Amit Verma[1], K. G. Upadhyay[2], M. M. Tripathi[3]

J. Electrical Systems 17-3 (2021): 324-337

Regular paper

Development of Artificial Intelligence Techniques for Solar PV Power Forecasting for Dehradun Region of India

Journal of Electrical Systems

The need of reducing emission of carbon dioxide became possible with the increased penetration of solar photovoltaic (PV) power generation. The variable and intermittent nature of solar PV power generation affects the stability of the distribution grid. It comes to be necessary to forecast the generated PV power to avoid such type of uncertain conditions. This paper presents the empirical comparisons of six developed supervised learning algorithms to predict the solar power generation for the Dehradun region in India. The algorithms namely multiple linear regression (MLP), ridge regression, decision tree (DT), random forest (RF), support vector machine (SVM) and K-nearest neighbor (KNN) are modified in accordance with the higher prediction accuracy. The detailed empirical comparisons of results are discussed on the basis of mean absolute percentage error (MAPE) and root mean squared error (RMSE). It is found that the best performance of RF method with MAPE as 2.2790% and RMSE as 0.8792%. Tree based algorithms have shown the improved performance among all the methods while SVM and ridge techniques perform quietly low.

Keywords: Supervised learning algorithms: Random forest; Decision Tree; mean absolute percentage error (MAPE); root mean squared error (RMSE)

1. Introduction

Now a day's climate change has become a major issue to almost all the countries due to the increasing global temperature by the emission of greenhouse gases. If we keep ignoring the present electricity generation by means of conventional sources such as thermal, nuclear etc., the temperature on the earth may reach to a high level causing the glaciers to melt and it will harm our ecosystem. The clean energy sources such as wind and solar photovoltaic (PV) power are proving to be better replacement of fossil fuel to fulfill our future electricity demands. Solar PV systems are easy to install and use also it would be less costly. Solar PV generation mainly depends on the solar direct and diffuse irradiance [1] which causes the solar panel to generate electric power. It is expected that in future a lot of solar PV generation will be available behind smart meters for trading. The local demand and supply will be met through these available powers and main power from grid will only be used to maintain the security of the grid.

In such emerging scenario the forecasting of solar PV generation is becoming a must to do task so as to increase the penetration of solar PV in distribution system grids. Due to intermittent and variable nature of solar irradiance the solar PV generation is not easy to forecast with high accuracy as per the expectation of the distribution grid operators. However, for proper operation and control of grid it is necessary that we forecast solar

* Corresponding author: Amit Verma, , Faculty of Technology, Uttarakhand Technical University, Dehradun, India amit.v.ashu@gmail.com

[1]Assistant Professor, Faculty of Technology, Uttarakhand Technical University, Dehradun, India

[2]Professor, Electrical Engineering Department, M.M.M.U.T., Gorakhpur, India

[3]Professor, Electrical Engineering Department, Delhi Technological University, Delhi, India

irradiance and electricity in a precisely accurate manner. Accurate forecasting will also help in load management and planning of new solar power plants.

Solar forecasting in context with distribution system market is discussed in [2] describing the forecasting techniques such as time series and ensemble method of prediction. Numerous algorithms of forecasting such as statistical regression, autoregressive (AR), K-nearest neighbors (KNN), decision tree (DT), support vector machine (SVM), artificial neural network (ANN) and combine methods are described in perspective of data driven approach [3]. Artificial intelligent (AI) techniques broadly categorized into three main types as supervised learning, unsupervised learning and reinforcement learning in which the most popular one is supervised learning that includes multiple linear regression, K-NN, SVM, Random forest (RF), DT etc. Performance comparison of ten supervised learning techniques are evaluated and found that RF and SVM shows excellent performance [4]. Gradient boosted regression tree (GBRT) algorithm [5] applied to the data collected from 42 individual PV rooftops and performance was compared with data of a single PV rooftop and it has been found that the combined data gives better result than single one. An investigation [6] of roof top solar PV (installed on small community of Switzerland) potential is done by applying a hybrid method of SVM and geographic information system (GIS). Solar irradiances (global and direct) are predicted by KNN algorithm and compared to the reference persistence method and found that 10% to 25% improvement [7]. Hourly global solar irradiance predicted by applying ANN, DT and SVM in Malaga, Spain [8]. Nonlinear autoregressive recurrent neural network by means of exogenous input (NARX) method has been compared [9] with multilayer perceptron (MLP), auto regressive moving average (ARMA) and persistence method by taking three years of data and concluded that NARX method shows better results in terms of root mean squared error (RMSE) compared to MLP, ARMA and persistence method. A stochastic state space model and uncertainty based model are introduced to predict solar irradiance and PV power by utilizing three different distributions named as Laplace, uniform and Gaussian distributions [10]. It has been seen that the performance of RF technique [11] gives excellent results when compared to the other techniques. A combined method of solar forecasting [12] including nonlinear regression, mean, median, linear regression and supervised learning algorithms applied to solar plant situated in Australia.

