Solar Power Forecasting based on fixed parameter weather type classification using SVR

#1Mr.V Badri Rama Krishnan
#2 Mr.Ch.Leelakrishna, #3 P.Buchibabu
#1, #2, #3Assistant Professor, Department of EEE, GNITS, Hyderabad

Abstract— **Energy forecasting is a technique to predict future energy needs to achieve demand and supply equilibrium. If the challenges arise from the resource, pre-anticipation would be sufficing to tackle the uncertainty. While wind energy forecasting research is considered mature, solar energy forecasting is witnessing a steadily growing attention from the research community. The aim of this project is to predict the solar power using the previous available data from the solar plant of 30kW Rating. The project utilizes an SVM regression model to predict the output with the available inputs. Inputs considered are type of Weather and AC output data available from the inverter of Solar Plant. By using the two inputs the model will predict the DC output power from the solar plant and plots a regression and relational curves for analysis of data. T-sne plots are plotted for the data to analyze the penetration of data with different parameters.**

Index Terms—**Solar Power, Forecasting, SVR, SVM, Machine learning**

I. INTRODUCTION

India has a target of 175 GW renewable energy generation by 2022, with the share of renewables to grow non-linearly, making the grid integration of these intermittent renewable resources into the country's energy mix critical and at the same time challenging. Due to the intermittent nature of wind and solar energy, forecasting and scheduling are essential for stable and efficient grid management. According to Mercom India Solar Project Tracker, India has a cumulative solar installation of 33.8 GW as of Q3 2019.

Due to dependencies on the weather, renewable energy generation, especially from wind and solar power, is variable. Weather forecasting organizations have the most important scheduling input. It is difficult for a renewable generator to apply specific schedules without an accurate weather forecast. Energy forecasts are accomplished by integrating the availability of plants with the weather forecasts for the location. The generators need to predict their demand and output for the next day based on modelling, historical data, machine learning, and weather forecasts. The scheduling body, which may be the state or regional load dispatch centers, must submit this information. The backbone of grid service, both at the intra-state and inter-state network levels, is the day-ahead scheduling process. Based on the data generated by the generators, the process aims to optimally plan and dispatch electricity for the next day. With the seasonal variations, the generation frequency of renewables changes drastically in a shorter period, which the grid operators do not have the visibility into since until recently, the wind and solar generators were exempt from any forecasting and scheduling responsibilities. The grid operators are responsible for large-scale grid integration of solar and wind generating stations while maintaining grid stability and security.

II. EXISTING WORK AND SYSTEM

In the solar case, the predictability of the movement of the cloud is very difficult. There are many forecasting companies, but they are unable to predict the period when a particular project would cross the cloud. A support vector regression was modelled to foretell solar power on a rolling basis for 24 hours ahead over an entire year, to represent the practical business of energy forecasting was shown [1]. Penalties are affected because of the cloud movement. Even though solar forecasting and scheduling are a bit easy because you can easily track the curve with sun irradiation and temperature expectation, if heavy clouds are moving, then it will be difficult for a model to identify, and this leads to penalization.

The grid operation in Australia is similar to India, which is delicate due to the vast distances covered by transmission lines, and grid congestions play a significant role in handling renewable energy production, states the report. For all major renewable energy markets, forecasting and scheduling have now become indispensable for effective renewable energy integration, and India is no exception. In view of the higher generation of renewable energy in the future, the forecast for renewable energy seems to be a price that is unavoidable in order to encourage higher renewable energy use and reduce cutbacks. A survey was attempted to provide a review and analysis of machine-learning models in renewable-energy predictions which, portrayed procedures, including data pre-processing techniques, parameter selection algorithms, and prediction performance measurements, used in machine-learning models for renewable-energy predictions [2].

Modelling of power prediction one-hour ahead using SVM and Random forest methods of ML technique shown in [3] gave nearest possible output. In parallel to the increased demand for solutions to forecast PV capacity, the means Compared to conventional time series predictive models, forecasting with the help of machine learning (ML) techniques has gained popularity in recent years. Although ML methods are nothing new, the improved computational capacity and greater quality data availability have made the techniques useful for forecasting. In context to the above mentioned reasons, corresponding project as small step towards problem solution, predicts solar power using the previously available data. The project considers certain parameters as a reference for learning and predicts the future power estimation.

