

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЗАПОРІЗЬКА ПОЛІТЕХНІКА»

Кафедра системного аналізу та обчислювальної математики
(найменування кафедри)

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Python для аналізу даних
(назва навчальної дисципліни)

Освітня програма: Інтелектуальні технології та прийняття рішень в складних системах
(назва освітньої програми)

Спеціальність: 124 – Системний аналіз
(найменування спеціальності)

Галузь знань: 12 – Інформаційні технології
(найменування галузі знань)

Ступінь вищої освіти: бакалаврський
(назва ступеня вищої освіти)

Затверджено на засіданні кафедри

СА та ОМ

(найменування кафедри)

Протокол №18 від 16/08/2021 р.

м. Запоріжжя 2021

1. Загальна інформація	
Назва дисципліни	<i>Python для аналізу даних</i>
Рівень вищої освіти	<i>Перший (бакалаврський) рівень</i>
Викладач	<i>Рябенко Антон Євгенович, к.ф.-м.н., доцент</i>
Контактна інформація викладача	<i>Телефон кафедри 769-8247, Телефон викладача +380680689709, rjabenkoae@gmail.com</i>
Час і місце проведення навчальної дисципліни	<i>Предметна аудиторія кафедри, комп'ютерні класи згідно розкладу занять</i>
Обсяг дисципліни	<i>розподіл годин: лекції-14, лабораторні заняття-28, вид контролю- залік,</i>
Консультації	<i>Згідно з графіком консультацій</i>
2. Пререквізити і постреквізити навчальної дисципліни	
<i>Пререквізити: основи математичного аналізу, лінійна алгебра, програмування та алгоритмічні мови</i>	
3. Характеристика навчальної дисципліни	
Інтегральна компетентність: Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми системного аналізу у професійній діяльності або в процесі навчання, що передбачають застосування теоретичних положень та методів системного аналізу та інформаційних технологій і характеризуються комплексністю та невизначеністю умов Загальні компетентності: – K01. Здатність до абстрактного мислення, аналізу та синтезу; – K02. Здатність застосовувати знання у практичних ситуаціях – K04. Знання та розуміння предметної області та розуміння професійної діяльності; – K07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел; – K10. Здатність працювати автономно; – K12. Здатність працювати в команді. Фахові компетентності: – K18. Здатність формалізувати проблеми, описані природною мовою, у тому числі за допомогою математичних методів, застосовувати загальні підходи до математичного моделювання конкретних процесів; – K22. Здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, оптимізації, обробки інформації,	

інтелектуального аналізу даних.

Результати навчання:

- ПР08. Володіти сучасними методами розробки програм і програмних комплексів та прийняття оптимальних рішень щодо складу програмного забезпечення, алгоритмів процедур і операцій.
- ПР09. Вміти створювати ефективні алгоритми для обчислювальних задач системного аналізу та систем підтримки прийняття рішень.
- ПР12. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

Результати навчання:

В результаті вивчення дисципліни студент повинен

знати:

- основні типи даних та базові синтаксичні конструкції Python;
- базові принципи структурування та роботи з даними;
- основні функції бібліотек NumPy, Pandas, SciPy;
- можливості візуалізації даних за допомогою бібліотек Matplotlib, Plotly, Seaborn
- основні методи навчання з учителем та без учителя
- архітектури штучних нейронних мереж

вміти:

- працювати у середі Jupiter Notebook
- зчитувати дані різних типів та з різних джерел
- проводити маніпуляції з даними засобами бібліотек Python
- візуалізувати дані
- реалізовувати на практиці моделі навчання з учителем та без учителя, оцінювати їх ефективність
- створювати штучні нейронні мережі та навчати їх

4. Мета вивчення навчальної дисципліни

Метою вивчення курсу «Python для аналізу даних» є оволодіння навичками програмування на мові Python для подальшого використання в інших курсах, а також ознайомлення з його можливостями у розв'язуванні задач Data Science.

5. Завдання вивчення дисципліни

Завданням вивчення навчальної дисципліни є набуття компетенцій, знань, умінь та навиків із використання Python як алгоритмічної мови програмування, а також

застосування Python у задачах Data Science.

