



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ ВСТУП ДО ТЕХНОЛОГІЇ DATASCIENCE

Галузь знань	12 Інформаційні технології
Шифр та назва спеціальності	123 Комп'ютерна інженерія
Назва освітньо-професійної програми	Комп'ютерні мережі та Інтернет
Рівень вищої освіти	Перший (бакалаврський)
Факультет	Інформаційних технологій та кібербезпеки
Кафедра	Комп'ютерної інженерії та інформаційних систем
Статус навчальної дисципліни	ОК-30 ОПП «Комп'ютерні мережі та Інтернет»
Форма навчання	Денна

Викладач

Тихонова Олена Вікторівна
elena.tykhonova@suitt.edu.ua



Старший викладач кафедри Комп'ютерної інженерії та інформаційних систем, кандидат технічних наук

Загальна інформація про дисципліну

Анотація до дисципліни	Програма вивчення навчальної дисципліни «Вступ до технології DataScience» складена відповідно до освітньо-професійної програми підготовки бакалаврів напряму (спеціальності) 123 Комп'ютерна інженерія. Дисципліна знайомить здобувачів вищої освіти із предметною галуззю науки про дані. Розглядається робочий процес обробки даних, необхідні професійні навички та якості для різних позицій у сфері data science. Здобувачі вищої освіти дізнаються про те, як мова програмування Python може допомогти в аналізі даних. В курсі розглядаються основи інтерактивної роботи з Python у блокноті Jupyter Notebook, дається необхідний мінімум синтаксичних конструкцій
-------------------------------	---

	Python для завдань обробки даних, розглядаються базові аналітичні пакети: NumPy, pandas, matplotlib, seaborn. Розглядаються питання завантаження даних різних форматів, очищення, дослідницького аналізу, візуалізації даних.
Мета дисципліни	Ознайомлення з основними концепціями в галузі науки про дані, отримання базових навичок аналізу даних з використанням Python та інтерпретації отриманих результатів.
Компетентності, формуванню яких сприяє дисципліна	ЗК-2. Здатність вчитися і оволодівати сучасними знаннями. ЗК-7. Вміння виявляти, ставити та вирішувати проблеми. СК-7. Здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема з метою підвищення їх ефективності.
Результати навчання	ПРН-3. Знати новітні технології в галузі комп'ютерної інженерії. ПРН-4. Знати та розуміти вплив технічних рішень в суспільному, економічному, соціальному і екологічному контексті. ПРН-8. Вміти системно мислити та застосовувати творчі здібності до формування нових ідей. ПРН-15. Вміти виконувати експериментальні дослідження за професійною тематикою. ПРН-20. Усвідомлювати необхідність навчання впродовж усього життя з метою поглиблення набутих та здобуття нових фахових знань, удосконалення креативного мислення.
Обсяг дисципліни	Загальний обсяг дисципліни: 4 кредити ЄКТС (120 годин). Для денної форми навчання: лекції –20 годин, практичні заняття –12 години, лабораторні заняття –12 години, самостійна робота – 76 годин.
Форма підсумкового контролю	Залік
Терміни викладання дисципліни	Дисципліна викладається у 7-му семестрі.

Програма дисципліни

Тема 1.	Основні поняття та визначення. Історичні передумови та основні етапи розвитку галузі Data Science. Тенденції, що визначають майбутнє у сфері Data Science. Посадові обов'язки, технічні навички та додаткові якості спеціалістів, затребуваних на сучасному ринку праці у сфері Data Science. Етичні міркування при роботі з даними, принципи безпеки та конфіденційності даних.
Тема 2.	Найпоширеніші інструменти, що використовують для вирішення завдань в області DataScience. Електронні таблиці (Google Sheets, Microsoft Excel), бази даних (Google Cloud, CloudSQL, Oracle, Microsoft SQL Server), інструменти візуалізації даних (Tableau, Matplotlib, Seaborn, Google Charts, InfoGram, ChartBlocks), інформаційні панелі (Tableau, LookerStudio, Microsoft PowerBI). Мови програмування R і Python, їх особливості з точки зору обробки даних. Інтерактивне середовище розробки Jupyter Notebook.

