Solar Prediction

Capstone Two

Introduction

- Solar radiation is a key parameter to study climate change, environmental pollution, crop production, food industry, and hydrology.
- Designing a solar energy field, requires the availability of global to predict solar radiation at the desire location.
- Solar radiation is known to have dependency on several parameters such as sun rise, sun set, temperatures, humidity and more.
- Solar radiation measurements are much more complex and expensive.
- The ability to predict solar radiation potential is at most importance to further build alternative energy sources and predicting the performance of solar energy equipment.

- The NASA data obtained from Kaggle.
- The data has 32686 samples.
- There are 11 columns.

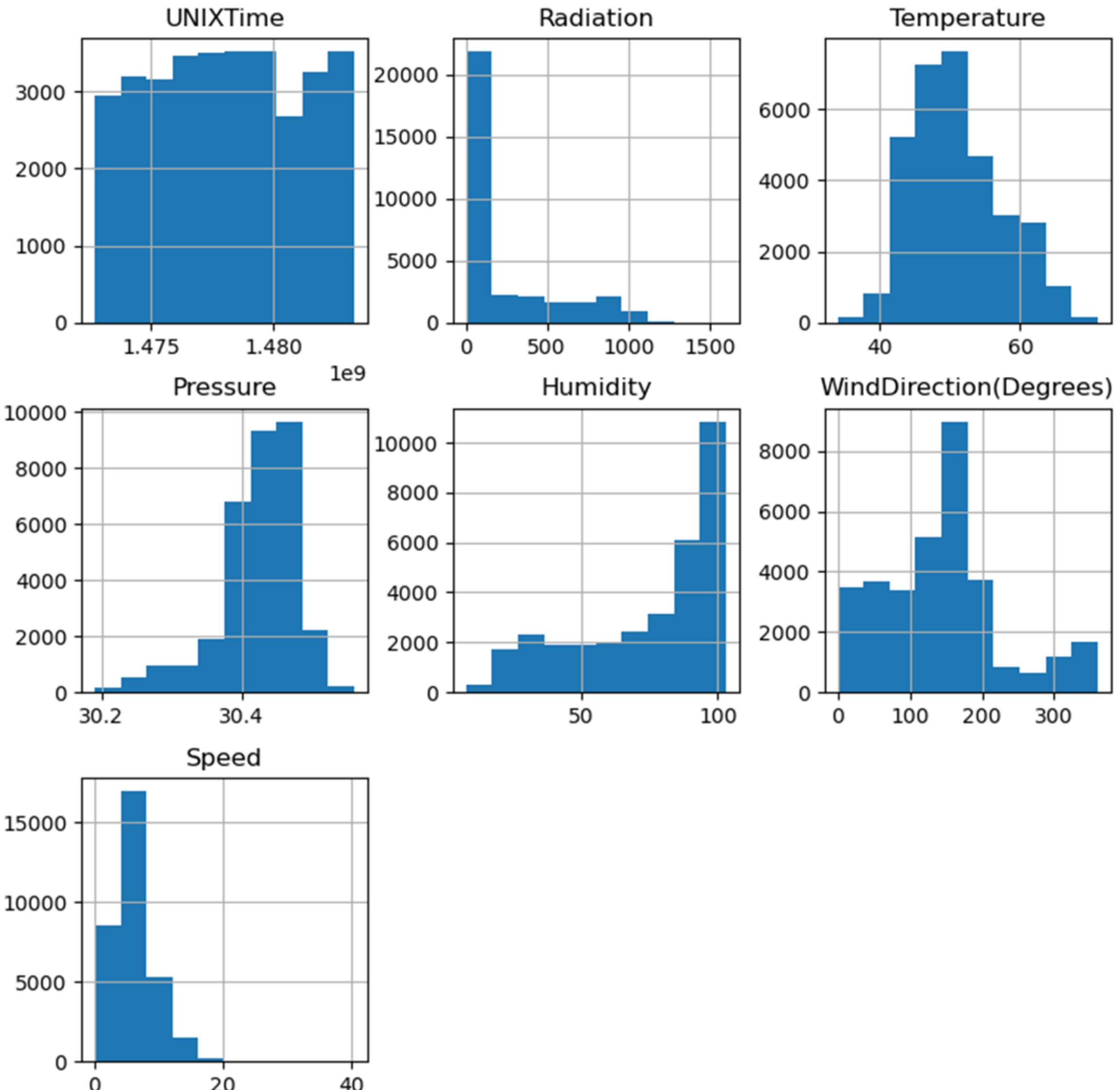
Data summery

```
In [92]: solar_prediction.describe()
```

