Weather-Driven Predictions of Solar PV Energy using Machine Learning

By

Manav Sanghi (1913040)

Mriganka Dhar (1913060)

Arif Mohammad (1913039)

Under the guidance of

Asso. Prof. Dulal Chandra Das

In Partial Fulfilment of the Fourth year UG Project Work

Department of Electrical Engineering

NIT Silchar, Assam

Table of Contents

Introduction

- Climate change and energy crisis have motivated the use and development of sustainable energy sources.

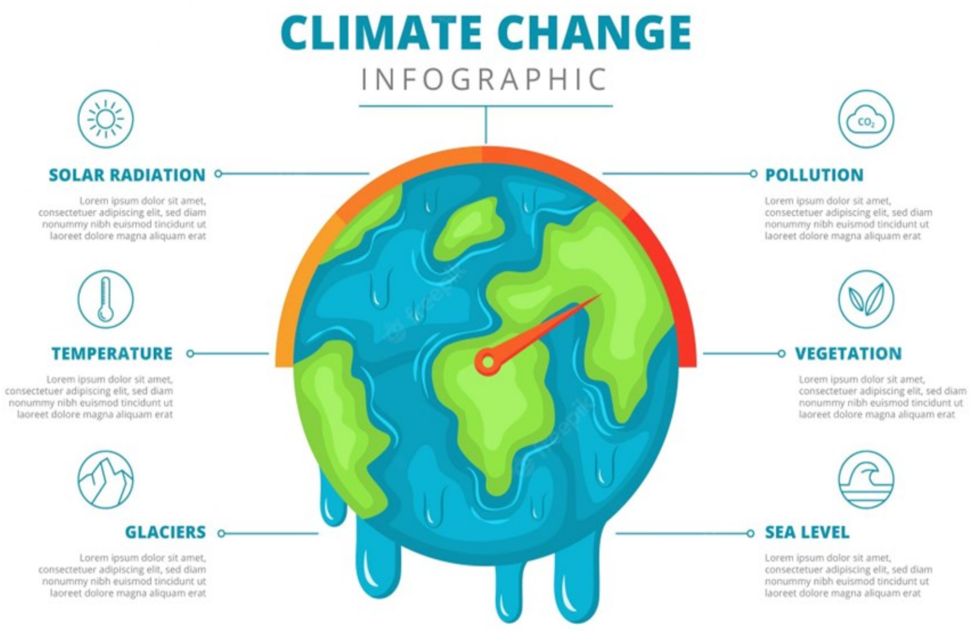


Fig 1 : Effects of climate change

Source: https://www.freepik.com/free-vector/flat-climate-change-infographic-template_17895350.html

Problems !!

- Renewable energy sources power generation is fully dependent on weather and meteorological parameters.

- The variable output power results in voltage and frequency fluctuations, which poses challenges on grid operation especially at high PV penetration rates.

- In order to cope with the power variability and maintain reliable grid operation, large amounts of costly storage facilities, balancing energy and/or frequency reserves are required.

Literature Review:

Sl. no	Paper Name	Author Name	Summary of The Work
1	Uncertainty energy planning of net-zero energy communities with peer-to-peer energy trading and green vehicle storage considering climate changes by 2050 with machine learning methods.	• Jia Liu • Yuekuan Zhou • Hongxing Yang • Huijun Wu	• Provides the necessary statical models like regression, distribution fitting and random sampling required to analyze the past dataset of weather and to predict the weather in future with the level of uncertainty ▪ The aleatory and epistemic uncertainties of climate changes on the key weather parameters determining solar PV and wind power generation. ▪ Development of PV and wind power generation models to study the P2P energy trading to achieve carbon neutrality by 2050.
2	Impact of climate change on electricity systems and markets – A review of models and forecasts	• Torben K. Mideksa • Steffen Kallbekken	• This review highlights the research areas where the knowledge has advanced significantly like: temperature sensitive demand estimation, localized impact of climate variations on renewable energy, etc. • The research areas where the impacts are still unclear includes: system-wide impact on renewable and power generation; economic modeling of extreme events ; the effects of climate, policy and impact uncertainties on modeling outcomes etc.

