**Solar Irradiance Forecasting via Artificial Neural Network**

|  |  |  |
| --- | --- | --- |
| Manish Kumar Gupta  *Department of CST*  *University of Engineering and Management,* Kolkata, India  mkrock2397456@gmail.com | Sneha Kumari  *Department of CST*  *University of Engineering and Management,* Kolkata, India  kumarisnehasaw@gmail.com | Aastha Jaiswal  *Department of CST*  *University of Engineering and Management,* Kolkata, India  aasthakjaiswal@gmail.com |
| Ankit Shaw  *Department of CST*  *University of Engineering and Management,* Kolkata, India  ankitasn333@gmail.com | Anand Kumar  *Department of CST*  *University of Engineering and Management,* Kolkata, India  anandkumar55631@gmail.com | Sudipta Basu Pal  *Department of CST & CSIT*  *University of Engineering and Management,* Kolkata, India  sudipta\_basu68@yahoo.com |

*Department of CSIT, Institute of Engineering and Management, University of Engineering and Management, Kolkata, India,*

*2\*Department of CST & CSIT, Institute of Engineering and Management, University of Engineering and Management, Kolkata, India*

*Email: ………………………………………………………………………, sudipta\_basu68@yahoo.com*

*Corresponding Author:* [*sudipta\_basu68@yahoo.com*](mailto:sudipta_basu68@yahoo.com)

**ABSTRACT**

With rising global energy demands, the integration of renewable energy sources has become crucial for sustainable development. Solar photovoltaic (PV) systems, a widely adopted renewable source, play a significant role in supplementing traditional energy sources. However, the integration of solar PV systems into the national electricity grid requires accurate and reliable forecasts of solar irradiance, as fluctuations in solar energy production can disrupt grid stability. This study presents a predictive model for solar irradiance based on artificial neural networks (ANN), which uses various weather parameters, such as temperature, humidity, and cloud cover, as input features.

The ANN model is built from scratch, employing both gradient descent (GD) and stochastic gradient descent (SGD) optimization techniques to refine the model's performance. While both optimization methods demonstrate effective convergence, testing shows that the SGD method results in lower mean- squared error (MSE) values, indicating enhanced accuracy for predicting solar irradiance levels. To validate the ANN model's effectiveness, it was compared against a baseline linear regression model, where performance was evaluated using metrics such as mean-squared error (MSE), mean absolute error (MAE), and root mean-squared error (RMSE).

The results indicate that the ANN model outperforms linear regression, demonstrating lower error rates across all metrics. This study contributes to the field of renewable energy forecasting by providing a robust and efficient model for solar irradiance prediction, which could assist in better managing grid stability and optimizing energy distribution. A proof of-concept implementation of this project can be referenced for further details.

**Keywords— Solar Irradiance Forecasting, ANN, PV, GD, SGD, Pyranometer**

**1. Introduction**

Solar power is a great source of renewable power because it is not depleting and is widespread. However, the efficiency of solar power systems depends to a large degree upon the predictability of solar irradiance. Weather, cloud movements, and the status of the atmosphere influence solar radiation, thereby making predictability an issue [1]. Successful forecasting helps towards efficient operation of photovoltaic (PV) systems, improves the stability of the grid, as well as ensuring improved control of energy [2].

Different prediction methods have been created over the years. The traditional method consists of statistical methods such as linear regression and time series models, considering past trends [3]. The other approach is Numerical Weather Prediction (NWP), employing the atmosphere equations to forecast solar radiation [4]. Both of these approaches fail when they account for the variability and uncertainty of solar irradiance data.

With improvements in computer methods, Artificial Neural Networks (ANNs) are presently in vogue for predicting solar energy [5]. ANNs can identify complex patterns in big data sets and are therefore better than older models. Optimization methods such as Gradient Descent (GD) and Stochastic Gradient Descent (SGD) are used extensively to improve the accuracy of ANN predictions [6].

This study aims at improving solar irradiance prediction with the help of ANN models. By optimizing network parameters and using real meteorological data, the study aims at increasing prediction accuracy, thereby assisting in the efficient use of solar energy.

**2. BACKGROUND**

Solar Irradiance Data Plays a crucial role on various field like Climate Studies, agriculture Field which Help in optimizing solar power generation [7]. There is the vast use of Solar Irradiance but there should be some challenges faced like Calibration Error, different time zones, unfortunate changes in climate condition, less amount of dataset with is required to train the data. Due to this Optimizing the Solar Irradiance may be difficult for us. To improve or tackle the challenges we need to improve measurement technique, increase the amount of dataset and enhancement in equipment. We must work in AI and machine learning Integration and using of Advanced Monitoring System which help in future scope. For Solar Irradiance changes in climate should be improved and to create Space-Based Solar Power working 24/7.

**3. EXPERIMENTAL SETUP**

To conduct the experiment, following method were used to compare the model by finding the accuracy of the model

**ANN:** It is a machine learning model which has an input layer, one or more hidden layer and an output layer that rely on training data to learn and improve accuracy [8].

**MAE:** It is absolute difference between predicted and actual values that calculates the average magnitude [9].

**MSE:** It is a metric used for evaluating the performance of the model by measuring the average squared difference between the predicted and actual target values [10].

**RMSE:** It is used to measure the quality of the prediction by measuring the average error magnitude between the predicted and actual values [11].

**SGD:** It is an optimization algorithm used to train models efficiently usually used to deal with large datasets which randomly selects a small number of samples for each iteration and calculate the gradients for that [12].

**Linear Regression:** It is an algorithm that provides a linear relationship between an independent variable and dependent variable to predict the outcome [13].

Now, overview of our project where the entire process follows structured pipeline, our process starts with data collection where historical solar irradiance dataset was taken along with the environmental factors like temperature, humidity, pressure etc. Next step is data preprocessing where we clean and normalize the data to improve the accuracy. Then, model training, where multi-layer ANN was designed and were trained using historical data. And, then to ensure high performance model evaluation was done using MAE, RMSE metrics.

