**GROUP** Pietro Ferrazzi, Marigarita Hernandez

**TITLE** Analysis and forecasts of heart attacks (TO BE CHANGED)

**DATA**

The data set has been downloaded at <https://archive.ics.uci.edu/ml/index.php>. It contains **299 observations** of **13 variables**. Each statistic unit is one patient. What follows is a brief description of the variables:

* TIME: time under observation for the patients, i.e., follow-up time during which the patient was constantly monitored. If the event occurred after that time, there is not track of that in the data;
* Event: it is set to 1 is the heart attack occurred, 0 otherwise;
* Gender, Smoking, Diabetes, BP and Anaemia are **dichotomous variables**;
* Age, Ejection.Fraction, Sodium, Creatinine, Pletelets and CPK are **continuous variables**.

**WHAT TO BE DONE**

* Classic explorative analysis to identify important variables and possible correlations in order to estimate the survival ratio.
* Explorative analysis to identify which variables can be useful to provide forecasts about the variable *Event* treated as a *class.*
* Modelling throw ML techniques to perform classification of the data.

**WHY**

Integrate classical statistical analysis with ML techniques to see if they can perform well in an *adverse* scenario: few variables, few observations, and the fact that the data are of a survival type and then should be treated following strict procedures. In addition, we are really interested in data coming from the sanitary field.

**PREVIOUS AVAILABLE WORK**

Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA (2017) *Survival analysis of heart failure patients: A case study*. PLoS ONE 12(7): e0181001. https://doi.org/10.1371/journal.pone.0181001

Chicco and Jurman BMC Medical Informatics and Decision Making (2020), *Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone*, 20:16 https://doi.org/10.1186/s12911-020-1023-5