TITLE - SOLAR RESOURCE ASSESSMENT USING MACHINE LEARNING

Solar power generation forecasting using BiLSTM, SVR and DNN models.

Vishwesh Ingale

AIM: To predict the solar power generation capacity of a location using Machine Learning algorithms and statistical methods.

OBJECTIVES:

1) To study the given dataset

2) To build appropriate ML model(s) and train them using the dataset

3) To predict DNI and GHI values using the trained ML model(s)

WORK COMPLETED AS PER OBJECTIVES:

1) Based on the dataset, three ML models were created and trained using these techniques: Deep Neural Network (DNN), Support Vector Regression (SVR) and Bidirectional Long Short-Term Memory (BiLSTM).
2) DNI values were predicted via all 3 models and then using these predicted values, GHI values were predicted.

ABSTRACT

Solar radiation can be used for producing heat, causing chemical reactions, or generating electricity. Therefore, the amount of solar radiation at different times of the day must be determined to design and equip all solar systems. In Addition to that, it is necessary to go through the thorough understanding of different solar radiation components, such as Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), and Global Horizontal Irradiance (GHI).

We utilized various machine learning (ML) models for this purpose. This paper aims to develop a set of deep learning models through feature importance algorithms to predict the DNI data. The proposed models are based on historical data on meteorological parameters and solar radiation properties in a specific location in the region of Nagpur, Maharashtra, from June 18, 2019, to June 18, 2020. These predictive models play a crucial role in optimizing solar energy systems by providing actionable insights into solar energy potentials and system enhancements. Through the utilization of these data methodologies, stakeholders in the renewable energy sector can make well-informed decisions, significantly contribute to sustainable energy solutions, and pave the way for a greener future.

INTRODUCTION:

Solar Height angle

Factors like cloud cover, air pollution and the atmospheric conditions influence solar irradiance falling on a solar cell's surface. However, there is another important factor that influences it: angle of incidence of sunlight. It is the angle which depends on height of the sun above horizon. This height is given by the solar height angle, also called the elevation angle. This angle measures the angular position of the sun in the sky relative to horizon. The altitude angle varies throughout the day and is influenced by latitude of the observer and the day of the year.

As shown in Fig 1, the altitude angle is zero at sunrise and sunset, which results in almost no radiation to the sun. If atmospheric conditions do not change as the day progresses, the angle of elevation increases approximately in line with the sine curve. This leads to an increase in solar irradiance. At noon, when the sun seems to be reaching its highest point in the sky, it usually attains a peak solar radiation. This results in the progressive reduction of the angle of elevation between noon and sunset, resulting in an overall reduced solar irradiance. The altitude angle θ depends on the observer's latitude, time of day, and day of year, and it can be computed using the formula:

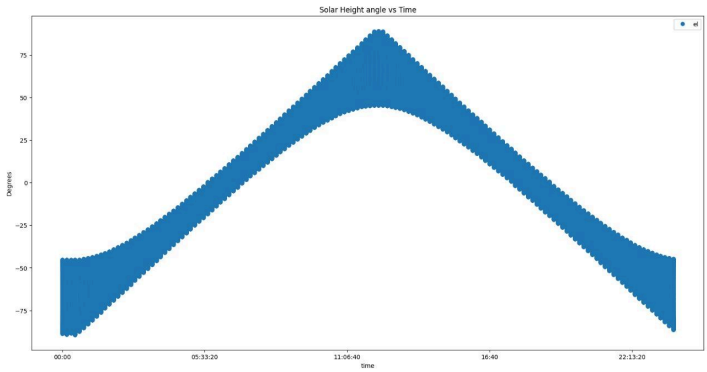
$\theta = \sin^{-1}((\sin \delta)*(\sin \phi) + (\cos \delta)(\cos \phi)*(\cos \omega))$

