PHOTOVOLTAIC (PV) SYSTEM OUTPUT POWER FORECASTING USING SUPPORT VECTOR MACHINES (SVM) TECHNIQUE

WONG WAI LEONG

BACHELOR OF ELECTRICAL ENGINEERING WITH HONORS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2019

PHOTOVOLTAIC (PV) SYSTEM OUTPUT POWER FORECASTING USING SUPPORT VECTOR MACHINES (SVM) TECHNIQUE”

WONG WAI LEONG

A report submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Electrical Engineering with Honours

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2019

DECLARATION

I declare that this thesis entitled "PHOTOVOLTAIC (PV) SYSTEM OUTPUT POWER FORECASTING USING SUPPORT VECTOR MACHINES (SVM) TECHNIQUE" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : ....

Date : ....

APPROVAL

I hereby declare that I have checked this report entitled “PHOTOVOLTAIC (PV) SYSTEM OUTPUT POWER FORECASTING USING SUPPORT VECTOR MACHINES (SVM) TECHNIQUE” and in my opinion, this thesis it complies the partial fulfillment for awarding the award of the degree of Bachelor of Electrical Engineering with Honours

Signature : ....

Supervisor Name : ....

Date : ....

DEDICATIONS

To my beloved mother and father

ACKNOWLEDGEMENTS

My Final Year Project is in smooth progress with the help and contribution from many of others. Therefore, I would like to show my appreciation and extend my sincere thanks to all of them. First of all, I would like to thank the Almighty God for granting me health and knowledge so that I could have finished Final Year Project.

Next, I would like to express my very great appreciation to my supervisor, En. Kyairul Azmi Bin Baharin for the support and guidance which can help me to finish my project on time. I very appreciate that he taught me a lot of knowledge about my project and lead me to the path of success.

Furthermore, I would like to offer my special thanks to all the lecturers for willing to support and guide me by teaching me a lot of knowledge and techniques to overcome the challenge and complete my Final Year Project.

Last but not least, I wish to thank my family and friends that support me and lend me a hand when I seeking for help. I appreciate that there are family and many friends who always got my back throughout the whole progress of Final Year Project.

ABSTRACT

The use of Solar Photovoltaic (PV) system for power generation have expanded rapidly for the past few years. However, the growth of the solar PV system is also causing problems for the management of the power distribution as the operator have to always maintain the stability of the power grid between power generation and power distribution. Therefore, solar power output forecasting have become an important task to focus on to overcome the problems of using solar PV system for power generation. A solar power output prediction model is developed in this project to predict the day ahead hourly power output by using the Support Vector Machines (SVM) method. The prediction model is developed based on the data and module technology of the Solar Lab of Faculty of Electrical Engineering (FKE) in University Teknikal Malaysia Melaka (UTeM). The prediction model is designed by training the prediction model using local data with regression learner application in MATLAB software version R2017b. The results indicate that using SVM model to forecast solar power output is valid and the accuracy of the prediction is satisfied. The predictor variables used to trained the predictive model is analyzed. Irradiance and Module Temperature are the most dominant variables that will give a large impact to the accuracy of the trained predictive model to perform day ahead solar PV power output forecating.

ABSTRAK

Penggunaan sistem Solar Fotovolta (PV) untuk penjanaan kuasa telah berkembang dengan pesat sejak beberapa tahun yang lalu. Walau bagaimanapun, pertumbuhan sistem PV solar juga menyebabkan masalah pengurusan pengedaran kuasa kerana pengendali harus sentiasa mengekalkan kestabilan grid kuasa antara penjanaan kuasa dan pengagihan kuasa. Oleh itu, tumpuan kepada peramalan hasil tenaga solar telah menjadi objektif yang penting untuk mengatasi masalah menggunakan sistem PV solar untuk penjanaan kuasa. Model ramalan hasil tenaga solar telah dibina dalam projek ini untuk meramal penghasilan kuasa setiap jam sehari ke depan dengan hanya menggunakan kaedah *Support Vector Machines* (SVM). Model ramalan dibina berdasarkan data dan modul teknologi yang didapati di Makmal Solar Fakulti Kejuruteraan Elektrik (FKE) di Universiti Teknikal Malaysia Melaka (UTeM). Model ramalan direka bentuk dengan melatihnya menggunakan data tempatan dengan aplikasi '*Regression Learner*' dalam MATLAB versi R2017b. Kesimpulannya, keputusan yang didapati adalah dengan menggunakan model SVM untuk ramalan penghasilan kuasa solar adalah sah dan ketepatan ramalan adalah memuaskan. Pembolehubah peramalan yang digunakan untuk melatih model ramalan dianalisa. Keamatan cahaya matahari dan suhu modul adalah pembolehubah yang paling dominan yang akan memberi impak yang terbesar kepada ketepatan model ramalan yang telah dilatih untuk melaksanakan ramalan hasil tenaga solar bagi sehari ke depan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

PV	-	Photovoltaics
GHI	-	Global Horizontal Irradiance
SVM	-	Support Vector Machines
SVR	-	Support Vector Regression
NNs	-	Neural Networks
RFs	-	Random Forests
HMM	-	Hidden Markov Model
GBR	-	Gradient Boosted Regression
PPF	-	Past-Predicts-Future
KNN	-	K Nearest Neighbor
AR	-	Autoregressive
MLR	-	Multi Linear Regression
ANN	-	Artificial Neural Network
GWC	-	Generalized Weather Classes
WSPR	-	Weather Statuses Pattern Recognition
LS	-	Least Square
RBFNN	-	Radial Basis Function Neural Network
ARIMA	-	Autoregressive Integrated Moving Average
SARIMA	-	Seasonal Autoregressive Integrated Moving Average
GRNN	-	Generalized Regression Neural Network
NWP	-	Numerical Weather Predictiion
ESDLS	-	Evolutionary Seasonal Decomposition Least Square
ELM	-	Extreme Machine Learning
WSVM	-	Weighted Support Machines
LSSVM	-	Least-Square Support Vector Machines
FFA	-	Firefly Algorithm
GP	-	Genetic Programming
LVQ	-	Learning Vector Quantization
WD	-	Wavelet Decomposition
WT	-	Wavelet Transform
GA	-	Genetic Algorithm
PSO	-	Particle Swarm Optimization
SOM	-	Self-organizing Map
MSE	-	Mean Square Error
RMSE	-	Root Mean Square Error
MAE	-	Mean Absolute Error
CSI	-	Clear Sky Index
RH	-	Relative Humidity

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Motivation

Nowadays, large scale power plant that generate electricity by renewable energy has been utilized all around the world. There are many types of renewable energy used to generate electricity in power plant and one of the most common used renewable energy for power generation for the power grid are solar energy as the solar photovoltaic (PV) system had growth rapidly in worldwide. Based on the review of the cumulative installed capacity for solar photovoltaic system, by the end of 2017 the cumulative installed capacity had reached 398 Gigawatts (GW) in whole world [1].

On the other hand, there are also solar PV systems used to generate electricity in Malaysia. Based on Sustainable Energy Development Authority Malaysia (SEDA), by the end of 2017 the cumulative installed capacity of solar PV systems had achieved around 380 Megawatts (MW) [2]. This shows that the solar PV systems also growth rapidly in Malaysia as Malaysia is a country that located at the equator causing Malaysia geographically exposed to the sunlight for quite a long period of time in a year.

However, using solar energy as the source of electricity creates quite some problems causing the management of the electricity become difficult. Therefore, solar power forecasting is used to improve the energy management system for grid planning, scheduling, maintenance and the balance between power generation and power consumption [3]. Solar forecasting is to predict or estimate the electricity power output generated in the future and this can be performed in several method and the method used in this project is machine learning method.

The machine learning method used in this project is Support Vector Machines (SVM) method. SVM is a supervised machine learning method that can used to solve classification or regression problems. Therefore, a prediction model that can predict the solar power output which coincided with the solar panel module technology of Malaysia is to build in this project through simulation by using MATLAB R2017b. The prediction model is trained with the data of predictor variables that obtained from local solar PV system and this prediction model should predict an accurate day ahead hourly power output based on give input data.

1.2 Problem Statement

The power output production of a photovoltaic (PV) system is unstable as the power output is mainly affected by the global solar irradiance and other factors such as weather condition and PV module characteristics. Therefore, the ability to forecast the power output precisely has become the main problem to overcome so that fluctuations can be anticipated before it occurs and necessary mitigation measures can be executed.

The main problem in this project is that currently there is no adequate investigation on the practicality of forecasting power output in tropical climates such as in Malaysia. PV performance depends strongly on local weather and environment. Predictive models used in other countries cannot be applied directly due to this limitation.

How to prioritize which parameter that must be used for prediction is one of the main problem too. This is important in designing a predictive model that corresponds with Malaysia PV systems for solar power forecasting.

There are a multitude of available machine learning methods from to choose from. One of them, SVM has been proven to perform well in solar power output forecasting in many previous researches. However, the performance of the prediction depends on the type of data and system information available. Hence, there is a need to evaluate the validity of using SVM to build a prediction model using local data.

1.3 Objective

There are a few objectives have to be achieved during the analysis of the validation of Solar PV Forecasting by SVM predictive model. Below shows the objectives of this project.

i) To analyze the potential of Support Vector Machine (SVM) in PV forecasting.

ii) To predict the power output of the PV system in the solar lab at FKE UTeM.

iii) To determine the validity and necessity of the predictor variables to train the predictive model.

1.4 Scope and Limitation

The analysis of the solar PV forecasting with machine learning is a big field to work into. Therefore in this project, there are some scope and limitation that are made to narrow down the scope of work as shown below.

a) To perform solar power forecasting by using Support Vector Machines (SVM) algorithms.

b) Perform simulation by using the regression learner in MATLAB software.

c) Using the data collected from the PV system of FKE in 2016 to train a predictive model with SVM.

d) The input data is restricted to two parts. The first part is the hourly power output from PV and module temperature. This data is obtained from the 6 kWp monocrystalline system in FKE. The second part comprise meteorological parameters such as irradiance, ambient temperature, relative humidity, wind speed and amount of rain.

CHAPTER 2

LITERATURE REVIEW

2.1 Solar Forecasting

As the solar photovoltaic (PV) technology had growth rapidly these few years, the photovoltaic has become the main target in the electricity market and this cause the solar photovoltaic power plant become more popular.

2.2 Machine Learning

Machine learning is an artificial intelligence that classified in the subfield of computer science that can learn and solve problems which are impossible to be represented by explicit algorithms or equations [4]. Machine learning algorithms can analyze and interpret the input and output data and determined the relation between the input data to perform classification or prediction based on given data. Machine learning can be classified into two categories which are supervised learning and unsupervised learning.

2.2.1 Unsupervised Learning

Unsupervised learning is a learning model that group and interpret the dataset consist only input data to determine the hidden pattern in the data. The main unsupervised learning technique is clustering, which group the data based on the common pattern or characteristics. Clustering also divided into two categories which are hard clustering and soft clustering. Hard clustering is that the every of the data point is either completely belong to a cluster or vice versa while soft clustering is to assign data point into more than one cluster with a probability [4]. Table 2.1 show the common hard clustering and soft clustering algorithms.

Table 2.1 Algorithms of Hard Clustering and Soft Clustering

Hard Clustering	**Soft Clustering**
k-Means	Fuzzy c-Means
k-Medoids	Gaussian Mixture Model
Hierarchical Clustering	
Self-Organizing Map	

2.2.2 Supervised Learning

Supervised learning is a model that required the known information of both input data and desired output data to learn and find a general relation between the inputs and outputs [5]. In supervised learning, the training data is obtained from a set data of training examples. The supervised learning model will analyze the training data and developed a predictive model. Supervised learning can classified into two big categories which are classification and regression. Classification techniques is used to predict discrete responses while regression techniques is used to predict continuous responses. Table 2.2 Show the algorithms that categorized under classification and regression.

Table 2.2 Algorithms for Classification and Regression

Classification	**Regression**
Support Vector Machine	Linear Regression
Discriminant Analysis	Generalized Linear Model
Naive Bayes	Support Vector Regression
Nearest Neighbor	Gaussian Process Regression model
	Ensemble Methods
	Decision Trees
	Neural Networks

2.2.3 Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised learning model under classification categories that can analyzed the input data and determined the linear decision boundary (hyperplane) to classify all data of one class from another class [6], [7]. After the data sets is separated linearly, the best hyperplane is achieved by finding the largest margin for the nearest data point between two classes and the classification model is achieved. The SVM classification algorithms can be modified and extended with more function in machine learning. The use of SVM for regression is first introduced by Vladimir N. Vapnik in 1996 and this method is known as Support Vector Regression (SVR)[5]. Figure 2.1 shows how the SVM separated the non-linear input data into two category by hyperplane.

