

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Приватний заклад вищої освіти
«ХАРКІВСЬКИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ «ШАГ»

«ЗАТВЕРДЖУЮ»

Голова Вченої ради



ХАРКІВСЬКОГО ТЕХНОЛОГІЧНОГО
УНІВЕРСИТЕТУ «ШАГ»

В. Є. Зайцев

Наказ № 001/2302-пк від 07.02.2023

РОБОЧА ПРОГРАМА

курсу підвищення кваліфікації за напрямом:

ВСТУП В ПРОГРАМУВАННЯ МОВОЮ
PYTHON ДЛЯ BIG DATA ТА DATA SCIENCE

2023 р.

Робоча програма курсу підвищення кваліфікації за напрямом: «Вступ в програмування мовою Python для Big Data та Data Science» для викладачів закладів професійно-технічної, фахової передвищої, вищої та післядипломної освіти.

Розроблено компанією ТОВ «ГРІД ДІНАМІКС УКРАЇНА»

Розробники: канд. техн. наук, доцент Метельов В.О., Бондарчук А.С.

Робочу програму курсу затверджено:

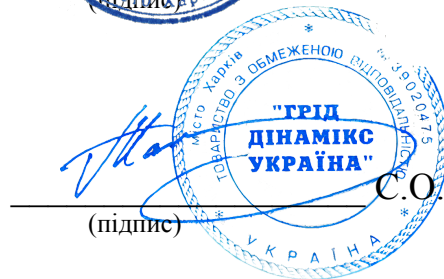
Виконавча директорка

ГС «Харківський кластер
інформаційних технологій»



О. С. Шаповал

Генеральний директор
ТОВ «ГРІД ДІНАМІКС УКРАЇНА»



С.О. Тарадай

1 МЕТА, НАПРЯМИ ТА ЗАВДАННЯ КУРСУ

Мета: курс спрямовано на формування системи знань і формування навичок в області застосування багатoproфiльної мови програмування Python (широко використовується у багатьох IT-професiях: написанi скриптiв, сервернiй розробцi, автоматизацiї, створенню дашбордiв тощо) з метою побудови ефективних алгоритмiв обробки даних та їх програмної реалiзацiї - тобто знайомить здобувача з такими специфiчними галузями IT як наука про данi та великi данi на рiвнi, достатньому для роботи в IT компанiї на посадi стажера.

Напрями пiдвищення квалiфiкацiї:

- розвиток професiйних компетентностей;
- формування професiйних компетентностей галузевого спрямування, опанування новiтнiми технологiями, ознайомлення iз сучасним станом i тенденцiями розвитку галузi.

Завдання: за результатом вивчення курсу учасники повиннi:

знати:

- Основи програмування мовою Python;
- Принципи та методи розробки програмного забезпечення використовуючи парадигму ООП та систему контролю версiй git;
- Збирати, обробляти та аналізувати статистичнi данi, науково-аналiтичнi матерiали, необхiднi для вирiшення комплексних економiчних, технiчних, соцiальних завдань використовуючи iнструменти Data Science та Big Data на початковому рiвнi (стажер, trainee).

вмiти:

- збирати, аналізувати та обробляти статистичнi данi, науково-аналiтичнi матерiали, якi необхiднi для розв'язання комплексних економiчних, технiчних, соцiальних проблем, робити на їх основi обгрунтованi висновки;

- використовувати сучаснi iнформацiйнi технологiї, методи та прийоми дослідження економiчних та соцiальних процесiв, адекватнi встановленим потребам дослідження;

- створювати та оцiнювати моделi економiчних процесiв як аналітично так i з використанням унiверсальних програмних засобiв i аналітичних платформ, що застосовуються для аналізу даних.

володiти: системою теоретичних та практичних знань та навичок щодо застосування механiзмiв iнтелектуального аналізу для вирiшення прикладних задач обробки великих даних.

вдосконалити компетенцiї:

загальнi, до яких вiдноситься:

- здатнiсть до абстрактного мислення, аналізу та синтезу;
- здатнiсть застосовувати знання у практичних ситуацiях;
- знання та розумiння предметної областi та розумiння професiйної дiяльностi;
- здатнiсть вчитися й оволодiвати сучасними знаннями;

- здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- здатність генерувати нові ідеї (креативність);
- здатність оцінювати та забезпечувати якість виконуваних робіт;

спеціальні (фахові), до яких відносять:

- здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо;
- здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем;
- здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризику;
- здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів;
- здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління;
- здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач;
- здатність реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень при розробці й експлуатації розподілених систем паралельної обробки інформації.

2 ПРОГРАМА КУРСУ

Змістовий модуль 1. Мова програмування Python.

Тема 1. Вступ. Python 3. Встановлення Git.

Тема 2. Змінні. Типи даних.

Тема 3. Типи даних II.

Тема 4. Розгалуження.

Тема 5. Петлі.

Тема 6. Функції.

Тема 7. Функції II.

Тема 8. Стил ь кодування.

Тема 9. IO.

Тема 10. Модулі.

Тема 11. ООП.

Тема 12. ООП II.

Тема 13. Ітератори.

Тема 14. Помилки.

Тема 15. Модульне тестування.

Змістовий модуль 2. Вступ до Data Science.

Тема 16. CRISP-DM. Ділове розуміння.

Тема 17. Основи статистики.

Тема 18. SQL.

Тема 19. Підготовка даних. Pandas.

Тема 20. Основи ML.

Змістовий модуль 3. Вступ до Big Data.