**Workflow**

**Fig.1.Workflow of the Experimental Setup**

**Data Collection:** Here we taken solar irradiance and weather data from database.

**Data Preprocessing:** Here we have done data preprocessing in which we take input and model.

**Feature Engineering:** We have taken the input such as Humidity, temperature, sunlight, etc.

**Model Selection:** Select the model like Linear Regression and Neural Network.

**Model Training & Optimization:** Train the data on these model for working of better performance.

**Model Evaluation:** Train the model using metrices like RMSE, MAE, MSE, R-2 Score

**Solar Irradiance Forecast Output:** Train the data and then predict the own value for decision making and generated the final value.

**4. EXPERIMENTAL MECHANISM**

**(a) DATASET DESCRIPTION**

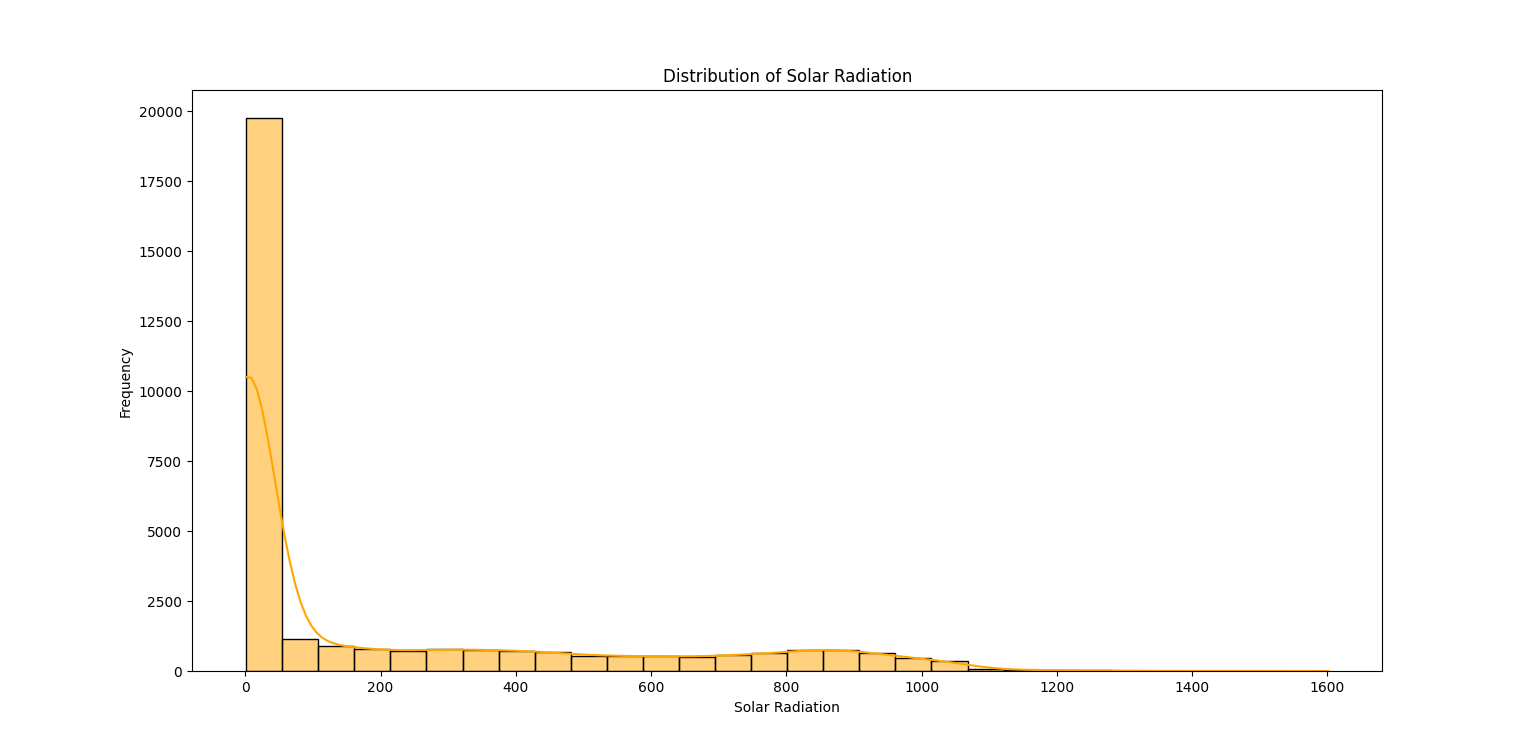
The unused dataset is obtained from Kaggle.[14] The dataset contains several meteorological factors used in predicting solar irradiance, which are as follows:

* Solar Radiation: Extracted in watts per meter squared (W/m²).
* Temperature: Expressed in degrees Fahrenheit (°F).
* Humidity: Presented in percentage form.
* Barometric Pressure: Extracted in inches of mercury (Hg).
* Wind Direction: Expressed in degrees.
* Wind Speed: Extracted in miles per hour (mph).
* Sunrise and sunset times.

The real-world raw data undergo various transformations before being represented to improve practicality. The data in question is stored in tables, and the weather conditions are represented by rows while each column represents a certain meteorological feature. The dataset has an approximate size of 2.96 MB. The dataset has known, performed preprocessing steps like missing value imputation, feature scaling, outlier filtering, or inclusion to fulfill data quality standards.