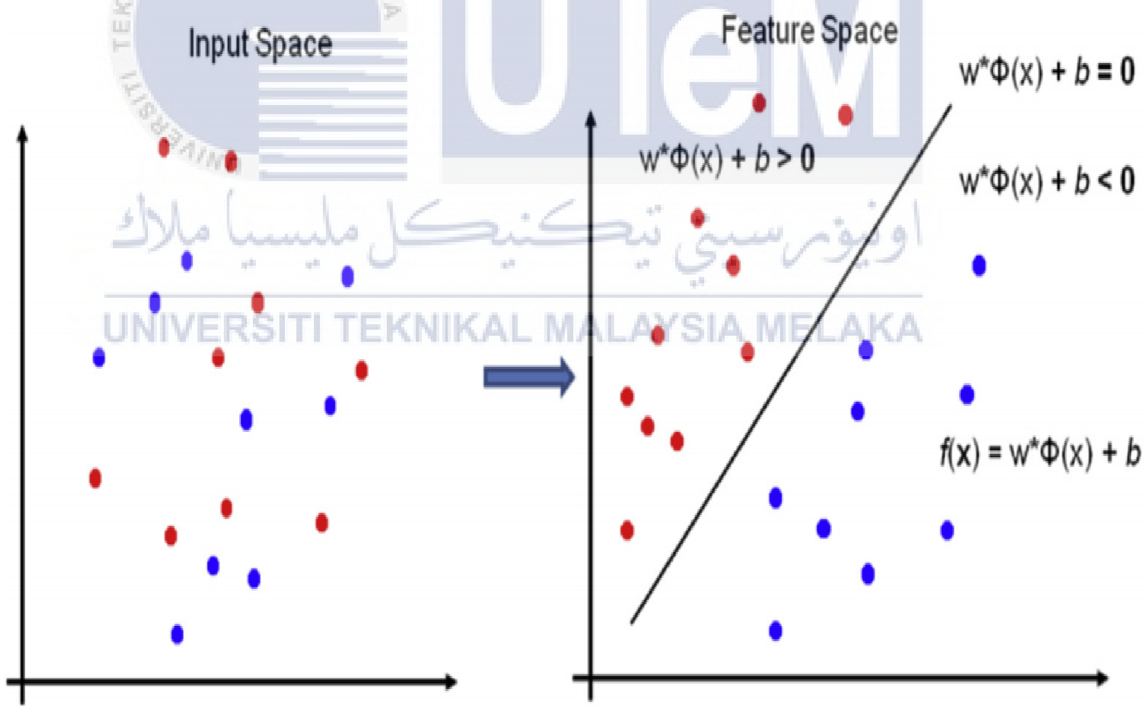

Figure 2.1 Classification of Non-Linear Input by Support Vector Machine [6]

2.2.4 Support Vector Regression (SVR)

The Support Vector Regression (SVR) is the predictive model that solve regression problem based on SVM algorithms [5]. SVR show an excellent performance in non-linear regression and time series prediction tasks [5], [7], [8]. The SVR developed a model that depend and learn on a subset of training data with a large number of predictor variables that allow the SVR model to determine the non-linear decision boundaries and predict a continuous response [8]. The purpose of using SVR algorithms is same as SVM that is minimize the error and maximize the hyperplane margin in term of prediction. Therefore, SVR algorithms is chosen as the method to use in this project as SVR is a suitable method to apply in solar power forecasting to predict the power output of the power plant.

2.3 Solar Forecasting Predictive Model

There are a lot of existed machine learning based predictive model for the power output or solar irradiance forecasting such as Neural Networks (NNs), Random Forests (RFs), Hidden Markov Model (HMM),Gradient Boosted Regression (GBR), Support Vector Machine (SVR) and etc. On the other hand, there are also some other predictive model based on Extreme Machine Learning (ELM) and Quantile Regression Forests as machine learning tools. From the previous research, there are comparison done by the researcher to determine which method is the better predictive model for solar forecasting.

2.3.1 Comparision between SVR and other Predictive Model

As mentioned above, Support Vector Regression (SVR) is one of the most common solar forecasting predictive model for power output or solar irradiance prediction. However, from the previous research we know that there are some other ready exist predictive model too. Therefore, it is important for us to determine which method give the best result in solar forecasting. In this part, SVR method is used to compare with other method based on other researches to analyze which give the more accurate solar forecasting result as in Table 2.3.

Table 2.3 Comparison of SVR with other Prediction Model

Ref	Objective	Method To Compare	Results & Comment
[6]	The objective of this paper is to develop a short-term solar irradiance forecasting algorithms. The forecasting algorithms is designed to predict future 5-30min solar irradiance under different weather condition.	Support Vector Machines Regression and Hidden Markov Model is used in this paper as short-term solar irradiance forecasting model.	Both method give positive results that can give accurate forecasting at different weather condition for the forecasting time of 5 minutes, 15 minutes and 30 minutes.
[7]	The purpose of this paper is to develop a predicting model for solar power generation from weather forecasts based on machine learning.	The main methods focus in this paper are Linear Least Square Regression and SVR. The developed predictive model also used to compare with existing models such as past-predicts-future model (PPS) and a simple model name *cloudy* that done the prediction based on sky condition.	By comparing the SVM-RBF with existing model, SVM-RBF with four dimensions is 27% more accurate than the simple cloudy model and 51% more accurate than the PPF model.
[8]	The task of forecasting the power output of the photovoltaic system for forecasting horizon of 5-60min is done in this paper.	The prediction model that used in this paper is an ensemble of Neural Networks (NNs) developed by the researchers and SVR.	The ensemble NNs model outperformed SVR for the forecasting horizon of 5 to 60 minutes can be predicted accurately.
[9]	In this paper, a day-ahead solar power production forecast is developed by considering different methods and inputs.	Support Vector Machines Regression (SVR) and Random Forests (RFs) is used in this paper to forecast the solar power output.	SVR method give the best results as the SVR method yield a minimum RMSE which mean the accuracy of the solar forecasting is the greatest.
[10]	A one-day-ahead PV power output forecasting	Support Vector Machines is applied to regression	Four model of SVM are set up and the

	model for a single station is derived based on the weather forecasting data.	prediction based on four different weather condition as model.	results find that the SVM models are effective and promising for grid-connected PV systems.
[11]	Daily and monthly global solar irradiance prediction on horizontal surface of Algeria is focus in this paper.	An application of SVM-based models for the prediction is used in this paper.	The SVM is effective and show good accuracy with only few simple parameter required.
[12]	The global horizontal irradiance for horizon of 1 hour is predicted with two types of prediction model.	SVR-based prediction model and RFs prediction model is used and compared along with the classic linear regression and kNN in this paper.	Both method are valid and show significant improvement for the forecasting compared to the previous prediction model.
[13]	The purpose of this paper is to compare SVR for PV power output forecasting with a physical modeling approach for 1 hour ahead forecasting.	A SVR is developed with a large number of data from PV data measurement, numerical weather prediction and satellite-based cloud motion vector forecast. This model is then compared with a physical modeling approach which is statistically enhanced prediction model.	Both model give the almost same prediction with three combine input.
[14]	The main purpose in this paper is to propose a few supervised machine learning method for hour ahead Global Horizontal solar Irradiance (GHI) forecast.	The method used in this paper included NNs, Gaussian processes and SVM and also a simple linear autoregressive model.	The performance of this three supervised learning method are considered equivalent for hour ahead forecasting but this three method still outperformed the linear AR.
[15]	A day-ahead short-term solar pv power forecasting is performed using prediction models based on weather classification model.	SVM and KNN is used in this paper to compare which give a better prediction based on weather classification model.	SVM give better results for small sample data while KNN performed better with larger number of sample.
[16]	This paper proposed to develop a prediction	SVR method is used by giving twelve weather	The SVR model is outperformed the

	model by using new predictor variables to predict 24 hours ahead over entire year.	variables and adding new variables which are heat index and speed of wind. The SVR model is then compared with ANN and multi linear regression (MLR) model.	ANN and MLR. By adding the two new variable, the accuracy of the model is influenced as it might be improved or become less accurate.
[17]	This paper aim to develop a forecasting model to predict one day ahead power output with an interval of 15 minutes.	The forecasting model is built by using SVR method.	The quality of this model is good as the prediction are accurate.
[18]	This paper intend to build an accurate short term solar irradiance forecasting model.	SVR method is used with the utilized of clearness index conversion, ramp-down event forecasting and solar irradiance refinement procedures.	This SVR model give a highly accurate prediction as the used of the error corrected function enhanced the accuracy of the model.
[19]	The aim of this paper is to carry out an hourly power output forecast for 1 year.	The SVR model with numerically predicted cloudiness is used in this paper.	The model with the used of cloudiness improved the performance of SVR model.
[20]	This paper propose to forecast hourly solar power output.	The method used in this paper is multi input SVR and it is compared with analytical method.	SVR show a slightly higher accuracy compared to analytical method.
[21]	This paper proposed to forecast hour ahead solar PV power output by using SVR.	SVR method is then compared with Polynomial Regression and Lasso.	SVR outperformed the other two models with a better accuracy.
[22]	This paper aim to develop a global solar radiation forecasting model with SVR.	The SVR is develop by training with measured air temperature and relative humidity.	The SVR is capable to give accurate prediction based on the predictor variables.
[23]	A weather based forecasting model is proposed in this paper to perform a short-term PV power forecasting for time interval of 24 hours.	A solar irradiance feature extraction and SVM based weather statuses pattern recognition (WSPR) model is built. The model also included four generalized weather classes (GWC) for a better performance.	In conclusion, this model performance effectively for PV power forecasting although there are some missing data of the weather type of historical data.

2.3.2 Comparison between modified SVR and other Predictive Model

Modified Support Vector Machines (SVM) model for regression problems is one of the method that can predict the solar PV power output. The SVM model is modified into different predictive model such as Weighted Support Machines (WSVM) and Least-Square Support Vector Machines (LSSVM) that consider the different samples that influences the prediction [24], [25]. In this part of view, based on some others researcher they find that the modified SVM model also give better results than other solar power prediction model.

Table 2.4 Comparison of Modified SVR with other Prediction Model

Ref	Objective	Method To Compare	Results & Comment
[24]	To perform short term PV power forecast with data of 5 most similar days.	Weighted Support Vector Machines is used in this paper and compared with ANN method.	WSVM is validate and it is more efficient than the ANN.
[25]	The purpose of this paper is to build a short term solar power prediction with the historical data of atmospheric transmissivity.	The Least-square (LS) Support Vector Machines (SVM) is used to build the prediction model and it is compared with reference autoregressive model and the radial basis function neural network (RBFNN).	It is found out that LSSVM is outperformed reference autoregressive (AR) model and better than radial basis function neural network (RBFNN).
[26]	This paper proposed to forecast monthly solar power output.	The model is developed by using an evolutionary seasonal decomposition least-square support vector regression (ESDLS-SVR).	This model give a better performance compared to autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), generalized regression neural network (GRNN) and LS-SVR.

2.3.3 Comparison between Hybrid SVR and other Predictive Model

A hybrid Support Vector Machines (SVM) is the forecasting model that combined the SVM techniques with any others ensemble learning algorithms that could benefit and improved the performance of the prediction model. The hybrid SVM is that combining the strength of each of the learning algorithms and developed a better accuracy forecasting model. The SVM can combined with many of the learning algorithms such as Particle Swarm Optimization (PSO), Random Forests (RFs), Self-organizing Map (SOP) and many others more. Below Table 2.5 shows the researchers finding on the hybrid SVR and the performance of the hybrid SVR.

Table 2.5 Comparison of Hybrid SVR with other Prediction Model

Ref	Objective	Method To Compare	Results & Comment
[27]	To improve 3-h accumulated radiation forecasts provided by Numerical Weather Prediction (NWP) system with several methods of machine learning.	Three approach of machine learning method used to improve the NWP forecast which is Support Vector Regression (SVR), Gradient Boosted Regression (GBR) and Random Forest Regression (RFR). A hybrid model is built by combining the three approach.	The three machine learning method and the hybrid model built by combining these three method can improved the solar radiation forecasting. The hybrid model performed better than the individual three model.
[28]	To examine the accuracy of solar radiation prediction based on meteorological data by using hybrid machine learning.	A new hybrid model of Support Vector Machine (SVM) with Firefly Algorithm (FFA) is used to predict the monthly mean horizontal global solar radiation.	The SVM-FFA hybrid model is validate and have a more accurate and precise prediction compared to Artificial Neural Network (ANN) and Genetic Programming (GP) model.
[29]	This paper proposed a new hybrid model for	The hybrid is built by combining two method	The hybrid model show a better

	short-term power forecasting of one hour ahead.	which are SARIMA and SVM.	performance compared to the individual model of SARIMA and SVM.
[30]	A weather-based hybrid model is built to forecast one day ahead hourly PV power output to make the management system of power grid easier.	The hybrid model is built with self-organizing map (SOM), learning vector quantization (LVQ) for classification, SVR to train the model and forecast with fuzzy inference method.	This hybrid model is outperformed the simple SVR and traditional ANN methods.
[31]	The day ahead solar power forecast is done by building an ensemble model of random forest and SVR.	The ensemble model is built by combining random forest and SVR together and its performance is compared with others SVR combined methods.	This ensemble model performed better compared to others method.
[32]	The hybrid model is built to perform PV output power forecasting for five forecasting horizons from 1 hour up to 24 hours.	The hybrid model is built by combining LS-SVM with wavelet decomposition (WD) and it is compared with traditional ANN and normal LS-SVM.	The hybrid model give a better performance on prediction compared to other methods.
[33]	This paper proposed a hybrid forecasting model for one day ahead power output forecasting.	The hybrid model is built by combining wavelet transform, particle swarm optimization and support vector machine (Hybrid WT-PSO-SVM). This method is then compared with others seven method.	This paper concluded that this hybrid model performed way better than those seven method with better accuracy.
[34]	In this paper, an optimized SVR is proposed to forecast the daily solar radiation by training the model with the data of similar day in previous year.	Two optimized SVR which are SVR optimized with hyper parameter using Genetic Algorithms (SVRGA) and SVR with Particle Swarm Optimization (SVRPSO) is used to compare with SVR.	All the model have high performance in short term forecasting and SVRPSO outperformed the other two model in this experiment.
[35]	This paper focus on the forecast of hourly solar irradiance time series using a novel hybrid model.	The novel hybrid model is designed based on self-organizing maps (SOM), SVR and particle swarm optimization (PSO).	This hybrid is found that performed better than the traditional forecasting models.

