International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Support Vector Machines Analysis of Photovoltaic Power Dataset Utilizing Internet of Things

Khaydarov Sherzod, Zhang Jinjiang, Raza Ahmad

School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, 310023, China

Abstract: ***The recent years have been evidencing a huge leap forward in the application of internet of things in assessment of PV systems. It altered the way we store the real - time monitoring data from the sensors on the Internet Cloud by utilizing sensor friendly Raspberry Pi computers. The detection of the unforeseeable conditions of the PV plants that effect on the reliability of the power generation being one of the contemporary problems of our times has found the solution by monitoring the weather condition in real time applying Internet of things. Relying on the weather data acquired from sensors connected to Raspberry Pi statistical model that predict the solar power generation is built. This paper proposes the 2kW photovoltaic station power performance and implements predictions by means of support vector machines (SVM) and analyses the results derived from applying different kernels.In order to assess and forecast the dataset, the models were built in MATLAB software. The most actual result was achieved in Medium Gaussian SVM model.***

Keywords: support vector machines, PV forecasting, Thingspeak

1. Introduction

The alarming report from the Scripps Institution of Oceanography released in June 4, 2019 surpassed a new milestone of 415 ppm of global CO2 levels in the global atmosphere compared with the level of before the Industrial Revolution which remained stable around 280 ppm for hundreds of thousand years indicates how serious the situation of climate change on Earth [1]. This is caused undoubtedly by human activities, mainly by fossil fuel burning in an overwhelming amount [2]. Moreover, the disastrous catastrophe in March 2011 in Fukushima nuclear power plant once more showed us the consequences of nuclear pollution [3]. Due to this fact, only renewable energy sources (RES) can provide us with the energy we need and leave out the emissions we do not.

Among renewables, solar energy, also called photovoltaic energy is abundant and offers significant potential for near-term and long-term environmental pollution mitigation [4]. In the light of the photovoltaic (PV) development, the amount of construction of solar plants is increasing where predictive analytics will play a significant role towards real-time optimal management of energy use that secure the operation of power systems and maintain balance between consumption and production demand [5]. For this reason the field of forecasting by means of machine learning has been in a high priority among scientists. Since the forecasting is dependent on historical observations in time period, time series forecasting technique is applied to extrapolate the future attitude of the predicted variable [6].

Research on forecasting techniques of solar thermal and renewable energy generation stations has been conducted in past few decades in large scale.

1) Groups of multi-locus approaches have been developed for the PV power generation forecasting using ML and DL tools. Among different ML approaches, Wang et alproposed a method to assess the daily energy generation of PV plants based on Principal Component Analysis (PCA) as a data preprocessing method and Support Vector Machines to develop a classifier. The proposed method could detect deviations between estimated and the measured daily energy generation of 10% [7].
2) Ahmad *et al* put forward the ensemble methods (EM) that reduces variance and bias, such as random forest (RF) and extra trees (ET) for predicting hourly PV generation output. Their results indicated that ET outperformed RF and SVR in terms of computational cost [8].
3) Meteorological conditions, especially solar irradiation [9] plays a pivotal role in determining PV power [10] that requires different tools to acquire the weather database. Wang et al. predicted PV power from satellite based weather data utilizing SVM based on adapting data fusion (SVM-DF). Results showed that the SVM-DF model performed better than ANN model [11]

Mu et al. carried out PV power generation online observation with Artificial Neural Network (ANN) to stabilize load frequency in the micro-grids and indicated its effectiveness over Proportional Integral (PI) and fuzzy controllers [12]

2. PV Forecasting Data Features

The selection of input variables and prediction horizon affects the accuracy of the developed prediction model. In general, some significant variables are used as input of forecasting models but are not bound to the following factors [13]:

- Historical data of photovoltaic generation;
- Historical explanatory variables, which are relevant to the meteorological variables, consisting of global horizontal irradiance, temperature, cloud cover, humidity, wind speed, and so on.

2.1 Data collection

The methodology is developed and tested using historical measurements from 2 kW roof-mounted PV plant of the Zhejiang University of Science and Technology located in Hangzhou, Zhejiang province of China (longitude 120°02'54''E, latitude 30°22'28''N, altitude 32 m).