Тема 3.	Аналітика даних, мета аналізу даних. Задачі збирання, очищення, зберігання даних, дослідницького аналізу даних, побудови статистичних моделей, моделей машинного навчання, представлення результатів аналізу даних. Види даних: якісні та кількісні, номінальні та впорядковані, дискретні та безперервні. Загальні методи візуалізації даних: таблиці, кругові діаграми та стовпчасті діаграми з накопиченням, лінійні діаграми та діаграми з областями, гістограми, діаграма розсіювання, теплові карти, карти дерев.
Тема 4.	Робочий процес з обробки даних. Основні етапи проекту з обробки даних. Методологія дослідження даних CRISP-DM. PACE (Plan, Analyze, Construct, Execute) - методологія дослідження даних від компанії Google. Вхідні дані та очікувані результати окремих етапів та проекту в цілому. Важливість розробки та представлення візуалізацій під час робочого процесу з обробки даних.
Тема 5.	Додаткові бібліотеки мови Python, що використовуються для аналізу даних. Робота з багатовимірними масивами за допомогою бібліотеки NumPy. Векторизовані обчислення. Обробка табличних даних за допомогою бібліотеки Pandas. Основні структури даних в Pandas - Series та DataFrame. Візуалізація даних. Побудова графіків і діаграм з використанням бібліотек matplotlib, pandas і seaborn.
Тема 6.	Дослідницький аналіз даних (Exploratory Data Analysis, EDA). Важливість процесу EDA для розробки моделей машинного навчання. Шість основних практик EDA: виявлення (discovering), структурування (structuring), очищення (cleaning), об'єднання (joining), перевірка (validating), представлення (presenting). Засоби мови Python для реалізації практик EDA.

Список рекомендованих джерел

1. Фостер Провост, Том Фоусетт. Data science для бізнесу: як збирати, аналізувати і використовувати дані (переклад з англ.). Київ : Наш формат, 2020. 392 с.
2. D.Gielen, A. Meysman, M. Ali. Introducing Data Science: Big Data, machine learning, and more, using Python tools. Manning; First Edition, 2016. 320 p.
3. Jake VanderPlas. Python Data Science handbook: essential tools for working with data. O'Reilly Media; 1st edition (2017). 546 p.
4. Олещенко Л.М. Технології оброблення великих даних: конспект лекцій. КПІ ім. І. Сікорського, 2021 р. 227 с.
5. Peter Wentworth, Jeffrey Elkner, Allen B. Downey, Chris Meyers. Learn Python the right way: how to think like a computer scientist. Ritza, 2022. 457 p.
6. A Practical Guide To Data Visualization: Part 1.Using Python, Seaborn, and Matplotlib. Режим доступу: <https://towardsdatascience.com/practical-guide-to-data-visualization-83e375b0037>.
7. Joel Grus. Data Science from Scratch, 2nd Edition. O'Reilly Media, Inc., 2019.
8. Joe Reis, Matt Housley. Fundamentals of Data Engineering. O'Reilly Media, Inc., 2022.

Інформація про консультації

Згідно визначеного розкладу: ауд. 402 або онлайн за посиланням

<https://us04web.zoom.us/j/3185149804?pwd=TmUybHZZZYzBRK2dleUQrNVhPaG1wdz09>

Загальна схема оцінювання

Сума балів за всі види навчальної діяльності	Шкала ЄКТС	Оцінка за національною шкалою		Н а р а х у в а н н я б а л і в	Бали нараховуються таким чином: <i>Оцінювання знань здобувачів вищої освіти здійснюється за 100-бальною шкалою і становить: за поточну успішність (участь у практичних заняттях, виконання практичних завдань та контрольних робіт) – до 80 балів, за результати заліку – до 20 балів.</i>
		для іспиту	для заліку		
90-100	A	Відмінно	зараховано		
82-89	B	Добре			
74-81	C				
64-73	D				
60-63	E	Задовільно			
35-59	FX	Незадовільно з можливістю повторного складання	Не зараховано з можливістю повторного складання		
0-34	F	Незадовільно з обов'язковим повторним вивченням дисципліни	Не зараховано з обов'язковим повторним вивченням дисципліни		

Політика опанування дисципліни

Відвідування занять: відвідування здобувачами навчальних занять є обов'язковим, запізнення на заняття на 15 хвилин і більше не допускається. При проведенні занять в онлайн режимі присутність здобувача зараховується у разі включення ним камери та/або мікрофона.

Умови зарахування пропущених занять: зарахування пропущених практичних/лабораторних занять здійснюється за умови виконання та захисту відповідних завдань. До екзамену допускаються здобувачі, які виконали практичні та лабораторні завдання. Здобувач, який не з'явився на екзамен або не був допущений на момент його проведення, має право повторно його пройти у визначений викладачем термін.

Дотримання принципів академічної доброчесності: Підготовка усіх завдань, письмових робіт і т. ін., що виконуються в межах дисципліни, здійснюється здобувачем вищої освіти самостійно, на засадах академічної доброчесності. У разі порушення здобувачем принципів академічної доброчесності робота оцінюється незадовільно та має бути виконана повторно.

Інші вимоги: Загальна оцінка з дисципліни – максимум 100 балів. У випадку отримання менше 60 балів, здобувач обов'язково здійснює перескладання для ліквідації академічної заборгованості.