Improved Machine Learning Model Selection Techniques for Solar Energy Forecasting Applications

Zaim Zulkifly *, Kyairul Azmi Baharin *‡, Chin Kim Gan *

* Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia

(zaimzulkifly27@gmail.com, kyairulazmi@utem.edu.my, ckgan@utem.edu.my)

‡ Corresponding Author; Kyairul Azmi Baharin, Faculty of Electrical Engineering,Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100, Durian Tunggal, Melaka, Malaysia, Tel: +60 270 2141, Fax: +60 270 1044, kyairulazmi@utem.edu.my

Received: 08.02.2021 Accepted:05.03.2021

Abstract- Grid-Connected Photovoltaic System (GCPV) in Malaysia had become vital due to its usages and contribution to the community. One of the advanced technologies that has been implemented in the solar field is the forecasting of PV power output and comes with a great challenge to produce high accuracy. This paper focuses on developing a ranking system to evaluate the performance of selected machine learning models. In this paper four models are considered, namely Support Vector Machine (SVM), Gaussian Process Regression (GPR), Linear Regression, and Decision Tree. Utilizing high-resolution ground-based measurement of meteorological and PV system power output, evaluation metrics such as Root Mean Squared Error (RMSE), coefficient of determination (R^2), Mean Absolute Deviation (MAD), Mean Absolute Error (MAE), and computation time has been recorded to evaluate the performance of machine learning forecasting methods. Results show that the computation time is the primary criterion that differentiates the performance of forecast models. Other statistical metrics show only marginal differences in terms of performance. The ranking system developed can serve as an indicator for solar power output forecasters to determine the best model for their application.

Keywords solar energy, grid-connected photovoltaic, power output forecasting, machine learning forecasting.

1. Introduction

In recent years, the crisis of global warming and energy production have fostered research and development of alternative energy options, renewable and clean. One of the major factors that contribute to increasing the greenhouse effect is the industrial production of greenhouse gases. Therefore, one of the alternatives that can be implemented to reduce global warming cases through clean energy resources such as water that can contribute to hydropower stations and solar energy related to solar farms to produce energy. Photovoltaic (PV) systems are one of the basic technology for converting solar energy into electrical energy. Germany had installed 38.24 GW and China had installed 28.05 GW [1].

Malaysia is committed to reduce carbon dioxide (CO2) emission through The Copenhagen Accord (COP) signed in 2015. A continuation of the accord conducted in Paris stated that Malaysia aims to reduce greenhouse gas emissions by 35% from 2005 levels by 2030 for each unit of economic growth with international support, which could rise to 45% by 2030. Several initiatives have been done by Malaysia to support the growth of renewable energy such as Feed-in Tarif (FiT), Large Scale Solar (LSS) and Net Energy Metering (NEM) as part of the strategy to reduce carbon dioxide emissions. The purposed of FiT scheme introduced by the Malaysian government is to encourage people to change their major electricity supply to the solar grid power system compared by using the electricity supply from the grid [2].

Malaysia also has enormous potential for solar power generation due to the location itself that received a vast amount of sunlight every year. An early assessment of the potential and status of renewable energy in Malaysia has been conducted with the purpose of supporting the government's goal of achieving a 20% renewable energy target by 2025 [3]. Malaysia has decided to concentrate on the growth of the economic and manufacturing sector as a priority. Planning and development in solar PV related fields is done by moving beyond the production of solar panels to the production of solar energy. SunPower and First Solar are a few of the established large-scale solar manufacturing companies in Malaysia. 400–600 MJ/m2 is the average amount of solar radiation received by Malaysia [3].

Renewable energy depends on renewable resources such as the sun and wind. PV power output relies on the amount of solar irradiance received. A lot of factors can affect the amount of PV output power produced such as the presence of clouds in the sky which may cause significant changes in the incident solar irradiance [4]. An equally unpredictable change in PV output came from various factors such as a sudden change in weather conditions, including insolation levels, and temperature. A lot of factors could affect PV system performance, resulting in unexpected fluctuations in PV output [5]. This situation will also cause problems for solar energy developers to predict the next PV power output. Since weather cannot be controlled, other alternatives need to be considered. The best alternative would be to have fast-response energy storage [6]. Ideally, it should be able to compensate for a sudden loss or increase in PV power output instantaneously. However, the cost is still prohibitive at this stage in time and unless there are breakthroughs in energy storage technology or competition to drive the prices down, this option is not feasible.

The next best alternative would be to anticipate the change through solar energy forecasting. Forecasting helps to the future PV power output with less cost compared to the fast-response energy storage approach. Furthermore, from the forecasted data there are possibilities of the PV power output producing lower or higher than expected real power output, hence encouraging those solar power developers to prepare hedging strategies to compensate for power drops from the solar power plants. Therefore, forecasting PV power output acts as an important approach that should be implemented by solar power developers.

The first section of the organization of the paper discusses the issue and the problems faced by the forecasters. Section 2 of the paper will discuss the important subtopics related to forecasting such as their types, categories in forecasting techniques, and time horizons related to forecasting. Furthermore, artificial intelligence, machine learning, deep learning, weather variables, and the benefits of forecasting will also be discussed in this chapter. Section 3 will explain the flowchart of the project, methods of machine learning forecasting, data that has been used and ranking method developed including the evaluation metrics. In section 4, the discussion revolves around the relationship between weather variables, PV power output, and the strength of each machine learning method concerning the evaluation metrics. Based on the selected criteria, the ranking method developed is elaborated.

2. Literature Review

2.1. Previous Research

A research has proposed a techniques of data mining solar radiation prediction [7]. The research showed that the prediction of solar radiation using the ANN-based prediction offer a better accuracy compared to statistical, conventional, linear, nonlinear and fuzzy logic models. The paper also highlighted that solar radiation is more important compared to other renewable energy resources.

Other research works related to neural networks also indicate favourable results. The use of a dynamic recurrent neural networks (DRNN) was found to outperform a simple feed-forward multilayer perceptron (MLP) in predicting global solar radiation forecast using the same meteorological data [8]. The critical element in producing good forecasts apart from the method is the quality of data itself. Suitable data mining methods must be utilized to select and filter relevant historical data [9].

For sites where actual data is not available, validated synthetic data can serve as a reliable alternative in estimating solar irradiance especially during the planning phase of a solar farm construction [10]. This paper used synthetic data together with statistical forecasting methods. Four time series structures were involved: auto regressive, auto regressive integrated, auto regressive moving average and auto regressive integrated moving average. Other than comparing the mean of the synthetic series with the data set means, coefficient of determination (R^2) was involved to obtain the performance of the proposed method.

2.2. Introduction to Forecasting

The method of making forecasts about future events is forecasting. With absolute certainty, they are rarely made and can at best be represented in probabilistic terms, as new discoveries and improved models can lead to updated predictions [11]. Involved quantitative forecasts could be primarily divided into two categories—explanatory models and time series. Explanatory models can be explained by the forecast variable that has an explanatory relationship with one or more other variables, whereas time series forecasting uses only knowledge about the forecast variable and does not attempt to discover the factors that influence its conduct [12]. As a feature of past data, quantitative forecasting techniques are used to predict future data, and when past data is available, it will be sufficient. Moving Averages (MA) and Autoregressive Integrated Moving Average are examples of quantitative forecasting techniques (ARIMA) [13]. Therefore, time series forecasting is the most suitable forecasting concept to be applied in this project.

Qualitative forecasting relies mostly on knowledge, experience, or intuition-based human judgment. Qualitative features consist mainly of subjective inputs, mostly of definitions that are not numerical [14]. It can be used in a wide range of situations where historical data is not required or circumstances that change so rapidly that a previous data-based mathematical forecasting model can become obsolete

[15]. Summarization of both methods had been shown in Table 1.