where δ = declination angle, ϕ = latitude angle and

ω = hour angle

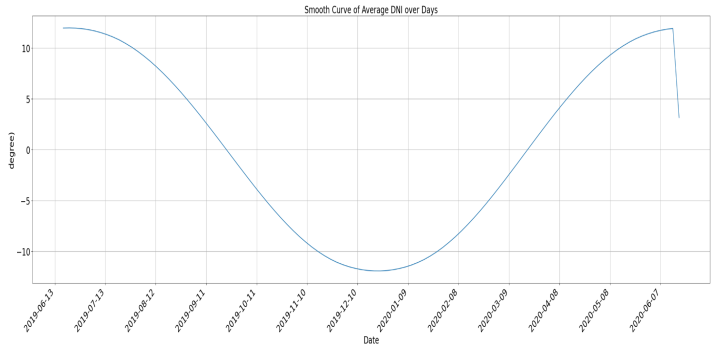
Axial tilt of earth is approximately 23.5 degrees concerning its orbit around the sun. This is the cause of the sinusoidal variation in the solar height angle for the year, as shown in Fig.2. Its lowest point is in January and the highest is in June. The radiation strikes Earth at a lower angle in January when the north hemisphere moves away from the sun, which results in reduced height angle of the sun, shortened days, and lighter sunlight. On the other hand, in June, the Northern Hemisphere tilts towards the sun, which results in a

higher solar height angle, longer days, and more direct sunlight. Throughout the year, this cyclical pattern influences weather patterns, climate, and the amount of sunlight reaching Earth's surface.



Graph 1. Solar Height Angle Vs Time

Graph 1 shows the variation in solar height angle concerning the time of a day. The time is plotted in HH:MM:SS format on the X-axis, where H represents hour, M represents minute and S represents seconds. The solar height angle is plotted on the Y-axis in degrees.



Graph 2. Solar Height Angle Vs Date

This graph shows the variation of solar height angle (which shows the angular position of the sun above the horizon) with the dates of a Gregorian calendar. The dates are plotted on the X-axis, ranging from 18 June 2019 to 18 June 2020. The solar height angle is plotted on the Y-axis, in degrees, and ranges from -20 degrees to +20 degrees.

Solar Azimuth angle [3]

The solar azimuth angle (S) is the angle made in the horizontal plane between the line due south and the projection of the line of sight of the sun on the horizontal plane. Thus, it gives the direction of the

shadow of a vertical rod, cast in a horizontal plane. By convention, it is typically measured from true north, with 0 degrees representing north, 90 degrees representing east, 180 degrees representing south, and 270 degrees representing west. It is taken to be positive if the projection of the line of sight is east of south and negative if west of south. The expression used for the calculation of this angle is

$\cos \gamma = (\cos\theta_z \sin\varphi - \sin\beta)/\sin\theta_z \cos\varphi$

Direct Solar Irradiance

According to National Renewable Energy Laboratory (NREL), “Direct irradiance is quotient of radiant flux on a given plane receiver surface received from small solid angle centered on sun's disk to the area of that surface.” The intensity of the sun depends mainly on the position of the sun in the sky with respect to the observer on Earth's surface. At higher zenith angles, light goes through more atmosphere as compared to when the sun is directly overhead. It is calculated using the steps described in the process below.

The fraction of direct solar radiation that passes through the atmosphere and reaches the surface of the Earth without being scattered is known as solar beam direct transmittance. It is calculated using the optical properties of atmosphere and the path length of solar radiation.

Steps to calculate solar direct beam transmittance:

1. Calculate Solar Zenith Angle (θ):

The solar zenith angle (θ) is the angle between the zenith (a point directly overhead) and the position of the sun. It can be calculated using the solar height angle:

$\theta_z = 90^\circ - \theta$

2. Calculate Optical Air Mass (AM):

The optical air mass (AM) quantifies the path length of solar radiation through the atmosphere. It can be calculated from the solar zenith angle using an approximation formula such as the Kasten-Young Air Mass Model:

$$AM = \frac{1}{cos(\theta z)}$$

3. Calculate Extraterrestrial Radiation:

This represents the solar irradiance reaching the top of the Earth's atmosphere on a surface perpendicular to the solar beam or the radiation that reaches the outer part of Earth's atmosphere, which varies slightly throughout the year. To account for the eccentricity of the Earth's orbit around the sun, the extraterrestrial radiation is calculated with a yearly varying term [5] :

$$I_0 = 1367.7\left(1 + 0.033cos\left(\frac{2\pi}{365} * DOY\right)\right)$$

4. Calculating DNI

There are two very simple models for determining components of GHI (i.e., DNI) using only the zenith angle. Two of these models are

Meinel Model (1976) [6]

$DNI = I_0 \times 0.7^{AM\ \hat{}0.678}$

Laue Model (1970) [7]

$DNI = I_0 \times ((1-0.14\times h)\times 0.7^{AM\hat{}0.678} + 0.14\times h)$

transmittance.

