Міністерство освіти і науки України
Центральноукраїнський національний технічний університет
Механіко-технологічний факультет
Кафедра кібербезпеки та програмного забезпечення

Machine Learning

Методичні вказівки до виконання лабораторних робіт для студентів денної та заочної форми навчання другого рівня вищої освіти, галузі Інформаційні технології.

ЗАТВЕРДЖЕНО
на засіданні кафедри кібербезпеки та програмного забезпечення, протокол № 1 від 26.08.2025 року

Кропивницький
2025

УДК 004.5

Machine Learning: Методичні вказівки до виконання лабораторних робіт для студентів денної та заочної форми навчання другого рівня вищої освіти, галузі Інформаційні технології. / М-во освіти і науки України, Центральноукр. нац. техн. ун-т; [уклад. Є.В. Мелешко, О.М. Дрєєв, Р.О. Ткачук] – Кропивницький: ЦНТУ, 2025. – 35с.

Укладачі: Мелешко Є.В. доцент кафедри кібербезпеки та програмного забезпечення, доктор техн. наук, доцент;
Дрєєв О.М. доцент кафедри кібербезпеки та програмного забезпечення, кандидат технічних наук;
Ткачук Р.О. Senior DevOps Engineer EPAM Systems /
асистент кафедри кібербезпеки та програмного забезпечення

Рецензенти: Смірнов О. А., докт. техн. наук, професор, завідувач кафедри;
Якименко Н.М., к. ф.-м. наук, доцент.

© Центральноукраїнський національний технічний університет, 2025

ЗМІСТ

ВСТУП……..…..….4
Лабораторна робота №1……..….8
Лабораторна робота №2. …….…….……11
Лабораторна робота №3. ……...13
Лабораторна робота №4.……..……..…16
Лабораторна робота №5.....18
Лабораторна робота №6.....20
Лабораторна робота №7.....22
Лабораторна робота №8.....24
Список використаної літератури……..……...30

ВСТУП

Курс «Machine Learning» призначений для набуття теоретичних та практичних знань з використання, розробки та реалізації алгоритмів і моделей машинного навчання в складі інтелектуальних систем для задач аналізу даних, класифікації, прогнозування та підтримки прийняття рішень.

Метою викладання дисципліни «Machine Learning » є формування у здобувачів знань і навичок проєктування та програмної реалізації рішень машинного навчання для інтелектуальних систем у складі комп'ютерних систем і мереж на різних програмних платформах. Дисципліна забезпечує опанування підходів до побудови архітектури програмного забезпечення ML компонентів, уміння аналізувати та формулювати задачі, обирати ефективні методи їх розв'язання, критично оцінювати результати й аргументувати інженерні рішення на основі сучасних наукових здобутків. Передбачено виконання проєктів з урахуванням міждисциплінарних аспектів, пошук і оцінювання інформації з різних джерел, а також опрацювання апаратно-програмної реалізації та обчислювального прискорення з урахуванням ресурсних і енергетичних обмежень. Результати навчання включають здатність представляти розробки та дослідження у звітах, статтях і презентаціях.

Метою вивчення навчальної дисципліни є забезпечення поглибленої теоретичної та практичної підготовки здобувачів у сфері комп'ютерної інженерії з орієнтацією на сучасні наукові підходи, інженерні методи та міждисциплінарну інтеграцію знань. Дисципліна спрямована на розвиток системного мислення, здатності до аналітичної діяльності та формування наукового підходу до розв'язання складних технічних і прикладних завдань.

У межах навчального курсу розглядаються принципи розроблення алгоритмічного та програмного забезпечення, побудови архітектури комп'ютерних систем і мереж, створення системного та прикладного

програмного забезпечення, а також підходи до проєктування інтелектуальних апаратно-програмних рішень. Значна увага приділяється використанню сучасних методів програмування, інструментів автоматизації проєктування та дослідницьких методик у галузі інформаційних технологій.

Опрацювання дисципліни передбачає формування здатності здійснювати науковий пошук, аналізувати інформаційні ресурси, критично оцінювати отримані результати та приймати обґрунтовані інженерні рішення. Розглядаються питання реалізації проєктів у сфері комп'ютерної інженерії з урахуванням технічних, соціальних, економічних і правових чинників, а також особливостей інтеграції апаратних і програмних компонентів інтелектуальних систем.

У результаті вивчення дисципліни здобувачі набувають систематизованих знань щодо застосування математичних, природничих та інженерних методів у комп'ютерній інженерії, опановують підходи до розроблення та реалізації складних апаратно-програмних рішень, зокрема з урахуванням обчислювальних, енергетичних і ресурсних обмежень сучасних комп'ютерних архітектур.