After getting raw data, outliers were detected through data preprocessing method. The database from June 17, 2019 was collected from Raspberry Pi embedded homemade prototype created using several sensors, while the output power was measured in a counter by Growatt company.

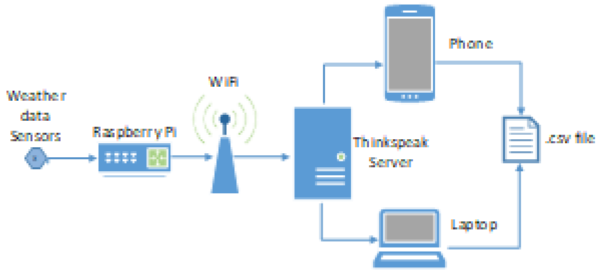

Figure 1: Flowchart of the creation of the database by Raspberry Pi embedded prototype

In this study the constructed SVM model predicts power generation values using dependent variables announced for the future period.

Table 1: Dependent and Independent Variables of the database

	Variable name	Description	Device or sensor type
Dependent variable	PV power	Solar power generation (kWh)	Growatt
	lux	Luminosity (lux)	TSL2561
Independent variable	temp	Temperature (°C)	DHT-11(R)
	hum	Humidity (%)	DHT-11
	press	Atmospheric pressure (hPa)	BME280

2.2. Internet of things analytics using Thingspeak platform

Thingspeak is an open sourceInternet of Thisgs(IoT) application to store and retrieve data from things, using REST and MQTT APIs. ThingSpeak enables the creation of sensor logging applications, location tracking applications, and a social network of things with status updates. In this project the temperature, humidity, air pressure and luminosity were uploaded and the data was created as comma separated value (.csv) file. It is also possible to observe the real-time data online through thelinkhttps://thingspeak.com/channels/950669. The resultant graph for the weather parameters was illustrated in Figure 2. While the power output in Figure 3 was measured through the Growatt technology. The power outcome shows the peak power was reached in the midday and the day was sunny. [14]