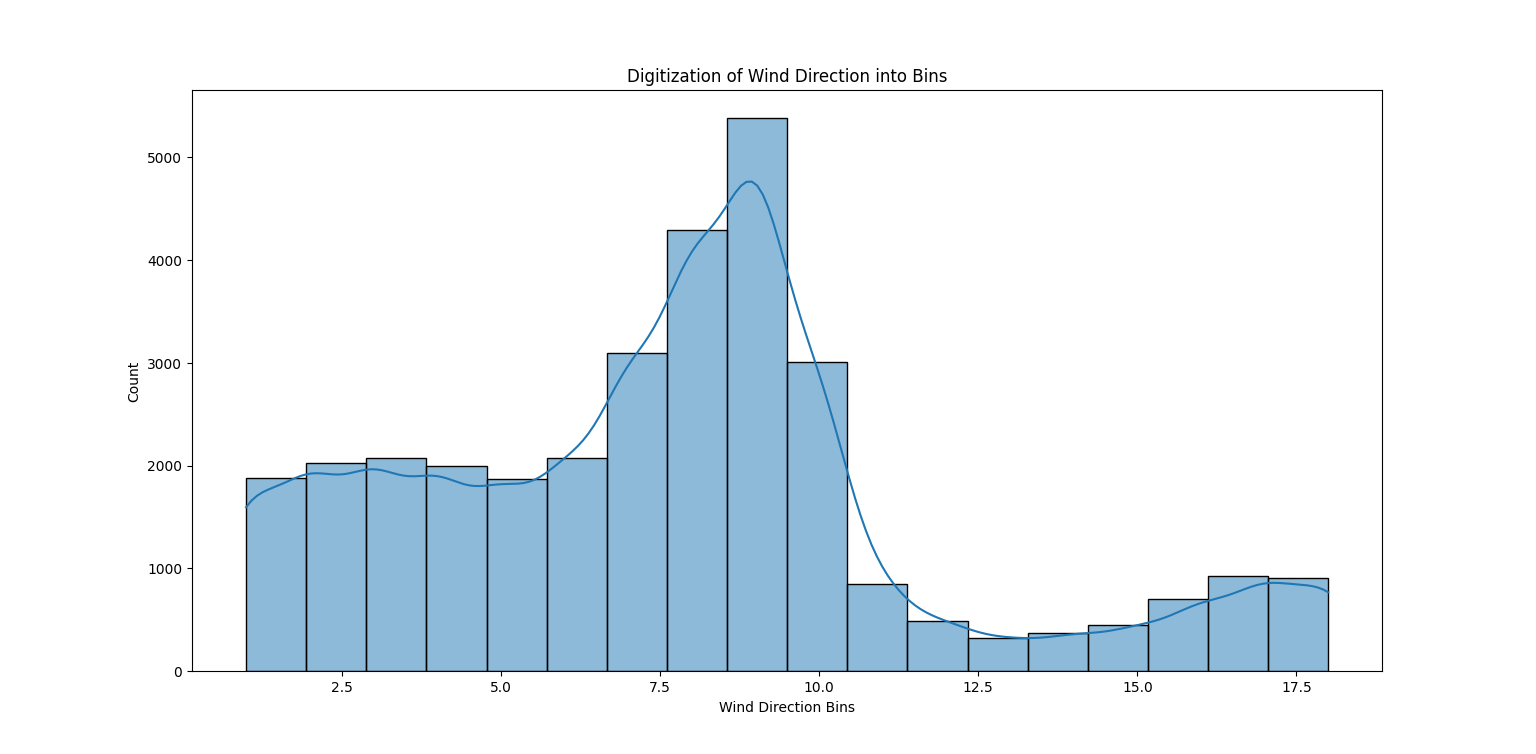
One of the beneficial attributes of this dataset is its application towards machine learning solar forecasting. This dataset was created to be used with many different predictive models, such as ANNs, decision trees, random forests, and even Transformer-based models. It assists in efficiently managing solar power by improving the accuracy of solar irradiance predictions.

**(b) EXPERIMENTAL PROCEDURAL & RESULTS**



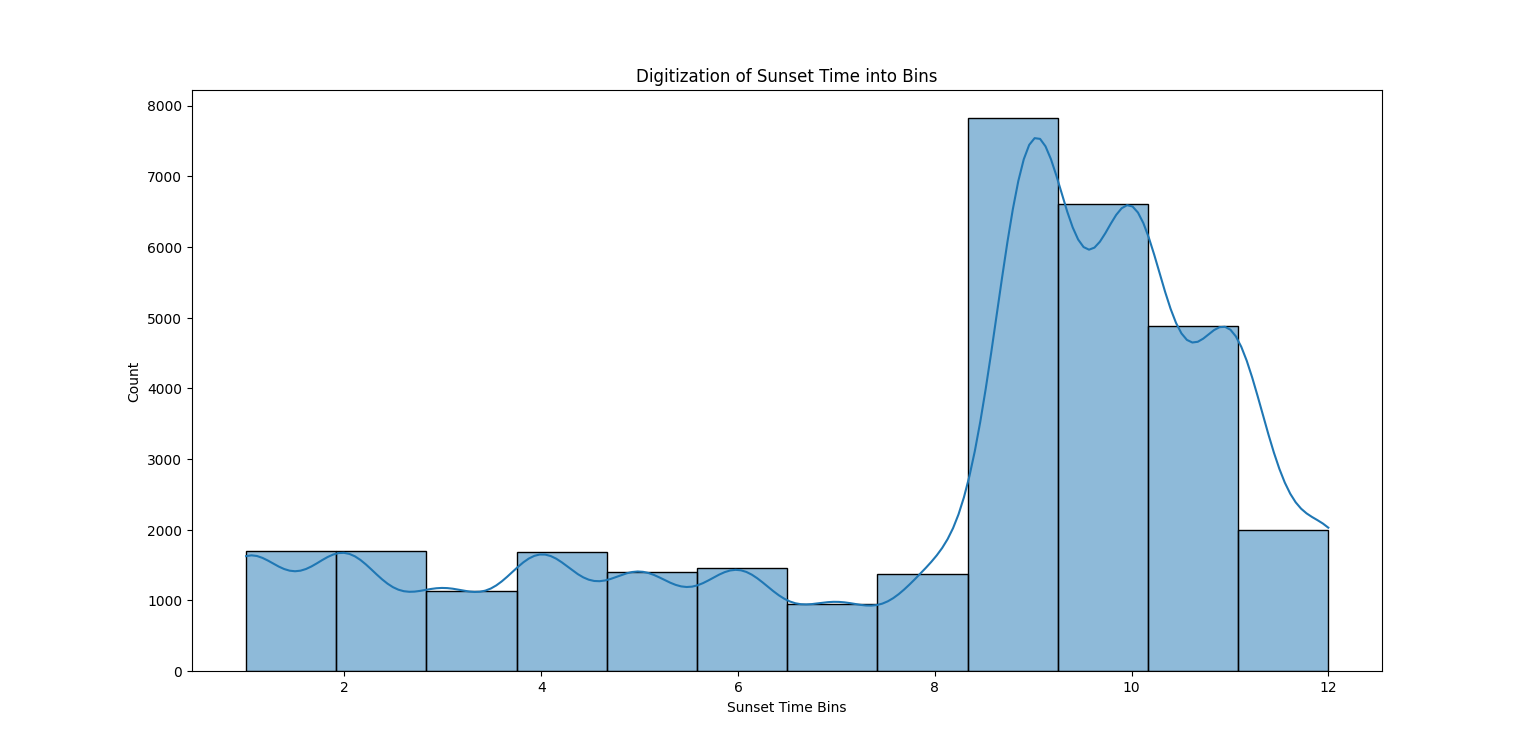
**Fig.2. Plot on Solar Radiation Distribution**