III. PROPOSED SYSTEM

The project utilizes data of 30kW solar plant available in an Educational Institution as a source for collection of data. Totally it has a set of around 10 to 12 panels connected in parallel and series combination with a nominal voltage of 700Vp and a Current rating of 20A at each node. Typically, the data is forwarded from the inverter to cloud, where the data is centralized and can be downloaded. Data obtained from the centralized inverter contained around 20 features out of which only 2 quantities viz. Type of Weather and AC output Power from the inverter were provided as an input to the model. As the dataset obtained was an Excel sheet with lot of data, which was reorganised to form columns of 2 features as inputs and third column with output for verification as shown in Fig 1 and Fig 2.

Time	DC Voltage PV4(V)	DC Current4(A)	Total DC Input Power(W)	AC Output Total Power (Active)(W)	Inverter Temperature(℃)	Generation Yesterday(kWh)	System Time
2020/09/28 06:09	65.50	0.10	33	20	24.70	118.40	20-9-28 6:32:26
2020/09/28 06:14	200.10	0.10	115	20	26.00	118.40	20-9-28 6:38:3
2020/09/28 06:19	216.00	0.20	329	280	26.70	118.40	20-9-28 6:43:34
2020/09/28 06:25	440.10	0.30	746	560	27.20	118.40	20-9-28 6:49:4
2020/09/28 06:30	675.90	0.30	909	710	27.70	118.40	20-9-28 6:53:28
2020/09/28 06:35	687.50	0.30	1101	920	28.40	118.40	20-9-28 6:59:2
2020/09/28 06:40	688.10	0.50	1665	1550	29.00	118.40	20-9-28 7:4:32
2020/09/28 06:46	699.50	0.70	2297	2180	29.70	118.40	20-9-28 7:10:3
2020/09/28 06:51	691.30	0.90	2757	2540	30.30	118.40	20-9-28 7:14:27
2020/09/28 06:56	688.30	1.10	3178	3010	31.00	118.40	20-9-28 7:20:0
2020/09/28 07:01	692.70	1.20	3753	3550	31.80	118.40	20-9-28 7:25:31
2020/09/28 07:07	699.90	1.50	4500	4260	32.80	118.40	20-9-28 7:31:1
2020/09/28 07:12	696.20	1.60	4656	4530	33.40	118.40	20-9-28 7:35:26
2020/09/28 07:17	708.00	1.70	5171	4950	34.30	118.40	20-9-28 7:40:59
2020/09/28 07:22	708.50	1.90	5906	5650	35.20	118.40	20-9-28 7:46:29
2020/09/28 07:28	712.50	2.00	6284	6070	36.20	118.40	20-9-28 7:52:0
2020/09/28 07:33	712.30	2.30	7141	6920	37.10	118.40	20-9-28 7:56:24
2020/09/28 07:38	712.20	1.90	6282	5990	37.70	118.40	20-9-28 8:1:58
2020/09/28 07:43	704.20	2.80	8855	8580	38.90	118.40	20-9-28 8:7:28
2020/09/28 07:49	692.20	2.70	8205	8020	39.10	118.40	20-9-28 8:12:59
2020/09/28 07:54	699.80	3.00	9280	9040	40.20	118.40	20-9-28 8:17:23
2020/09/28 07:59	691.80	3.00	9228	8940	40.70	118.40	20-9-28 8:22:57
2020/09/28 08:04	691.90	3.10	9731	9510	41.00	118.40	20-9-28 8:28:27
2020/09/28 08:10	688.30	3.50	10899	10560	42.00	118.40	20-9-28 8:33:58
2020/09/28 08:15	687.90	3.70	11920	11550	43.00	118.40	20-9-28 8:38:22
2020/09/28 08:20	692.60	3.80	11328	11010	42.30	118.40	20-9-28 8:43:55

Sheet0

Fig 1. Sample Unorganised data from cloud

The data collected was of 3 months' data for the months of October, November and December 2020. October was considered as autumn and November and December as Winter. Since the weather type data was not available, random numbers were assigned for the weather data as constants (Autumn = 2 and Winter = 1) for the dataset as one of the input. There were 9000 samples generated with 6 minutes' gap, with a day's data starting from morning 6:30 am to evening 6:30 pm. Data scaling was in the range of -2 to +2, as the AC and DC outputs vary from 0 to 25000 W. Due to large difference in the values, prediction of data can have some scaling error, which can be reduced by selection of weather data in the same range as the other input variable. System time can also be considered by utilizing time forecasting models of SVM.

