

Solar Radiation Prediction:

Solar radiation (also known as or sunlight) is the electromagnetic radiation emitted by the sun. Solar Radiation has high impact on people lives and, on the environment, Solar radiation is also a key parameter to study climate change, environmental pollution, crop production, food industry, and hydrology. Solar radiation measurements are much more complex and expensive, the ability to predict solar radiation on different parameters that are simpler to measure can help increase the further build of alternative energy sources.

Data:

The data is from NASA, the dataset was imported from Kaggle.

Link: [Solar Radiation Prediction | Kaggle](#)

This dataset contains such columns as: "wind direction", "wind speed", "humidity" and temperature. The response parameter that is to be predicted is: "Solar radiation", It contains measurements for 4 months.

The data have 32686 rows x 11 columns

Criteria for success:

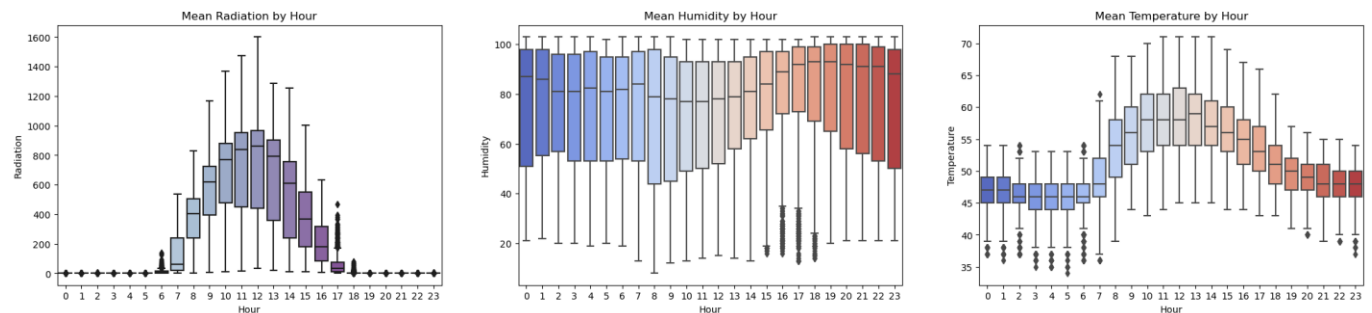
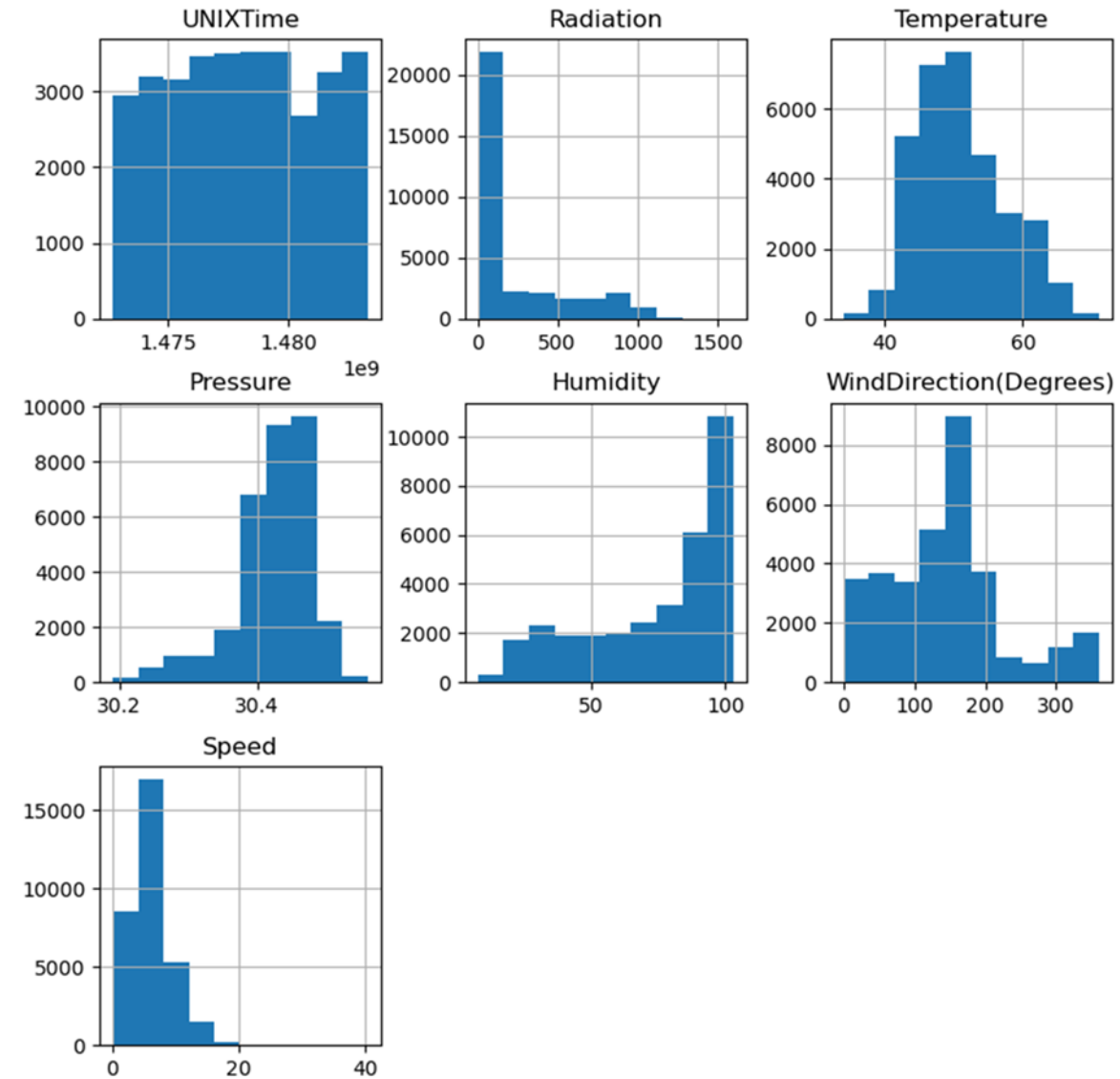
The ability to predict solar radiation with high coefficient of determination(R^2).

