Overview of Group 26 Operations

* Our dataset consists of data for 2 solar generation power plants in India, gotten from a kaggle task:

<https://www.kaggle.com/anikannal/solar-power-generation-data>

* Each station has 2 csv tables of data, one with data relating to weather conditions (irradiation, temperature etc) and one with data relating to power production in the inverters themselves (DC Power, AC Power, Energy Yield per day etc).
* Each table had data taken over 34 days, and observations were recorded every 15 minutes throughout each day.
* Some values in both tables were missing, and some were present in only one table but not in the other. The missing values were filled in with zeros, and a record placed as to whether that value was there originally or artificially filled in by us.
* Some additional features were added to help us predict power production, for example production values over the last three days, time of day, and whether a given observation occured during the day or at night.
* We split the data into 2 sets, a train and test set, where train consisted of 26 days’ worth of data, and test of 8 days’ worth.
* We trained 3 regressors on the data, a Random Forest Regressor, an XG Boost regressor and linear regressioon, taking DC Power produced as the variable top be predicted
* We also trained a baseline model, which predicted each value as simply the running average of the previous 3 days’ values.
* After choosing the best hyperparameters for each regressor separately, as well as some dimensionality reduction we used the models to predict each day’s production using a mixture of observations from the previous three days. The metrics measured were Root Mean Squared Error, Mean Absolute Error and R2 score. We made predictions for DC Power values at every 15 minute intervall, as well as total DC Power as a sum for the whole day and these are summarized in the table below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |