**SOLAR ENERGY PRODUCTION FORECASTING USING AI/ML**

***B.TECH SEM – VII Mini PROJECT***

***Dept. of Computer Science & Engineering***

By

**GROUP ID : 4**

**SAUMYA SHAH 18BCP102**

**SAGAR SINHA 18BCP094**

**Under the Supervision**

**Of**

**Dr. Rajeev Kumar Gupta**



**SCHOOL OF TECHNOLOGY**

**PANDIT DEENDAYAL ENERGY UNIVERSITY**

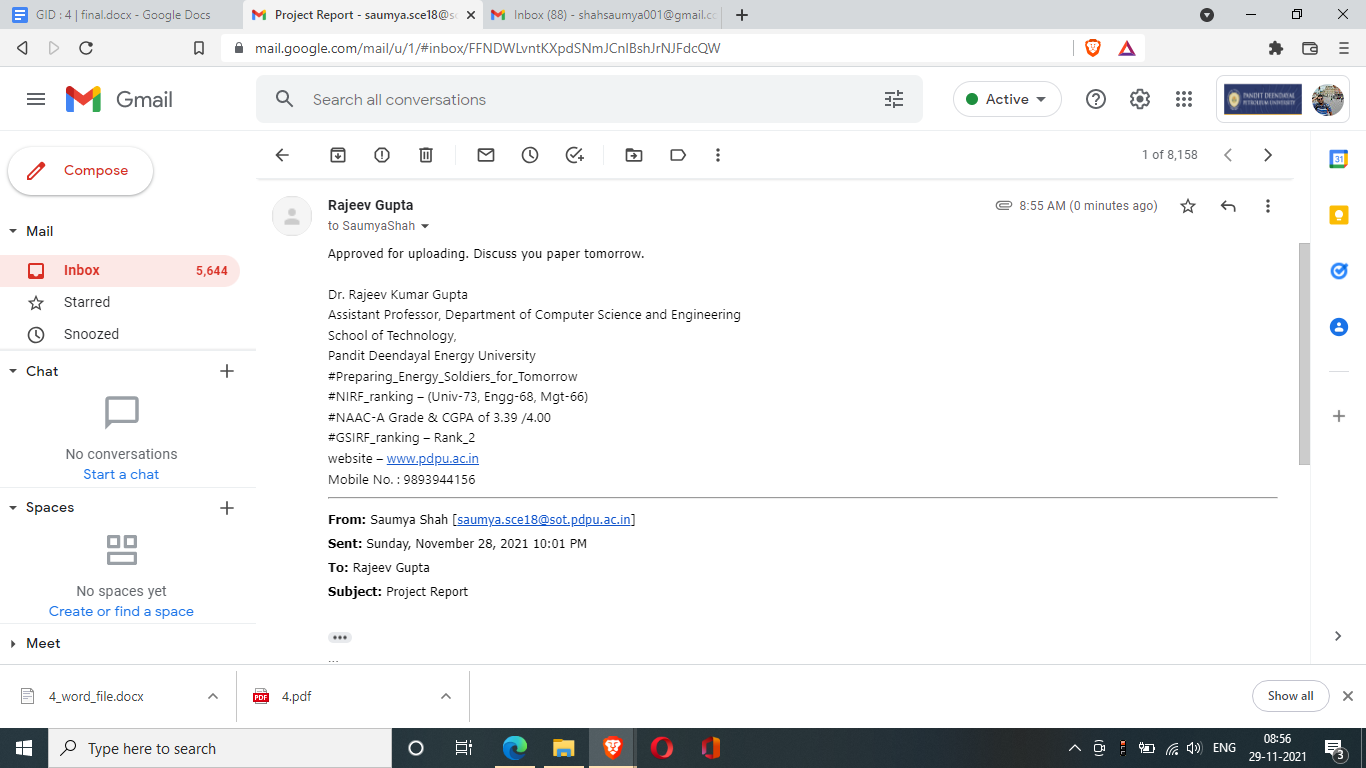
**GANDHINAGAR, GUJARAT, INDIA**

**July – December, 2021**

**CONTENTS**

| **Sr. no.** | **Topic** | **Page no.** |
| --- | --- | --- |
| 1 | Mentor's Digital Signature | 3 |
| 2 | Abstract | 3 |
| 3 | Introduction | 4 |
| 4 | Literature Review | 8 |
| 5 | Comparative Analysis | 10 |
| 6 | Dataset | 12 |
| 7 | Proposed Work | 14 |
| 8 | Implementation Details | 14 |
| 9 | Results | 18 |
| 10 | Challenges | 18 |
| 11 | Limitations and Future Directions | 18 |
| 12 | References | 19 |

**Mentor’s Digital Signature** *(Dr. Rajeev Kumar Gupta)*

**



**Abstract**

The global energy system is transforming with the consideration of a growing population with rising living standards that will need more energy. Simultaneously, the world must find ways to reduce greenhouse gas emissions and provide sustainable energy production. As electricity is the fastest-growing part of the energy system and thus shifting to renewable energy production systems is the need of the hour. IEA reports that the Sun could be one of the largest resources for energy production for the upcoming ‘net zero goals’ of the world. Rapid decrease in the cost of installation and availability of solar radiation are major advantages for adoption, although this mode of generation of energy has a variable output which is influenced by several parameters. Hence solar energy forecasting is crucial to understand and predict the output of production, the demand of the consumers, and optimizing the dispatch of the electricity from grids to users. Here, we have analyzed innovative methodologies for forecasting solar energy radiation. Moreover, it provides a review and comparative analysis on various ANN, Probabilistic and Statistical, and Time Series based models used for estimating solar irradiation. Each of these models have been trained and tested in different geographical regions, weather conditions, and corresponding relevant parameters, which would provide an insight on efficient model selection and optimization for predicting short and long-term solar irradiation. Further, we have discussed some challenges while forecasting solar irradiation and future research direction in this domain.

Keywords: solar irradiation, solar energy, AI models, analysis, ANN, Statistical, Probabilistic, Time series, short- and long-term forecasting

**Introduction**

Solar Panels are made of numerous semiconductor cells in an array, which when struck by photons from the Sun would knock electrons loose from their atoms. Conductors of electricity are provided at the sides of these cells which collect electrons and transfer them to wires. The solar inverter then converts DC electricity from solar modules to AC electricity, which is used by appliances. Excess electricity is then fed back to the grid. There are some alternatives to Photovoltaic solar energy generation, such as solar thermal systems, concentrated solar power, and passive solar gain.

A major concern around the solar energy production systems is the variability and unpredictability of environmental factors and other parameters it depends on. This poses issues with the grid reliability and expenses associated with operating the technical infrastructure. Moreover, peak demands and user consumption patterns are uncertain. Therefore, solar energy forecasting addresses this issue by developing models to predict the energy demand, production and irradiation from the Sun. Forecasting systems can help regulate PV systems and determine the dispatching of the energy created.

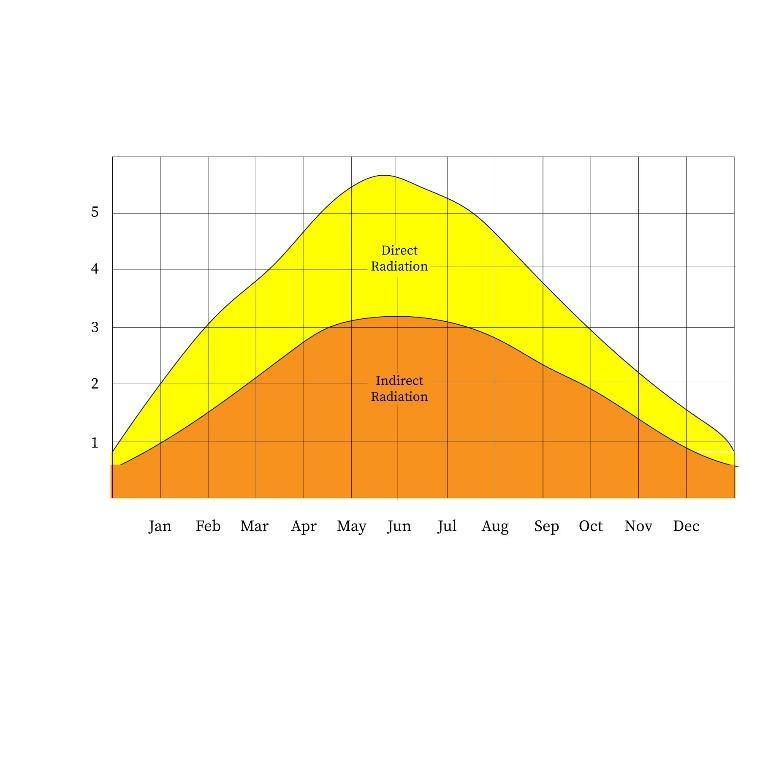
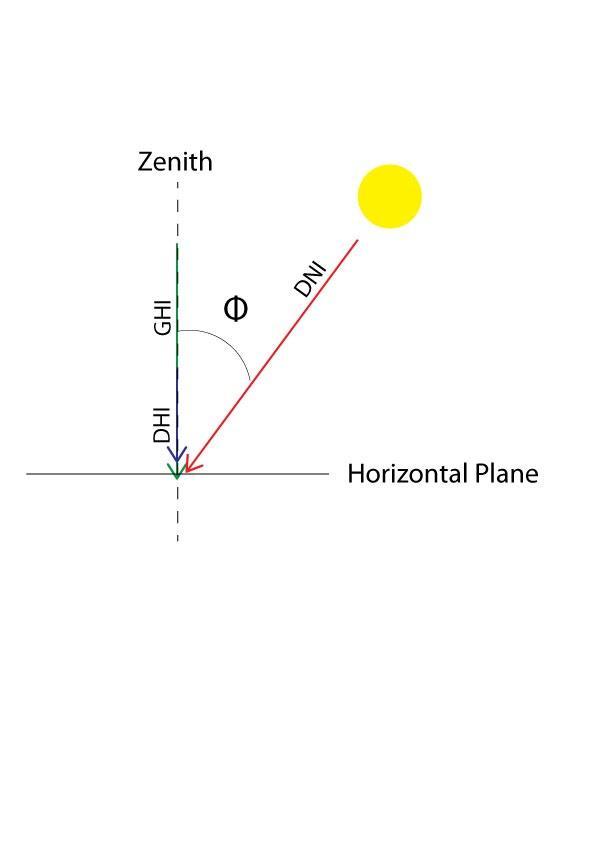
Here, we will particularly focus on Photovoltaic solar systems and the comparative analysis of methodologies used for forecasting solar irradiation. Below listed are the considered parameters for data collection and model training.

