

Notes:

The Weather class requires a key from wunderground.com, the key allows a limited number of queries per day and limited rate per minute.

To run properly the Hadoop streaming scripts additional configurations could be needed depending by different factors (inputs file, RAM available, hd space, cores available , etc...)

The classes assume that the field name will be the ones choosen by safecast and wunderground, if different datasets are used the column names of the dataframes should be changed accordingly.

Python Version: 2.7

lpython version: 3.2.1

Versions of the libraries:

- pandas 0.16.2
- numpy1.9.2
- matplotlib 1.4.3
- sklearn 0.16.1
- urllib2 2.7
- json 2.0.9

Hadoop version: 2.6.0.2.2.4.2-2

Pig version: 2.1.0.1470

Andrea Mattera