Out[92]:

	UNIXTime	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	MonthOfYear	DayOfYear	WeekOfYear
count	3.268600e+04	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000
mean	1.478047e+09	207.124697	51.103255	30.422879	75.016307	143.489821	6.243869	10.526066	306.110965	43.871015
std	3.005037e+06	315.916387	6.201157	0.054673	25.990219	83.167500	3.490474	1.096691	34.781367	4.963061
min	1.472724e+09	1.110000	34.000000	30.190000	8.000000	0.090000	0.000000	9.000000	245.000000	35.000000
25%	1.475546e+09	1.230000	46.000000	30.400000	56.000000	82.227500	3.370000	10.000000	277.000000	40.000000
50%	1.478026e+09	2.660000	50.000000	30.430000	85.000000	147.700000	5.620000	11.000000	306.000000	44.000000
75%	1.480480e+09	354.235000	55.000000	30.460000	97.000000	179.310000	7.870000	11.000000	334.000000	48.000000
max	1.483265e+09	1601.260000	71.000000	30.560000	103.000000	359.950000	40.500000	12.000000	366.000000	52.000000

Data summery

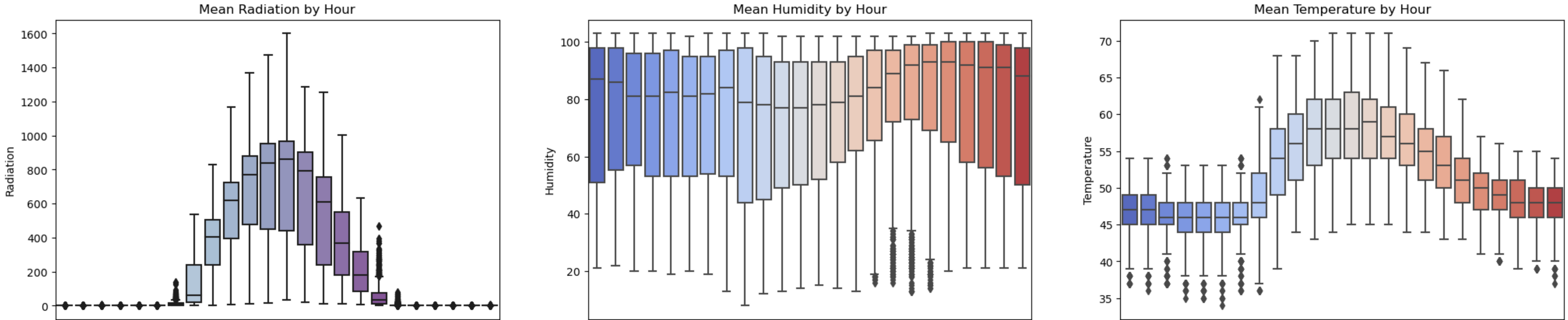


Correlation

