

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Основи аналітики в Python»



Ступінь освіти	Бакалавр
Галузь знань	F Інформаційні технології, всі спеціальності
Тривалість викладання	1 семестр
Заняття:	3, 4 чверть (скорочена форма навчання)
лекції	2 год./тижд.
практичні роботи	1 год./тижд.
Мова викладання	українська

Передумови для вивчення: вивчення дисципліни «Основи аналітики в Python» у встановлених відповідною робочою програмою обсягах передбачає розуміння основ програмування та структур даних.

Сторінка курсу в СДО НТУ «ДП»:

<https://do.nmu.org.ua/course/view.php?id=5692>

Консультації: за окремим розкладом, що попередньо погоджений зі здобувачами освіти.

Онлайн-консультації: MS Teams, електронна пошта

Інформація про викладачів:



Викладач:

Хом'як Тетяна Валеріївна

к.ф.-м.н., доцент, доц. каф. САУ

Посилання на профіль:

Сторінка кафедри САУ:

<https://sau.nmu.org.ua/ua/kadry/khomyak.php>

Orcid ID:

<https://orcid.org/0000-0002-6177-2827>

Scopus ID:

<https://www.scopus.com/authid/detail.uri?authorId=56997472200>

Google scholar:

<https://scholar.google.com.ua/citations?user=iOXX6uAAAAAJ&hl=ru>

1. Анотація курсу

Щоденна робота аналітика складається з шаблонних, повторюваних процедур збору, аналізу та візуалізації даних, які можна автоматизувати за допомогою мови програмування Python. Python – у топ-5 мов за популярністю у світі, що використовується аналітиками даних, програмними інженерами та розробниками для автоматизацій процесів, побудови функціоналу додатків та заглиблення у машинне навчання. Такі компанії як Google, SpaceX та Instagram використовують його для очищення даних, впровадження моделей та додатків зі штучним інтелектом та ін.

Дана дисципліна створена для тих, хто хоче здобути навички аналітика або розширити свій аналітичний інструментарій, а саме - аналізувати дані, візуалізувати дані, приймати обґрунтовані рішення у різних сферах бізнесу. Отримані знання стануть фундаментом для поглибленого вивчення аналізу даних та машинного навчання.

2. Мета та завдання навчальної дисципліни

Мета дисципліни – надання знань та вмінь щодо застосування ефективних інструментальних засобів аналізу даних та візуалізації даних.

Завдання курсу:

– опанування теоретико-понятійної бази курсу;

- розгляд інструментів для інтерактивної роботи з Python;
- проведення наукових обчислень;
- ознайомлення з пакетом для візуалізації даних;
- отримання практичних навичок з використання пакету Pandas.

3. Результати навчання

Знати, розуміти та вміти використовувати у практичній діяльності:

- розробляти програмний код в інтерактивному середовищі Google Colaboratory, Ipython.;
- проводити аналіз даних та наукові обчислення в бібліотеці NumPy;
- здійснювати візуалізацію даних з використанням бібліотеки Matplotlib;
- проводити фільтрацію даних, групування, сортування за допомогою бібліотеки Pandas;
- проводити статистичну обробку даних та здійснювати візуалізацію даних за допомогою бібліотеки Seaborn.