The graph (Fig.2.) illustrates the distribution of solar radiation values using a histogram. The graph shows that the distribution is skewed to the right, with most of the values clustered around zero, indicating lower solar radiation values are more frequent. In addition, a density curve is placed over the histogram, illustrating a better representation of the trend of distribution.



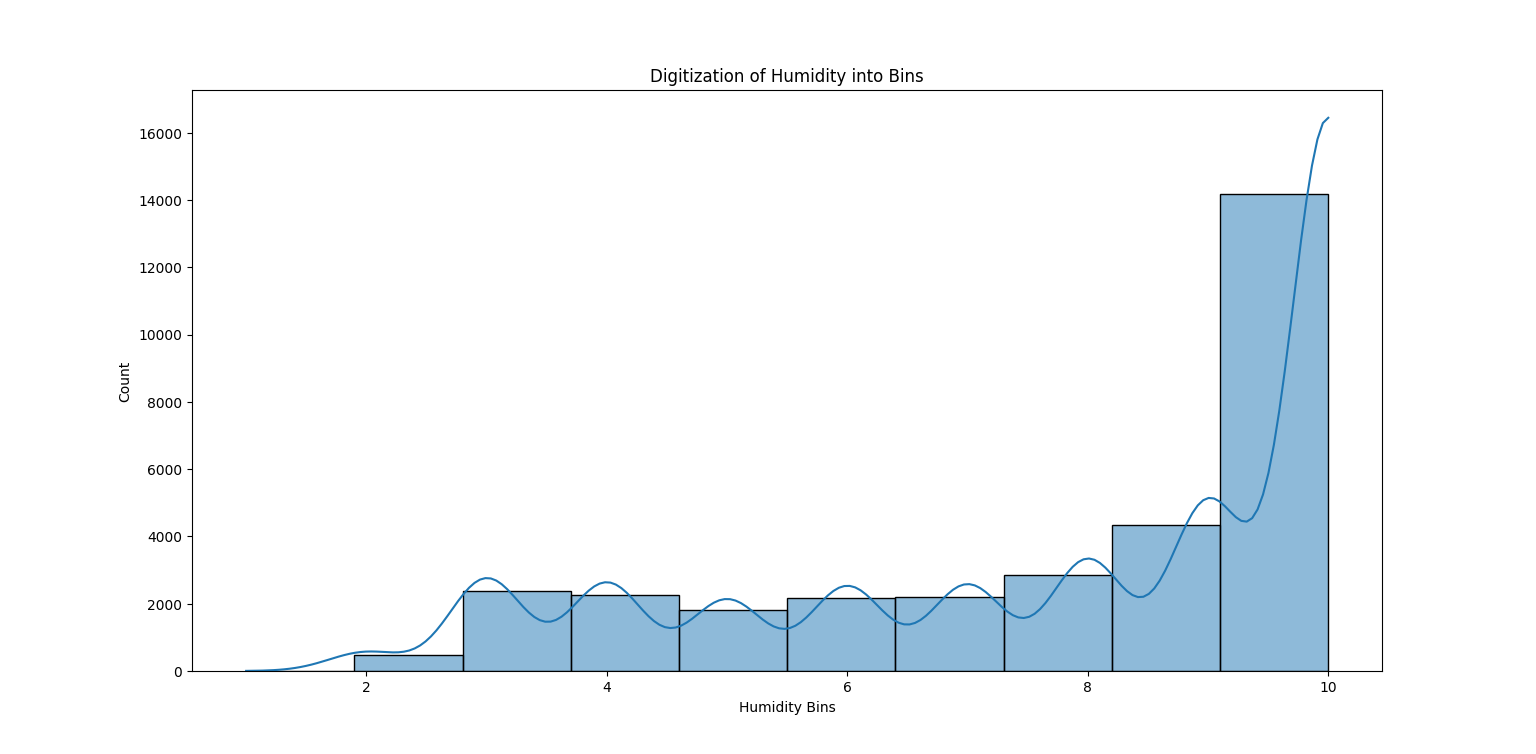
**Fig.3. Plot on Wind Direction into Bins**

The histogram (Fig.3.) illustrates the way wind direction is grouped, indicating how frequently each wind direction occurs. The data appears to be centered on a central peak, with the most frequent directions in the middle groups, indicating a central wind direction. The smooth line at the top provides a better picture of the distribution, highlighting changes and trends in the data.

