



Інженерні методи аналізу даних для автоматизації та робототехніки

Силабус

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>17 Електроніка, автоматизація та електронні комунікації</i>
Спеціальність	<i>174 Автоматизація, комп'ютерно-інтегровані технології та робототехніка</i>
Освітня програма	<i>Комп'ютерно-інтегровані системи та технології в приладобудуванні</i>
Статус дисципліни	<i>За вибором</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>3 курс, весняний семестр</i>
Обсяг дисципліни / кредитного модуля	<i>4 кредити ЕКТС (120 годин), 54 години аудиторної роботи, 66 годин самостійної роботи</i>
Семестровий контроль/ контрольні заходи	<i>МКР, залік / поточний контроль</i>
Розклад занять	<i>Згідно з розкладом на сайті http://roz.kpi.ua/</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>к.т.н., доцент Лисенко Юлія Юріївна, j.lysenko@kpi.ua ас. Драчук Олеся Олександрівна, o.drachuk@kpi.ua</i>
Розміщення курсу	<i>https://classroom.google.com</i>

Програма навчальної дисципліни

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

Навчальна дисципліна “Інженерні методи аналізу даних для автоматизації та робототехніки” є продовженням дисципліни з програмування на Python, з акцентом на вивчення та використання бібліотек для роботи з числовими даними та даними у різних форматах. Він охоплює методологію створення зрозумілих та інформативних візуалізацій, а також розглядає такі теми, як лінійна регресія, кластеризація та прийняття рішень. Особлива увага приділяється практичним прикладам використання цих методів у галузі автоматизації та робототехніки.

Мета курсу: Забезпечити студентів затребуваними навичками програмування, аналізу та візуалізації даних за допомогою сучасних інструментів Python, а також навчити їх застосовувати ці навички для вирішення реальних проблем в автоматизації та робототехніці.

Предмет вивчання: програмування на Python, використання бібліотек NumPy і pandas для роботи з числовими даними, методології візуалізації даних за допомогою matplotlib, seaborn та plotly, а також техніки та підходи до аналізу даних у контексті автоматизації та робототехніки.

Результатом вивчення дисципліни є формування у студентів **компетентностей**:

- навички використання інформаційних і комунікаційних технологій;
- здатність застосовувати методи системного аналізу, математичного моделювання, ідентифікації та числові методи для розроблення математичних моделей окремих елементів та систем автоматизації в цілому, для аналізу якості їх функціонування із використанням новітніх комп'ютерних технологій;
- здатність обґрунтовувати вибір технічних засобів автоматизації на основі розуміння принципів їх роботи аналізу їх властивостей, призначення і технічних характеристик з урахуванням вимог до системи автоматизації і експлуатаційних умов; налагоджувати технічні засоби автоматизації та системи керування;
- здатність використовувати для вирішення професійних завдань новітні технології у галузі автоматизації та комп'ютерно-інтегрованих технологій, зокрема, проектування багаторівневих систем керування, збору даних та їх архівування для формування бази даних параметрів процесу та їх візуалізації за допомогою засобів людино-машинного інтерфейсу.

Згідно з вимогами освітньо-професійної програми студенти після засвоєння дисципліни мають продемонструвати такі **результати навчання**:

- знати лінійну та векторну алгебру, диференціальне та інтегральне числення, функції багатьох змінних, функціональні ряди, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію функції комплексної змінної, теорію ймовірностей та математичну статистику, теорію випадкових процесів в обсязі, необхідному для користування математичним апаратом та методами у галузі автоматизації;
- вміти застосовувати сучасні інформаційні технології та мати навички розробляти алгоритми та комп'ютерні програми з використанням мов високого рівня та технологій об'єктно-орієнтованого програмування, створювати бази даних та використовувати інтернет-ресурси;
- розуміти суть процесів, що відбуваються в об'єктах автоматизації (за галузями діяльності) та вміти проводити аналіз об'єктів автоматизації і обґрунтовувати вибір структури, алгоритмів та схем керування ними на основі результатів дослідження їх властивостей. Вміти застосовувати методи теорії автоматичного керування для дослідження, аналізу та синтезу систем автоматичного керування;
- вміти застосовувати методи системного аналізу, моделювання, ідентифікації та числові методи для розроблення математичних та імітаційних моделей окремих елементів та систем автоматизації в цілому, для аналізу якості їх функціонування із використанням новітніх комп'ютерних технологій.

