

Data Mining in R

ASSIGNMENT #4

The goal of this assignment is to test your knowledge on **model comparison and selection** using R and the package **performanceEstimation**.

Your starting point is the task and data you have used in assignment #3. Your goal is to provide justified advice on which model you recommend for this task. With this goal in mind you should produce a R markdown dynamic report where you describe the steps you have taken in terms of model comparison and selection to reach your recommendation.

Submission instructions: You should deliver your assignment as ZIP file containing a folder where you put : (i) a R markdown file (*.Rmd) that I can execute in my computer through R; (ii) the report produced from this markdown file - you should select HTML as the target report format; and (iii) whatever files your report requires to be executed (for instance if your report reads the data from a file, you should also include that file). The ZIP file and the folder contained in the ZIP file **must** be named after your first and last names. For instance, John Smith would create in his computer a folder name SmithJohn containing his submission files, then would create a ZIP file named SmithJohn.ZIP with the contents of this folder, and submit this file. As a general rule you should make sure that if I unzip your submitted file and open your R markdown file in Rstudio, I will be able to produce your report by simply clicking the Knit button at Rstudio interface. A frequent source of errors (on my side) is that you submit a report that uses absolute folder paths of your computer (for instance when loading a data set) that will not work in my computer – avoid this, by using relative paths or simply by putting everything you need for your report on the same folder where the report is, and thus avoid paths.

June 2017
Luis Torgo