Table 1. Summarization of Qualitative and Quantitative Methods

Qualıtatıve Method	Quantıtatıve Method
• Inputs: Non-numerical description, e.g. expertise, experience, human judgment. • Does not consider numerical measures.	• Inputs: Numerical data based on measurement of related variables depending on the scope of the project.
• Enable risk assessment to be conducted intuitively in a short time.	• Obtained more accurate prediction of risk and generally more expensive with demanding greater experience and advanced tools.
Applications: Law enforcement.	Applications: Finance, weather prediction.

2.3. Categories in Forecasting Techniques

Forecasting techniques can be divided into three major categories, which are time series, hybrid and physical methods. The time-series approach relies only on previously collected data and does not include any details about the PV systems or the location where they are located [16]. Time series has been the most popular method in the past despite its drawbacks. This includes usage complexity, time-consuming, and not entirely adaptable to large, dynamic data inputs [17].

It is possible to classify these approaches as statistical or artificial intelligence (AI) oriented. Examples of statistical approaches are autoregressive (AR), moving average autoregressive (ARMA), whereas examples of the latter are Artificial Neural Networks (ANN), machine learning, and deep learning. Many research works have utilized various forms of statistical techniques. For example, a simple and succinct methodology has been developed using the ARMA model to forecast hourly power output from a photovoltaic solar generator. Meanwhile, ANN can be categorized as machine learning methods which include several different types such as Multilayer Perceptron (MLP) and Support Vector Machine (SVM). Regression and persistence models are typically used as reference-based models [18]. The drawback of this method is that with forecasting time intervals, the prediction error increases. Statistical time series and neural network approaches are mostly intended for short-term forecasts [19]. Depending on the model architecture collection, a random initial dataset may decrease the reliability of the forecasted results.

Physical methods consist of a series of mathematical equations that explain the physical state and dynamic movement of the atmosphere, focusing primarily on the numeric weather forecast, satellite imaging and sky imagery [20]. The benefits of physical techniques are demonstrated by higher results when weather conditions are stable and that accuracy is highly influenced by sharp shifts in meteorological variables [21]. Total Sky Images benefit from very exact information about the scale, composition, and movement of existing clouds at the prediction stage, the method currently does not account for cloud production or dispersal or significant cloud geometry changes [22].

Models of Numerical Weather Prediction (NWP) are valuable resources in the process of producing wind and solar power output forecasts from a plant. These basic concepts include atmospheric motion, sources of observation and consistency, data assimilation, and the need for model production post-processing, the importance of probabilistic forecasts, and how to validate and check the forecast [23].

A variation of each of the other methods of forecasting is hybrid methods. The hybrid method's basic concept is to combine various models with specific features to overcome the shortcomings of individual techniques, thereby increasing the performance forecast [24]. Hybrid methods can be used along with historical data of meteorological variables, such as numerical weather forecasts, which can lead to better forecasting performance [21]. Fig. 1 shows the summarization of forecasting techniques based on time-series, a physical and hybrid method that has been mentioned earlier.

2.4. Introduction to Artificial Intelligence

Artificial Intelligence (AI) is the science and engineering of enabling machines to demonstrate intelligence in areas such as visual identification, speech recognition, and decision-making. In essence, it is the artificial version of human intelligence done by machines, in particular computer systems [25]. Recently, artificial intelligence is reflected as the imitation of the human brain which tries to stimulate their learning process to mimic the human brainpower [26]. In simple terms, AI aims to extend and augment the capacity and efficiency of mankind in tasks of remaking nature and governing society through intelligent machines. Machine learning (ML) is a branch of Artificial Intelligence that pushes forward the idea that, by giving access to the right data, machines can learn by themselves how to solve a specific problem [27]. In its most fundamental function, machine learning is the practice of using algorithms in various fields such as finance, marketing, development, and weather to parse data, learn from it, and then determine or predict performance. The computer is then "trained" with a specific set of instructions to accomplish a specific task using vast quantities of data and algorithms that allow it to learn how to execute the task [28].

The learning potential of deep learning is higher compared to conventional machine learning techniques and can allow greater use of datasets for extracting features [29]. In fields such as medical applications, self-driving vehicles, and predictive forecasting. It has already had a significant effect and one of the examples of deep learning is the Deep Neural Network (DNN). Fig. 2 shows the overall idea of artificial intelligence, machine learning, and deep learning.

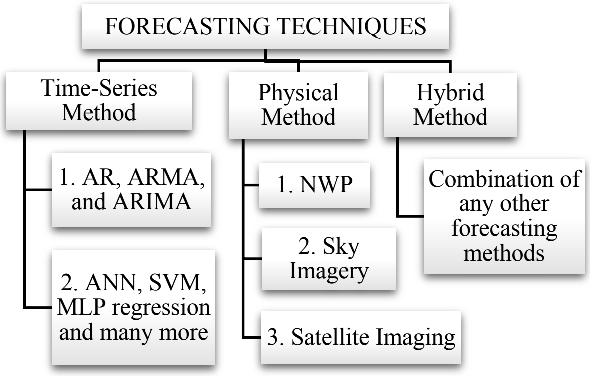

Fig. 1. Summarization of Forecasting Techniques.

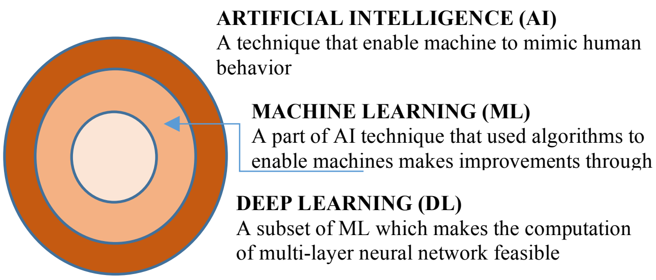

Fig. 2. Illustration of Artificial Intelligence, Machine Learning, and Deep Learning

2.5. Time Horizons in Forecasting

There are generally three types of time horizons in forecasting which are short-term, medium-term, and long-term. Short-term forecasting can be in the form of hours ranging from 0 up to 6-8 hours [30]. Business conditions and services around the world are also increasingly evolving along with the rapid growth of the internet and economic exchanges, and therefore predicting accuracy has been influenced by further uncertainties. As a result, many strategists have embraced short-term forecasting strategies to adapt in real-time to the shift in market environments. Short-term forecasts have been commonly used in small-size power systems in demand prediction, inventory management, finance, and scheduling [31].

The medium-term outlook varies from 0 hours to 7 days in advance, covering many days. This type of forecast is helpful in many ways such as electricity markets trading, and maintenance planning. Finally, the long-term prognosis can be a weekly forecast. This range encompasses applications that can start from 1-2 weeks for planning maintenance to months for planning hydro-storage. Some end-users need forecasts for up to 2 years [30]. A summary of time horizon in forecasting has been showed in Table 2.

2.6. Importance of Weather Forecasting

Forecasting played an important role and very useful in various fields such as finance, production, and service providers—weather conditions and demographic variables. There are also adverse effects of major PV fluctuations on an unprepared grid, in addition to the technological difficulties of a grid-connected PV system, including voltage breaches, flicker, and other power quality problems, reverse power flow and safety synchronization problems, increased wear on utility equipment, and actual and reactive power imbalances [32]. From the precise viewpoint, it is very important to contribute and improve the efficiency of electricity supplied to the grid and, therefore, to reduce the supplementary costs associated with general fluctuations [33]. Any climate change, ranging from agriculture transportation to water resource management, will affect many social, economic, and tourism activities [34].

