Ricardo Aler Mur

- The aim of this text is to introduce the Python notebooks that the student must carry out on his own, in order to learn Python, Scikit-learn, and Pyspark.
- *Python tutorial* is a practical introduction to Python and Jupyter notebooks. The material for this tutorial includes its own introduction.
- The notebook *introDecisionTrees.ipynb* introduces Scikit-learn, which is the standard Machine Learning library for Python, and also introduces Decision Trees, one of the Machine Learning methods explained in the theoretical lectures.

- The notebook *decisionTreesTrainTest.ipynb* explains how to evaluate decision tree models by means of a test set (holdout partition).
- By going through notebook *decisionTreesHyperparameters.ipynb*, the student will learn how to tune decision tree hyperparameters by means of grid search.
- Notebook *attribute_selection_reduced.ipynb* introduces how to do attribute selection in Scikit-learn.
- Notebook *mispark-kmeansComplete.ipynb* introduces Spark, and gives one programming example: the k-means method programmed within the Spark framework.