Exploratory data analysis:

The data described below

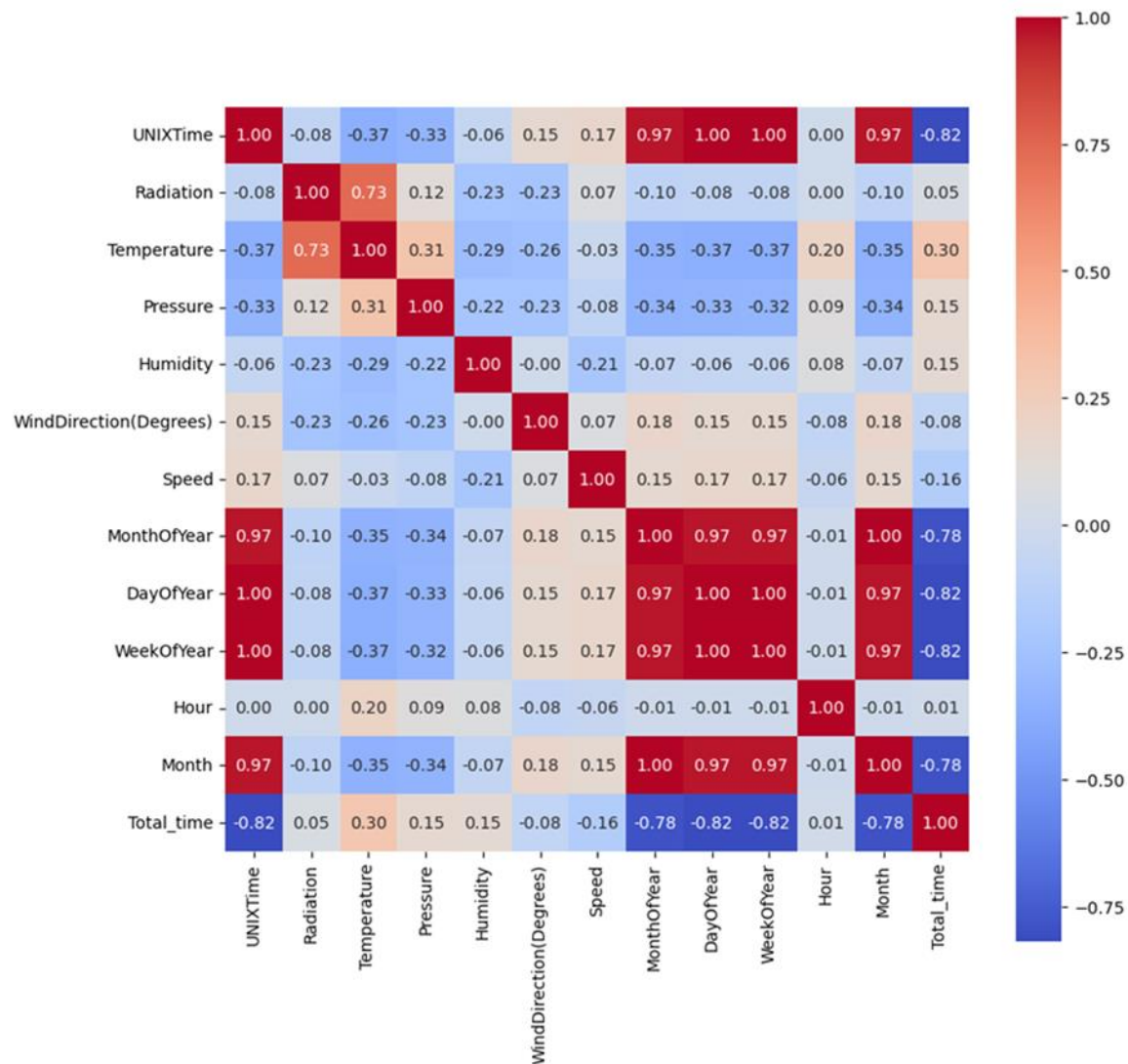
solar_prediction.describe()										
	UNIXTime	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	MonthOfYear	DayOfYear	WeekOfYear
count	3.268600e+04	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000	32686.000000
mean	1.478047e+09	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

The columns histogram presented below



During the cleaning I did not find any null values and the data did not show any large deviation needed to be explored.

The correlation map between the different parameters shown below:



Hi correlation was found between solar radiation and temperatures as can be expected.

Algorithms & Machine Learning:

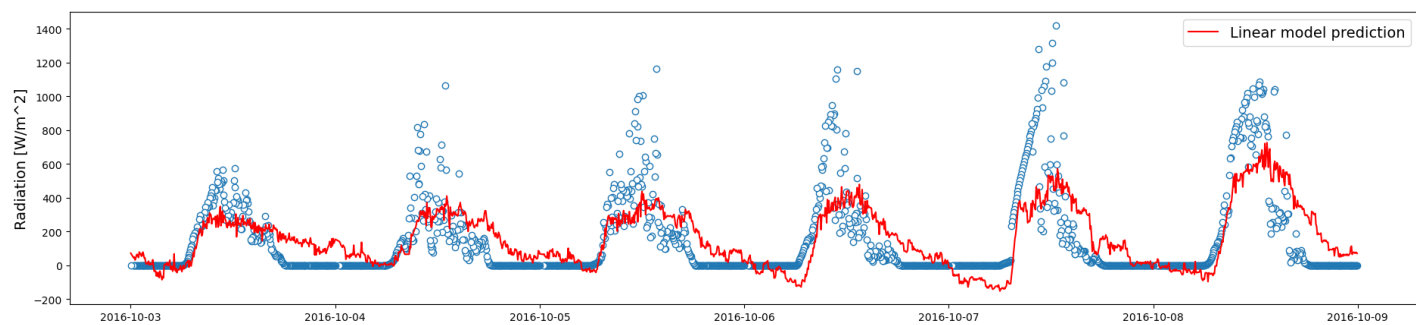
I worked with training parameters: Test Size=0.3 and Random State=42. I was working with 3 different models: linear regression model, Random Forest regressor and Gradient Boosting regressor.