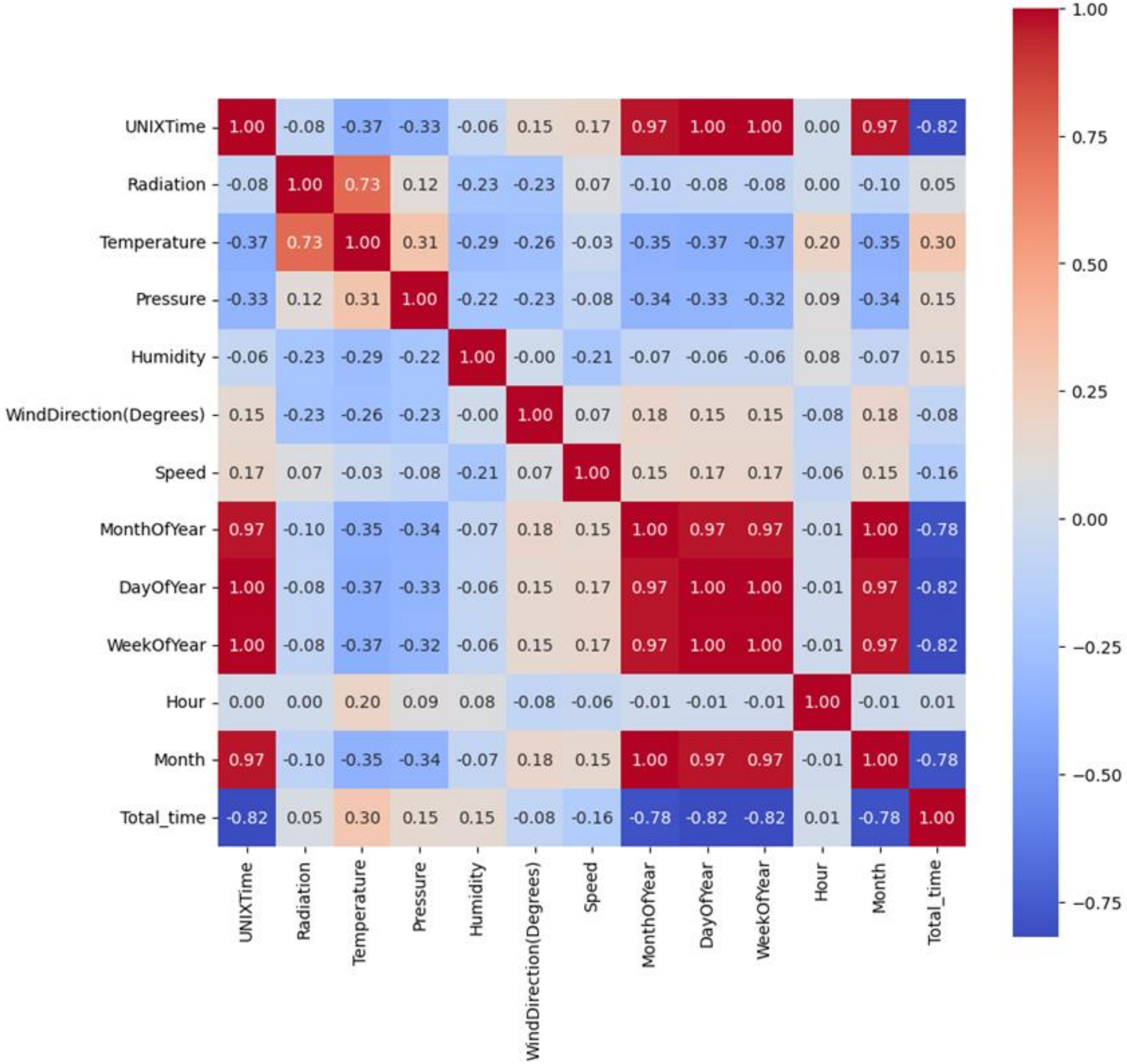
In [92]:
```
solar_prediction.describe()
```

Out[92]:

	UNIXTime	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	MonthOfYear	DayOfYear	WeekOfYear
count	3.268600e+04	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000
mean	1.478047e+09	207.124697	51.103255	30.422879	75.016307	143.489821	6.243869	10.526066	306.110965	43.871015
std	3.005037e+06	315.916387	6.201157	0.054673	25.990219	83.167500	3.490474	1.096691	34.781367	4.963061
min	1.472724e+09	1.110000	34.000000	30.190000	8.000000	0.090000	0.000000	9.000000	245.000000	35.000000
25%	1.475546e+09	1.230000	46.000000	30.400000	56.000000	82.227500	3.370000	10.000000	277.000000	40.000000
50%	1.478026e+09	2.660000	50.000000	30.430000	85.000000	147.700000	5.620000	11.000000	306.000000	44.000000
75%	1.480480e+09	354.235000	55.000000	30.460000	97.000000	179.310000	7.870000	11.000000	334.000000	48.000000
max	1.483265e+09	1601.260000	71.000000	30.560000	103.000000	359.950000	40.500000	12.000000	366.000000	52.000000