Sl. no	Paper Name	Author Name	Summary of The Work
3.	A systematic analysis of meteorological variables for PV output power estimation	• Soumyabrata Dev • Lennard Visser • Murhaf Hossari • Wilfried van Sark	• In this paper, we present a systematic approach to perform an analysis on different meteorological variables, namely temperature, dew point, temperature, relative humidity, visibility, air pressure, wind speed, cloud cover, wind bearing and precipitation, and assess their impact on PV output power estimation ▪ Analysis covers the correlation and interdependence among the meteorological variables. ▪ Machine learning-based regression methods, we identify the primary meteorological variables for PV output power estimation. ▪ Finally, the paper concludes that the impact of using a lower-dimensional subspace of meteorological variables per location, as input for the regression methods, results in a similar estimation accuracy in the two case studies.
4	Regression model in estimating solar radiation in Perlis.	• I. Daut • Y.M Irwan • M. Irwanto • Z. Farhana	• This study presents an investigation of a relationship between solar radiation and temperature in Perlis, Northern Malaysia for the year of 2006. • Since the scatter plots represent the straight line, the regression model was selected to estimate the solar radiation.
5	Predictive model for PV power generation using RNN (LSTM)	• Min Kyeong Park • Jong Man Lee • Won Hee Kang • Jong Min Choi	• In this study, a recurrent neural network (RNN) was utilized in predicting photovoltaic (PV) power generation. An RNN is an artificial neural network in which the connection between units is composed of a cyclic structure that can reflect the characteristics of time series. • This study uses previously measured weather data and PV power generation data to predict the future energy.

Objective

1. To identify which factors of weather effect the most for solar energy prediction.
2. To identify which ML algorithms are the most efficient and accurate in predicting weather-driven solar energy output .

Factors affecting PV power generation taken into consideration : -

Cloud Coverage

- The study of clouds, where they occur, and their characteristics, play a key role in the understanding of climate change. Low, thick clouds primarily reflect solar radiation and cool the surface of the Earth.
- Whether a given cloud will heat or cool the surface depends on several factors, including the cloud's altitude, its size, and the make-up of the particles that form the cloud.
- On average, solar panels will generate 10 to 25% of their normalpower output on days with heavy cloud coverage.

Table 9-10. Sky cover. Oktas=eighths of sky covered.

Sky Cover (oktas)	Sym-bol	Name	Abbr.	Sky Cover (tenths)
0		Sky **Clear**	SKC	0
1		**Few*** Clouds	FEW*	1
2				2 to 3
3		**Scattered**	SCT	4
4				5
5		**Broken**	BKN	6
6				7 to 8
7				9
8		**Overcast**	OVC	10
(9)		Sky **Obscured**		un-known
(/)		Not Measured		un-known

* "Few" is used for (0 oktas) < coverage ≤ (2 oktas).

Fig 2 : Sky coverage

Temperature

- Most of us would assume that stronger and hotter the sun is, the more electricity our solar panels will produce. But that's not the case. Although the temperature doesn't affect the amount of sunlight a solar cell receives, it does affect how much power is produced.

- Semiconductors are sensitive to temperature changes. Temperatures above the optimum levels decrease the open circuit voltage of solar cells and their power output, while colder temperatures increase the voltage of solar cells.

- According to the manufacture standards, 25 °C or 77 ° temperature indicates the peak of the optimum temperature range of photovoltaic solar panels.

.

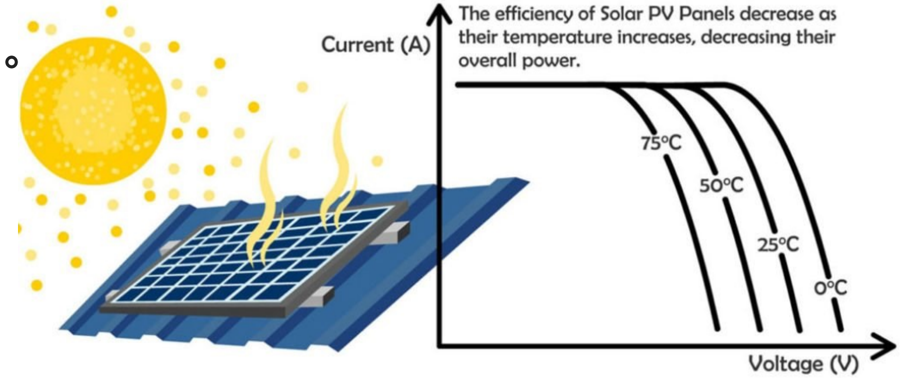


Fig 3 : Temperature effect on solar energy

Source: https://couleenergy.com/datasheet-values-rating-of-a-solar-panel/

Wind

- When a solar panel is too hot, it reduces efficiency due to the science behind a solar panel generating electricity. On the other hand, cooler solar panel temperatures improve efficiency.

