SOLAR RADIATION FORECASTING USING MACHINE LEARNING TECHNIQUES

A PROJECT REPORT

Submitted by

SARATH SASIDHARAN

TKM20EEPS14

to

the APJ Abdul Kalam Technological University

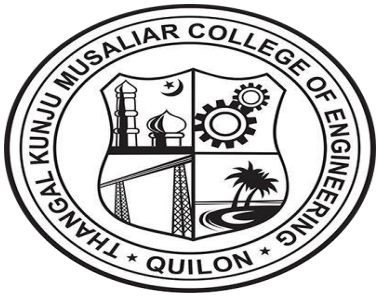
in partial fulfillment of the requirements for the award of the

Degree of

MASTER OF TECHNOLOGY

In

POWER SYSTEMS



DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

T.K.M COLLEGE OF ENGINEERING

KOLLAM-5

2020-2022

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

THANGALKUNJUMUSALIAR COLLEGE OF ENGINEERING

KOLLAM

CERTIFICATE

This is to certify that the project report entitled **"SOLAR RADIATION FORECASTING USING MACHINE LEARNING TECHNIQUES"**is a bonafide record of Project done by **Sarath Sasidharan** ,under our advisory and guidance. This is being submitted to Electrical and Electronics Department of TKMCE, under APJ Abdul Kalam Technological University in partial fulfillment of the requirement for the awardof the Masters in Power Systems.

PROF.BAIJU R NAINA
Associate Professor [Internal Supervisor]
Department of Electrical and Electronics

Dr. SABEENA BEEVI K
Associate Professor [HOD]
Department of Electrical and Electronics

PROF.SHANAVAS T N
Associate Professor [PG- coordinator]
Department of Electrical and Electronics

External Examiner

ACKNOWLEDGEMENT

I am obediently thankful to **God Almighty**, praise and glory is to Him, for all His uncountable bounties and guidance, without which, this project would have never been a reality.

It is my privilege and pleasure to express my gratitude and indebtedness to **Dr. T A Shahul Hameed**, Principal of TKM College of Engineering, and **Dr. Sabeena Beevi.K**,Associate Professor, Head of the Department, Dept of Electrical and Electronics Engineering for providing all the necessary facilities and support.

My heartfelt gratitude to **Dr.Sabeena Beevi K**, Associate Professor, HOD and Internal Supervisor, Department of Electrical and Electronics Engineering, for her constructive guidance and suggestions in designing and implementing this project.

I am greatly obliged to **Prof. Shanavas T N,** Associate Professor, PG Coordinator, Department of Electrical and Electronics, for his encouragement and support.

I express my thanks to **Prof. Jibi P Mathew**, Asst. Professor, Project coordinator, Department of Electrical and Electronics, for his constant support and technical guidance provided during this project work.

I show my extreme gratitude to all Faculty member and Technical staffs in Electrical and Electronics Dept. for providing all the help and necessary facilities to present this project and my deep hearted cheers to my parents and all my friends who extended their support and co-operation towards the successful presentation of the project.

SARATHSASIDHARAN

ABSTRACT

A well-known statistical modelling method named ARIMA has been used to forecast the total daily solar radiation generated by a solar panel located in a research facility. The beauty of the ARIMA model lies in its simplicity and it can only be applied to stationary time series. So, our time series data, which is seasonal and non-stationary, is transformed into a stationary one for applying the ARIMA model. The model is developed using sophisticated statistical techniques. The optimum model is chosen and validated using Akaike information criterion (AIC) and residual sum of squares (SSE). Another method used for solar radiation prediction is LSTM. Long short-term memory (LSTM) models based on specialized deep neural network-based architecture have emerged as an important model for forecasting time-series. existing models are not good at learning long-term historical dependencies, lead to compromise in modeling non- linear solar irradiance patterns. In this paper, a novel prediction model Long Short Term Memory (LSTM) from deep neural network family is used to predict hourly solar irradiance with enhanced prediction accuracy by considering long-term historical data dependencies. The proposed model is compared with Random forest and Extreme Gradient Boost (XGBoost).

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

ARIMA	Auto Regressive Integrated Moving Average
AR	Auto Regression
MA	Moving Average
ACF	Auto Correlation Function
PACF	Partial Auto Correlation Function
AIC	Akaic Information Criterion
LSTM	Long Short Term Memory
XGBOOST	Extreme Gradient Boosting

CHAPTER1

INTRODUCTON

The world's finite fossil fuel reserves cannot keep up with the rising demand for electricity. Additionally, fossil fuels are bad for the environment and a major contributor to the problem of climate change. Increased use and penetration of distributed renewable energy resources are now possible thanks to needs and technical breakthroughs. Microgrid deployment will be widely used, which will only speed up the process. Solar energy, which is based on photovoltaics, has the potential to take the lead in the production of renewable energy in the near future. The process of producing solar energy is stochastic, though, as it depends heavily on environmental circumstances. As a result, forecasting is crucial to the effective management, upkeep, and planning of solar energy sources. Better demand side management is ensured by accurate forecasting. Time series offers a straightforward foundation for predicting even though there are numerous variables involved. Stochastic time series modelling can be used to solve the forecasting problem for solar energy production. When the exploratory variables and their effects on the outcome are unclear, forecasting using time series is especially helpful. As time is the only independent variable and past values are what determine the output, a time series is a fairly straightforward way to illustrate a process. A time series process is modelled using a training dataset, and the same model is then applied to forecast future values. It has been extensively researched how to forecast time series using traditional statistical models. One of the mostpopular methods for predicting time series is the auto regressive integrated moving average (ARIMA). Due to the simplicity of its Box-Jenkins methodology-based implementation, ARIMA has gained popularity. The assumption that the output valuesof the past and present are linearly connected accounts for the simplicity. The simplicity of ARIMA may make it an appealing option, but it is also the reason why it performs poorly in complicated systems, which are scarcely linear

Other non-linear statistical models and cutting-edge machine learning techniques can be used because of the linear assumption of the ARIMA model. Despite this flaw, the ARIMA model can be utilised as a foundation for more complex methodologies' application and comparison

1.1 MOTIVATION

Numerous solar energy applications, as well as economic and environmental factors, are based on information about the world's sun radiation. Accurate worldwide solar radiation data, however, are either difficult to collect or impossible to obtain due to solar radiation variations and measurements not always being accessible. On the other hand, machine learning models can solve extremely nonlinear issues. They are highly valued by academics around the world and have a wide range of possible uses. Any solar project must have access to data on global solar radiation that is at least reasonably accurate. However, because to the expensive cost of measuring equipment and the inability to calibrate it technically, only a small number of meteorological stations worldwide are able to collect these data. Engineers and researchers have created a variety of data generation options in an effort to overcome this difficulty. In order to categorise them and highlight their benefits and drawbacks, we evaluated the techniques used to produce synthetic global solar radiation in this work.

1.2 OBJECTIVE

The main objective of my project is to learn different machine learning techniques for solar radiation prediction and find the most accurate method. For this I want to compare the accuracy of ARIMA , LSTM , XGBOOST , RANDOM FOREST to get better model among these and my final objective is to model a solar panel to to obtain average predicted power output

1.3 ORGANIZATION OF THESIS

El The thesis consist of 7 chapters. chapter 1 gives a brief introduction , motivation and objective of project. Chapter 2 deals with literature review. Chapter 3 deals with ARIMA. chapter 4 deals with LSTM and Core Concept of LSTM . Chapeter 5 deals with Random Forest and chapter 6 deals with XGBOOST. Finally chapter 7 deals with the results and discussion

CHAPTER 2

LITERATURE REVIEW

3.1] Sharif Atique, Subrina Noureen, Vishwajit Roy, Stephen Bayne "Time series forecastingof total daily solar energy generation: A comparative analysis between ARIMA and machinelearning techniques,". 978-1-7281-5017-8/20/$31.00 ©2020 IEEE Transactions on IndustrialInformatics

In this study, the possibility for time series forecasting of the total daily solar energy generation using machine learning-based algorithms has been investigated. The well-known classical approach auto regressive integrated moving average (ARIMA) is first used to model the time series, and its performance is then contrasted with that of two more widely used machine learning techniques, support vector machine (SVM), and artificial neural network (ANN)

3.2] I. Majumder, M. K. Behera, and N. Nayak, "Machine learning methods for solarradiation forecasting ,Power and Computing Technologies *(ICCPCT), April 2017, pp. 1–6.*Weather circumstances have a significant impact on solar forecasting accuracy, hence weather awareness forecasting models are anticipated to enhance predicting performance. In this study, a short-term global horizontal irradiance (GHI) forecasting technique based on uncontrolled clustering (UC-based) is created. First, an Optimized Cross-validated ClUsteRing (OCCUR) approach is used to cluster the daily GHI time series, which establishes the ideal number of clusters and produces the best clustering outcomes

3.3] M. Z. Hassan, M. E. K. Ali, A. B. M. S. Ali, and J. Kumar, "Forecastingday-ahead solarradiation using machine learning approach," *in 2017 4th Asia-Pacific World Congress onComputerScienceandEngineering(APWC onCSE),Dec2017, pp. 252–258*.