Тема 21. Джерела даних, 6V.

Тема 22. Пайплайни.

Тема 23. SQL.

Тема 24. PySpark.

Тема 25. Хмари.

3 СТРУКТУРА КУРСУ

Назви тем	Кількість годин				
	Усьог о	Ле кці ї	Се мін ари	Кон суль та-ц ії	Само стійн а робот а
1	2	3	4	5	6
Змістовий модуль 1. Мова програмування Python.					
Тема 1. Вступ. Python 3. Встановлення Git.	6	2	0	0	4
Тема 2. Змінні. Типи даних.	6	2	0	0	4
Тема 3. Типи даних II.	6	2	0	0	4
Тема 4. Розгалуження.	6	2	0	0	4
Тема 5. Петлі.	6	2	0	0	4
Тема 6. Функції.	6	2	0	0	4
Тема 7. Функції II.	6	2	0	0	4
Тема 8. Стиль кодування.	6	2	0	0	4
Тема 9. ІО.	6	2	0	0	4
Тема 10. Модулі.	6	2	0	0	4
Тема 11. ООП.	6	2	0	0	4
Тема 12. ООП II.	6	2	0	0	4
Тема 13. Ітератори.	6	2	0	0	4
Тема 14. Помилки.	6	2	0	0	4
Тема 15. Модульне тестування.	6	0	2	0	4
Разом за змістовим модулем 1	90	28	2	0	60

1	2	3	4	5	6
Змістовий модуль 2. Вступ до Data Science.					
Тема 16. CRISP-DM. Ділове розуміння.	6	2	0	0	4
Тема 17. Основи статистики.	6	2	0	0	4
Тема 18. SQL.	6	2	0	0	4
Тема 19. Підготовка даних. Pandas.	6	0	2	0	4
Тема 20. Основи ML.	6	2	0	0	4
Разом за змістовим модулем 2	30	8	2	0	20
Змістовий модуль 3. Вступ до Big Data.					
Тема 21. Джерела даних, 6V.	6	2	0	0	4
Тема 22. Пайплайни.	6	2	0	0	4
Тема 23. SQL.	6	0	2	0	4
Тема 24. PySpark.	6	2	0	0	4
Тема 25. Хмари.	6	2	0	0	4
Разом за змістовим модулем 3	30	8	2	0	20
Випускний проєкт	30	2	0	2	26
Усього годин	180	46	6	2	126
Усього за курс, у кредитах ЄКТС	6				

4 МЕТОДИ НАВЧАННЯ

1. Методи організації та здійснення навчально-пізнавальної діяльності за допомогою слайд-лекцій, пояснень на багатьох практичних прикладах.
2. Методи стимулювання й мотивації навчально-пізнавальної діяльності у виконанні власних проєктів з практичної реалізації завдань курсу.
3. Методи контролю (самоконтролю, корекції (самокорекції)), за ефективністю навчально-пізнавальної діяльності, ці методи спрямовані на самостійну, творчу пізнавальну діяльність, особливо при створенні власних проєктів.

5 МЕТОДИ КОНТРОЛЮ

Практичний контроль – передбачає виявлення вмінь і навичок, що набуті під час практичної діяльності (семінарських занять, роботи над власним проектом). Така перевірка дає змогу виявити, на якому рівні було засвоєно теоретичні основи курсу.

6 МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ТА РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. С. Дєви, М. Арно, А. Мохамед. Основи Data Science и Big Data. ВХВ, 2017, 336 с.
2. Черняк О.І., Захарченко П.В. Інтелектуальний аналіз даних: Підручник, К., 2014, 599 с.
3. А. О. Олійник, С. О. Субботін, О. О. Олійник. Інтелектуальний аналіз даних : навчальний посібник, Запоріжжя : ЗНТУ, 2012, 278 с.

7 ІНФОРМАЦІЙНІ РЕСУРСИ

1. What Is the Python Global Interpreter Lock (GIL)? (<https://realpython.com/python-gil/>).
2. Memory Management And Garbage Collection In Python (<https://towardsdatascience.com/memory-managementand-garbage-collection-in-python-c1cb51d1612c>).
3. How to Use Generators and yield in Python (<https://realpython.com/introduction-to-python-generators/>).
4. Hadoop - MapReduce (https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm).
5. Variables in Python (<https://realpython.com/python-variables/>).
6. Sets in Python (<https://realpython.com/python-sets>).
7. PEP 8 – Style Guide for Python Code (<https://www.python.org/dev/peps/pep-0008/>).
8. PEP 257 – Docstring Conventions (<https://www.python.org/dev/peps/pep-0257/>).
9. Python Function Tutorial – Type of Functions in Python(With Example) (<https://data-flair.training/blogs/python-function/>).
10. Python monkey-patching like a boss (<https://medium.com/@chipiga86/python-monkey-patching-like-a-boss-87d7ddb8098e>).

11. Primer on Python Decorators
(<https://realpython.com/primer-on-python-decorators/>).
12. Python Classes and Objects
(<https://www.geeksforgeeks.org/python-classes-and-objects/>).
13. RDDs vs DataFrames vs DataSets: The Three Data Structures of Spark
(<https://www.wisewithdata.com/2020/05/rdds-vs-dataframes-vs-datasets-the-three-data-structures-of-spark/>).
14. PySpark Documentation
(<https://spark.apache.org/docs/latest/api/python/index.html>).
15. How to Import and Export Datasets in Google Colab
(<https://towardsdatascience.com/google-colab-import-and-export-datasets-eccf801e2971>).