TECHNOCOLABS DATA ANALYSIS INTERNSHIP

PROJECT REPORT

TITLE: Forecasting Solar Energy Prediction Using Weather Data

AIM: The main aim of our project is to perform data analysis and train a model using the most accurate Machine Learning algorithm in order to predict solar radiation for the next 48 hours.

ABSTRACT: Solar radiation prediction has a great importance in electricity generation from solar energy and helps to size photovoltaic power systems.

DATASET:

The database actually includes several types of solar radiation measurements; we specifically choose to analyze global horizontal irradiance (GHI), as it incorporates both direct incident radiation and ambient solar radiation reflected from nearby surfaces and atmospheric particles. This makes it a good indicator for solar panel readings

Original format of the dataset: csv file

A brief explanation of every column in the dataset is as follows:

1. **Year-**

   We limit our analysis to the data collected over the entirety of 2017, 2018 and 2019. This leaves us with distinct observations, each with the 14 features a corresponding measure of GHI.

2. **Surface Albedo-**

   The surface albedo quantifies the fraction of the sunlight reflected by the surface of the Earth. Different albedo concepts are defined: The directional albedo or directional-hemispherical reflectance (also called black-sky albedo) is the integration of

the bi-directional reflectance over the viewing hemisphere.The overall albedo of the Earth - measured to be 0.30 - has a significant effect on the equilibrium temperature of the Earth as it changes how much solar energy is reflected by the Earth as opposed to how much is absorbed.

3. **Cloud Type-**

The foundation consists of 10 major cloud types. In addition to cirrus, stratus, cumulus, and nimbus clouds, there are cirrostratus, cirrocumulus, altostratus, altocumulus, stratocumulus, nimbostratus, and cumulonimbus clouds. Clouds can sometimes increase solar output.

4. **Dew Point-**

The dew point is the temperature the air needs to be cooled to (at constant pressure) in order to achieve a relative humidity (RH) of 100%. ... The higher the dew point rises, the greater the amount of moisture in the air.Dew is repelled from the glass surface of solar panels produced by reputable manufacturers by way of a hydrophobic coating.

5. **Zenith Angle-**

The zenith angle is the angle between the sun and the vertical. The zenith angle is similar to the elevation angle but it is measured from the vertical rather than from the horizontal, thus making the zenith angle = 90° - elevation.The intensity of solar radiation is largely a function of the angle of incidence, the angle at which the Sun's rays strike the Earth's surface. If the Sun is positioned directly overhead or 90° from the horizon, the incoming insolation strikes the surface of the Earth at right angles and is most intense.

6. **Pressure-**

Pressure is defined to be the amount of force exerted per area. P = F A {\Large P=\dfrac{F}{A}} P=AF. So to create a large amount of pressure, you can either exert a large force or exert a force over a small area (or do both).Effect of air pressure on the output of photovoltaic panels and solar illuminance/intensity have been done. Data analysis spells that solar illuminance/intensity, output current and voltage rise with increase in air pressure.

7. **Wind Direction-**

Wind direction is defined as the direction the wind is coming from. If you stand so that the wind is blowing directly into your face, the direction you are facing names the wind.Specifically, the effect of wind direction and tilt angle of the PV module is investigated. By increasing the tilt angle at wind velocities higher than 1 m/s, the average PV temperature increases by about 4 K; as a result, its efficiency decreases.

8. **Wind Speed-**

Wind speed, or wind flow speed, is a fundamental atmospheric quantity caused by air moving from high to low pressure, usually due to changes in temperature. Wind speed is now commonly measured with an anemometer.panel outputs (short circuit current and open circuit voltage) and solar illuminance/intensity are favoured by increase in wind speed: that is, when the wind is towards the front of an observer (or panel) with the sun,

some distance away in front and when the wind is towards the back of the observer (or panel)

9. **Relative Humidity-**

   Relative humidity is a way of describing how much humidity is present in the air, compared to how much there could be. ... If the actual amount of vapor is compared to the total amount there could be as a fraction, then the number tells if the air feels dry or moist.Humidity readily affects the efficiency of the solar cells and creates a minimal layer of water on its surface. It also decreases the efficiency by 10-20% of the total power output produced.

