Notes Python

**Configuration choisie**

* Anaconda pour les librairies standards
* PyCharm comme IDE. Si trop lourd revenir sur Sublime. Spyder a une bonne visualisation de dataset

**Random Forest**

* Need float values, no missing data
* <http://www.analyticbridge.com/profiles/blogs/random-forest-in-python>

**Selectionner les lignes 1 à 4**

Table[1 :4]

**Selectionner les columns**

Table[]

**Kicker les lignes contenant des valeurs NaN**

Table.dropna()

**Create new dataframe from existing dataframe column**

Nrow

len(df.index)

**Rename column**

train10=train10.rename(columns = {'P1':'P1(W)'})

**MSE automatique**

**http://stackoverflow.com/questions/17197492/root-mean-square-error-in-python**

from sklearn.metrics import mean\_squared\_error

from math import sqrt

rms = sqrt(mean\_squared\_error(y\_actual, y\_predicted))

Selectionner des variables au dessus d’un seuils dans une liste

import numpy as np

j = np.array(j)

sum(j > i)

**Cross validation**

<http://scikit-learn.org/stable/modules/cross_validation.html>

**Model selection**

<http://scikit-learn.org/stable/modules/model_evaluation.html>

Attention à la MSE comme indicateur : <https://github.com/scikit-learn/scikit-learn/issues/2439>

**Regression polynomiale**

<http://stackoverflow.com/questions/10988082/multivariate-polynomial-regression-with-numpy>

**Cookbook HeatMap**

<http://stackoverflow.com/questions/14391959/heatmap-in-matplotlib-with-pcolor>

**Concatenate dataframes**

<http://pandas.pydata.org/pandas-docs/stable/merging.html>

**Split string**

**PyPlot tutorial**

<http://matplotlib.org/users/pyplot_tutorial.html>

**Handling date and time**

* Universal time: seconds since Unix Epoch Time (1Jan1970)

**i) Conversion for using the datetime package**

* datetime.datetime.fromtimestamp(1284286794)

**ii) Conversion with pandas.to\_datetime**

**Managing Timestamp type** – from pd.to\_datetime

**Get a file from Internet to your current directory**

Curl –O <https://github.com/FerdinandL/Playground/20160228_processTables.py>

**Null bytes in csv**

* Find them <http://stackoverflow.com/questions/4166070/python-csv-error-line-contains-null-byte>