CHAPTER 3

METHODOLOGY

3.1 Project Overview

Figure 3.1 shows the flowchart of the simulation to develop a SVM predictive model with MATLAB version R2017b. First start with filtering and choosing the input data, then run the Regression Learner application in MATLAB and generate a predictive model. The predictive model is then used to perform the forecasting of the day ahead solar PV power output. The error and validation of the predictive model is determined.

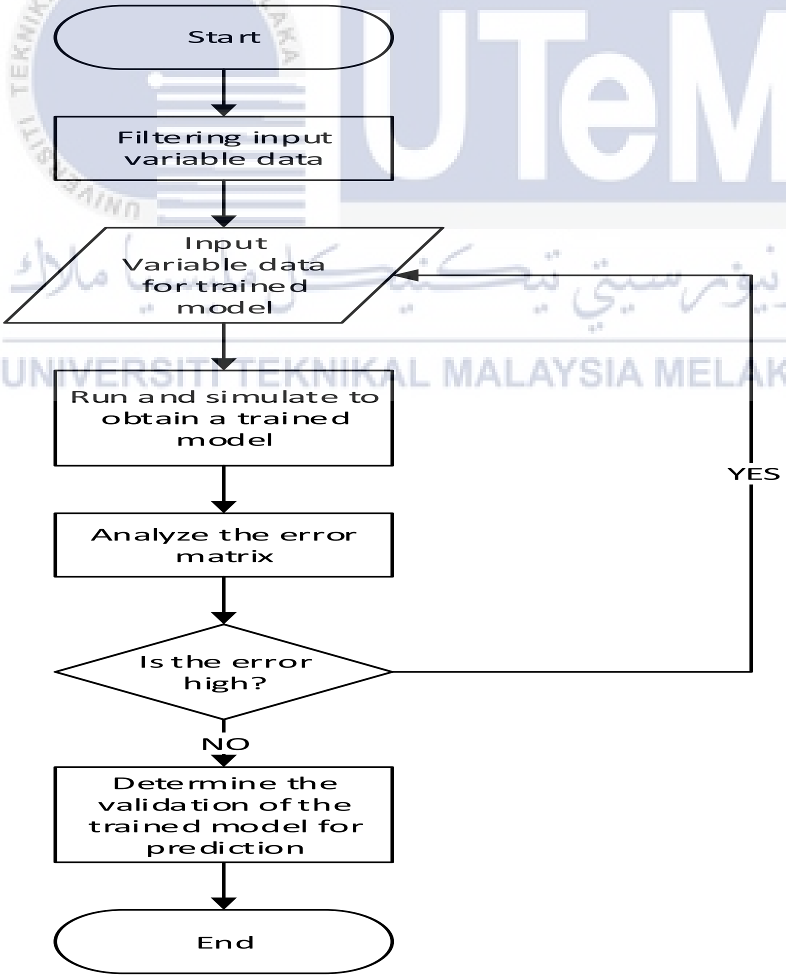

Figure 3.1 Flow Chart of Simulation

Refer to Figure 3.2, the Regression Learner in the red box highlighted is choose to start up the predictive model training window.

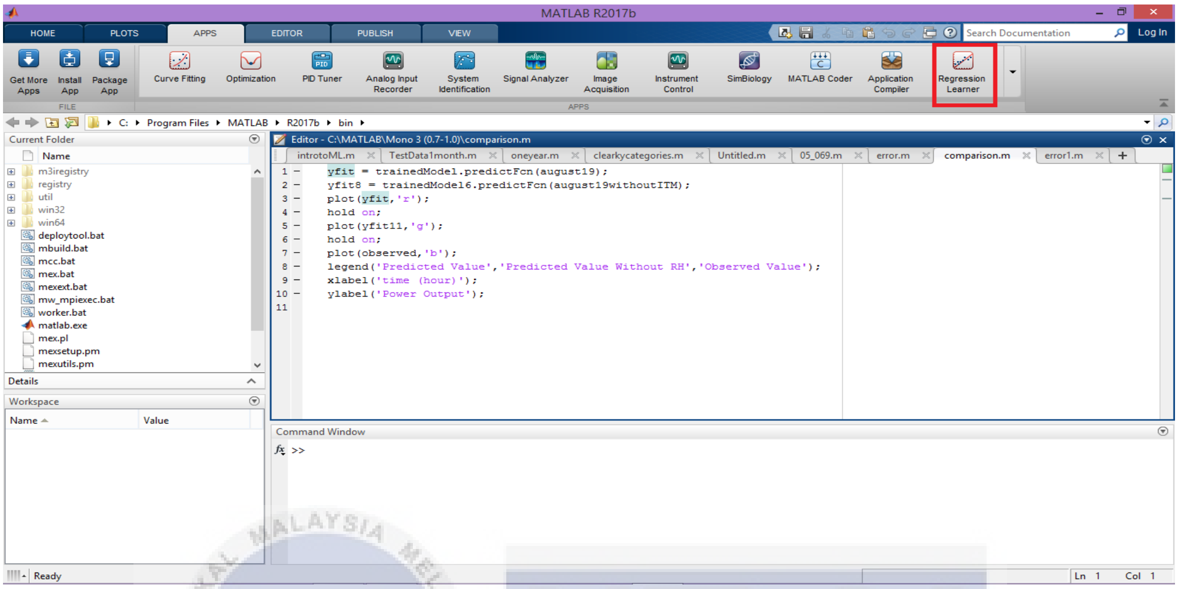

Figure 3.2 Regression Learner Application in MATLAB

Based on Figure 3.3, the Predictors and Response are choose under this section to use to train the predictive model. The cross validation is used to examine the predictive accuracy of the fitted models.

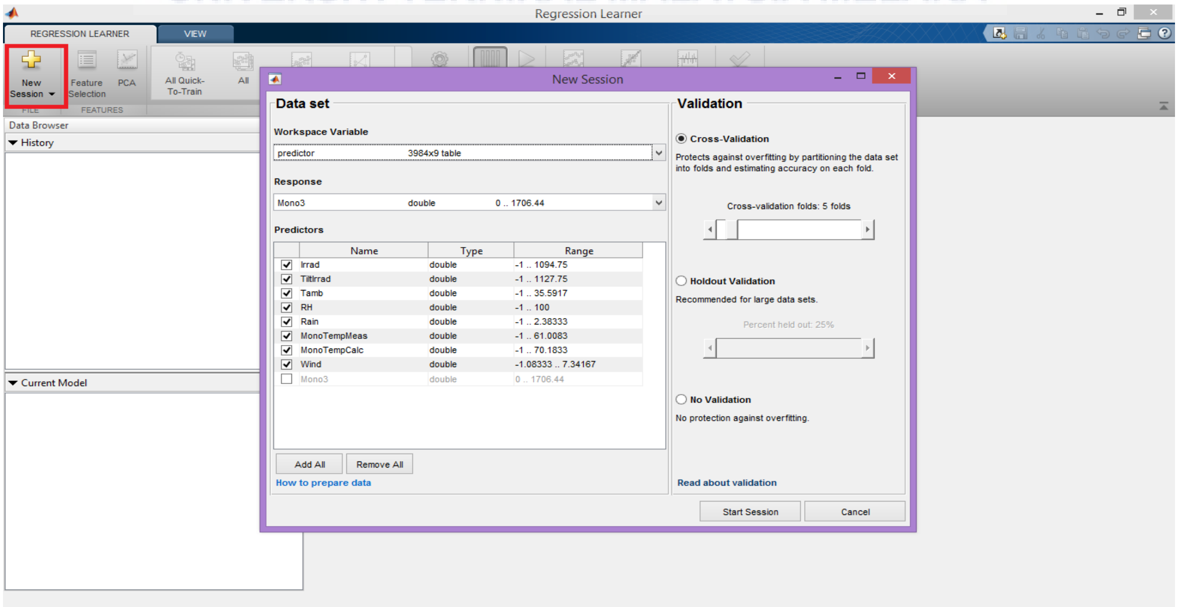

Figure 3.3 Predictor and Response of Trained Model

Refer to Figure 3.4, item 1 is the type of SVM algorithms that used to train the predictive model and the item 2 is the Error Matrix of the corresponding SVM predictive model. The Response Plot is used to justify the differences between the true value and also the predicted value as shown in item 3 and item 4. After choosing the convincing predictive model, the model needed to export to MATLAB by clicking item 6 for further prediction.

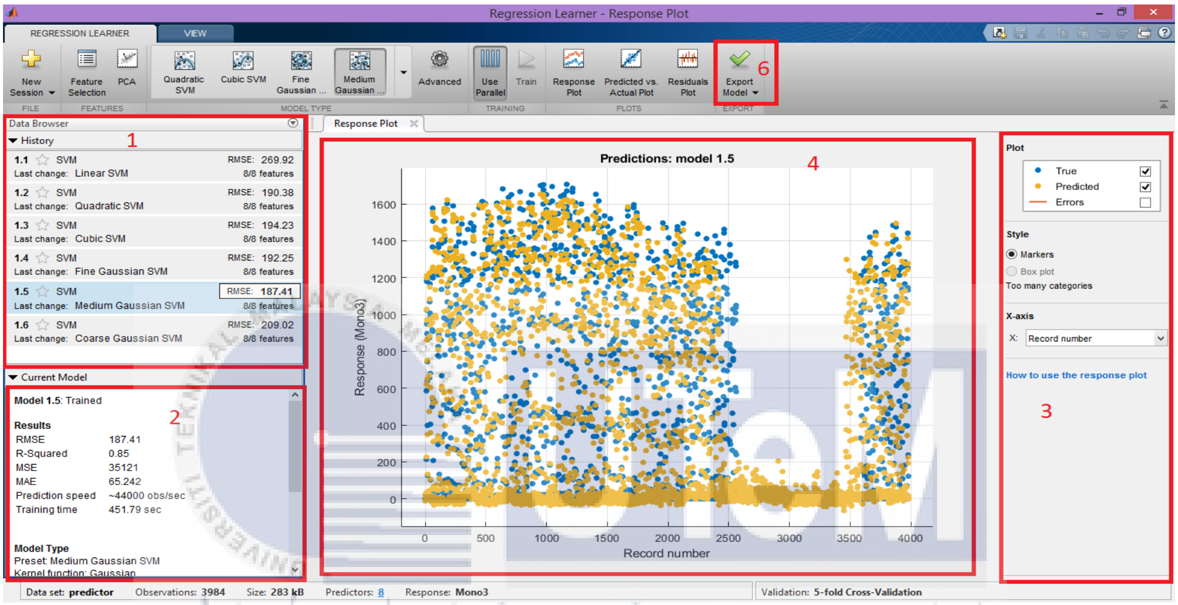

Figure 3.4 The Interface of Regression Learner

Refer to Figure 3.5, the predicted versus actual plot is shown and this plot is used to determine the accuracy of the predictive model. The predictive model is said to be accurate if the most of the observation dots are located close to the perfect prediction line.

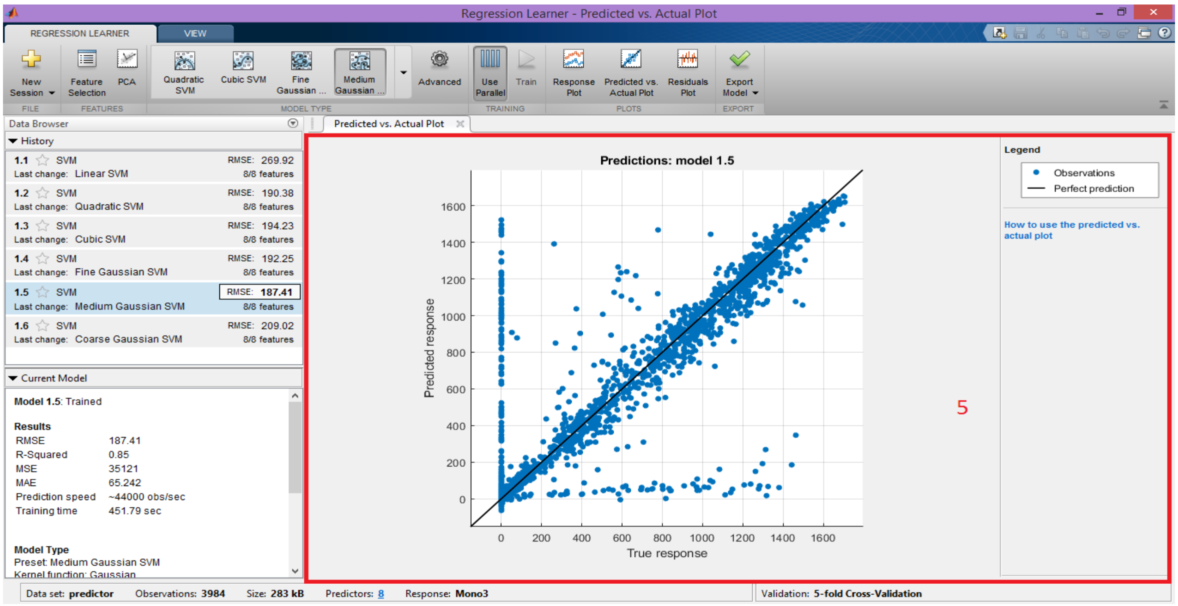

Figure 3.5 Predicted vs Actual Plot

The residual plot is used to determine how closed the predicted value is to the true response as shown in Figure 3.6. When most of the residuals are located around the x-axis, it can be concluded that the accuracy of predictive model is convincing.