Diffuse Horizontal Irradiance

The sunlight that is scattered in the atmosphere is scattered in all directions. A part of this radiation is redirected towards Earth and is called diffuse irradiance. During overcast days, the solar power comes almost completely from this component. Diffuse irradiance also includes reflections from the ground, which depends on the surface albedo. There is a prominent increase in this value when there is snow.

As per Daneshyar–Paltridge–Proctor (DPP) model (1978) [8, 9]:

Diffuse = 14.29 +21.04(π/2- $z\pi$/180);

Graph 3 depicts the fluctuation in Direct Normal Irradiance (DNI) throughout a year together with variations in solar radiation intensity. The y-axis of the graph indicates DNI levels in measures as watts per square meter (W/m2), while the x-axis represents time in months. During the winter months, DNI variance appears to be smaller. This drop can be attributed to winter conditions, which include higher clouds and lower sun elevation angles. These differences highlight how essential it is to take atmospheric conditions into account when calculating solar radiation statistics for the production of energy and the assessment of resources. It is crucial to comprehend how solar radiation and atmospheric phenomena interact in order to accurately measure performance and make decisions on the use of solar energy.

Global Horizontal Irradiance

The total solar radiation on a horizontal surface is called global horizontal irradiance (GHI). GHI is the sum of the diffuse radiation incident on a horizontal surface plus the direct normal irradiance projected onto the horizontal surface

$$GHI = Diffuse + DNIcos(z)$$

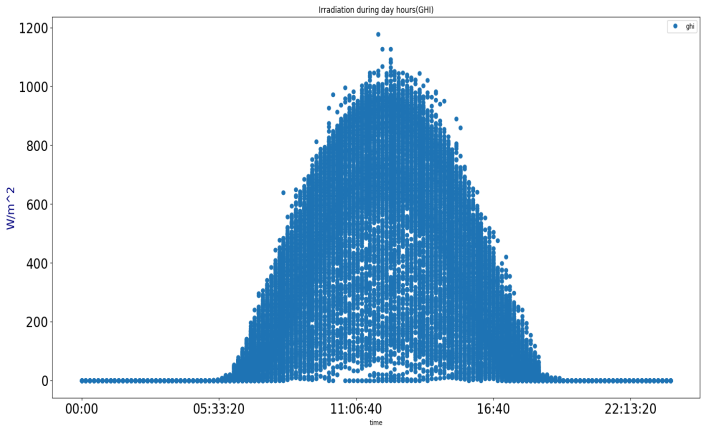
where z is the solar zenith angle. Figure 2 depicts the variation in Global Horizontal Irradiance (GHI) over the span of a year. The x-axis represents time, likely in months, while the y-axis represents GHI values in units such as watts per square meter (W/m²).

The graph illustrates distinct fluctuations in GHI throughout the year, with noticeable differences particularly pronounced during the transition between seasons. During the winter months, the variation in GHI appears to be less pronounced compared to other seasons. This phenomenon is characteristic of regions experiencing winter conditions, where the sun's angle relative to the Earth's surface tends to be lower, resulting in reduced solar radiation intensity.

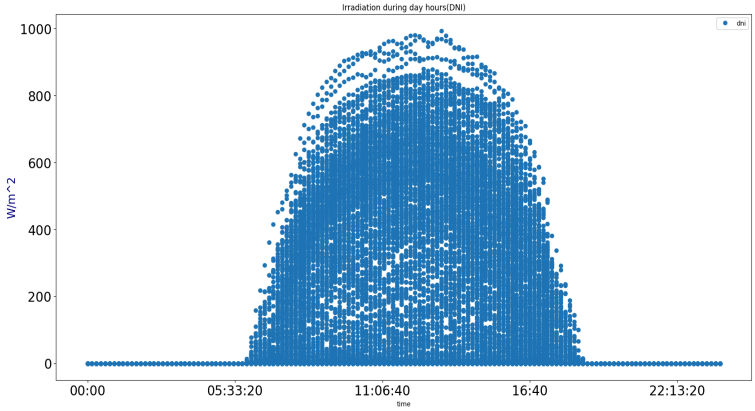
The reduced variability in GHI during winter months can be attributed to several factors, including shorter daylight hours, lower solar elevation angles, and increased cloud cover or atmospheric attenuation.

These factors collectively contribute to a more consistent GHI pattern during winter, with lower overall irradiance levels compared to other seasons.