Correlation map



Training Models

- LinearRegression
- RandomForestRegressor
- GradientBoostingRegressor

❑Test Size=0.3, Random State=42

❑Cross validation for RFR and GBR

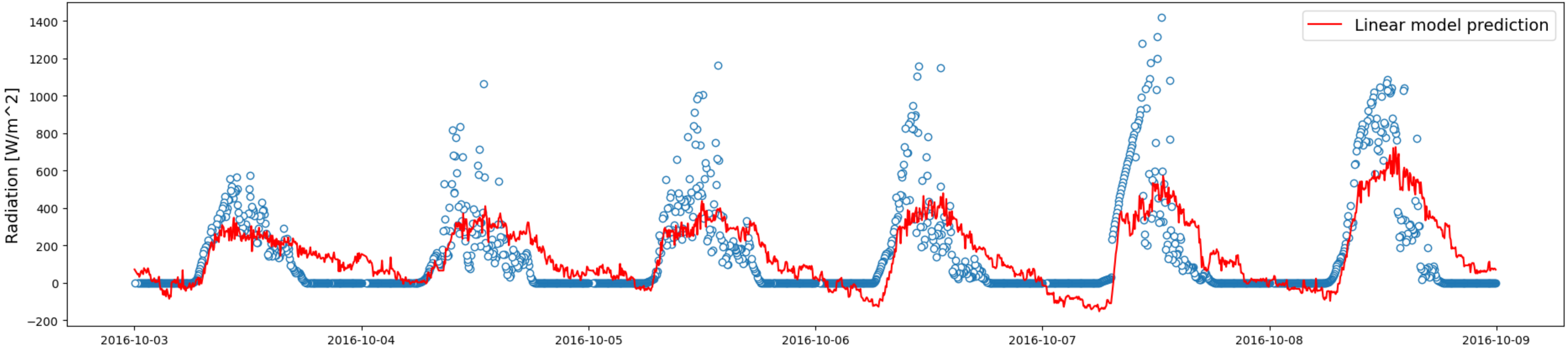
❑Finding the optimal hyperparameters for RFR and GBR

❑Parameter : Temperature, Pressure, Humidity, Wind Direction(Degrees), Speed, Month Of Year, Day Of Year, Week Of Year, Total time

Results Liner model

- Linear model, R^2 training set:0.61
- Linear model, R^2 test set:0.59

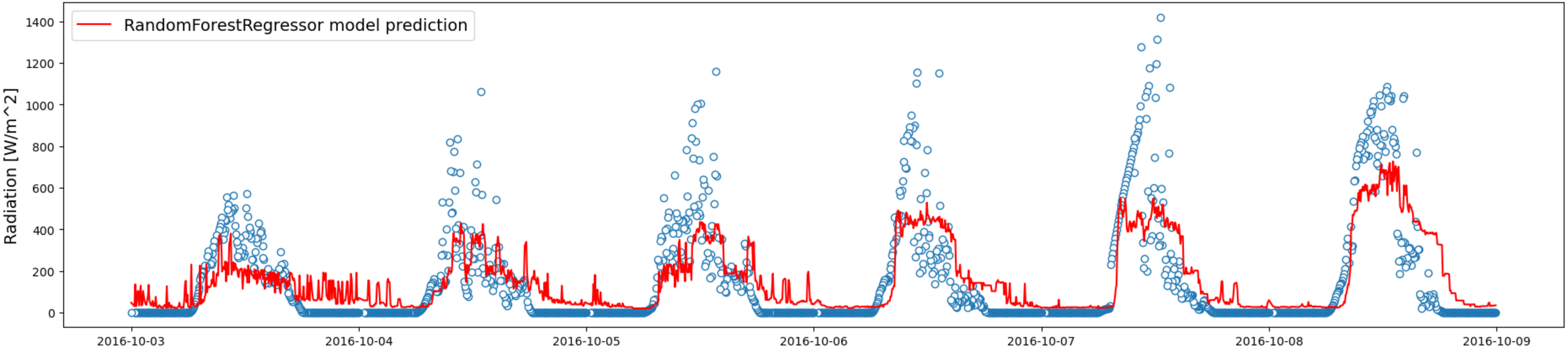
5 Days prediction



Results Random Forest Regressor

- Random Forest, R^2 score training set:0.65
- Random Forest, R^2 score test set:0.64

