випадку, якщо він навів повний та правильний розв'язок задачі або припустився незначних помилок/описок, які не вплинули суттєво на розв'язок.

Меншу ніж максимальну кількість балів (як правило, половину від цієї максимальної кількості) студент отримує у випадку, коли наведений ним розв'язок є правильним, проте неповним (не всі правильні відповіді були зазначені), або хід розв'язку є правильним, проте студент припустився помилок, які суттєво вплинули на відповідь.

Нуль балів студент отримує у випадку, коли задача взагалі не розв'язана, або наведений хід розв'язку містить грубі помилки, або на задачу наведена тільки відповідь, без її обґрунтування.

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 5-ти 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 матеріаломнараховується загалом до 5-ти 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 комп'ютерних практикумів 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. Визнання результатів неформальної освіти відбувається у порядку визначеному у відповідному Положенні КПІ ім. Ігоря Сікорського: <https://osvita.kpi.ua/node/179>

Розрахунок шкали (R) рейтингу

Сума вагових балів контрольних заходів протягом семестру складає:

$$RC = R_{Lab} + R_{CW} = 70 \text{ балів} + 30 \text{ балів} = 100 \text{ балів.}$$

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 нижче наведеної 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 колоквіум 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;

18–19 балів — відповідь правильна, змістовна, але має незначні вади;

14–17 балів — відповідь містить незначні помилки чи є неповною;

12–13 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 завідувачем кафедри прикладна математика, д.т.н., проф. ЧЕРТОВИМ О. Р.

Ухвалено кафедрою ПМА (протокол № 18 від 10.06.2024).

Погоджено Методичною комісією факультету прикладної математики (протокол № 12 від 21.06.2024).