Type of season	AC Output Total Power (Active)(W)	Total DC Input Power(W)
2	20	33
2	20	115
2	280	329
2	560	746
2	710	909
2	920	1101
2	1550	1665
2	2180	2297
2	2540	2757
2	3010	3178
2	3550	3753
2	4260	4500
2	4530	4656
2	4950	5171
2	5650	5906
2	6070	6284
2	6920	7141
2	5990	6282
2	8580	8855
2	8020	8205
2	9040	9280
2	8940	9228
2	9510	9731
2	10560	10899
2	11550	11920
2	11010	11328
2	12040	12435
2	12490	12925
2	12000	12389

consolidated112

Fig 2. Organised Data for the model requirement

A. *SCALING DATA and NORMALIZATION*

The data obtained for inputs was AC power output in the range of 0 to 25000W and DC output power in the similar range. The weather data was introduced as an extra column, which is an assumed quantity with 2 variations of weather as Autumn and Winter having values 2 and 1 respectively. By observation it was found that the weather level and AC output power was having lot of value difference. The huge gap of values between two features can create a marginal error in the output, so all the values were scaled to a single reference as shown in fig 3 and fig 4. Totally 9000 samples of each with 6 minutes delay from each column were considered for prediction with one month of Autumn and two months of Winter.

X - NumPy array

	0	1
0	1.2874	-1.29963
1	1.2874	-1.29963
2	1.2874	-1.26479
3	1.2874	-1.22726
4	1.2874	-1.20716
5	1.2874	-1.17902
6	1.2874	-1.09459
7	1.2874	-1.01016
8	1.2874	-0.961917
9	1.2874	-0.898931

Format Resize Background color Save and Close Close

Fig 3. Scaled Input data

Fig 4. Scaled Output data

B. *MODELLING and DEVELOPMENT*

The model was developed using Support Vector Machine Regression algorithm in python environment. Support Vector Machine, abbreviated as SVM can be used for both regression and classification tasks. But, it is widely used in classification objectives. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N - the number of features) that distinctly classifies the data points as shown in fig 5.

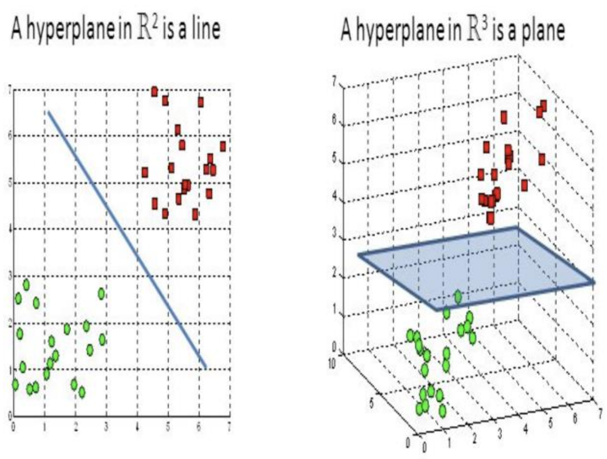

Fig 5. Hyperplane in 2D and 3D(Courtesy : towardsdatascience.com)

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

Support Vector Regression (SVR) uses the same principle as SVM, but for regression problems. The problem of regression is to find a function that approximates mapping from an input domain to real numbers on the basis of a training sample.

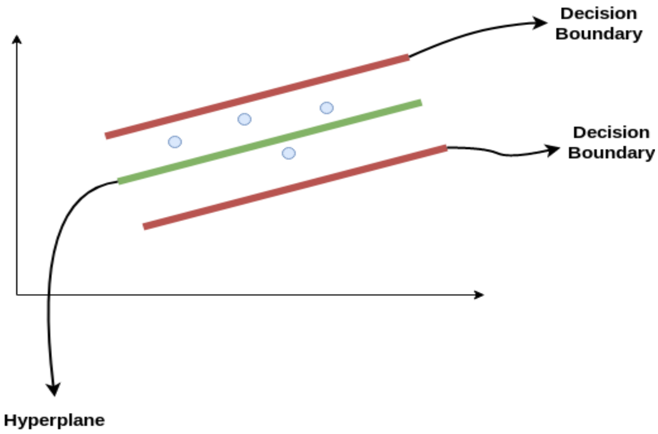

Fig 6. SVR hyperplane

Consider the above fig 6 two red lines as the decision boundary and the green line as the hyperplane. The objective, when moving on with SVR, is to basically consider the points that are within the decision boundary line and best fit line is the hyperplane that has a maximum number of points. The distance between green and red line is called epsilon. Assuming that the equation of the hyperplane is as follows:

Y = wx+b (equation of hyperplane)
Then the equations of decision boundary become:
wx+b= +a
wx+b= -a
Thus, any hyperplane that satisfies SVR should satisfy:
-a < Y- wx+b < +a

The data model considered here was from sklearn library with various other libraries as default. An organised dataset with inputs was given to the SVR model, which evaluated to the nearest results out of training and manual verification. SVM algorithms uses a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. The project used both linear and Radial basis Function Kernel to evaluate the performance of the model. A snippet of code is provided in fig 7 as a sample code to analyse the model parameters.

```
# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc_X = StandardScaler()
sc_y = StandardScaler()
X = sc_X.fit_transform(X)
y = sc_y.fit_transform(y)
# Training the SVR model on the whole dataset
from sklearn.svm import SVR
regressor = SVR(kernel = 'linear')
regressor.fit(X, y)
```

Fig 7. Code snippet from Spyder

IV. RESULTS

Various plots were obtained from the model to analyse the behaviour of the model. Plot between Weather Type and the Output DC power was observed to be liner as shown in fig 8 and the plot between AC output power and DC output power as shown in fig 9 was also observed to be linear. Hence the result showcased a small error with accuracy of 0.98 floating values were obtained. T-sne plot was also plotted from various dataset points existing, by varying the perplexity of 30 and 50. Result showed that resolution could be more to make data more readable, Hence other plots with variation in the iterations from 1000 to 5000 was obtained as shown from fig 10 to fig 13. Further analysis showed that predictions for data in the middle range values (10000 W to 25000 W) gave more accuracy than the lower range,one sample prediction is shown in fig 14.

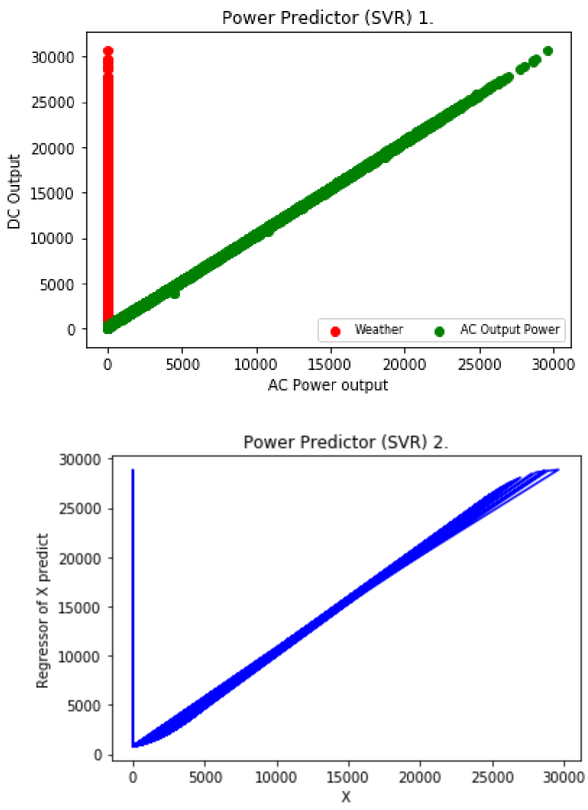

Fig 8 Radial Basis Function as Kernel with Output and input variables plotted separately