This paper presents the forecasting of the solar PV generation for the Dehradun region situated in northern part of India by applying six different supervised learning techniques namely multiple linear regressions (MLP), Ridge regression, DT, RF, SVM and KNN. The correlation between dependent variable solar PV electricity and independent variable direct irradiance, diffuse irradiance and temperature is calculated by using Python software. The performances of all the methods are compared to find out suitable method of solar PV forecasting in Dehradun region. Rest of the paper is organized as follows: section 2 explains the data preprocessing which included the data collection, data cleaning, data selection, data transformation and data mining. Section 3 discusses the developed methodology of six algorithms such as multiple linear regression, ridge regression, decision tree, random forest, SVM and KNN. Section 4 describes the empirical result comparisons of all six applied algorithms and found their performance ranking. Section 5 gives the conclusions drawn from this work.

2. Data description

The data used for this work is obtained from an open source website which includes six different parameters such as time, local time, and temperature, irradiance-direct, irradiance-diffuse, and solar PV power. The following steps are carried out at this stage of research:

- Data collection
- Data cleaning
- Data selection
- Data transformation
- Data mining

2.1. Data collection

One-year data from January 1, 2019 to December 31, 2019 from Dehradun region of India is utilized in this paper to prepare the forecasting model. Total data have 6 columns and 8764 rows in which 6 parameters are arranged columns wise and time series with 1 hour steps in each row. The data contains total six parameters in which three parameters are independent like temperature, irradiance-direct and irradiance-diffuse and is taken as input to the proposed forecasting models. The solar PV generation is a dependent parameter which is taken as target while performing training of different algorithms.

2.2. Data cleaning

Data cleaning includes the procedure to develop data model in consistent format which searches the missing data, finding duplicate values, filling and removing zero values etc. In used data set several rows contain Nan value so we dropped these rows using dropna () method inbuilt in Pandas Library in Python. Some rows comprise zero values, first we counted the total number of zero rows and found that the zero count is not very large so removed the rows which comprises zero value. If zero count would have been large then we could have replaced these zeros by applying other techniques such as mean and median. After cleaning the data new commas separated value (CSV) file is created and clean data is stored in it.

2.3. Data selection

At this stage data decided to analyze linear regression because dependent variable solar PV electricity is a continuous variable. Data set comprises three major attributes for prediction of electricity such as temperature, direct irradiance and diffuse irradiance. Some plots resembling regression plots and heat map are generated by Python to understand data set clearly. Correlation between attributes and target is represented in Fig.1. For better analyzing of data three regression plots are drawn taking the final data sets which shows

Fig. 2, Fig 3 and Fig 4 as the regression plot between solar PV electricity versus irradiance-direct, irradiance diffuse and temperature which clearly indicates the good correlation between these three variables.

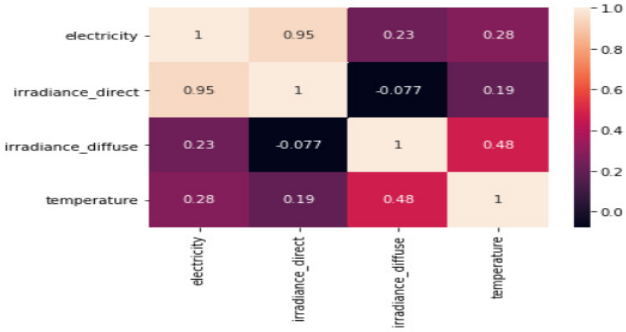


Fig. 1: Heat map showing the correlation between attributes and target

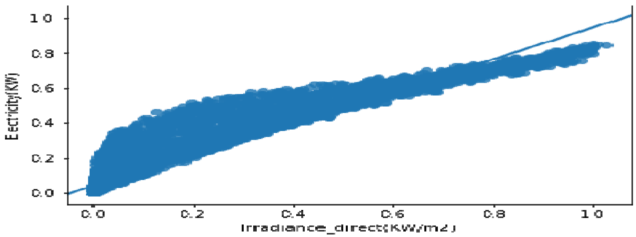


Fig. 2: Regression plot between solar PV electricity and irradiance direct

The fig.1 shows the generated heat map in which the correlation between electType equation here.ricity, irradiance-direct, irradiance-diffuse and temperature are presented. The correlation between electricity and irradiance-direct is highest with a value 0.95 while the correlation between irradiance-diffuse and electricity has the value of 0.23. We have used the Spearman Rank Correlation coefficient 'r' as shown in eq. 1 below.

$$r = \frac{\sum (I - I_{mean})(O - O_{mean})}{\sqrt{\sum (I - I_{mean})^2 \sum (O - O_{mean})^2}} \qquad (1)$$

Where I is the different input variables, I_{mean} is the mean value of input variable, O indicates the dependent output variable and O_{mean} denotes the mean value of the dependent output variable.