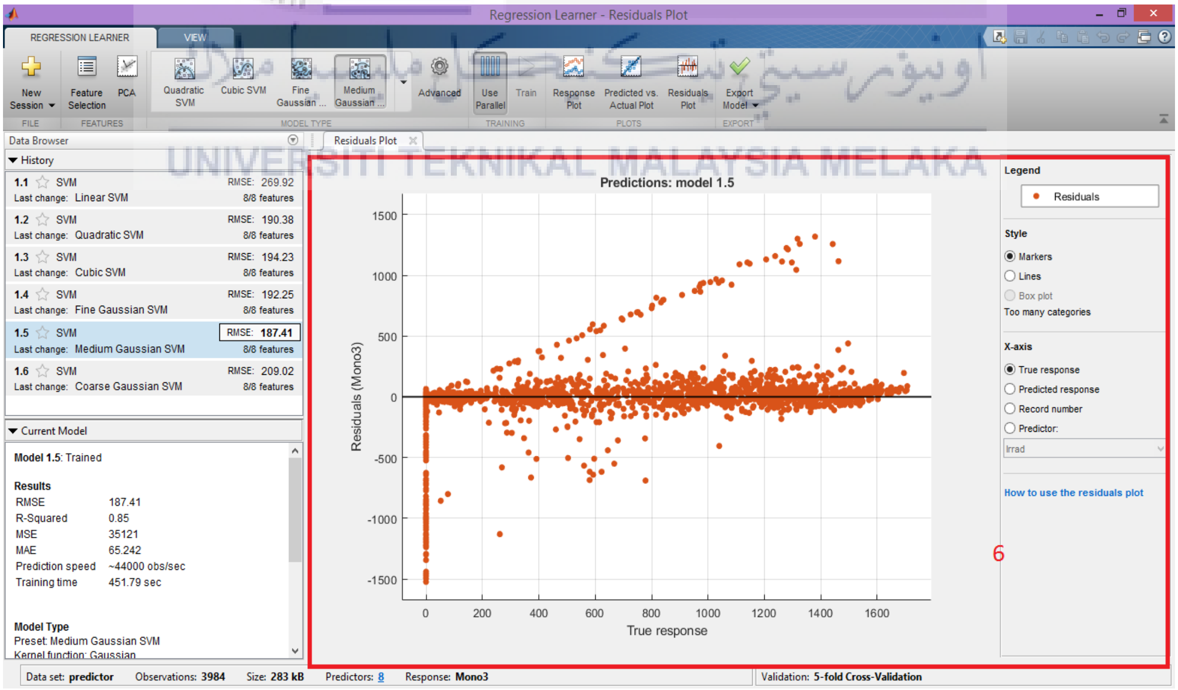

Figure 3.6 Residual Plot

The coding highlighted as shown in Figure 3.7 is run in the MATLAB to generate a set of predicted values based on the input data. By inserting the relevant input variables data, the coding can come out with a set of predicted values and the predicted values is compared with the observed value to justify the accuracy of the predictive model.

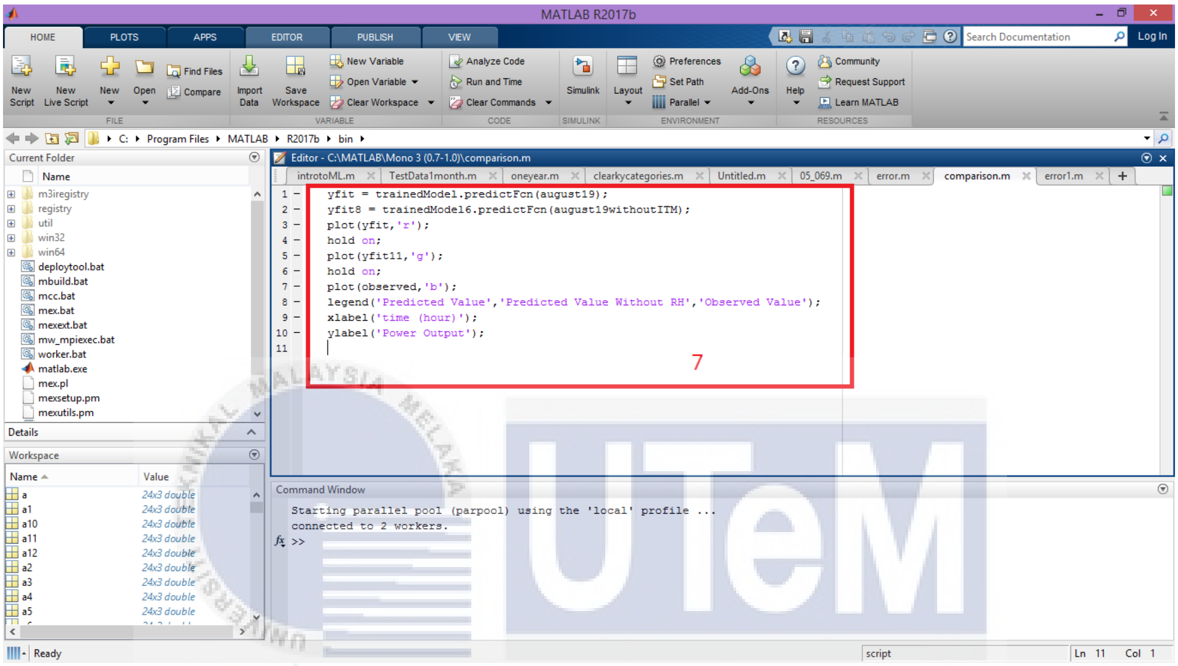

Figure 3.7 Coding used to Simulate Prediction of PV Power Output

3.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) is the main method that concerned in this project. SVM is a supervised machine learning model that used for classification analysis. The basic concept of SVM is to determine a boundary line that show a clear gap to categorize data into different classes. Therefore, the main function of SVM is to learn the relation between the input data of different categories and determine a boundary to classify the data into two different classes. Since SVM is machine learning model, it solve the classification problem by determined where the new input data are belongs to between one of the two classes which refer to the boundary line named as Hyperplane.

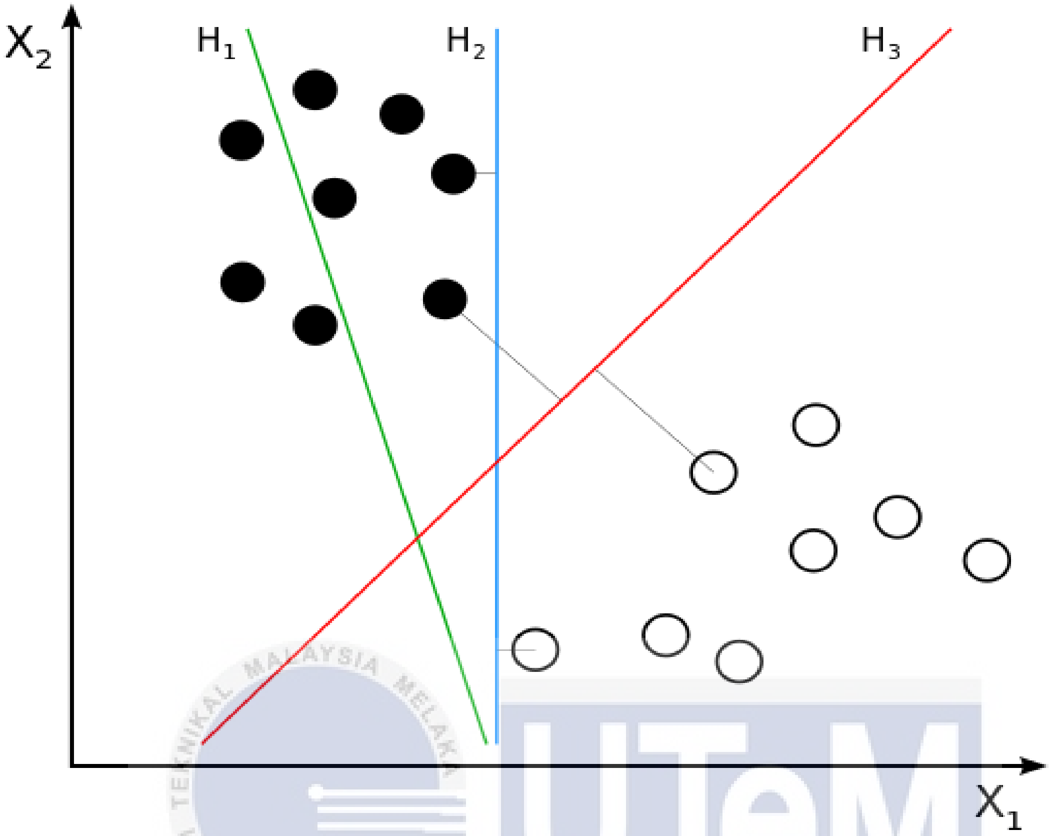

Figure 3.8 The boundary line to classify data in categories

The best hyperplane is the largest margin or separation that separate the two classes. Therefore, the best hyperplane is determined by finding the one data point from each class that is closest to the line which give the largest distance. Refer to the Figure 3.8 above, line H_1 does not classified the data perfectly while line H_3 and line H_2 do classified the data into two categories. The line H_3 is the best hyperplane compared to line H_2 as the distance of the point from two classes that closest to the line H_3 is larger than the distance of the point that closest to line H_2.

3.3 Linear Support Vector Machines (LSVM)

The most common SVM method is Linear Support Vector Machines. Linear SVM is a classifier that classified data into two categories linearly by defined a maximum-margin hyperplane. This hyperplane is achieved when the distance of the data point from each class that is closest to the margin is maximum. Refer to formula below, we can find out how linear SVM determined the margin or hyperplane with mathematical method.

Given a training dataset of number n in the form:

$$(\overrightarrow{x_1}, y_1), \ldots (\overrightarrow{x_n}, y_n) \tag{3.1}$$

Where $\overrightarrow{x_i}$ is the real vector data point and y_i is either 1 or -1 which is the boundary indicate which class the data point located at.

The hyperplane is determined when the equation below is satisfied.

$$\vec{w} \cdot \vec{x} - b = 0 \tag{3.2}$$

Where $\vec{w}$ is the unnecessary normal vector to the hyperplane.

3.3.1 Hard Margin

Hard margin is used when the data set could separate linearly. The maximum-margin hyperplane can be defined as the hyperplane at midpoint that lied in the region between two parallel hyperplane called 'margin' that classified the dataset into two classes. These parallel hyperplane is determine by the following equation below when the dataset is normalized.

$$\vec{w} \cdot \vec{x} - b = 1 \tag{3.3}$$

And

$$\vec{w} \cdot \vec{x} - b = -1 \tag{3.4}$$

The data point that above or below the boundary of 1 and -1 is classified into one class from another class.