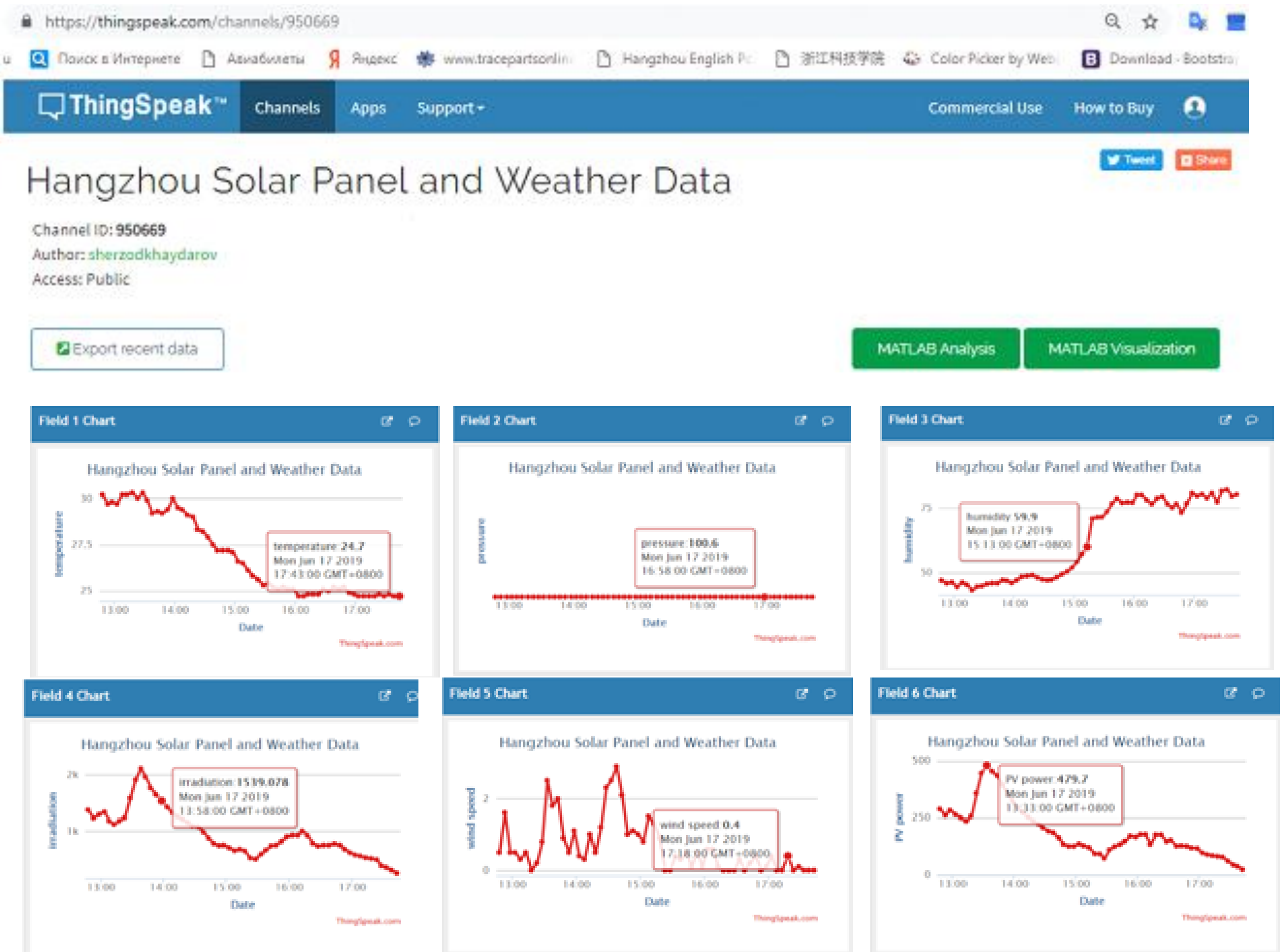

Figure 2: Weather parameters for the database in Hangzhou, China. 17/06/2019

Paper ID: SR20612151700 DOI: 10.21275/SR20612151700

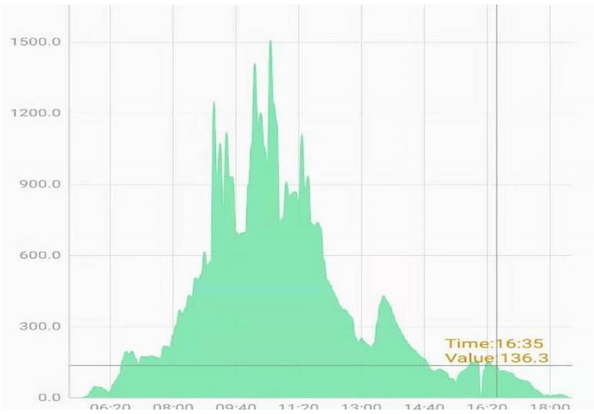

Figure 3: PV power diagram

Figure 4: Solar plant location

2.3. MATLAB

MATLAB (*matrix laboratory*) is a multi-paradigmnumerical computing environment and proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages.

Current paper utilizes MATLAB to compare predicted values and true values according to the different SVM kernel functions [15]

3. Model definition and kernel functions

3.1 Support Vector Machines (SVM)

A Support Vector Machines (SVM) is a kernel based supervised learning tool used in classification, prediction, pattern recognition and regression problems. It was initiated by V. Vapnik in 1986 [16].

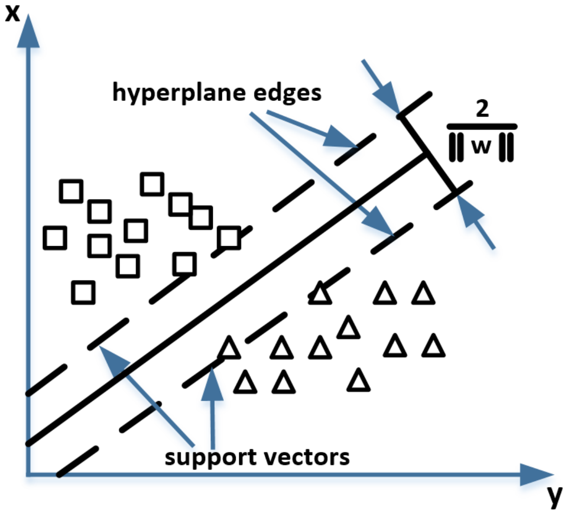

Figure 5: SVM classification, support vectors and hyper-plane edges

The kernel functions namely Gaussian radial basis function (RBF), polynomial functions, and hyperbolic tangent are used to transform the nonlinear classes separated by lines to linearly classified space. The idea that lies behind this method is mapping the input vectors x into a high dimensional space Z constructing the separating hyperplane (Figure 5).