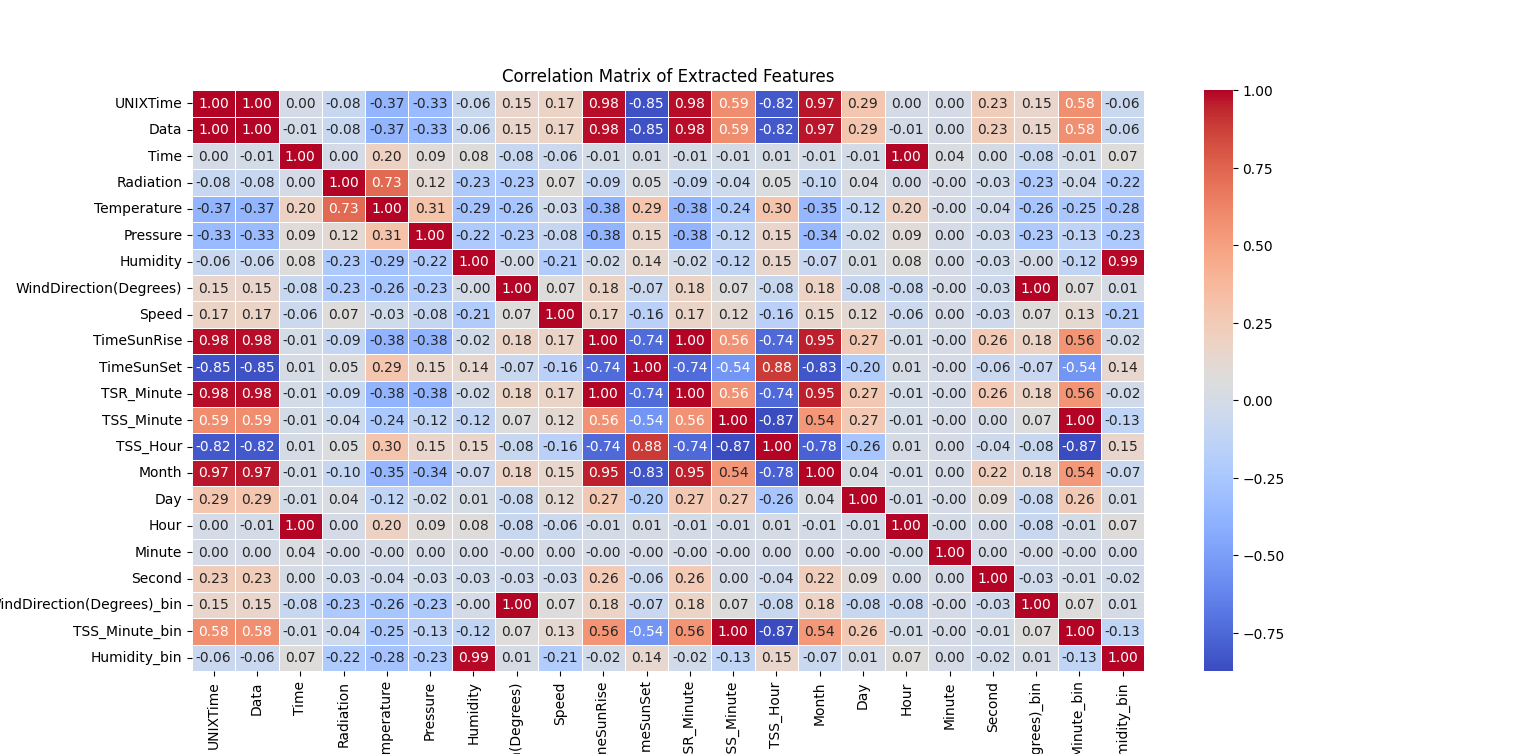


**Fig.4. Plot on Sunset Time into Bins**

The histogram (Fig.4.) shows how the times of sunset are binned, showing how often each occurred. The peak towards the later bins makes it obvious that most sunset times are somewhere within some range. The smooth curve on top makes it easier to see the overall pattern and trend of sunset times.

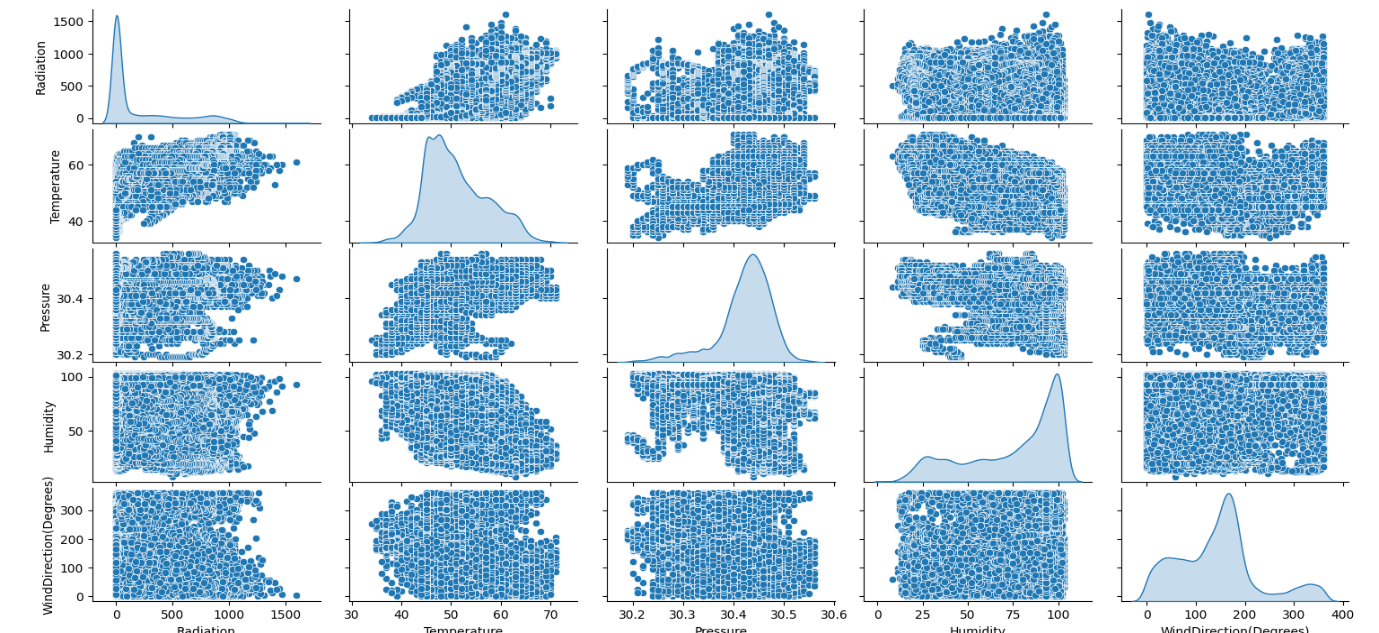
**Fig.5. Plot on Humidity into Bins**

This graph (Fig.5.) depicts the distribution of the humidity values in bins. The x-axis is the bins of humidity, and the y-axis is how frequently each bin occurs. The distribution skews to increased humidity values, with a steep rise in frequency in the higher bins. This shows there are more data points in that range.



