

Institute of Information and Communication Technology (IICT)



Shahjalal University of Science and Technology, Sylhet

Course Title: Python Programming and Basic Data Science

Course Description: The objective is to train the students basic skills of Python Programming in the context of Data Science. In this course they will learn how to handle with different kind of data and apply appropriate model for analysis or predictive modeling. The course is designed for participants with no or very beginners level programming skills. They will also receive introductory knowledge about various theoretical models of data science.

Course Goals: The participants will be able --

- 1. to describe, analyze, and code various forms real-world projects relevant to data science
- 2. to identify problems and challenges and reflect on potential solutions
- 3. to independently and together with others plan and work in data science projects

Teaching methods: The class will primarily follow a workshop format, incorporating Power point presentations, reflections, and discussions. Various learning activities will be facilitated, including individual tasks (preparatory work), group engagement, and participation in smaller workgroups. Participants will receive course materials and information in advance.

Course Plan

Lecture 1

- Installing Jupyter(/Spider/Pycharm) Notebook Server
- Running Jupyter Notebook Server
- Common Jupyter Notebook Commands
- Jupyter Notebook Components
- Jupyter Notebook Dashboard
- Jupyter Notebook User Interface
- Creating a new notebook

Lecture 2

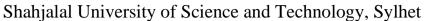
- Shell or Notebook?
- Help and Documentation in IPython
- Keyboard Shortcuts in the IPython Shell
- IPython Magic Commands
- Input and Output History

Lecture 3

- IPython and Shell Commands
- Shell-Related Magic Commands
- Errors and Debugging
- Profiling and Timing Code



Institute of Information and Communication Technology (IICT)





Lecture 4

- What is Python
- Python Expressions
- Python Statements
- Python Comments
- Python Data Types

Lecture 5

- Casting Data Types
- Python Variables
- Python List
- Python Tuple
- Python Dictionaries
- Python Operators

Lecture 6

- Python Conditional Statements
- Python Loops
- Python Functions

Lecture 7

- Understanding Data Types in Python
- The Basics of NumPy Arrays
- Computation on NumPy Arrays: Universal Functions
- Aggregations: Min, Max, and Everything in Between

Lecture 8

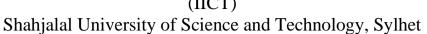
- Computation on Arrays: Broadcasting
- Comparisons, Masks, and Boolean Logic
- Fancy Indexing
- Sorting Arrays
- Structured Data: NumPy's Structured Arrays

Lecture 6

- Installing and Using Pandas
- Introducing Pandas Objects
- Data Indexing and Selection
- Operating on Data in Pandas



Institute of Information and Communication Technology (IICT)





Lecture 10

- Handling Missing Data
- Trade-Offs in Missing Data Conventions
- Missing Data in Pandas
- Operating on Null Values
- Hierarchical Indexing

Lecture 11

- Combining Datasets: Concat and Append
- Combining Datasets: Merge and Join
- Aggregation and Grouping
- Vectorized String Operations
- High-Performance Pandas: eval() and query()

•

Lecture 12

- General Matplotlib Tips
- Two Interfaces for the Price of On
- Simple Line Plots
- Simple Scatter Plots
- Visualizing Errors

Lecture 13

- Density and Contour Plots
- Histograms, Binnings, and Density
- Customizing Plot Legends
- Customizing Colorbars
- Multiple Subplots

Lecture 14

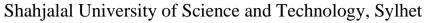
- Text and Annotation
- Customizing Ticks
- Customizing Matplotlib: Configurations and Stylesheets

Lecture 15

- Three-Dimensional Plotting in Matplotlib
- Geographic Data with Basemap
- Visualization with Seaborn



Institute of Information and Communication Technology (IICT)





Lecture 16

- What is data science?
- Impact of data science
- Data science life cycle
- Terminology
- Kaggle Data Sets

Lecture 17

- What is data cleaning
- Basic data cleaning process

Lecture 18

- What Is Machine Learning?
- Introducing Scikit-Learn

Lecture 16

• Hyperparameters and Model Validation

Lecture 20

- Feature Engineering
- Naive Bayes Classification
- Linear Regression

Lecture 21

• Support Vector Machines

Lecture 22

- Decision Trees and Random Forests
- Principal Component Analysis

Lecture 23

- In-Depth: Manifold Learning
- In Depth: k-Means Clustering
- In Depth: Gaussian Mixture Models
- In-Depth: Kernel Density Estimation

Lecture 24-26

- Project on application of Data Science
- Application: A Face Detection Pipeline