6. Зміст навчальної дисципліни

У навчальній дисципліні «Python для аналізу даних» розглядаються основи програмування на мові Python, інструменти зчитування, обробки та візуалізації даних. Курс містить огляд базових можливостей мови, а також функціоналу актуальних бібліотек для розв'язування задач Data Science.

7. План вивчення навчальної дисципліни

№ теми	Назва теми	Форми організації навчання	Кількість годин
1.	Математика і Python для аналізу даних. Основи роботи в середовищі Jupiter Notebook. Основи мови програмування Python. Оператори, синтаксичні конструкції та структури мови Python. Огляд бібліотек NumPy, matplotlib, pandas, SciPy, seaborn, scikit-learn. Створення масивів NumPy ndarray та операції з ними. Створення структур Series та DataFrame pandas. Очищення даних. Ієрархічне індексування. Завантаження даних з файлів Excel та CSV. Відображення, фільтрація та групування даних. Зведені таблиці. Застосування векторизованих функцій на масивах NumPy. Функції агрегації в одно- та багатовимірних масивах NumPy та датафреймах pandas. Зміна розмірності масивів. Математичні та логічні операції з масивами різної розмірності. Сортування масивів.	Лекції, лабораторні заняття, самостійна робота	10
2	Візуалізація даних та статистика. Візуалізація даних matplotlib та seaborn: лінійні та стовпчасті діаграми, діаграма розкиду, пузирчаста діаграма, кругова	Лекції, лабораторні заняття, самостійна робота	8

	діаграма, полярна діаграма, гістограма, таблична діаграма, льодяникова діаграма, коробковий графік. Інтерактивна візуалізація даних з plotly. Розрахунок статистичних показників в numpy та pandas. Субмодуль статистичних функцій scipy.stats. Генерація випадкових даних за різними розподілами імовірностей. Розподіл Стюдента. Розрахунок довірчого інтервалу.		
3	Методи навчання з вчителем. Лінійні моделі класифікації і регресії. Їх реалізація в sklearn.linear_model. Класи LinearRegression, Ridge, LogisticRegression. Прогнозування. Оцінка точності. Нормалізація та стандартизація даних. Функціонал якості і градієнтний спуск. Вимірювання якості моделей. Застосування метрик якості.	Лекції, лабораторні заняття, самостійна робота	8
4.	Методи навчання без учителя. Кластеризація методом k–середніх та спектральна кластеризація. Побудова дендрограми. Детекція аномалій даних за допомогою метода опорних векторів. Зниження розмірності і візуалізації багатовимірних даних (методи PCA та t-SNE).	Лекції, лабораторні заняття, самостійна робота	8
5.	Нейронні мережі. Побудова перцептрона. Простий приклад розпізнавання зображень. Огляд бібліотек Keras та TensorFlow.	Лекції, лабораторні заняття, самостійна робота	8

	Підходи до реалізації різних архітектур: згорткова нейронна мережа, рекурентна нейронна мережа, та сучасні архітектури.		
8. Самостійна робота			
<i>Необхідним елементом успішного засвоєння навчального матеріалу дисципліни є самостійна робота студентів з вітчизняною та закордонною спеціальною літературою. Самостійна робота є основним засобом оволодіння навчальним матеріалом у час, вільний від обов'язкових навчальних занять.</i> <i>Основні види самостійної роботи, запропоновані студентам:</i> - <i>вивчення лекційного матеріалу;</i> - <i>робота з рекомендованою літературою;</i> - <i>вивчення окремих тем або питань, що передбачені для самостійного опрацювання;</i> - <i>вивчення основних термінів та понять з галузі регресійного аналізу;</i> - <i>підготовка до лабораторних занять;</i> - <i>підготовка до проміжного та підсумкового контролю;</i> - <i>контрольна перевірка кожним студентом особистих знань за питаннями для самостійного поглибленого вивчення та самоконтролю;</i> - <i>робота над індивідуальним завданням.</i>			
9. Система та критерії оцінювання курсу			
<i>Проміжний контроль знань студентів включає такі заходи:</i> – <i>виконання та захист лабораторних робіт, що передбачені учбовим планом;</i> – <i>аудиторні контрольні роботи (2 за семестр), що проводяться перед кожним підсумковим контролем.</i> <i>Підсумковий контроль знань студентів включає: іспит.</i> <i>Оцінювання знань при проміжному контролі можливо таким чином:</i> - <i>за відмінний захист лабораторних робіт – 50 балів;</i> - <i>при відмінному виконанні всіх завдань підсумкової контрольної роботи (контрольної роботи для заочної форми) – 50 балів.</i>			
10. Політика курсу			
<i>Академічна доброчесність. Студенти не видають за свої результати роботи інших людей. При використанні чужих ідей і тверджень у власних роботах обов'язково посилаються на використані джерела інформації. Під час оцінювання результатів навчання не користуються недозволеними засобами, самостійно виконують навчальні завдання, завдання поточного та підсумкового контролю.</i>			