* 1. Parameters
* Temperature(C) - Temperature of the environment around the solar panels' installation. Temperature increase causes the bandgap of semiconductors to diminish, affecting its material properties. Moreover, the open-circuit voltage is the characteristic in a solar cell that is most impacted by temperature changes.
* Humidity(mm)- The amount of solar radiation is influenced by the atmosphere's relative humidity level since increased water vapor concentration produces an increase in reflected refractive radiation. The intensity of the solar radiation absorbed by the photovoltaic cell is lowered in this scenario, which is reflected in the cell's productivity.
* Clear sky index - Cloud cover affects the solar radiation reaching the solar panels, and this index estimates atmospheric attenuation due to clouds by measuring the ratio of surface solar radiation G to the solar radiation that would be received under a clear(cloudless) sky, Gcs.
* Wind speed(m/s)- There is an anticorrelation between wind speed and solar irradiation, and the wind speed range varies according to the geographical location and weather conditions.
* The angle of incidence(ω) - The solar incidence angle is the angle between the sun's rays and the normal on the horizontal surface.
* Azimuth angle(φ) - This is the angle between the projection of the sun's center onto the horizontal plane and due south direction.
* Zenith angle(Z)- This is the angle between the sun and the vertical.
* Atmospheric pressure(atm) - The pressure produced by the weight of air in the Earth's atmosphere is known as atmospheric pressure, and the weight of air is gravitational. This force increases as altitude decreases, exerting a stronger downward pull on the photons, raising solar intensity and the output current and voltage.
* Albedo - This is the amount of light reflected by a surface as is a crucial metric to determine the material of semiconductors in the solar panels. Moreover, the surroundings of the solar panels can diffuse solar radiation, and hence this effect can affect the performance of solar cells.
* Solar insolence(kWh/m^2) - Solar insolation is the measure of solar radiation incident upon a unit horizontal surface over a specified time for a given locality. It depends strongly on the solar zenith angle and also on the ratio of the actual distance to the mean distance of the Earth from the Sun. (S=solar constant, Zenith angle).

𝐼 = 𝑆 \* cos 𝑍

* Tilt angle(α) - The angle of tilt of the solar panels is the tilt angle which is largely determined by the latitude values of the location of the solar panel.
* Terrain Elevation(m) - Higher altitudes plants with optimum conditions result in efficient solar energy production than at sea levels.
* Solar irradiance(W/m^2) - This is output power per unit area received from the Sun in the form of electromagnetic radiation in photons.
* Global Horizontal Irradiance(W/m^2)- GHI is the total amount of shortwave radiation received from above by a surface horizontal to the ground. This value includes the Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DIF). DNI is solar radiation that comes perpendicular to the surface from the direction of the sun. DIF is solar radiation that does not arrive on a direct path from the sun but has been scattered by molecules and particles in the atmosphere.

𝐺𝐻𝐼 = 𝐷𝑁𝐼 \* *cos* θ + *DHI*



*Figure 1. (a)GHI and DNI, DHI relationship, (b)Diffuse and Indirect solar radiation*

* Size of PV installation(kW) - These can be broadly categorized as small residential, medium and commercial, ground-mounted, and floating large-scale solar panel systems.

Data collected for prediction is not and undergoes various pre-processing processes most of which are discussed below:

* 1. Pre-processing techniques

Data collection of the parameters above from the sensors undergo various manipulation to remove outliers and nulls, scale the data and aggregate it if necessary. Sometimes the data are not based on the same dimensionality and need a reduction. Time series data can be difficult to manipulate and scale and thus mathematical representation is attained by Wavelet transform.

* Imputation: It refers to filling up missing values in the dataset. The most commonly used technique for imputation is Interpolation. It is used to increase the size of the dataset keeping in mind the dimensionality, and the variance of the data. This is a very crucial step as ML modelling is only possible on data with same dimensionality
* Normalization (min-max scaling): In this pre-processing technique, the data points are scaled to unit vectors and the data range is shifted between 0 and 1. It is done to prevent some features with larger variance overshadow the ones with smaller variance during the ML modelling.



* One-Hot-Encoding: One-hot encoding is a common approach for dealing with categorical data in machine learning. Categorical variables must be converted in the pre-processing section since many machine learning models need numeric input variables.
* Aggregation: It is a very important pre-processing technique in ML modelling. The most used types of aggregation are:
  + Mean: It gives us a central tendency of the data.
  + Median: It gives us the central tendency of sorted data.
  + Mode: It gives us the category having the highest frequency count.
* PCA: It is a dimensionality reduction technique used for training conventional ML models. The transformed lower dimensions give an abstract representation of the variance of the data. It is based on the concept of Diagonalization and Eigen Value Decomposition.
* Wavelet Transform: It is a pre-processing technique similar to the Fourier transform (or much more to the windowed Fourier transform), but with a completely different merit function. The major distinction is that the Fourier transform decomposes the signal into sines and cosines, or functions localized in Fourier space; on the other hand, the wavelet transform utilizes functions localized in both real and Fourier space.

1.3 Performance Metrics

There are multiple metrics used to measure the performance of an ML model on the testing data. Some of the important metrics are:

* MSE: The average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value—is measured by the mean squared error or mean squared deviation of an estimator. MSE is a risk function that represents the squared error loss's anticipated value.

* RMSE: The root-mean-square deviation, also known as root-mean-square error, is a commonly used metric for comparing values predicted by a model or estimate to values observed.

* MAE: It is the average of all the absolute errors or alternatively the difference between two given continuous variables.

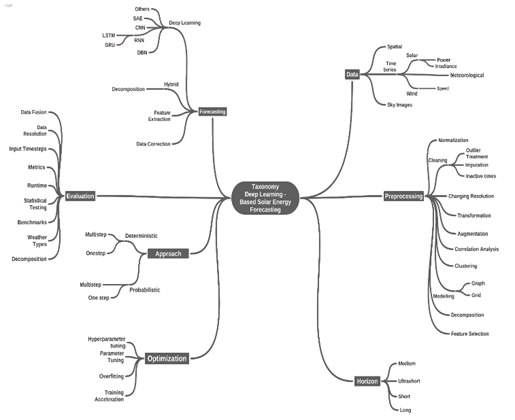
* MAPE: In statistics, the mean absolute % error, also known as mean absolute percentage deviation, is a measure of a forecasting method's prediction accuracy.

* R - Squared: The coefficient of determination, is the percentage of variance in the dependent variable that can be predicted by the independent variable.

* Correlation Index: The Pearson product-moment correlation coefficient, often known as Pearson's r, is a measure of linear correlation between two sets of data.

* Kurtosis - a measure of how "tailed" a real-valued random variable's probability distribution is.

* Phillips’s perron test- In time series analysis, it is used to unit root test the null hypothesis that a time series is integrated of order one.

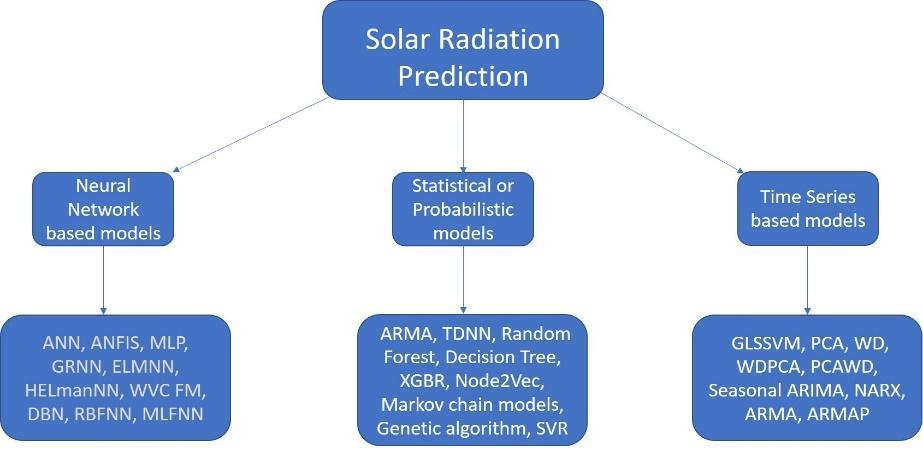


*Figure 2. Taxonomy of methods used for solar energy forecasting*

**Literature Review**

Quej et al. [1], worked on predicting daily global solar radiation at Yucatán Peninsula, Mexico. The data required for training ML models was collected from Mexican National Meteorological Service and Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (INIFAP). Yadav & Sethy [2], presented methods for predicting solar irradiation for the USA. The dataset used was collected by the National Renewable Energy Laboratory, USA. The research by Suyono et al. [3], compared the Solar Radiation Intensity predicted by ANFIS and MLR for Basel City, Switzerland. The dataset used for the study was taken from NASA Meteoblue Climatology website Alqudah et al. [4], worked on Prediction of Solar Radiation Based on Spatial and Temporal Embeddings for San Antonio, Texas. The National Solar Radiation Database was used for training ML models. Ivanova et al. [5], proposed the usage of NARX Model for Solar Radiation Prediction at Barcelona. The data was provided by an industrial company located in Northern Barcelona. Gupta & Singhal [6], worked on Prediction of Global Solar Radiation of Pune, India. The data for the research was collected from the Indian Meteorological Department, Pune (IMD, Pune). Noriega-Angarita et al. [7], proposed methods for Solar Radiation Prediction for Dimensioning Photovoltaic Systems. The data collection was carried out at Columbia weather stations in the Atlantic Coast of Columbia. Works by Madugu et al. [8], focused on Solar radiation prediction for household purposes. The study was completed in Kano city of Nigeria. The dataset provided by the Nigerian Meteorological Agency (NiMet) was used in ML modelling. Bendali et al. [9] studied the comparison of models with genetic algorithms by using the Time series data from the solar plants at Fes, Morocco for hourly forecasting of solar irradiation. Ali-Ou-Salah et al. [10] proposed a new hybrid model for predicting one hour-ahead global solar radiation in the city of Evora located south of Portugal. The weather data is collected from Evora city’s meteorological station using the Eppley Pyranometer. Alsharif et al. [11] study aimed to develop a new model for forecasting solar irradiation whose data is obtained from the Korean Meteorological Administration over a span of daily and monthly average prediction output. Liu et al. [12] studied the short-term forecasting of solar irradiation and analyzed its uncertainties by using historical data from the PV power plant located in Ashland, US. Ghimire et al. [13] proposed the design of a deep learning model from NASA’s GIOVANNI repository of the MODIS satellite and the land observed solar irradiation values were fetched from the Long Paddock SILO database that cover 4 of Australia’s solar cities to forecast long term solar irradiation. Malvoni et al. [14] developed a hybrid Time series model for daily prediction of solar irradiation from the data acquired from PV systems from University of Salento, Italy. Urrego-Ortiz et al. [15] worked on day-ahead forecasting of hourly solar irradiation with the data provided from SIATA station located in Medellin, Colombia.

Ensemble approaches have been a key enhancement in terms of prediction accuracy and have grabbed the attention of many researchers. For example, AlKandari et al. [16] proposed an architecture of model wherein LSTM model and a statistical model such as the Theta model are trained separately on the same dataset and the results are combined to forecast the final forecast. Sometimes, ground data of solar radiation forecasting are not enough as they can be uncertain in some geographical regions, therefore an innovative approach was presented by Carrière, T. [17] where they address this limitation by combining data derived from satellite, such as cloud motion vectors and clear sky index which is then estimated by adding Gaussian noises. Time series forecasting is a great method to model solar energy forecasting but existing methods fail to fully exploit latent spatial dependencies between pairs of variables in multivariate time series. But recent research proves that Graph Neural networks show a high capability to handle relational dependencies, for instance the study developed by Wu et al. [18] models the graph neural networks which automatically extracts the uni-directed relations and a novel mix-hop propagation layer and a dilated inception layer that are further proposed to capture the spatial and temporal dependencies within the time series. There are multiple researches proposing methodologies for predicting direct or beam solar radiation. But predicting diffused solar radiation is quite difficult. In the research by Lou et al. [19], a logistic regression algorithm was employed to predict the horizontal sky-diffuse irradiance and conduct sensitivity analysis for the meteorological variables. The study focused on some important parameters for predicting diffused solar radiation such as clearness index, solar altitude, air temperature, cloud cover and visibility. For Hong Kong and Denver, USA, the mean absolute error (MAE) was less than 21.5 W/m2 and 30 W/m2, respectively. The suggested model is suited to estimate long-term diffuse solar radiation and research climate change. Prediction of Global solar radiation (GSR) is a critical variable for designing photovoltaic systems. Dhakal et al. [20], proposed a cost-effective approach to this problem. Readily available meteorological data at Biratnagar Airport, Nepal was utilized to predict GSR. Extraterrestrial solar radiation, sunshine duration, maximum and minimum ambient temperatures, precipitation, and relative humidity were all factors considered in the study. An Artificial Neural Network along with five other statistical ML models was used for prediction of GSR. The R-Squared calculated was 0.8870 which is very good for cost effective data collection and computationally efficient statistical ML model training.



*Figure 3: Flowchart of the model approaches used for comparative analysis*

**Comparative Analysis**

Artificial Neural Networks (ANN) are multi-layer fully-connected neural nets that consist of an input layer, multiple hidden layers, and an output layer. Every node in one layer is connected to every other node in the next layer and it takes the weighted sum of its inputs, and passes it through a non-linear activation function. Some innovative ANN models are developed to forecast solar irradiation, as given below:

Table 1. Neural Network based models.

| Research paper cite | Targeted Radiation | Forecast Period | Approach | Performance |
| --- | --- | --- | --- | --- |
| Yadav et al. [2] | Solar Irradiation | 5 days average | ANN, Satellite Imaging,  Multi – layered perceptron (MLP) | MSE = 0.0355 |
| Suyono et al. [3] | Solar radiation | Hourly, Daily, Monthly | Adaptive Network based fizzy inference system (ANFIS), MLP | ANFIS is better RMSE = 131.68 MAE = 95.60 |
| Gupta & Singhal [6] | Global radiation | Monthly | ANN | RMSE = 0.5  R-Squared = 0.86 to 0.93 |
| Noriega-Angarita et al. [7] | Global radiation | Monthly | ANN | Correlation index (r) = 0.7 |
| Liu et al. [12] | Global radiation | Hourly | General regression neural network (GRNN), Elman neural network (ELmanNN), Extreme learning machine neural network (ELMNN), Weight carrying combination forecast mode (WVCFM) | WVCFM best; R- Squared = 0.9512 MAPE= 6.05% to 7.07% for all weather types |
| Ghimire et al. [13] | Global radiation | Monthly | Deep belief network (DBN), XGBoost regressor (XGBR), Random Forest (RF), Gradient boosting machine (GBM), Decision tree (DT), Deep neural network (DNN), ANN | DBM best; RMSE=0.609 MAE=0.5 |

Table 2. Time – Series based models.

| Research paper cite | Targeted Radiation | Forecast Period | Pre - processing | Approach | Performance |
| --- | --- | --- | --- | --- | --- |
| Ivanova et al. [5] | Solar radiation | Hourly | Pre-processed data | Nonlinear autoregressive model with exogenous  inputs (NARX) | MSE = 0.0348,  MAE = 0.1360, RMSE = 0.1864 |
| Madugu et al. [8] | Solar radiation | Daily | Taking aggregation (mean) of daily data | ARMA, Auto regressive  moving average process (ARMAP) | ARMAP is better;  RMSE < 0.0004;  SSE < 0.03885 |
| Alsharif et al. [11] | Solar radiation | Daily, Monthly | Removal of outliers, treatment of zero readings, interpolation of missing data | Seasonal Auto regressive integrated moving  average (ARIMA) | RMSE(monthly) = 33.18, R^2(monthly) = 79% RMSE(daily) = 104.26 ; R^2(daily) = 68% |
| Malvoni et al. [14] | Solar radiation | Daily | wavelet decomposition, Principal Component Analysis | Group least square  support vector machine (GLSSVM), PCA, wavelet decomposition (WD),  WDPCA, PCAWD | GLSSVM performs best |

Probabilistic models use statistical techniques to account for the events and these observe the inputs probability distribution to forecast the future events. These also help in understanding and measuring the uncertainty of the predictions, data and models. Uncertainty is prevalent in solar irradiation forecasting and these models have a number of advantages. Some probabilistic models are compared in the table below:

Table 3. Statistical and Probabilistic models.

| Research paper cite | Targeted Radiation | Forecast Period | Pre - processing | Approach | Performance |
| --- | --- | --- | --- | --- | --- |
| Quej et al. [1], | Global solar  radiation | Daily | Interpolation  imputation | Sine wave formula, Cosine wave formula, Gaussian formula | Gaussian formula is  best estimator;  R2=0.868,  RMSE=1.191,  MBE=0.006,  MABE=0.928,  MPE=-0.385%,  MAPE=5.09 % |
| Alqudah et al. [4] | Solar radiation  based on  spatio-temporal embeddings | Seasonal | Min-max scaling  (Normalization), missing value removal,  1-hot encoding | Auto regressive moving average (ARMA), Time  delay neural network  (TDNN), RF, Node2Vec | Random Forest best;  R-Squared for summer=0.91,  winter=0.85, global=0.89 |
| Bendali et al. [9] | Solar irradiation | Hourly | Stationary, historical  lag identification, normalization,  wavelet transform, self-organizing map | Recurrent neural network (RNN), Long short term  memory (LSTM),  Genetic Algorithm | LSTM-GA best; MSE=0.0015, MAE = 0.027 |
| Ali-Ou-Salah et al. [10] | Solar radiation | Hourly | Not mentioned | ANN, GB, Daily  Classification technique + Support vector regression (SVR),  RF | Daily Classification technique + SVR, RF best;  SVR + Daily classification RMSE (sunny)=38.8,  MAE (sunny)=19.72; RMSE (cloudy)=87.05,  MAE (cloudy)=25.37  RF + Daily classification  RMSE (sunny)=39,  MAE (sunny)=21.03;  RMSE (cloudy)=85.77, MAE (cloudy)=50.19 |
| Urrego-Ortiz et al. [15] | Solar irradiation | Hourly | normalization | Markov chain models  Persistence based, 2 chain Markov model | 2 chain Markov model best; RMSE=214 MBE=33.9 |

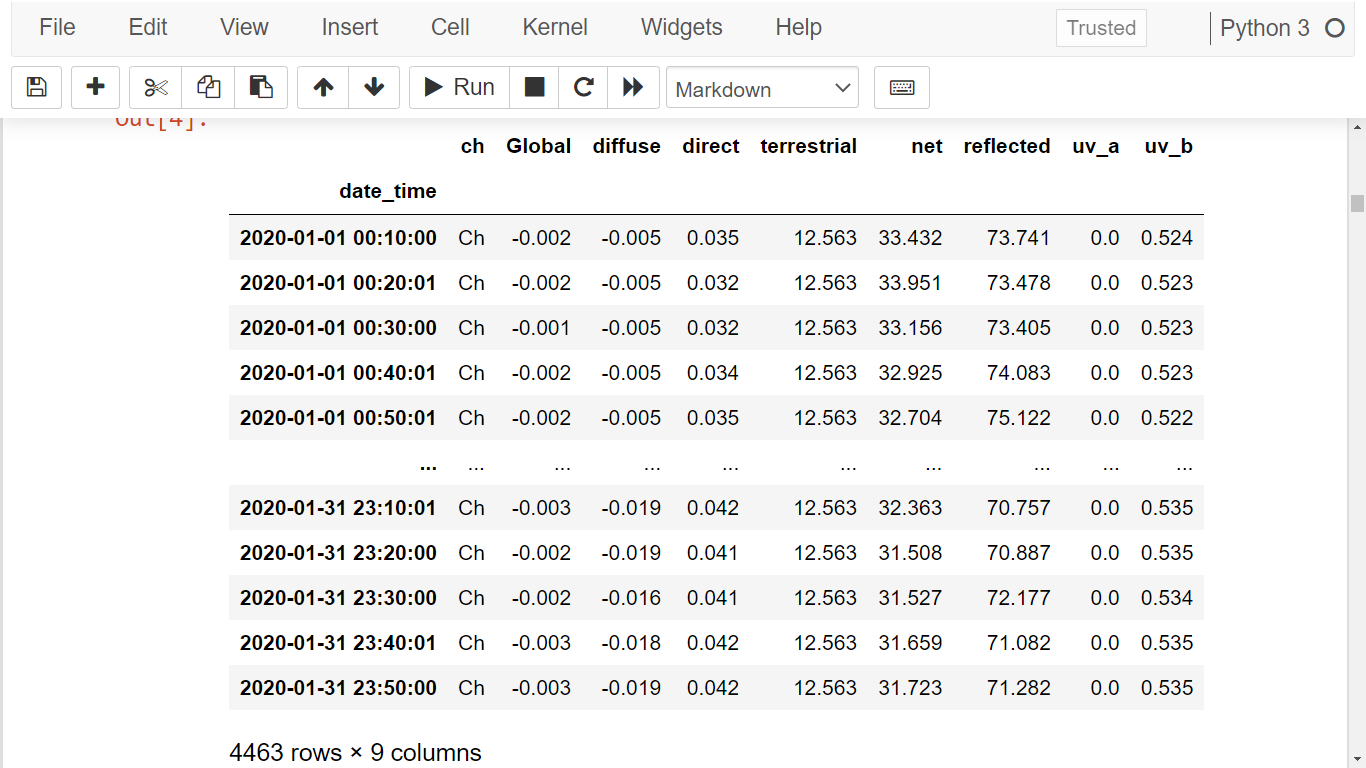
**Dataset (*parameters considered for the study)***

Solar Energy is harvested from Solar Panels installed to benefit from the solar radiations coming from the Sun. These radiations are of multiple forms :

* *Global Radiation* is the solar radiation which falls perpendicularly on a horizontal ground surface.
* *Diffuse Radiation* is the solar radiation which reaches the ground surface after diffusing from dust particles and aerosols.
* *Direct Radiation* is referred as the radiation which reaches the ground surface without diffusion
* *Terrestrial Radiation* is the portion of the natural background radiation that is emitted by naturally occurring radioactive materials, such as uranium, thorium, and radon in the earth.
* *Net Radiation* is the balance between incoming and outgoing energy at the top of the atmosphere. It is the total energy that is available to influence the climate.
* *Reflected radiation* describes sunlight that has been reflected off of non-atmospheric things such as the ground
* *UV Radiation (type A)* causes skin aging. Absorbing higher amounts might lead to skin cancer.
* *UV Radiation (type B)* causes skin burning. It is not as harmful as the UV radiation of type A.

Out of these, Global Radiation contributes most to the production of Solar energy. In technical terms, it is also known as the Global Horizontal Index (GHI). Solar panels have a fixed range of output i.e. 18-22% of the incident GHI. This is a linear conversion rate which can be easily calculated. To accurately forecast solar energy production, we need to predict incident solar radiation. Pyranometer is the device which is used to get accurate readings of incident solar radiation.

Since solar irradiation depends heavily on weather and meteorological data, we got our data from the Indian Meteorological Dept. of Bhopal. Under the mentorship of Dr. Rajeev Gupta and Ved Prakash Sir (data scientist at IMD, Bhopal), we learnt time-series modelling and analytics of solar energy production.

We worked on sample data collected during the year 2020. The time-series data comprises 10 min interval readings of all the above-mentioned radiations along with readings of UV radiations (both A and B type). This work can be further extended to model and predict the higher amounts of UV radiations of type A. 

