

El Nino Dataset

Cotting Matthieu

Bonelli Alberto

Allani Mohamed

Keywords: Data science

During this project, we would like to investigate the data set of oceanographic and surface meteorological readings taken from a series of buoys positioned throughout the equatorial Pacific, from the dataset "El Nino Dataset" of the website *kaggle*.

This dataset contains many different measures such as date, latitude, longitude, zonal winds, meridional winds, relative humidity, air temperature and sea surface temperature.

We would like to

- Perform data cleaning because some data are missing and there might be outliers.
- know if the temperature influences the wind direction and/or its intensity, and investigate other possible relationship
- know if it is possible to define regions in the pacific area with common meteorological characteristics
- Trying to predict the evolution of the different parameters, given the year, the month and the day and the previous values of the parameters
- Evaluate the performances of our algorithm by taking only a part of our data for the training and then test with the rest of the data

For all these tasks we would use several algorithms for the machine learning part and different metrics to assess the eventual correlations between the data. Then we would choose which ones suit the best our dataset depending on the intermediate results that we would get.