Comparative Study of Supervised Learning Algorithms for PV Power Prediction

Rachana A Bujurke
Dept. of Electrical and Electronics
Dayananda Sagar College of Engineering
Bangalore, India
bujurkerachana@gmail.com

Satish B A
Dept. of Electrical and Electronics
(Asst. Professor)
Dayananda Sagar College of Engineering
Bangalore, India
satish.eee@dayanandasagar.edu

Divya D A
Dept. of Electrical and Electronics
Dayananda Sagar College of Engineering
Bangalore, India
divyadhawle@gmail.com

Adarsh D Shetty
Dept. of Electrical and Electronics
Dayananda Sagar College of Engineering
Bangalore, India
adarshshetty2805@gmail.com

Aayush Basavesh
Dept. of Electrical and Electronics
Dayananda Sagar College of Engineering
Bangalore, India
aayushbasavesh@gmail.com

***Abstract*—This paper aims at predicting Solar PV output power and managing DC load usage. With the data available from the source, output power of the PV system is determined. Variations in power patterns are studied with respect to time, temperature, light intensity. Machine learning techniques with supervised learning processes are used to make predictions. Successful implementation of Linear Regression(LR) and K-Nearest Neighbors(KNN) prove their applicability in maintaining system stability and efficient renewable power consumption.**

***Keywords* — PV system, machine learning, linear regression, K Nearest neighbors, renewable power**

I. INTRODUCTION

Solar power is a clean and inexpensive energy source which can be harnessed easily from anywhere in the world due to its abundance. Over the past few years solar systems are being extensively used to power small loads typically installed on roof tops of buildings. On site solar power for businesses and non-profits with small scale utilities have a large number of customers in the power market. Residential owners can install solar systems and design their premises to take full advantage of solar technology. Subsidies provided by the government for the installation of solar panels also add to their favour. Several factors have to be considered before employing a PV system like location, weather conditions, capital cost and maintenance. However these measures do not always yield state of the art performance in recent times. In the present scenario solar energy is directly fed to the grid without checking the status of the grid. Power generated from the distributed solar panels doesn't exceed the limit of penetration of the grid. However, in the future almost every house will be using DC loads. But the grid is not designed such as to integrate large amounts of varying solar power into it. So in order to face this issue and to control the excess power, utilities should have an estimate of how much amount of solar power will be added to the grid after domestic usage. This will assist the utility to make use of these resources to its maximum and maintain the stability of the grid. Predicting the power output of a solar panel also helps in optimal load scheduling by individuals, bidding in electricity markets etc. But the power output from a solar panel can change rapidly due to unexpected weather conditions. This makes forecasting the power output of a solar panel even more challenging.

II. CURRENT TECHNIQUES USED FOR PREDICTION

A. *Nowcasting:*

This is a method which concentrates on predicting the power in the next few minutes up to four to six hours ahead[6]. Nowcasting is based on time series processing of data[7][10].

B. *Short-term solar power forecasting:*

This type of forecasting predicts power up to seven days ahead. This type of prediction helps in making decisions regarding grid operations. The models used for short-term prediction are Global Forecast System(GFS) or the information provided by the European Centre for Medium Range Weather Forecasting(ECMWF). These two models are the state of the art of short-term prediction method[8].

C. *Long-term solar power forecasting:*

This type of forecasting refers to prediction of power monthly or annually. These models are run with mesoscale models which are fed with reanalysis data[9]. The output of these models are processed using statistical approaches.

III. METHODS OVERVIEW

There are two main ways of forecasting solar power. They are statistical approaches and physical approaches. Physical approach is done by using weather data and statistical approach is done by using historical time series of data. A well-known physical approach is calculating the solar irradiance falling on earth[3]. Short-term power forecasting models were used to predict hourly values of power generated from the solar panel. This model will adapt to the variations in climatic conditions since it is trained in real time. This will help the model to adapt with the possible weather changes that can happen in the next hour[2].

Statistical methods based on multi-regression analysis are developed in order to predict power of a PV plant[1]. This model is trained using three different input vectors. This study demonstrates data driven management of PV system to improve performance and grid stability.

