Deloitte.

The Future of Renewable Energy - Solar Energy

By: Augusth Koppal

Table of Contents:

Business Understanding

Who are the stakeholders?

The stakeholders are venture capitalists from a large firm.

Which energy industry/sector is best to invest in for the future?

The renewable energy industry is one industry to focus on. In particular, the solar energy sector.

Which assets in the solar energy are most enticing?

Photovoltaic (PV) Systems from solar power plants.

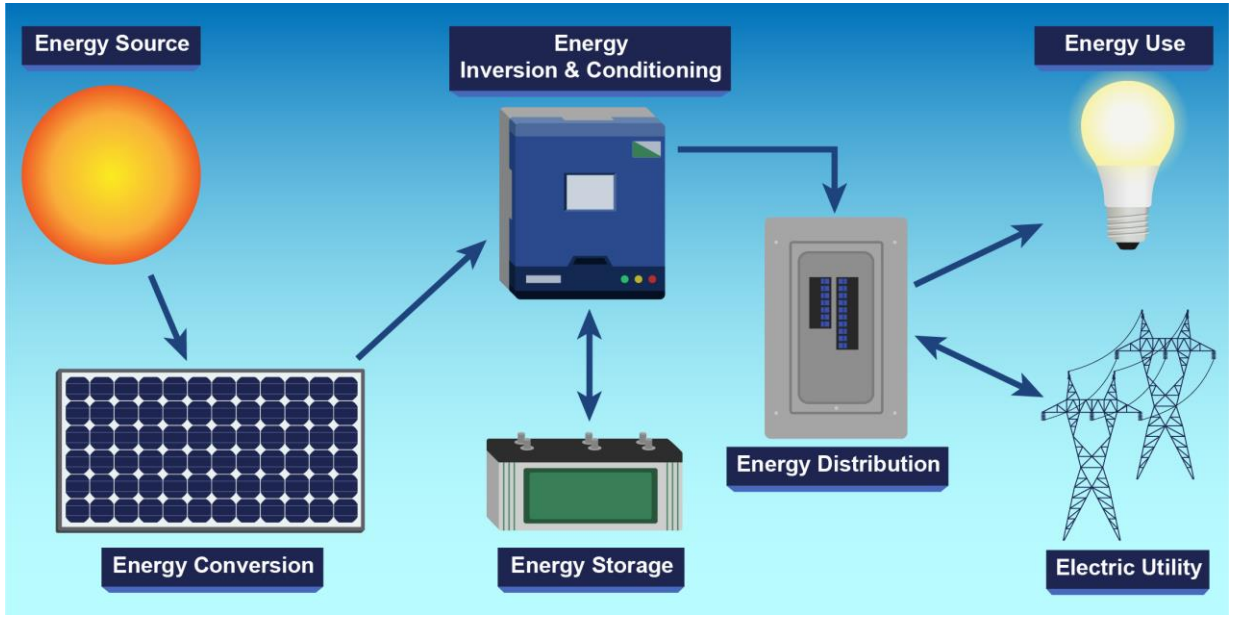
What is the focus of this project?

The model predicts the power output of PV systems in the short term.

Tremendous for managing power grid production daily/hourly.

Also, helpful for resource planning & or energy storage/delivery.