- In short, the effect of temperature on solar cell performance is this: cooler panels allow more energy to get through like an electric current than hot panels do.

- Here’s where the wind comes in. The wind cools solar panels. Though it won’t make or break your solar panel production overall, it does make a difference. Solar panels cooled by 1 degree Celsius are 0.05 percent more efficient. This percentage adds up over time.

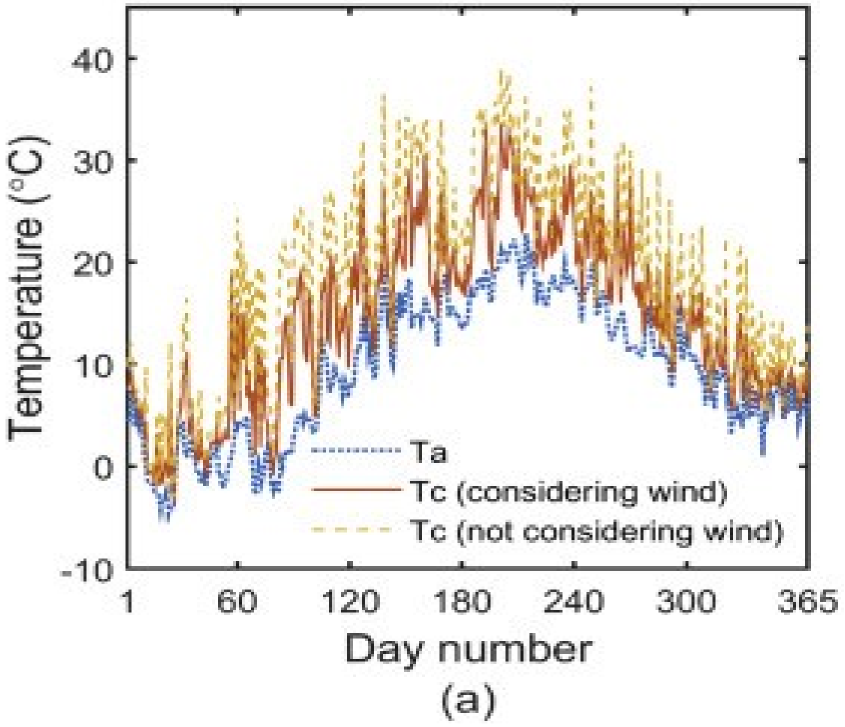


Fig 4 : Wind effect on solar energy

Source: https://www.sciencedirect.com/science/article/pii/S0960148116300179

Humidity

Humidity can slow efficiency in two ways.

1. Tiny water droplets, or water vapor, can collect on solar panels and reflect or refract sunlight away from solar cells. This reduces the amount of sunlight hitting them and producing electricity.
2. Consistent hot, humid weather can degrade the solar panels themselves over their lifetime. This is true for both crystalline silicon cells and thin film modules, but cadmium telluride (thin film) solar cells perform about 5 percent better in tropical climates.