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

Навчальна дисципліна «Інженерні методи аналізу даних для автоматизації та робототехніки»:

- 1) базується на знаннях, здобутих студентами при вивченні таких дисциплін, як вища математика, програмування; 2) є основою для дисциплін, пов'язаних з комп'ютерним моделюванням процесів та систем; 3) може бути використана під час виконання дипломного проєкту (роботи).

3. Зміст навчальної дисципліни

Дисципліна складається з 5-ти розділів. Тематичний план дисципліни:

Тема 1. Аналіз даних інструментами Python.

Тема 2. Візуалізації даних інструментами Python.

Тема 3. Основи машинного навчання.

Тема 4. Алгоритми машинного навчання.

Тема 5. Аналіз великих даних із застосуванням Spark.

4. Навчальні матеріали та ресурси

Базові джерела

- 1) Підручник з Python. – 2024, Python Software Foundation. Доступ: <https://docs.python.org/uk/3/tutorial/index.html>
- 2) W. McKinney. (2022) Python for Data Analysis, 3rd Edition, O'Reilly Media.
- 3) S. Mukhopadhyay, P. Samanta (2023) Advanced Data Analytics Using Python: With Machine Learning, Deep Learning and NLP Examples, Apress.
- 4) J. VanderPlas. (2022) Python Data Science Handbook: Essential Tools for Working with Data, 2nd Edition, O'Reilly Media.

Допоміжна

- 1) Raschka, S. and Mirjalili, V. (2019) Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn and TensorFlow 2. 3rd Edition, Packt Publishing, Birmingham.
- 2) Martinez, D. and Jesús, S. D. (2020) Applied Computational Thinking with Python: Design Algorithmic Solutions for Complex and Challenging Real-world Problems (1 ed.), Packt Publishing.
- 3) Christoph Dürr and Jill-Jênn. (2020) Competitive Programming in Python (1 ed.), Cambridge University Press.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Основні форми навчання – лекції, комп'ютерні практикуми, самостійна робота студентів.

Лекційний курс розрахований на вивчення основ аналізу даних, візуалізації даних та основи машинного навчання з використанням Python, включаючи роботу з бібліотеками NumPy, Pandas, Matplotlib, Seaborn, Plotly. Курс охоплює як теоретичні основи, так і практичні аспекти застосування технологій для аналізу великих даних, створення інтерактивних візуалізацій, зокрема для бізнес- та наукових застосувань.

Тема 1. Аналіз даних інструментами Python.

Лекція 1. Вступ до аналізу даних.

Розглядаються основи використання найважливіших бібліотеки для роботи з даними (NumPy і Pandas), базові принципи обробки та маніпулювання даними.

Тема 2. Візуалізації даних інструментами Python.

Лекція 2. Основи візуалізації даних з Matplotlib та Seaborn

Ознайомлення з основами візуалізації даних за допомогою бібліотек Matplotlib та Seaborn, що дозволяє створювати графіки та діаграми для візуального аналізу даних. Особлива увага приділяється налаштуванню графіків і інтерактивності візуалізацій.

Лекція 3. Інтерактивні візуалізації з Plotly та Cufflinks

Вивчаються інструменти для створення інтерактивних візуалізацій даних, що дозволяють користувачам взаємодіяти з графіками через сторонні засоби, створення динамічних діаграм та інтерактивних карт.

Тема 3. Основи машинного навчання.

Лекція 4. Основи статистики та регресії для аналізу даних

Розглядаються основні статистичні методи, такі як описова статистика, лінійна регресія та методи оцінки моделей, вивчається застосування цих методів для аналізу та інтерпретації даних.

Тема 4. Алгоритми машинного навчання.

Лекція 5. Введення в алгоритми класифікації

Вивчаються основи машинного навчання, зокрема алгоритми класифікації (логістична регресія та метод найближчих сусідів, розглянуто процес побудови моделей для передбачення категорій).

Лекція 6. Кластеризація та зменшення вимірів

Розглядаються методи кластеризації та оптимізації вимірів, що дозволяють знаходити структуру в даних та зменшувати складність моделей.

Лекція 7. Введення в нейронні мережі

Розглядаються основи нейронних мереж, їх архітектура та принципи роботи, проста нейронна мережа для класифікації та регресії.

Тема 5. Аналіз великих даних із застосуванням Spark.

Лекція 8. Обробка природної мови (NLP)

Ознайомлення з основами обробки природної мови (NLP), основними методами для аналізу текстових даних (токенізація, стемінг та векторизація тощо).

Лекція 9. Великі дані та обробка з Apache Spark

Розглядається робота з великими даними та з використанням платформи Apache Spark для масштабованого аналізу даних, вивчається обробка великі обсягів за допомогою бібліотек для розподіленого обчислення. **Залік**

Мета проведення **практичних занять** – закріплення на практиці теоретичних знань, набутих студентами під час лекційних занять з дисципліни.