The global solar radiation forecasting can be performed by several methods; the two big Categories are the cloud imagery combined with physical models, and the machine learning model. The objective of this paper is to give an overview of forecasting methods of solar irradiation using machine learning approaches The performance ranking different methods is complicated due to the diversity of the data set, timestep, forecasting horizon, setup and performance indicators

3.4] V. P. Singh, V. Vijay, M. S. Bhatt, and D. K. Chaturvedi, "Generalized neural networkmethodology for short term solar power forecasting," ***in 2013 13thInternationalConferenceonEnvironment andElectricalEngineering(EEEIC),Nov2013,pp. 58–62***

The study illustrates the processes employed in machine-learning models for forecasting renewable energy, including data pre-processing approaches, parameter selection algorithms, and prediction performance measures. Analysis of renewable energy sources, mean absolute percentage error values, and coefficient of determination values were performed

3.5]J.WuandC.K.Chan,"Thepredictionofmonthlyaveragesolarradiationwithtdnnand arima," ***in 2012 11th International Conference on Machine Learning and Applications, vol. 2,Dec2012, pp. 469–474.***

The simplicity of the ARIMA model, which can only be used with stationary time series, is what makes it so attractive. In order to use the ARIMA model, our time series data, which is seasonal and non-stationary, is turned into a stationary one. Complex statistical methods are used to create the model. The Akaike information criterion (AIC) and residual sum of squares are used to choose and assess the best model (SSE). To show the effectiveness of the suggested method, error analysis is conducted.

CHAPTER 3

ARIMA

3.1 SeasonalARIMA(SARIMA)

The anticipated output value in an ARIMA model is made up of previous output values and residual errors. Before being modelled using ARIMA, a time series must meet one fundamental requirement: it must be stationary. One of the two ARIMA model variantsseasonal or non-seasonal—is chosen depending on the properties of the time series. Only SARIMA would be discussed in this part because this work only covers the seasonal form. A time series is said to be seasonal when it contains a pattern that repeats itself and S, which expresses seasonality, is numerically equivalent to the periodicity. Equation below uses a product model of seasonality and non-seasonality to mathematically represent SARIMA

$\mathrm{ARIMA}(p,d,q)*(P,D,Q)s$ ----------(1)

P, D, and Q stand for seasonal autoregression, differencing, and moving average orders, respectively, while p, d, and q stand for non-seasonal autoregression, differencing, and moving average orders, respectively. As was already indicated, stationary time series are the only ones that can be used with ARIMA models. To check for time series stationarity, the autocorrelation function (ACF), partial autocorrelation function (PACF), and augmented Dickey Fuller (ADF) test are first utilised. If a non-stationary time series is extracted from the dataset, the required transformations must be carried out to make the series stationary. Only differencing has been used as a transformation approach in this paper. The aforementioned steps can be used to estimate p, P, q, Q, d, D, and S roughly. However, after testing many models under specified constraints, the final values are measured. The preferred model is the one with the lowest value of the Akaike information criterion (AIC), which is used to choose and assess the optimal model.

3.2 ADF TEST

An often-used statistical test to determine whether a particular Time series is stationary or not is the Augmented Dickey Fuller test (ADF Test). When examining the stationary of a series, it is one of the statistical tests that is most frequently applied. The ADF test is a member of the test subset known as "Unit Root Test." In the equation below, when alpha = 1, a unit root is said to exist in a time series.

$$Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon \quad \text{.....(2)}$$

A Dickey-Fuller test is a unit root test that tests the mull hypothesis that $\alpha=1$ in the following model equation

$$y_t = c + \beta t + \alpha y_{t-1} + \phi \Delta y_{t-1} + e_t \quad \text{---------- (3)}$$

- y(t-1)=lag1 of time series
- deltaY(t-1) =first difference of thes eries at time(t-1)

The p-value should be smaller than the significance level (0.05) in order to reject the null hypothesis since the null hypothesis presupposes that unit root, or =1, is present. The test's underlying premise is that, if a unit root process characterizes the series, then the lagged level of the series will not provide any further useful information for predicting the change in addition to that found in the lagged changes. The null hypothesis is not disproved in this instance. The lagged level will provide useful information in predicting the change of the series, however the null of a unit root will indicate that the process is stagnant and demonstrates reversion to the mean when it has no unit root. We discovered how the Augmented Dickey Fuller Test functions and how to apply statistical models to it. Now that you have any time series, you should be able to use the ADF Test to determine whether or not the series is stationary.

3.3 ACF ANDPACF PLOTS

We can use the autocorrelation function (ACF), a statistical method, to determine the degree of correlation between the values in a time series. The lag, which is expressed in terms of a certain number of periods or units, is plotted against the correlation coefficient using the ACF. A lag is the period of time after which the first value in a time series is observed. The correlation coefficient might be between -1 (which indicates a complete negative connection) to +1. (a perfect positive relationship). There is no link between the variables if the coefficient is 0. Additionally, the most common methods of measurement are the Spearman rank correlation coefficient or the Pearson correlation coefficient. It is most frequently used to analyse numerical sequences resulting from random processes, such as those utilised in scientific or economic measurements. Additionally, associated data sets like stock prices or climatic observations can be used to find systematic patterns. An ACF plot shows error bands in blue bars; anything inside these bars is not statistically significant. It implies that correlation values outside of this range are highly likely the result of a correlation rather than a statistical anomaly. By default, the confidence interval is set at 95%. Because the signal is always completely correlated with itself, it is interesting to notethat for lag zero, ACF is always equal to one. By analysing ACF and PACF plots, as we said above, it can often be very difficult and time-consuming to discover the correct order of and for the ARMA model. As a result, there are some simpler methods for tweaking thismodel. Nowadays, the majority of statistical tools provide integrated capability known as "auto ARIMA.". For instance, the auto ARIMA method in Python and R will produce the ideal and parameters by itself, which are suitable for the data set and will result in better predictions. Similar high-level reasoning underlies the hyperparameter adjustment of every other machine learning model. We need to experiment with different combinations of and parameters, then use a validation set to compare the outcomes. We can use a common method for hyperparameter optimization known as grid search because our search space is not large and most values are not higher than 10.Grid search is just a thorough search of a manually chosen portion of a learning algorithm's hyperparameter space. In essence, it means that this method will test every possible combination of and from the subset that we specified. Additionally, we require some objective function that will gauge model performance on a validation set in order to determine the ideal pairing of and. AIC and BIC are typically effective tools for this.

