

## Abstract

The project is about performing some operations on CMIP (Coupled-Model Inter-comparison Project models phase -6) climate models on any parameter like wind. To explore how to access '.nc' files in Python and extent of Machine learning on it.

The libraries included would mostly be among pandas, statsmodels.formula.api, sklearn.model\_selection.train\_test\_split, seaborn and more as we go ahead. These libraries we had used while learning  $R^2$  models. If data is suitable for linear regression, this  $R$  squared model can be formulated with train\_test\_split.

We would import the data, try to plot the parameters with seaborn/matplotlib. We would like to establish hypothesis functions for given data so as to predict the value for given inputs using methods like gradient descent. Correlation coefficient for linear regression of data can be printed using statsmodels.

Extra things we will have to learn are to handle '.nc' files. We have learned Machine learning algorithms in MATLAB, not in Python, so we would have to figure out the changes and go on.

The motivation behind the project is this mini-project would be helpful for our MTP projects involving large CMIP data objected to compare past historical data and climate projected future data, to form basics to analyze '.nc' file in Python, perform the learned Machine learning algorithms like gradient decent, techniques from Neural networks. Also, we would try to establish correlations between different parameters like wind, salinity, sea surface elevation, if possible, in given time window of two weeks.

### References:

1. <https://realpython.com/linear-regression-in-python/>
2. [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)
3. <https://seaborn.pydata.org/>