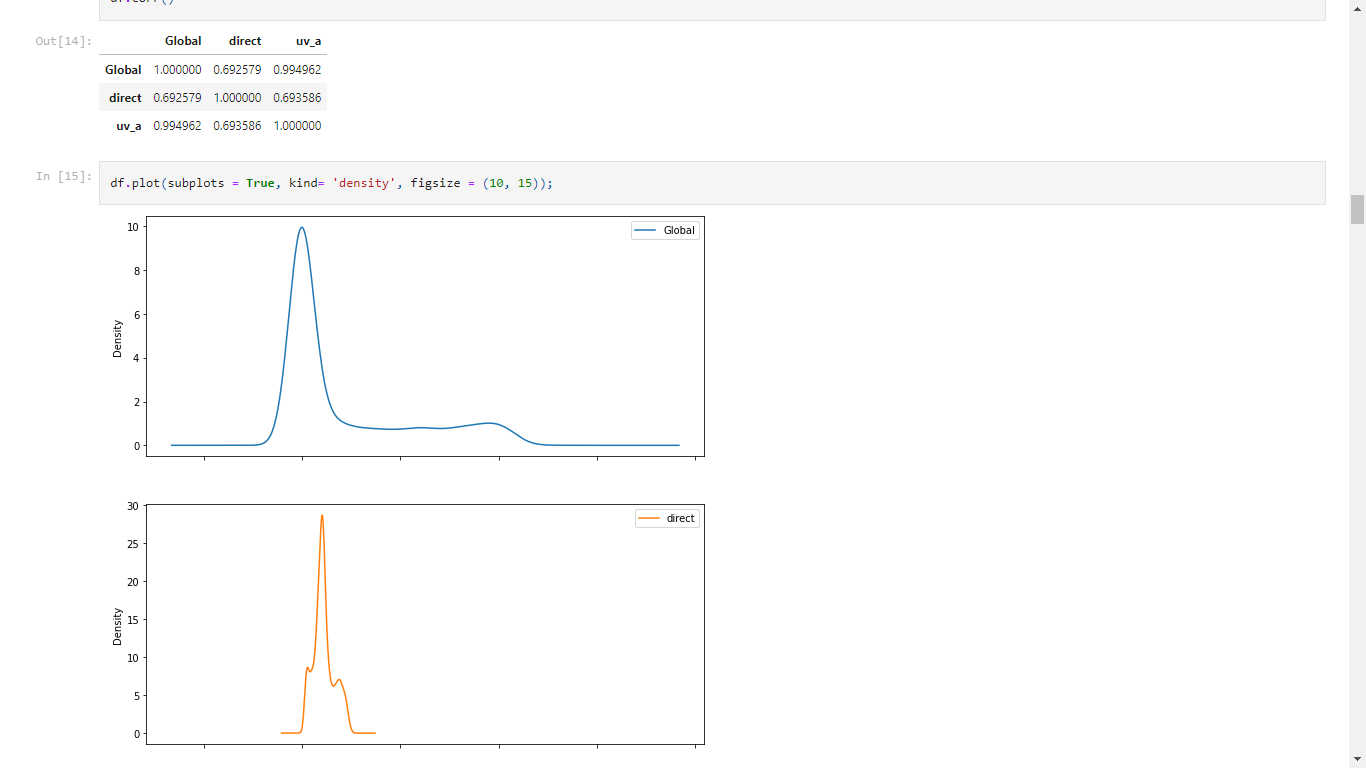
**Proposed Work**

We performed a univariate forecasting on a sample data containing readings of different types of radiation for every 10 minutes in the month of January 2020. In total, 4319 samples were used for model training and 144 samples were used for model validation purposes. Exponential Moving average with exponential smoothing is the statistical model used for training. Reasons for using a statistical model :

* It has way better explainability as compared to other Deep Learning models.
* The data passes the KPSS test as well as ADFuller test of stationarity which proves that the statistical model would outperform the Deep Learning models (LSTM, RNN, Facebook Prophet,…).

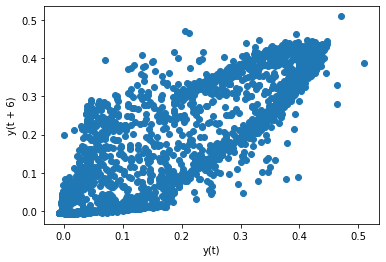
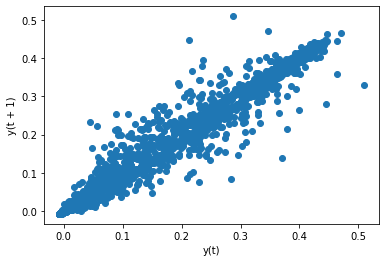
**Implementation Details**

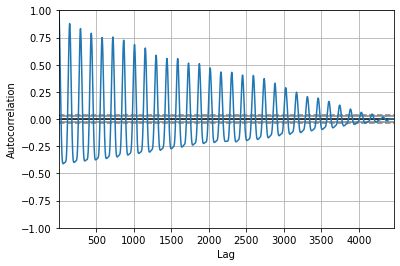
Step 1 : It includes filtering out only the useful data for modelling purposes. We also perform some exploratory data analysis to get in-depth understanding of the data. We also transform some features to fit it appropriately into our model.



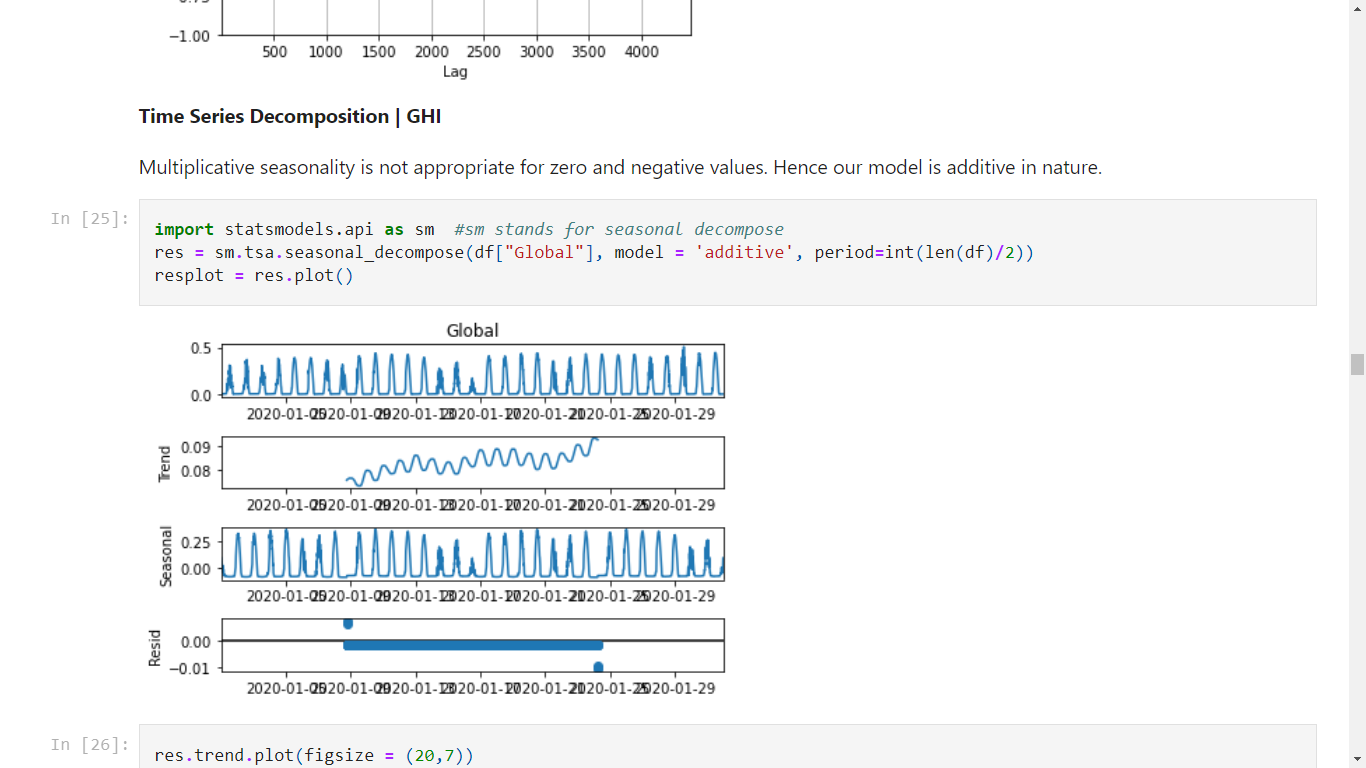


Step 2 : Then we perform lag plot analysis along with plotting auto-correlation plots to check if the data is ready for feeding into auto-regressive models.

