

Industrial Data Science Workshop

19 - 21 September 2023

Speakers



Dr Francisco
Navarro

IFF



Dr Mattia
Vallerio

Solvay



Dr. Carlos
Perez

Solvay

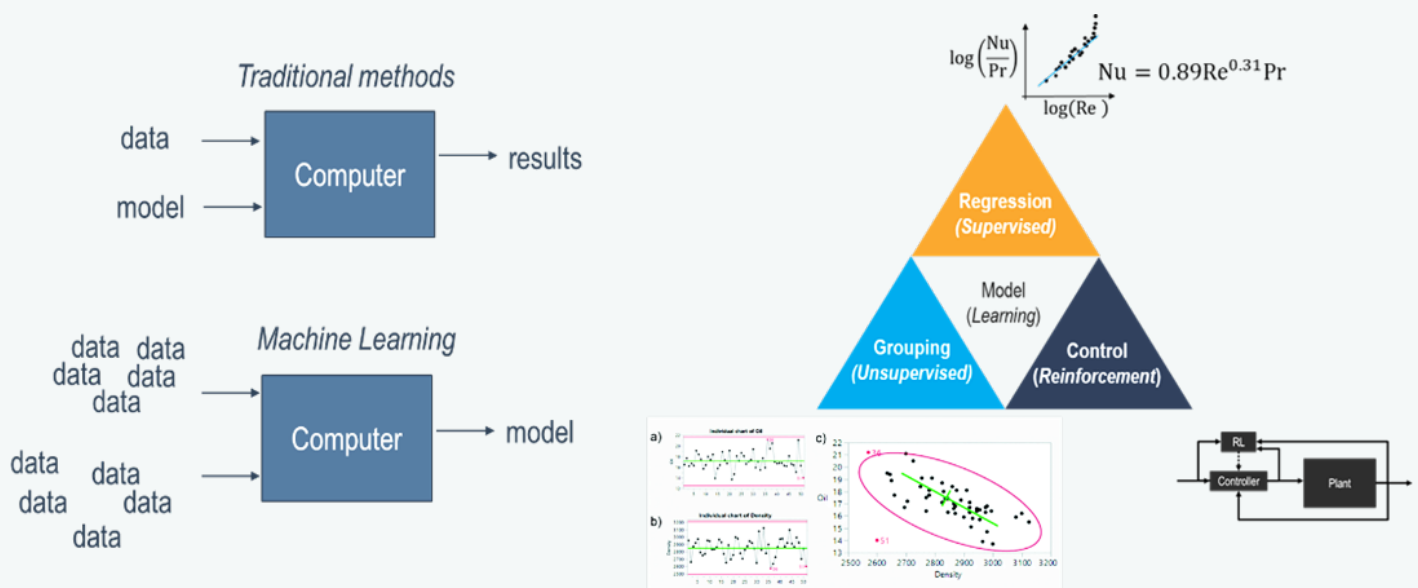
Course participants will be introduced to process data challenges and how to solve them with data science. The syllabus is geared towards general Machine Learning concepts for regression models (supervised learning) and anomaly detection (unsupervised learning).

The focus, however, will be around the application of industrial data science techniques to continuous and batch process data, so a subject domain expert (e.g., process engineers) can perform the following data-driven analysis independently:

- 1) Extract and transform process data
- 2) Quantify variability and identify relevant process changes
- 3) Quickly find potential causes to improve processes

At the end of each day, there will be sessions to perform individual hands-on exercises allowing participants to use their own datasets in their own laptops (in Excel, CSVs, or from AspenTech IP.21 and Osisoft PI systems).

Basic knowledge of statistics (e.g., six sigma training) can be helpful to fully understand the concepts of this course. A programming background (Python, MATLAB...), is not required.



Day 1 – Industrial data science

- Distillation tower (full example)
- Industrial databases (tags, historians, and automation pyramid)
- Contextual data (asset hierarchies, batch events)
- Quality and tabular data (LIMS, ERP)
- Data democratization and software alternatives
- Hands-on session (connect to databases with Excel, ODBC, and RestAPIs).

Day 2 – Monitoring assets

- Batch dryer example
- Defining KPIs for continuous and batch processes (feature engineering)
- Tracking variability (visual analytics, statistical process control, robust statistics)
- Batch data alignment (e.g., time warping)
- Machine learning for anomaly detection (KNN, PCA, Autoencoders)
- Identifying plant changes in the Tennessee Eastman Process
- Hands-on session (Bring your own data!)

Day 3 – Troubleshooting processes

- Problem definition
- Screening process variables (bootstrap forest, decision trees, and boosted trees)
- Improving processes (sensitivity analysis, explainable AI with SHAP)
- Modeling processes (missing data, Lasso regression, and neural networks)
- Industrial applications (inferential sensors and digital twins)
- Hands-on session (Bring your own data!)

Practical information

Venue:

The Sargent Centre for Process Systems Engineering
Imperial College London
Roderic Hill Building
South Kensington Campus
London SW7 2BB

Closest Underground Stations are South Kensington or Gloucester Road.



Affordable summer accommodation walking distance from the Roderic Hill Building, can be enquired about via this [link](#). The accommodation to enquire about is Prince's Gardens or Beit Hall.

Industrial price: £1050

Academic price: £250

Sargent Centre Academic price: £200

Industrial Consortium Partners (One place per company only): Free

To register, please follow this [link](#) .