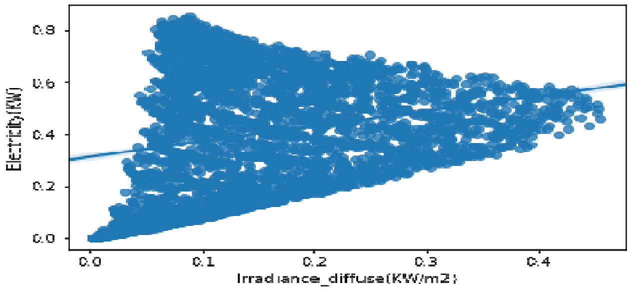


Fig. 3: Regression plot between solar PV electricity and irradiance diffuse

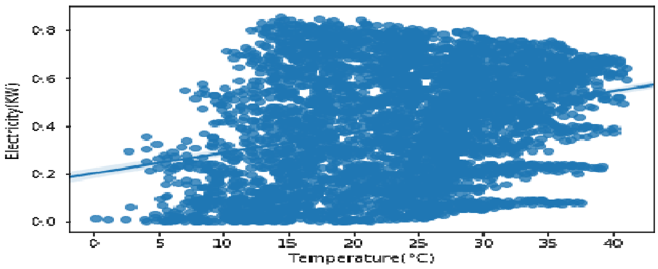


Fig. 4: Regression plot between solar PV electricity and temperature

Regression plots have drawn by taking the processed data shows the best regression line between solar PV electricity and direct radiation.

2.4. Data transformation

Data transformation is also known as data normalization. At this stage selected data is transformed between the ranges of 0 to 1which makes training less sensitive for computation. Fortunately, our data of solar PV electricity, direct irradiance and diffuse radiation was already in the format between ranges of 0 to 1. We applied the min-max scalar transformation formula to temperature (°C) data as given by eq. 2.

$$Nx = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (2)$$

Where X and Nx is the actual and normalized value while min(X) and max(X) are the minimum and maximum value of the temperature. This task is carried out with the help of sklearn library and data is saved in csv file format. The preprocessed data are ready for the data mining purpose.

2.5. Data mining

This is the last stage of data preparation in which the final normalized meteorological data is sectionalized in three parts as percentage of 70%, 10% and 20% for the training, cross validation and testing purpose.

3. **Performance of applied algorithms**

The various algorithms used in the proposed work are discussed below along with their performance.

3.1. Multiple linear regression

Linear regression is the most popular and frequently used algorithm that is generally used for developing a forecasting model. Regression estimates the dependent variable on the basis one or more independent variable by plotting the most fit line on the basis of the available data. The fig. 3 illustrates example of best fit line where the plot between actual electricity generation and predicted electricity generation is obtained. The line can be a straight line or curved type depending the algorithm used. The best fit line can furthermore be the quadratic or polynomial which gives us to better predictions. Here we used the "LinearRegression()" method from sklearn module to generate linear regression object. The applied function namely "fit()" is used to draw the best fit regression line, considering the dependent and independent variable data. This function is responsible for training the model and doing remodeling and concatenation operation. The "predict()" method is used to forecast the testing data set and comparing actual and predicted value to obtain the mean squared error (RMSE) and mean absolute percentage error (MAPE) indicating the performance of the algorithm. Fig. 5 shows the regression plot between electricity and irradiance direct after applying linear regression

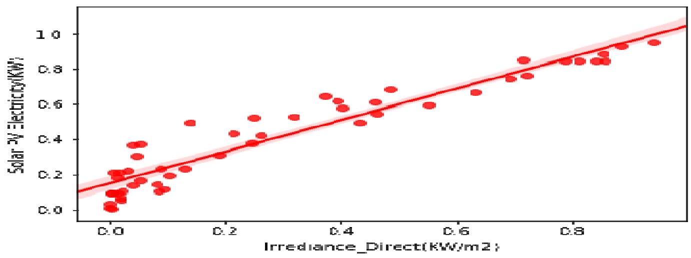


Fig. 5: Plot after applying multiple linear regression

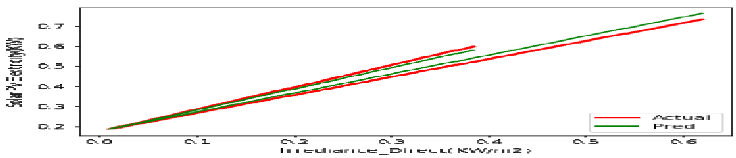


Fig. 6: Plot after applying the multiple linear regression algorithm

Fig.6 illustrates the Solar PV electricity forecasting after applying the multiple linear regression algorithm. The multiple linear regression equation for one output variable and three input variable can be given by eq. 3.

$$y = m_0 + m_1 x_1 + m_2 x_2 + m_3 x_3 + \varepsilon \quad (3)$$

where x_1, x_2 and x_3 are the input variable and m_0, m_1, m_2 and m_3 are regression coefficients, y is dependent variable and ε is error term.