4. Структура курсу

Види та тематика навчальних занять	Внесок в загальну оцінку, %
ЛЕКЦІЇ	40
1. Інструменти для інтерактивної роботи з Python Інтерактивна оболонка мови Python – Ipython (поєднання можливостей інтерактивної консолі Python і командного рядка Unix). Інструмент для розробки та представлення проектів Data Science в інтерактивному вигляді - Jupyter Notebook. Безкоштовне інтерактивне середовище для роботи з кодом - Google Colaboratory.	8
2. Аналіз даних та наукові обчислення з використанням бібліотеки NumPy Інсталяція та знайомство з бібліотекою NumPy. Створення масивів в NumPy. Арифметичні операції над масивами в NumPy Індексція масива в NumPy. Агрегація в NumPy. Створення матриць в NumPy на прикладах. Арифметичні операції над матрицями в NumPy. Скалярний добуток в NumPy. Індексція матриць та агрегування матриці в NumPy. Транспонування та зміна форми матриць в NumPy. Таблиці NumPy — приклади використання таблиць. Аудіо та часові ряди в NumPy. Обробка зображень та тексту в NumPy на прикладах.	8
3. Візуалізація даних з використанням бібліотеки Matplotlib Графік вигляду $y = f(x)$, діаграма у вигляді кола, трьовимірні графіки. Гістограми, лінії рівня. Оформлення графіків: засоби задання кольору, стиль ліній на одновимірному графіку, легенда до графіку, положення легенди. Формули в нотації TeX. Стрілки на графіках та анотації. Лінії та геометричні фігури на графіку. Оформлення ліній за замовченням. Розташування графіків. Декілька графіків в одному вікні. Засіб розташування графіків за допомогою subplot2grid. Використання класу GridSpec для розташування графіків. Параметри налаштування вісей. Інтерактивні можливості. Виростання пакету wxPython. Анімовані графіки. Елементи інтерфейсу засобами пакету Matplotlib.	8
4. Основи аналітики в бібліотеці Pandas Інсталяція бібліотеки Pandas. Знайомство з Jupyter Notebook. Використання структур даних Series та DataFrame. Завантаження файла .csv в DataFrame. Відбір даних з DataFrame. Вивід всіх даних із DataFrame. Вивід частини даних із DataFrame. Відображення відповідних стовпчиків із DataFrame. Фільтрація значень в DataFrame. Агрегація та групування даних в Pandas. Методи форматування даних. Засоби об'єднання: inner, outer, left, right. Сортування	8

Види та тематика навчальних занять	Внесок в загальну оцінку, %
в Pandas.	
5.Статистична обробка даних та візуалізація даних за допомогою бібліотеки Seaborn Інсталяція бібліотеки Seaborn. Основні функції Seaborn. Структури даних, що використовуються в Seaborn. Візуалізація статистичесних залежностей, статистичних розподілів, категоріальних даних, часових рядів, регресійних моделей, матричних моделей. Налаштування елементів графіка.	8
ПРАКТИЧНІ ЗАНЯТТЯ	60
Практична робота № 1. Особливості роботи в інтерактивному середовищі - Google Colaboratory. Мета: закріпити теоретичні знання і розвинути практичні навички роботи в інтерактивному середовищі.	10
Практична робота № 2 Наукові обчислення з використанням бібліотеки NumPy. Мета: закріпити теоретичні знання і розвинути практичні навички роботи з бібліотекою NumPy.	10
Практична робота № 3 Обробка часових рядів в NumPy. Мета: закріпити теоретичні знання і розвинути практичні навички з обробки часових рядів в NumPy.	10
Практична робота № 4 Візуалізація даних з використанням бібліотеки Matplotlib. Мета: закріпити теоретичні знання і розвинути практичні навички з візуалізації даних в Matplotlib, налаштування параметрів графіків.	10
Практична робота № 5 Основи аналітики в бібліотеці Pandas. Мета: закріпити теоретичні знання і розвинути практичні навички з методів фільтрації даних в Pandas.	10
Практична робота № 6 Статистична обробка даних та візуалізація даних за допомогою бібліотеки Seaborn. Мета: закріпити теоретичні знання і розвинути практичні навички з методів обробки даних в Seaborn.	10
РАЗОМ	100

5. Технічне обладнання та/або програмне забезпечення

Технічні засоби навчання: мультимедійні та комп'ютерні пристрої.

Засоби дистанційної освіти: Moodle, MS Teams.

Пакети приладних програм: Python (безкоштовна версія), Jupyter Notebook та Google Colaboratory (безкоштовна версія).

6. Система оцінювання та вимоги

6.1. Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
75-89	добре
60-74	задовільно

6.2. Здобувач ступеня освіти «Бакалавр» може отримати підсумкову оцінку з навчальної дисципліни на підставі поточного оцінювання знань за умови, якщо набрана кількість балів з поточного тестування та самостійної роботи складатиме не менше 60 балів. Поточна успішність складається з успішності за теоретичну частину курсу (максимум – 36 балів) та оцінок за виконання практичних робіт (максимум 8 балів за кожну роботу та максимальною сумарною оцінкою за всі роботи – 64 бали). Отримані бали за теоретичну частину курсу та практичні роботи додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за поточною успішністю здобувач вищої освіти може набрати 100 балів.