IV. Data Collection from Source

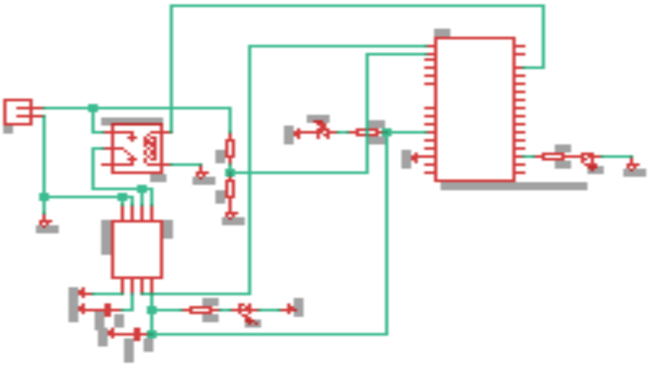

Fig. 1 Source side data collector circuit

A data logger is designed for the PV system using Arduino and Raspberry Pi. Voltage and current are measured with the use of resistors and current sensors. The parameters (date, time, voltage, current.) are recorded in an SD-card. The data is collected every 3 minutes and is saved as a .CSV file. To edit the .CSV file spreadsheet software MS-EXCEL is employed.

V. Preparing Data Set

The data acquired from source is raw and needs to be cleaned and validated before using it for analysis. External parameters like Sun angle, temperature and light intensity are added. The power for each data set. Missing or invalid data can give less accuracy of prediction.

```
datetime,Voltage,Current,Temperature,Degree,Power,Lintensity
25-10-2019 17:22:57,18.68,0.22,22,256.79,4.1096,115.848571428571
25-10-2019 17:23:15,18.68,0.22,22,256.79,4.1096,115.848571428571
25-10-2019 17:23:33,18.7,0.23,22,256.79,4.301,116.021428571429
25-10-2019 17:23:51,18.67,0.22,22,256.79,4.1074,115.787142857143
25-10-2019 17:35:48,18.21,0.16,22,257.49,2.9136,112.661428571429
25-10-2019 17:36:06,18.16,0.16,22,257.49,2.9056,112.354285714286
25-10-2019 17:36:24,18.09,0.14,22,257.49,2.5326,111.824285714286
25-10-2019 17:36:42,18.03,0.14,22,257.49,2.5242,111.455714285714
25-10-2019 17:37:00,17.97,0.13,22,257.49,2.3361,111.037142857143
25-10-2019 17:37:18,17.91,0.13,22,257.49,2.3283,110.668571428571
25-10-2019 17:37:36,17.88,0.13,22,257.49,2.3244,110.484285714286
25-10-2019 17:37:54,17.81,0.12,22,257.49,2.1372,110.004285714286
25-10-2019 17:38:12,17.76,0.13,22,257.49,2.3088,109.747142857143
25-10-2019 17:38:30,17.72,0.11,22,257.66,1.9492,109.401428571429
25-10-2019 17:38:48,17.66,0.11,22,257.66,1.9426,109.032857142857
25-10-2019 17:39:06,17.62,0.1,22,257.66,1.762,108.737142857143
```

Fig. 2 Represents the attributes of the used data set

VI. Machine Learning Techniques

This subsection briefly explains the machine learning methods tested in this study. The type of data processing used is supervised learning which is used to develop predictive models using both input and output data. It is called supervised learning because the process of learning from the dataset is similar to a teacher supervising the learning process. The goal is to develop a mapping function from (X) to (Y) so that when a new input data is given an output can be predicted. Learning continues until the algorithm reaches an acceptable level of performance.

A. Linear Regression

The In this method we can predict the target using one variable. Regression is a parametric approach as it assumes a linear functional form for f(X). It is used when one needs to estimate a small number of coefficients. This method is easy to fit and easy to interpret.

The general equation of Regression line is

$$y = mx+c \qquad \text{where,}$$

y = Dependent Variable
x= Independent Variable
c = y-Intercept

The accuracy and goodness of fit is measured by loss, R squared value. Least square method is used to determine the regression line. Squaring the distance between the data points and minimizing the sum of squares gives the best-fit regression line.

