

Workshop:

Getting the Most Out of Your Data

| Who is this workshop for?

- Managers who consider introducing machine learning to their company's business
- Tech executives (CTO, CIO, etc.)
- Tech leads

This workshop can be delivered either on-premise or online, by arrangement.

As someone who supervises a data science-based project, you really want to know the basics of getting the most out of your data. It may seem not so challenging to come up with a sensible model when your data is perfect – but in real life it almost never happens.

99% of the time real data has missing values, noise, outliers, excessive information, and all those things that make it harder to use.

For that reason, data scientists consider data preprocessing is the most time consuming – and therefore expensive – part of a DS-based project. Properly processed, well laid-out data and efficient, domain-specific features are key success factors of the project.

At this workshop, you will learn some simple but effective techniques of dealing with realistic datasets. Also, it will be demonstrated how data processing techniques can affect the performance of different models.

To run the code, we will employ Microsoft Azure Notebooks free service, available from any browser, that allows to develop and run Jupyter Notebooks in the cloud. Data for the workshop would be downloaded from a datastore on Microsoft Azure Cloud.



Why learn with WaveAccess?

WaveAccess has been doing Machine Learning since 2014. Our projects have been awarded Microsoft Partner of the Year in Artificial Intelligence (2018) and in Business Analytics (2017).

With our team of data scientists, mathematicians, algorithm engineers, and developers, we use technical expertise to increase business efficiencies, optimize slow or unreliable systems, and bring ambitious ideas to life.



2019 Partner of the Year
Media & Communications Award

2018 Partner of the Year
Artificial Intelligence Award

2017 Partner of the Year
Business Analytics Award

Gold Application Development
Silver DevOps
Silver Cloud Platform
Silver Datacenter

Audience & Requirements

- Basic knowledge of Python and Machine Learning
- Microsoft ID to gain access to Microsoft Azure Notebooks

If you do not have any coding experience, you are still welcome to join us to get a sense of how much effort is invested to develop a machine learning project. Also, all the solutions will be provided at the end.

Content & Process

- 1 Preliminary Data Analysis and Processing:** data cleaning, dealing with errors, type conversion
- 2 Exploratory Data Analysis:** finding correlations, relationships, and distribution of variables
- 3 Feature Engineering:** getting rid of the least useful features and building new ones
- 4 Machine Learning model training and evaluation:** training different machine learning models and seeing how data processing and feature engineering affects the accuracy and model training time.

Deliverables

Having completed this workshop, you will understand:

- the challenges your machine learning team faces
- why data preprocessing is so challenging and important
- how you can help them achieve the best results.