



Кам'янець-Подільський національний університет
імені Івана Огієнка
Фізико-математичний факультет
Кафедра комп'ютерних наук

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Прикладне машинне навчання

1. Загальна інформація про курс. Викладачі

Іванюк Віталій Анатолійович, доктор технічних наук, доцент, завідувач кафедри комп'ютерних наук

E-mail: wivanyuk@kpmu.edu.ua

Профайл: <https://cs.kpmu.edu.ua/2019/11/04/ivaniuk-vitalij-anatolijovych/>

Консультації: щочетверга – 16.00-18.00 (очно та онлайн)

Пилипук Тетяна Михайлівна, доцент, доцент кафедри комп'ютерних наук

E-mail: pylypyuk.tetiana@kpmu.edu.ua

Профайл: <https://cs.kpmu.edu.ua/2019/10/30/pylypiuk-tetiana-mykhajlivna/#more-649>

Мова викладання: українська

2. Анотація до курсу

Курс зосереджений на практичному застосуванні ключових методів та алгоритмів машинного навчання для вирішення реальних задач прогнозування та класифікації. Курс є логічним продовженням дисципліни «Інженерія даних» та використовує підготовлені дані для побудови та оцінки моделей.

Студенти детально вивчать реалізацію основних родин алгоритмів (напр., лінійні моделі, дерева рішень, ансамблі) за допомогою бібліотеки Scikit-learn. Критична увага приділяється методології коректної оцінки якості моделей, вибору адекватних метрик, технікам валідації (напр., крос-валідація) та інтерпретації результатів для уникнення поширених помилок, таких як перенавчання.

3. Мета і завдання курсу

Метою курсу є формування у студентів глибокого розуміння фундаментальних методів машинного навчання та надання практичних навичок для їх застосування з використанням Python. Курс фокусується на вирішенні задач класифікації та регресії, припускаючи, що дані вже пройшли етап первинної підготовки (який є предметом курсу "Інженерія даних з Python").

Завдання курсу:

- Систематизувати знання про основні алгоритми машинного навчання (напр., лінійні моделі, дерева рішень, ансамблеві методи), їхні математичні основи, переваги, недоліки та сфери застосування.
- Навчити практично будувати, навчати та налаштовувати (hyperparameter tuning) моделі машинного навчання, використовуючи стандартні бібліотеки (зокрема, Scikit-learn).
- Сформувані навички коректної валідації моделей (cross-validation) та вибору адекватних метрик якості (accuracy, precision, recall, F1, RMSE) для об'єктивної оцінки їх ефективності та уникнення перенавчання.
- Ознайомити з використанням візуалізації даних (Matplotlib/Seaborn) як інструменту для дослідження результатів моделювання (напр., аналіз помилок, матриці плутанини) та інтерпретації моделей.
- Розвинути вміння інтерпретувати результати роботи моделей та формулювати висновки для прийняття рішень на основі даних.

4. Результати навчання

Після успішного завершення курсу студент зможе:

1. Формалізувати прикладні задачі як задачі машинного навчання (класифікація/регресія) та обґрунтовано обирати відповідні алгоритми (Scikit-learn) для їх вирішення.

2. Практично реалізовувати повний цикл побудови моделі: навчати алгоритми на тренувальних даних, налаштовувати їх гіперпараметри та прогнозувати результати на нових даних.
3. Здійснювати коректну валідацію моделей, застосовуючи відповідні метрики якості (напр., Accuracy, Precision, Recall, F1, RMSE) та методи (напр., крос-валідація) для об'єктивної оцінки їх ефективності та уникнення перенавчання.
4. Використовувати інструменти візуалізації даних для аналізу результатів моделювання (напр., матриці плутанини, криві навчання) та інтерпретації поведінки моделі.

5. Формат курсу

Онний курс з елементами дистанційного навчання в системі Moodle.

6. Обсяг і ознаки курсу

Найменування показників	Характеристика навчальної дисципліни
Кількість кредитів ЄКТС	3/4
Загальний обсяг годин	90/120
Кількість годин навчальних занять	30/40
Лекційні заняття	12
Практичні заняття	0
Семінарські заняття	0
Лабораторні заняття	18/28
Самостійна та індивідуальна робота	60/80
Форма підсумкового контролю	Залік

7. Пререквізити і кореквізити курсу

Перелік дисциплін, які бажано, але не обов'язково, вивчити раніше: Програмування, Вища математика, Теорія ймовірностей та математична статистика, Інтелектуальний аналіз даних, Методи та системи штучного інтелекту, Основи програмування мовою Python, Інженерія даних з Python.

8. Технічне й програмне забезпечення, обладнання

Обладнання: ПК (Windows/macOS/Linux) з доступом до Інтернету.

Основне ПЗ: Python 3.9+, Git.

Середовище: VS Code (або JupyterLab).