Table 2. Time Horizon in Forecasting

Time Horizon	Period	Example	Description
Short Term Forecast	0 up to 6-8 hours	• 5-minutes ahead • 30-minutes ahead • Hourly	Utilized in demand estimation, inventory control, finance, and scheduling in small size power systems.
Medium-Term Forecast	0 hours up to 7 days	• 2-hours ahead • Daily • 2-days ahead	useful for functions such as unit commitment, trading in electricity markets, and even maintenance planning
Long Term Forecast	1-2 weeks up to months	• Weekly • Monthly	Useful for maintenance planning up to months for hydro-storage planning.

The supply and transportation of energy are critical to the future of all developing countries. It is reasonable to predict that energy demand will continue to rise in the near future. Sustainability elements related to renewable energy have become one of the big concerns for both consumers and energy providers. A balanced partnership between society and the environment requires the deployment of renewable energy sources. Renewable energy also provides customers with safe and affordable choices for living in a healthy and environmentally sustainable world [35]. Lastly, weather forecasting provides a lot of advantages in many aspects such as daily activities, electrical, construction, transportation, agriculture, marine, and military. Thus, unprepared individuals and businesses can end up in risky situations and may end up being harmed without proper weather forecasts [36].

2.7. Weather Variables

The status of the atmosphere refers to weather conditions, defining the status of the atmosphere either it is hot or cold, calm or stormy, and many more. Weather variables such as solar irradiance, wind speed, cloud cover, and ambient temperature can affect the PV power output produced. For instance, a sudden change in solar irradiance on the surface of the solar panels will also make instantaneous changes in the PV power output. Therefore, issues such as power fluctuation occurred at the grid was one of the effects of a sudden change in solar irradiance [37]. All the weather variables mentioned earlier can affect the machine learning forecasting accuracy and error of the data. A lot of forecasting techniques have been done in other countries and the process of collecting and filtering input data was very important before forecasting.

The scattered radiation directly from the sun on the earth's surface is called direct solar radiation. Solar radiance is a calculation of the photovoltaic module's power output. However, since solar irradiance is affected by air substances, the solar photovoltaic is unable to generate stable power output and the module is therefore shielded from the clouds, predicting solar irradiance accurately is an essential problem [38].

To have a better understanding of climate variability, climate dynamics, and radiation balance, high-resolution space-time surface solar irradiance data is necessary [39]. The previous research paper, clearly showed that solar irradiance is the most important element that can affect the photovoltaic power output to be produced.

The wind is also one of the energies that can be harnessed and converted by turbines called wind turbines into electricity and the amount of electricity generated depends on its size and wind speed. Wind speed is significant because wind speed and velocity largely decide the amount of electricity that wind turbines can produce [40]. In the aspect of solar energy production, wind speed has been one of the factors that affect the power output produced from the solar panels. Wind speed can cause other variables such as solar irradiance and panel temperature change from the expected value based on the weather condition hence decreasing the forecasting power output from the solar power plant.

Ambient temperature means the surrounding temperature of a machine or equipment. In this research, ambient temperature refers to the temperature around the solar panels. Ambient temperature can also increase or decrease the panels' temperature hence, it will also affect the power output produced by the panels. Research on ambient temperature has been done and claimed that the prediction of energy consumption and solar-radiation are also related to ambient air temperature forecasting [41]. Ambient temperature will affect the temperature of solar panels hence will be related to the PV power output productions. However, ambient temperature has a positive association with the system's power production, so the temperature can be used in the design and estimation of the amorphous silicon photovoltaic solar system's success in the field of research.

3. Methodology

3.1. Data

The measured Global Horizontal Irradiance (GHI) and the values of atmospheric temperature used in this article are obtained from a weather station situated at 2.32ºN, 102.3ºE, on the rooftop of the Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka. There are two stations mounted on the rooftop for redundancy. For each second, the data is collected and stored in a main and backup server. The average amount of data for this paper is every five minutes and the period is from 1 January 2014 to 31 December 2016. The data were split into two categories, with training data from 1 January 2014 to 31 December 2015 and data from 1 January 2016 to 31 December 2016 being used for training.

3.2. Flowchart of Data Processing

Fig. 3 shows the flowchart of data processing that has been done in this project. Collecting and filtering data is the first stage to gather and analyze the data then followed by separating the data into two categories that are training data and testing data. After the process of prediction takes place by the machine learning, evaluation metrics will be used to evaluate the performance of each of the machine learning models. Last but not least, a ranking system has been implemented based on the computation time between those machine learning methods.

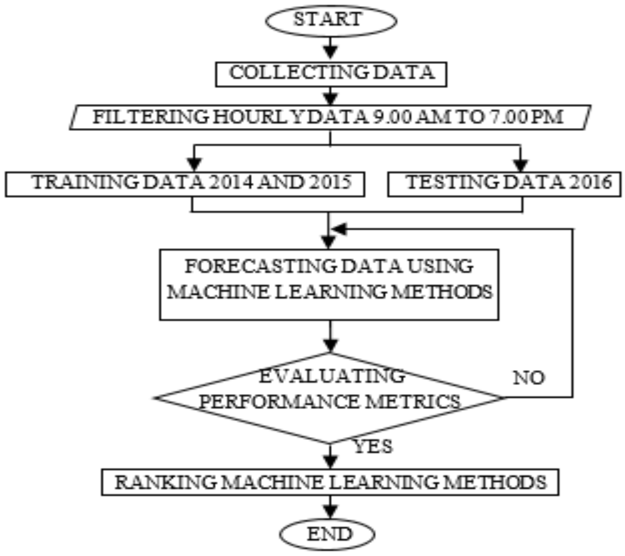

Fig. 3. Flowchart of Data Processing

3.2.1. Collecting and Filtering Data

In this project, data has been collected for the years 2014, 2015, and 2016. All the variables such as solar irradiance, tilt irradiance, wind speed, humidity, and power output for each inverter installed in the solar laboratory were being tabulated and stored in the Excel file. Furthermore, all the collected data will be filtered based on weather conditions. Some of the power output data were lost due to the activities in the lab such as maintenance of the system. During the maintenance, some inverters needed to be shut down for services. Another reason can cause incomplete data such as maintenance activities at the ground-based measurement station, and broken sensors. In the year 2016, the amount of data loss was higher compared to 2015 and 2014.

3.2.2. Training and Testing Data

In this project, data that has been collected for the years 2014 and 2015 has been used as training data for machine learning forecasting. The input variables involved in the training data were ambient temperature, solar irradiance, and historical PV power output. A total of 12,048 hourly data were recorded for the years 2014, 2015, and 2016. From this portion, 8033 training data from the years 2014 and 2015 were used for training while 4015 data that has been collected in 2016 were used for testing. The same input variables involved in the training data has been used in the testing data.

3.3. Machine Learning

Machine learning methods used in this project are GPR, (SVM), Linear Regression, and Decision Tree. The selection is based on commonly used methods for regression applications. To validate the performance, a series of evaluation metrics are needed to benchmark each model's accuracy and margin of error. The terms used to define the characteristics of a given model has been showed in Table 3.

Table 3. Terms of Machine Learning Models

Term	Description
Interpretability	The ability of a prediction model to explain the prediction in understandable terms to human.
Flexibility	The ability of a prediction model to vary within a given set of training data.
Overfitting	As a consequence of an overly complex model, the algorithm captures the noise of results.
Underfitting	As a result of an overly simple model, the algorithm can not capture the basic pattern of data.