Провідною формою навчання є лекція, яка забезпечує систематизований виклад теоретичних основ машинного навчання, логіки побудови моделей, принципів оцінювання їх якості та інтерпретації результатів. У процесі викладання застосовуються аналітичний, індуктивний та дедуктивний методи пізнання, що дозволяють поєднати математичний апарат із практичними прикладами використання алгоритмів.

Супровідною формою навчання є лабораторні роботи, які спрямовані на закріплення теоретичних знань та формування практичних навичок роботи з реальними наборами даних. Основна дидактична мета лабораторних занять – формування компетентностей у сфері проєктування ML-експериментів,

підготовки даних, навчання моделей, оцінювання їх ефективності та документування результатів.

Лабораторні роботи виконуються у такій послідовності:

- вивчення теоретичного матеріалу за темою роботи;

- підготовка до виконання завдання (аналіз вхідних даних, ознайомлення з інструментами та методами);

- реалізація алгоритмів у програмному середовищі;

- аналіз отриманих результатів та формування висновків;

- оформлення та захист звіту.

По завершенні кожної лабораторної роботи студент готує звіт, який повинен містити:

- тему та мету роботи;

- постановку задачі та опис використаних методів;

- фрагменти програмного коду;

- результати експериментів (таблиці, графіки, метрики якості моделей);

- аналіз отриманих результатів і висновки.

Перелік тем лабораторних робіт

- Лабораторна робота №1. Постановка та дизайн ML-експерименту.
- Лабораторна робота №2. Вивчення та візуалізація даних (EDA).
- Лабораторна робота №3. Розбиття даних, нормалізація та конструювання ознак.
- Лабораторна робота №4. Регресія та класифікація на класичних моделях.
- Лабораторна робота №5. Настройка гіперпараметрів і крос-валідація.

- Лабораторна робота №6. Підбір ознак та регуляризація.
- Лабораторна робота №7. Кластеризація та зниження розмірності.
- Лабораторна робота №8. Прогнозування часових рядів.

Оцінювання результатів виконання лабораторних робіт здійснюється з урахуванням рівня теоретичної підготовки, правильності реалізації алгоритмів, обґрунтованості вибору методів, якості аналізу результатів та оформлення звіту. Оцінки за лабораторні роботи враховуються при виставленні підсумкової оцінки з дисципліни.

Таким чином, дисципліна «Machine Learning» поєднує фундаментальні теоретичні засади з практичною реалізацією алгоритмів аналізу даних, забезпечуючи підготовку фахівців, здатних ефективно застосовувати сучасні методи машинного навчання у професійній діяльності.

ЛАБОРАТОРНА РОБОТА №1

ТЕМА: ПОСТАНОВКА ТА ДИЗАЙН ML-ЕКСПЕРИМЕНТУ

МЕТА: Сформувати навички формулювання ML-задачі, постановки гіпотез, вибору метрик та планування експерименту.

ЗНАТИ: Типи ML-задач, принципи формулювання гіпотез, критерії успіху, базові підходи до валідації.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв'язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Getting Started https://scikit-learn.org/stable/getting_started.html
- Data Splitting and Validation https://scikit-learn.org/stable/modules/cross_validation.html

Додаткова документація:

– https://pandas.pydata.org/docs/

– https://numpy.org/doc/

– https://developers.google.com/machine-learning/glossary

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python (локально або на ресурсах кафедри). Під час онлайн-навчання формат взаємодії погоджувати з викладачем.

ЗАВДАННЯ

Необхідно спланувати ML-експеримент для обраного датасету та задокументувати постановку задачі.

Підготовка до виконання

1. Обрати датасет (відкритий або запропонований викладачем) та описати його предметну область.
2. Визначити тип задачі (класифікація, регресія, кластеризація тощо).

Формулювання експерименту

3. Сформулювати гіпотези щодо можливих факторів, що впливають на цільову змінну.
4. Обрати основні та допоміжні метрики якості, описати критерій успіху.
5. Запропонувати план розбиття даних (train/val/test) з обґрунтуванням.

Планування та ризики

6. Описати очікувані ризики (шум, пропуски, дисбаланс) та способи їх мінімізації.
7. Скласти короткий план експериментів (мінімум 3 сценарії).

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Опис датасету, постановку задачі, гіпотези, метрики, план експерименту та ризики.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Які основні типи задач машинного навчання?
2. Як сформулювати гіпотезу для ML-експерименту?
3. Які метрики доречно застосовувати для регресії та класифікації?
4. Навіщо потрібен план експерименту та критерій успіху?
5. Чим відрізняються train/val/test набори?
6. Які ризики найчастіше зустрічаються в реальних даних?

7. Як обґрунтувати вибір метрик для конкретної задачі?
8. Що таке baseline і для чого він потрібен?

ЛАБОРАТОРНА РОБОТА №2

ТЕМА: ВИВЧЕННЯ ТА ВІЗУАЛІЗАЦІЯ ДАНИХ (EDA)

МЕТА: Розвинути навички дослідження структури даних, виявлення закономірностей і аномалій, побудови базових візуалізацій.