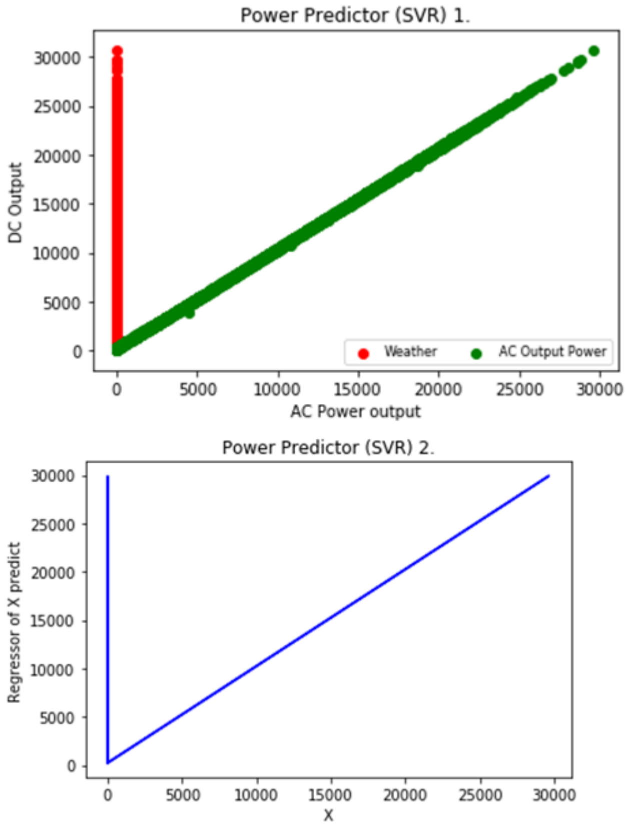

Fig 9. Linear Kernel with Output and input variables plotted separately

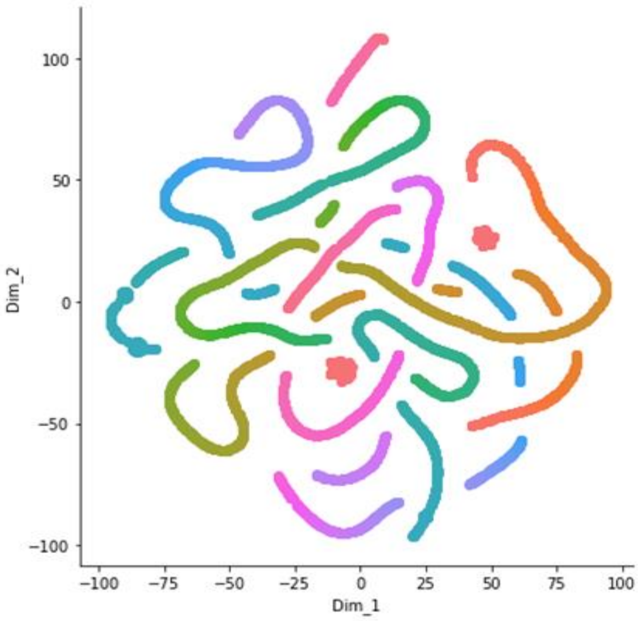

Fig 10. T-sne plot of input variables and output

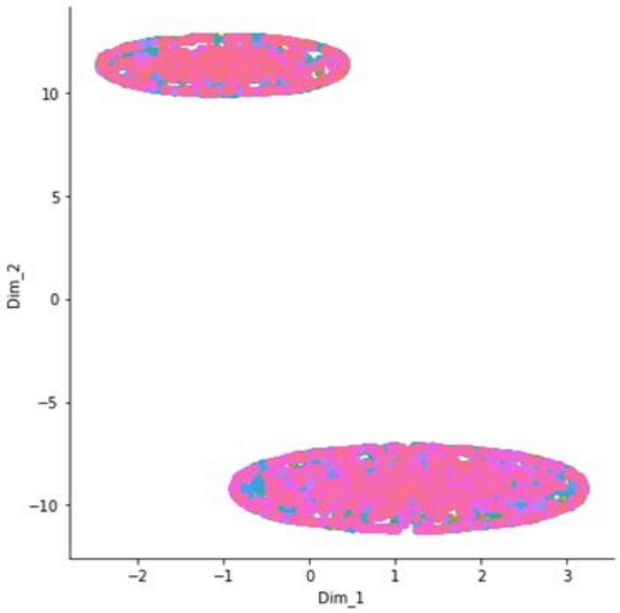

Fig 11. Weather Vs DC Output

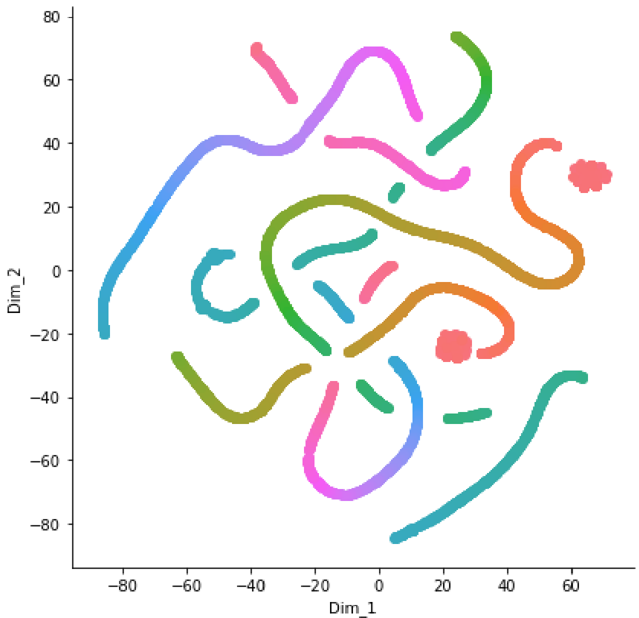

Fig 12. T-sne plot with perplexity = 50 and iteration of 1000

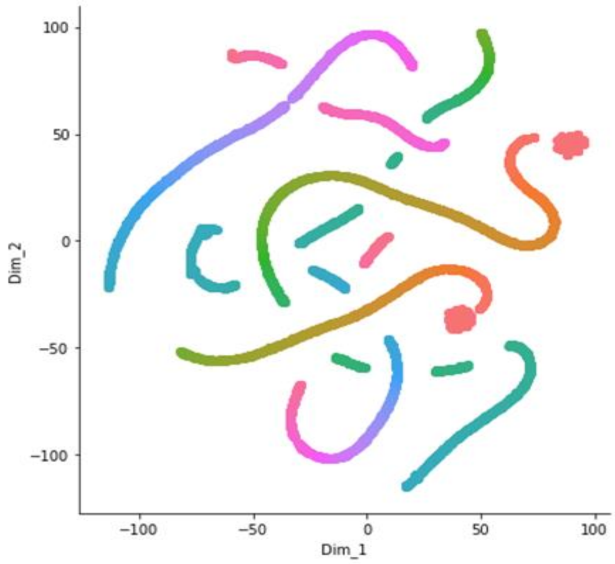

Fig 13. T-sne plot with perplexity = 50 and iteration of 5000

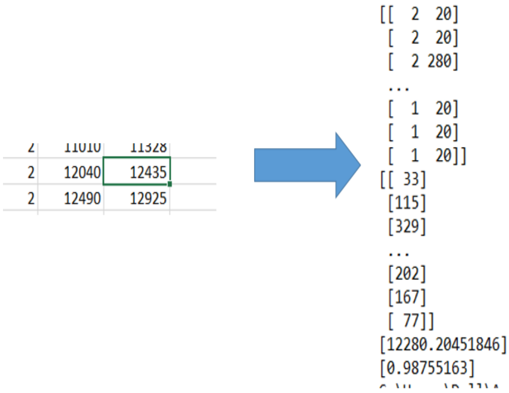

Fig 14. Sample Prediction of output with verification

V. CONCLUSION

It can be concluded that the project was a small step towards addressing major challenges. The plots gave a vague idea about the behavior of solar plant and its ancillary systems. Only two features were considered for the prediction of output values, as there was very less scope for selection of other features without effective parameter valuation. Though the values obtained were accurate up to 98%, but in practical scenario this may vary with all other parameters included in the model.