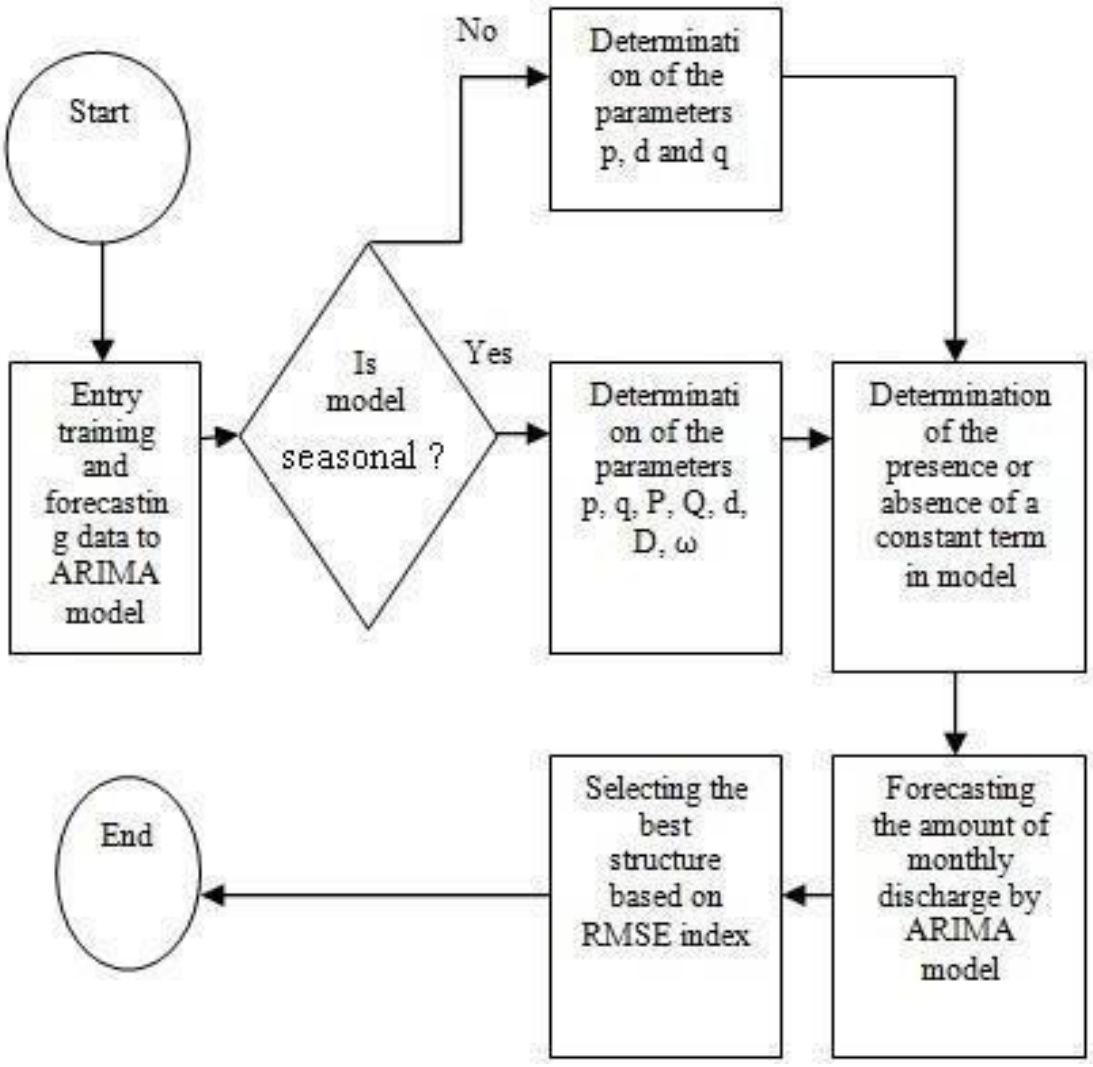


Fig1,Flowchart of ARIMA model

CHAPTER-4

LSTM

Solar The vanishing gradient problem is resolved in a specific type of recurrent neural network, or RNN. In recent years, it has grown immensely popular, and other scholars have further tweaked and enhanced the strategy. The typical RNN, which has a chain-like topology, loops back to the preceding layer. In contrast, LSTMs have four layers that network and communicate with one another. Through specific "gates," errors can spread back over many virtual layers. The reduction of long-term dependency issues served as the driving force behind the introduction of LSTMs. The issue with long-term reliance is that some neural networks struggle to remember and retrieve information after extended periods of inactivity. It involves sampled performance based on a variety of adjusted parameters, including batch and epoch sizes, size and number of hidden layers, and cross-validation techniques.The Recurrent Neural Networks make use of a series of repeating modules. Multiple chains of these repeating modules make up each layer of these recurrent neural networks. Given that it is a repetitive implementation of the atanh layer, as illustrated in the figure, the repeating module's structure is rather straightforward.

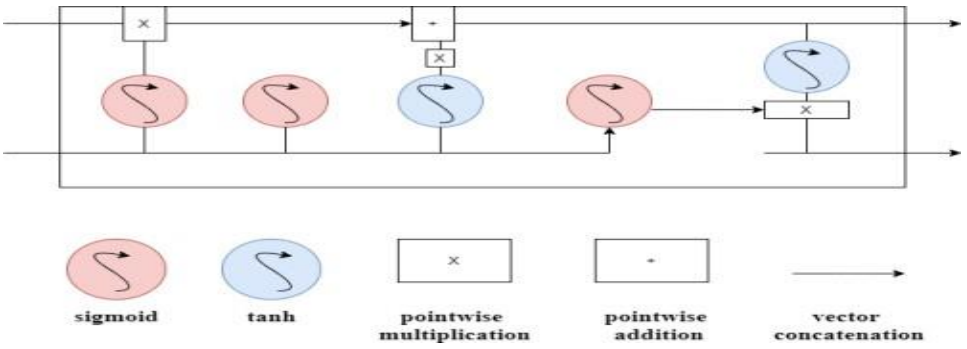


Fig 2, LSTM model functions

The movement of a full vector from one segment to another, where valuable information can be added to the state of the cell and undesirable information is removed, is shown by the straight arrowed lines.The pink circles depict the vector addition and other pointwise operations. The sigmoid activation function, which is similar to the tanh activation function, is located in the gates, and the forked line represents the content being duplicated in multiple spots while the merging lines show concatenation. For the output layer, we have multiplication between the tanh (blue circle) and sigmoid (red circle) layers. The only difference is that the data are compressed between 0 and 1 rather than -1 to 1.By multiplying the quantities by 0, this makes it possible to forget the values. The results of division by one are retained. The outcome after subjecting the values 5, 0.1, and -0.5 after squashing is shown in the figure.

The LSTMs use four different neural network layers instead of just one, each of which interacts with the others in a different way. We can speculate that an LSTM would be the best to capture the trends in time-dependent data because it has feedback loops demonstrating memorization of the temporal behaviour based on the parameters and linking the matrix among the variables

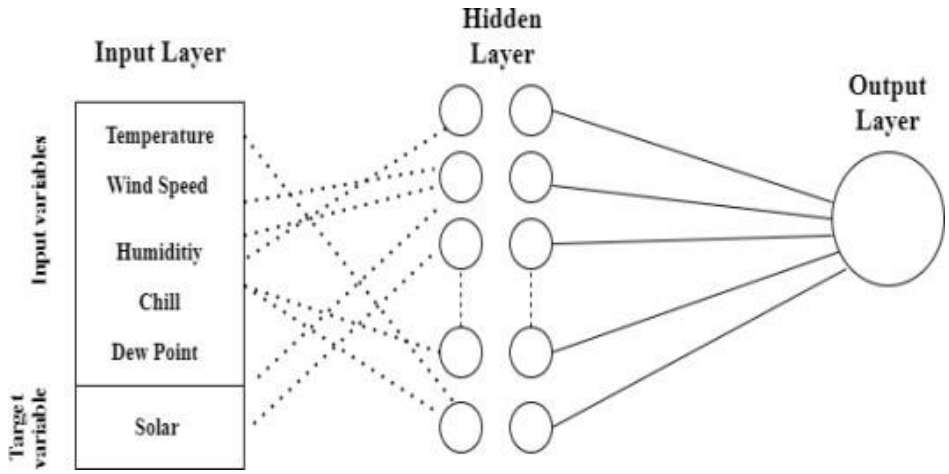


Fig 3, input / output model of LSTM

4.1 CORE CONCEPT OF LSTM

The fundamental idea behind the LSTM approach, along with the multiple gates, is the cell state. The expected output is calculated during the training procedure by allocatingcomparing the weighted average to the solar value. replication through theThe output layer of the network is where the error is calculated. Each time, the result isupdated weights are calculated to reduce the following step

The cell's primary goal is to act as a conduit for the flow of necessary information down the chain. The LSTM network's memory can be informally compared to the cell state. The cells send information to each step in the chain in an impartial manner, making older knowledge accessible to subsequent steps in the chain, which can counteract the benefits of short-term memory. We calculate the gradient of the weight by multiplying weight’s delta and input activations then subtract a ratio of this gradient from the weight to update them. The speed and quality of training affects the ratio it. This ratio is called Learning Rate. The network learns faster if the learning rate is high, but its learning will be more accurate if the network is slow. Learning process is repeated until the accuracy or loss meets a threshold. For the LSTM process, all the data was interpolated and normalized based on the time.