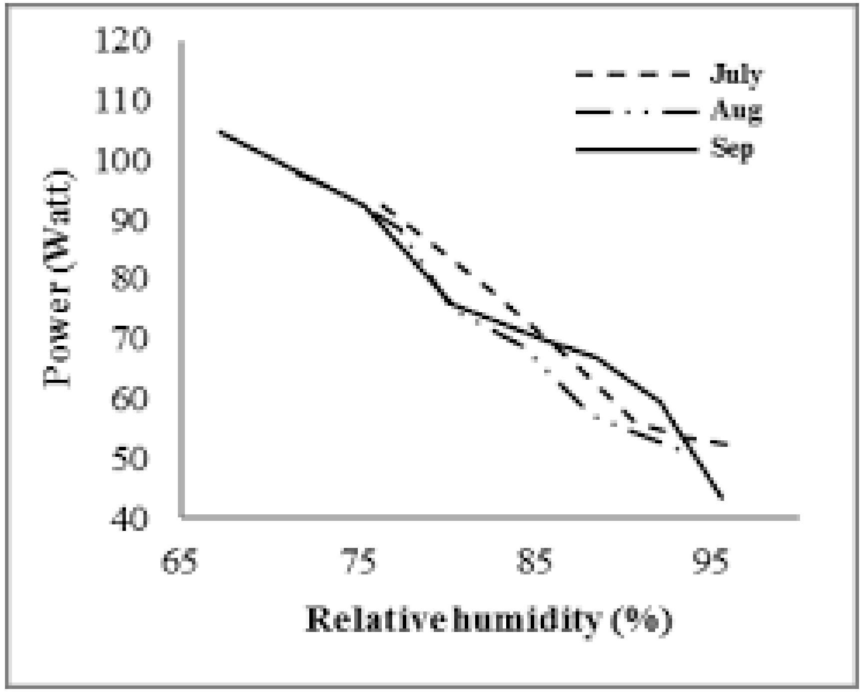


Fig 5 : Humidity effect on solar energy

Source: https://www.researchgate.net/publication/289546072_Effect_of_humidity_on_photovoltaic_performance_based_on_experimental_study