3.3.1. Support Vector Machine (SVM)

In machine learning, SVM is defined as supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Research projects have explored the use of support vector machines to estimate solar irradiance hourly in a tropical country [42]. The evaluation tests for the performance verifications for the individual models and the result showed that the SVM performs considerably better than standard technology. The characteristics of support vector machine (SVM) models have been presented in Table 4. Apart from linear SVM, all other SVM variants have poor interpretability and varying ranges of flexibility.

3.3.2. Linear Regression

As these regression estimates are used to describe the association between the variables, linear regression is a simple and widely used form of predictive analysis. A regression curve can be used to establish a correlation between an interesting variable and one or more associated variables that are supposed to affect the problem of forecasting. If data is abundant, the regression curve reflects a curve moving through the bulk of the data [43]. Normally, the first few subjects selected by individuals in researching predictive modelling is the linear regression. The characteristics of linear regression models has been explained in Table 5. In general, all its variants have good interpretability but low flexibility.

3.3.3. Gaussian Process Regression (GPR)

Over the last decade, Gaussian Processes (GP) have received considerable attention from many authors due to their frequent appearance in many areas of statistical theory and practice, particularly in machine learning approaches [44]. A powerful statistical learning tool for modeling non-linear mapping is Gaussian process regression and it is supervised learning and can be divided into two categories: classification and regression. Discrete class labels are the outputs for classification, regression concerns the prediction of continuous quantities [45]. The characteristics of the Gaussian Process Regression models in MATLAB showed in Table 6. The different kernel functions used by each model determine the differences in data interpretation.

Table 4. Characteristics of the Support Vector Machine (SVM) models

Model Type	Interpretability	Flexibility
Linear SVM	Easy	Low
Quadratic SVM	Hard	Medium
Cubic SVM	Hard	Medium
Fine Gaussian SVM	Hard	High
Medium Gaussian SVM	Hard	Medium
Coarse Gaussian SVM	Hard	Low

Table 5. Characteristics of Linear Regression models

Model Type	Interpretability	Flexibility
Linear	Easy	Very Low
Interactions Linear	Easy	Medium
Robust Linear	Easy	Very low
Stepwise Linear	Easy	Medium

Table 6. Characteristics of the Gaussian Process Regression models

Model Type	Interpretability	Flexibility
Exponential GPR	Hard	Automatic
Matern 5/2 GPR	Hard	Automatic
Squared Exponential GPR	Hard	Automatic
Rational Quadratic GPR	Hard	Automatic

3.3.4. Decision Tree

A map of decisions about an item is utilized by the decision tree as a forecasting model to conclude about the target value of the item. In contrast to other classification methods used in analytics, data mining, and machine learning, it is one of the predictive modeling approaches that can be developed relatively quickly [46]. The tree structure represents three important items those are nodes represent objects, divergence paths represent the attribute values possibility, and each leaf node corresponds to the value of the entity represented by the path to the leaf node from the root node [47]. Fig. 4 shows a simple example to illustrate the decision tree concept and Table 7 summarizes the interpretability and flexibility of the models.

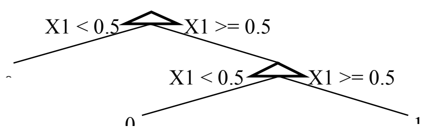

Fig. 4. Illustration of the Decision Tree Concept.

Table 7. Characteristics of the Regression Tree models

Model Type	Interpretability	Flexibility
Fine Tree	Easy	High
Medium Tree	Easy	Medium
Coarse Tree	Easy	Low

3.4. Ranking Method and Evaluation Metrics

Material and decision selection is an important process that needs to be carefully carried out to increase the probability of success in decision making [48]. Decision-making involves many parameters and criteria to be considered to find the best solution. Multicriteria Decision-making (MCDM) refers to the prioritization, ranking, or selection of human judgment-based alternatives from A finite set of alternatives for decisions in terms of many typically contradictory parameters [49].

Evaluation metrics are the formulae used in this paper to determine the accuracy and error obtained from the machine learning forecast. As a common statistical metric, the RMSE has been applied to calculate model efficiency in meteorology, climate science studies, and other applications. While the MAE is another useful measuring tool commonly applied in model assessments [50]. RMSE is a metric that is often used interchangeably as a measure of forecast accuracy [51].

Furthermore, some previous researches also applied the same evaluation metrics. For example, a research on the characterization of solar parameters played an important role in the installation of solar energy. In this paper, three different curve fitting methods, Fourier, sum of sines and smoothing spline were being applied to model global solar radiation and air temperature parameters. Hence, there were two evaluation metrics used those are coefficient of determination (R^2) and the root mean squared error (RMSE) to determine the accuracy of the proposed method [52].

Besides, the determination coefficient is the proportion of the variance in the dependent variable that can be estimated by one or more independent variables. It gives a measure of how the model replicates well-observed findings, based on the proportion of total variance illustrated by the outcomes. The R2 determination coefficient will have a value ranging from 0 to 1. A value of R2 near 1 indicates that the different input values explain the majority of the variation of the response results, while a value of R2 near 0 indicates that little of the variation is explained by the different input values [53].

Mean Average Deviation (MAD) is often regarded as a robust measure of the scale of distribution [54]. By averaging the alleged mistake, MAD tests the accuracy of the forecast by (the absolute value of each error) [55]. The following formulae are the formula of MAE, RMSE, $R^{2,}$ and MAD.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|e_i^2|} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{2}$$

$$R^2 = \frac{n\,(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum x^2 - (\sum x)^2]}} \tag{3}$$

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n}|xi - \bar{x}|}{n} \tag{4}$$

4. Results

4.1. Weather Variables Against PV Power Output

As mentioned earlier, weather variables are important input data for the process of forecasting in machine learning. A good relationship between the variables and PV power output will help in generating a well-organized forecasting method. Therefore, the selection of correct inputs for machine learning is very important to produce less error and high accuracy.

Fig.5 depicts the PV power output variation with the change in ambient temperature and the PV performance increases with ambient temperature linearly and vice versa. The correlation between both variables represented by the figure. There is also a field study that claimed both variables had an efficient correlation with each other [56]. The research has been done on an online platform of a weather forecast and ambient temperature is one of the important inputs for the ANN in the research. Therefore, ambient temperature could be included as input to produce a high accuracy forecasting method.

The next weather variable examined in this paper is the wind speed. Based on the theory, wind speed plays an important role in heat dissipation which reduces or decreases the cell temperature. Wind speed in this area varies and could be one of the variables that would influence the PV power output. Fig. 6 showed that the PV power output pattern has a low correlation with the pattern of wind speed. Even though wind speed can reduce the temperature of solar panels for optimum absorption of solar irradiance, it is not reflected in the data since the relationship is indirect. A weak correlation indicates that if used as an input, it will lead to a decrease in forecast accuracy.

Fig. 7 shows that PV power output had the best correlation with the solar irradiance compared to ambient temperature and wind speed. The pattern of solar irradiance against the PV power output is well-mapped between each other and this points out that solar irradiance could be the highest impact on input data for forecasting. Therefore, solar irradiance is best recommended as one of the important inputs for machine learning forecasting. Solar irradiance has the strongest influence in accurately forecasting PV power output.

In addition, other weather parameters such as cloud cover, shading, and humidity also affect the PV power output. Higher forecasting accuracy needs a larger number of inputs to the forecast models. However, the cost and complexity of model computation will also be increased due to the process of aggregating a vast amount of input parameters. Hence, a forecast model with optimal inputs will be a huge contribution to overcome the problems of expensive cost and complexity while at the same time producing good forecasting performance with high accuracy and low error.