CHAPTER 5

RANDOM FOREST

5.1 FEATURE EXTRACTION

General procedure for the related amount of solar radiation data component breakdown, recombination, and selection is called feature extraction. It is an important component of random forests since it affects not only the robustness, complexity, and efficiency of the system as a whole but also how well interpretable the results of further processing are. In order to produce a score for each feature, eliminating those significant signs, we apply the single variable feature extraction techniques to determine the link between a feature and regression variables. Here, the chi-square approach is used to test the features score.A frequent distribution hypothesis testing technique based on is the chi-squaretest. There is no difference between observed frequency and expected frequency, which is the null hypothesis. The test's fundamental premise is that the value will be determined based on the assumption that H0is true. This shows the degree of difference between the observed values and the theoretical value. We can accurately compute the present statisticsin the case of H0 and the likelihood under more extreme conditions using the Chi-square distribution and degrees of freedom. The degree observation value compared to the theoretical value is too great if the P value is low, hence the null hypothesis should be rejected.We have normalised the original variables, which have been turned into dimensionless data, in order to increase the accuracy of the model predictions because the difference in order of magnitude between these variables is greater. All of the data is mappedto their original values in the range of zero to one using the Min-Max normalisation method.

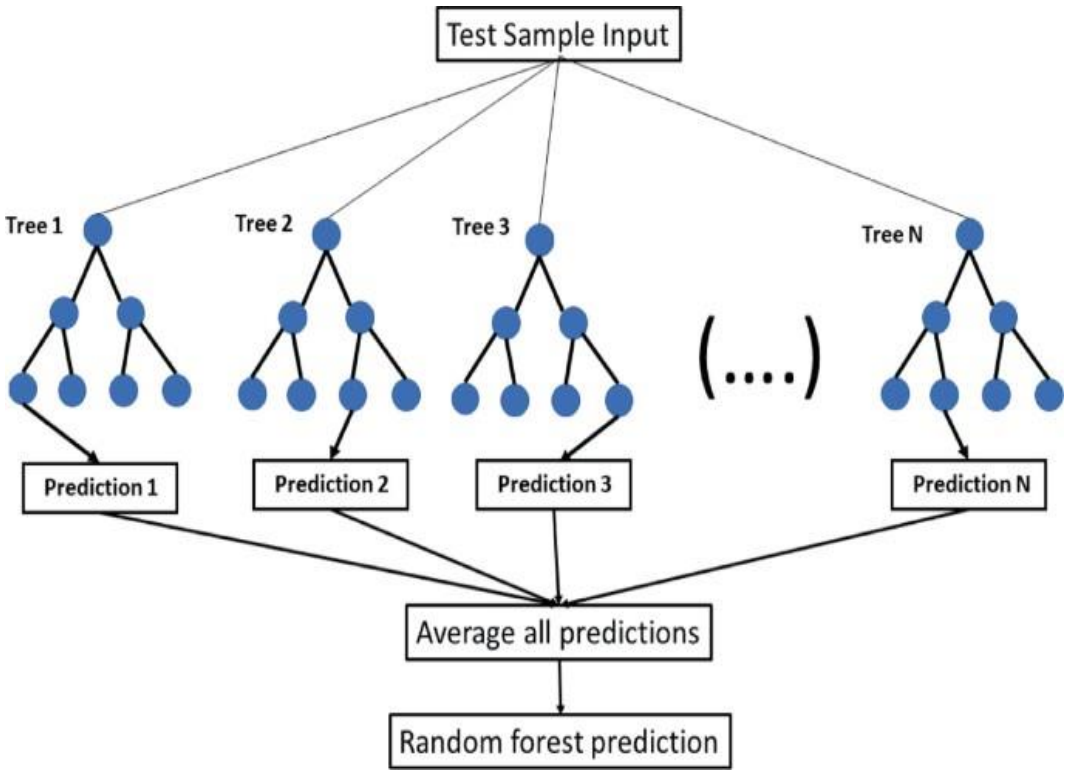


Fig 4, Random forest prediction

We'll engage in some feature engineering. The number of daylight hours available can be represented by compressing TimeSunRise and TimeSunSet into a single column. This feature should be known as DayTime. We can also utilise Time as a feature, but we'll change Time's time format (hh:mm:ss) to the total amount of seconds to make it a numeric variable. We will refer to this feature as Time because it can be completed for us by nesting the lubridate package's hms and period to seconds functions. The day of the year, which can be derived from the UNIX epoch time that we already know and which we will refer to as feature Year.Day, is another feature that we can generate in this context.

CHAPTER 6

XGBOOST

The gradient boosted trees approach is widely used and well implemented in open-source software called XGBoost. Gradient boosting is a supervised learning process that combines the predictions of a number of weaker, simpler models to attempt to properly predict a target variable.Regression trees serve as the weak learners when utilising gradient boosting for regression, and each one of them associates each input data point with a leaf that holds a continuous score. With a convex loss function (based on the difference between thepredicted and target outputs) and a penalty term for model complexity, XGBoost minimises a regularised (L1 and L2) objective function (in other words, the regression tree functions). Adding new trees that forecast the residuals or errors of earlier trees, which are then integrated with earlier trees to produce the final prediction, is how the training process is carried out iteratively. Because the loss when introducing new models is minimised, the technique is known as gradient boosting.

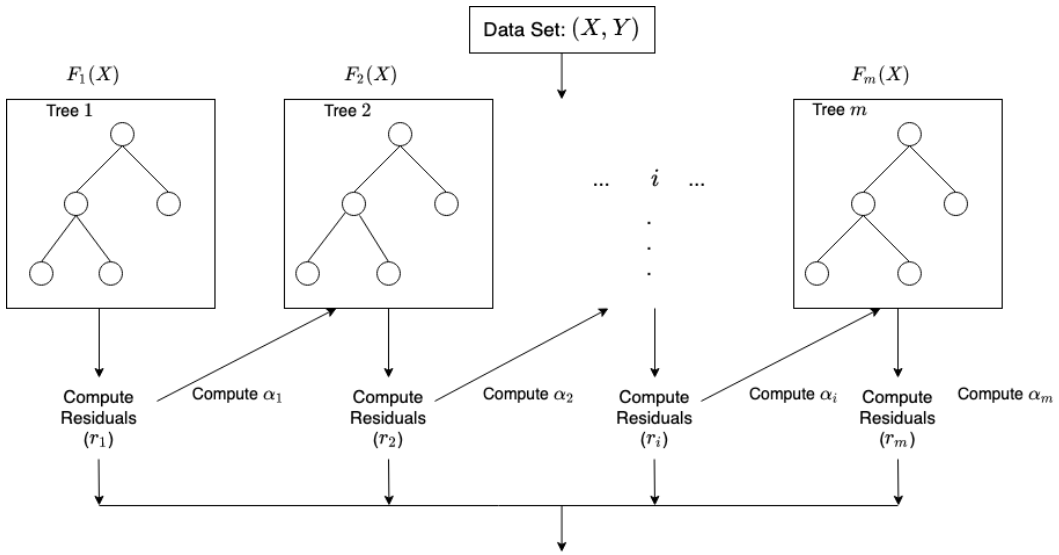


Fig 5, illustration of XGBoost

One of the recent powerful machine learning methods is called Xg-boost (EXtreme gradient boosting). Its accuracy and quickness have conquered the world. In contrast to previous ensemble techniques, it uses parallel and distributed computing, which makes learning extremely quick. The generalised gradient boosting algorithm has been tweaked to create this algorithm. The gradient boosting algorithm creates one type of tree, while the xg-boost algorithm creates another. The split in xg-boost is determined using similarity score and gain. The regularisation option is utilised to prevent the split from being over-fit. The classic gradient boosting approach is used when the regularisation parameter is zero. Two further methods,in addition to regularisation, prevent overfitting.