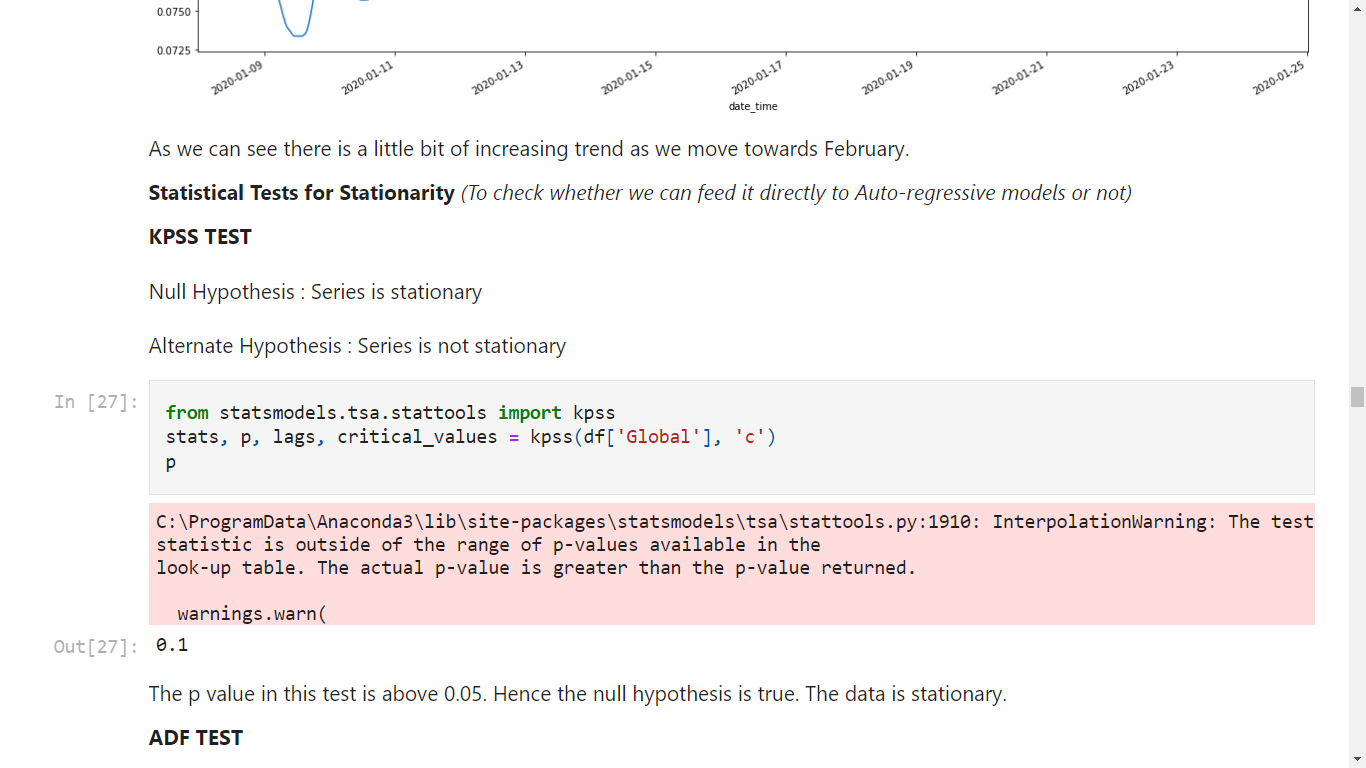


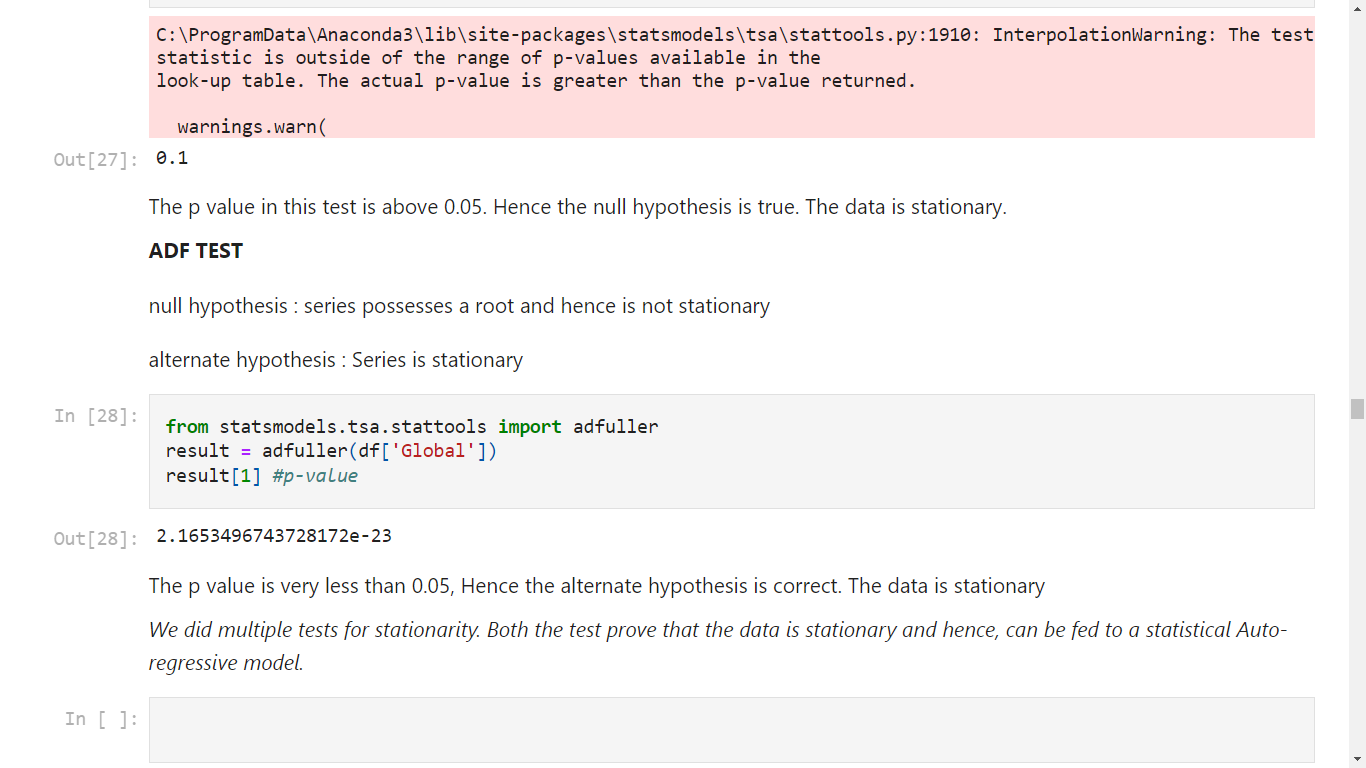


Step 3 : Every time-series data decomposes into 4 components, namely - Trend, Seasonality, Cyclic patterns, and Residual (error). We now decompose our data to get more information.