The results are shown in the table below:

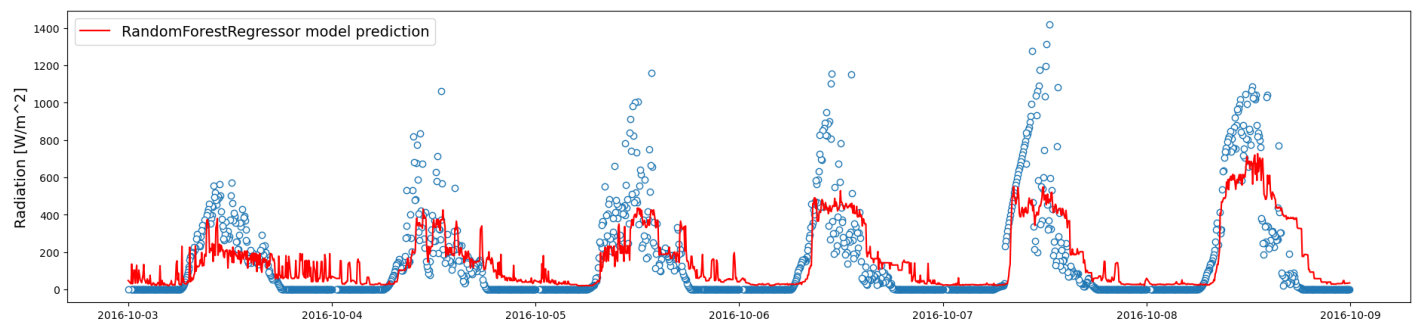
Model	CV MSE	MSE train	MSE test	R^2 score train	R^2 score test
Linear model		39312.84	40445.61	0.61	0.59
Random Forest regressor	35295.42	34996.37	35827.54	0.65	0.64
Gradient Boosting regressor	22373.20	21347.49	22817.60	0.79	0.77

The prediction plots for five days shown below:

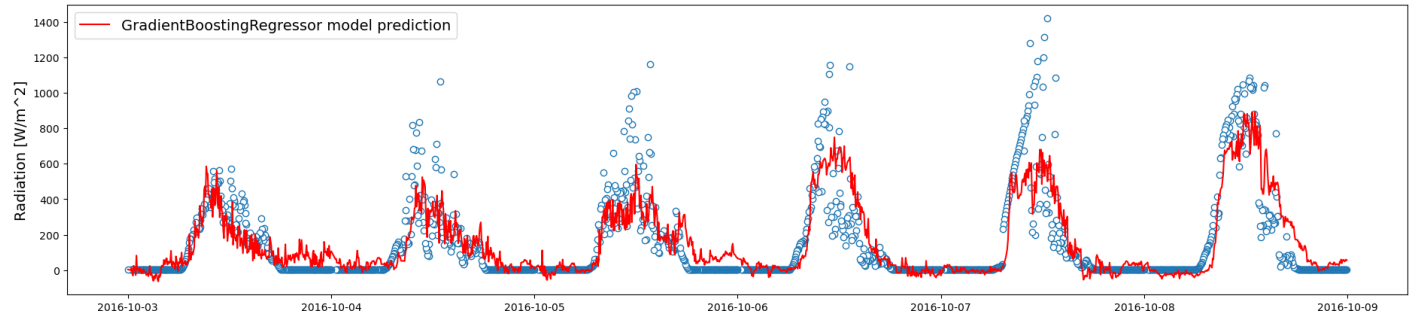
Linear model



Random Forest regressor



Gradient Boosting regressor



The best results achieved from the Gradient Boosting regressor.

The dataset I used in this prediction was: Temperature, Pressure, Humidity, Wind Direction (Degrees), Speed, Month Of Year, Day Of Year, Week Of Year and Total time.

Conclusion:

During this project I found a way to predict solar radiation using different parameters with R^2 score of 0.77. More complex models can be added in order to increase the R^2 score, choosing different parameters can also affect the R^2 score and give us more exact prediction.