

Fourth Industrial Revolution (4IR) Summer School

Data Preparation

Period	1 Week (20 Hours) 7:30 am to 11:30 pm
Location	College of Petroleum Engineering and Geoscience
Instructor	Dr. Hamdi Al-Jamimi aljamimi@kfupm.edu.sa
Office Hour	11:30 pm – 12:00 pm (or by appointment)

Pre-requisites:

Module 1: Python Programming

Description:

It is undeniable that a significant chunk of the data scientist's time and effort is spent in collecting, cleaning and preparing the data for analysis because datasets come in various sizes and are different in nature. It is extremely important for a data scientist to reshape and refine the datasets into usable datasets, which can be leveraged for analytics. In this module, the participant gains an understanding of the core concepts of data preparation illustrated using the Python language. To start dealing with real data, the participant learns loading the data from different sources (including files, HTML, XML and JSON), to prepare and assemble the data in the form suitable for NumPy arrays and Pandas data structures. In addition to exploring tools for detecting and handling missing data, outliers, and duplicated data. The participant learns, also, how to use different methods for data preprocessing, manipulation, transformation, aggregation and grouping.

Objectives:

Participants should be able to:

- Understand the goal of data preparation and its importance.
- Understand the NumPy and Pandas data structure.
- Access and Load data from different sources.
- Apply the statistical analysis methods on data.
- Apply the normalization and scaling methods on data.
- Detect, filter, and handle the missing data and outliers.
- Understand how to Combine, join, and rearrange data.
- Understand transforming, aggregating and grouping data.

Topics:

- Introduction to Data Analytics
- NumPy and Pandas Data Structures
- Data Loading
- Data Exploring and Statistical Analysis
- Data Preprocessing
- Data Cleaning
- Data Manipulation
- Data Transformation
- Data Aggregation and Grouping

Means:

- Lectures
- In-class examples
- Hands-On problem solving

Required:

- Personal laptop for Hands-On sessions
- Google account to have access to Colab environment
- Python programming