When using the SVM for regression, the target is to minimize the distance between the hyperplane edges and this method is called Support Vector Regression (SVR) [17]

The forecast for this method calculated for the input test $x_*$is provided as:

$$\hat{y} = \sum_{i=1}^{n} \alpha_i k_{rbf}(x_i, x_*) + b \qquad (1)$$

Here as we can see SVR uses RBF kernel defined as:

$$k_{rbf}(x_p, x_q) = \exp\left[\frac{-(x_p - x_q)^2}{2\sigma^2}\right] \qquad (2)$$

In these equations, the inputs x_p and x_q are the pth and qth dimensions and the σ is the kernel width. The coefficients α_i are the difference of two Lagrange multipliers, which are the solutions of a quadratic programming (QR) problem, while b is the bias parameter.

Table 1: SVM model kernel functions classification

Regression Model Type	Interpre-tability	Model flexibility
Linear SVM	Easy	Low
Quadratic SVM	Hard	Medium
Cubic SVM	Hard	Medium
Fine Gaussian SVM	Hard	High Allows rapid variations in the response function. Kernel scale is set to sqrt(P)/4, where P is the number of predictors
Medium Gaussian SVM	Hard	Medium Gives a less flexible response function. Kernel scale is set to sqrt(P)
Coarse Gaussian SVM	Hard	Low Gives a rigid response function. Kernel scale is set to sqrt(P)*4.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

The selection of suitable kernel function for the dataset is an essential task to be performed. Linear SVM functions decode the values easily, but less accurately compared to Gaussian or Radial Basis Function (RBF) kernel functions (Table 2)

4. Results and Discussions

4.1 Error Metrics Formulation

Forecasting improvement, especially time series forecasting accuracy is an important yet often difficult task facing decision makers in many areas[18].Despite the numerous machine learningmodels available, the research for improving the effectiveness of forecasting models has never stopped. Several large-scale forecasting competitions with a large number of commonly used machine learning [19,20] forecasting models conclude that combining forecasts from more than one model often leads to improved performance, especially when the models in the ensemble are quite different,root mean square error (RMSE)[21, 22]is calculated through the equations:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}_j - \mathbf{t}_j)^2} \quad (3)$$

RMSE shows the difference between the forecasted and real values and gives an idea about the accuracy of the model according to the lower the better principle.

R^2 measures the whether the forecasted values are fitting the real values using the predictive model. The measurement range is [0,1], where R^2 tends to 1 indicates the ideal fit of the model and the accuracy gets lower as the value decreases.

$$R^2 = 1 - \frac{var\,(y - \hat{y})}{var\,(\hat{y})} \quad (4)$$

MAPE is equal to the ratio of the error to true value in percentage.

$$MAPE = \frac{1}{N} \times \sum_{i-1}^{N} \left|\frac{(\hat{y}(i) - y(i))}{y(i)}\right| \times 100\% \quad (5)$$

MAE proposes the evaluation of uniform prediction errors by measuring the average distance between the measured values and predictive model.

$$MAE = \frac{1}{N} \times \sum_{i-1}^{N} |\hat{y}(i) - y(i)| \quad (6)$$

where N denotes the number of samples, $\hat{y}$ (or predicted time series) with observed data y (or measured time series).

4.2. Support Vector Machines Models Results

Six SVM models with six different kernels were evaluated in this work for the PV power generation prediction (Figure 6). All the models were validated against the experimental data of irradiation and PV power as a target data and the comparisons based on statistical measures such as RMSE, MSE, MAE, and correlation coefficient (R^2). For higher modeling accuracy RMSE, MSE and MAE indices should be closer to zero but R^2 value should be closer to 1. In addition, the training time and prediction speed value were estimated to determine whether the model is feasible to apply in the real world applications.