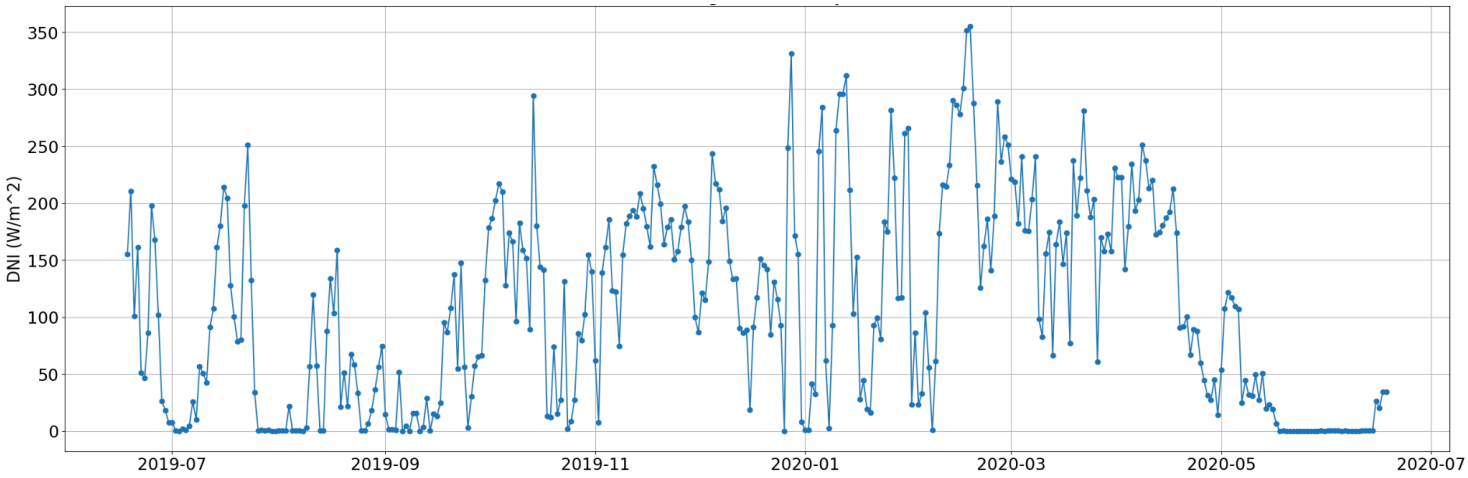
Reflecting the sinusoidal fluctuation inherent in solar radiation dynamics, Figures 3a. , 3c. and 3b. show a similar diurnal pattern over a day for both Global Horizontal Irradiance (GHI) and Direct Normal Irradiance (DNI). The time is shown by the x-axis, which is usually divided into hourly segments, and the irradiance values are represented by the y-axis, which is expressed in watts per square meter (W/m²). When the sun is above the horizon during the day, both GHI and DNI achieve their greatest levels. During the night, when solar radiation is absent, these values progressively decrease to zero. At noon, when the sun is at its highest elevation angle and the irradiance is at its maximum, the curves reach their zenith. This periodic variation highlights the need to take the time of day into account in solar energy applications and the fluctuations in solar radiation availability.



Graph 3b. GHI(W/m^2) V/s time



Graph 3a. DNI(W/m^2) V/s Months



Graph 3c. DNI(W/m^2) V/s Months

DATASET

	#date time	el	az	ghi	wb	precip	press	dp	rh	dhi_flag	dni	ws	maxws	airtemp
0	2019-06-18 00:00:00	-45.2	353.6	0	24.2	0.0	967	20.3	47	10004	0	1.5	3.5	32.9
1	2019-06-18 00:10:00	-45.4	356.8	0	24.2	0.0	967	20.4	47	10004	0	1.5	3.3	32.8
2	2019-06-18 00:20:00	-45.5	180.1	0	24.3	0.0	967	20.6	48	10004	0	2.2	4.4	32.6
3	2019-06-18 00:30:00	-45.4	3.4	0	24.3	0.0	967	20.7	49	10004	0	1.9	5.3	32.6
4	2019-06-18 00:40:00	-45.2	6.6	0	24.4	0.0	967	20.8	49	10004	0	1.4	2.8	32.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
52767	2020-06-18 10:30:00	64.5	79.7	750	26.0	0.0	966	24.4	69	10000	224	2.0	7.2	30.5
52768	2020-06-18 10:40:00	66.8	79.6	636	26.0	0.0	966	24.5	69	10000	100	1.9	5.9	30.5
52769	2020-06-18 10:50:00	69.1	79.5	532	25.9	0.0	966	24.4	70	10000	9	2.2	6.2	30.4
52770	2020-06-18 11:00:00	71.4	79.2	847	25.9	0.0	967	24.4	70	10000	237	2.6	7.3	30.3
52771	2020-06-18 11:10:00	73.7	78.7	896	26.3	0.0	966	24.7	68	10000	314	1.7	5.2	31.1