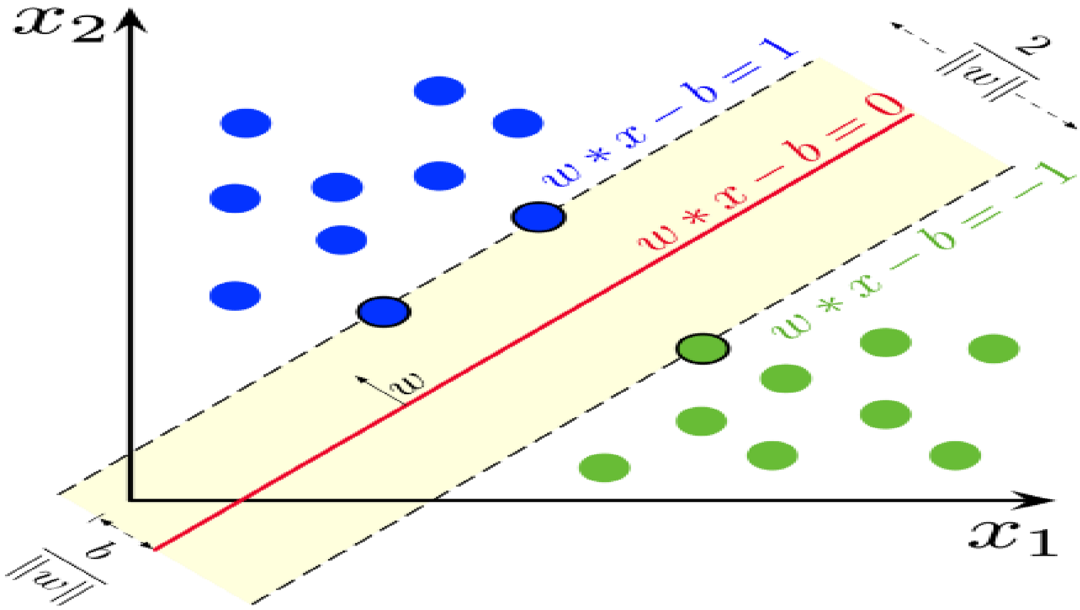

Figure 3.9 The maximum-margin hyperplane

From the above diagram can find that the maximum-margin hyperplane is lied between two parallel hyperplane obtained from equation (3.3) and equation (3.4). Also, from the Figure 3.9 can see that the distance between two parallel hyperplane is $\frac{2}{||\vec{w}||}$. Therefore, the $||\vec{w}||$ needed to be minimized in order to maximize the maximum-margin hyperplane. In order to ensure that the data point does not lied in the margin region and lied on the correct side of classification, the following restrict is added as shown below.

$$\vec{w} \cdot \vec{x} - b \geq 1, \text{ if } y_i = 1 \tag{3.5}$$

Or

$$\vec{w} \cdot \vec{x} - b \leq -1, \text{ if } y_i = -1 \tag{3.6}$$

Then the above formula can written as below,

$$y_i(\vec{w} \cdot \vec{x} - b) \geq 1, \text{ for all } 1 \leq \text{i} \leq \text{n} \tag{3.7}$$

For a conclusion, it can concluded that minimizing $\|\vec{w}\|$ subjected to $y_i(\vec{w} \cdot \vec{x} - b) \geq 1$. From the above, the $\vec{w}$ and b is used to justify the classifier and the maximum-margin hyperplane is depends on the data point $\vec{x_i}$ which is closest to the hyperplane and these $\vec{x_i}$ is called support vectors.

3.4 Support Vector Regression (SVR)

Support Vector Regression is that using Support Vector Machines classification algorithms to solve regression problem to make prediction. SVR developed a model that depend and learn on a subset of training data with a large number of predictor variables that allow the SVR model to determine the non-linear decision boundaries and predict a continuous response. Since the SVR prediction model is built by using SVM algorithms, therefore SVR can train by using parameter in SVM as shown as below.

Minimize $\frac{1}{2}\|\vec{w}\|^2$

$$\text{Subject to} \begin{cases} y_i - < w_i, x_i > - b \leq \varepsilon \\ < w_i, x_i > + b - y_i \leq \varepsilon \end{cases} \tag{3.8}$$

Where,

$\boldsymbol{x_i}$ is a training sample with target value $\boldsymbol{y_i}$.

$< \boldsymbol{w_i}, \boldsymbol{x_i} > + \boldsymbol{b}$ is the prediction for the sample.

$\boldsymbol{\varepsilon}$ is the parameter act as a threshold.

Based on the equation (3.8), it shows that the prediction value should be lied within the range of $\boldsymbol{\varepsilon}$ to secure the accuracy of prediction. The Epsilon, $\boldsymbol{\varepsilon}$ is act as a reference of accuracy for the prediction model.

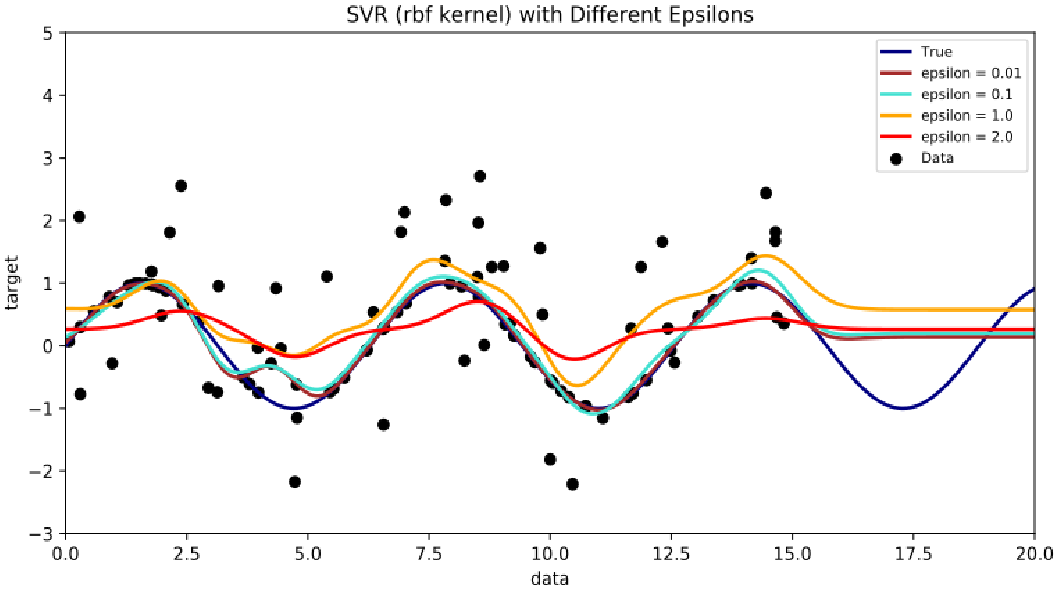

Figure 3.10 SVR Model with Different Epsilons

Based on Figure 3.10, it shows that by comparing SVR model of different values of Epsilons with the true value of data, it can found that when the value of Epsilon increases, the accuracy of the prediction value decreases. Thus, this mean that the smaller the value of Epsilon, the more sensitive the prediction to error which mean the more accurate the prediction.

3.5 Error Matrix

Error is the difference between the predicted value and true value. The error matrix are the parameters used to justify the accuracy of the solar power output forecasting which used SVR by determined the error between the predicted values and true values. From the errors, the accuracy of the predicted values can be obtained as the error reflected the accuracy of the prediction model. The error matrix as in Figure 3.11 shows the errors that used to justify the accuracy of the prediction model in this project.

Results	
RMSE	74.09
R-Squared	0.97
MSE	5489.3
MAE	46.838
Prediction speed	~2700 obs/sec
Training time	1.4525 sec

Figure 3.11 Error Matrix

The error parameter that used to determine or justify the accuracy of the forecast predicted value are Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R-squared.

3.5.1 Mean Square Error (MSE)

MSE is the measured of the average of the square of the error of predictor. MSE is to enlarge the error to make the average error more obvious to see and it is always a non-negative value that greater than zero due to randomness of the dataset. Thus, the smaller the value of error, the more accurate the predicted values are.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3.9}$$

From equation (3.9), it show that the mean square error of n data where $\hat{Y}_i$ is the predicted value and Y_i is the observed value.

3.5.2 Mean Absolute Error (MAE)

MAE is the measured of average of the absolute error for predictor. Absolute error is the measured the difference between predicted value and observed value. Since MAE is absolute value, therefore MAE is a non-negative value and the closer the value to zero, the better the forecast value.

$$|e_i| = |Y_i - X_i| \tag{3.10}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|Y_i - X_i|}{n} = \frac{\sum_{i=1}^{n}|e_i|}{n} \tag{3.11}$$

Equation (3.11) show that the absolute error of n data where Y_i is the predicted value and X_i is the real value, and $|e_i|$ is the absolute error.

3.5.3 Root Mean Square Error (RMSE)

Root Mean Square Error is the square root of the Mean Square Error. RMSE also always used to measure the accuracy of the value predicted by forecasting model. RMSE is the square root of MSE as to obtain the proportional size of the MSE value as depicted in equation (3.12). RMSE also a non-negative value and the zero value of RMSE shows that the predicted value match perfectly with the observed value. Therefore, the lower the value of RMSE indicated the better accuracy of prediction model.

$$\text{RMSE} = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{3.12}$$

3.5.4 R-squared

In regression analysis, R-squared is often used to determine how accurate the predicted value match to real value. R-squared is explained as the percentage of the variance of output variable that depends on the input variables. The R-squared is always in the form of percentage which is from 0% to 100% and generally this indicated that the larger the value of percentage the better the result of the predictive model.

Given mean of the observed data Y_i of n set of data,

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{3.13}$$

Below show that the sum of squares formula to determine R-squared:

1) The total sum of squares

$$SS_{tot} = \sum_{i} (Y_i - \bar{Y})^2 \tag{3.14}$$

2) The residual sum of squares

$$SS_{res} = \sum_{i} (Y_i - \hat{Y}_i)^2 = \sum_{i} e_i \tag{3.15}$$

Where $\hat{Y}_i$ is the predicted value and Y_i is the observed value.

Then R-squared is determined as shown below:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3.16}$$

From the above, the R-squared is determined by the ratio of residual sum of squares and total sum of squares. The SS_{res} is the variance of dependent variable and the SS_{tot} is the variance of independent variable which is the observed data. From equation (3.16), the value of R-squared is set that between 0 and 1. Therefore, if the prediction value fit the observed value perfectly, the R-squared value will be 1 or vice versa.

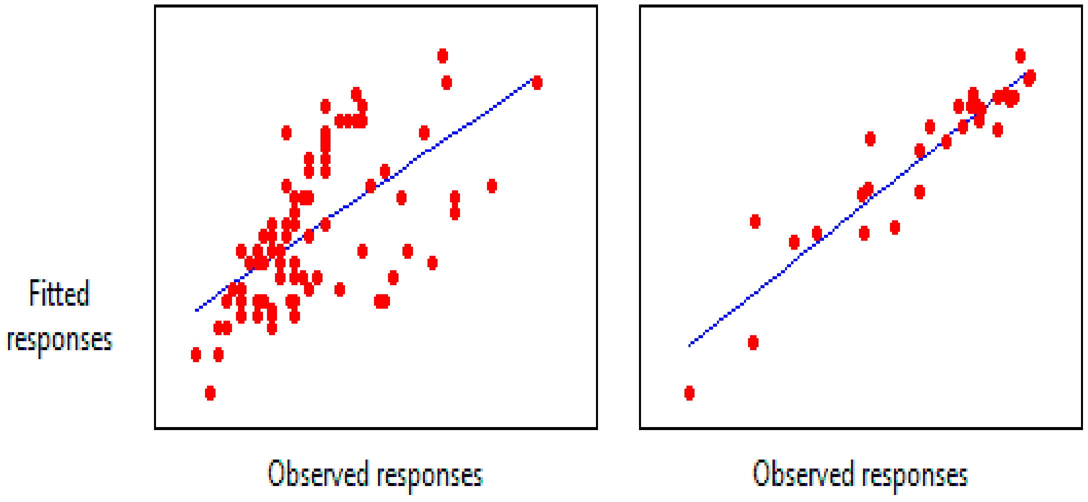

Figure 3.12 The Plot of Fitted Response versus Observed Response

Refer to Figure 3.12, the blue line is defined as the fitted regression line. The predicted data is spread around the fitted regression line and the more accurate the predicted data the closer the red data points to the line. Therefore, the model that have high accuracy will give a plot that all the data points will fall on the fitted regression line. Generally, the R-squared value is close to 1 or 100% when the prediction model is highly accurate while there is some cases that the inaccurate model give a R-squared value near to 1 or 100% too. This is because the model might lack of some important predictor variable or it is not suitable for non-linear regression or any other reasons.

3.6 Predictor Variable

To design a predictive model, the predictor variables are important variable that decide the accuracy of the model to predict an outcome. This is because to predict an output variable, the output is depends on the predictor variables and this shows that the predictor variables used must be suitable. Therefore, the predictor variables used must be precise and reasonable. In this project, the predictor variables that used to develop the predictive model are as shown below.