ЗНАТИ: Базові статистики розподілу, методи пошуку пропусків і викидів, правила інтерпретації графіків.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв’язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Pandas: User Guide https://pandas.pydata.org/docs/user_guide/index.html
- Seaborn: Tutorials https://seaborn.pydata.org/tutorial.html

Додаткова документація:

– https://matplotlib.org/stable/users/index.html

– https://scikit-learn.org/stable/auto_examples/index.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно виконати EDA для обраного датасету та підготувати короткі висновки.

Підготовка до виконання

1. Завантажити датасет і описати його структуру (кількість рядків/стовпців, типи даних).

Базова статистика та якість даних

2. Розрахувати основні статистики числових ознак.
3. Виявити пропуски та оцінити їх частку.
4. Перевірити наявність викидів та запропонувати способи їх обробки.

Візуалізація та інтерпретація

5. Побудувати розподіли для ключових ознак (гістограми, boxplot).
6. Проаналізувати кореляції та побудувати heatmap.
7. Сформулювати не менше 3 висновків щодо структури даних.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Таблиці описової статистики, графіки, висновки та короткі рекомендації щодо підготовки даних.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Що включає поняття EDA?
2. Які базові статистики описують розподіл ознаки?
3. Як виявляти пропуски та оцінювати їх вплив?
4. Чим відрізняються boxplot і гістограма?
5. Як інтерпретувати кореляційну матрицю?
6. Навіщо перевіряти наявність викидів?
7. Які ризики неправильного EDA?
8. Що таке data leakage і як його уникати на етапі аналізу даних?

ЛАБОРАТОРНА РОБОТА №3

ТЕМА: РОЗБИТТЯ ДАНИХ, НОРМАЛІЗАЦІЯ ТА КОНСТРУЮВАННЯ ОЗНАК

МЕТА: Навчитися коректно розбивати дані, застосовувати масштабування та конструювати інформативні ознаки.

ЗНАТИ: Методи масштабування, основи feature engineering, правила уникнення data leakage.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв’язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Preprocessing https://scikit-learn.org/stable/modules/preprocessing.html
- Scikit-learn: Feature Engineering https://scikit-learn.org/stable/modules/feature_extraction.html

Додаткова документація:

– https://pandas.pydata.org/docs/

–https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно підготувати датасет до моделювання з коректним розбиттям та трансформаціями.

Підготовка до виконання

1. Визначити цільову змінну та набір ознак.
2. Виконати розбиття на train/val/test із фіксацією random seed.

Масштабування та кодування

3. Застосувати масштабування (Standard або MinMax) до числових ознак.
4. Закодувати категоріальні ознаки (One-Hot або Ordinal) з обґрунтуванням.

Конструювання ознак

5. Побудувати щонайменше 2 нові ознаки (поліноміальні або взаємодії).
6. Оцінити вплив нових ознак на якість базової моделі.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Опис розбиття, застосовані трансформації, приклади нових ознак та короткий аналіз їхнього впливу.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Навіщо потрібне розбиття на train/val/test?
2. У чому різниця між StandardScaler і MinMaxScaler?
3. Які ризики data leakage при масштабуванні?
4. Коли доречно використовувати One-Hot кодування?
5. Що таке feature engineering та навіщо він потрібен?
6. Як перевірити корисність нових ознак?

7. Які підходи існують для роботи з пропусками?
8. Чому важливо фіксувати random seed?

ЛАБОРАТОРНА РОБОТА №4

ТЕМА: РЕГРЕСІЯ ТА КЛАСИФІКАЦІЯ НА КЛАСИЧНИХ МОДЕЛЯХ

МЕТА: Реалізувати та порівняти декілька класичних моделей для задач регресії або класифікації.

ЗНАТИ: Базові алгоритми регресії та класифікації, принципи оцінки якості моделей.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв'язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Supervised Learning
- https://scikit-learn.org/stable/supervised_learning.html
- Scikit-learn: Model Evaluation https://scikit-learn.org/stable/modules/model_evaluation.html

Додаткова документація:

– https://scikit-learn.org/stable/modules/linear_model.html

– https://scikit-learn.org/stable/modules/tree.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно навчити та порівняти 3-4 класичні моделі на обраному датасеті.

Підготовка до виконання

1. Визначити тип задачі (регресія або класифікація) та метрики оцінки.
2. Підготувати датасет з попередньої лабораторної роботи.

Навчання моделей

3. Обрати щонайменше 3 моделі (наприклад, лінійна/логістична регресія, kNN, дерево, ансамбль).
4. Навчити моделі та отримати результати на validation або test наборі.

Порівняння та інтерпретація

5. Порівняти якість моделей за обраними метриками.
6. Побудувати матрицю помилок (для класифікації) або графік залишків (для регресії).
7. Зробити висновки щодо сильних та слабких сторін кожної моделі.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Опис моделей, таблицю метрик, візуалізації та висновки.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Які базові моделі використовують для регресії та класифікації?
2. Чим відрізняється logistic regression від linear regression?
3. Які метрики є ключовими для класифікації?
4. Навіщо потрібна матриця помилок?
5. Як інтерпретувати bias і variance на практиці?
6. Коли варто використовувати ансамблі?
7. Як визначити baseline модель?
8. Які типові джерела помилок у класичних моделях?

ЛАБОРАТОРНА РОБОТА №5

ТЕМА: НАСТРОЙКА ГІПЕРПАРАМЕТРІВ І КРОС-ВАЛІДАЦІЯ

МЕТА: Навчитися систематично підбирати гіперпараметри та оцінювати стабільність моделі.

ЗНАТИ: Grid Search, Random Search, принципи крос-валідації, криві навчання.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв'язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Hyperparameter Search https://scikit-learn.org/stable/modules/grid_search.html
- Scikit-learn: Cross-Validation https://scikit-learn.org/stable/modules/cross_validation.html

Додаткова документація:

– https://scikit-learn.org/stable/modules/learning_curve.html

– https://scikit-learn.org/stable/modules/model_evaluation.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно підібрати гіперпараметри для однієї моделі та оцінити результат через крос-валідацію.

Підготовка до виконання

1. Обрати модель з попередньої лабораторної роботи для оптимізації.

2. Визначити простір гіперпараметрів для пошуку.

Підбір параметрів

3. Виконати Grid Search або Random Search із крос-валідацією.
4. Зафіксувати найкращі параметри та результати.

Оцінка стабільності

5. Побудувати криві навчання або валідації.
6. Порівняти якість до і після тюнінгу.
7. Зробити висновки щодо стабільності моделі.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Таблицю параметрів, графіки кривих, результати крос-валідації та висновки.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Що таке гіперпараметри і чим вони відрізняються від параметрів моделі?
2. Коли варто використовувати Grid Search, а коли Random Search?
3. Навіщо потрібна крос-валідація?
4. Які типові помилки при підборі гіперпараметрів?
5. Що показують криві навчання?
6. Як інтерпретувати розкид результатів між фолдами?
7. Що таке переобучення в контексті тюнінгу?
8. Як обрати метрику для пошуку параметрів?

ЛАБОРАТОРНА РОБОТА №6

ТЕМА: ПІДБІР ОЗНАК ТА РЕГУЛЯРИЗАЦІЯ

МЕТА: Навчитися зменшувати перенавчання через відбір ознак і регуляризацію.

ЗНАТИ: L1/L2 регуляризація, фільтраційні та обгорткові методи відбору ознак.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв'язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Feature Selection https://scikit-learn.org/stable/modules/feature_selection.html
- Scikit-learn: Regularization https://scikit-learn.org/stable/modules/linear_model.html

Додаткова документація:

– https://scikit-learn.org/stable/modules/permutation_importance.html

– https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно застосувати відбір ознак та регуляризацію для покращення узагальнення моделі.

Підготовка до виконання

1. Вибрати модель, чутливу до перенавчання (лінійна, логістична або дерево).

2. Підготувати набір ознак з попередніх робіт.

Відбір ознак

3. Застосувати фільтраційний метод (кореляція, chi2 або mutual information).
4. Використати обгортковий метод (RFE або permutation importance).

Регуляризація

5. Порівняти моделі з L1 і L2 регуляризацією.
6. Оцінити зміни в якості та інтерпретувати важливість ознак.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Порівняльні таблиці якості, обрані ознаки, пояснення ефекту регуляризації.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Чому перенавчання пов’язане з кількістю ознак?
2. У чому різниця між L1 і L2 регуляризацією?
3. Які методи відбору ознак називають фільтраційними?
4. Коли застосовують RFE або permutation importance?
5. Як регуляризація впливає на інтерпретованість моделі?
6. Чому важливо оцінювати модель на відкладених даних?
7. Як визначити оптимальну кількість ознак?
8. Які ризики неправильного відбору ознак?

ЛАБОРАТОРНА РОБОТА №7

ТЕМА: КЛАСТЕРИЗАЦІЯ ТА ЗНИЖЕННЯ РОЗМІРНОСТІ

МЕТА: Освоїти базові методи кластеризації та зниження розмірності, навчитися інтерпретувати результати.

ЗНАТИ: k-means, DBSCAN, PCA, методи оцінки якості кластерів.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв'язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Scikit-learn: Clustering https://scikit-learn.org/stable/modules/clustering.html
- Scikit-learn: PCA https://scikit-learn.org/stable/modules/decomposition.html#pca

Додаткова документація:

–https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

– https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно виконати кластеризацію та зниження розмірності для обраного датасету.

Підготовка до виконання

1. Підготувати датасет (масштабування, обробка пропусків).