MACHINE LEARNING WORKFLOW

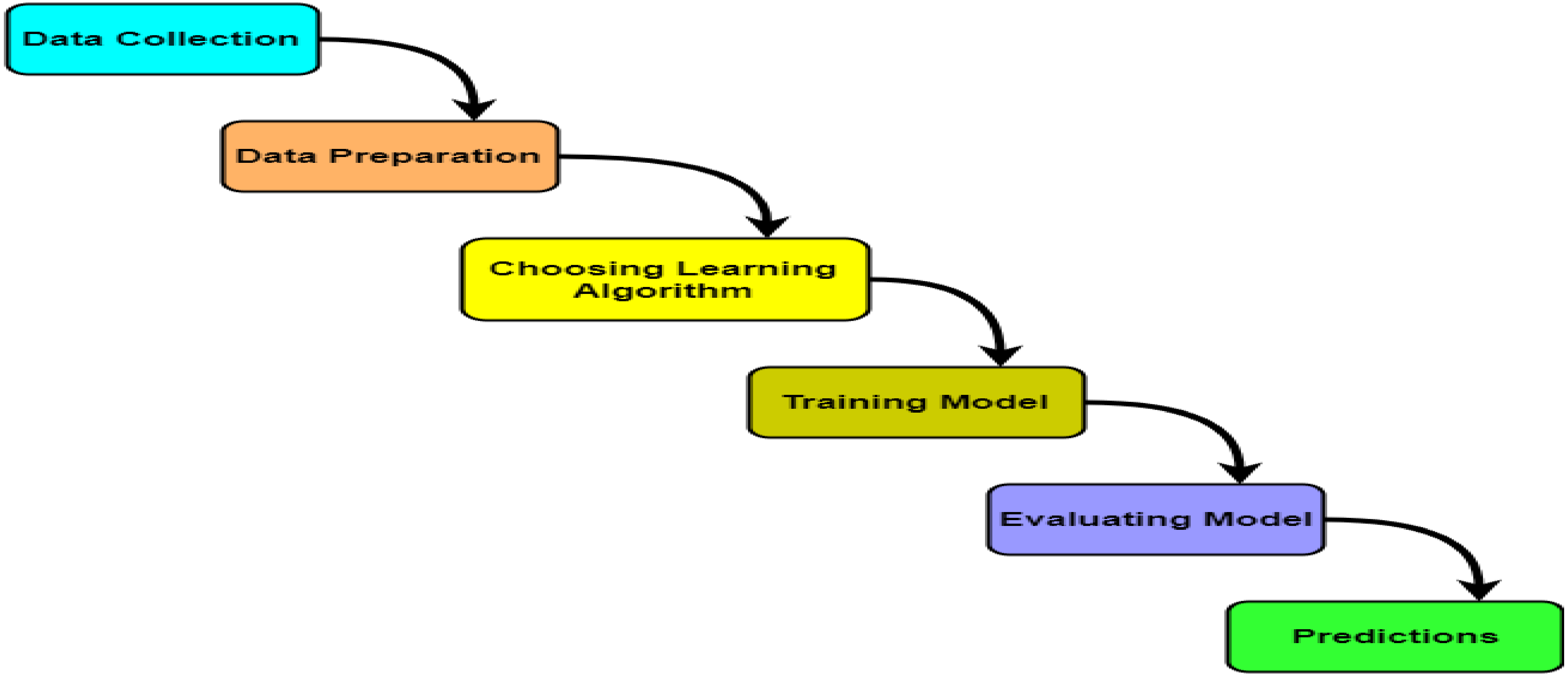


Fig 6 :ML Model flowchart

Source: https://www.gatevidyalay.com/tag/machine-learning-process-diagram/

Data Set :

Dataset:

-Power output from photovoltaic panels located at 12 Northern hemisphere sites of North America over 14 months

- Independent variables in each column include: location, date, time sampled, latitude, longitude, altitude, year and month, hour, season, humidity, ambient temperature, power output from the solar panel, wind speed.

	Location	Date	Time	Latitude	Longitude	Altitude	YRMODAHRMI	Month	Hour	Season	Humidity	AmbientTemp	PolyPwr	Wind.Speed	Visibility	Pressure	Cloud.Ceiling
0	Camp Murray	20171203	1145	47.11	-122.57	84	2.017120e+11	12	11	Winter	81.71997	12.86919	2.42769	5	10.0	1010.6	722
1	Camp Murray	20171203	1315	47.11	-122.57	84	2.017120e+11	12	13	Winter	96.64917	9.66415	2.46273	0	10.0	1011.3	23
2	Camp Murray	20171203	1330	47.11	-122.57	84	2.017120e+11	12	13	Winter	93.61572	15.44983	4.46836	5	10.0	1011.6	32
3	Camp Murray	20171204	1230	47.11	-122.57	84	2.017120e+11	12	12	Winter	77.21558	10.36659	1.65364	5	2.0	1024.4	6
4	Camp Murray	20171204	1415	47.11	-122.57	84	2.017120e+11	12	14	Winter	54.80347	16.85471	6.57939	3	3.0	1023.7	9

Table 1: Weather parameter

Source : https://www.kaggle.com/datasets/saurabhshahane/northern-hemisphere-horizontal-photovoltaic?resource=download

Libraries

```
[78] import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns

     from sklearn.model_selection import train_test_split, RandomizedSearchCV
     from sklearn.metrics import explained_variance_score, mean_absolute_error, mean_squared_error, mean_absolute_percentage_error, r2_score
     from sklearn.preprocessing import StandardScaler,MinMaxScaler
     from sklearn.neighbors import KNeighborsRegressor
     from sklearn.pipeline import Pipeline
```

Data Processing

Identifying and handling the missing values

- Identifying missing values is an important part of data pre-processing.

- Otherwise, we may draw inaccurate and false inferences from our data.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21045 entries, 0 to 21044
Data columns (total 17 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Location       21045 non-null  object
 1   Date           21045 non-null  int64
 2   Time           21045 non-null  int64
 3   Latitude       21045 non-null  float64
 4   Longitude      21045 non-null  float64
 5   Altitude       21045 non-null  int64
 6   YRMODAHRMI     21045 non-null  float64
 7   Month          21045 non-null  int64
 8   Hour           21045 non-null  int64
 9   Season         21045 non-null  object
 10  Humidity       21045 non-null  float64
 11  AmbientTemp    21045 non-null  float64
 12  PolyPwr        21045 non-null  float64
 13  Wind.Speed     21045 non-null  int64
 14  Visibility     21045 non-null  float64
 15  Pressure       21045 non-null  float64
 16  Cloud.Ceiling  21045 non-null  int64
dtypes: float64(8), int64(7), object(2)
memory usage: 2.7+ MB
```

Feature Engineering:

- Cyclic features