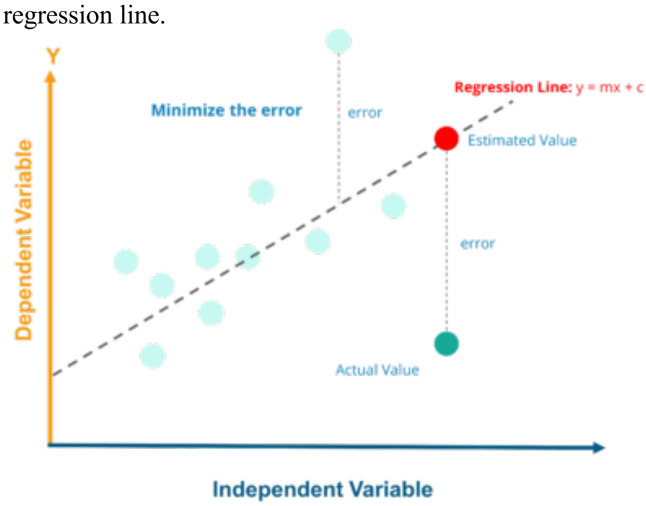

Fig. 3 Nature of data points and regression line

R–squared value is the statistical measure to know how close the data points are located around the obtained regression line. R–squared value should be low for a good model.

$$R^2 = 1 - \frac{\sum (y_p - \bar{y})^2}{\sum (y - \bar{y})^2}$$

Linear regression is implemented using a machine learning library called scikit-learn.

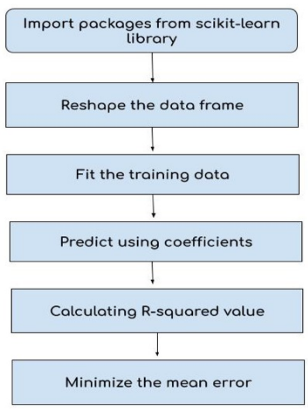

Fig. 4 Steps to Implement Linear regression

Test error is also evaluated using a predefined formula. Testing accuracy is measured using the mean error. Training error and training accuracy are also evaluated.

B. K-Nearest Neighbours

KNN is a non- parametric method which does not assume an explicit form of f(X). This is more complex than parametric methods. The algorithm involves following steps:

1) It Assume a value for the number of nearest neighbors K and a prediction point xo.

2) KNN identifies the training observations No closest to the prediction point xo.

3) KNN estimates f(xo) using the average of all the responses in No, i.e.

$$\hat{f}(x_o) = \frac{1}{K}\sum_{x_i \in N_o} y_i.$$

A small value for K provides the most flexible fit, which is not biased but has large variance. Hence larger values of K provide good fit; the prediction in a region is an average of several points, and so changing one observation has a smaller effect.

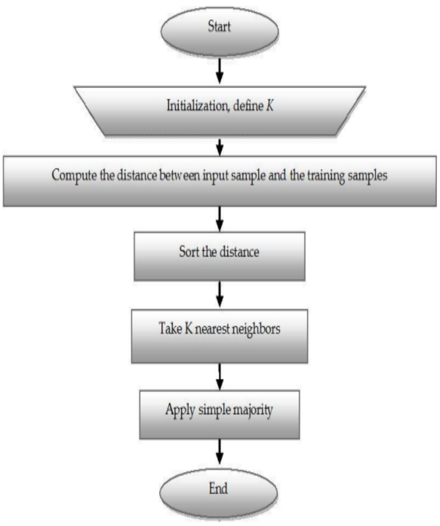

Fig. 5 Workflow of KNN algorithm

To perform Knn we use the Knn() function which predicts using a single command with four inputs. A matrix containing training data; a matrix containing the data for which predictions are to be made; a vector containing the training observations and a value of K the number of nearest neighbours to be used.

We split the data into train and test sets using the scikit-learn library. The below script splits the dataset into 80% train data and 20% test data. We first import the KNeighborsClassifier class from the sklearn.neighbors library. The parameter n_neigbours is used to initialise the value of K. To start out 5 seems to be the most commonly used value for the KNN algorithm.

One way to find the value of K is to find the mean error of predicted values of the test sets for all K values between 1 and 100 and plotting the error values against the K values. The script below explains this process.