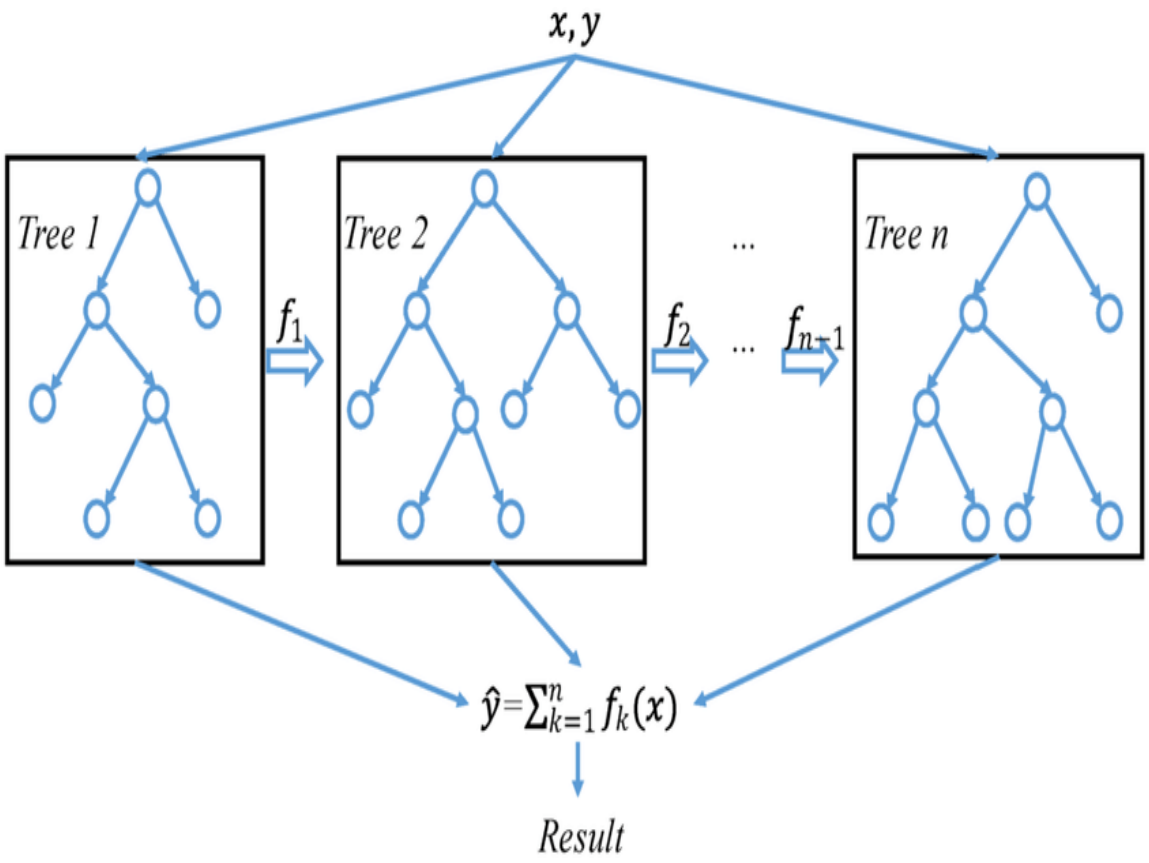


Fig 6, General diagram of XGBoost

CHAPTER 7

RESULTS AND DISCUSSION

7.1 ARIMA RESULTS

In the case of ARIMA first of all we should have to know that the series is stationary or not.So to check the stationarity of the series we should have to conduct ADF testTheADFstatistics and P value from test is shownbelow

So clearly we can say that the series is stationary because the ADF static is negative and p value is less than 0.05.Next we should have to observe the autocorrelation and partial auto correlation plots to obtain the order of auto regression and moving average terms

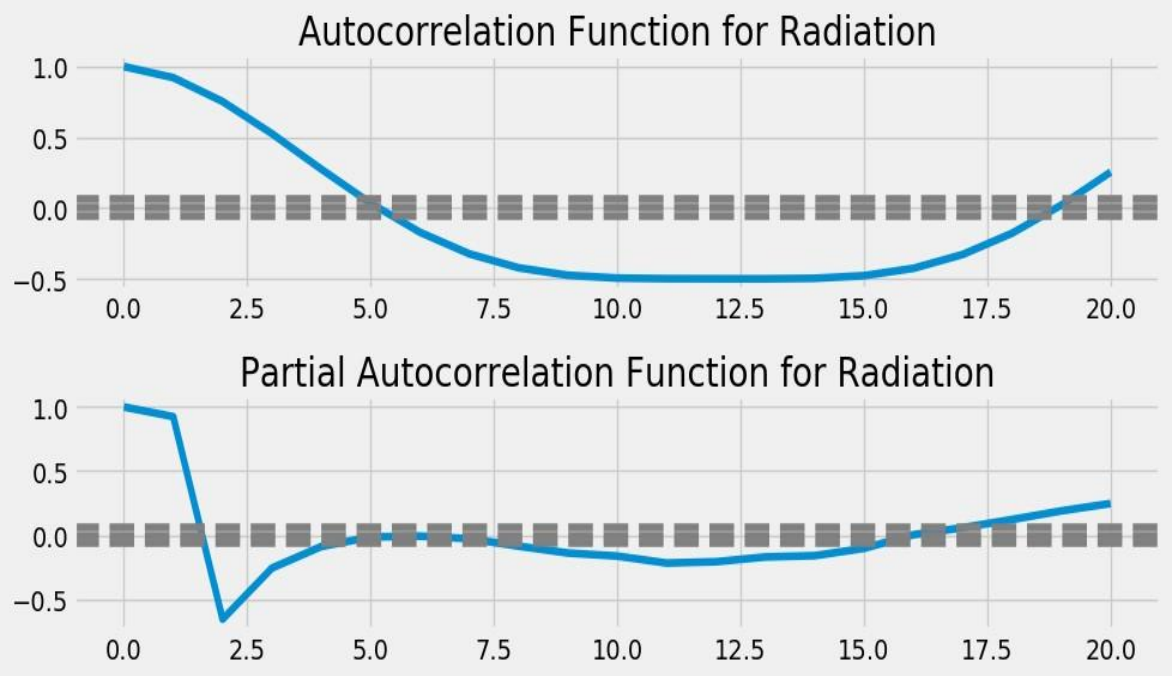


Fig 7 , ACF and PACF plot

From the above plot of ACF and PACF it is clear that the order of auto regression and moving average terms are both unity. Next step is ARIMA modelling, now let us see the actualvspredicted

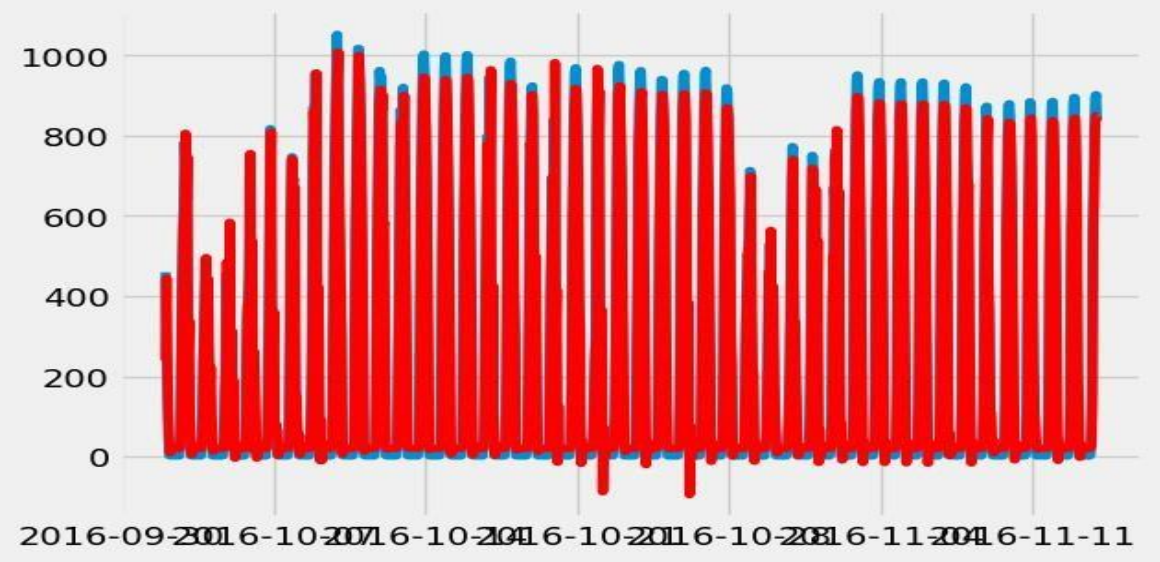


Fig 8 , Actual vs predicted graph of ARIMA

In the above graph actual radiation is shown by blue color and predicted radiation is shown by red color. In most of the cases predicted vs actual curves are closer together but from the curve we can clearly observe that the predicted radiation goes to below zero this condition is practically impossible so clearly we can conclude that the accuracy of ARIMAmodel is less