3.2. Ridge regression

Ridge regression is also known as Tikhonov regularization. A regression model is solved by using this model in which a loss function is linear least square type and regularization is specified by the I^2 norms. This estimator has built-in support for multivariate regression. Fig. 7 shows the data plot between solar PV electricity and actual solar PV electricity.

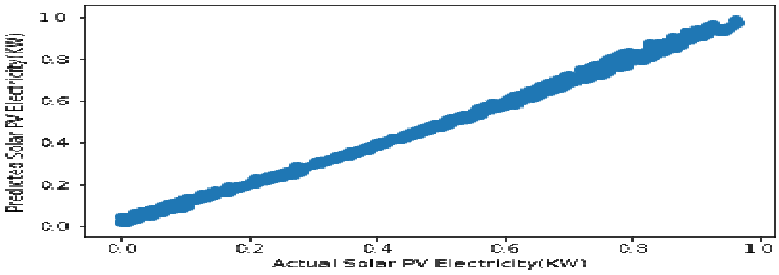


Fig.7: Forecasting of solar PV electricity by applying ridge regression

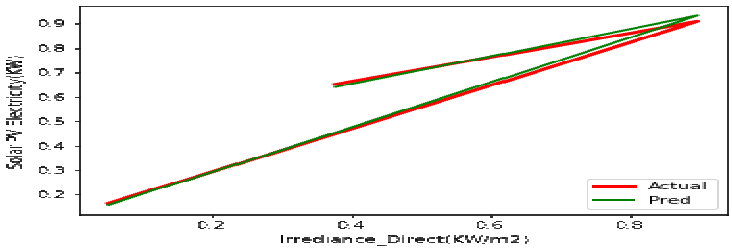


Fig. 8: Plot after applying ridge regression

Result shown in Fig. 8 is between actual and predicted value after applying ridge regression. The result improves slightly in terms of RMSE, when Ridge regression is applied. Hyper-parameter of Ridge symbolized as alpha (α) which means that these models are not automatically learned whenever they have to be set manually. Here we consider α equal to 1. It can be understood that as the value of α increases the magnitude of coefficient decreases where the value approaches to zero. If RMSE is calculated for each value of α we can observe that the value of RMSE will be optimum when α is equal to 1. So its value is chosen wisely by iterating it through a variety of values and using the one that gives us lowest error. The cost function 'C' of Ridge regression can be taken as given in eq. 4.

$$C = \min \left(\|Y - X(\theta)\|_2^2 + \alpha \|\theta\|_2^2 \right) \tag{4}$$

The extra term denoted here is the penalty term which can be controlled by α. For large value of α, penalty will be high and hence reduce the coefficient value.

3.3. Decision tree

Decision tree is a supervised learning algorithm that utilizes a tree like structure in which node represents the test of any decision, outcome denotes the branches and leaf as a target. The independent variable data set can be spitted into subsets based on a certain test.

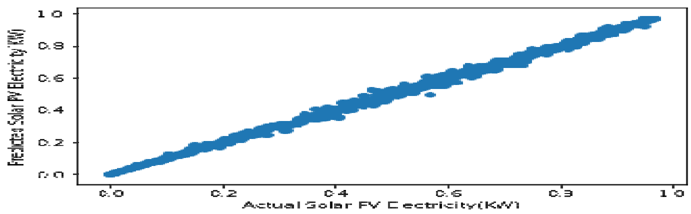


Fig. 9: Forecasting of solar PV electricity by applying decision tree algorithm

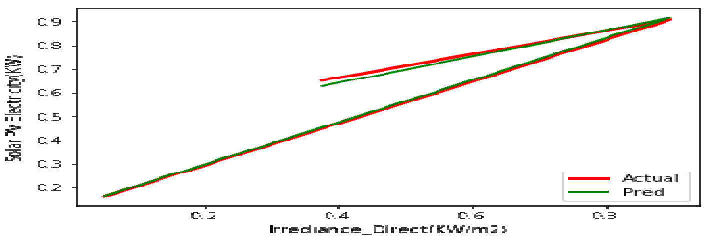


Fig. 10: Plot after applying decision tree

This process is iteratively repeated for each step of test until the target value achieved. DTs can be designated by various statistical and computational methods to assistance the categorization of datasets. Classification and regression tree (CART) is the most popular term used in tree type of structures in which classification indicates the category and regression includes the numerical data. The DT algorithms for specific conditions are named as Iterative dichotomiser 3 (ID3), C4.5, multivariate adaptive regression splines (MARS), chi- squared automatic interaction detector (CHAID) etc. (Rachit Srivastava et al., 2019). Training data sets are split into many nodes and homogeneous splits are preferred which can be tested by node impurity metrics such as Gini index, Entropy and Misclassification error. We used here the “DecisionTreeRegressor” library to predict the solar PV electricity. The output generated is shown in Fig. 10, which shows better results at lower and higher values of direct irradiance while in middle values the error is large. The plot between the predicted and actual electricity in kilowatt (KW) after applying decision tree algorithm is shown in Fig. 9.

3.4. Random forest

Ensemble learning techniques are the combination of many forecasting models which provide better results as compared to the single model. Random Forest (RF) is an ensemble learning which combine the various decision trees to boost their performance when they are randomly selected. The main features of RF is to decide number of regression trees, their max depth and the finally leaf node. It also controls the problem of over-fitting. If bootstrap equals true then sub-sample size can be controlled by the “max_samples” in Python programming, otherwise total dataset is used to construct each tree. This technique is employed by the “RandomForestRegressor()” library in Python. The output (Y) of RF can be calculated by averaging the prediction of selected regression trees t_n of N bootstrap samples and is given by eq. 5.

$$Y = \frac{1}{N}\sum_{n=1}^{N} t_n \tag{5}$$

The generated prediction result after applying the RF algorithm is shown in fig. 9. Predicted solar PV electricity (in KW) and irradiance direct (in KW/m^2) varies directly in proportion as the predicted test values and actual test values mostly coincide at the lower direct irradiance while for other higher values it shows some error. Fig. 11 shows the plot after applying random forest algorithm. The forecasting result is shown in Fig. 12 after applying random forest method.

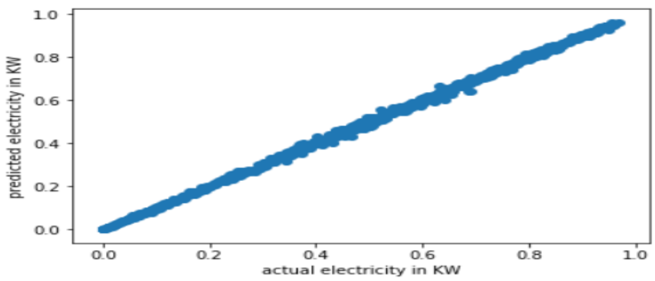


Fig. 11: Solar PV electricity prediction by applying random forest algorithm

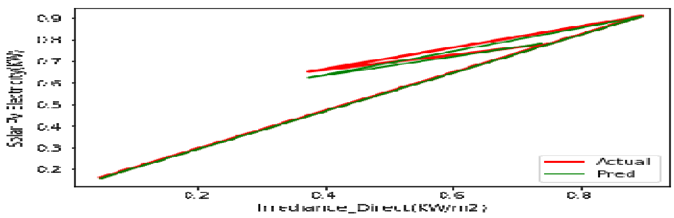


Fig. 12: Plot after applying random forest algorithm

3.5. Support vector machine

Support vector machine (SVM) is the type of regression and classification technique which separates the training dataset by creating the two dimensional or multidimensional hyper-planes. SVM utilizes the Kernel technique which converts low dimensional non-linear data sets into high dimensional vector spaces. The basics of SVM in context with regularized risk factor and error penalty factor is clearly elucidated in the research article

(Lanre Olatomiwa et al., 2015). Solar PV electricity prediction for the test data set by applying SVM algorithm is shown by fig. 10. The result is not showing good response as the actual and predicted electricity lines are far away to each other. It reflects that SVM is not a worthy performer for every data set. Fig. 13 is plotted between predicted and actual value of solar PV electricity after applying SVM method. Forecasting result by SVM is shown in Fig.14.