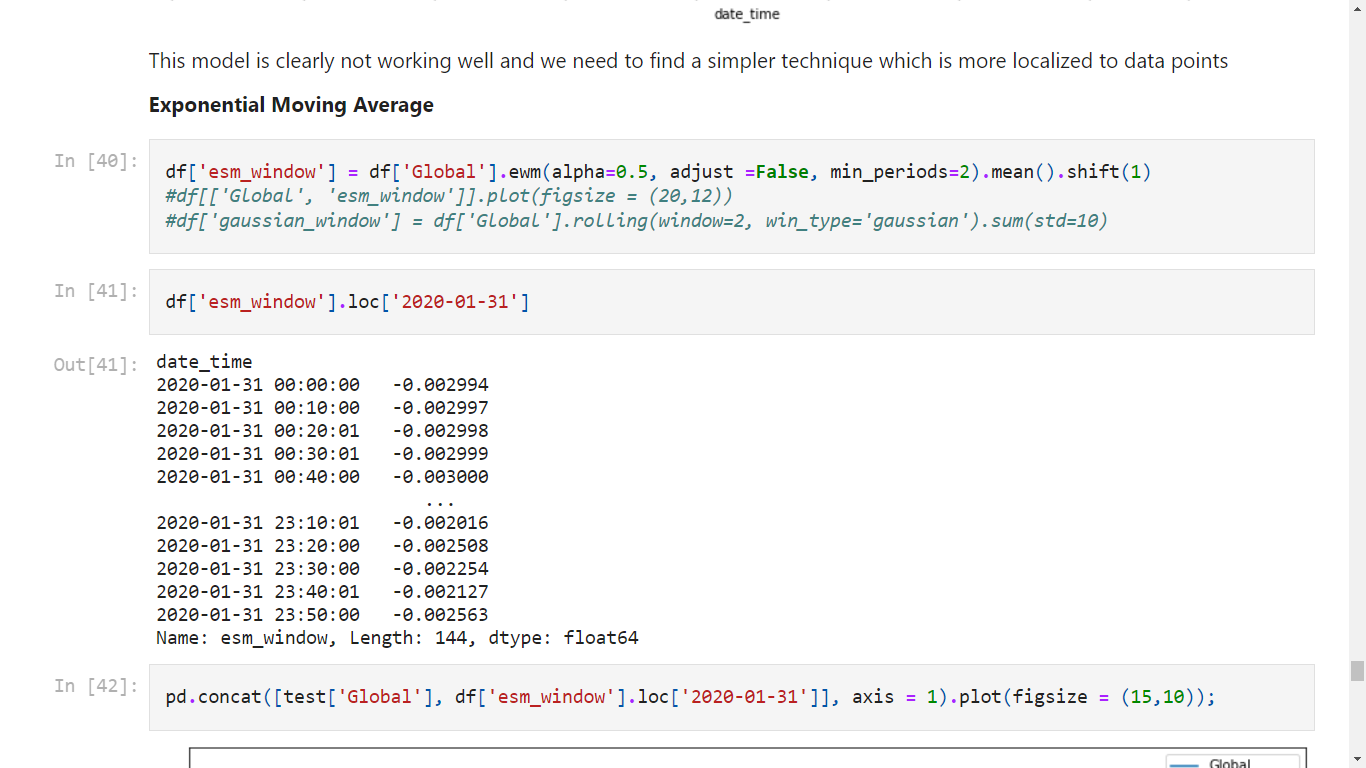


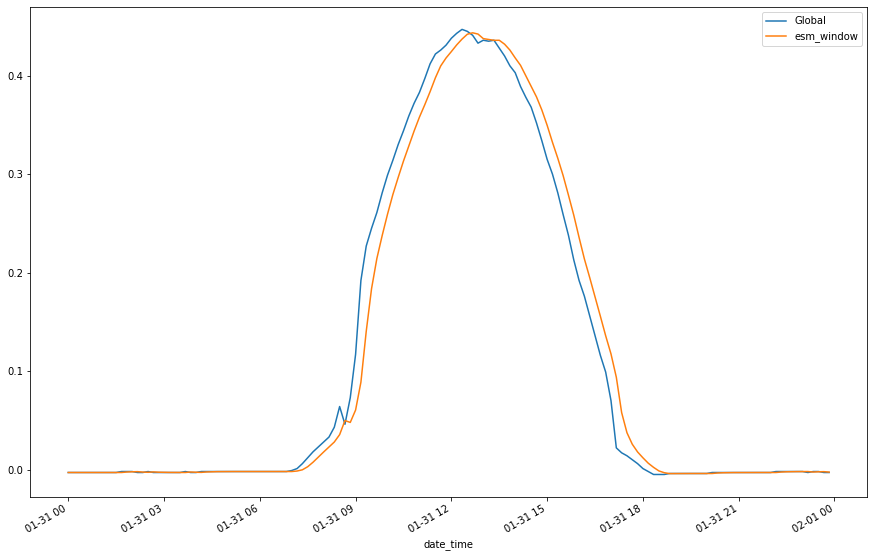
Step 4 : Here we perform statistical tests such as KPSS test and ADFuller test to check whether our data is stationary over our decomposed trend or not. Depending on its result, we select our model.





Step 5 : Depending on the results of above mentioned analysis, we have selected our model. An Exponential Moving Average model with an exponential smoothing of 0.5 and a window size of two worked well.

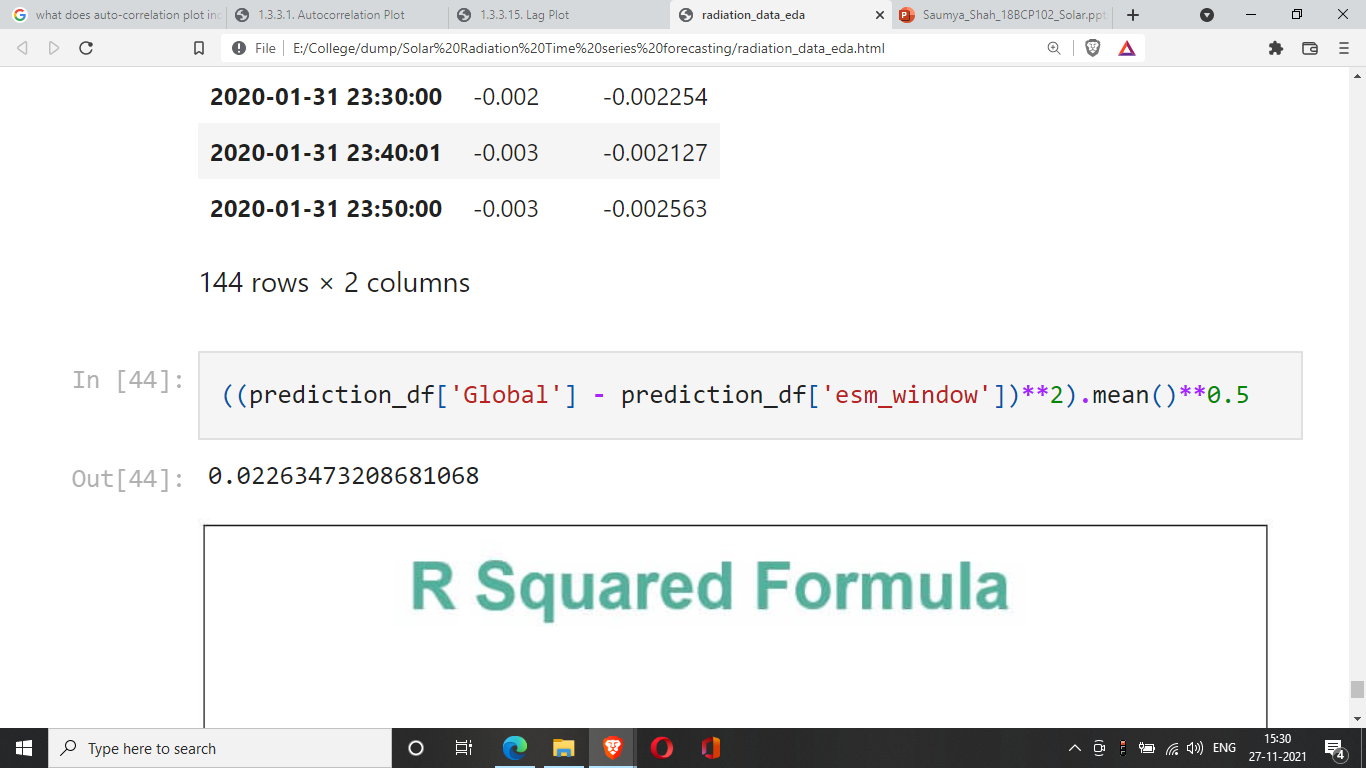




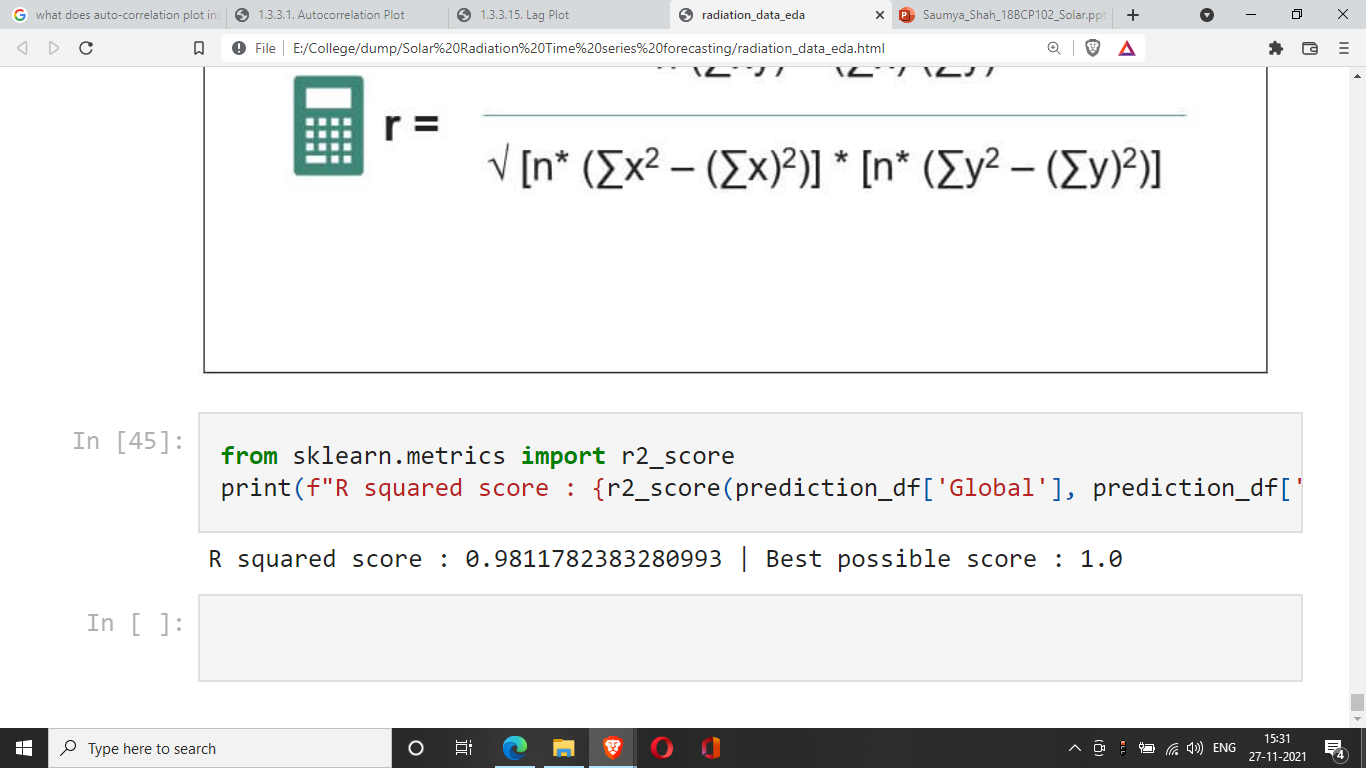
**Results**

The exponential moving average model (trained on the first 30 days of January 2020) was then validated on the dataset containing records of 10 min interval data of 31st January 2020 (144 records). The Root Mean Squared Error (RMSE) and the R-Squared score was observed as follows :