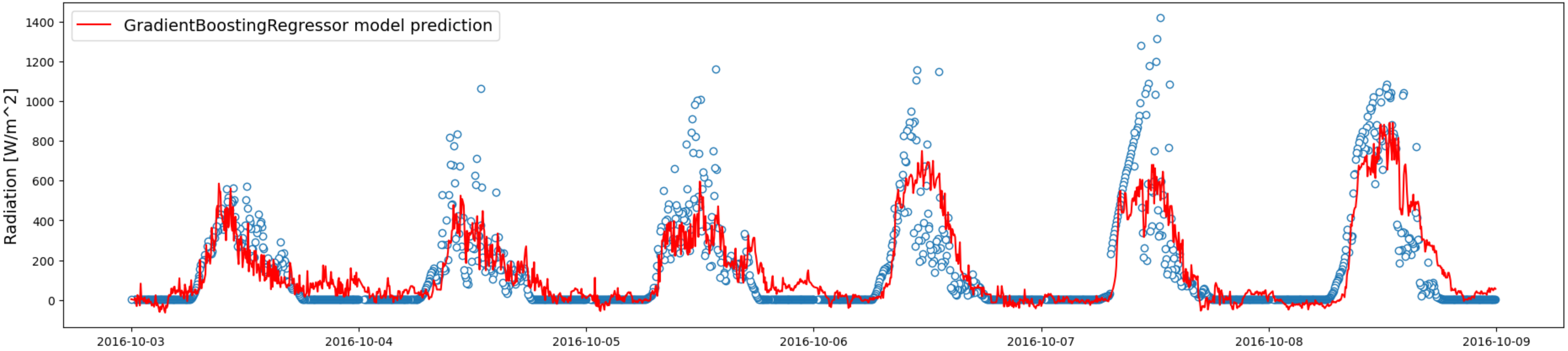
5 Days prediction



Results Gradient Boosting Regressor

- Gradient Boosting, R^2 score training set:0.79
- Gradient Boosting, R^2 score test set:0.77

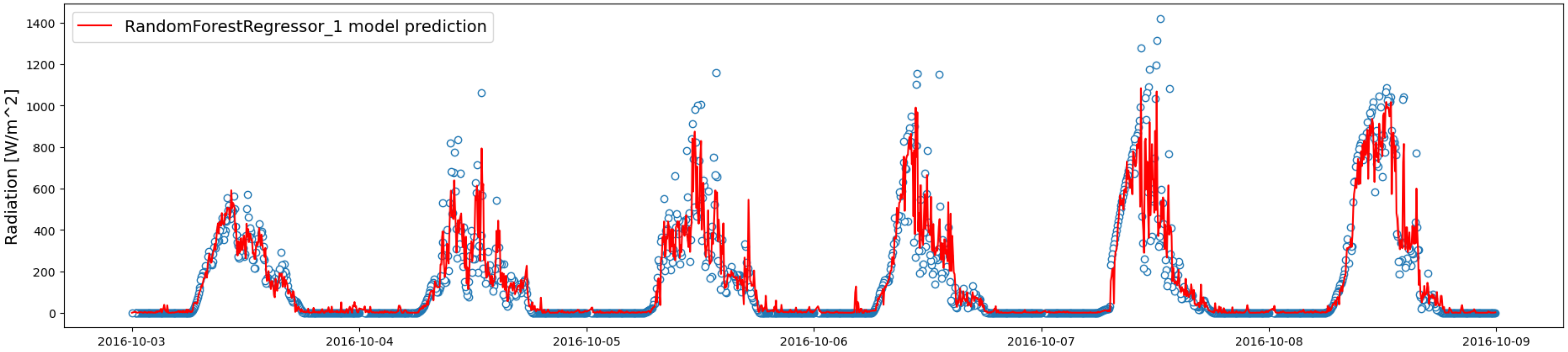
5 Days prediction



Results Random Forest Regressor

- Changing Max Depth=25 and Random State=3
- Random Forest, R^2 score training set:0.98
- Random Forest, R^2 score test set:0.88

5 Days prediction



Conclusion

- Based on random state =42 an on the selected parameters, the best model to choose for solar prediction is Gradient Boosting Regressor, with R^2 = 0.77
- Changing parameters, both columns selection and models parameters can results in better prediction.