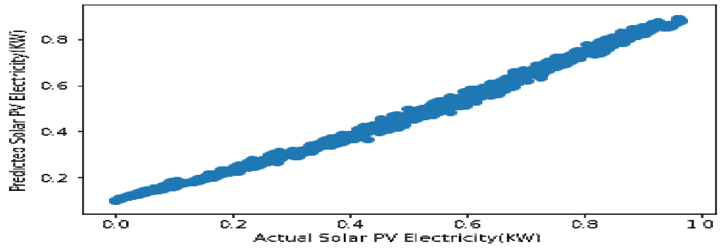


Fig. 13: Solar PV electricity predictions by applying support vector machine

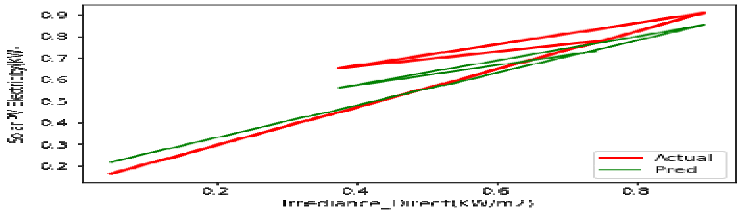


Fig.14: Solar PV electricity forecasting after applying SVM algorithm

3.6. K – nearest neighbor

KNN algorithm can solve both regression and classification type forecasting problems. Industry uses it widely for classification problems. It is generally used because of its low calculation time and easy illustration. It uses feature similarity for forecasting corresponding to any new data points. The preliminary step is to find out the distance between each training data point to the new point. There are many methods for computing the distance such as (i) Euclidian distance (ii) Manhattan distance and (iii) Hamming distance. Euclidian distance between two point's p(x_1, y1) and q(x_2, y2) is given by eq. 6.

$$\text{Euclidian distance} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (6)$$

The second step is to choose the value of K, which determine the number of neighbors in between training dataset. Forecasting results very much depend on the value of k that's why care should be taken while deciding the value of k. In this work two methods are applied to find out the value of K such as to plot elbow curve and to apply grid search in the Python programming. Elbow curve is plotted between root mean square error (RMSE) and the value of K which is shown in fig. 15.

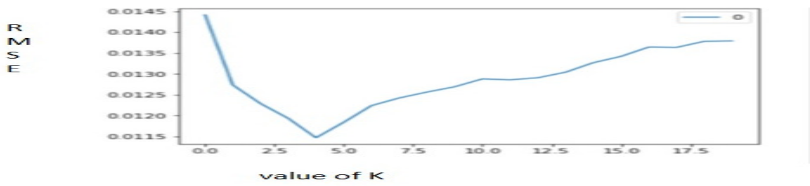


Fig.15: Selection of K-value

The model overfits the training data for a low value of K such as K=0.5 and K=1 while for model underfits at high value of K as 10 or 12. The optimum of point of RMSE reaches near the value of K at 2.8. To find the best value of K, grid search method is applied in the Python programming which results K=6. Plot between predicted and actual value of solar PV electricity by applying KNN is shown in Fig. 16 and Fig.17 shows actual and predicted value of solar PV electricity after applying KNN algorithm.

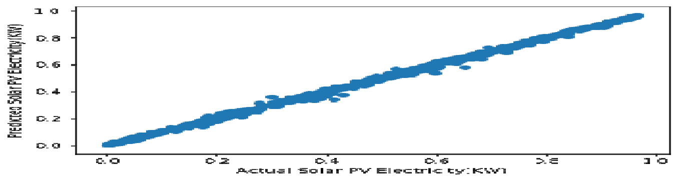


Fig. 16: Plot by applying KNN algorithm

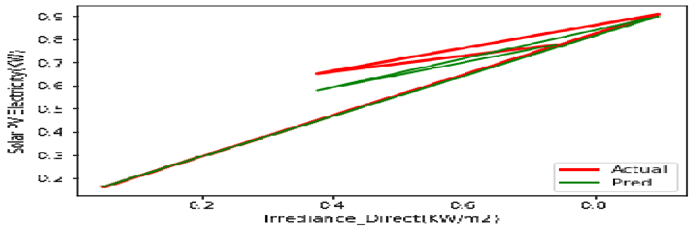


Fig. 17: Plot after applying KNN algorithm.

4. Performance evaluation factors

Performance assessments of applied algorithm are made on the basis of the factors MAPE and RMSE which is defined below by eq. 7 and eq. 8:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A-F}{A}\right| \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(A-F)^2}{n}} \quad (8)$$

Where A is the actual value, F is the forecasted value and n is the total number of samples.

5. Results

After applying all the six algorithms the final results is obtained and shown below. Generated solar PV electricity is taken on y-axis while time in minutes taken on x-axis. The result obtained after applying RF algorithm shows best among all. Fig. 18-23 shows the forecasting results of solar PV electricity (KW) with respect to time (in minutes) after applying the six algorithm: multiple linear regression, ridge regression, decision tree, random forest, SVM and KNN.

