



Силабус курсу

Засоби аналізу даних на Python

Ступінь вищої освіти-бакалавр

Освітньо-професійна програма: «Інженерія програмного забезпечення»

Спеціальність – 121 Інженерія програмного забезпечення

Рік навчання: 2, Семестр: 4

Кредитів: 3 Мова викладання: українська

Керівник курсу

ППП

д. ф., ст. викладач Тимчишин Василь Степанович

Контактна інформація

kn@wunu.edu.ua, +38 (0352) 23-60-29

Опис дисципліни

Дисципліна спрямована на формування теоретичних знань та практичних навичок використання мови програмування Python для роботи з даними. У ході вивчення курсу студенти опановують основні принципи обробки, аналізу та візуалізації даних, ознайомлюються з поняттями статистичної обробки, підготовки даних, а також базовими методами машинного навчання. Особлива увага приділяється застосуванню найпоширеніших інструментів екосистеми Python, зокрема бібліотек NumPy, SciPy, Pandas, Matplotlib, Scikit-learn. Результатом навчання є здатність студентів самостійно виконувати аналітичні завдання на основі реальних наборів даних, застосовуючи сучасні засоби Python.

Структура курсу

Години (лек./пр.)	Тема	Результати навчання	Завдання
2/1	Тема 1. Основні поняття аналізу та обробки даних	Знання етапів аналізу та обробки. Опанування базових структур даних у Python та принципів роботи з ними. Ознайомлення з інструментами та бібліотеками Python для аналізу даних.	Практична робота
2/1	Тема 2. Джерела, формати та якість даних	Знання основних джерел отримання даних та форматів їх подання. Навички імпорту та експорту даних у Python. Забезпечення цілісності та якості наборів даних.	Практична робота
2/1	Тема 3. Методи та алгоритми статистичної обробки даних	Застосування описової статистики для аналізу вибірок. Проведення перевірки статистичних гіпотез. Здійснення інтерполяції та екстраполяції даних.	Практична робота
2/1	Тема 4. Методи попередньої обробки даних	Масштабування, нормалізація та стандартизація змінних. Кодування категоріальних ознак. Формування тренувальних і тестових вибірок.	Практична робота

3/1	Тема 5. Опрацювання даних засобами Python	Використання бібліотек NumPy і Pandas для опрацювання даних. Створення та модифікація таблиць DataFrame. Фільтрація, сортування, групування та агрегація даних.	Практична робота
3/1	Тема 6. Методи візуалізації та інтерпретації даних	Створення візуалізацій у Matplotlib, Seaborn, Plotly. Побудова діаграм, гістограм, boxplot, heatmap, pairplot. Візуальна інтерпретація трендів і кореляційних зв'язків.	Практична робота
3/1	Тема 7. Методи групування та кластеризації даних	Вибір метрик відстаней для оцінки схожості об'єктів. Оцінювання якості кластеризації (silhouette score, inertia). Візуалізація кластерів та групування об'єктів.	Практична робота
3/1	Тема 8. Аналіз часових рядів і прогнозування	Виявлення тренду, сезонності та шуму в часових рядах. Застосування методів згладжування часових рядів. Побудова моделей ARIMA, SARIMA, Prophet.	Практична робота
3/1	Тема 9. Вступ до інтелектуального аналізу даних та штучного інтелекту	Розуміння відмінностей між Data Mining, Machine Learning та Artificial Intelligence. Ознайомлення з напрямками застосування ШІ. Знання основних етапів побудови рішення на основі ШІ.	Практична робота
3/2	Тема 10. Методи машинного навчання з учителем і без учителя	Опанування методів supervised та unsupervised learning. Застосування алгоритмів класифікації та регресії. Побудова моделей у Scikit-learn та оцінювання їх ефективності.	Практична робота
3/2	Тема 11. Глибинне навчання та нейронні мережі	Розуміння архітектури нейронних мереж. Створення моделей у TensorFlow/Keras або PyTorch. Застосування Dense, CNN і RNN шарів у прикладних задачах.	Практична робота
3/1	Тема 12. Оцінка, оптимізація та пояснення моделей штучного інтелекту	Вибір та розрахунок метрик оцінювання моделей (accuracy, precision, recall, F1, ROC-AUC). Оптимізація гіперпараметрів (Grid Search, Random Search). Пояснення моделей за допомогою SHAP, LIME. Урахування етичних аспектів застосування ШІ.	Практична робота

Літературні джерела

1. McKinney W. Python for Data Analysis. – 3rd ed. – O'Reilly Media, 2022. – 550 p.
2. VanderPlas J. Python Data Science Handbook. – 2nd ed. – O'Reilly Media, 2023. – 740 p.
3. Downey A. B. Think Stats: Probability and Statistics for Programmers. – 3rd ed. – O'Reilly Media, 2022. – 320 p.
4. Kelleher J. D. Data Science: An Introduction. – CRC Press, 2021. – 420 p.
5. Miller J. D. Statistics for Data Science. – Packt Publishing, 2021. – 390 p.
6. Seaborn & Matplotlib Documentation. Data Visualization Tools for Python. – Online documentation. – 2023.
7. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. – 3rd ed. – O'Reilly Media, 2023. – 880 p.