- Irradiance
- Tilt Irradiance
- Ambient Temperature
- Relative Humidity (RH)
- Rain
- Mono Temperature Measured (Module Temperature)
- Mono Temperature Calculated (Module Temperature)
- Wind

The above predictor variables are used to design a trained model to predict the solar power output. These predictor variables are collected and obtained from the FKE Solar Lab as the value of these variables are valid.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Timestamp	Year	Month	Day	Hour	Min	Irrad	TiltIrrad	Tamb	RH	Rain	Mono Temp Meas	Mono Temp Calc	Wind	Mono1	Mono2	Mono3
2	2016-01-01 01:00:00 MYT	2016	01	01	01	00	5.833333	5.333333	23.7	100	0.35	22.2	23.88333333	1.625	0	0	0
3	2016-01-01 02:00:00 MYT	2016	01	01	02	00	5.75	5.666667	23.61667	100	0.35	22.175	23.79166667	1.225	0	0	0
4	2016-01-01 03:00:00 MYT	2016	01	01	03	00	5.833333	5.666667	23.425	100	0.4	21.98333333	23.60833333	1.291667	0	0	0
5	2016-01-01 04:00:00 MYT	2016	01	01	04	00	5.333333	5.25	23.45	100	0.283333	22.00833333	23.58333333	1.766667	0	0	0
6	2016-01-01 05:00:00 MYT	2016	01	01	05	00	5.333333	5.25	23.5	100	0.2	21.95833333	23.63333333	0.95	0	0	0
7	2016-01-01 06:00:00 MYT	2016	01	01	06	00	6	6	23.51667	100	0.016667	22	23.71666667	0.866667	0	0	0
8	2016-01-01 07:00:00 MYT	2016	01	01	07	00	6	5.916667	23.55	100	0.466667	22	23.75	0.708333	0	0	0
9	2016-01-01 08:00:00 MYT	2016	01	01	08	00	8.083333	7.916667	23.49167	100	0.25	22.04166667	23.725	0.7	0	0	0
10	2016-01-01 09:00:00 MYT	2016	01	01	09	00	33.41667	32	23.575	100	0.083333	22.45833333	24.69166667	0.516667	27.51833	29.25833	35.34417
11	2016-01-01 10:00:00 MYT	2016	01	01	10	00	91.75	90.08333	23.75	100	0.05	23.86666667	26.85	0.675	132.9667	139.1258	142.7083
12	2016-01-01 11:00:00 MYT	2016	01	01	11	00	135.5833	135.8333	24.15	100	0.35	24.725	28.75833333	0.791667	240.2475	255.3625	256.9992
13	2016-01-01 12:00:00 MYT	2016	01	01	12	00	377.75	387	25.125	100	0	29.16666667	38.06666667	1.525	688.905	711.0417	716.2717
14	2016-01-01 13:00:00 MYT	2016	01	01	13	00	454.6667	459.6667	26.23333	99.78333	0	34.925	41.81666667	2.708333	782.2642	807.8858	815.4817
15	2016-01-01 14:00:00 MYT	2016	01	01	14	00	438.4167	435.3333	27.03333	96.54167	0	35.09166667	42.05	3.808333	751.2608	772.7683	777.8667
16	2016-01-01 15:00:00 MYT	2016	01	01	15	00	344.1667	338	26.90833	95.80833	0	32.96666667	38.69166667	3.141667	584.7208	596.7967	602.1683
17	2016-01-01 16:00:00 MYT	2016	01	01	16	00	244.8333	241.25	26.34167	99.59167	0.016667	29.44166667	34.7	3.558333	421.175	435.0292	440.95
18	2016-01-01 17:00:00 MYT	2016	01	01	17	00	243.3333	242.5	26.11667	100	0	29.51666667	34.44166667	2.325	400.3225	413.7117	419.3342
19	2016-01-01 18:00:00 MYT	2016	01	01	18	00	150.3333	150.25	26.925	100	0	28.93333333	32.03333333	2.458333	230.9742	239.0942	243.3817
20	2016-01-01 19:00:00 MYT	2016	01	01	19	00	58.41667	57.83333	27.01667	100	0	26.69166667	28.98333333	1.758333	59.0725	61.87417	65.915
21	2016-01-01 20:00:00 MYT	2016	01	01	20	00	10.41667	10.41667	26.825	100	0	24.70833333	27.15	1.241667	0.9875	1.335	1.829167
22	2016-01-01 21:00:00 MYT	2016	01	01	21	00	5.583333	5.5	26.60833	100	0	24.175	26.76666667	1.358333	0	0	0
23	2016-01-01 22:00:00 MYT	2016	01	01	22	00	5	5	26.25	100	0	23.78333333	26.35	1.758333	0	0	0
24	2016-01-01 23:00:00 MYT	2016	01	01	23	00	5	5	26.13333	100	0	23.54166667	26.23333333	2.283333	0	0	0
25	2016-01-02 00:00:00 MYT	2016	01	02	00	00	4.75	4.666667	25.60833	100	0	22.98333333	25.70833333	2.441667	0	0	0

Figure 3.13 The Datasheet of Predictor Variables and Output Variables

Figure 3.13 shows the datasheet that contained the values of both predictor variables and output variables for every hour that obtained at FKE Solar Lab in 2016.

3.7 Clear Sky Index (CSI)

Clear Sky Index is the ratio of the measured global horizontal irradiance (GHI_{actual}) over the global horizontal irradiance under clear sky condition ($GHI_{clearsky}$) which used to differentiate the clear sky condition from cloudy sky condition. The Clear Sky Index is often used to identify the sky condition of certain day by referring to the value of Clear Sky Index. Based on the ratio, the value of Clear Sky Index must be closer to 1 to justify the clear sky condition while the closer the value of Clear Sky Index to 0, the cloudy the sky condition is. Below shows the equation of ratio calculation of the Clear Sky Index.

$$\text{Clear Sky Index (CSI)} = \frac{GHI_{actual}}{GHI_{clear\ sky}} \tag{3.17}$$

The Clear Sky Index is applied to the classification of the data set that obtained from FKE Solar Lab which classified the days accordingly into three categories which are clear sky (0.70-1.00), partially cloudy (0.50-0.69) and overcast (<0.50) to ensure the trained predictive model can performed day ahead forecast the Solar PV power output more precisely in corresponding to the sky condition of the day that wanted to forecast.

Time	G clear sky	Actual G	CSI
1-Jan			
8:00	98.61	8.083333	0.081973
9:00	317.8	33.41667	0.10515
10:00	545.99	91.75	0.168043
11:00	741.29	135.5833	0.182902
12:00	876.61	377.75	0.430921
13:00	934.29	454.6667	0.486644
14:00	915.58	438.4167	0.47884
15:00	818.41	344.1667	0.420531
16:00	649.98	244.8333	0.376678
17:00	433.91	243.3333	0.560792
18:00	202.78	150.3333	0.741362
19:00	16.8	58.41667	3.477183
SUM	6552.05	2580.75	0.393884

Figure 3.14 The Data of GHI_{actual} and $GHI_{clearsky}$

Based on Figure 3.14, the clear sky hourly average irradiance is obtained from formulated data sheet while the actual hourly average irradiance is obtained from datasheet from FKE Solar Lab. The CSI value is calculated by:

$$\text{Clear Sky Index (CSI)} = \frac{2580.75}{6552.05} = 0.3939 \qquad (3.18)$$

This CSI value of 0.3939 indicate that 1st of January is categorized under the category of overcast. This process is repeated for one year data so that classification of 365 days according to clear sky index into three categories are done.

CHAPTER 4

RESULTS

4.1 Simulation to Develop Trained Model

A few predictive models trained with different predictor variables are developed by using the regression learner in the MATLAB software R2017b. The regression learner application can developed a trained prediction model that could predict an outcome by only inserting the dataset of predictor variables. In this part, the trained models only apply variants of SVM algorithm (e.g. linear, quadratic, cubic, Gaussian etc) in the regression learner. The results below shows the findings from the prediction of

4.2 Classification of Data

The data obtained from the FKE Solar Lab is classified into three categories according to clear sky index which are clear sky (0.70-1.00), partially cloudy (0.50-0.69) and overcast (<0.50). The data is classified into three categories so that the day ahead forecasting can be done more precisely according to the weather condition. Table below shows the days categories according to clear sky index.

Table 4.1 Categories of Clear Sky Index

Clear Sky Index (CSI)	**Number of days**
clear sky (0.70-1.00)	237
partially cloudy (0.50-0.69)	77

overcast (<0.50)	51

Based on Table 4.1, 70% of data of each of the categories is used to train the predictive model by using the SVM algorithms in the Regression Learner in the MATLAB software R2017b while the remaining 30% of data is used as test data. Since the quantity of data of overcast category is insufficient which will cause the trained predictive model to be inaccurate, therefore only clear sky and partially cloudy categories is focused in this project.

4.3 Baseline Error

The baseline error is set to determine whether the predictive model trained with SVM algorithm can performed accurate prediction. A model that is trained with Linear Regression algorithm is used to perform forecasting and the error of this prediction is assumed as the baseline error of the SVM trained models. Below Table 4.2 shows the baseline error for the categories of clear sky and partially cloudy.

Table 4.2 Baseline Error for Clear Sky and Partially Cloudy

CATEGORY	RMSE	MAE
Clear sky	210.81	147.25
Partially cloudy	134.27	92.08

4.4 Validation of Predictive Model trained with SVM Algorithms

The predictive model trained with SVM algorithms for clear sky category and partially cloudy category are used to perform a few forecasting for the random days chosen from the remaining 30% of test data. This is to determine the validation of the SVM trained predictive model whether the prediction done by this model can be performed accurately. Below Table 4.3 shows the results of the random days picked from the clear sky category and partially cloudy category.

Table 4.3 Validation for ClearSky and Partially Cloudy

Clear Sky Category			
Date	**Error**		
	Linear Regression	**SVM**	**Percentage of Improvement**
Sept 30		RMSE = 58.07 MAE = 38.11	RMSE: 72.45% MAE: 74.12%
Oct 18		RMSE = 55.52 MAE =37.81	RMSE: 73.66% MAE: 74.32%
Nov 15	RMSE = 210.81 MAE = 147.25	RMSE = 93.18 MAE = 63.43	RMSE: 55.80% MAE: 56.92%
Dec 3		RMSE = 54.03 MAE = 39.57	RMSE: 74.37% MAE: 73.13%
Dec 21		RMSE = 45.39 MAE = 28.02	**RMSE: 78.47%** **MAE: 80.97%**
Partially Cloudy Category			
Date	**Error**		

	Linear Regression	SVM	Percentage of Improvement
Nov 7		RMSE = 54.78 MAE = 34.63	RMSE: 59.20% MAE: 62.39%
Nov 28		RMSE = 38.36 MAE = 25.05	**RMSE: 71.43% MAE: 72.80%**
Dec 8	RMSE = 134.27 MAE = 92.08	RMSE = 57.82 MAE = 31.43	RMSE: 56.94% MAE: 65.87%
Dec 25		RMSE = 76.03 MAE = 39.49	RMSE: 43.38% MAE: 57.11%
Dec 31		RMSE = 59.76 MAE = 39.53	RMSE: 55.49% MAE: 57.07%

From the above Table 4.3, the error of each of the random day for clear sky category is all below the baseline error for clear sky category obtained from Table 4.2. Same to partially cloudy, all the error of the days tested are below the baseline error for partially cloudy category obtained from Table 4.2 too. The percentage of improvement by using SVM algorithms instead of Linear Regression are significant. This shows that the predictive model trained with SVM algorithm is valid to use as a forecasting model as the predictive model can performed accurate prediction.

4.5 Trained Prediction Model of Clear Sky Category

In clear sky category, 70% of the data is used to train the predictive model while the remaining 30% of data is used as test data. A random day which is 19 August 2016 is chosen and the variables data is used to forecast the solar PV power output. Figure 4.1 is categorized under case 1, Figure 4.2 - 4.7 are categorized under case 2 and Figure 4.8 – 4.12 are categorized under case 3. Below shows the results of forecasting with different predictor input of predictive model.

Table 4.4 Different Cases of Trained Predictive Model for Clear Sky Category

Case 1:Trained Prediction Model with all Predictor Variables	**Case 2: Trained Prediction Model without one of the Dominant Predictor Variables**	**Case 3: Trained Prediction Model without two or more Dominant Predictor Variables**
With all Variable	Without Irradiance	Without Irradiance & Mono Temperature
	Without Mono Temperature	Without Irradiance & Ambient Temperature
	Without Ambient Temperature	Without Mono & Ambient Temperature
	Without Rain	Without Irradiance Mono & Ambient Temperature
	Without Wind	Without Wind and Rain
	Without Relative Humidity	