Практичні заняття проводяться у формі комп'ютерних практикумів. Основне завдання практикумів – оволодіння практичними вміннями застосовувати чисельні методи для вирішення інженерних задач.

№	Назва комп'ютерного практикуму	К-ть ауд. годин
1	Вступне заняття. Знайомство з робочим місцем. Техніка безпеки. Вступ до аналізу даних у Python. Бібліотеки NumPy та Pandas, введення в базові операції з даними. Виконання операцій з багатовимірними масивами: маніпулювання даними, арифметичні операції, індексація та фільтрація	2
2	Робота з таблицями в Pandas: імпорт/експорт даних, очищення даних, групування, агрегація та маніпулювання таблицями.	2
3	Індивідуальне завдання 1	2
4	Основи візуалізації даних з Matplotlib. Створення базових графіків, налаштування параметрів та стилів.	2
5	Створення комплексних графіків із Seaborn. Використання бібліотеки Seaborn для побудови карт, парних графіків, графіків кореляцій.	2
6	Індивідуальне завдання 2	2
7	Побудова регресійних моделей. Введення в лінійну регресію, аналіз точності моделі, використання основних метрик оцінки.	2
8	Алгоритми класифікації. Робота з алгоритмами логістичної регресії та методу найближчих сусідів, оцінка моделей.	2
9	Індивідуальне завдання 3	2
10	Кластеризація даних за допомогою K-Means. Аналіз та групування даних з використанням кластеризації, візуалізація результатів.	2
11	Зменшення вимірів за допомогою PCA. Вивчення методу головних компонент для оптимізації структури даних.	2
12	Індивідуальне завдання 4	2
13	Основи обробки природної мови (Natural Language Processing). Робота з текстовими даними: токенізація, стемінг, векторизація тексту	2
14	Побудова нейронних мереж для аналізу даних. Вивчення простої нейронної мережі для задач класифікації або регресії з використанням Keras.	2
15	Індивідуальне завдання 5	2
16	Робота з великими наборами даних за допомогою PySpark, виконання паралельних обчислень. Підготовка даних для масштабованого аналізу	2
17	Модульна контрольна робота	2
18	Підсумкове заняття. Обговорення результатів контрольної. Пояснення на прикладах помилок, які виникли. Аналіз результатів роботи протягом семестру	2

6. Самостійна робота студента

У відповідності до робочого навчального плану передбачено 66 годин самостійної роботи студентів, з яких 6 годин – на підготовку до заліку і 54 годин на підготовку до аудиторних занять, опрацювання матеріалів лекцій, модульної контрольної, самостійний розв'язок додаткових задач, оформлення звітів про результати виконання комп'ютерних практикумів та ознайомлення із джерелами відповідно до списку. Самостійна робота спрямована на засвоєння та поглиблення вивченого матеріалу, а також на підготовку до аудиторних занять і семестрового контролю.

7. Політика навчальної дисципліни

Перед студентом ставляться наступні вимоги:

- **правила відвідування занять:**
 - у режимі очного навчання заняття відбуваються в аудиторії відповідно до розкладу занять;
 - у режимі дистанційного навчання заняття відбуваються у формі онлайн-конференцій у програмі Zoom або Google Meet; сервіс організації відео-конференцій та посилання на конференцію видається на початку семестру.
- **правила поведінки на заняттях:**
 - не рекомендується займатися будь-якою діяльністю, яка прямо не стосується дисципліни або може зашкодити здоров'ю;
 - рекомендується обмежити використання засобів зв'язку під час занять;
 - забороняється будь-яким чином зривати проведення занять;
 - під час занять та консультацій рекомендується дотримуватися положень Кодексу честі КПІ ім. Ігоря Сікорського; детальніше: <https://kpi.ua/code>.
- **правила зарахування індивідуальних завдань комп'ютерних практикумів:**
 - для зарахування комп'ютерних практикумів необхідно представити викладачу:
 - друкований або електронний протокол-звіт, що відображає назву, мету, порядок виконання роботи, індивідуальне завдання відповідно до варіанту (за заліковою книжкою студента), алгоритм виконання завдання та його реалізацію у вигляді програмного коду однією з мов програмування (Python, C++, Java), контрольні приклади та розрахунки;
 - результати виконання роботи у відповідному середовищі розробки, з наглядною демонстрацією коректності роботи програми;
 - результати захисту практикуму шляхом виконання завдання іншого варіанту (іншого методу) у вигляді алгоритму.
 - зарахування комп'ютерних практикумів проходить під час проведення практичних занять, а у випадку дистанційного навчання – у режимі онлайн-конференцій у відведений розкладом час;
 - у окремих випадках допускається можливість захисту під час проведення консультацій.
- **правила призначення заохочувальних та штрафних балів:**
 - докладна інформація щодо штрафних та заохочувальних балів наведена у п. 8 «Види контролю та рейтингова система оцінювання результатів навчання»;
 - максимальна кількість заохочувальних та штрафних балів визначається відповідно до Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського, Положення про поточний, календарний та семестровий контроль результатів навчання в КПІ ім. Ігоря Сікорського, а також інших Положень та рекомендацій, які діють в КПІ ім. Ігоря Сікорського.
- **політика дедлайнів та перескладань:**
 - дедлайни зарахування всіх видів робіт визначаються термінами календарного та семестрового контролю;