Кластеризація

2. Виконати кластеризацію методом k-means і визначити оптимальну кількість кластерів.
3. Виконати кластеризацію методом DBSCAN та порівняти результати.

Зниження розмірності

4. Застосувати PCA для 2D або 3D візуалізації.
5. Побудувати графіки кластерів у зменшеному просторі.

Оцінка якості

6. Розрахувати silhouette score або інші метрики якості.
7. Зробити висновки щодо структури даних і придатності методів.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Порівняння методів кластеризації, графіки та інтерпретацію результатів.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Чим відрізняються k-means та DBSCAN?
2. Як обрати кількість кластерів у k-means?
3. Що таке silhouette score і як його інтерпретувати?
4. Навіщо потрібне зниження розмірності?
5. Які обмеження має PCA?
6. Як впливають масштабування та шум на кластеризацію?
7. Що таке щільнісні методи кластеризації?
8. Як оцінити стабільність результатів кластеризації?

ЛАБОРАТОРНА РОБОТА №8

ТЕМА: ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

МЕТА: Навчитися готувати часові ряди до моделювання та будувати базові моделі прогнозу.

ЗНАТИ: Поняття тренду, сезонності, основи ARIMA, принципи оцінки прогнозу.

ТЕОРЕТИЧНІ ВІДОМОСТІ

У зв’язку з великим обсягом інформації використовувати електронну документацію (погоджувати з лектором).

Основна документація:

- Statsmodels: Time Series Analysis https://www.statsmodels.org/stable/tsa.html
- Forecasting Principles https://otexts.com/fpp3/

Додаткова документація:

– https://pandas.pydata.org/docs/user_guide/timeseries.html

– https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html

УМОВИ ВИКОНАННЯ Л/Р

Лабораторна робота виконується у середовищі Python. Формат подання результатів погоджувати з викладачем.

ЗАВДАННЯ

Необхідно побудувати базову модель прогнозування для обраного часового ряду та оцінити її якість.

Підготовка до виконання

1. Обрати часовий ряд та описати його частоту і горизонт прогнозу.
2. Виконати розбиття на train/test з урахуванням хронології.

Аналіз та моделювання

3. Декомпозувати ряд на тренд/сезонність (за можливості).
4. Побудувати модель ARIMA або іншу базову модель.

Оцінка прогнозу

5. Розрахувати метрики якості (MAE, RMSE або MAPE).
6. Побудувати графік фактичних і прогнозних значень.
7. Зробити висновки щодо якості прогнозу та можливих покращень.

ОФОРМЛЕННЯ ЗВІТУ

У звіт лабораторної роботи помістити:

– Завдання л/роботи.

– Опис часового ряду, параметри моделі, метрики та графіки прогнозу.

Лабораторну роботу оформляти за загальноприйнятою формою (ДСТУ) та аналогічно до інших предметів кафедри.

КОНТРОЛЬНІ ЗАПИТАННЯ

1. Чим відрізняється розбиття для часових рядів від стандартного train/test?
2. Що таке тренд і сезонність?
3. Які основні параметри має модель ARIMA?
4. Які метрики застосовують для оцінки прогнозів?
5. Навіщо потрібна декомпозиція часового ряду?
6. Які ризики викликає нерівномірна частота даних?
7. Як визначити горизонт прогнозу?
8. Які способи покращення базового прогнозу?

Контроль знань

Експрес-опитування, письмові контрольні завдання, поточний контроль у

Види контролю: поточний, підсумковий.

Методи контролю: спостереження за освітньою діяльністю здобувачів, усне опитування.

Форма підсумкового контролю: екзамен.

Контроль знань і вмінь здобувачів (поточний і підсумковий) здійснюється згідно з кредитною трансферно-накопичувальною системою організації освітнього процесу. Рейтинг здобувача вищої освіти визначається за 100 бальною шкалою: складається з рейтингу із поточної академічної активності впродовж семестру, для оцінювання якої призначається 100 балів (по 30 балів за кожен змістовний модуль та 40 балів за екзамен - оцінка здобувача вищої освіти).

Оцінювання здійснюється відповідно до вимог Положення про організацію освітнього процесу в ЦНТУ (розділ 8, стор. 28-30, 32-33; апеляція – стор. 34-35) за такими **критеріями оцінювання академічних досягнень** здобувачів вищої освіти:

- Вчасність,
- ініціативна дослідна робота в межах СРС,
- активність, ініціативність на заняттях.

Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота							Сума балів за рубіжні контролі	Підсумковий контроль (екзамен)
І рубіж				ІІ рубіж				
Т1	Т2	Т3	Т4	Т5	Т6	Т7	60	100
7	7	8	8	10	10	10		
30				30				

Шкала оцінювання знань. Шкала оцінювання: національна та ЄКТС

Сума балів за всіма видами навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
1-34	F	незадовільно з обов’язковим повторним вивченням дисципліни	не зараховано з обов’язковим повторним вивченням дисципліни

При оцінюванні знань студентів використовуються такі засоби контролю: усне опитування перед допуском до виконання лабораторної роботи – здійснюється на їх початку; якість виконання, набуття теоретичних знань і практичних навичок перевіряється шляхом захисту кожної лабораторної роботи згідно з робочою програмою дисципліни.

Оцінка, яка виставляється за лабораторне заняття, складається з таких елементів: усне опитування студентів перед допуском до виконання лабораторної роботи; знання теоретичного матеріалу з теми; якість оформлення звіту; вміння студента обґрунтувати прийняті конструктивні рішення; покращувати результат після зворотного зв'язку; своєчасний захист лабораторної роботи. Для виконання програми дисципліни студент повинен отримати оцінки за всі лабораторні роботи.

Термін захисту лабораторної роботи вважається своєчасним, якщо студент захистив її на наступному після виконання роботи занятті. За несвоєчасний захист лабораторної роботи з неповажної причини студент за позитивну відповідь отримує оцінку «задовільно».

Пропущене лабораторне заняття студент повинен відпрацювати в лабораторіях кафедри у встановлений викладачем термін, але не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Відповідність шкали оцінювання ЄКТС національній системі оцінювання визначена у розділі 8 Положення про організацію освітнього процесу в ЦНТУ: таблиця 1, стор. 33 www.kntu.kr.ua/file/content/424/polozhennia-pro-orhanizatsiyu-osvitnoho-protsesu-v-tsntu.pdf

Критерії оцінки іспиту:

оцінку «відмінно» (90-100 балів, А) заслуговує студент, який:

- всебічно, систематично і глибоко володіє навчально-програмовим матеріалом;

- вміє самостійно виконувати завдання, передбачені програмою, використовує набуті знання і вміння у нестандартних ситуаціях;

- засвоїв основну і ознайомлений з додатковою літературою, яка рекомендована програмою;

- засвоїв взаємозв'язок основних понять дисципліни та усвідомлює їх значення для професії, яку він набуває;

- вільно висловлює власні думки, самостійно оцінює різноманітні життєві явища і факти, виявляючи особистісну позицію;

- самостійно визначає окремі цілі власної навчальної діяльності, виявив творчі здібності і використовує їх при вивченні навчально-програмового матеріалу,проявив нахил до наукової роботи.

оцінку « добре» (82-89 балів, B) – заслуговує студент, який:

- повністю опанував і вільно (самостійно) володіє навчально-програмовим матеріалом, в тому числі застосовує його на практиці, має системні знання достатньому обсязі відповідно до навчально-програмового матеріалу, аргументовано використовує їх у різних ситуаціях;

- має здатність до самостійного пошуку інформації, а також до аналізу, постановки і розв'язування проблем професійного спрямування;

- під час відповіді допустив деякі неточності, які самостійно виправляє, добирає переконливі аргументи на підтвердження вивченого матеріалу;

оцінку «добре» (74-81 бал, C) заслуговує студент, який:

- в загальному роботу виконав, але відповідає на екзамені з певною кількістю помилок;

- вміє порівнювати, узагальнювати, систематизувати інформацію під керівництвом викладача, в цілому самостійно застосовувати на практиці, контролювати власну діяльність;

- опанував навчально-програмовий матеріал, успішно виконав завдання, передбачені програмою, засвоїв основну літературу, яка рекомендована програмою;

оцінку «задовільно» (64-73 бали, D) – заслуговує студент, який:

- знає основний навчально-програмовий матеріал в обсязі, необхідному для подальшого навчання і використання його у майбутній професії;

- виконує завдання, але при рішенні допускає значну кількість помилок;

- ознайомлений з основною літературою, яка рекомендована програмою;

- допускає на заняттях чи екзамені помилки при виконанні завдань, але під керівництвом викладача знаходить шляхи їх усунення.

оцінку «задовільно» (60-63 бали, E) – заслуговує студент, який:

- володіє основним навчально-програмовим матеріалом в обсязі, необхідному для подальшого навчання і використання його у майбутній професії, а виконання завдань задовольняє мінімальні критерії. Знання мають репродуктивний характер.

оцінка «незадовільно» (35-59 балів, FX) – виставляється студенту, який:

виявив суттєві прогалини в знаннях основного програмового матеріалу, допустив принципові помилки у виконанні передбачених програмою завдань.

оцінку «незадовільно» (35 балів, F) – виставляється студенту, який:

- володіє навчальним матеріалом тільки на рівні елементарного розпізнавання і відтворення окремих фактів або не володіє зовсім;

- допускає грубі помилки при виконанні завдань, передбачених програмою;

- не може продовжувати навчання і не готовий до професійної діяльності після закінчення університету без повторного вивчення даної дисципліни.