Case 1: Trained Prediction Model with all Predictor Variables

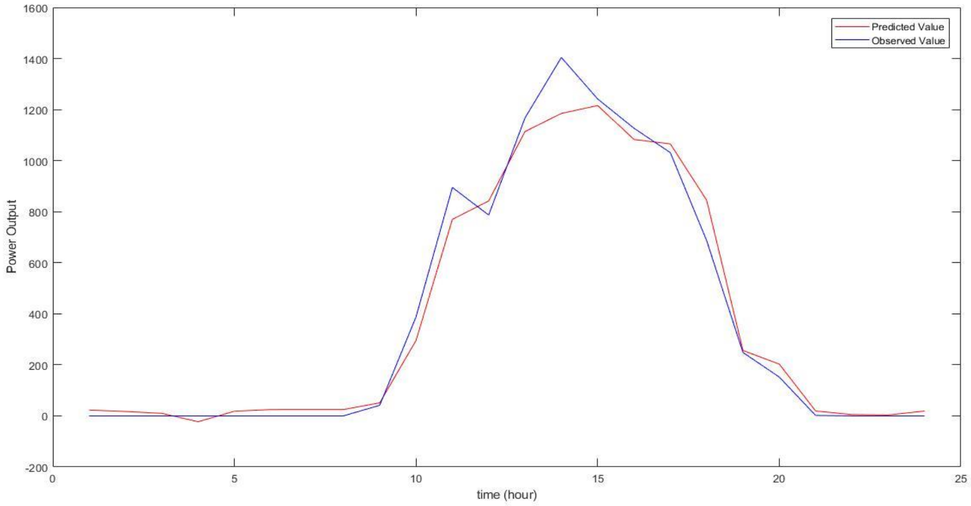

Figure 4.1 Graph of Predicted Value vs Observed Value

Case 2: Trained Prediction Model without one of the Dominant Predictor Variables

1) Without Irradiance

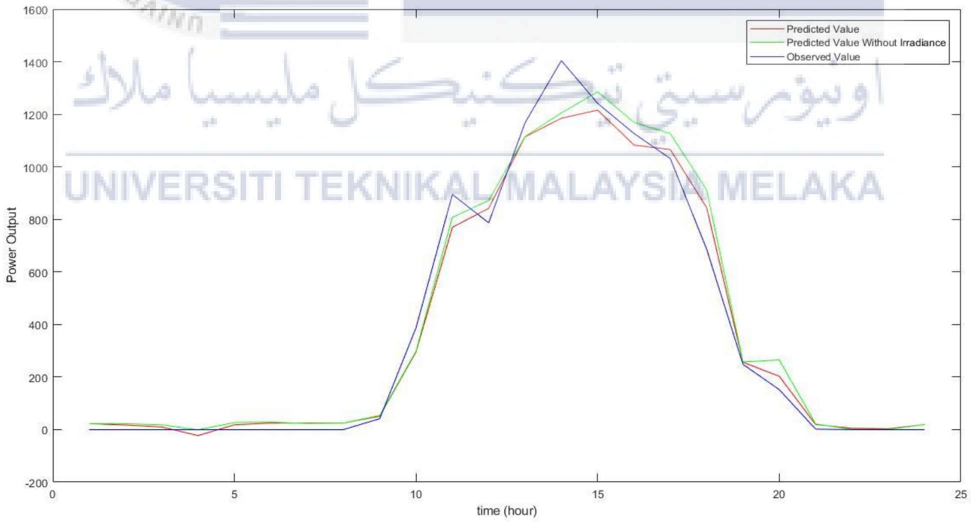

Figure 4.2 Comparison of Predicted Values with and without Irradiance

2) Without Mono Temperature

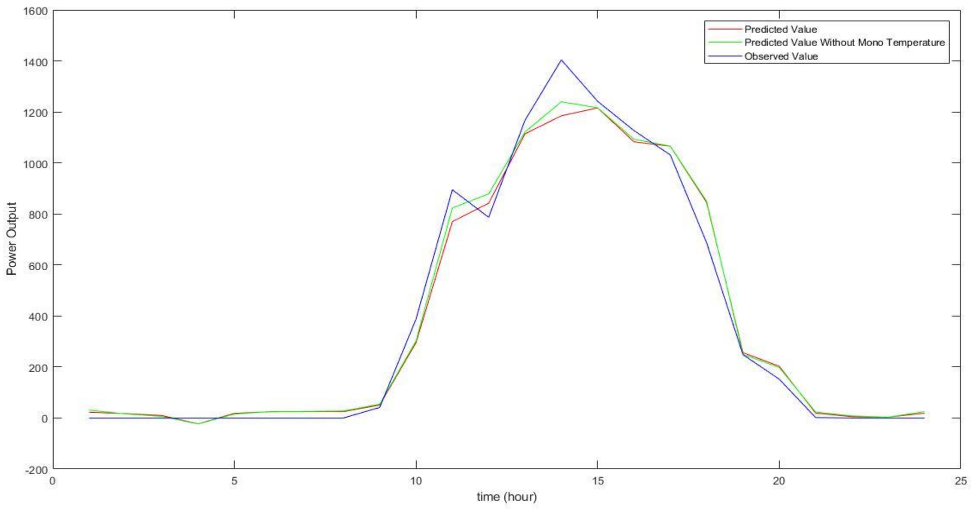

Figure 4.3 Comparison of Predicted Values with and without Mono Temperature

3) Without Ambient Temperature

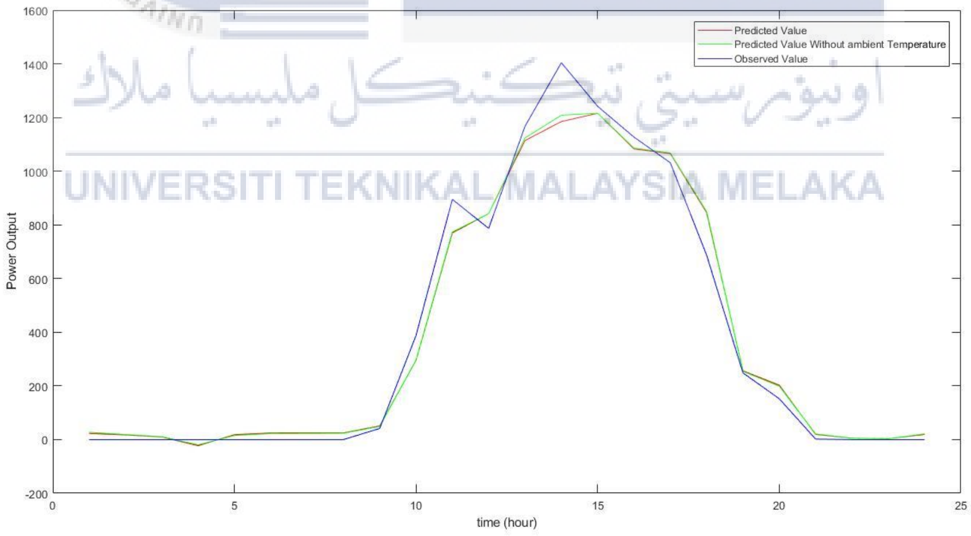

Figure 4.4 Comparison of Predicted Values with and without Ambient Temperature

4) Without Rain

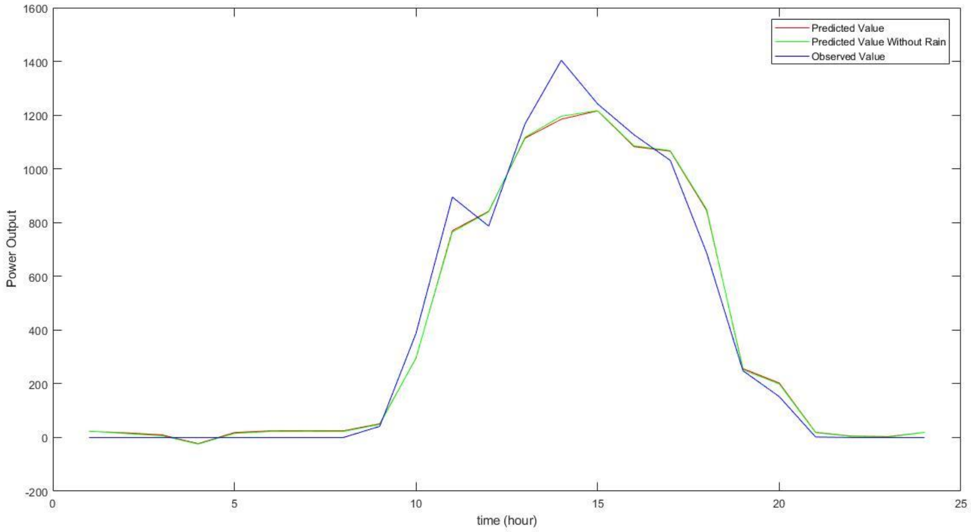

Figure 4.5 Comparison of Predicted Values with and without Rain

5) Without Wind

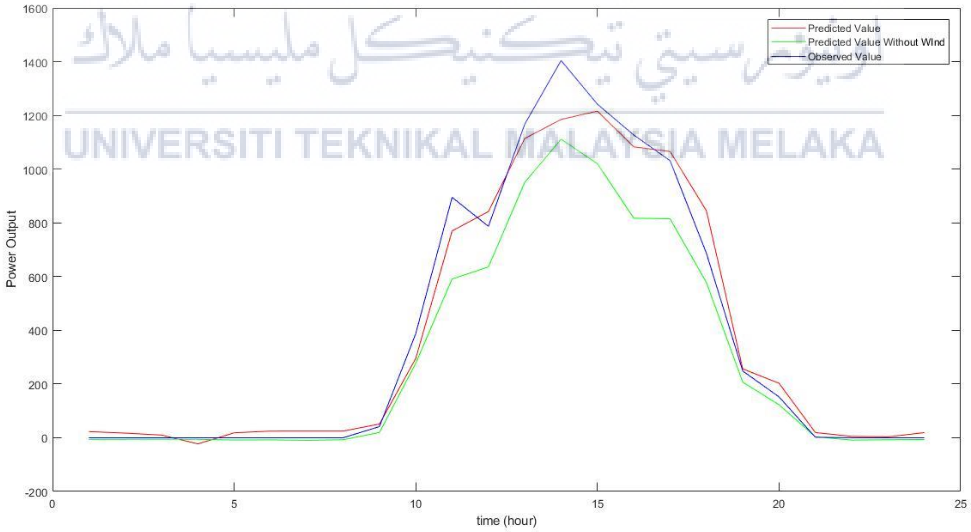

Figure 4.6 Comparison of Predicted Values with and without Wind

6) Without Relative Humidity (RH)

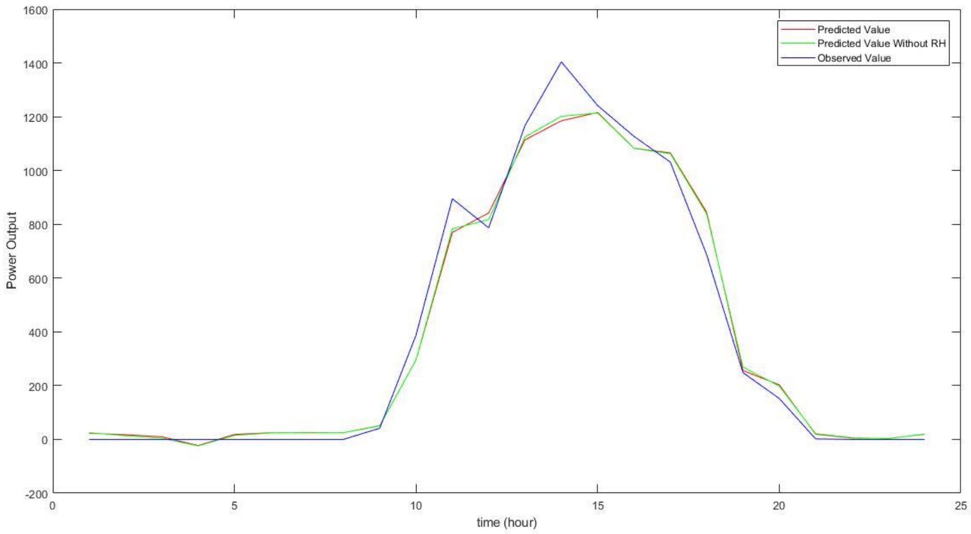

Figure 4.7 Comparison of Predicted Values with and without RH

Case 3: Trained Prediction Model without two or more Dominant Predictor Variables

1) Without Irradiance and Mono Temperature

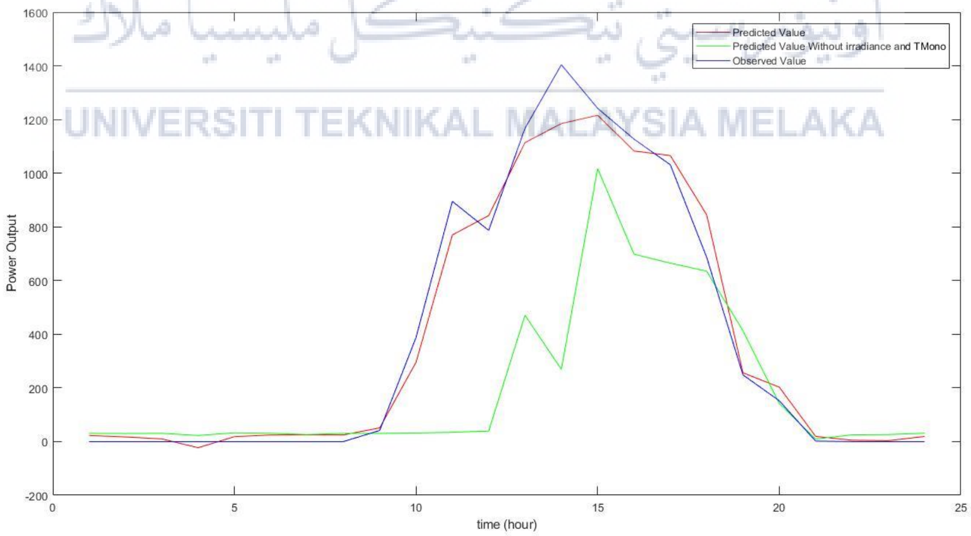

Figure 4.8 Comparison of Predicted Values with and without Irradiance and Mono Temperature