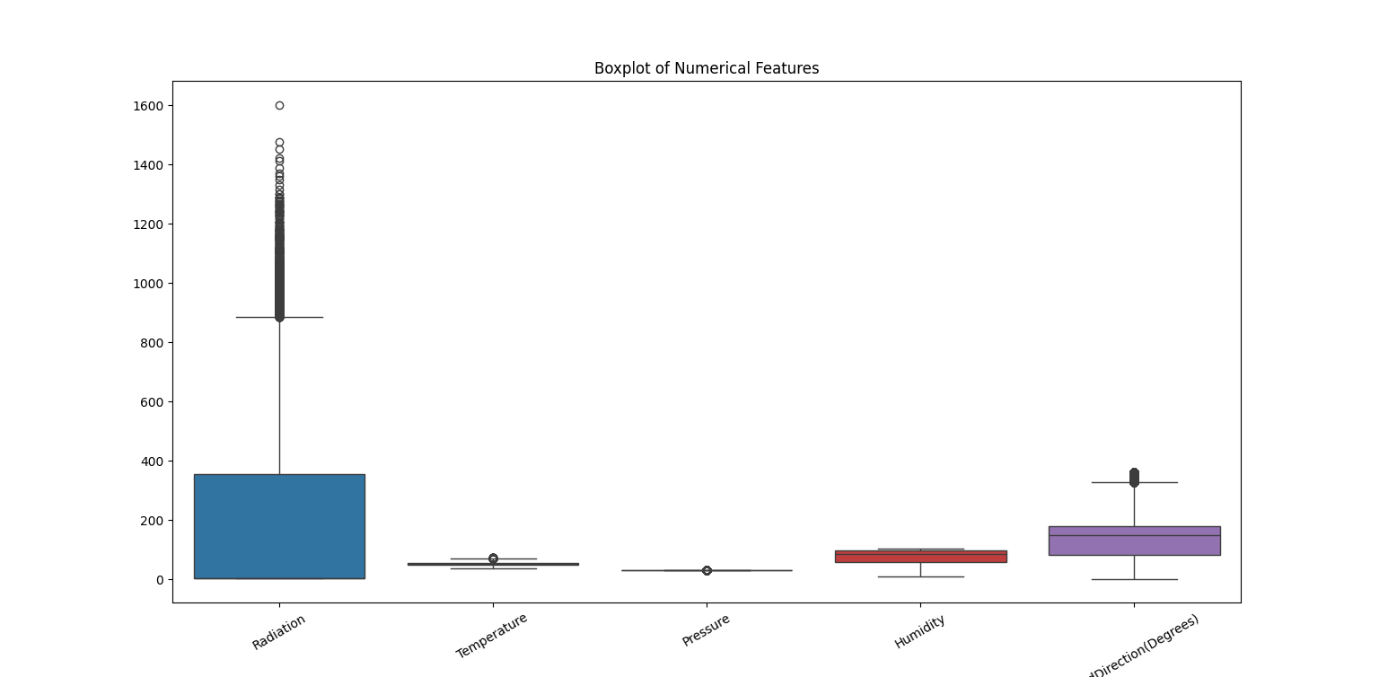
**Fig.6. Correlation Matrix of Extracted Features**

The image (Fig.6.) is a feature correlation matrix, presented in a heatmap that has values varying from -1 to 1. Each entry in the table is the correlation coefficient between the two features. Negative correlations have blue colors and positive correlations have red colors. Sometime features and humidity buckets have high correlations, while other features have weak or no correlations.



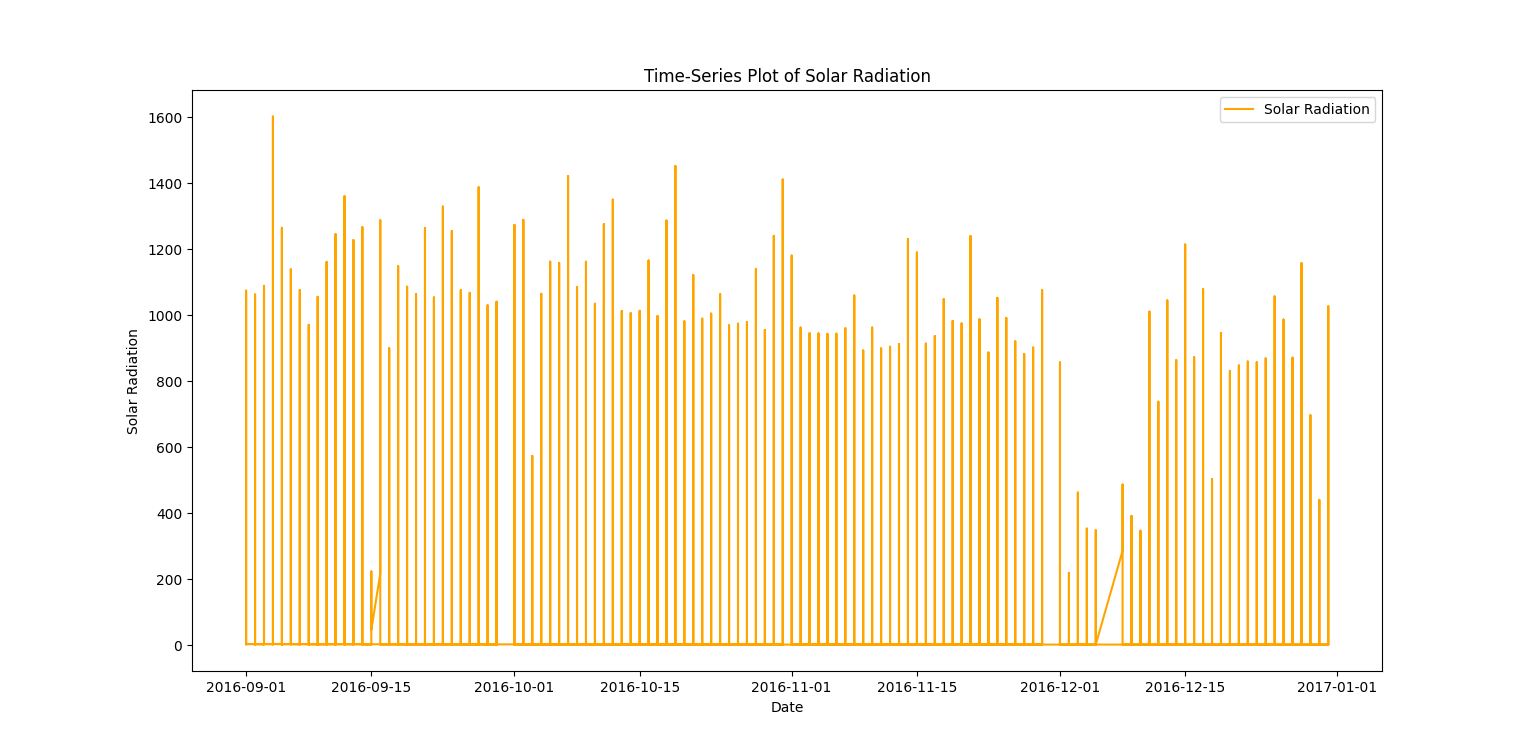
**Fig.7. Pairplot to visualize relationships between key numerical features**

