Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра комп'ютерних наук та інформаційних технологій (№ 302)

ЗАТВЕРДЖУЮ

Голова НМК 1
_____ Сергій НИЖНИК

Голова НМК 2
_____ Дмитро КРИЦЬКИЙ

Голова НМК 3
_____ Ганна ЛІХОНОСОВА

«30» 08 2024 р.

СИЛАБУС
ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Основи Data Science з мовою Python
(назва навчальної дисципліни)

Галузі знань: усі галузі, за якими відбувається підготовка здобувачів в університеті

Спеціальності: усі спеціальності, за якими відбувається підготовка здобувачів в університеті

Освітні програми: усі освітні програми відповідних спеціальностей, за якими відбувається підготовка здобувачів в університеті

Форма навчання: денна

Рівень вищої освіти: другий (магістерський)

Вводиться в дію з «01» вересня 2024 р.

Харків 2024 рік

Розробник: Олена ЯШИНА, доцент, к.т.н., доцент
(прізвище та ініціали, посада, науковий ступінь та вчене звання) (підпис)

Силабус розглянуто на засіданні кафедри комп'ютерних наук та інформаційних технологій (№ 302)

Протокол № 671/07 від 27 серпня 2024 р.

Завідувач кафедри д.т.н., проф. Олег ФЕДОРОВИЧ
(науковий ступінь та вчене звання) (підпис) (ініціали та прізвище)

Погоджено з представником здобувачів освіти:

студент групи 356 Артем ГРОМЕНКО
(підпис) (ініціали та прізвище)

1. Загальна інформація про викладача

Яшина Олена Сергіївна, к.т.н., доцент. З 1998 з року викладає в університеті наступні дисципліни:
- Проектування інформаційних систем;
- Управління ІТ-проектами;
- Структури даних.

Напрями наукових досліджень: проектування інформаційних систем, методи та технології обробки даних, розробка баз даних та програмного забезпечення

2. Опис навчальної дисципліни

Семестр, в якому викладається дисципліна – 1 семестр.
Обсяг дисципліни:
5 кредитов ЄКТС (150 годин), у тому числі аудиторних – 64 годин, самостійної роботи здобувачів **–** 86 годин.
Форми здобуття освіти
Денна, дистанційна, дуальна.
Дисципліна – вибіркова.
Види навчальної діяльності – лекції, практичні заняття, самостійна робота здобувача.
Види контролю – поточний, модульний та підсумковий (семестровий) контроль (іспит).
Мова викладання – українська.

3. Мета та завдання навчальної дисципліни

Мета вивчення – надання студентам знань, уміння, інформаційних технологій та інструментальних засобів для виконання досліджень з різноманітними наборами даних.
Завдання. Курс присвячений вивченню методичних прийомів та інструментальних засобів, необхідних для програмної реалізації моделей та алгоритмів перетворення, дослідження та візуалізації даних за допомогою мови Python. Вивчаються основні можливості бібліотек Pandas, NumPy, Regex, Matplotlib, Seaborn. Практичні завдання передбачають роботу з публічними наборами даних: пошук наборів даних в інтернеті, завантаження, очищення та обробка даних, а також візуалізація та репрезентація результатів дослідження.

Очикувані результати наавчання:

За результати вивчення курсу студенти повинні **знати**:
- засоби та методичні прийоми дослідження даних;
- можливості основних модулів та бібліотек мови Python для обробки, очищення та візуалізації даних;
- основні вимоги до дослідження даних.

Уміти:

- обирати інструментальні засоби для отримання, збереження та пертворення даних;
- аналізувати, структурувати та обробляти дані з зовнішніх джерел, в тому числі зашумлених;
- візуалізувати та репрезентувати результати дослідження даних;
- розробляти програмне забезпечення для дослідження даних;
- самостійно виконувати дослідження даних.

4. Зміст навчальної дисципліни

Змістовний модуль №1. Засоби та технології Data Science.

Тема 1. Вступна лекція

- *Форма занять: лекція, самостійна робота.*
- *Обсяг аудиторного навантаження: 1-2 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з підключенням до інтернет.*

Предмет вивчення і задачі дисципліни «Основи Data Science з мовою Python». Data Science як междисциплінарна наукова та практична галузь. Зв'язок з іншими дисциплінами. Інструментальні засоби. Блокноти Jupyter. Середовище Google Colab.

- *Обсяг самостійної роботи здобувачів: 2-3 години.*

Опрацювання матеріалу лекцій. Формування питань до викладача.