10. **Temperature-**

    As the temperature of the solar panel increases, its output current increases exponentially, while the voltage output is reduced linearly. In fact, the voltage reduction is so predictable that it can be used to accurately measure temperature. As a result, heat can severely reduce the solar panel's production of power.

11. **Precipitation Water-**

    When rain falls, the solar panels continue to generate electricity from the force of the falling rain on their surface. Compared to sunny days, rainy days have 90 percent less sunlight. With less sunlight, the solar panels produce an equal drop in electricity.

12. **GHI-**

    Global Horizontal Irradiance (GHI) is the total solar radiation incident on a horizontal surface. It is the sum of Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance, and ground-reflected radiation. HOMER uses Solar GHI to compute flat-panel PV output.GHI is the amount of solar radiation received per unit area by a horizontal surface from the hemisphere above. It comprises Direct Normal Irradiance, corrected for the angle of incidence on the surface, and Di use Horizontal Irradiance.

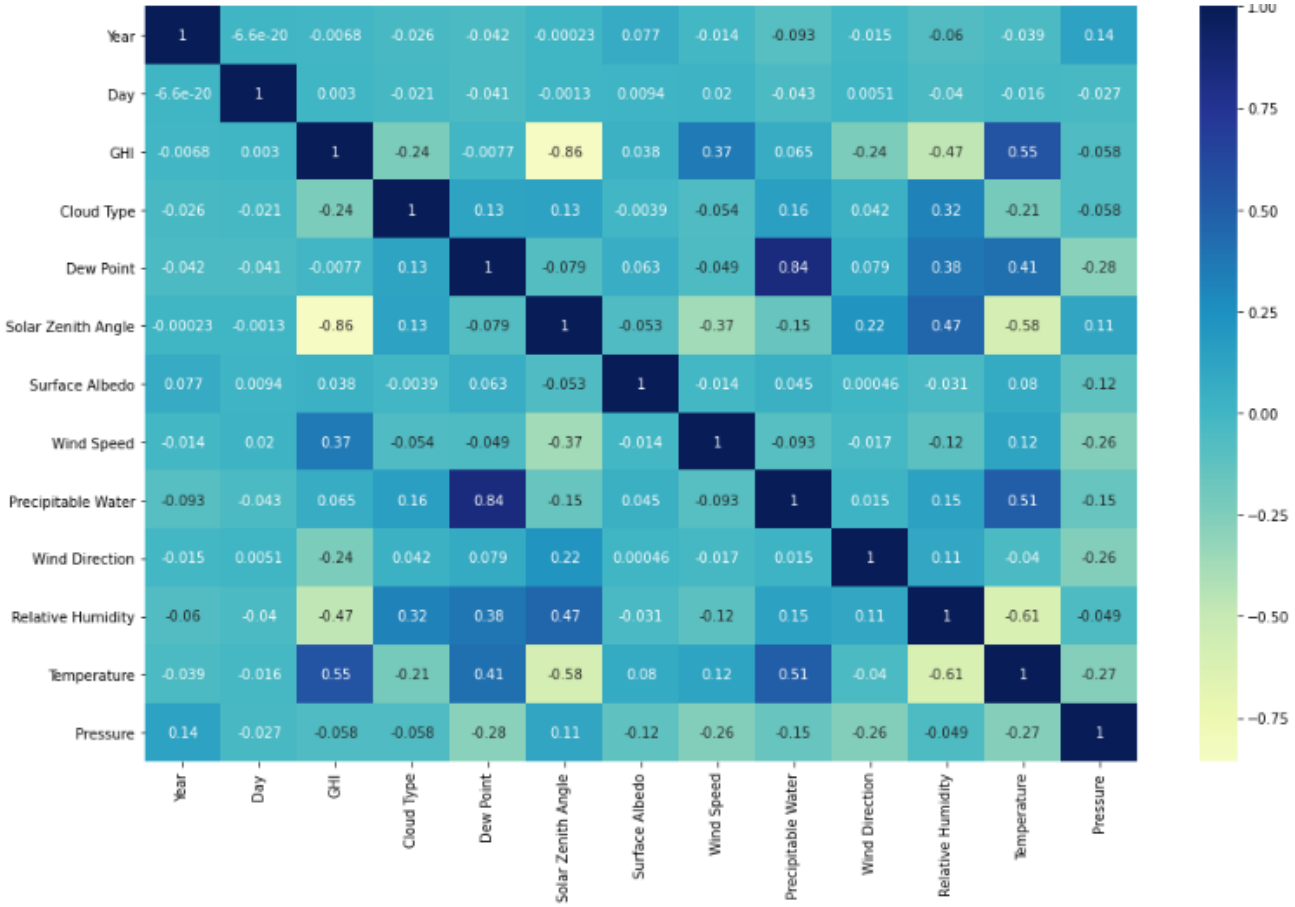
DATA SEGMENTATION AND DATA CLEANING:

- The dataset chosen contains weather data of 3 yrs, 2017, 2018, 2019.
- All the 3 datasets are concatenated to form a single dataset with 52560 rows and 18 columns.
- There are no null values in the dataset.
- Dropped unnecessary columns : Clearsky GHI','Fill Flag.
- Combine the Year, Month, Day, Hour, Minute column to form a perfect DateSeries in the format: YYYY-MM-DD HH:MM:SS.

```
datetime_series = pd.Series(pd.date_range(start = '2017-01-01',periods = 52560,freq =
'30min'),name = 'DateSeries')
```

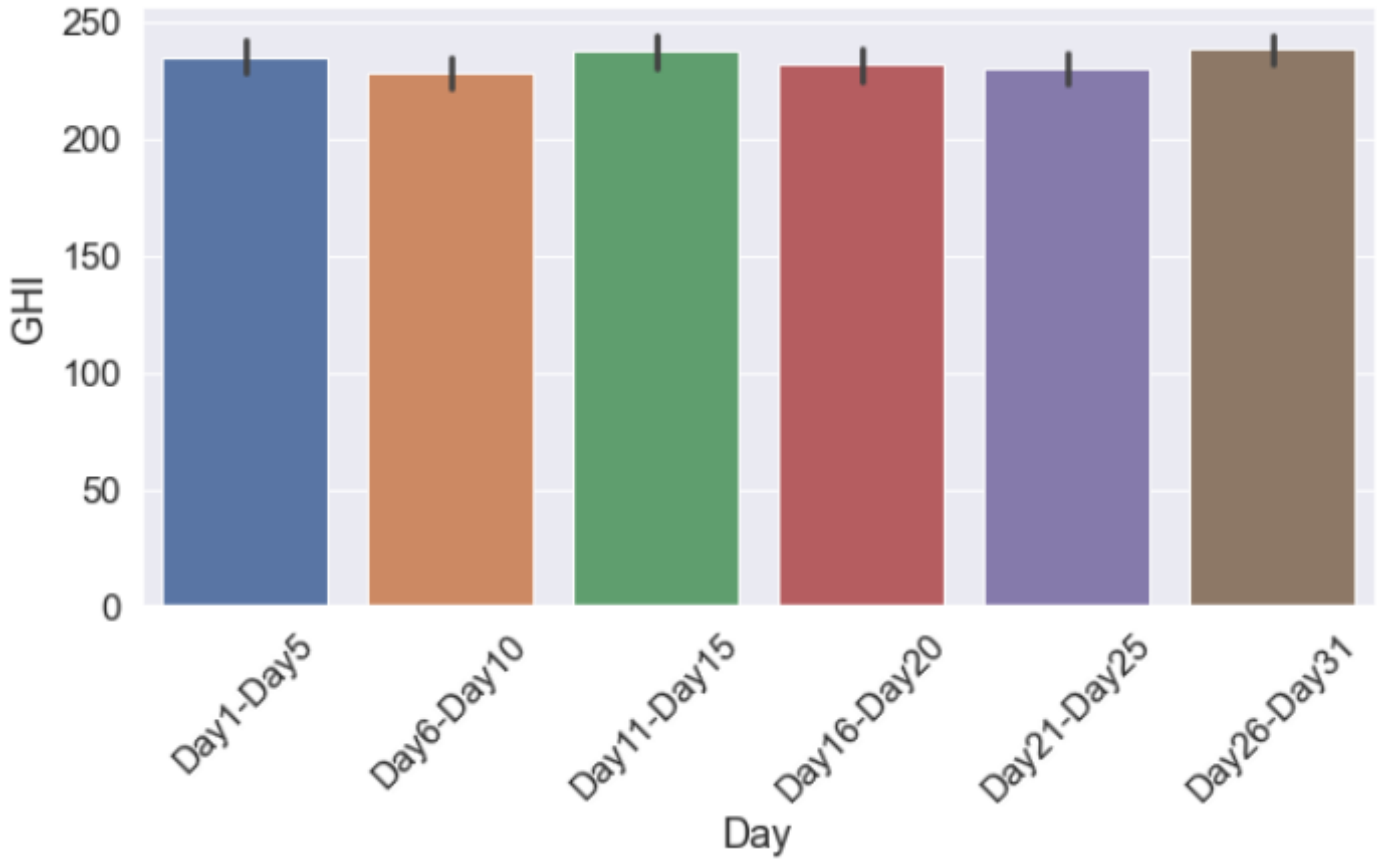
EXPLORATORY DATA ANALYSIS:

Plot 1: Correlation Study



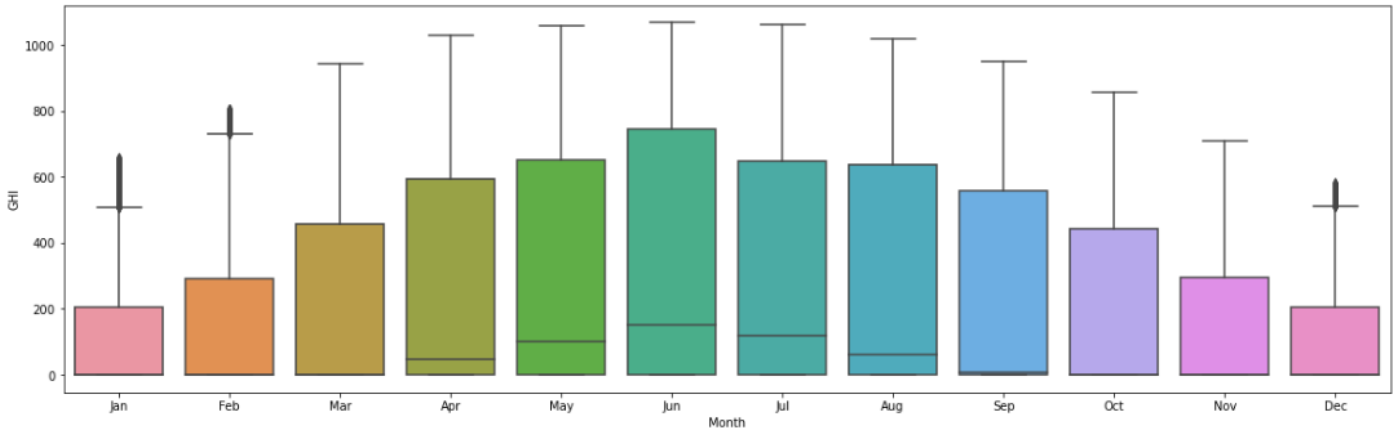
The heat map above shows a correlation between the different variables available in the data set. The highest positive correlation is seen between the variables Precipitate water and Solar Zenith Angle. There are some notable correlations like Temperature to GHI as well as with Precipitable Water. An inverse relationship is highly evident in the case of Solar Zenith Angle to GHI. The negative correlation also appears to be the case with Temperature to Relative Humidity and Solar Zenith Angle.