Fig 1. Dataset

This is the dataset collected by Nagpur University at intervals of 10 minutes of a year (from 6/18/2019 0:00 to 6/18/2020 11:10). It consists of the following factors:-

1. Sun azimuth angle (Degrees).

2. Direct normal irradiance (DNI) (W/m²).

3. Diffuse horizontal irradiance (DHI) (W/m²).

4. Horizontal wind speed (10m) (m/s).

5. Max. Horizontal wind speed (10m): This is the maximum horizontal wind speed observed at the specified height of 10 meters during a specific period.

6. Wind direction (degrees clockwise from north (0°) to 360°.)

7. Air temperature (°C)

8. Relative humidity

9. Barometric pressure

10. Precipitation

11. Dew point temperature (°C)

12. Wet bulb temperature (°C)

13. DateTime

As these parameters affect the availability and intensity of solar radiation as well as factors including temperature, humidity, or wind speed that are also important in affecting the performance of photovoltaic panels, and can be a valuable source of information to predict how much energy will be produced.

METHODOLOGY AND VALIDATION:

Data loading and processing: The process starts with the solar dataset from Nagpur University, named as 'Solar Dataset_Solar' from MEDA-1.XLSX[] (see fig 1). We have used Pandas, a robust Python library, for data manipulation and analysis. Some columns marked as flags or indicators, such as 'ghi_flag', 'dni_flag', 'dhi_flag', etc., are deleted to simplify the dataset. To ensure data integrity, rows containing missing values (NAN) are systematically deleted. The missing values are replaced with 0 to avoid any inconsistency in the dataset.

Exploratory Data Analysis (EDA): Basic information about the dataset is introduced via info() which provides a simple summary of each column's nonnull acquisition counts and data types. Using the astype() function, some features such as 'dni' - Direct Normal Irradiance, 'ws' - horizontal wind speed and 'rh' - relative humidity, are converted to float data type to facilitate numerical analysis.

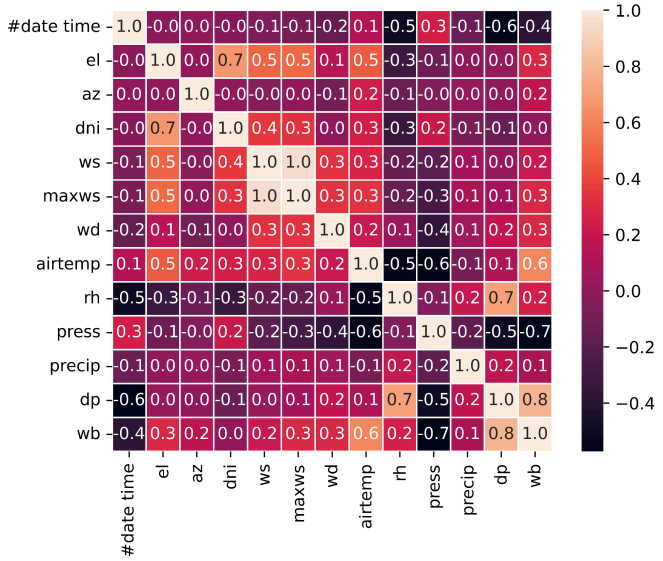


Fig 2. Heatmap

To get relationships between pairs of variables, Spearman correlation coefficients are calculated. We created a correlation matrix using the Pandas corr() function which records the correlation between all pairs of columns in the dataset. A heatmap visualizer using Seaborn is created to assist in the identification of