4.2. Forecasting Result of Machine Learning Methods

The results of forecasting for each of the machine learning methods have been recorded and tabulated in Table 8. Standard evaluation metrics such as the RMSE, MAE, MAD, and R2 have been calculated to evaluate the performance of each machine learning and computation time that has also been recorded in the same table.

First of all, the coefficient of determination, R2, and the RMSE for each machine learning method has been recorded in Table 8. In this project, the coefficient of determination (R2 > 85%) and root mean squared error (RMSE < 10%) has been chosen as the reference indicator for each machine learning methods. Based on the tabulated results, the value of R2 and RMSE for each machine learning methods evaluation metrics exceed the minimum reference value that has been set. Therefore, both evaluation metrics showed that all the machine learning methods are well-forecast using the same input data set.

The overall best performance in terms of RMSE is Fine Tree (5.83%) which also obtained the best coefficient of determination (R2 = 95.91%) compared to other machine learning methods. Fig. 8 shows the scatter plot (R2) of Fine Tree, Cubic SVM, Interaction Linear Regression, and Rational Quadratic GPR. From the four methods that have been chosen, Fine Tree has the highest value of R2 which means that it had the highest accuracy among the others. It is shown that the amount of the data from Fine Tree that plotted away from the linear line is less than the data from Squared Exponential SVM plotted away from the linear line. The best value of R2 can be obtained if all the forecast data superimposes the input data. Therefore the distance between every point towards the linear regression line plays an important role in representing the accuracy of the model that has been developed.

Based on Table 9, the main reason Fine Tree has the best evaluation metrics performance is due to the ability to interpret data easier and has high flexibility in terms of processing the algorithm. Another proof can be seen in the result of MAD, Fine Tree has the lowest value of MAD (3.7%). Furthermore, the value of MAE for each machine learning method was almost the same and can be round off to 35%. MAE that has been calculated showed the model evaluation that exactly represented the whole system performance based on the error calculated.

A summary of the evaluation metrics results indicates that there are insignificant differences between the performance metrics. This is because of the criteria of the data that were used in this paper and the ability of the evaluation metrics to process the data to obtain the average or mean error. Different data sets may produce results with favor other types of models depending on the context and expectation. Therefore, this paper suggests the use of computation time as the primary indicator to determine the appropriate model selection.

4.3. Ranking of Forecasting Methods

The Computation time is the time for machine learning methods to complete the process of forecasting PV power output. From Table 8, Rational Quadratic GPR has the longest time while Interaction Linear Regression has the shortest time taken to complete the process mentioned earlier. The cons for Rational Quadratic GPR is the time computation recorded that is 15032 seconds which is over four hours to complete forecasting. This case will be a disadvantage for the users to predict the value of PV power output for the next hour. As shown in Table 8, machine learning methods from the group of Gaussian Process Regression have hard data interpretability. This is one of the factors that affect forecasting computation time.

The ranking method is recorded in Table 8. The lowest computation time will be first placed and vice versa. The lowest computation time in this project comes from Interaction Linear Regression. From Table 5, the group of Linear Regression has easy data interpretability to process the input data. This is the main factor that contributes to having the shortest computation time in forecasting. Therefore, Interaction Linear Regression has great potential to overcome a sudden change in climate and forecast the next PV power output. Fig. 10 shows the hourly forecasted of Fine Tree, Squared Exponential SVM, Robust Linear Regression, and Exponential GPR with the real PV power output from 1st January 2016 until 4th January 2016. All of the machine learning methods that have been chosen in this figure has the lowest percentage of RMSE compare to other methods in each of the groups. It clearly showed from the figure, that there is no big difference between the forecasted PV power outputs with the real PV power output but there is a big difference in computation time as shown in Fig. 9.

5. Conclusion

This research work has been done to highlight that the average result of forecasting could be obtained from the same input data for weather forecast but there is one important aspect that should be considered to forecast weather data that is computation time. The forecast model should have the flexibility to accommodate new data when there is a sudden change in weather profile. In this scenario, the aspect of computation time should be the best criterion to be highlighted in the PV forecasting field.

Table 8. Result of RMSE, MAE, R^2, Computation Time(s) and Weightage Ranking Method

Method		RMSE (%)	MAE (%)	R2 (%)	MAD (%)	Computation Time (s) Hour (h):Minutes (m):Seconds (s)	Ranking
Gaussian Process Regression (GPR)	Matern 5/2 GPR	7.95	34.93	92.19	5.2	2:29:54	15
	Rational Quadratic GPR	8.01	34.93	92.05	5.2	4:10:32	17
	Exponential GPR	7.78	34.93	92.56	5.1	2:0:21	16
	Squared Exponential GPR	8.01	34.93	92.04	5.2	1:30:49	14
Linear Regression	Linear Regression	8.14	34.93	92.13	5.6	0:0:9	2
	Interaction Linear Regression	7.91	34.95	92.79	4.8	0:0:6	1
	Robust Linear Regression	7.66	35.37	92.40	4.8	0:0:14	4
	Stepwise Linear Regression	7.91	34.95	92.79	5.3	0:0:15	6
Regression Tree	Coarse Tree	7.56	34.93	92.96	4.9	0:0:13	3
	Fine Tree	5.83	34.93	95.91	3.7	0:0:18	7
	Medium Tree	6.93	34.93	94.09	4.5	0:0:14	5
Support Vector Machine (SVM)	Linear SVM	7.70	35.20	92.32	4.9	0:13:53	11
	Quadratic SVM	7.55	35.33	92.86	4.6	0:14:17	12
	Cubic SVM	7.57	34.93	92.88	4.6	1:30:29	13
	Fine Gaussian SVM	7.43	35.57	93.34	4.5	0:5:0	8
	Medium Gaussian SVM	7.47	35.56	93.19	4.5	0:9:18	9
	Coarse Gaussian SVM	7.51	35.50	93.02	4.5	0:13:30	10

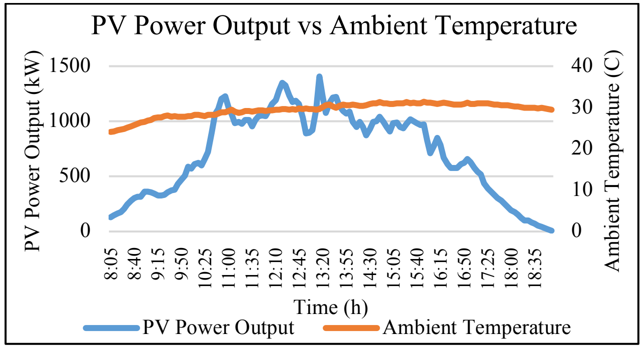

Fig. 5. PV Power Output against Ambient Temperature

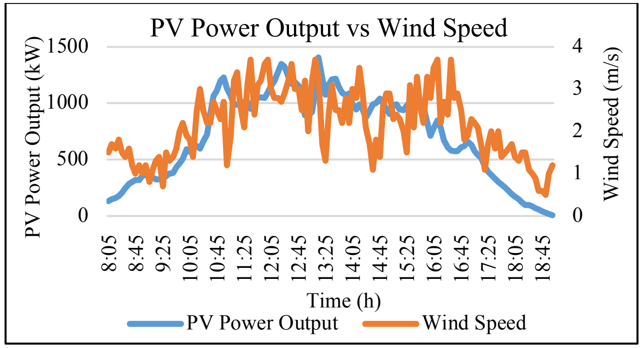

Fig. 6. PV Power Output against Wind Speed

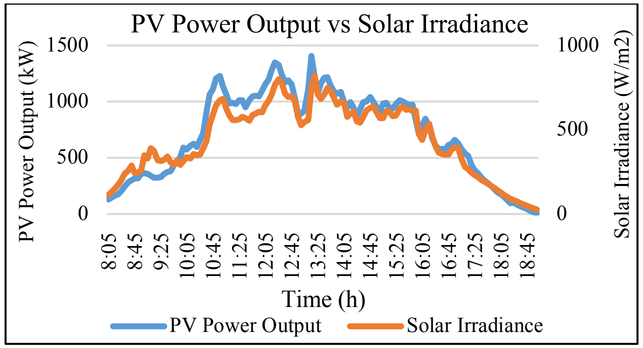

Fig. 7. PV Power Output against Solar Irradiance.