- перескладання будь-яких контрольних заходів передбачено тільки за наявності документально підтверджених вагомих причин відсутності на заняттях;
- перескладання з метою підвищення рейтингових балів не передбачені.
- **політика округлення рейтингових балів:**
 - округлення рейтингового балу відбувається до цілого числа за правилами округлення.
- **політика оцінювання контрольних заходів:**
 - оцінювання контрольних заходів відбувається відповідно до Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського, Положення про поточний, календарний та семестровий контролі результатів навчання в КПІ ім. Ігоря Сікорського, а також інших Положень та рекомендацій, які діють в КПІ ім. Ігоря Сікорського;
 - нижня межа позитивного оцінювання кожного контрольного заходу має бути не менше 60% від балів, визначених для цього контрольного заходу;
 - негативний результат оцінюється в 0 балів.

Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі КПІ ім. Ігоря Сікорського. Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі КПІ ім. Ігоря Сікорського. Детальніше: <https://kpi.ua/code>.

Оскарження результатів контрольних заходів

У випадку незгоди із результатами контрольних заходів студенти можуть виконувати їх у присутності комісії, яка формується із викладачів кафедри автоматизації та систем неруйнівного контролю.

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

8.1. Рейтинг студента з дисципліни складається зі 100 балів. Семестровий рейтинг складається з балів, що студент отримує за:

- виконання п'яти індивідуальних завдань комп'ютерних практикумів (80 балів);
- модульну контрольну роботу (20 балів).

8.2. Критерії нарахування балів

8.2.1. Виконання індивідуальних завдань комп'ютерних практикумів

Виконання індивідуальних завдань шести комп'ютерних практикумів оцінюється максимум у **80 балів**. Виконання кожного індивідуального завдання оцінюється за шкалою «0,9–0,75–0,6–0».

Критерії оцінювання:

16-15 балів – робота виконана в повному обсязі, продемонстровані та пояснені отримані результати, наданий протокол-звіт, зауважень до результатів виконання роботи немає або є незначні;

14-12 балів – робота виконана в повному обсязі, наданий протокол-звіт, продемонстровані та пояснені отримані результати, є значні зауваження до результатів виконання роботи;

11-10 балів – робота виконана в повному обсязі, наданий протокол-звіт.

0 балів – робота не виконана.

8.2.2. Виконання модульної контрольної роботи (МКР)

Виконання контрольної роботи оцінюється максимум у **20 балів**. Оцінювання відбувається за шкалою «0,9–0,75–0,6–0».

Критерії оцінювання МКР:

20-18 балів – роботу виконано у повному обсязі (завдання розкрито за допомогою графічного, текстового та алгоритмічного пояснення), зауважень щодо результатів виконання роботи немає або є незначні;

17-15 балів – роботу виконано у повному обсязі (завдання розкрито за допомогою графічного, текстового та алгоритмічного пояснення), проте є значні зауваження щодо результатів виконання роботи;

14-12 балів – роботу виконано частково (завдання розкрито за допомогою графічного або текстового пояснення);

0 балів – контрольна робота не виконана або виконана невірно.

8.3. Штрафні та заохочувальні бали за:

- за участь у факультетській олімпіаді з дисципліни або споріднених дисциплін, модернізації лабораторних робіт з курсу нараховується від 1 до 5 заохочувальних балів;

- за активність під час практичних занять (розв'язування задач) нараховується 1 бал за кожен розв'язану задачу.

Сума заохочувальних балів не може перевищувати «+5».

8.4. Умови позитивної проміжної атестації (календарний контроль)

Умовою першої атестації є отримання не менше 16 балів за виконання індивідуальних завдань комп'ютерних практикумів (з урахуванням заохочувальних балів).