The plot (Fig.7.) is a pair plot that illustrates the relationship between radiation, temperature, pressure, humidity, and wind direction. The plot on the diagonal presents the distribution of a single variable, and the plots in between present how two variables relate to one another. The diagonal density plots present information regarding how each feature is distributed, and the scatter plots present potential trends, groups, or patterns in the data.



**Fig.8. Box plot of numerical features**

The graph (Fig.8.) shows a box plot of comparisons of distribution of numerical properties, i.e., radiation, temperature, pressure, humidity, and wind direction. Differences in data range, median, and presence of outliers are clear from the plot, particularly for the radiation property, which contains numerous extreme values. The remaining variables have comparatively tight distributions with fewer outliers, showing more uniform data ranges.

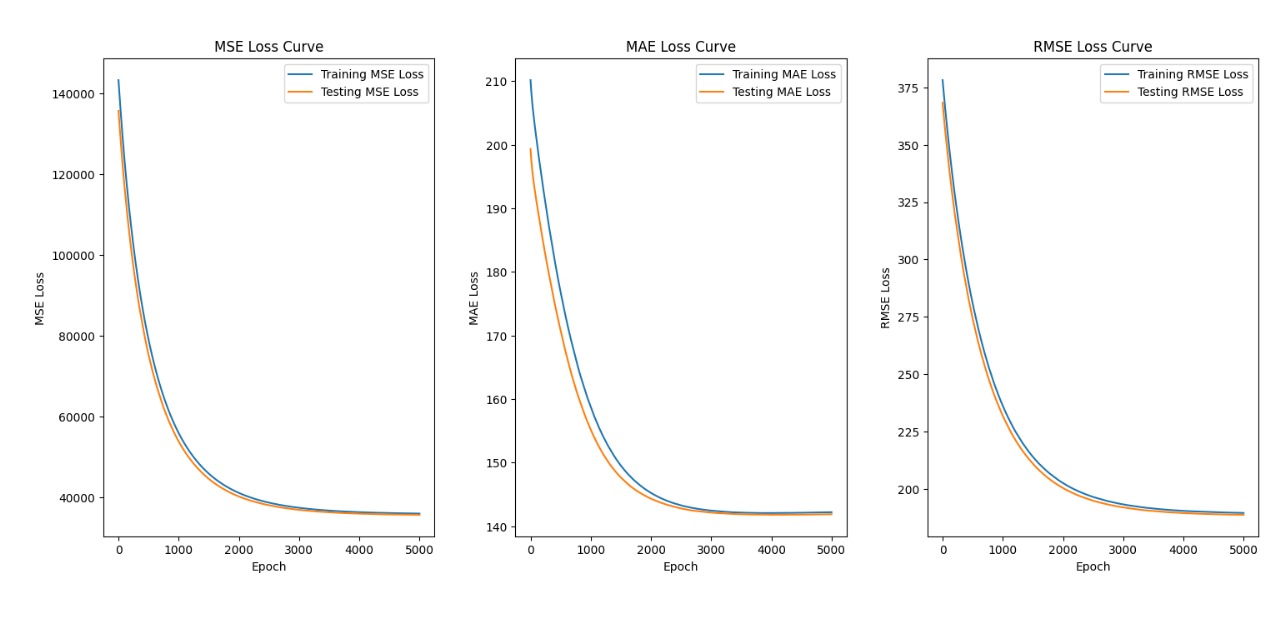
 **Fig.9. Time series plot of solar radiation**