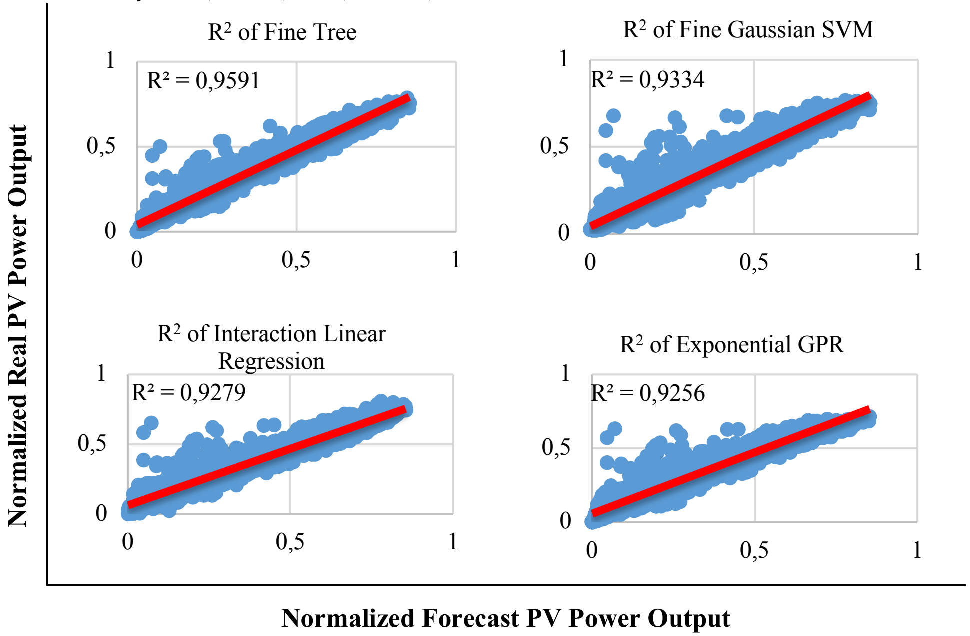

Fig. 8. R^2 of Fine Tree, Squared Exponential SVM, Robust Linear Regression and Exponential GPR.

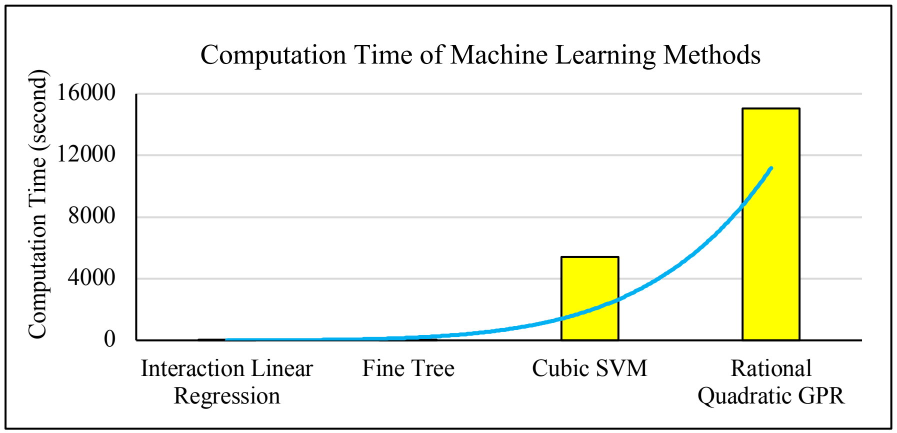

Fig. 9. Computation Time of Fine Tree, Cubic SVM, Interaction Linear Regression, and Rational Quadratic GPR.

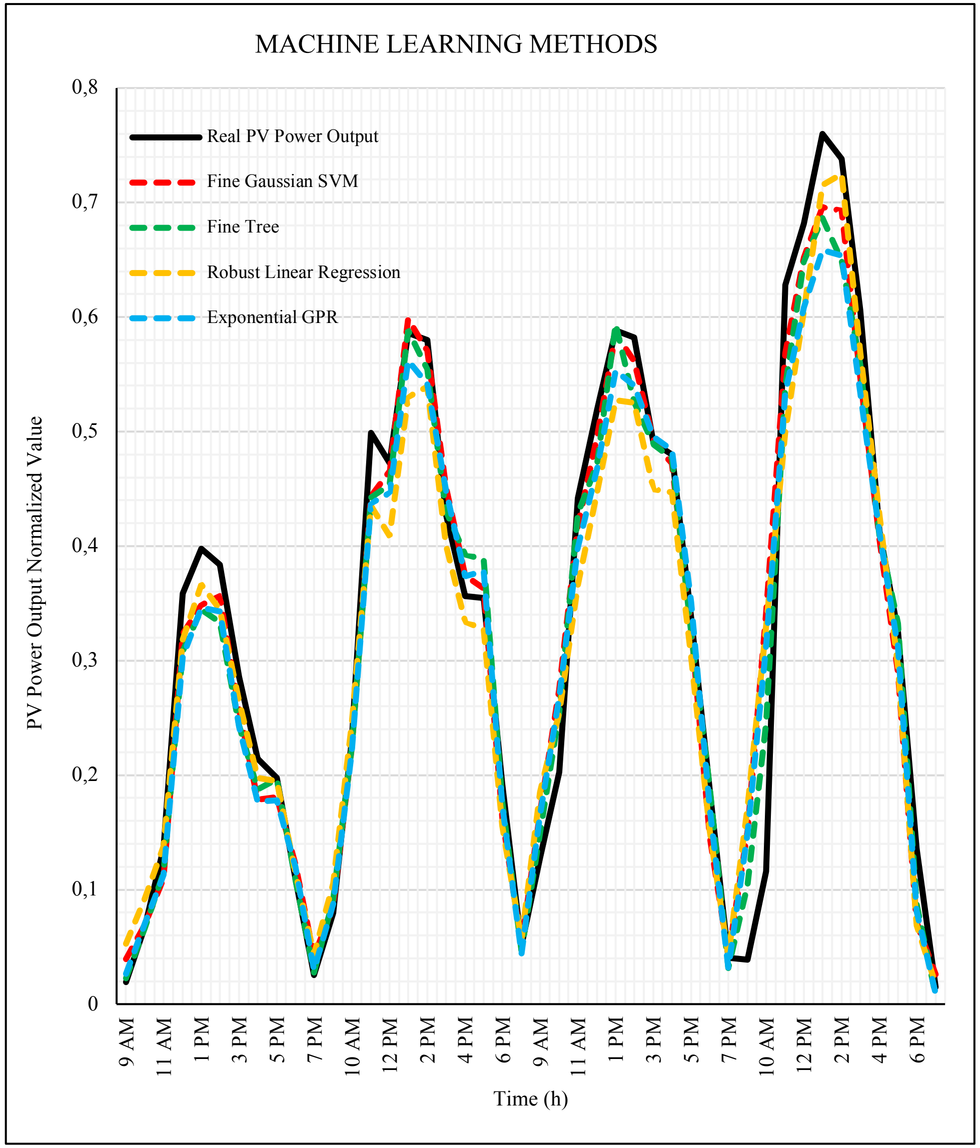

Fig. 10. Hourly Forecasted of Fine Tree, Squared Exponential SVM, Robust Linear Regression, and Exponential GPR.