7.2 RESULTS OF RANDOM FOREST

The actual radiation vs forecasted radiation of RANDOM FOREST is shown below

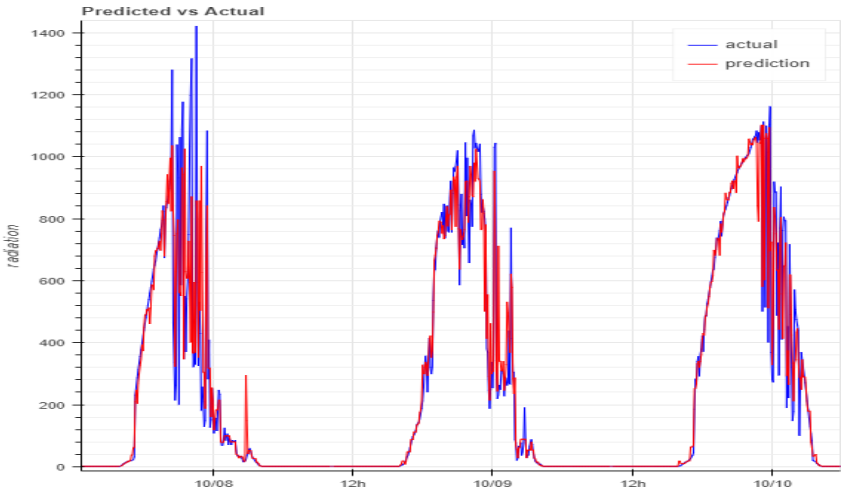


Fig 9 , Actual vs Predicted graph of RANDOM FOREST

From the above graph it is clear that the accuracy of RANDOM FOREST is better than ARIMA because no negative values of radiation. But in the graph some cases actual value of actual value of solar radiation is not close together to predicted value we can observe a significant variation so let us go through another machine leaning model for solar radiation prediction called XGBOOST

7.3 XGBOOST RESULTS

For XGBoost regression we need to find out how all parameters in data set respond to radiation to find the feature values for extreme gradient boost prediction

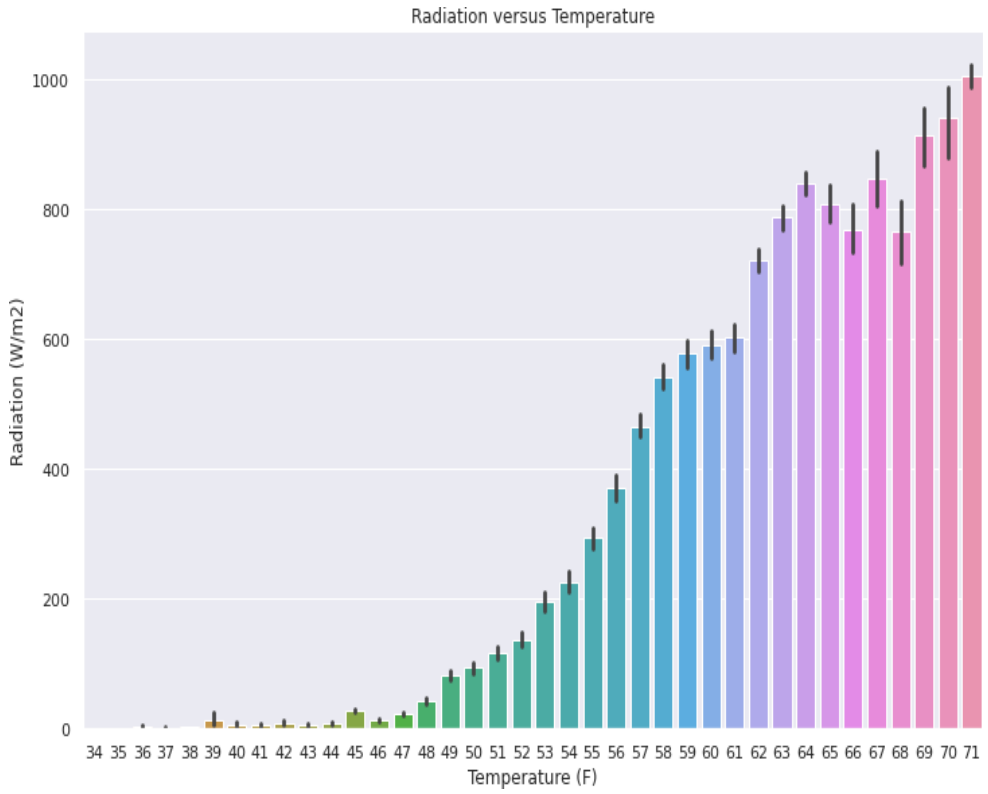


Fig 10, temperature vs radiation

From this graph it is clear that temperature shows a directly proportional relationship with radiation

The graph is logically correct because temperature increases linearly with increase in the availability of solar irradiance. Now let us see how temperature depends on different months in dataset

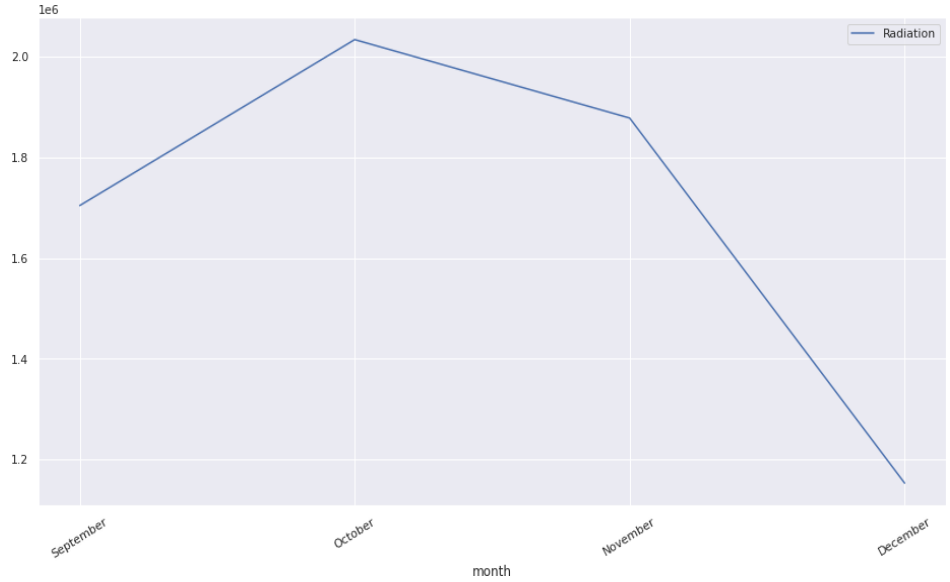


Fig 11, month vs radiation

From this graph clearly we can understand that availability of solar radiation is maximum at the month of October.Now let us see how solar radiation depends on time of the day hour

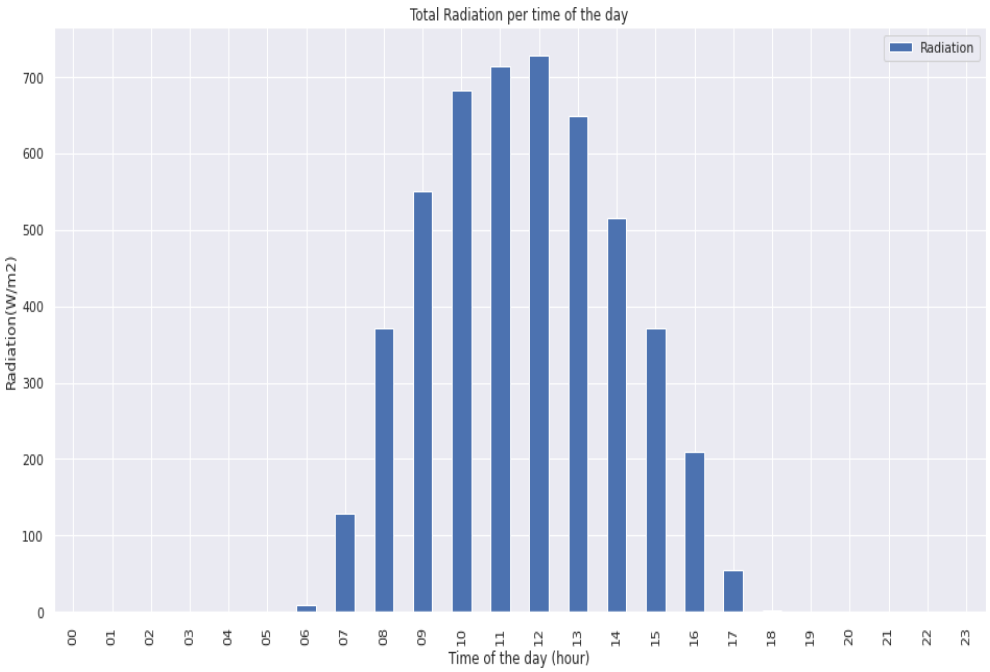


Fig 12, Time of the day vs Radiation

This graph is also logically correct because the availability of solar radiation is maximum at 12.00 pm