We utilized the Statistics and Machine Learning toolbox of Matlab2019a to train and predict SVM regression model. It implemented linear epsilon-insensitive SVM (ε-SVM) regression to ensure that each training point X was as flat as possible. The main goal of ε-SVM regression is to find a function g(x) that diverges from $y_n$with the less value than ε.

The results in Fig. 7 and Fig.8 show that Medium Gaussian SVM predicted more accurately (0.87) in comparison with other five models in all measurement indices and its training time was also in the medium level. On the contrary, the slowest kernel function was found to be linear function with spending twice more time (14.298 sec.) than the remaining model functions.

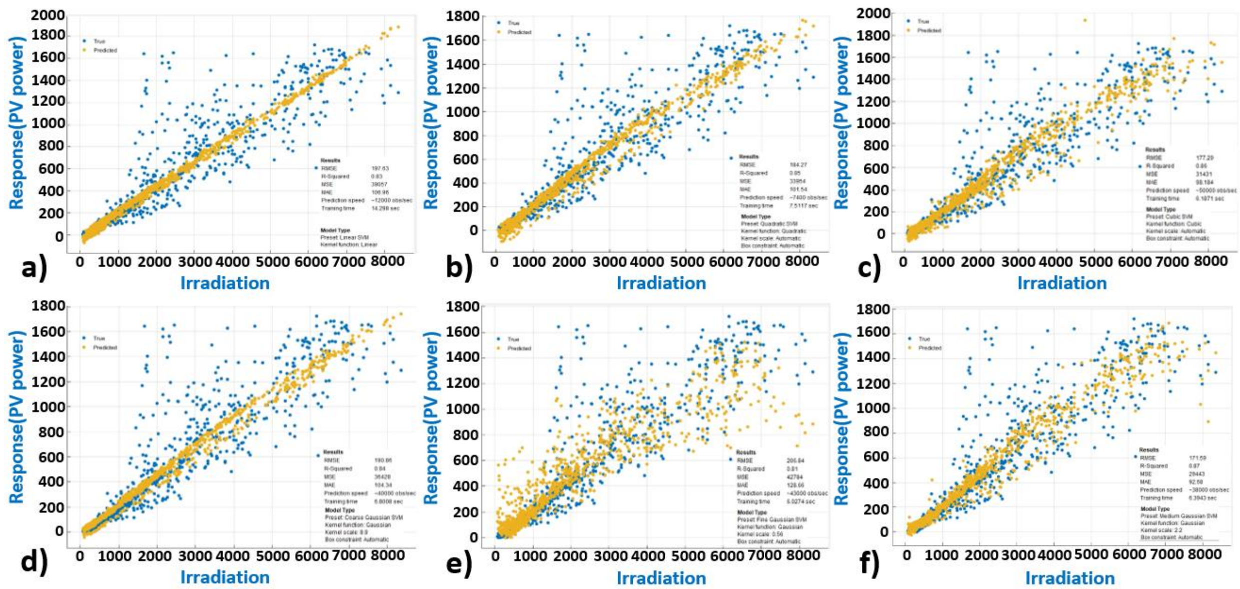

Figure 6: True (blue) and predicted values (orange) for different SVM models. a) Linear SVM b) Quadratic SVM c) Cubic SVM d) Coarse Gaussian SVM e) Fine Gaussian SVM f) Medium Gaussian SVM