For example, the result in Table 8 showed that both Interaction Linear Regression and Rational Quadratic Gaussian Process Regression have RMSE value 7.91% and 8.01% respectively. Even though there is a slight difference in the result of evaluation metrics but a greater difference in computation time. Interaction Linear Regression only need six seconds while Rational Quadratic Gaussian Process Regression required more than two hours to complete the forecasting process. Hence, it is worth to be considered in the forecasting process. Therefore, this project highlights the potential of computation time in the machine learning method in forecasting PV power output.

For future work, the study should consider different data duration so that the machine learning model can tune its algorithm more precisely. Furthermore, if data with shorter or longer time intervals such as five-minute, one-minute, daily or two-hour data is available, this should also be investigated as the input variables help increase the accuracy and reduce the error of PV power output forecasting. Lastly, hybrid models or multilayer of machine learning should also be studied to determine whether they produce similar results and identify which method is the most suitable for localized PV system forecasting.

Acknowledgements

This research work was sponsored under Universiti Teknikal Malaysia Melaka (UTeM) Fundamental Research Grant No. FRGS/2018/FKE-CERIA/F00351. The main author would like to thank Universiti Teknikal Malaysia Melaka and the Ministry of Higher Education Malaysia for supporting his Masterwork.

Reference

[1] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, 2016, doi: 10.17775/cseejpes.2015.00046.

[2] "Feed-in Tariff in Malaysia : An Economic Assessment of Renewable Energy policy."

[3] M. R. Islam, R. Saidur, N. A. Rahim, K. H. Solangi, and E. Tower, "Usage of Solar Energy and Its Status in Malaysia," vol. 5, no. 1, pp. 6–10, 2010.

[4] I. Ranaweera, "SHORT-TERM INTERMITTENCY OF SOLAR IRRADIANCE IN SOUTHERN," no. August 2016, 2014.

[5] M. S. Elnozahy, "Technical Impacts of Grid-Connected Photovoltaic Systems on Electrical Networks — A Review Technical impacts of grid-connected photovoltaic systems on electrical networks — A review Additional information on J . Renewable Sustainable Energy," no. March 2016, 2013, doi: 10.1063/1.4808264.

[6] N. Ertugrul, "Battery storage technologies, applications and trend in renewable energy," *IEEE Int. Conf. Sustain. Energy Technol. ICSET*, vol. 0, no. August, pp. 420–425, 2017, doi: 10.1109/ICSET.2016.7811821.

[7] A. Qazi, H. Fayaz, and R. G. Raj, "Discourse on Data Mining Applications to Design Renewable Energy Systems," no. August 2017, 2014, doi: 10.15242/iie.e0314204.

[8] R. Al-Hajj, A. Assi, and M. M. Fouad, "A predictive evaluation of global solar radiation using recurrent neural models and weather data," in *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, pp. 195–199, doi: 10.1109/ICRERA.2017.8191265.

[9] M. Yesilbudak, M. Çolak, and R. Bayindir, "A review of data mining and solar power prediction," in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, 2016, pp. 1117–1121, doi: 10.1109/ICRERA.2016.7884507.

[10] L. F. N. Lourenço, M. B. de Camargo Salles, M. M. F. Gemignani, M. R. Gouvea, and N. Kagan, "Time Series modelling for solar irradiance estimation in northeast Brazil," in *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, pp. 401–405, doi: 10.1109/ICRERA.2017.8191093.

[11] S. M. Hohle and K. H. Teigen, "Forecasting forecasts: The trend effect," *Judgm. Decis. Mak.*, vol. 10, no. 5, pp. 416–428, 2015.

[12] R. J. Hyndman, "Forecasting: An Overview," *Int. Encycl. Stat. Sci.*, pp. 536–539, 2011, doi: 10.1007/978-3-642-04898-2_256.

[13] S. N. Kane, A. Mishra, and A. K. Dutta, "Preface: International Conference on Recent Trends in Physics (ICRTP 2016)," *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016, doi: 10.1088/1742-6596/755/1/011001.

[14] B. Rama Sanjeeva Sresta and V. Arundhati, "Analysis of Forecasting Sales By Using Quantitative and Qualiative Methods," *Int. J. Eng. Res. Appl.*, vol. 6, no. 9, pp. 83–86, 2016.

[15] Cochran, James J. Cox, Louis A. Keskinocak, Pinar Kharoufeh, Jeffrey P. Smith, J. Cole Wang, Shouyi Chaovalitwongse, Wanpracha Art., "Evaluating and Comparing Forecasting Models," *Wiley Encycl. Oper. Res. Manag. Sci.*, no. March 2018, 2011, doi: 10.1002/9780470400531.eorms0307.

[16] Nespoli, Alfredo Ogliari, Emanuele Leva, Sonia Pavan, Alessandro Massi Mellit, Adel Lughi, Vanni Dolara, Alberto, "Day-ahead photovoltaic forecasting: A comparison of the most effective techniques," *Energies*, vol. 12, no. 9, pp. 1–15, 2019, doi: 10.3390/en12091621.

[17] M. Shah and R. Agrawal, "A Review On Classical And Modern Techniques with Decision Making Tools For Load Forecasting," vol. 6, no. 3, pp. 174–184, 2013.

[18] D. B. De Alencar, C. De Mattos Affonso, R. C. L. De Oliveira, J. L. M. Rodríguez, J. C. Leite, and J. C. R. Filho, "Different Models for Forecasting Wind Power Generation: Case Study," *Energies*, vol. 10, no. 12, 2017, doi: 10.3390/en10121976.

[19] W.-Y. Chang, "A Literature Review of Wind Forecasting Methods," *J. Power Energy Eng.*, vol. 02, no. 04, pp. 161–168, 2014, doi: 10.4236/jpee.2014.24023.

[20] C. Monteiro, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, A. Muñoz-Jimenez, and P. M. Lara-Santillan, "Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques," *Math. Probl. Eng.*, vol. 2013, no. July 2016, 2013, doi: 10.1155/2013/767284.

[21] S. Al-Dahidi, O. Ayadi, J. Adeeb, M. Alrbai, and B. R. Qawasmeh, "Extreme Learning Machines for Solar Photovoltaic Power Predictions," *Energies*, vol. 11, no. 10, 2018, doi: 10.3390/en11102725.

[22] P. Luukkonen, P. Bateman, J. Hiscock, Y. Poissant, and D. H. and L. Dignard-Bailey, "International Energy Agency Co-Operative Programme on Photovoltaic Power Systems," p. 2013, 2013, Accessed on: May 5, 2020. [Online]. Available: http://www.cansia.ca/sites/default/files/201306_cansia_2012_pvps_country_report_long.pdf.

[23] E. N. Lorenz, "Energy and Numerical Weather Prediction," *Tellus*, vol. 12, no. 4, pp. 364–373, 1960, doi: 10.3402/tellusa.v12i4.9420.

[24] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Math. Comput. Simul.*, vol. 131, pp. 88–100, 2017, doi: 10.1016/j.matcom.2015.05.010.