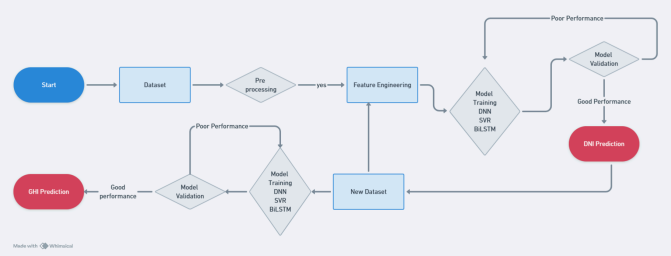


Fig 3a. Methodology WorkFlow

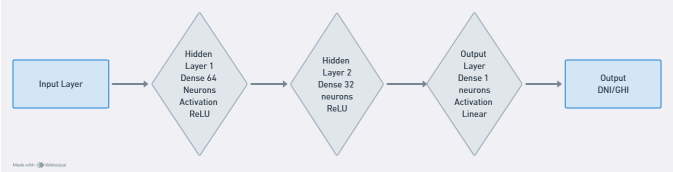


Fig 3b. DNN Model Architecture

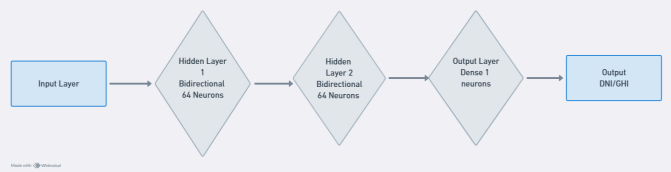


Fig 3c. BILSTM Model Architecture

patterns and relations between variables as shown in Fig 2. Particular pairs having a correlation coefficient above 0.5 are highlighted, indicating a moderately positive correlation.

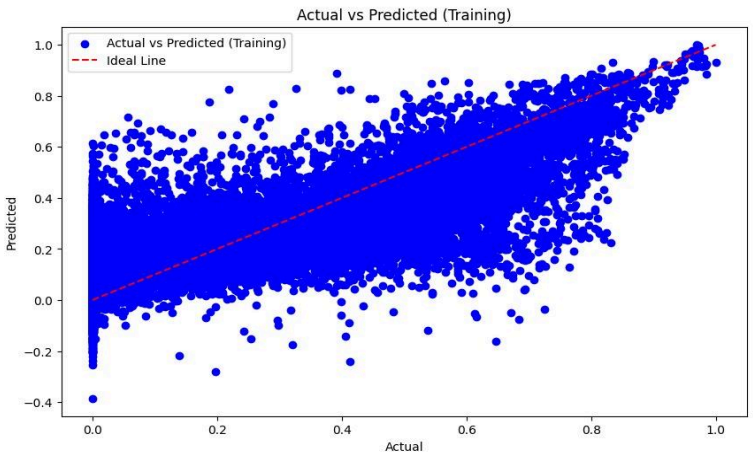
To predict direct normal irradiation DNI, the set of features is carefully prepared for the machine learning models. It involves selecting important features for training the model, such as 'el' - solar height angle, 'az' - azimuth angle, 'wb' - wet bulb temperature, 'precip' - precipitation, 'press' - barometric pressure, 'dp' - dew point temperature, 'rh' - relative humidity, 'air temp' - air temperature, 'ws' - wind speed, and 'maxes' - max horizontal wind speed. The target variable 'dni' is scaled into a normalized range (0 to 1) for better accuracy.

DNN Model Training: The DNN model is built and trained using TensorFlow and Keras. The model architecture includes three dense layers: two hidden layers comprising 64 and 32 neurons with ReLU activation functions, and an output layer featuring a linear activation function as shown in Figure 3(b). Root Mean Squared Error (RMSE) serves as the metric to evaluate the model's performance and the Adam optimizer is characterized by a learning rate of 0.002. The model is trained for over 50 epochs with a batch size of 32, during which the loss decreases from 0.1339 to 0.0839 for the training set and from 0.1121 to 0.0870 for the validation set, indicating effective learning and generalization.

BiLSTM Model Training: A Bidirectional Long Short-Term Memory (BiLSTM) model is built and trained using TensorFlow and Keras. The model architecture incorporates two BiLSTM layers, each comprising 64 and 32 neurons with ReLU activation functions as shown in Figure 3(c). This bidirectional structure captures both forward and backward temporal dependencies in the data. Mean Squared Error (MSE) is designated as the metric for evaluating the model's performance and the Adam optimizer, with a learning rate set at 0.002. The model was tuned using hyperparameters and is trained for over 50 epochs with a batch size of 32 (Graph 4, 5). The model subsequently improved in accuracy and the variation narrowed in the later epochs during which the loss decreased from 0.0171 to 0.0070 for the training set and from 0.0144 to 0.0081 for the validation set, indicating effective learning and generalization.