Plot 2: Day Trends



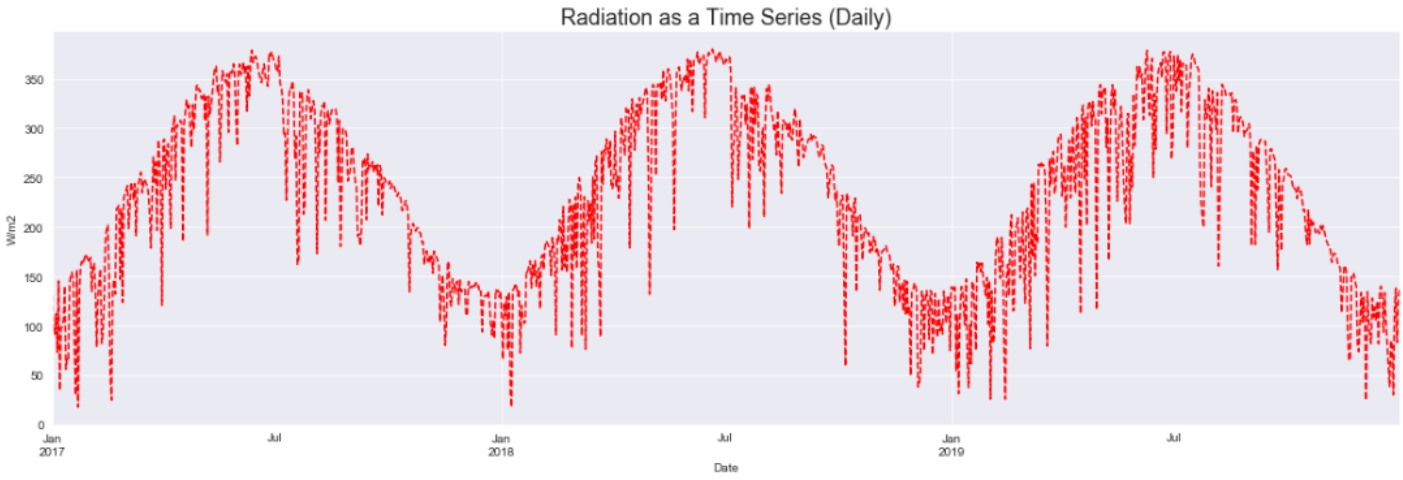
The plot describes the cumulative spread of GHI in the average days. The ranges differ by negligible values. On careful observation, the ranges of days between 1-5, 11-15, 26-31 have the highest values of GHI.

Plot 3: The Month Trend



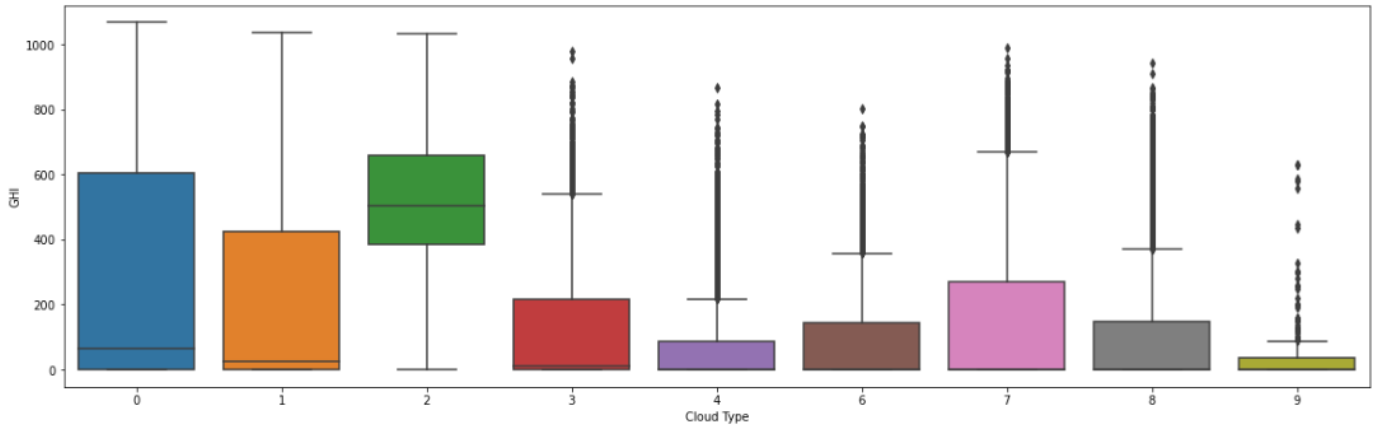
The plot shows a relationship between the target variable GHI with Months. Clearly, the median value of GHI is highest in the middle of the year (May, June, July and August), while the start and end of a year has negligible values of GHI, as inferred from the provided data.