[25] I Yu, Chaochao Chen, Rui Li, Juan Jinxing Li, Juan Jinxing

Drahansky, Martin Paridah, M.t Moradbak, Amin Mohamed, A.Z Owolabi, FolaLi, Hongqhan abdulwahab taiwo Asniza, Mustapha Abdul Khalid, Shawkataly H.P Sharma, Taruna Dohare, Neeraj Kumari, Meena Singh, Upendra Kumar Khan, Abbul Bashar Borse, Mahendra S Patel, Rajan Paez, Angela Howe, Anna Goldschmidt, Degussa Corporation, Chemical Coates, John Reading, Further., "We are IntechOpen , the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %," *Intech*, vol. i, no. tourism, p. 13, 2012, doi: 10.1016/j.colsurfa.2011.12.014.

[26] J. Shabbir and T. Anwer, "Artificial Intelligence and its Role in Near Future," vol. 14, no. 8, pp. 1–11, 2018, Accessed on: January 15, 2020. [Online]. Available: http://arxiv.org/abs/1804.01396.

[27] Musumeci, Francesco Rottondi, Cristina Nag, Avishek Macaluso, Irene Zibar, Darko Ruffini, Marco Tornatore, Massimo, "An Overview on Application of Machine Learning Techniques in Optical Networks," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2019, doi: 10.1109/COMST.2018.2880039.

[28] A. Habeeb, "Artificial intelligence Ahmed Habeeb University of Mansoura," *Res. Gate*, vol. 7, no. 2, 2017, doi: 10.13140/RG.2.2.25350.88645.

[29] S. W. X. Du, Y. Cai, "Overview of deep learning." 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, pp. 159–164, 2016.

[30] G. Giebel and G. Kariniotakis, "Best practice in short-term forecasting. A users guide," *Eur. Wind Energy Conf. Exhib. 2007, EWEC 2007*, vol. 2, no. May, pp. 850–854, 2007.

[31] H. S. Dang, Y. F. Huang, C. N. Wang, and T. M. T. Nguyen, "An application of the short-term forecasting with limited data in the healthcare traveling industry," *Sustain.*, vol. 8, no. 10, 2016, doi: 10.3390/su8101037.

[32] K. Vaibhav, "Solar Power forecasting : The State-Of-The-Art Solar Power forecasting : The State-Of-The-Art," no. June 2014, 2015.

[33] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manag.*, vol. 156, no. May 2017, pp. 459–497, 2018, doi: 10.1016/j.enconman.2017.11.019.

[34] L. Al-Matarneh, A. Sheta, S. Bani-Ahmad, J. Alshaer, and I. Al-Oqily, "Development of temperature-based weather forecasting models using neural networks and fuzzy logic," *Int. J. Multimed. Ubiquitous Eng.*, vol. 9, no. 12, pp. 343–366, 2014, doi: 10.14257/ijmue.2014.9.12.31.

[35] H. I. Bulbul, M. Colak, A. Colak, and S. Bulbul, "Special session 1: Public awareness and education for renewable energy and systems," in *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, p. 12, doi: 10.1109/ICRERA.2017.8191076.

[36] M. Biswas, T. Dhoom, and S. Barua, "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data," *Int. J. Comput. Appl.*, vol. 182, no. 34, pp. 20–24, 2018, doi: 10.5120/ijca2018918265.

[37] W. Omran, M. Kazerani, and M. M. A. Salama, "Investigation of Methods for Reduction of Power Fluctuations Generated From Large Grid-Connected Photovoltaic Systems," *Energy Conversion, IEEE Trans.*, vol. 26, pp. 318–327, Apr. 2011, doi: 10.1109/TEC.2010.2062515.

[38] Y. Chang, K. Mu, Y. Leu, and L. Chou, "Solar Irradiance Forecasting Using Intelligent Technology," no. 10, pp. 15–19, 2017.

[39] R. Mueller, T. Behrendt, A. Hammer, and A. Kemper, "A New Algorithm for the Satellite-Based Retrieval of Solar Surface Irradiance in Spectral Bands," pp. 622–647, 2012, doi: 10.3390/rs4030622.

[40] T. N. Project, "Wind at A Glance," pp. 47–50, 2018.

[41] M. Afzali, A. Afzali, and G. Zahedi, "Ambient Air Temperature Forecasting Using Artificial Neural Network Approach," *4th Int. Conf.*, vol. 19, pp. 176–180, 2011, Accessed on: July 27, 2020. [Online]. Available: http://www.ipcbee.com/vol19/34-ICECS2011R30008.pdf.

[42] K. A. Baharin, H. A. Rahman, M. Y. Hassan, and C. K. Gan, "Hourly irradiance forecasting in Malaysia using support vector machine," in *2014 IEEE Conference on Energy Conversion (CENCON)*, 2014, pp. 185–190, doi: 10.1109/CENCON.2014.6967499.

[43] K. Albran, "What Are Regression Models? A model with a single independent variable is called a simple regression model. Multiple regression refers to a model with one dependent and two or more independent variables," 2016, Accessed on: September 27, 2020. [Online]. Available: http://cpdftraining.org/downloads/Levenbach_Causal2017.pdf.

[44] M. T. Alodat and M. K. Shakhatreh, "Gaussian process regression with skewed errors," *J. Comput. Appl. Math.*, vol. 370, no. December, p. 112665, 2020, doi: 10.1016/j.cam.2019.112665.

[45] C. K. I. W. C. E. Rasmussen, *Gaussian Processes for Machine Learning*, 2006 Massa., vol. 7, no. 5. 2006.

[46] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016, doi: 10.21275/v5i4.nov162954.

[47] Xin, Yang Kong, Lingshuang Liu, Zhi Chen, Yuling Li, Yanmiao Zhu, Hongliang Gao, Mingcheng Hou, Haixia Wang, Chunhua, "Machine Learning and Deep Learning Methods for Cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018, doi: 10.1109/ACCESS.2018.2836950.

[48] A. Jahan, F. Mustapha, S. M. Sapuan, M. Y. Ismail, and M. Bahraminasab, "A framework for weighting of criteria in ranking stage of material selection process," *Int. J. Adv. Manuf. Technol.*, vol. 58, no. 1–4, pp. 411–420, 2012, doi: 10.1007/s00170-011-3366-7.

[49] E. Roszkowska, "Rank Ordering Criteria Weighting Methods – a Comparative Overview," *Optimum. Stud. Ekon.*, vol. 5, no. 5(65), pp. 14–33, 2013, doi: 10.15290/ose.2013.05.65.02.

[50] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[51] G. Brassington, "Mean absolute error and root mean square error: which is the better metric for assessing model performance?," *Eur. Geosci. Union Gen. Assem.*, vol. 19, pp. EGU2017-3574, 2017, Accessed on: November 13, 2020. [Online]. Available: http://meetingorganizer.copernicus.org/EGU2017/EGU2017-3574.pdf%0Ahttp://meetingorganizer.copernicus.org/EGU2017/meetingprogramme.

[52] M. Yesilbudak, M. Colak, R. Bayindir, and H. I. Bulbul, "Very-short term modeling of global solar radiation and air temperature data using curve fitting methods," in *2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, 2017, pp. 1144–1148, doi: 10.1109/ICRERA.2017.8191233.

[53] S. Boslaugh, "Coefficient of Determination," *Encycl. Epidemiol.*, 2012, doi: 10.4135/9781412953948.n78.

[54] E. A. H. Elamir, "Mean absolute deviation about median as a tool of explanatory data analysis," *Lect. Notes Eng. Comput. Sci.*, vol. 2197, no. July 2012, pp. 324–329, 2012.

[55] U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012002.

[56] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, pp. 2856–2870, Nov. 2011, doi: 10.1016/j.solener.2011.08.027.