**Solar Panel Weather Prediction in Aswan**

**Problem:**

In order to regulate energy consumption, it is essential to know the power output of solar panels. Since Egypt is facing a climate change problem, a new solar panel consumption study needs to be administered.

**Data Summary:**

The data gives several features regarding the weather in Aswan, ranging from 4/30/2021, until 4/1/2022. The data includes the average temperature, average dew, humidity, wind, pressure, and solar(PV).

Solar(PV) stands for photovoltaics, which is the conversion of light into electricity, this will be our main feature that we surround the study around.

**Data Discovery and Preprocessing:**

In our code, the data will be described, and multiple preprocessing steps will be taken if necessary. Steps such as dropping useless columns, missing values, creating dummy variables, taking care of missing data, conversion of data frame into NumPy, and dividing the data into training and testing.

The Data was discovered and the values were plotted in a 2d graph, in correlation to the date, etc.

Some of these methods will not be useful for our case since for example, we do not have missing or null values.

We drop first column, because it is just for numbering, then we drop date column since it isn’t useful for predicting Solar(PV), we convert data into NumPy array, set X to our values and Y to Solar(PV) and drop it from X, then we split data into training and testing

We then plotted a pair plot graph for better visualization of the data

We then applied variance threshold

**Data Predictions and Analysis:**

We did some more data discovering then applied multiple linear regression. Since we want to predict a numerical output using multiple numerical inputs.

We calculate the predicted values and their SSE

We show the regression coefficients and intercepts and the variance score, with 1 giving a flawless prediction

Accuracy is determined using the explained variance score

Then, we plot for residual error

The residual error plot showed that the data wasn’t revolved around the horizontal access, which indicated that a non-linear model would fit more

Afterwards we did a Random Forest Regressor and got the RMSE (squared of MSE), and the value was way above 1, meaning that the model couldn’t find an optimized solution