Plot 4: Time Series Trend



The plot justifies the previous plot on month trend. The trend is same throughout the years from 2017 to 2019, being the peak GHI value is at the middle of the year,

Plot 5: Cloud Type

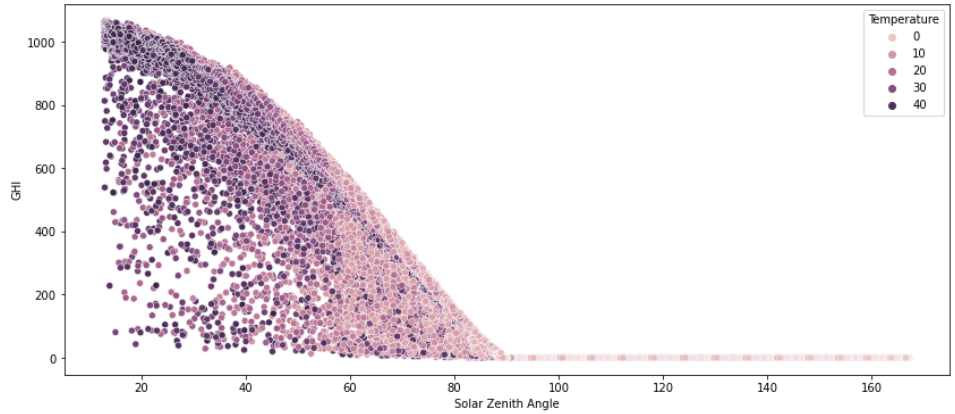


This is a plot that defines the relationship between GHI and Cloud Type. As inferred from the plot, the cloud type 2 has the highest mean GHI. Another important aspect to be jotted is that Cloud Type 0 has the highest ranges (25 % - 75%) of GHI values.

Plot 6: SolarZenithAngle vs GHI

```
plt.figure(figsize=(14,6))
sns.scatterplot(x=new_df['Solar Zenith Angle'],y=new_df['GHI'],hue=new_df['Temperature'])
```

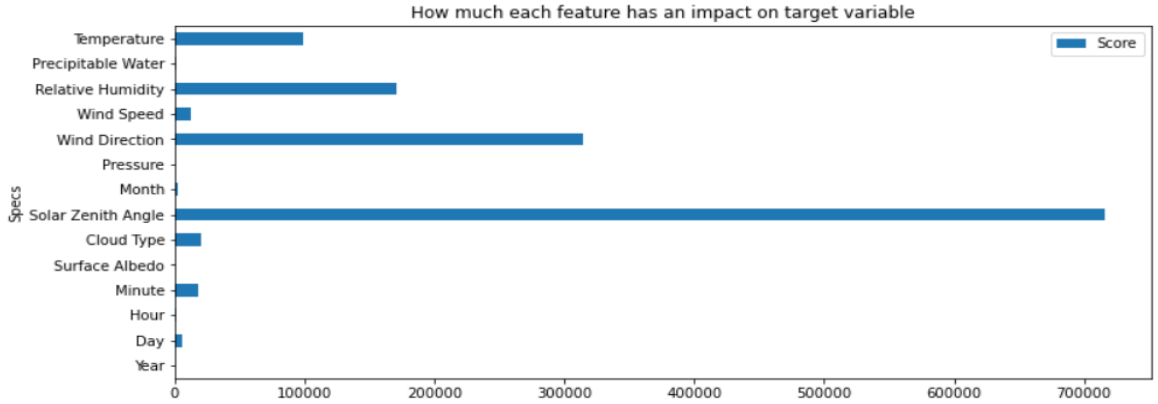
```
<AxesSubplot:xlabel='Solar Zenith Angle', ylabel='GHI'>
```



This plot is the scatter plot that shows the relation between SolarZenithAngle(SZA) and the target variable GHI in graphical form.From the Plot1 correlation plot we found the correlation between GHI and SolarZenith angle is (0.86) which is maximum for any. It states that our target variable GHI was mostly dependent on SZA.It shows us for SZA>90 degrees the GHI was equal to zero,to obtain maximum GHI we need to have minimum SZA angle.