Now let us see the feature importance score of XGBoost

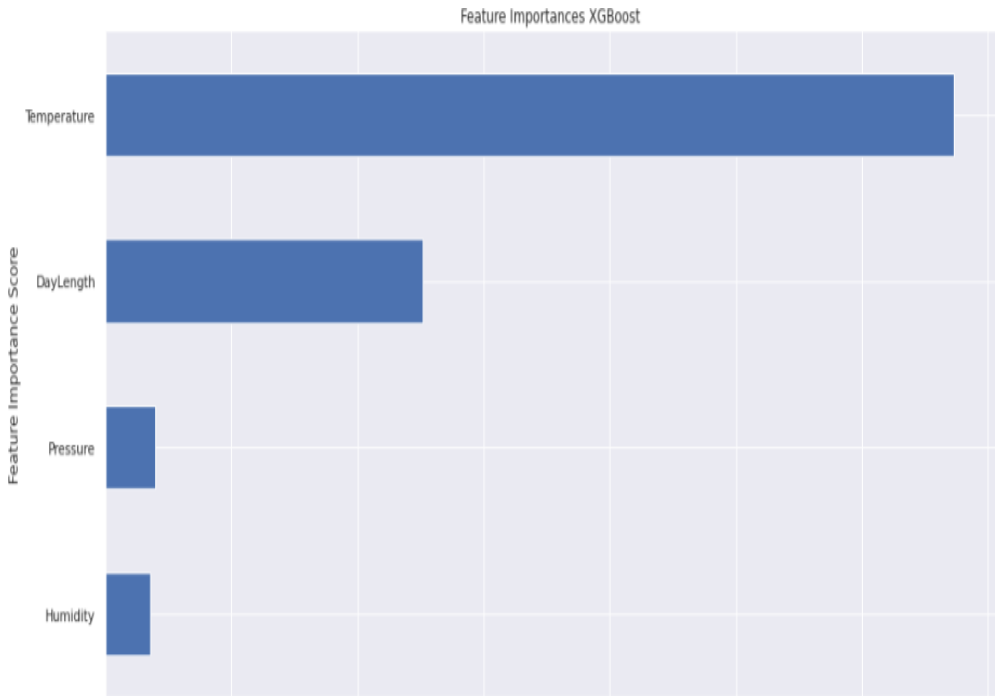


Fig 13, Feature importance of XGBoost

So temperature has more feature importance with radiation than other parameters. Now let us see the actual vs predicted graph of XGBoost

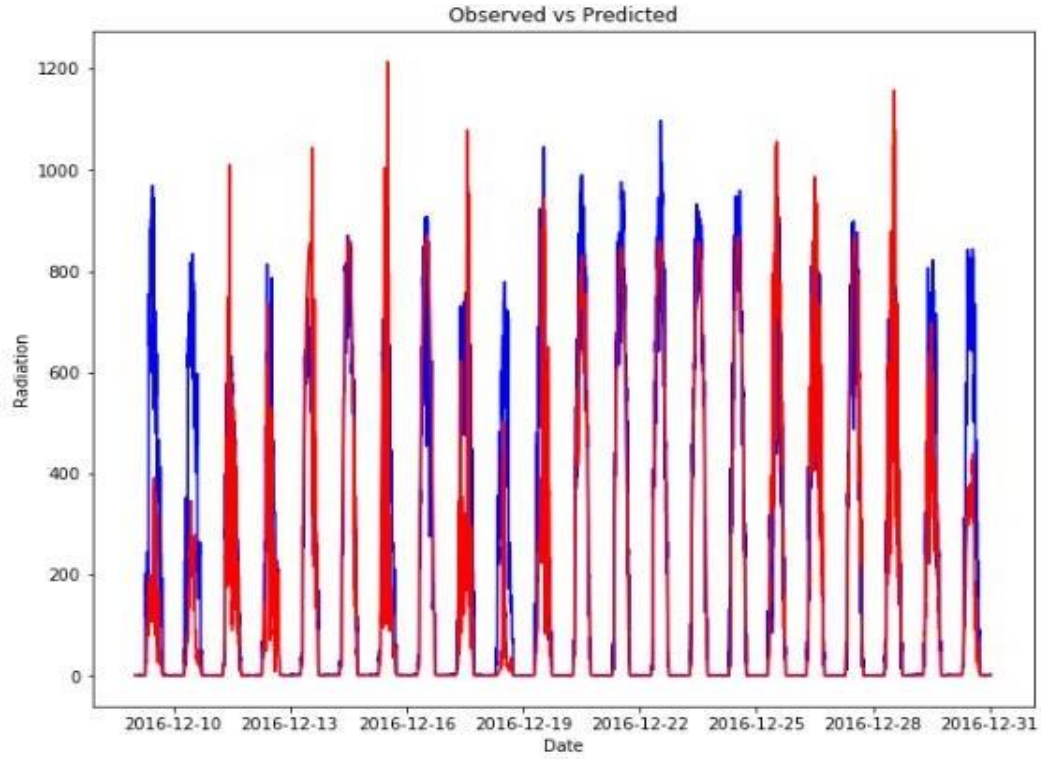


Fig 14, Actual vs predicted solar radiation

The actual value is more closer to predicted value and clearly we can say that XGBoosT method is more accurate than both ARIMA and RANDOM FOREST. Now let us see the results of LSTM

7.4 LSTM RESULTS

In the case of LSTM 70% of the data is given for testing and 30% of the data is given forvalidation

The actual vs forecasted result of LSTM is shown below

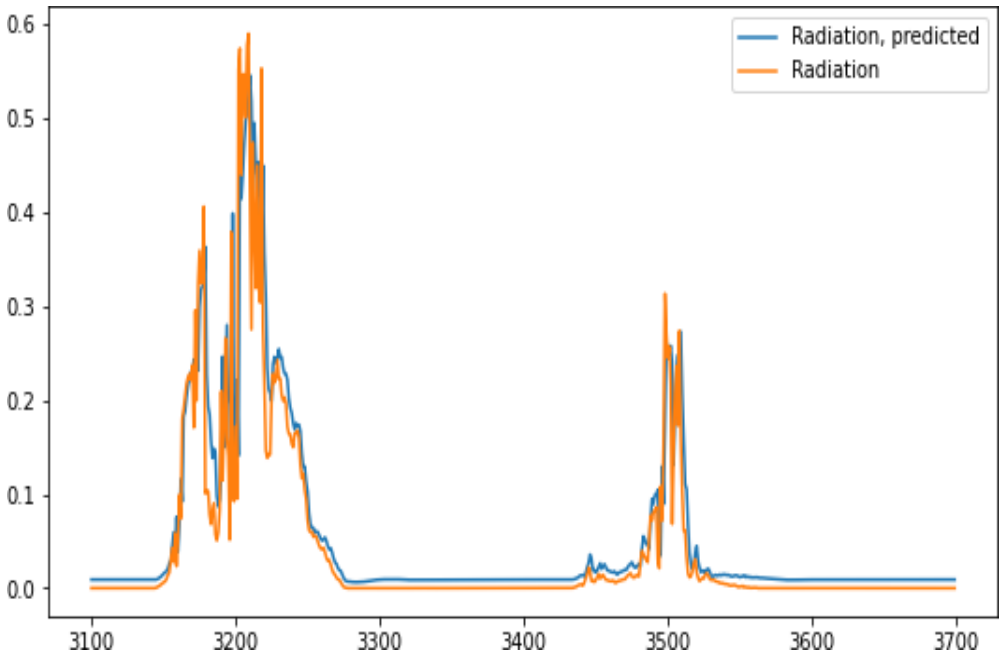


Fig 15, Actual vs predicted graph of LSTM

So clearly from above graph we can say that actual values are very much closer to the predicted Values. The accuracy of LSTM model can be analyzed with the help of training and validation loss. If the training loss is greater than validation loss clearly we can say that the model is accurate. The graph showing training and validation loss is shown below