При виставленні оцінки враховуються результати навчальної роботи студента протягом семестру

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

Базова

1. Kevin P. Murphy. Probabilistic Machine Learning: Advanced Topics. The MIT Press, 2023. ISBN 0262048434.

2. Kevin P. Murphy. Probabilistic Machine Learning: An Introduction. The MIT Press, 2022. ISBN 0262046822.

3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012. ISBN 0262018020.

4. Читальний зал № 1 (ЦНТУ) Фісун М. Т., Кравець І. О., Казмірчук П. П., Ніколенко С. Г.. Інтелектуальний аналіз даних: практикум : навч. посіб. Львів : Новий Світ 2000, 2023. 161 с. ISBN 978-966-418-287-7.

5. Читальний зал № 1 (ЦНТУ) Висоцька В. А., Досин Д. Г., Микіч Х. І. та ін. Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій : монографія. Львів : Новий Світ 2000, 2019. 236 с. ISBN 978-617-7519-47-7.

6. Читальний зал № 1 (ЦНТУ) Нікольський Ю. В., Пасічник В. В., Щербина Ю. М.. Системи штучного інтелекту : навч. посіб. 2-ге вид., випр. та доп. Львів : Магнолія, 279 с. ISBN 978-617-57-40-11-4.

7. Alexander Jung. Machine Learning: The Basics. Springer, 2022. ISBN 9811681929.

Допоміжна

8. Кавун С.В., Смірнов О.А.,Сорбат І.В., Мелешко Є.В., Коваленко О.В. Системи штучного інтелекту // Навчальний посібник – Кіровоград: КНТУ 2013. – 335 с.

9. Інженерія програмного забезпечення: Навчальний посібник / О.А. Смірнов, О.В. Коваленко, Є.В. Мелешко, Л.В. Константинова, А.С. Кожанова // Кіровоград: КНТУ 2013. – 335 с.

10. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar. Foundations of Machine Learning. The MIT Press, 2018. ISBN 0262039400.
11. John D. Kelleher, Brian Mac Namee, Aoife DArcy. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. The MIT Press, 2020. ISBN 0262044692.
12. Ruye Wang. Introduction to Machine Learning: From Math to Code. Cambridge University Press, 2025. ISBN 1316519503.
13. Ethem Alpaydin. Introduction to Machine Learning. The MIT Press, 2020. ISBN 0262043793.
14. Christopher M. Bishop, Hugh Bishop. Deep Learning: Foundations and Concepts. Springer, 2023. ISBN 3031454677.
15. Andriy Burkov. The Hundred-Page Machine Learning Book. Andriy Burkov, 2019. ISBN 1999579518.
16. Richard S. Sutton, Andrew G. Barto. Reinforcement Learning: An Introduction. The MIT Press, 2018. ISBN 0262039249.
17. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006. ISBN 0387310738.
18. Jeffrey A. Fessler, Raj Rao Nadakuditi. Linear Algebra for Data Science, Machine Learning, and Signal Processing. Cambridge University Press, 2024. ISBN 1009418149.
19. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009. ISBN 0387848576.
20. Andreas Lindholm, Niklas Wahlström, Fredrik Lindsten, Thomas B. Schön. Machine Learning: A First Course for Engineers and Scientists. Cambridge University Press, 2022. ISBN 1108843603.
21. Christian Kästner. Machine Learning in Production: From Models to Products. The MIT Press, 2025. ISBN 0262049724.

Наукові публікації

22. Mohammed A.S., Meleshko, Y., Balaji S.B., Semenov S. Collaborative Filtering Method with the use of Production Rules Proceedings of ICCIKE, Amity University DubaiDubai; United Arab Emirates. –2019. – с. 387-391 **(SCOPUS) (Web of Science)** URL: https://www.scopus.com/record/display.uri?eid=2-s2.0-85080933551&origin=resultslist

23. Мелешко Є.В. Методи оцінки якості роботи рекомендаційних систем // Збірник наукових праць Системи управління, навігації та зв'язку. – Полтава: ПНТУ, 2018. - Вип. 5 (51). - С. 92-97. **(фахове видання)** – URL: http://journals.nupp.edu.ua/sunz/article/view/1271 **(Index Copernicus, Google Scholar)**

24. Мелешко Є.В. Проблеми сучасних рекомендаційних систем та методи їх рішення // Збірник наукових праць Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ, 2018. – Т. 4 (50). – С. 120-124. **(фахове видання)** – URL: http://journals.nupp.edu.ua/sunz/article/view/1211 **(Index Copernicus, Google Scholar)**

25. Meleshko Ye. Method of collaborative filtration based on associative networks of users similarity // Науковий журнал Сучасні інформаційні системи. – Харків: НТУ "ХПИ", 2018. - Т. 2, № 4. - С. 55-59. **(фахове видання)** – URL: http://ais.khpi.edu.ua/article/view/2522-9052.2018.4.09 **(Google Scholar)**