8. Pandas Documentation. Working with Structured Data in Python. – Online documentation. – 2024.
9. Brownlee J. Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End. – Machine Learning Mastery, 2022. – 575 p.
10. Goodfellow I., Bengio Y., Courville A. Deep Learning. – MIT Press, 2023. – 800 p.
11. Ng A. Machine Learning Yearning. – DeepLearning.ai, 2022. – 120 p.
12. Chollet F. Deep Learning with Python. – 2nd ed. – Manning Publications, 2021. – 504 p.
13. Burkov A. The Hundred-Page Machine Learning Book. – 2nd ed. – Andriy Burkov, 2022. – 160 p.
14. Raschka S., Mirjalili V. Python Machine Learning. – 4th ed. – Packt Publishing, 2023. – 750 p.
15. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. – 2nd ed. – Independently Published, 2024. – 330 p.

Політика оцінювання

– *Політика щодо дедлайнів і перекладання.* Для виконання усіх видів завдань здобувачами освіти і проведення контрольних заходів встановлюються конкретні терміни. Перекладання модулів проводиться в установленому порядку.

– *Політика щодо академічної доброчесності.* Списування під час проведення контрольних заходів заборонені. Під час контрольного заходу учасник може користуватися лише дозволеними допоміжними матеріалами або засобами, йому забороняється в будь-якій формі обмінюватися інформацією з іншими учасниками, використовувати, розповсюджувати, збирати варіанти контрольних завдань.

– *Політика щодо відвідування.* За об'єктивних причин (наприклад, карантин, воєнний стан, хвороба, закордонне стажування) навчання може відбуватись в дистанційній формі за погодженням із керівником курсу з дозволу дирекції факультету.

Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Засоби аналізу даних на Python» визначається як середньозважена величина, залежно від питомої ваги кожної складової залікового кредиту:

Модуль 1		Модуль 2	Модуль 3
40 %	40 %	5 %	15%
Поточне оцінювання	Модульний контроль	Тренінги	Самостійна робота
Оцінка за поточне оцінювання визначається, як середнє арифметичне з оцінок, отриманих під час занять	Середнє арифметичне із оцінок, отриманих за виконання практичних робіт 1-4	Визначається як єдина оцінка, отримана під час презентації проєкту, розробленого під час тренінгу	Визначається як єдина оцінка по 100-бальній шкалі за результат самостійної роботи за обраною тематикою

Виконання лабораторних робіт:

90-100 балів (Відмінно) - здобувач освіти самостійно, без помилок, виконав усі кроки в рамках лабораторної роботи, правильно задокументував етапи та вільно оперує поняттями та принципами, що відносяться до теми дисципліни.

75-89 балів (Добре) - здобувач освіти виконав завдання лабораторної роботи, проте в процесі виконання допустив кілька дрібних помилок, які не вплинули на кінцевий результат (наприклад, неточна послідовність дій), в процесі роботи виникали додаткові запитання.

60-74 балів (Задовільно) - здобувач освіти виконав завдання лабораторної роботи, але з суттєвими помилками, наприклад, не з першого разу чи не до кінця. Розуміння поставлених у лабораторній роботі завдань є поверхневим та неповним.

1-59 балів (Незадовільно) - здобувач освіти не зміг виконати завдання або результати були повністю невірними. Не продемонстрував базових навичок роботи з програмним забезпеченням.

Тренінг:

90-100 балів (Відмінно) - здобувач освіти самостійно, без помилок, виконав усі етапи завдання, правильно задокументував усі етапи роботи, та вільно оперує поняттями та принципами дисципліни.

75-89 балів (Добре) - здобувач освіти виконав завдання, але з кількома дрібними помилками, які не вплинули на кінцевий результат, в процесі роботи виникали додаткові запитання.

60-74 балів (Задовільно) - здобувач освіти виконав завдання, але з суттєвими помилками, наприклад, не з першого разу. Розуміння поставлених у тренінгу завдань є поверхневим.

1-59 балів (Незадовільно) - здобувач освіти не зміг виконати завдання або результати були повністю невірними. Не продемонстрував достатній рівень навичок роботи з апаратним та програмним забезпеченням.

Самостійна робота: оцінюється у вигляді єдиної підсумкової оцінки за 100-бальною шкалою на основі результатів самостійної роботи студента з обраної теми.

Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)