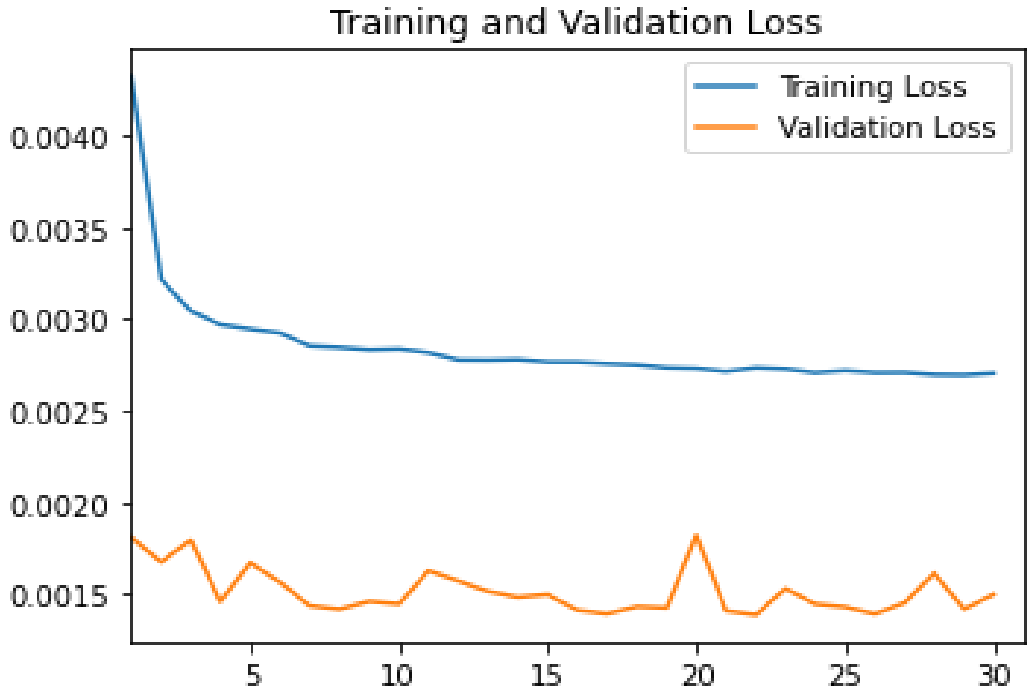


Fig 16 , Training and validation loss

7.5 MODELLING OF SOLAR PANEL FOR AVERAGE POWER OUTPUT

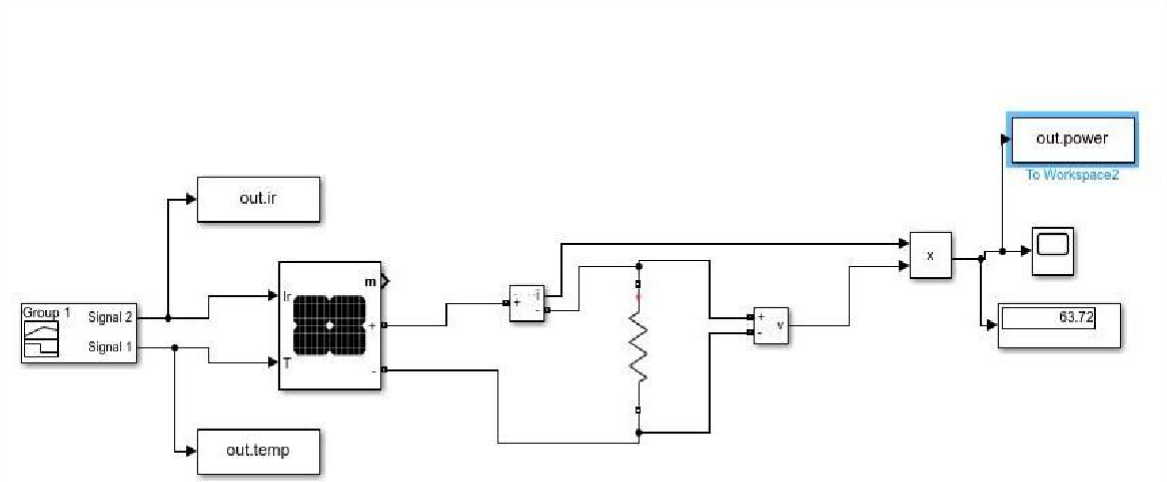


Fig 17 , simulation of solar panel

The data of predicted solar radiation from most accurate method from my models(LSTM) is obtained by using web plot digititizer and predicted output power of two weeks is shown by the following graph

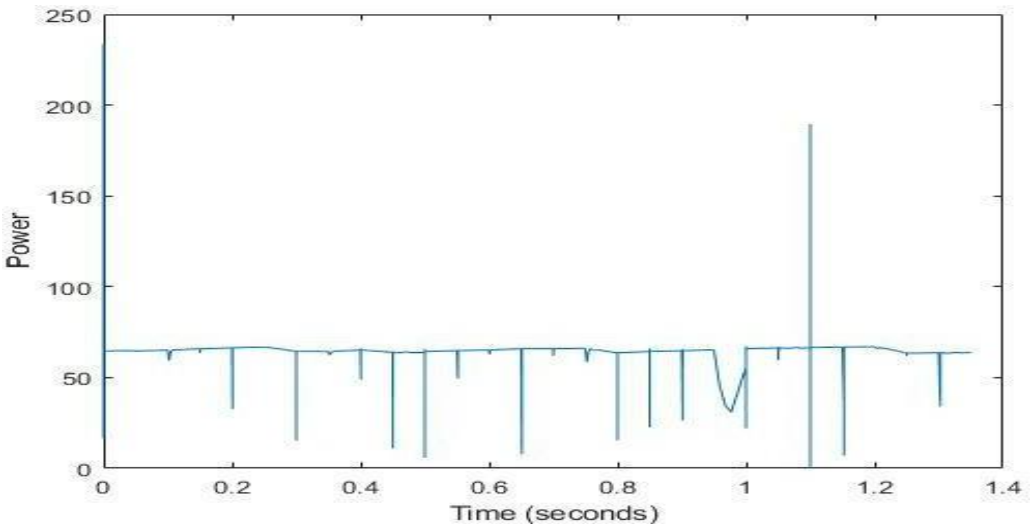


Fig 18 , predicted output power from solar panel

CONCLUSION

Solar radiation prediction using different machine learning techniques like ARIMA (Auto Regressive Integrated Moving Average) , LSTM(Long Short Term Memory) , Random Forest , XGBoost (eXtreme Gradient Boost) is completed. Each models are validated by using corresponding validation techniques .Observed all actual vs predicted graphs of these models and identified that LSTM is the best model for solar radiation prediction. For predicting average power output of a solar panel required modelling and simulation is done in matlab platform .

REFERENCES

[1]SharifAtique,SubrinaNoureen,VishwajitRoy,StephenBayne"Timeseriesforecasting oftotaldailysolarenergygeneration:AcomparativeanalysisbetweenARIMAandmachine learningtechniques,".978-1-7281-5017-8/20/$31.00©2020IEEE, IEEETransactionsonIndustrialInformatics

[2]I. Majumder, M. K. Behera, and N. Nayak, "Machine learning methods for solar radiationforecasting,Powerand ComputingTechnologies*(ICCPCT),April 2017, pp. 1–6*.

[3]M. Z. Hassan, M. E. K. Ali, A. B. M. S. Ali, and J. Kumar, "Forecasting day-ahead solarradiation using machine learning approach," *in 2017 4th Asia-Pacific World Congress onComputerScienceand Engineering(APWC onCSE),Dec2017, pp.252–258*.

[4]V. P. Singh, V. Vijay, M. S. Bhatt, and D. K. Chaturvedi, "Generalized neural networkmethodology for short term solar power forecasting," *in 2013 13th International ConferenceonEnvironment and ElectricalEngineering(EEEIC), Nov2013, pp. 58–62*.

[5] J. Wu and C. K. Chan, "The prediction of monthly average solar radiation with tdnn andarima," *in 2012 11th International Conference on Machine Learning and Applications,vol.2,Dec2012,pp.469–474.*