Умовою другої атестації є отримання не менше 32 балів за виконання індивідуальних завдань комп'ютерних практикумів (з урахуванням заохочувальних балів).

8.5. Умови допуску до семестрової атестації

Необхідною умовою допуску до семестрової атестації є зарахування трьох і більше індивідуальних завдань комп'ютерних практикумів. У випадку, якщо протягом семестру не було зараховано жодного комп'ютерного практикуму, студент не допускається до семестрової атестації.

Сума рейтингових балів, отриманих студентом протягом семестру, переводиться до підсумкової оцінки згідно з таблицею 1. Кількість індивідуальних завдань, які необхідно виконати, визначаються студентом самостійно, виходячи з пріоритетів щодо рейтингової оцінки (див. табл. 1).

Якщо сума балів менша за 60, студенти виконують залікову контрольну роботу. У такому разі бали за МКР анулюються, а сума балів за виконання комп'ютерних практикумів та залікової контрольної роботи переводиться до підсумкової оцінки згідно з таблицею 1.

Студенти, які протягом семестру отримали більше 60 рейтингових балів, але бажають підвищити семестрову рейтингову оцінку, виконують залікову контрольну роботу. У такому разі бали за МКР анулюються, а сума балів за виконання комп'ютерних практикумів та залікової контрольної роботи переводиться до підсумкової оцінки згідно з таблицею 1.

Виконання залікової роботи оцінюється максимум у 20 балів. Критерії оцінювання залікової роботи наступні:

20-18 балів – роботу виконано у повному обсязі (завдання розкрито за допомогою графічного, текстового та алгоритмічного пояснення), зауважень щодо результатів виконання роботи немає або є незначні;

17-15 балів – роботу виконано у повному обсязі (завдання розкрито за допомогою графічного, текстового та алгоритмічного пояснення), проте є значні зауваження щодо результатів виконання роботи;

14-12 балів – роботу виконано частково (завдання розкрито за допомогою графічного або текстового пояснення);

0 балів – залікова робота не виконана або виконана невірно.

Таблиця 1

Таблиця переведення рейтингових балів до оцінок

Бали	Оцінка
100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно
КР не зараховано	Не допущено

9. Додаткова інформація з дисципліни

9.1. Умови зарахування сертифікатів дистанційних та онлайн курсів

В межах опанування дисципліни «Інженерні методи аналізу даних для автоматизації та робототехніки» допускається можливість повного або часткового (за окремими темами) зарахування сертифікатів проходження дистанційних або онлайн курсів за відповідною тематикою, за умови попереднього узгодження з викладачем.

9.1.Орієнтовний перелік питань до модульної контрольної роботи та до залікової контрольної роботи

1. Що таке аналіз даних і які етапи він включає?
2. Назвіть основні типи даних у Python, які використовуються для обробки даних.
3. Як імпортувати бібліотеки NumPy та Pandas у Python?
4. Як створити масив у NumPy і які основні операції можна виконувати з масивами?
5. Яка різниця між DataFrame і Series у Pandas?
6. Поясніть, як завантажити дані з файлу CSV у Pandas DataFrame.
7. Що таке очищення даних, і які основні методи очищення доступні в Pandas?
8. Як створити простий графік у Matplotlib? Наведіть приклад.
9. Які типи графіків можна створювати за допомогою Seaborn? Наведіть приклади.
10. Як створити інтерактивний графік за допомогою Plotly?
11. Що таке лінійна регресія, і для яких задач вона використовується?
12. Опишіть основні кроки побудови моделі лінійної регресії в Python.
13. Яка різниця між класифікацією та регресією?
14. Як працює алгоритм k-найближчих сусідів?
15. Що таке крос-валідація, і чому вона важлива?
16. Опишіть принцип роботи алгоритму кластеризації K-Means.
17. Що таке метод головних компонент (PCA) і для чого він використовується?
18. Які основні етапи обробки тексту в задачах NLP?
19. Що таке токенизація, і як вона реалізується в Python?
20. Як працюють нейронні мережі, і які задачі вони вирішують?

Робочу програму навчальної дисципліни (силабус):

Складено Лисенко Ю.Ю., к.т.н., доцент, доцент кафедри автоматизації та систем неруйнівного контролю

Ухвалено кафедрою автоматизації та систем неруйнівного контролю
(протокол №5 від 13.11.2024 р.)

Погоджено методичною комісією приладобудівного факультету
(протокол №9/24 від 15.11.2024 р.)