

On line Basic Course:

# Data Science applied to Chemical/Energetic Industry

## About this course

The demand for Data Science experts in industry is rapidly growing. This course will introduce you to the basics of R programming and its role in industrial processes decision. Engineers face problems in the productive sector in their professional lives against which they must address solutions. In that context, data analysis becomes a tool to generate arguments when testing hypotheses and making decisions. In this course, you will learn how to visualise and analyse data to improve an industrial process by using real chemical/energetic processes dataset and R programming language as a tool. This language offers an environment oriented to statistical computing and is a good starting point for get into data science.

## What you'll learn

R fundamentals: vectors, data-frames, functions, Data management...  
Data visualization, R plots, How to use ggplot2 to create custom plots  
Tidyverse, Regression, Lineal models and GAMLSS  
Real chemical/energetic dataset analysis  
Practical examples: Equations of state, Spanish Electrical system, prediction of thermodynamic properties, : Fault detection in a chemical plant, etc.



1 ECTS for students from G9 universities

	Length:	6 weeks
		From 30 <sup>th</sup> october to 10 <sup>th</sup> december
	Level:	Introductory
	Language:	English
	Video transcript:	English/Spanish
	Methodology:	Self-pace on your time
	Price:	150 €

Enroll from 15<sup>th</sup> to 29<sup>th</sup> october, [click here](#)

# Data Science applied to Chemical/Energetic Industry

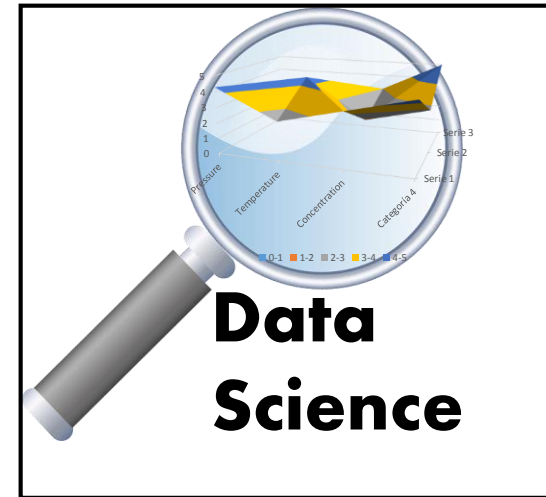
Instructor:



**Dr. Jaime Mosquera**

Chemical Engineer

Adjunct Professor (Design of Experiments) at Universidad de Antioquia, Statistical Advisor



Program:

## UNIT 0. ¿What is *data science*?

- Definition of *data science*.
- Difference between data science (DS), machine learning (ML) and artificial intelligence (AI).

## UNIT 1. R Foundations

- Data structures
  - Vectors.
  - Matrices. Arrays

- Data frames.
- Lists.
- Data tables
- Functions and operators
- [Application example 1.1: Functions and data structures: Equations of State.](#)
- Data preprocessing
  - Data import.
  - Missing data.
  - Data partition.
- Data frames manipulation
  - Tidyverse.
  - Visualizations: basic graphics in R and ggplot packages.

## Program (cont.):

- Data exploration with R
  - Descriptive Statistics
  - [Application example 1.2: Exploration of time series of Spanish Electric System](#)

### UNIT 2. Supervised Learning

- Regression
  - Linear Regression.
  - Regression Trees.
  - Non-linear regression.
  - [Application example 2.1: Evaluating results of an experiment.](#)
  - [Application example 2.2: Prediction of thermodynamic properties.](#)
- Classification
  - Logistic Regression.
  - Support Vector Machines (SVM).
  - Special topics in modeling (GAMLSS)
  - [Application example 2.3: Red wine classification according to its quality.](#)

### UNIT 3. Unsupervised Learning

- Dimensionality reduction
  - Principal Component Analysis (PCA).
  - Discriminant Analysis.
  - [Application example 3.1: Fault detection in a chemical plant](#)