Тема 2. Основи синтаксису Python.

- *Форма занять: лекція,практичне заняття, самостійна робота.*
- *Обсяг аудиторного навантаження: 4-6 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Основні синтаксичні конструкції. Засоби відладки. Отримання довідкової інформації. Функції. Основи функціонального підходу.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 3. Робота з текстом у Python.

- *Обсяг аудиторного навантаження: 6-8 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Тип даних class 'str' . Методи роботи зі строками. Регулярні вирази, пакет regex. Основи Text Mining.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 4. Списки та колекції.

- *Обсяг аудиторного навантаження: 4-6 годин.*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Списки, кортежі, множини, словники. Операції з колекціями. Генерування та перетворення списків. Нарізка списків

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 5. Масиви Numpy.

- *Обсяг аудиторного навантаження: 6-8 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Бібліотека numpy. Масиви. Генерування масивів. Методи роботи з масивами. Агрегатори. Арифметичні операції (векторизація). Нарізка масивів. Фільтрація.

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Підготовка до модульного контролю.

Модульний контроль 1

- *Форма занять: виконання тестових завдань (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 2 години*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, підключення до інтернету.*
- *Обсяг самостійної роботи здобувачів – за необхідністю.*

Змістовий модуль 2. Дослідження даних

Тема 6. Бібліотека Pandas. Серії даних

- *Обсяг аудиторного навантаження: 4-6 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Бібліотека Pandas. Структура Series. Індексування серій. Операції над серіями. Запити ло серій.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 7. Структура DataFrame.

- *Обсяг аудиторного навантаження: 4-6 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Структура даних pandas DataFrame. Індексація, мультиіндекси. Запити до фреймів. Конкатенація та злиття фреймів.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 8. Робота з публічними наборами даних.

- *Обсяг аудиторного навантаження: 4-6 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Завантаження даних з публічних джерел (Укрстат, Wikipedia та ін.). Очищення даних. Виконання запитів.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 9. Основи машинного навчання.

- *Обсяг аудиторного навантаження: 6-7 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Навчання із вчителем (контрольоване) та без вчителя (неконтрольоване). Методи класифікації та кластеризації. Основи глибокого навчання.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.

Тема 10. Візуалізація даних.

- *Обсяг аудиторного навантаження: 6-7 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Бібліотека Matplotlib. Базові види діагарм: графіки, стовбчаті діаграми, гістограми, діаграми розсіювання. Оформлення діаграм

Тема 11. Складні візуалізації.

- *Обсяг аудиторного навантаження: 6-7 годин.*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер з встановленим програмним забезпеченням: Python, VS Code або інша IDE, можливе виконання завдань в середовищі Google Colab.*

Принципи Каїро. Чорнила даних. Робота з субплотами. Бібліотека Seaborn.

- *Обсяг самостійної роботи здобувачів: 4-8 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Робота з програмним забезпеченням, написання програмного коду відповідно до теми заняття. Оформлення звіту з практичного завдання та підготовка до його здачі.
Підготовка до модульного контролю.

Модульний контроль 2

- *Форма занять: виконання тестових завдань (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 2 години*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер, підключення до інтернету.*
- *Обсяг самостійної роботи здобувачів – за необхідністю.*

5. Індивідуальні завдання

Не передбачено навчальним планом

6. Методи навчання

Проведення лекцій, лабораторних занять, індивідуальні консультації (при необхідності), самостійна робота студентів за матеріалами, опублікованими кафедрою (методичні посібники).

7. Методи контролю

Поточний контроль (теоретичне опитування й розв'язання практичних завдань), модульний контроль (тестування за розділами курсу) та підсумковий (семестровий) контроль (іспит).

8. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Виконання і захист практичних завдань	0…10	3	0…30
Модульний контроль	0…15	1	0…15
Змістовний модуль 2			
Виконання і захист практичних завдань	0…10	4	0…40
Модульний контроль	0…15	1	0…15
За семестр			**0…100**

Прийнята шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка для екзамену, *курсового проекту* (роботи), практики
90-100	відмінно
75-89	добре
60-74	задовільно
01-59	незадовільно з можливістю повторного складання

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування й за наявності допуску до іспиту. Під час складання семестрового іспиту студент має можливість отримати максимум 100 балів.

Іспит проводиться у вигляді тестування. Тест складається з 30 питань закритого типу (максимальна кількість балів за відповідь на одне питання – 2 або 3 бала залежно від складності та 2 питань відкритого типу різного рівня складності (максимальна кількість балів за відповідь на одне питання, залежно від складності – 10 та 20 балів).