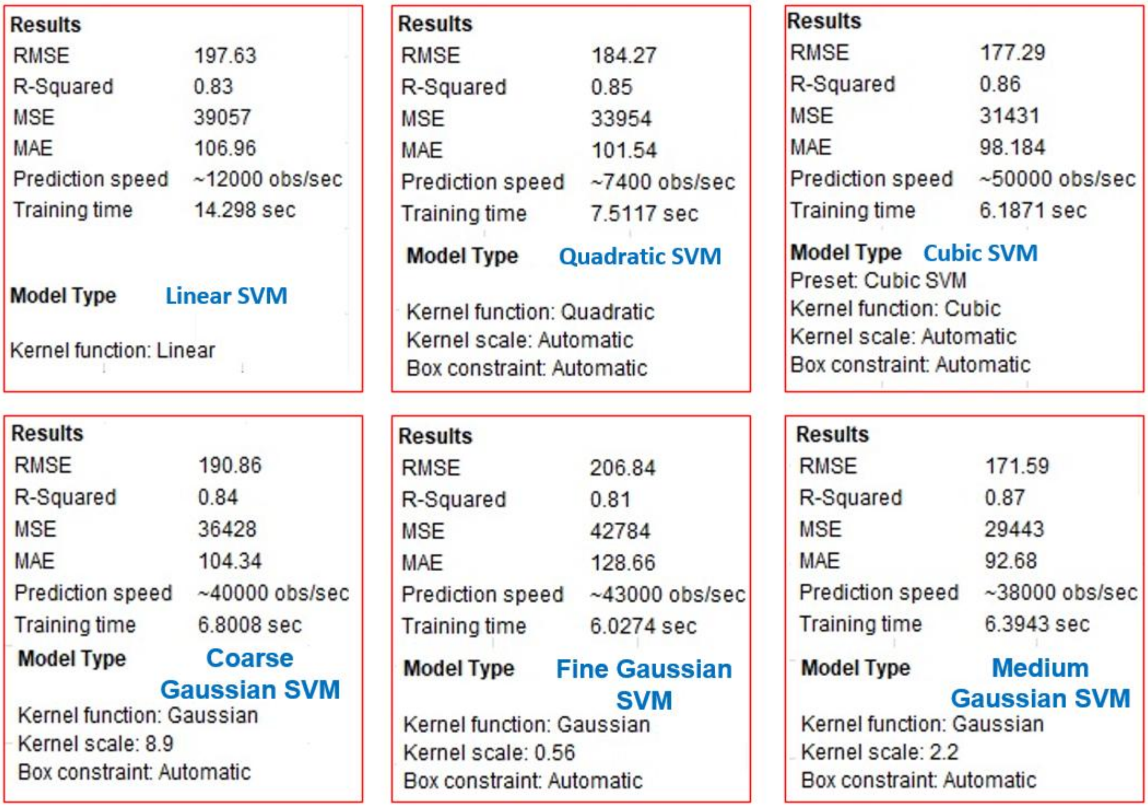

Figure 7: Benchmarking results of different SVM models

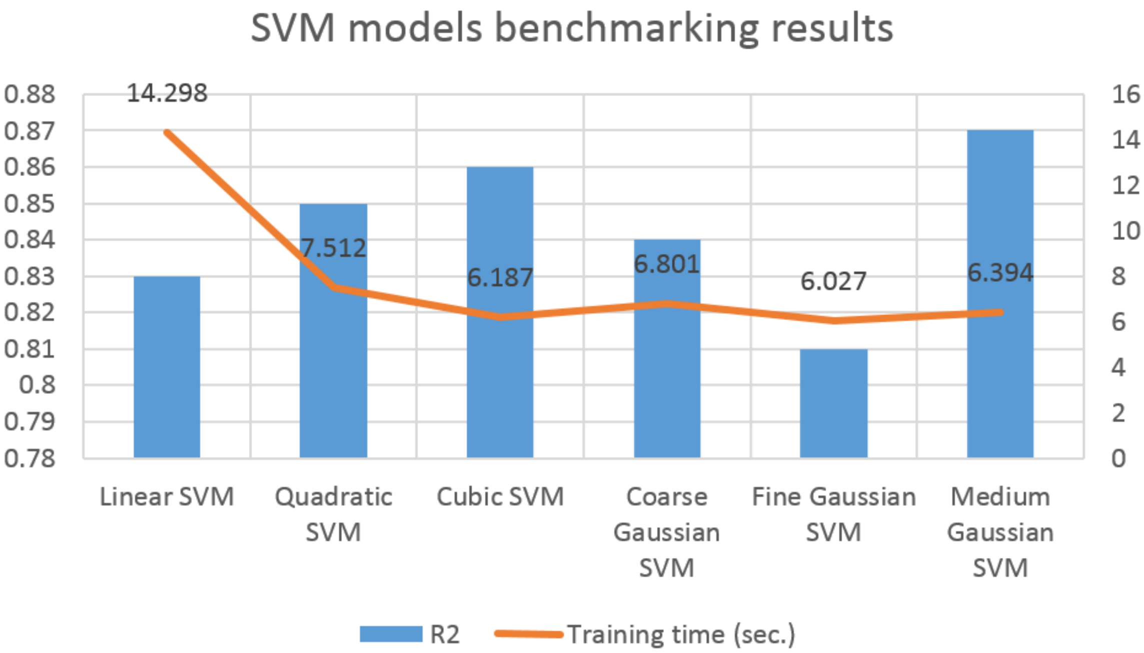

Figure 8: The comparison between R^2 and training time among SVM models

References

[1] https://scripps.ucsd.edu/programs/keelingcurve/2019/06/04/animation-of-keeling-curve-history-updated-to-include-2019-milestone/
[2] https://scripps.ucsd.edu/programs/keelingcurve/2018/09/19/is-the-current-rise-in-co2- definitely-caused-by-human-activities/
[3] Iinuma, K. , Ando, Y., East Japan disaster: lessons from the facts and preparation for future, *No to hattatsu. Brain and development*, 2012, 44. 149-52.
[4] Arvizu, D., et.al, 2011: Direct Solar Energy. In IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
[5] Diagne M, David M, Boland J, Schmutz N, Lauret P. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. *Energy Procedia*, 2014;
[6] C. Voyant *et al.*, "Machine learning methods for solar radiation forecasting : A review," *Renew. Energy*, vol.

105, pp. 569–582, 2017.

[7] J. Wang, Z. Qian, H. Zareipour, and D. Wood, “Performance assessment of photovoltaic modules based on daily energy generation estimation,” *Energy*, vol. 165, pp. 1160–1172, 2018.

[8] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression,” *Energy*, vol. 164, pp. 465–474, 2018.

[9] Yu, Y., Cao, J., & Zhu, J. "An LSTM Short-Term Solar Irradiance Forecasting under Complicated Weather Conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019

[10] Huang, W., Zhang, C., Zhang, X., Meng, J., Liu, X.,& Yuan, B. Performance assessment of photovoltaic modules based on daily energy generation estimation*IEEE Sustainable Power and Energy Conference (iSPEC)*, 1596-1600, 2019

[11] Wang, B., Cao, J., Wang, B., & Feng, S." A Solar Power Prediction Using Support Vector Machines Based on Multi-source Data Fusion," *2018 International Conference on Power System Technology (POWERCON)*, 4573-4577, 2018