Шкала оцінювання (зазначено максимально можливі бали):

Теоретична частина	Практичні роботи		Разом
	При своєчасному складанні	При несвоєчасному складанні	
40	60	40	100

6.3 Критерії оцінювання поточного та підсумкового контролю:

– підсумкове оцінювання відбувається у формі диференційованого заліку у форматі тесту, який складається з 16 завдань (15 запитань із вибором варіанту відповіді – 2 бали за правильну відповідь; 1 завдання у формі задачі – максимум 6 балів, якщо надано повністю правильну і обґрунтовану відповідь);

– поточне оцінювання практичних робіт відбувається шляхом захисту звіту з відповідної роботи (максимальний бал – 8, який формується наступним чином: 50 % – правильність і повнота викладення матеріалу в звіті, 50 % – захист індивідуальної роботи шляхом відповіді на контрольні питання).

7. Політика курсу

7.1. Політика щодо академічної доброчесності. Академічна доброчесність студентів є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). У НТУ «Дніпровська політехніка» політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка":

http://www.nmu.org.ua/ua/content/activity/us_documents/System_of_prevention_and_detection_of_plagiarism.pdf.

У разі порушення студентом академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікаційна політика. Студенти повинні мати активовану університетську (корпоративну на домені @ntu.one) пошту. Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання. Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4. Відвідування занять. Для студентів денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, відрядження, які необхідно підтверджувати документами у разі тривалої (два тижні) відсутності. Про відсутність на занятті та причини відсутності студент має повідомити викладача або особисто, або через старосту. Якщо студент захворів, ми рекомендуємо

залишатися вдома і навчатися за допомогою дистанційної платформи. Студентам, чий стан здоров'я є незадовільним і може вплинути на здоров'я інших студентів, буде пропонуватися залишити заняття (така відсутність вважатиметься пропуском з причини хвороби). Лабораторні заняття не проводяться повторно, ці оцінки неможливо отримати під час консультації. **За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватися в он-лайн формі за погодженням з керівником курсу.**

7.5. Участь в анкетуванні. Наприкінці вивчення курсу та перед початком сесії студентам буде запропоновано анонімно заповнити електронні анкети (MS Office 365), які буде розіслано на ваші університетські поштові скриньки. Заповнення анкет є важливою складовою вашої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати ваші пропозиції стосовно покращення змісту навчальної дисципліни.

8. Рекомендовані джерела інформації:

1. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. Ч.: ФОП Баликіна С.М., 2020. -180 с.

2. Основи програмування. Python. Частина 1 [Електронний ресурс]: підручник для студ. спеціальності 122 "Комп'ютерні науки", А.В. Яковенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 1,59 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2018. – 195 с.

3. Васильєв О. М. Програмування мовою Python. Тернопіль: Навчальна книга – Богдан, 2019. – 504с.

4. Копей В. Б. Мова програмування Python для інженерів і науковців : навч. посіб. / В. Б. Копей. - Івано-Франківськ : ІФНТУНГ, 2019. - 272 с.

5. Коряшкіна Л.С. Бази даних в професійних задачах аналітики [Електронний ресурс]: методичні рекомендації до виконання практичних робіт для здобувачів ступеня магістра спеціальності 124 Системний аналіз / уклад.: Л.С.Коряшкіна, Т.В. Хом'як, Д.М. Гаранжа, А.В. Малієнко ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро: НТУ «ДП», 2024. – 47 с. – Режим доступу: <https://ir.nmu.org.ua/entities/publication/f0efd83c-24ba-4ea0-9c87-dc417662fc74>

6. Онлайн аналітика. [Електронний ресурс]: методичні рекомендації до виконання індивідуальних завдань для здобувачів ступеня магістра галузі знань / А.В. Малієнко, Д.М.Гаранжа, Т.В.Хом'як, О.С. Мінєєв; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2023. – 100 с.

7. Руденко В., Жугастров О. Інформатика. Основи алгоритмізації та програмування мовою Python. Харків: Ранок, 2019. – 192 с.

8. Хабарлак К.С. Аналіз та обробка великих даних [Електронний ресурс] : конспект лекцій для здобувачів ступеня магістра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / К.С. Хабарлак, Т.В. Хом'як ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2024. – 111 с. – Режим доступу: <https://ir.nmu.org.ua/server/api/core/bitstreams/8e7bcd29-9bae-40ef-9393-cf3c8d526c99/content>

Інформаційні ресурси в Інтернеті

1. Python Підручник - <https://www.w3schools.com/python/default.asp>
2. https://pandas.pydata.org/docs/user_guide/10min.html
3. <https://www.python.org/>
4. <https://docs.python.org/3/library/sys.html>
5. <https://docs.python.org/3/library/os.html>
6. <https://jupyter.org/>
7. <https://realpython.com/pandas-dataframe/>