```
df['Sine_Month'] = np.sin((df.Month - 1)*np.pi/11)
df['Cos_Month'] = np.cos((df.Month -1)*np.pi/11)
df['Sine_Hour'] = np.sin(df.Delta_Hour*np.pi/(max_hour - min_hour))
df['Cos_Hour'] = np.cos(df.Delta_Hour*np.pi/(max_hour - min_hour))
df.head()
```

Sine_Month	Cos_Month	Sine_Hour	Cos_Hour
5.665539e-16	-1.0	0.587785	0.809017
5.665539e-16	-1.0	0.951057	-0.309017
5.665539e-16	-1.0	0.951057	-0.309017
5.665539e-16	-1.0	0.951057	0.309017
5.665539e-16	-1.0	0.587785	-0.809017

Encoding the categorical data:

The machine learning model does not understand the categorical data so we need to encode it.

Types of encoding:-

1. Label Encoding
2. One Hot Encoding

Location_MNANG	Location_Malmstrom	Location_March AFB	Location_Offutt	Location_Peterson	Location_Travis	Location_USAFA	Season_Spring	Season_Summer	Season_Winter
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	1

Co- Relation Analysis

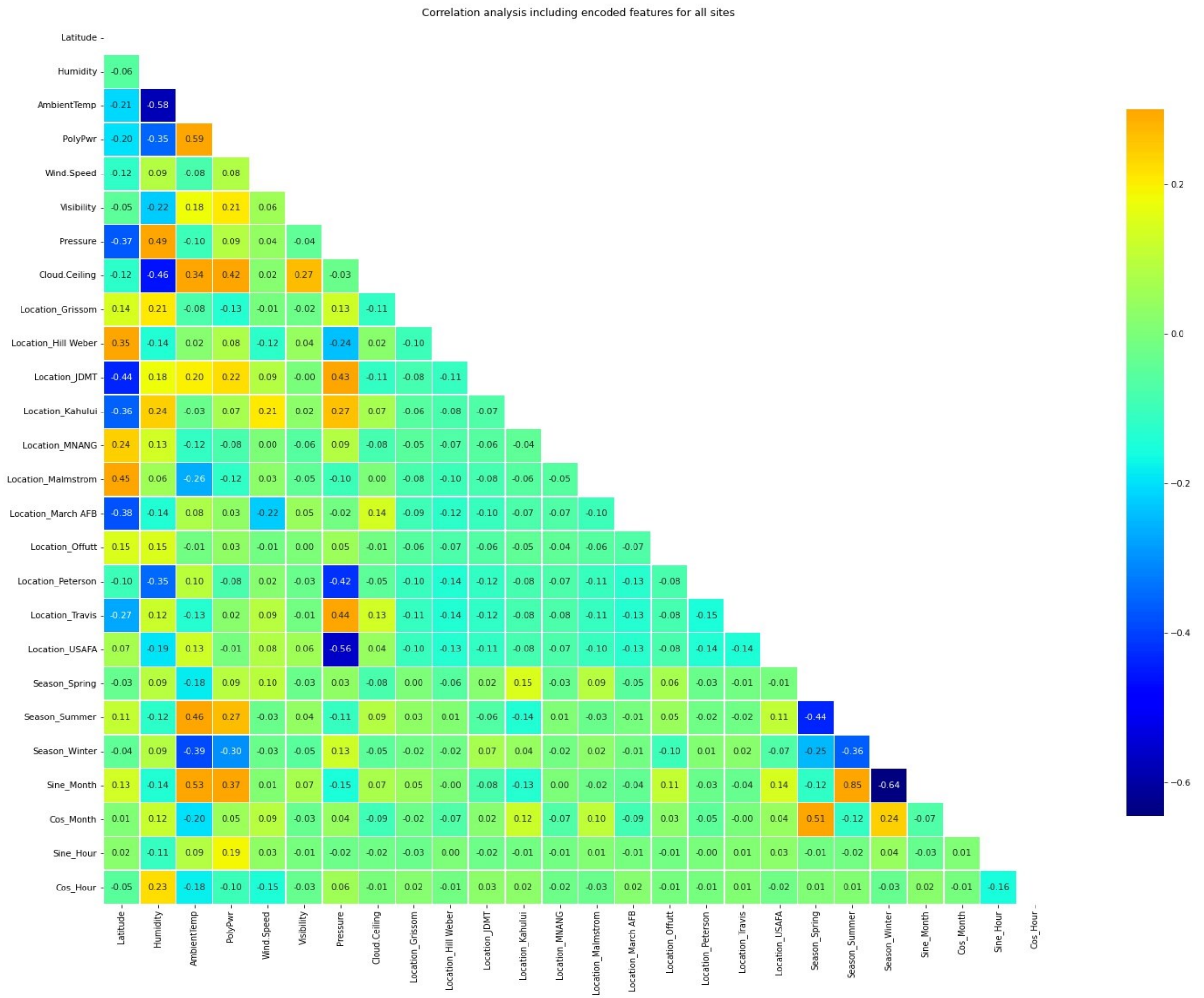


Fig 7 : Heat map for co-relation analysis

Summary of Feature selection

1. Altitude is dropped because it has a high correlation with Pressure and it does not change for a given location while pressure is more dynamic.

2. Longitude is dropped because it has zero correlation with the target variable.

3. Time, Hour, Month and Date are dropped because they have strong correlations with the engineered cyclic features but low correlation with the target variable.

Model Algorithm – KNN Regression

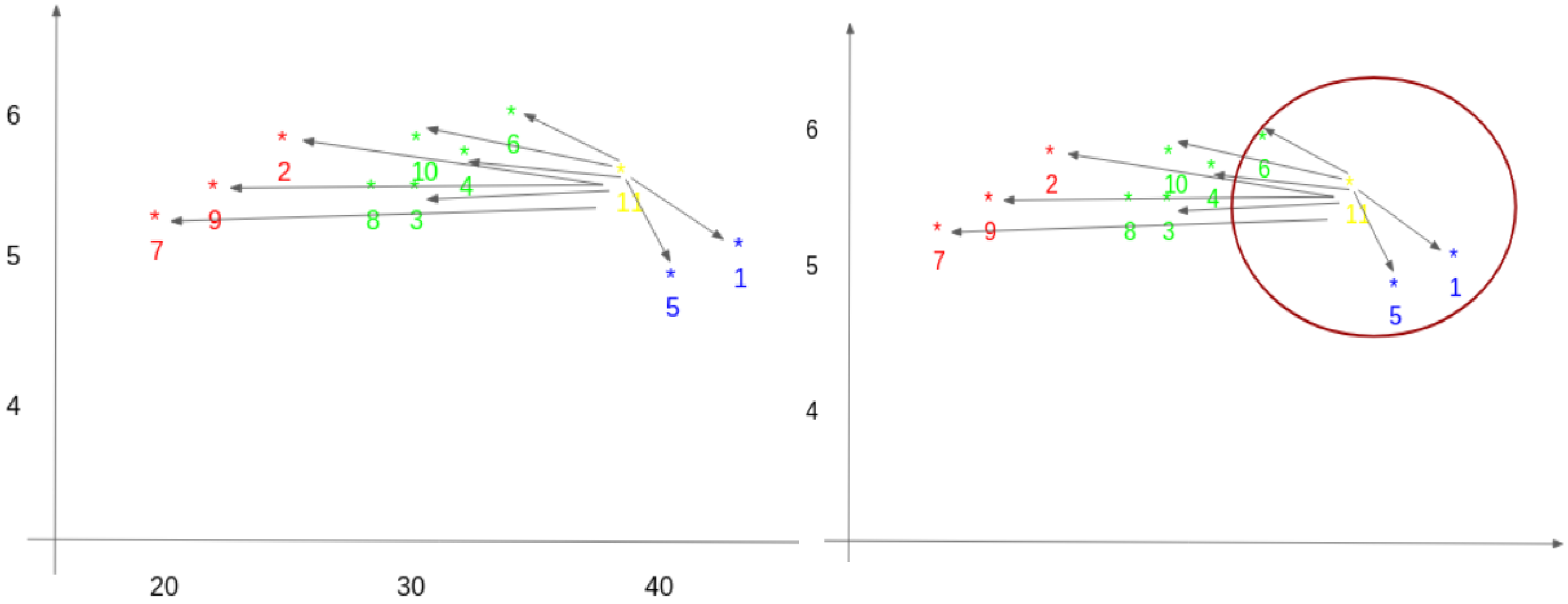


Fig 8 : KNN algorithm working

Source: https://www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/

Parameter selection for model

- n_neighbors
- algorithm : auto , ball_tree , kd_tree , brute
- leaf_size : default=30
- p : default=2

