

Data Science Foundation for Programmers

Big Data Workshop

Akram Alkouz

Princess Sumaya University for Technology

Computer Science Department

October 8, 2017

Abstract

“Data Science foundation for Programmers” is a one day workshop that introduces Programmers to developing Data Science applications. Attending this workshop helps technology enthusiasts to build their career in this technology. The hands-on workshop uses the Python Programming language to introduce machine learning algorithms.

Keywords: Data Science, Machine Learning, Data Engineering, Visualization

Description

This specially tailored workshop is to familiarize people with Data Exploration, Visualization and Feature Engineering. Participants will be required to spend 4-5 hours in pre-requisite learning to make the best out of the workshop. In particular, it will be shown:

- What is Data Science
- The Python Programming Language
- Exploratory Data analysis
- Data Visualization
- Statistical Programming
- Machine Learning

- Python in the Big Data world – ex Apache Spark
- Deep learning
- Python and R

Requirements for the participants

To be able to participate in the hands-on sessions during the workshop you will need a PC with installed Anaconda and Jupyter Notebook. You will also be required to download a number of Python libraries from the workshop git Hub, the full list will be announced two weeks before the workshop.

Provided materials

The presentations, all the instructions and code will be publicly available on a dedicated Github repository. The repository will be announced two weeks before the workshop. A Slack project will be created for the workshop. You are invited to contribute, open issues and ask questions there after the workshop. The final materials will be published after the workshop day.

Timetable

- **9:00-10:00** Introduction: Data Science and Python programming language
- **10:00-11:00** Hands-on: Configuring Jupyter Notebook and doing some ETL
- **11:00-11:15** Coffee Break
- **11:15-11:45** Exploratory Data Analysis and Visualization
- **11:45-12:15** Introduction to Statistical Programming
- **12:15-13:00** Hands-on:
- **13:00-14:00** Lunch
- **14:00-15:00** Machine Learning Algorithms
- **15:00-16:00** Hands-on: Doing some data science in Big Data environments
- **16:00-16:30** Python and R

Outcomes

By the end of this workshop attendees will be able to:

- Install and configure a Data Science Environment
- Understand the data

- Prepare the Data
- Build a prediction model
- Evaluate his prediction model
- Deploy his model on Big Data Platform

About Presenter

Akram Alkouz is an assistant professor at Princess Sumaya University for Technology. He has a Ph.D. in Information Retrieval and Machine Learning from Technical University of Berlin.

Akram Alkouz has been part of Arabic Text Mining and Analysis (ATMA) research group for the last four years contributing to the projects related to Arabic semantic graph, building Arabic resources, benchmarking and education. For the last two years, Akram has been participating Jordan Open Source Association (JOSA), he contributed to the capacity building and awareness spreading of the Data Science and Big Data technologies in Jordan.