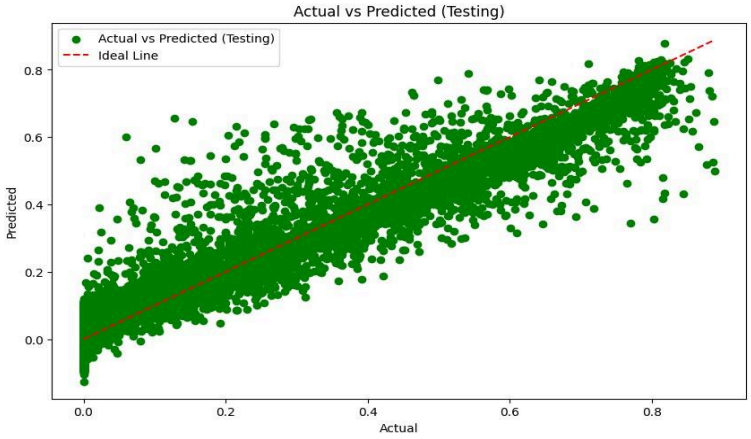
SVR Model Training: A different approach is explored utilizing the Support Vector Regression (SVR) model with a radial basis function (RBF) kernel and default parameters (C=1.0, epsilon=0.1), fitted to the scaled training data. The SVR model results in a training score of 0.9245 and a testing score of 0.9226 indicating high consistency in the dataset. This performance is supplemented with the DNN model trained for over 50 epochs with a batch size of 32. The loss decreases from 0.0171 to 0.0070 for the training set and from 0.0144 to 0.0081 for the validation set indicating promising performance.

Graph 4. Predicted V/s Actual for Training Dataset

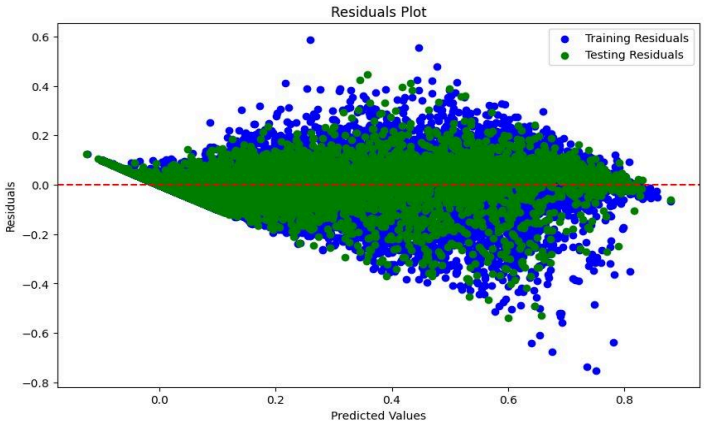


GHI Prediction: similar feature engineering methods are used to forecast GHI where DNI along with other features are fed to models such as DNN, SVR, and BiLSTM to predict the Global Horizontal Irradiance (GHI) values.

RESULTS AND DISCUSSION:

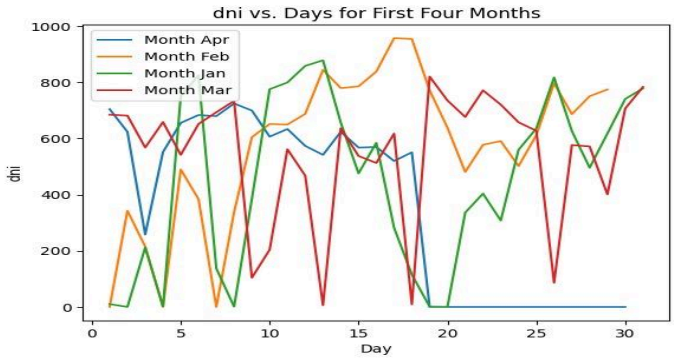


Graph 5. Predicted V/s Actual for Test Dataset



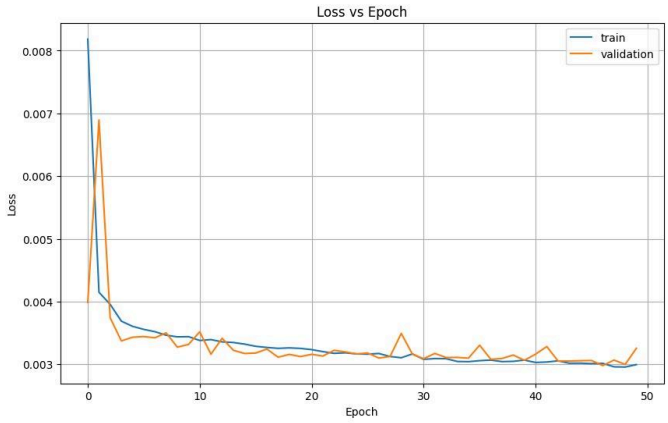
Graph 6. Residual Plot V/s Predicted Values for Test and Train Dataset