TRAINING THE MODEL

In our project we have built a Decision Tree model where we have used one of the libraries of scikit-learn called SelectKBest and came up with 10 best features which have a significant impact in predicting the target variable.



Decision Tree Regressor:

Decision Tree is a kind of supervised learning technique that may be used for each classification and Regression problems. We have used this algorithm because when we analysed the patterns & trends of data , we found it to be in non-linear in nature. Since Decision tree is well suited to work on non-linear data we moved ahead to implement this algorithm.

Accuracy of Model

Score of the Model

```
print('Score of model on training data :{}'.format(model1.score(x_train,y_train)))
```

```
Score of model on training data :1.0
```

```
print('Score of model on testing data :{}'.format(model1.score(x_test,y_test)))
```

```
Score of model on testing data :0.9732899068019265
```

```
from sklearn.metrics import r2_score
```

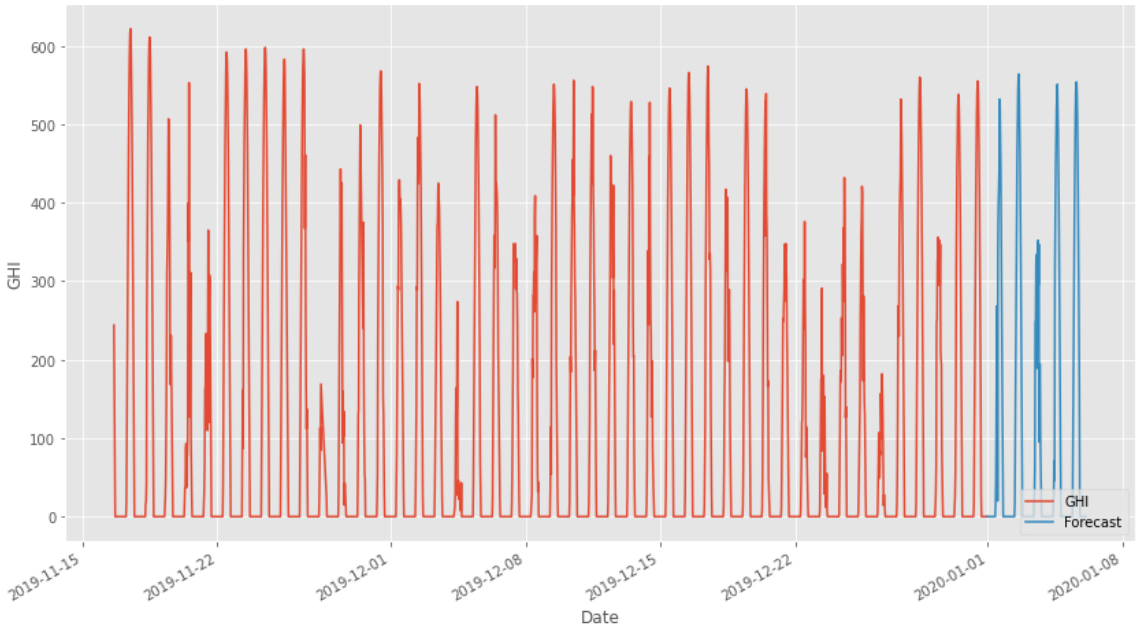
```
print('r2_score :{}'.format(r2_score(y_test,y_predict)))
```

```
r2_score :0.9732899068019265
```

Forecasting GHI

Based on our trained model with highest accuracy DecisionTreeRegressor(97.32%) we have developed a function which forecasts GHI based on the previous trend,on entering the timestamp(Date) it will return the forecast of that date.

```
next_48hrs_forecast=hr_forecast(1577817000)
```



Teammates:

Megha S
Mahima Chandwani
Lashman Pavan Kumar Gadde
Eesha Srivastava
Syed Khundmir Azmi
Pranoy G Bhadran
Naveen N S
Sudhanshu Pandey
Yash Lucas