26. Мелешко Є.В. Методи кластеризації графів соціальних мереж для побудови рекомендаційних систем // Збірник наукових праць Системи управління, навігації та зв'язку. Збірник наукових праць. – Полтава: ПНТУ, 2019. – Т. 2 (54). – С. 129-134. **(фахове видання)** – URL: http://journals.nupp.edu.ua/sunz/article/view/1421 **(Index Copernicus, Google Scholar)**

27. Meleshko Ye. Method of generating recommendations lists with considering activity indexes of users in a recommendation system // Науковий журнал Сучасні інформаційні системи. – Харків: НТУ "ХПИ", 2019. – Т. 3, № 1. – С. 43-47. **(фахове видання)** – URL: http://ais.khpi.edu.ua/article/view/2522-9052.2019.1.07 **(Google Scholar)**

28. Meleshko Ye. Computer model of virtual social network with recommendation system // Науковий журнал Сучасний стан наукових досліджень та технологій в промисловості. – Харків: ХНУРЕ, 2019. – Вип. 2(8). - С. 80-84 **(фахове видання)** – URL: http://journals.uran.ua/itssi/article/view/2522-9818.2019.8.080 **(Google Scholar)**

29. Meleshko Ye., Drieiev O., Drieieva H. Method of identification bot profiles based on neural networks in recommendation systems // Advanced Information Systems. – 2020. – Vol. 4, No. 2 – С. 24-28. **(фахове видання)** – URL: http://ais.khpi.edu.ua/article/view/2522-9052.2020.2.05 **(Google Scholar)**

30. Meleshko Ye., Drieiev O., Al-Oraiqat A.M. The improved model of user similarity coefficients computation for recommendation systems // Innovative technologies and scientific solutions for industries, Vol. 3 (13). – 2020. – pp. 52-57. doi: https://doi.org/10.20998/2522-9052.2020.3.06 **(фахове видання)** – URL: http://ais.khpi.edu.ua/article/view/2522-9052.2020.3.06 **(Google Scholar)**

31. Meleshko Ye. The method of detecting information attack objects in recommendation system based on the analysis of rating trends // Innovative technologies and scientific solutions for industries, Vol. 3 (13). – 2020. – pp. 52-57. doi: 10.30837/itssi.2020.13.052 **(фахове видання)** – URL: https://www.itssi-journal.com/index.php/ittsi/article/view/215 **(Google Scholar)**

32. Міхав В.В., Мелешко Є.В., Шимко С.В. Методи та структури даних для реалізації бази даних рекомендаційної системи соціальної мережі // Техніка в сільськогосподарському виробництві, галузеве машинобудування, автоматизація: збірник наукових праць Центральноукраїнського національного технічного університету.– Кропивницький: ЦНТУ.– вип. 4(35) – 2021. С. 8-16 **(Фахове видання категорії Б)** – URL: http://mapiea.kntu.kr.ua/archive/35.html **(Index Copernicus, ResearchBib, Google Scholar)**

33. Meleshko Y., Yakymenko M., Bosko V. A method of computer simulation modeling of user and bot behavior in a recommendation system using the graph database NEO4J // Innovative technologies and scientific solutions for industries, Kharkiv,

Ukraine. – 2021. – Vol. 3(17). – pp. 23-31. DOI: https://doi.org/10.30837/ITSSI.2021.17.023 (ISSN 2522-9818 (print)) **(Фахове видання категорії Б) (Index Copernicus, Google Scholar)**

34. Міхав В.В., Мелешко Є.В., Якименко М.С., Бащенко Д.В. Методи зберігання даних рекомендаційної системи на основі зв'язних списків // Системи управління, навігації та зв'язку – Полтава: ПНТУ, 2021. – Т. 4(66). – С. 59-62. – doi: https://doi.org/10.26906/SUNZ.2021.4.059. **(Фахове видання категорії Б)**

Інформаційні ресурси

35. Курс «Machine Learning» на сервері дистанційної освіта ЦНТУ. – URL: https://moodle.kntu.kr.ua/course/view.php?id=1087

36. https://prometheus.org.ua/ – українська платформа безкоштовних онлайн-курсів

37. https://www.tensorflow.org/ – An end-to-end open source machine learning platform

38. http://leenissen.dk/ – Fast Artificial Neural Network Library

39. https://www.codeproject.com/ – колективний блог з новинами та навчальними статтями про інформаційні технології та програмування.

40. http://stackoverflow.com/ – система питань і відповідей для професійних програмістів та новачків у програмуванні.

41. https://dou.ua/ – український веб-сайт з елементами колективного блогу, створений для розповсюдження новин, аналітичних статей та свіжої інформації пов'язаної із інформаційними технологіями.

42. http://www.algomation.com/ – це платформа для перегляду, обміну і створення візуалізацій алгоритмів.