

## **CREATING NEW TRANSFORMERS FOR SCIKIT-LEARN**

- You are going to create a new transformer for scikit-learn, similar to the *SimpleImputerQuartile()* developed in the theory slides.
- Create a matrix  $X$  with a few instances with some nan values in it. Create also a response variable vector  $y$ . You can put any values in both  $X$  and  $y$ .
- Following the structure explained for *SimpleImputerQuartile*, define a new transformer that replaces a nan value at location  $X[i,j]$  by a random value sampled from a normal distribution with mean  $m_j$  and standard deviation  $s_j$ . You can use the *numpy.random.normal* for sampling random values from a gaussian distribution. Check the *numpy.random.normal* documentation.
- Show that it works for your data matrix (executing the `.fit` and the `.transform` methods).
- Show that it works in a pipeline with knn (executing the `.fit` and the `.predict` methods of the pipeline).
- To hand in:
  - The code and a report showing that your program works
  - Or, a jupyter notebook with comments.

**ATTRIBUTE NAMES** (more info at <https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest/data>)

Variable	Description	Units
apcp_sfc	3-Hour accumulated precipitation at the surface	kg m-2
dlwrf_sfc	Downward long-wave radiative flux average at the surface	W m-2
dswrf_sfc	Downward short-wave radiative flux average at the surface	W m-2
pres_msl	Air pressure at mean sea level	Pa
pwat_eatm	Precipitable Water over the entire depth of the atmosphere	kg m-2
spfh_2m	Specific Humidity at 2 m above ground	kg kg-1
tcdc_eatm	Total cloud cover over the entire depth of the atmosphere	%
tccl_eatm	Total column-integrated condensate over the entire atmos.	kg m-2
tmax_2m	Maximum Temperature over the past 3 hours at 2 m above the ground	K
tmin_2m	Minimum Temperature over the past 3 hours at 2 m above the ground	K
tmp_2m	Current temperature at 2 m above the ground	K
tmp_sfc	Temperature of the surface	K
ulwrf_sfc	Upward long-wave radiation at the surface	W m-2
ulwrf_tatm	Upward long-wave radiation at the top of the atmosphere	W m-2
uswrf_sfc	Upward short-wave radiation at the surface	W m-2