This graph 6 illustrates outliers for certain values that deviate from the actual values which signify that the model is not doing well in unusual or extreme data points and can be accounted for by the inconsistencies present in the data used. Patterns in the graph indicate very little to less bias in the predicted values.



Graph 7. Direct Normal Irradiance V/s Days of Month

The graph 7 shows the fluctuations in Direct Normal Irradiance (DNI) across the four months, showing variations in solar irradiance levels. The average value of DNI was higher for months March and April which indicates higher predicted power output.



Graph 8. Loss V/s No. of Epochs for DNN Model Prediction on Test and Train Dataset.

The graph 8 shows the convergence behavior of the model indicating its improving predictive performance. We observed a minimal gap between the training and test loss curves indicating its good generalization ability over unseen data. The Peaky nature of the validation set indicates that the model is consistently adjusting the parameters to reach smoothness and in hand minimization of loss.

CONCLUSIONS:

The experimental prediction of DNI and GHI values using three different machine learning models was done. The losses (deviation of actual values from theoretical values of GHI and DNI) were noted and conclusions are as follows:

- The BiLSTM model took the most amount of processing time and gave great results. It reduced the losses from 0.0171 to 0.0070 for the training set and from 0.0144 to 0.0081 for the validation set.
- The DNN model, working on the RMSE basis, decreased the losses from 0.1339 to 0.0839 for the training set and from 0.1121 to 0.0870 for the validation set.
- The SVR model works on the radial basis function (RBF) kernel. It resulted in the reduction of losses from 0.0171 to 0.0070 for the training set and from 0.0144 to 0.0081 for the validation set.
- The SVR model results in a training score of 0.9245 and a testing score of 0.9226 and is the most efficient model out of the 3 models used.

Future Work Plan:

1. Predicting GHI and DNI values with other machine learning methods such as ARIMA, SARIMA, and Facebook Prophet and comparing their performance with our models.
2. Calculation of actual power generation capacity using DNI and GHI values using different mathematical models of predicting solar power.

REFERENCES:

1. Herrería-Alonso S, Suárez-González A, Rodríguez-Pérez M, Rodríguez-Rubio RF, López-García C. A Solar Altitude Angle Model for Efficient Solar Energy Predictions. Sensors (Basel). 2020 Mar 4;20(5):1391. doi: 10.3390/s20051391 PMID: 32143294; PMCID: PMC7085731.
2. Amar, S., Bahich, M., Bentahar, Y., Afifi, M. and Barj, E., 2021. A study of the temperature influence on different parameters of mono-crystalline silicon photovoltaic modules. *Journal of Power and Energy Engineering*, *9*(6), pp.29-42.
3. S P Sukhatme, Solar Energy Principles of Thermal Collection and Storage, pp.76
4. A. B. Meinel and M. P. Meinel, Applied solar energy. Reading, MA: Addison-Wesley Publishing Co., 1976.
5. R. G. Allen, Environmental, and E. Water Resources Institute. Task Committee on Standardization of Reference, The ASCE standardized reference evapotranspiration equation. Reston, Va.: American Society of Civil Engineers, 2005
6. A. B. Meinel and M. P. Meinel, Applied solar energy. Reading, MA: Addison-Wesley Publishing Co., 1976.
7. E. G. Laue, "The measurement of solar spectral irradiance at different terrestrial elevations," Solar Energy, vol. 13, pp. 43-50, IN1-IN4, 51-57, 1970.
8. M. Daneshyar, "Solar radiation statistics for Iran," Solar Energy, vol. 21, pp. 345-349, 1978
9. G. W. Paltridge and D. Proctor, "Monthly mean solar radiation statistics for Australia," Solar Energy, vol. 18, pp. 235-243, 1976.
10. Solar Dataset_Solar from MEDA-1.XLSX, University of Nagpur.

Google Drive link - RES Group 11B