DATA SCIENCETECH INSTITUTE

***MSc in Applied Data Science & Big Data***

# Advanced Statistical Analysis and Machine Learning
# *(MSc in Applied Data Science & Big Data - DSBD2-002)*

## Volume of classes hours: 50 hrs *(± same personal work expected)*

## Course summary:

This course focuses the students' knowledge on two main areas: the CART algorithm and random forests in the context of Big Data and being used for a Map/Reduce system (first week), feature selection and engineering, covering techniques such as bagging, boosting, cross-validation and concluding with model competition and selection and an overview of various faulty implementations of techniques in "point-and-click" tools.

**Topics to be covered are:**

1. The CART algorithm
2. Random Forest (RF)
3. CART & RF for Map/Reduce: the mathematical challenges
4. Feature selection using CART & RF
5. Feature engineering using CART & RF
6. Models competition and selection

## Course objectives:

- to mathematically understand leading statistical techniques.
- to adapt and use these techniques with Map/Reduce technologies
- to be able to put multiple models and tools in competition and select the most mathematically valid one.

## Course mini-projects description:

Students will be working on an industry-driven project coming from the research laboratory contracts of the professor in charge of the class.
A Non-Disclosure Agreement may be required and would be given to the students to sign.

*According to DSTI Scientific Advisory Board policy for ever-evolving programmes, this syllabus may be subject to adaptations and changes when the class will be delivered by the selected Professor(s).*

## Theoretical background used:

Foundation of Statistical Analysis using R class

## Technologies used:

A complete specification list for the machine configuration may be specified later, but at minimum, students should have Chrome or Mozilla Firefox installed: the R language, R 3.1.1 or later, RStudio and the specific libraries will be loaded and installed by the students during the course *(instructions will be provided).*

*According to DSTI Scientific Advisory Board policy for ever-evolving programmes, this syllabus may be subject to adaptations and changes when the class will be delivered by the selected Professor(s).*