Відвідування занять. Очікується, що всі студенти відвідають усі лекції та практичні заняття курсу. За згоди декана та викладача дозволяється перейти на індивідуальний графік занять. У будь-якому випадку студенти зобов'язані дотримуватися термінів виконання усіх видів робіт, передбачених робочою програмою курсу.

Література

1. Гэددис Т. *Начинаем программировать на Python*. – 4-е вид. – СПб.: BHV, 2019. – 768 с.
2. Лутц М. *Изучаем Python. Том 1*. – 5-е вид. – М.: Диалектика, 2019. – 832 с.
3. Любанович Б. *Простой Python. Современный стиль программирования*. – 2-е вид. — СПб.: Питер, 2021. — 592 с.
4. Седер Н. *Python. Экспресс-курс*. – 3-е вид. — Переклад с англ. Е. Матвеев. — Manning Publications Co.; СПб.: Питер, 2019. — 479 с.: ил.
5. Маккинни У. *Python и анализ данных*. – М.: ДМК Пресс, 2015. — 486 с.
6. Свейгарт Э. *Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих* – М.: Вильямс, 2017. — 573 с.
7. Элбон К. *Машинное обучение с использованием Python. Сборник рецептов* – Пер. с англ. — СПб: БХВ-Петербург, 2019. — 384 с.: ил.
8. VanderPlas J. *Python Data Science Handbook*. – <https://jakevdp.github.io/PythonDataScienceHandbook/>.

Публікації з тематики курсу

1. Shyrokorad, D. V. Evolution of the Ni-Al Janus-like clusters under the impacts of low-energy Ar and Ar13 projectiles / D. V. Shyrokorad, G. V. Kornich, S. G. Buga // *Materials Today Communications*. – V. 23. (2020) 101107-12 (Изд. Elsevier) – doi: 10.1016/j.mtcomm.2020.101107.
2. Shyrokorad D.V. Formation of the core-shell structures from bimetallic Janus-like nanoclusters under low-energy Ar and Ar13 impacts: A molecular dynamics study / D. V. Shyrokorad, G. V. Kornich, S. G. Buga // *Computational Materials Science*. – 2019. – V. 159. – P. 110-119.
3. Dumin O., Plakhtii V., Pochanin G., Shyrokorad D.. Object classification using artificial neural network processing of data obtained by impulse GPR with 1 Tx+ 4Rx antenna system // *2020 IEEE Ukrainian Microwave Week (UkrMW)*. – Kharkiv, (Ukraine, September 21-25). – 2020. – P. 1140-1144.
4. Рябенко А.Е. Пинчук В.П. Кластеризация данных методом инкрементальных сфер // *Системний аналіз. Інформатика. Управління (САІУ-2013) : матеріали IV Міжнародної науково-практичної конференції (м. Запоріжжя, 13–16 березня 2013 року) / Міністерство освіти і науки, молоді та спорту України, Академія наук вищої школи України, Запорізька обласна державна адміністрація, Класичний приватний університет*. – Запоріжжя : КПУ, 2013. – С. 206-207.
5. Рябенко А.Є. Ляшенко А.Г. Створення функціонально-аналітичного модуля Python

для автоматизації роботи СУБД SQLite // Тиждень науки-2019. Факультет комп'ютерних наук і технологій. Тези доповідей науково-практичної конференції, Запоріжжя, 15–19 квітня 2019 р. [Електронний ресурс] / Редкол.: В. В. Наумик (відпов. ред.) Електрон. дані. – Запоріжжя : ЗНТУ, 2019. – С.179.