DATA SCIENCETECH INSTITUTE

Applied MSc in Data Engineering

"Statistical analysis of massive data

Pr. Charles BOUVEYRON"

Volume of classes hours: 24 hrs *(± some personal work expected)*

Course summary: The continuous increase of measurement capacities and the development of new uses of data (open data, social networks, Twitter ...) make it more and more necessary to have statistical tools capable of summarizing and extract the information contained in the data. However, the very nature of modern data (large size, mass of data) does not allow the use of most conventional statistical methods (tests, regression, classification). Indeed, these methods are not adapted to these specific conditions of application and suffer in particular from the plague of the dimension. The latest statistical tools for analyzing modern data and their implementation in realistic situations using the R software will be presented.

Course Objectives:
- Learn how to synthesize and extract information from masses of data or large data
- Master modern statistical tools for analysis and decision-making
- Know how to implement the recent techniques with the software R

Theoretical background used: Knowledge of basics in inferential statistics and data analysis. Knowledge of R software.

Technologies Used: The R software + packages

Course evaluation: One project