Бібліотеки (Інженерія даних): Pandas, SQLAlchemy (та драйвери БД, напр., psycopg2)

Бібліотеки (Прикладне МН): Scikit-learn, Matplotlib, Seaborn.

Для проведення лекцій використовується мультимедійне забезпечення. Проведення лабораторних робіт здійснюється в лабораторії із необхідним апаратним та програмним забезпеченням.

Програмне забезпечення: Python.

9. Політика курсу

Увесь навчальний контент розміщено в модульному середовищі навчання К-ПНУ імені Івана Огієнка – moodle. Підготовка та виконання завдань лабораторних робіт є обов'язковими для кожного студента.

Академічна доброчесність. Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел списування, втручання в роботу інших студентів становлять приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.

Відвідання занять. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати всі заняття. Виконання усіх лабораторних робіт є обов'язковим. Студенти зобов'язані дотримуватися термінів виконання усіх видів робіт, передбачених курсом.

Форми поточного та підсумкового контролю. Поточний контроль реалізується на лабораторних заняттях. Підсумковий контроль не здійснюється, а виставляється у формі заліку на основі поточного контролю.

10. Програма курсу

Модуль 1. Фундаментальні концепції ML

- **Тема 1. Вступ до Машинного навчання.**
 - **Зміст:** Типи задач (Supervised, Unsupervised). Огляд стеку (Scikit-learn, Matplotlib).
- **Тема 2. Життєвий цикл ML-проєкту.**
 - **Зміст:** Від бізнес-задачі до впровадження.
- **Тема 3. Ключова концепція: Валідація.**
 - **Зміст:** Поділ даних: Train-Test Split. Проблема **перенавчання** (Overfitting) та **недонавчання** (Underfitting).

Модуль 2. Задачі Регресії

- **Тема 4. Лінійна регресія.**
 - **Зміст:** Математичні основи (просто). Реалізація в Scikit-learn.
- **Тема 5. Метрики якості регресії.**
 - **Зміст:** MAE, MSE, RMSE, R-squared. Інтерпретація метрик.
- **Тема 6. Візуалізація результатів регресії.**
 - **Зміст:** Аналіз залишків. Графіки "прогнозовано vs. фактично".

Модуль 3. Задачі Класифікації

- **Тема 7. Логістична регресія.**
 - **Зміст:** Принцип роботи. `model.fit()`, `model.predict()`.
- **Тема 8. Дерева рішень та Ансамблі.**
 - **Зміст:** Decision Trees. Вступ до Random Forest (як приклад ансамблю).
- **Тема 9. Метрики якості класифікації.**
 - **Зміст:** Матриця плутанини (Confusion Matrix). Accuracy та її недоліки. Precision, Recall, F1-Score. ROC-AUC (оглядово).

Модуль 4. Покращення та валідація моделей

- **Тема 10. Крос-Валідація (Cross-Validation).**
 - **Зміст:** Чому Train-Test Split недостатньо? K-Fold Cross-Validation для надійної оцінки моделі.
- **Тема 11. Налаштування гіперпараметрів.**
 - **Зміст:** Що таке гіперпараметри? GridSearchCV для автоматичного пошуку найкращих параметрів.
- **Тема 12. Фінальний проєкт (ML Pipeline).**
 - **Зміст:** На основі даних з курсу "Інженерія даних", студенти будують, валідують та порівнюють 2-3 моделі (напр., Логістична регресія vs. Random Forest) для вирішення конкретної задачі. Обґрунтування вибору фінальної моделі на основі метрик.

11. Система оцінювання та вимоги

Курс складається з 1-го навчального (змістовного) модуля.

Поточний контроль полягає в перевірці теоретичних знань та практичних умінь і навичок під час лабораторних занять.

Результати перевірки самостійної роботи, в тому числі виконання домашніх завдань, студентів також входять до поточного контролю. Максимальний бал оцінки поточної успішності студентів на практичних заняттях приймається рівним 12.

Підсумковий контроль проводиться у формі заліку.

Рейтингова оцінка студентам виставляється відповідно до Положення про рейтингову систему оцінювання навчальних досягнень здобувачів вищої освіти Кам'янець-Подільського національного університету імені Івана Огієнка після проведення всіх навчальних занять та контрольних заходів з навчальної дисципліни.

Студенти, які мають академічну заборгованість за результатами підсумкового контролю, зобов'язані ліквідувати її в терміни, визначені графіком ліквідації академічної заборгованості.

12. Рекомендована література

Основна література та інформаційні ресурси:

1. Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd Edition). O'Reilly, 2022. (Український переклад: "Навчіться машинному навчанню з Scikit-Learn, Keras і TensorFlow")
2. Gareth James, et al. *An Introduction to Statistical Learning* (2nd Edition, with Python examples). Springer, 2021.
3. Andreas C. Müller & Sarah Guido. *Introduction to Machine Learning with Python*. O'Reilly, 2016.
4. Офіційна документація Scikit-learn (scikit-learn.org/stable/user_guide.html)
5. Kaggle (kaggle.com)