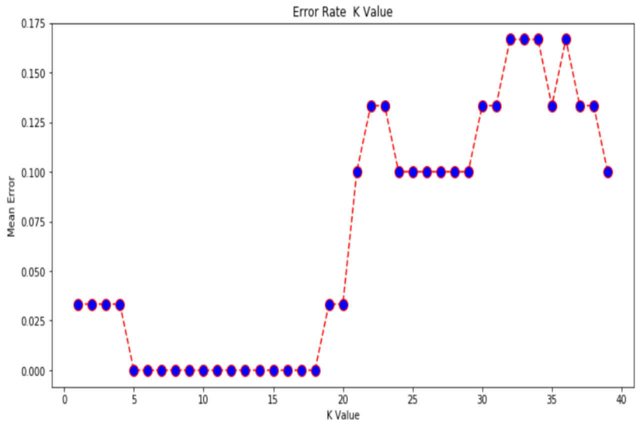

Fig. 6 Plot used to decide the best value of K

The plot indicates a range of K values for which the error is minimum. One can test these values and check the accuracy by applying it to the classifier.

VII. Results and Discussion

The above methods were tested for the obtained source data in Table 1. The results are mentioned below in Table 2 and compared using error and accuracy rates.

Metric	LR	K-NN
Training error	13.67%	1.370%
Training accuracy	86.33%	98.62%
Test error	7.88%	1.60%
Test accuracy	92.11%	98.39%

Table 1 Comparison of results

The results indicate that K-NN provides better accuracy than LR method. LR assumes a linear relationship between the variables. The effect of assumption is that the output gets biased. Whereas K-NN can produce non-linear solutions.

K-NN stores the entire training dataset and does not learn any model. K-NN makes real-time predictions by calculating the similarity between an input sample and each training instance.

VIII. Conclusion

The periodic characteristics of PV power is studied to analyse the correlation between power generation and weather factors. A forecasting model is proposed by analysing the historic data. Machine learning techniques are demonstrated to understand the nature of prediction. Power predicted using light intensity produced good efficiency with minimum modelling error. The purpose of PV power prediction is emphasized for safe and economical operation of power system.

REFERENCES

[1] Md Rahat Hossain, Amanullah Maung Than Oo, A. B. M. Shawkat Ali ‘Hybrid Prediction Method for Solar Power Using Different Computational IntelligenceAlgorithms’Article ID:28135 Smart Grid and Renewable Energy

[2] Z Wang, L He, X Cheng, J He ‘Method for short-term photovoltaic generation power prediction based on weather patterns IEEE China International Conference on Electricity Distribution (CICED) (2014), pp. 213-215

[3] K Malik, BA Bhatti, F Kamran ‘An approach to predict output of PV panels using weather corrected global irradiance’ IEEE International Conference on Intelligent Systems Engineering (ICISE) (2016), pp. 111-117

[4] HAARA Arto; KANGAS Annika. ‘Comparing K Nearest Neighbours Methods and Linear Regression’. Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS),feb 2012. ISSN 1946-7664.

[5] Sankara Subbu, Ramesh, "Brief Study of Classification Algorithms in Machine Learning" (2017). CUNY Academic Works

[6]Vorrath, Sophie (2019). "New APVI solar tool shows daily, time-based forecast for each state”.

[7]Espino, I., Hernández, M.. Nowcasting of wind speed using support vector regression. Experiments with Time Series from Gran Canaria. Renewable Energy and Power Quality Journal, N9, 12 May 2011.

[8]Bacher, P., Madsen, H., Nielsen H.A. Online short-term solar power forecasting. Solar Energy. Vol 83, Issue 10, October 2009: 1772-1783.

[9]Diagne, H.M., David, M., Lauret, P., Boland, J. Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids. In Proceedings of the World Renewable Energy Forum 2012 (WREF 2012), Denver, USA, May 2012.

[10]Luis Martín, Luis F. Zarzalejo, Jesús Polo, Ana Navarro, Ruth Marchante, Marco Cony, Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, Solar Energy, Volume 84, Issue 10, October 2010, Pages 1772-1781