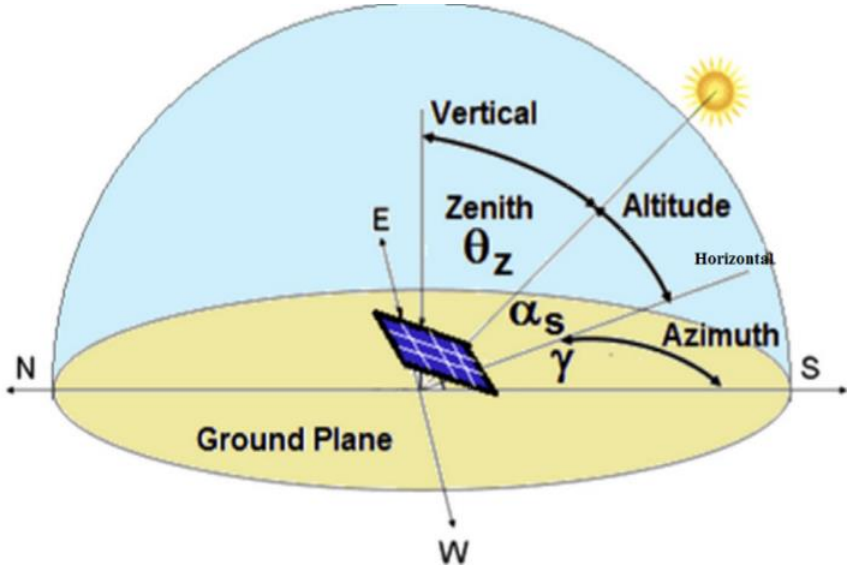
Project goal is to map an input value with a continuous target variable.



Data Understanding:

Environmental Attributes:

1. Temperature Above Ground
2. Relative Humidity Above Ground
3. Mean Sea Level Pressure
4. Total Precipitation
5. Snowfall Amount
6. Cloud Cover (Total, High & Low)
7. Shortwave Radiation (30% yield)
8. Wind Speed Above Ground (10, 80
9. Wind Direction Above Ground (10, 80 & 900 m.)
10. Wind Gust Above Ground (10 m.)



Angles of PV Systems:

1. Angle of Incidence
2. Zenith
3. Azimuth

Energy Generated :

1. Power Output (Kw)

Data Preparation:

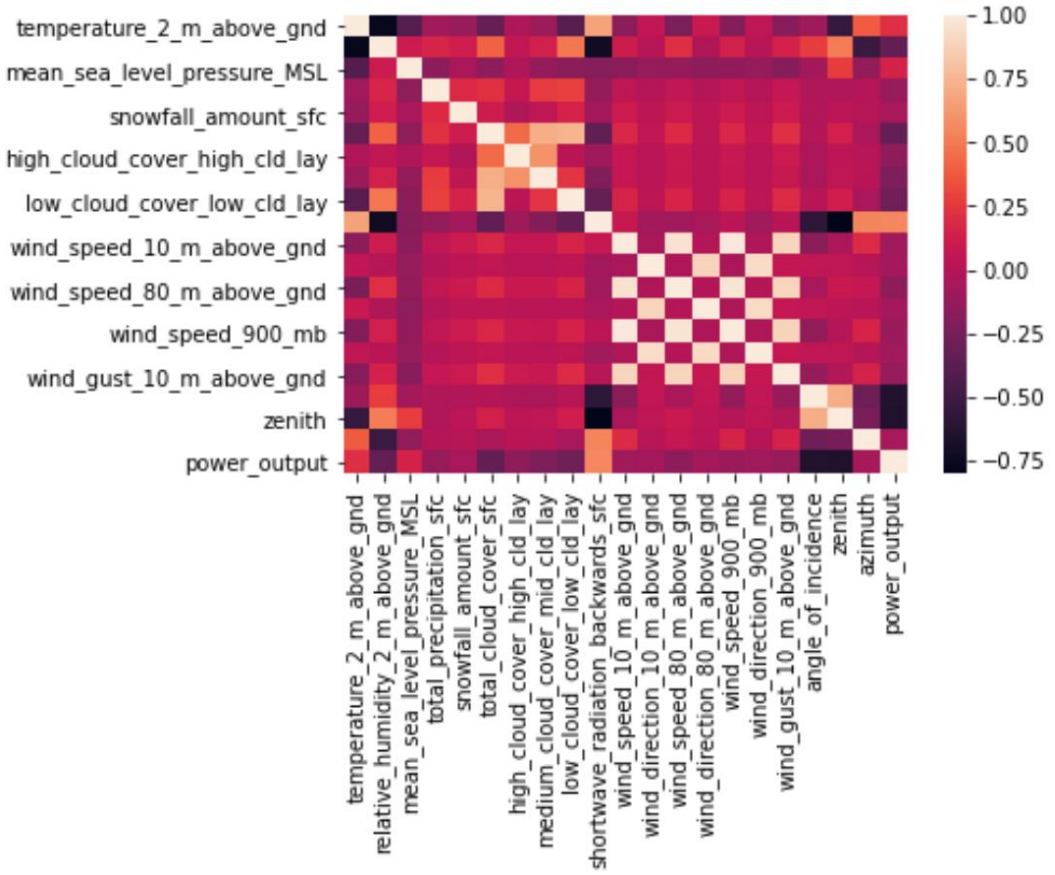
Feature Engineering:

- Feature engineering included taking the natural log transformation of the numeric variables.
- Dropped any null values.

Feature Selection:

1. Temperature Above Ground
2. Relative Humidity Above Ground
3. Mean Sea Level Pressure
4. Total Precipitation
5. Snowfall Amount
6. Shortwave Radiation (backwards)
7. Angle of Incidence
8. Zenith
9. Azimuth
10. Power Output (kW)

Heatmap for Data Correlation



Modeling Workflow & Results:

Machine Learning Algorithms

The algorithms used were specifically for regression tasks. Since the goal was to map an input value with a continuous target variable, these models were deemed to fit best:

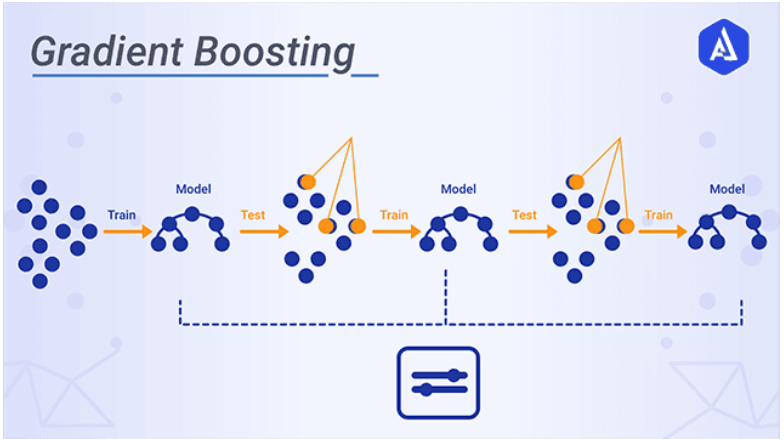
- **Linear Regression (396.3 MAE) & (513.8 RMSE)**
- **Gradient Boosted Regressor (315.8 MAE) & (453.3 RMSE)**
- **SVM (not ideal) (719.8 MAE) & (832.2 RMSE)**
- **Also, had a Naïve Baseline prediction before the data was split and got a MAE score of 839.58.**

ML Modeling Parameters

- 1 dimensional array
- X = Temperature Above Ground, Relative Humidity Above Ground, Mean Sea Level Pressure, Total Precipitation, Snowfall Amount, Shortwave Radiation (backwards), Angle of Incidence, Zenith, Azimuth.
- Y = Power Output
- Test Data = 30%
- Training Data = 70%

Hyper-Parameter Tuning of Gradient Boosting Regressor Model & Confidence Interval

- Hyper Parameters:
- estimators = [10, 50, 100, 150, 250, 300]
- GridSearchCV:
- Model, grid, cv = 5, verbose = True
- Fit 5 folds for each of the 6 Candidates, totaling 30 fits.
- Best Model Parameters chosen by GridSearchCV: n_estimators = 250 & loss='huber'.
- MAE score of 300.3.
- RMSE score of 453.3.
- Standard Deviation for MAE : |absolute value of (y_test – predictions) = 325.36
- Confidence Interval: There is a 95% chance of getting an error from 298 to 332.



The End

Q & A?