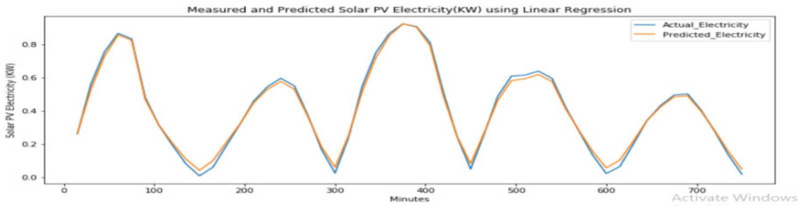


Fig. 18: Solar PV electricity versus time by applying multiple linear regression

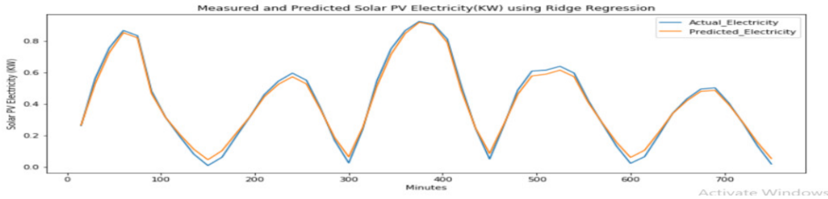


Fig. 19: Solar PV electricity versus time by applying ridge regression+

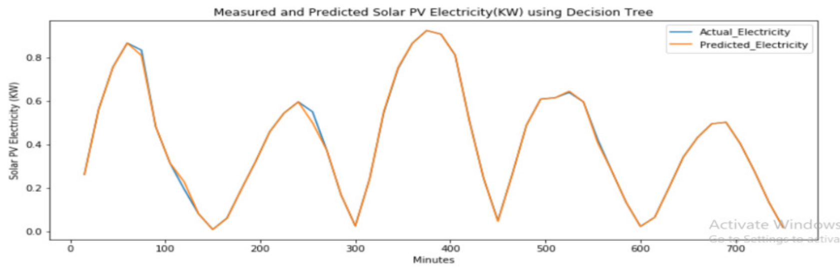
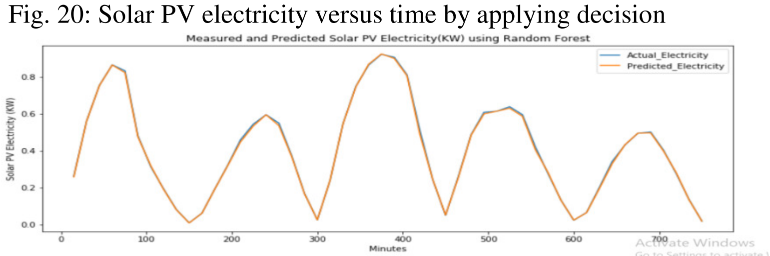


Fig. 20: Solar PV electricity versus time by applying decision



tree

Fig. 21: Solar PV electricity versus time by applying random forest

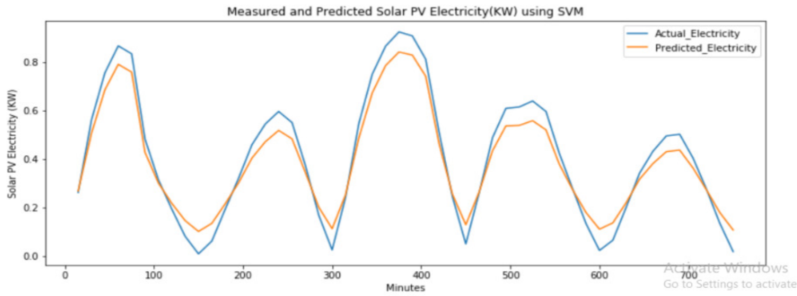


Fig. 22: Solar PV electricity versus time by applying SVM

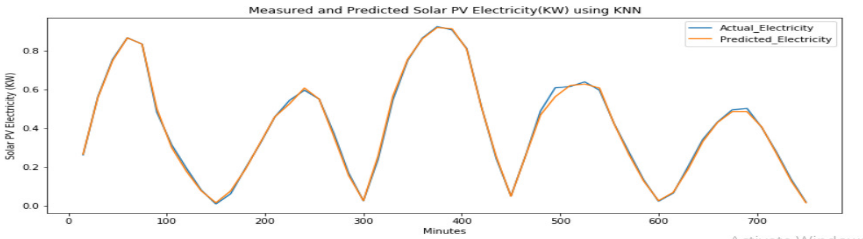


Fig. 23: Solar PV electricity versus time by applying KNN

Both MAPE and RMSE computed for each applied algorithm and results are compared on both the basis.

6. Conclusions

In this work six algorithms are applied to forecast solar PV electricity and a comparison chart is prepared on the basis of results obtained by mean absolute percentage error (MAPE) as shown by table 1. It can be seen that the performance of Random Forest method is best among all algorithm.

Table 1: Comparison Chart on the basis of MAPE

Rank	Applied Algorithm	MAPE	RMSE
1	Random Forest	0.022790	0.008792
2	Decesion Tree	0.038436	0.013651
3	KNN	0.060375	0.011843
4	Linear Regression	0.192710	0.018699
5	Ridge Regression	0.212317	0.018660
6	SVM	0.905273	0.055308

Once again comparisons of all six algorithms are made by root mean square error (RMSE) and it is shown by table 2.