*RMSE :*



*R - Squared Score :*



**Challenges**

Accurately predicting the solar irradiation involves uncertainties related to the characteristics of time series, optimization of ANN models and statistical hypothesis of the Probabilistic models and many weather conditions dependencies make these models highly volatile. Many factors are influenced by the geographical regions and seasons hence no particular model would be generalized. Although cloud imagery and hybrid models are paving the way for newer research models, emphasis on just improving accuracy is not adequate. Models that can optimize the input and relevant parameters without loss of precision could enable better performing forecasting in production. As the residential modules of solar energy generation increase, forecasting their demand and solar irradiation methods could be challenging due to irregular patterns of use and varied environmental factors. While the forecasting of solar energy generation is important, so is the optimization of the electricity grid that stores surplus electricity produced by solar PV. Long term forecasting has not significantly improved in terms of accuracy and precision due to high variance of climatic conditions and different parameters, while the short-term forecasting is observed to have a substantial development in research.

**Limitations and Future Directions**

*Selection Bias :* The proposed model has been trained on a sample time-series data of the month of January 2020. It significantly captures both - trends as well as seasonality components of the data. It can be used to forecast solar irradiation in the Spring and Winter season. But this model can perform very badly on the trends, and seasonality of the remaining part of the year.

*Observational Bias / Modelling Bias :* As we explore the data of the remaining months, we might have to switch from using a statistical model to a Deep Learning model and add multiple types of regularizations to it.

**References**

*[1]* Quej, V. H., Almorox, J., Ibrakhimov, M., & Saito, L. (2017). Estimating daily global solar radiation by day of the year in six cities located in the Yucatán Peninsula, Mexico. Journal of Cleaner Production, 141, 75–82. <https://doi.org/10.1016/j.jclepro.2016.09.062>

*[2]* Yadav, R. K., & Sethy, N. (2018). Solar Irradiance Prediction using Neural Model. International Journal of Engineering and Management Research, 8(3), 241–245. <https://doi.org/10.31033/ijemr.8.3.32>

*[3]* Suyono, H., Hasanah, R. N., Setyawan, R. A., Mudjirahardjo, P., Wijoyo, A., & Musirin, I. (2018).   
Comparison of solar radiation intensity forecasting using ANFIS and multiple linear regression methods.   
Bulletin of Electrical Engineering and Informatics, 7(2), 191–198. <https://doi.org/10.11591/eei.v7i2.1178>

*[4]* Alqudah, M., Dokic, T., Kezunovic, M., & Obradovic, Z. (2020). Prediction of solar radiation based on spatial and temporal embeddings for solar generation forecast. Proceedings of the Annual Hawaii   
International Conference on System Sciences, 2020-January, 2971–2980. <https://doi.org/10.24251/hicss.2020.363>

*[5]* Ivanova, N., Gugleva, V., Dobreva, M., Pehlivanov, I., Stefanov, S., & Andonova, V. (2016). We are IntechOpen , the world ’ s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %. Intech, i(tourism), 13.

*[6]* Gupta, R., & Singhal, S. (2016). Prediction of global solar radiation in india using artificial neural network. Journal of Sustainable Development of Energy, Water and Environment Systems, 4(2), 94–106. <https://doi.org/10.13044/j.sdewes.2016.04.0009>

*[7] Noriega-Angarita, E., Sousa-Santos, V., Quintero-Duran, M., & Gil-Arrieta, C. (2016). Solar radiation prediction for dimensioning photovoltaic systems using artificial neural networks. International Journal of Engineering and Technology, 8(4), 1817–1825.* [*https://doi.org/10.21817/ijet/2016/v8i4/160804234*](https://doi.org/10.21817/ijet/2016/v8i4/160804234)

*[8] Madugu, I. S., Olufeagba, B. J., Adediran, Y. A., Abdulkadir, F., Abdulkarim, A., Inaku, J. U., & Lawan, A. U. (2019). A novel model for solar radiation prediction. Telkomnika (Telecommunication Computing Electronics and Control), 17(6), 3100–3109.* [*https://doi.org/10.12928/TELKOMNIKA.v17i6.12729*](https://doi.org/10.12928/TELKOMNIKA.v17i6.12729)

*[9] Bendali, W., Saber, I., Bourachdi, B., Boussetta, M., & Mourad, Y. (2020). Deep Learning Using Genetic Algorithm Optimization for Short Term Solar Irradiance Forecasting. 4th International Conference on Intelligent Computing in Data Sciences, ICDS 2020, December.*[*https://doi.org/10.1109/ICDS50568.2020.9268682*](https://doi.org/10.1109/ICDS50568.2020.9268682)

*[10] Ali-Ou-Salah, H., Oukarfi, B., Bahani, K., & Moujabbir, M. (2021). A New Hybrid Model for Hourly Solar Radiation Forecasting Using Daily Classification Technique and Machine Learning Algorithms. Mathematical Problems in Engineering, 2021, 1–12.* [*https://doi.org/10.1155/2021/6692626*](https://doi.org/10.1155/2021/6692626)

*[11] Alsharif, M., Younes, M., & Kim, J. (2019). Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. Symmetry, 11(2), 240.* [*https://doi.org/10.3390/sym11020240*](https://doi.org/10.3390/sym11020240)

*[12] Liu, L., Zhao, Y., Chang, D., Xie, J., Ma, Z., Sun, Q., Yin, H., & Wennersten, R. (2018). Prediction of short-term PV power output and uncertainty analysis. Applied Energy, 228, 700–711.*[*https://doi.org/10.1016/j.apenergy.2018.06.112*](https://doi.org/10.1016/j.apenergy.2018.06.112)

*[13] Ghimire, Deo, Raj, & Mi. (2019). Deep Learning Neural Networks Trained with MODIS Satellite-Derived Predictors for Long-Term Global Solar Radiation Prediction. Energies, 12(12), 2407.*[*https://doi.org/10.3390/en12122407*](https://doi.org/10.3390/en12122407)

*[14] Malvoni, M., De Giorgi, M. G., & Congedo, P. M. (2017). Forecasting of PV Power Generation using weather input data‐preprocessing techniques. Energy Procedia, 126, 651–658.* [*https://doi.org/10.1016/j.egypro.2017.08.293*](https://doi.org/10.1016/j.egypro.2017.08.293)

*[15] Urrego-Ortiz, J., Martínez, J. A., Arias, P. A., & Jaramillo-Duque, Á. (2019). Assessment and Day-Ahead Forecasting of Hourly Solar Radiation in Medellín, Colombia. Energies, 12(22), 4402.* [*https://doi.org/10.3390/en12224402*](https://doi.org/10.3390/en12224402)

*[16] AlKandari, M., & Ahmad, I. (2020). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. Applied Computing and Informatics, ahead-of-print(ahead-of-print).* [*https://doi.org/10.1016/j.aci.2019.11.002*](https://doi.org/10.1016/j.aci.2019.11.002)

*[17] Carrière, T., Amaro e Silva, R., Zhuang, F., Saint-Drenan, Y.-M., & Blanc, P. (2021). A New Approach for Satellite-Based Probabilistic Solar Forecasting with Cloud Motion Vectors. Energies, 14(16), 4951.* [*https://doi.org/10.3390/en14164951*](https://doi.org/10.3390/en14164951)

*[18] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 753–763.* [*https://doi.org/10.1145/3394486.3403118*](https://doi.org/10.1145/3394486.3403118)

*[19] Lou, S., Li, D. H. W., Lam, J. C., & Chan, W. W. H. (2016). Prediction of diffuse solar irradiance using machine learning and multivariable regression. Applied Energy, 181, 367–374.* [*https://doi.org/10.1016/j.apenergy.2016.08.093*](https://doi.org/10.1016/j.apenergy.2016.08.093)

*[20] Dhakal, S., Gautam, Y., & Bhattarai, A. (2020). Evaluation of Temperature-Based Empirical Models and Machine Learning Techniques to Estimate Daily Global Solar Radiation at Biratnagar Airport, Nepal. Advances in Meteorology, 2020.* [*https://doi.org/10.1155/2020/8895311*](https://doi.org/10.1155/2020/8895311)