DATA SCIENCETECH INSTITUTE

MSc in Applied Data Science & Big Data

Survival Analysis

(Machine and Deep Learning Major - DSBD2-002)

Volume of classes hours: 25 hrs *(± same personal work expected)*

Course summary:

Analysis of survival data using parametric, nonparametric and semi-parametric methods.

Topics to be covered are:

1. Probabilistic description of Survival data, and parametric statistical methods
2. Nonparametric statistical methods for 1 and 2 groups
3. Semi-parametric regression, the Cox model: definition, interpretation, selection and diagnostics.
4. Applications to big data: penalized Cox regression and biomarkers validation

Course objectives:

- to understand the terminology describing survival data
- to be able to statistically analyze and model right-censored time-to-event data
- to apply these modeling techniques in high-throughput biomedical settings

Course mini-projects description:

During the class, there will be few applications to the analysis of real data.

Theoretical background used:

Classical Statistical Inference, data manipulation and visualization in R, basics of machine learning.

Technologies used:

A complete specification list for the machine configuration may be specified later, but at minimum, students should have installed: the R language, version 3.3.2 or later, RStudio or any other IDE the student is comfortable with, the R packages: asaur, ggplot2, glmnet, plyr, survivalROC. Additional R packages might be specified at the beginning of the class.