Table 2: Comparison Chart on the basis of RMSE

Rank	Applied Algorithm	RMSE	MAPE
1	Random Forest	0.008792	0.022790
2	KNN	0.011843	0.060375
3	Decision Tree	0.013651	0.038436
4	Ridge Regression	0.018660	0.212317
5	Linear Regression	0.018699	0.192710
6	SVM	0.055308	0.905273

It can be seen that Random Forest method gives the best performance on both the criteria of MAPE and RMSE. Tree based algorithm provides better results as it can be seen from both the tables that the performance of Decision Tree algorithm is almost follows the Random Forest method. Results obtained by KNN technique are also satisfactory. Performance of SVM technique has not shown the good results in terms of MAPE and RMSE for forecasting. In this paper supervised learning methods are taken into the consideration which can be further compared by unsupervised and reinforcement learning methods.

References

[1] J. Antononzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez - de- Pision and,F. Antonanzas-Torres, Review of photovoltaic power forecasting, *Solar Energy 136* (2016), pp. 78-111.

[2] A. Tuhay, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Moharlen, M. Lange, M.G. Casado, J. Black, M. Marquis and C. Collier, Solar forecasting, *IEEE Power and Energy Magazine,* Nov./Dec. 2015, pp. 50-59.

[3] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, *Sustainable Cities and Society*, 48 (2019) 101533.

[4] R. Caruana, and A. Niculescu-Mizel, An empirical comparison of supervised learning algorithms*, 23rd International conference on machine learning, Pittsburgh, PA,* 2006, pp. 161-168.

[5] C. Persson, B.Peder,T. Shiga, and H. Madsen, Multisite solar power forecasting using gradient boosted regression tree, *Solar Energy 150* (2017), pp. 423-436.

[6] D. Assouline, N. Mohajeri and J. L. Scartezzini, Quantifying rooftop photovoltaic solar energy potential: A machine learning approach, *Solar Energy 141* (2017), pp. 278-296.

[7] T. C. Hugo and F. M Coimbra Carlos, Nearest-neihger methodology for prediction of intra-hour global horizontal and direct normal irradiances' Renewable Energy 80 (2015), pp. 770- 742.

[8] P. F. Jimenez-Perez, and L. Mora-Lopez, Modeling and forecasting hourly solar global radiation using clustering and classification technique, *Solar Energy 135* (2016) , pp. 682-691.

[9] A. Ahmed, T. N. Anderson and T. T. Lie, Hourly global solar irradiation forecasting for New Zealand, *Solar Energy 122* (2015), pp. 1398-1408.

[10] J. Dong, M. M. Olama, A. M. Kuruganti Melin, S. M. Djouadi, Y. Zhang and Y. Xue, Novel stochastic methods to predict short term solar radiation and photovoltaic power, *Renewable Energy 145* (2020), pp. 333-346.

[11] R. Srivastava,A. N, Tiwari and V. K. Giri, Solar radiation forecasting using MARS, CART, MS and random forest model', *Heliyaon 5* (2019) e02692.

[12] C. L. Dewangan, S. N. Singh and S. Chakrabarty, Combining forecasts of daily-ahead solar power', *Energy 202* (2020) 117743.

[13] Ahmad Syahiman Bin Mohd Shah, Hiroki Yokoyama, and Naoto Kakimoto "High-Precision Forecasting Model of Solar Irradiance Based on Grid Point Value Data Analysis for an Efficient Photovoltaic System" IEEE Trans. on Sustainable Energy, Vol. 6, No. 2, April 2015.

[14] Yong Sheng Khoo, Andre Nobre, Raghav Malhotra, Dazhi Yang, Ricardo Ruther, Thomas Reindl, and Armin G. Aberle,"Optimal orientation and tilt angle for maximizing in-plane solar radiation for PV application in Singapur", IEEE Journal of Photovoltaic, Vol. 4, No. 2, March 2014,pp.647-653.

[15] Aidan Tuohy,John Zack, Sue Allen Haupt, Justin Sharp, Mark Ahlstrom, Skip Dies, Eric Girmit, Corinna Mohrlen, Matthias Lange, Mayte Garcia Casado, John Black, Melinda Marquis and Craig Collier, "Solar Forecasting-Method Challenges and Performance",IEEE Power & Energy Magazine, Vol. 13,Issue 6Nov./Dec. 2015, pp. 50-59.

[16] Faranak Golestaneh, Pierre Pinson and H. B. Gooi, "Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation-With Application to Solar Energy", IEEE Trans. Power Syst., Vol. 31,Issue-5, pp3850-3863, year2016.

[17] Yong Sheng Khoo, Andre Nobre, Raghav Malhotra, Dazhi Yang, Ricardo Ruther, Thomas Reindl, and Armin G. Aberle,"Optimal orientation and tilt angle for maximizing in-plane solar radiation for PV application in Singapur", IEEE Journal of Photovoltaic, Vol. 4, No. 2, March 2014,pp.647-653.