REFERENCES

[1] Mohamed Abuella, Badrul Chowdhury," Solar Power Forecasting Using Support Vector Regression", Proceedings of the American Society for Engineering Management International Annual Conference, 2016.
[2] Jung-Pin Lai, Yu-Ming Chang, Chieh-Huang Chen and Ping-Feng Pai," A Survey of Machine Learning Models in

Renewable Energy Predictions ",doi:10.3390/app10175975, Appl. Sci. 2020, 10.

[3] Stefan Preda, Simona-Vasilica Oprea,Adela Bâra and Anda Belciu (Velicanu)," PV Forecasting Using Support Vector Machine Learning in a Big Data Analytics Context ", doi:10.3390/sym10120748, Symmetry 2018, 10.

[4] Alireza Zendehboudi a, M.A. Baseer b, R. Saidur, " Application of support vector machine models for forecasting solar and wind energy resources: A review", https://doi.org/10.1016/j.jclepro.2018.07.164, Journal of Clean Production- 2018.

[5] Chih-Feng Yen, He-Yen Hsieh, Kuan-Wu Su, Min-Chieh Yu and Jenq-Shiou Leu, " Solar Power Prediction via Support Vector Machine and Random Forest", https://doi.org/10.1051/e3sconf/20186901004 , E3S Web Conferences - 2018.