[12] Mu, C., Tang, Y., & He, H. , "Observer-based sliding mode frequency control for micro-grid with photovoltaic energy integration," 2016 IEEE Power and Energy Society General Meeting (PESGM), 1-5, 2016

[13] M. W. Ahmad, J. Reynolds, and Y. Rezgui, “Predictive modelling for solar thermal energy systems : A comparison of support vector regression , random forest , extra trees and regression trees,” *J. Clean. Prod.*, vol. 203, pp. 810–821, 2018.

[14] https:// thingspeak.com

[15] https://www.mathworks.com/products/matlab.html

[16] Vapnik, V. "The Nature of Statistical Learning Theory," Springer, 2010

[17] Preda, S., Oprea, S., Bara, A., & Belciu, A."PV Forecasting Using Support Vector Machine,". *Symmetry*, 10, 748, 2018

[18] E. Ogliari, “Computational Intelligence Techniques Applied to the Day Ahead PV Output Power Forecast : PHANN , SNO and Mixed,” 2018.

[19] A. Dolara, F. Grimaccia, and E. Ogliari, “Comparison of Training Approaches for Photovoltaic Forecasts by Means of Machine Learning,” 2018.

[20] M. Kumar, I. Majumder, and N. Nayak, “Engineering Science and Technology , an International Journal Solar photovoltaic power forecasting using optimized modified extreme learning machine technique,” *Eng. Sci. Technol. an Int. J.*, vol. 21, no. 3, pp. 428–438, 2018.

[21] M. S. Al-musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, “Advanced Engineering Informatics Short-term electricity demand forecasting with MARS , SVR and ARIMA models using aggregated demand data in Queensland , Australia,” *Adv. Eng. Informatics*, vol. 35, no. April 2017, pp. 1–16, 2018.

[22] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, “Solar photovoltaic generation forecasting methods : A review Number of Day,” *Energy Convers. Manag.*, vol. 156, no. May 2017, pp. 459–497, 2020.