```
# Get optimal hyper-params
knn_random.best_params_
```

```
{'knn__weights': 'distance',
 'knn__p': 2,
 'knn__n_neighbors': 15,
 'knn__leaf_size': 30,
 'knn__algorithm': 'ball_tree'}
```

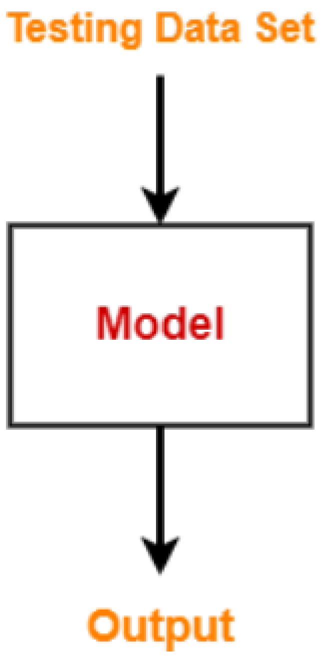
Train & Test model using optimal hyperparameters


```python
knn_model = knn_pipeline.set_params(**knn_random.best_params_)
```

```python
knn_model.fit(X_train, y_train)
```

Make model inferences on test set

```python
y_pred = knn_model.predict(X_test)
```



RESULTS :

After feeding the data to the model and then testing it we got the following scores -

- **Root Mean Squared Error**

  4.40285342242806

- **Mean Absolute Error**

  2.970990850370644

- **R2 Score**

  0.6184069962937428

Conclusion and Future Work

- ML Model was formed with MAE of 29% and R2 square of 62%.
- We can apply the same data set and features for neural network model.
- Compare the two models .
- Determine the best fit model from the two.

References

1. The impact of climate change on the electricity market: (Torben K. Mideksa, Steffen Kallbekken)
2. Impact of climate change on electricity systems and markets – models and forecasts Shankar N. Chandramowli , Frank A. Felder
3. Tackling Climate Change with Machine Learning DAVID ROLNICK, McGill University and MilaQuebec AI Institute PRIYA L. DONTI, Carnegie Mellon University
4. Linear Regression Model in Estimating Solar Radiation inPerlisS. Ibrahimb, I.Dauta, Y.M. Irwan , M. Irwantoa, N. Gomesha, Z. Farhanaa
5. Predictive model for PV power generation using RNN (LSTM) Min Kyeong Park1 , Jong Man Lee1 , Won Hee Kang1 , Jong Min Choi2 and Kwang Ho Lee3

6. C. Chen, S. Duan, T. Cai, B. Liu, Online 24-h solar power forecasting based on weather type classification using artificial neural network, Sol. Energy 85 (11) (2011).

7. Van der Meer, W. Dennis, Joakim Widen, M. Joakim, Review on probabilistic forecasting of photovoltaic power production and electricity consumption, Renew. Sustain. Energy Rev. 81 (2018)

8. M.P. Almeida, O. Perpi~nan, L. Narvarte, PV power forecast using a nonpara- metric PV model, Sol. Energy 115 (2015)

9. A. Mellit, S.A. Kalogirou, S. Shaari, H. Salhi, A.H. Arab, " Methodology for Predicting Sequences of Mean MonthlyClearness Index and Daily Solar radiation Data in Remote Areas: Application for Sizing a Standalone PV System", Renewable Energy, Science Direct, pp. 1570 – 1590, 2007

10. S.Ibrahim ,Daut Y.M.Irwan , M.Irwanto , N.Gomesh , Z.Farhana -Linear Regression Model in Estimating Solar Radiation in Perlis - Energy Procedia 18 (2012)

11. L. Visser, T. AlSkaif, W. van Sark, Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions, in: IEEE 46th Photovoltaic Specialist Conference (PVSC), IEEE, 2019