Критерії оцінювання роботи здобувача протягом семестру

Задовільно (60-74). Мати мінімум знань та умінь. Відпрацювати та захистити всі лабораторні роботи. Знати загальні характеристики основних методів проектування ІС. Вміти розробляти прості запити до даних мовами SQL та LINQ, автоматизувати прості CRUD операції.
Добре (75-89). Твердо знати мінімум знань, виконати усі завдання. Показати вміння виконувати та захищати всі лабораторні роботи в обумовлений викладачем строк з обґрунтуванням рішень та заходів, які запропоновано у роботах. Знати основні способи пошуку рішень при проектуванні ІС. Вміти обирати структуру та створювати програмне забезпечення у відповідності до вимог.
Відмінно (90-100). Повно знати основній та додатковий матеріал. Знати усі теми. Орієнтуватися у підручниках та посібниках. Досконально знати усі технології, які використовуються при проектуванні інформаційних систем. Вміти будувати і моделювати складні структури баз даних та засоби доступу до них.

9. Політика навчального курсу

Відпрацювання пропущених занять відбувається відповідно до розкладу консультацій, за попереднім погодженням з викладачем. Питання, що стосуються академічної доброчесності, розглядає викладач або за процедурою, визначеною у Положенні про академічну доброчесність.

10. Методичне забезпечення

1. Проектування багаторівневої архітектури інформаційних управляючих систем /О.Є. Федорович, О.С. Яшина, Л.М. Лутай. – Х.: Нац. аерокосм. ун-т ім. М.Є. Жуковського «Харк. авіац. ін-т», 2012. – 132 с.
2. Проектування інформаційних систем [Електронний ресурс] : навч. посіб. до лаб. практикуму / О. С. Яшина, О. В. Коновалова, К. О. Западня. – Харків : Нац. аерокосм. ун-т ім. М. Є. Жуковського «Харків. авіац. ін-т », 2021. – 79 с.
3. Сторінка дисципліни: https://mentor.khai.edu/course/view.php?id=5958.

11. Рекомендована література

Базова

1. Основи програмування. Python. Частина 1 [Електронний ресурс]: підручник для студ. спеціальності 122 "Комп'ютерні науки", спеціалізації "Інформаційні технології в біології та медицині" / А. В. Яковенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 1,59 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2018. – 195 с. https://ela.kpi.ua/bitstream/123456789/25111/1/Python.pdf
2. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. Національний університет «Чернігівський колегіум» імені Т.Г. Шевченка Ч.: ФОП Баликіна С.М., 2020. 180 с. http://stepanivka-school1.edukit.sumy.ua/Files/downloads/Book-Python.pdf.
3. Креневич А.П. Python у прикладах і задачах. Частина 2. Об'єктно-орієнтоване програмування. Навчальний посібник – К.: ВПЦ "Київський Університет", 2020. – 152 с. http://www.mechmat.univ.kiev.ua/wp-content/uploads/2021/01/python-u-prykladakh-i-zadachakh.-ch2-oop.pdf
4. Провост Фостер, Фоусетт Том. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані / пер. з англ. Анастасія Дудченко. — 2-ге вид. — К. : Наш формат, 2020. — 400 с. ISBN 978-617-7730-03-2 (паперове видання). ISBN 978-617-7730-04-9 (електронне

видання). https://nashformat.ua/pdf-preview/data-science-dlya-biznesu.-yak-zbyraty-analizuvaty-i-vykorystovuvaty-dani-709241

Допоміжна

1. Jake VanderPlas. Python Data Science Handbook. Released November 2016. Publisher(s): O'Reilly Media, Inc. ISBN: 9781491912058. https://jakevdp.github.io/PythonDataScienceHandbook/
2. Wes McKinney. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. 3rd Edition. Released August 2022. Publisher(s): O'Reilly Media, Inc. ISBN: 9781098104030. https://wesmckinney.com/book/

12. Інформаційні ресурси

1. Документація Python. https://docs.python.org/uk/3/index.html
2. Євген Голюк. Мапа розвитку в Data Science, або Як стати дослідником даних. https://dou.ua/lenta/articles/how-i-became-data-analyst/
3. https://medium.com
4. https://www.kaggle.com
5. https://pslcdatashop.web.cmu.edu/index.jsp
6. https://stackoverflow.com/questions/tagged/python