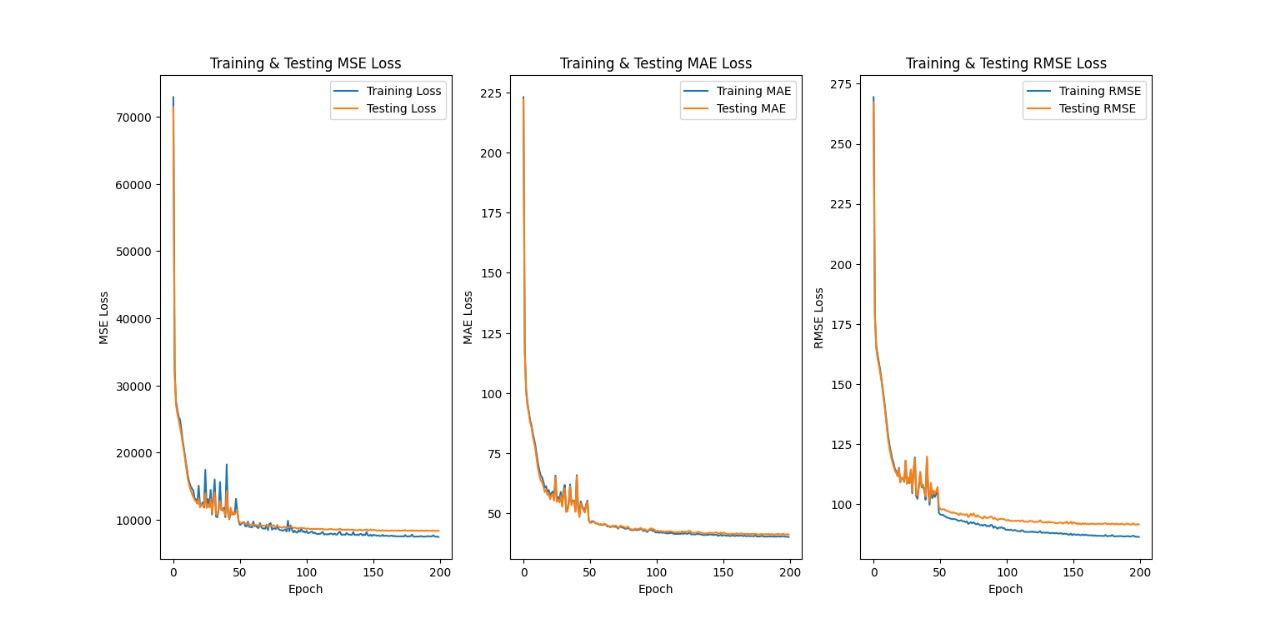
The image (Fig.9.) is depicting a time-series plot of solar radiation for a time frame between September 2016 and January 2017. The variation in the amount of solar radiation is depicted through vertical bars, with large variations and periodic decreases in intensity. The cyclical trend is representing the variation in solar exposure with time, possibly because of season or weather conditions.

**5. Result Analysis**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RMSE | MAE | MSE | R² SCORE |
| Linear Regression | 188.8341 | 141.9029 | 35658.3182 | 0.6339 |
| Neural  Network | 91.4252 | 41.2457 | 8358.5589 | 0.9142 |

In the above table we have calculated RMSE, MAE, MSE, R² SCORE on the Model of Linear Regression and Neural Network. In which we get RMSE, MAE, MSE, R² SCORE of Linear Regression is 188.8341, 141.9029, 35658.3182, 0.6339 respectively and after that we get the RMSE, MAE, MSE, R² SCORE of Neural Network is 91.4252, 41.2457, 8358.5589, 0.9142 and we that Neural Network perform better than Linear Regression. As R2-Score of Neural Network model is 0.9142 which is far better than the linear Regression (0.6339).

**Fig.10. MSE, MAE, RMSE Loss Curve of Training and testing data upto 5000 epoches of Linear Regression Model**



**Fig.11. MSE, MAE, RMSE Loss Curve of Training and testing data upto 200 epoches of Neural Network Model**

Here **fig.10** performing Linear Regression and the **fig.11** model performing Neural Network. The Neural Network perform better here.

**Real-time prediction on a recent weather data sample**

We have taken a sample of weather data of Bidhannagar (Kolkata) [15] for, Wednesday 2nd April 2025 at 03 PM (IST) and tries to predict the value of solar irradiance by using out trained model and our predicted result comes out to be 883.5321655273438 W/m² and the actual value at that time was approx. 925 W/m². So, we can see that out model is approx. 90-95% accurate.

**6. CONCLUSION**

Here in this project, we created an artificial neural network (ANN) to forecast solar irradiance based on different weather-related parameters. To set the model on the correct path, we initialized its weights with a specialized technique. We then trained the model using two optimization algorithms: gradient descent (GD) and stochastic gradient descent (SGD). To keep the model stable and prevent it from generating excessive values, we employed gradient clipping.

The results were encouraging—our ANN learned patterns in the data successfully. The SGD-based model was better, with a mean squared error (MSE) of approximately 9,000, while the GD-based model had a slightly worse MSE of approximately 11,000. When we compared our ANN with a linear regression model with fewer parameters, the ANN provided better performance.

In the future, improvements may include the addition of time-series data to capture trends over time more effectively. This might be achieved with more sophisticated models such as long short-term memory (LSTM) networks or temporal convolutional networks (TCN), which are better suited to sequential data.

References

1. Inman, R. H., Pedro, H. T., & Coimbra, C. F. (2013). *Solar forecasting methods for renewable energy integration*. Progress in Energy and Combustion Science, 39(6), 535-576.
2. Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-De-Pison, F. J., & Antonanzas-Torres, F. (2016). *Review of photovoltaic power forecasting*. Solar Energy, 136, 78-111.
3. Mellit, A., & Kalogirou, S. A. (2008). *Artificial intelligence techniques for photovoltaic applications: A review*. Progress in Energy and Combustion Science, 34(5), 574-632.
4. Qazi, A., Hussain, F., Rahim, N. A., Hardaker, G., & Alghazzawi, D. (2020). *Towards sustainable energy: A systematic review of renewable energy sources, technologies, and public opinions*. IEEE Access, 8, 1-24.
5. Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. (2018). *History and trends in solar irradiance forecasting: A review and future recommendations*. Renewable and Sustainable Energy Reviews, 89, 268-306.
6. Tao, Y., Shan, Y., Zhang, Z., & Ding, Y. (2019). *A hybrid deep learning framework for solar irradiance forecasting*. Energy Conversion and Management, 201, 112154.
7. Solar Energy Applications in Protected Agriculture: A Technical and Bibliometric Review of Greenhouse Systems and Solar Technologies <https://www.mdpi.com/2073-4395/14/12/2791>
8. Neural Network: <https://www.ibm.com/think/topics/neural-networks>
9. MAE: <https://www.sciencedirect.com/topics/engineering/mean-absolute-error>
10. MSE: <https://encord.com/glossary/mean-square-error-mse/>
11. RMSE: <https://www.deepchecks.com/glossary/root-mean-square-error/>
12. SGD: <https://h2o.ai/wiki/stochastic-gradient-descent/>
13. Linear Regression: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/>
14. Dataset Description: <https://www.kaggle.com/datasets/dronio/SolarEnergy?resource=download>
15. Bidhannagar, weather data <https://www.wunderground.com/history/daily/in/bidhannagar/yesterday>