2) Without Irradiance and Ambient Temperature

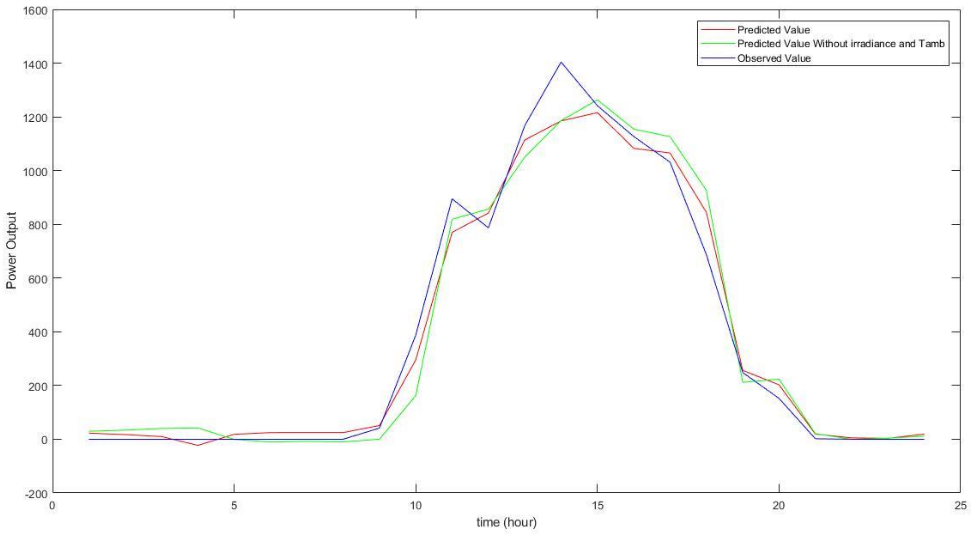

Figure 4.9 Comparison of Predicted Values with and without Irradiance and Ambient Temperature

3) Without Mono Temperature and Ambient Temperature

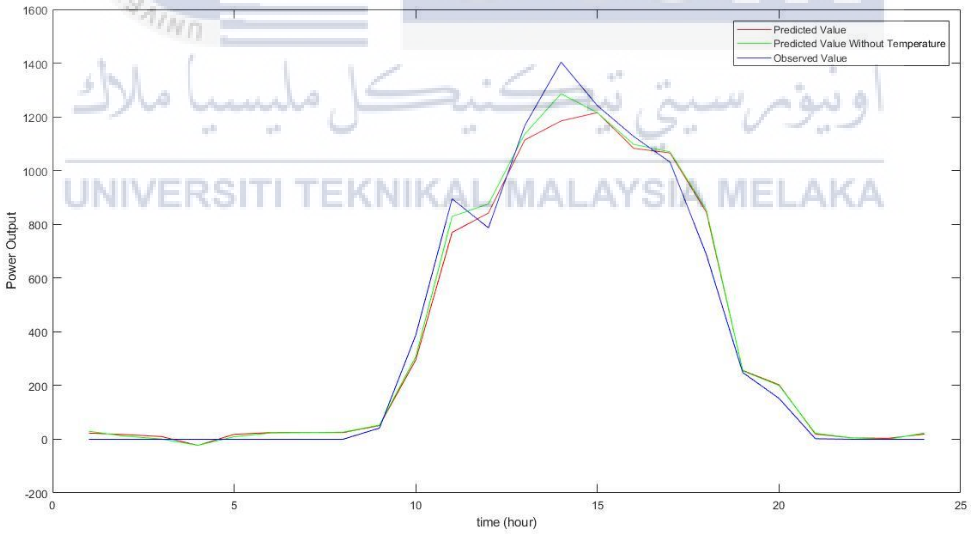

Figure 4.10 Comparison of Predicted Values with and without Mono Temperature and Ambient Temperature

4) Without Irradiance, Mono and Ambient Temperature

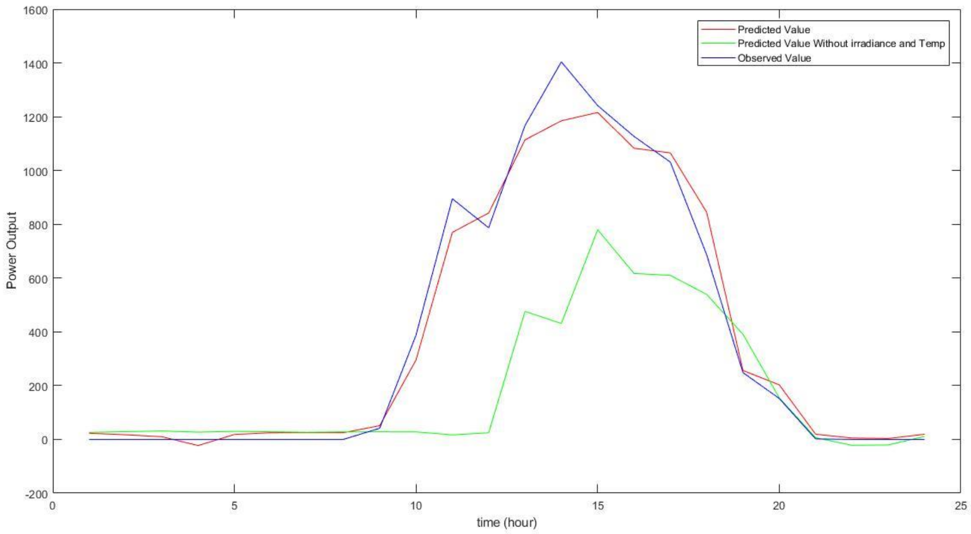

Figure 4.11 Comparison of Predicted Values with and without Irradiance, Mono Temperature and Ambient Temperature

5) Without Wind and Rain

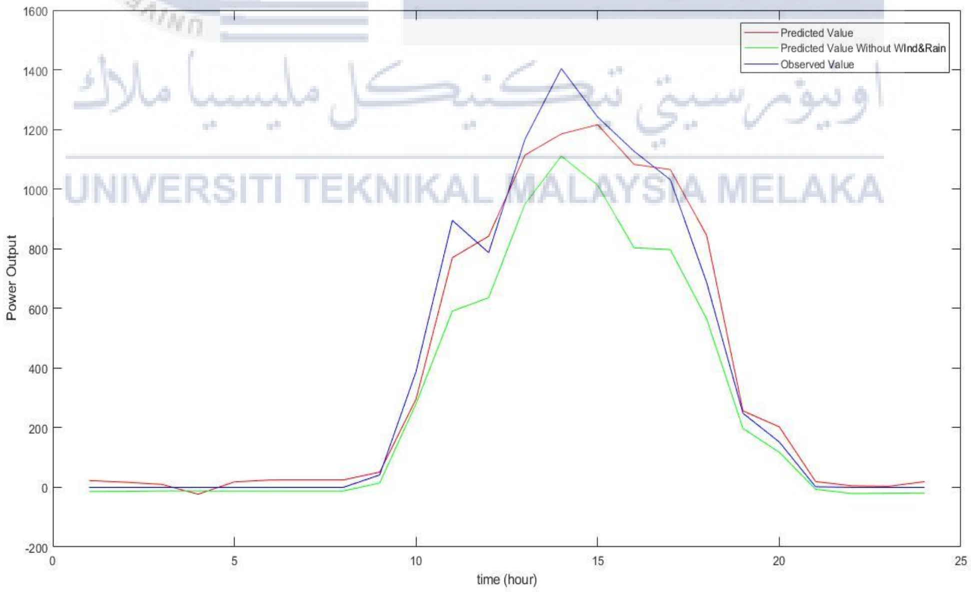

Figure 4.12 Comparison of Predicted Values with and without Wind and Rain

4.5.1 Error for each Predictive Model without one or more Dominant Predictor Variable

The table below shows the RMSE and MAE of each predictive model that can justify the accuracy of predictive model.

Table 4.5 Error for each Predictive Model for Clear Sky

Predictive Model	RMSE	MAE
With all variables	68.89	45.26
Without Irradiance	78.13	52.54
Without Mono Temperature	60.26	41.92
Without Ambient Temperature	65.55	43.47
Without Rain	67.70	43.70
Without Wind	139.54	87.66
Without Relative Humidity	63.99	42.28
Without Irradiance and Mono Temperature	387.65	224.02
Without Irradiance and Ambient Temperature	92.46	60.48
Without Ambient Temperature and Mono Temperature	54.51	37.81
Without Irradiance, Mono Temperature and Ambient Temperature	387.63	235.30
Without Wind and Rain	143.44	94.20

4.5.2 Summary

Based on all the graphs and error under the section of 4.3, it can be found that irradiance and module temperature are the most dominant variable which give the largest impact to the predictive model to perform accurate prediction for clear sky category. The predictive model shown in Figure 4.10 gives the most accurate prediction according to the lowest RMSE and MAE value obtained compared to other predictive models. This shows that when the model is trained with only one dominant variable, irradiance, it can perform better than other predictive models as the model trained considered only one variable rather than few variables that can give impact on the accuracy of prediction. However, in reality it is difficult to know the future irradiance profile beforehand because of its volatile nature of fluctuation.

Refer to Figure 4.3-Figure 4.5, Figure 4.7 and Figure 4.10 which these trained predictive model that excluded one or more predictor variables except irradiance, the predictive model still gives an accurate prediction as the RMSE and MAE of these predictive model is better than the RMSE and MAE of predictive model trained with all variables. Therefore, it can be conclude that irradiance is the most dominant variable that enough to use to train an accurate predictive model although without the data of some of the other variables.

Moreover, module temperature also one of the dominant variable. Refer to Figure 4.2 and Figure 4.9, these predictive models without irradiance variable still provide convincing prediction as the RMSE and MAE are not too high over the error of predictive model trained with all variable which set as reference baseline. On the other hand, predictive model that trained without variables of irradiance and module temperature performed poorly as shown in Figure 4.8 and Figure 4.11 and the error of these predictive model are extremely high.

Last but not least, wind variable also shown its influences on the prediction model for clear sky category. When the predictive model is trained without wind variable data, the accuracy of the model dropped drastically as the RMSE and the MAE increased double compared to the based predictive model.

As a result, irradiance and module temperature are the most dominant variables as the irradiance direct affect the current output of the PV module and the module temperature affect the voltage output of the PV module. Therefore without the irradiance and module temperature data will decreased the performance of the trained predictive model. Next, wind will be the second most dominant variable after irradiance and module temperature that give huge impact to the performance of predictive model. This is because of that for clear sky category, wind will affect the movement of cloud and the cloud will caused shading that seriously affect irradiance, thus wind had to take consideration when training a predictive model. The remaining variables does not impact the predictive model under clear sky category.

4.6 Trained Prediction Model of Partially Cloudy Category

In partially cloudy category, 70% of the data is used to train the predictive model while the remaining 30% of data is used as test data. A random day which is 2 November 2016 is chosen and the variables data is used to forecast the solar PV power output. Figure 4.13 is categorized under case 1, Figure 4.14 - 4.19 are categorized under case 2 and Figure 4.20 – 4.24 are categorized under case 3. Below shows the results of forecasting with different predictor input of predictive model.

Table 4.6 Different Cases of Trained Predictive Model for Partially Cloudy Category

Case 1:Trained Prediction Model with all Predictor Variables	**Case 2: Trained Prediction Model without one of the Dominant Predictor Variables**	**Case 3: Trained Prediction Model without two or more Dominant Predictor Variables**
With all Variable	Without Irradiance	Without Irradiance & Mono Temperature

	Without Mono Temperature	Without Irradiance & Ambient Temperature
	Without Ambient Temperature	Without Mono & Ambient Temperature
	Without Rain	Without Irradiance Mono & Ambient Temperature
	Without Wind	Without Wind and Rain
	Without Relative Humidity	

Case 1: Trained Prediction Model with all Predictor Variables

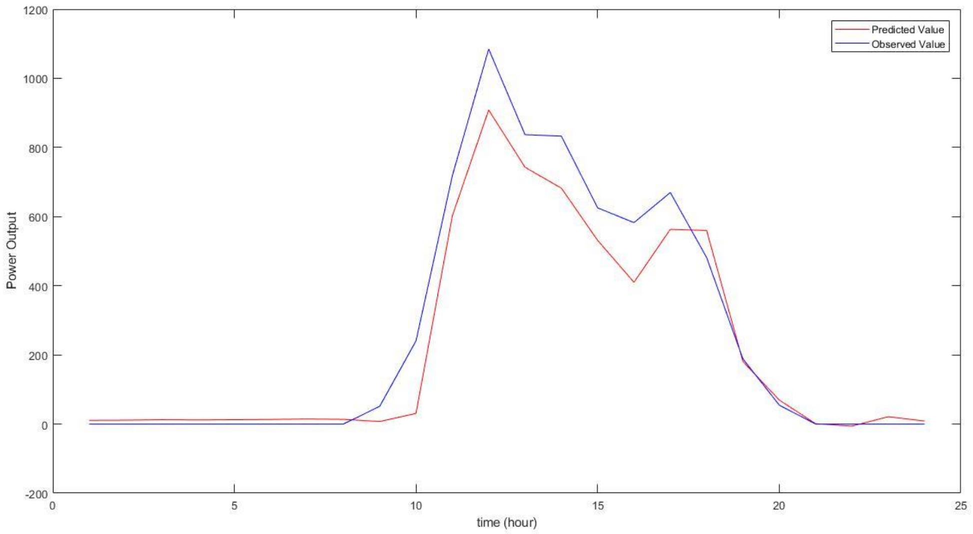

Figure 4.13 The Plot of Fitted Response versus Observed Response

Case 2: Trained Prediction Model without one of the Dominant Predictor Variables

1) Without Irradiance

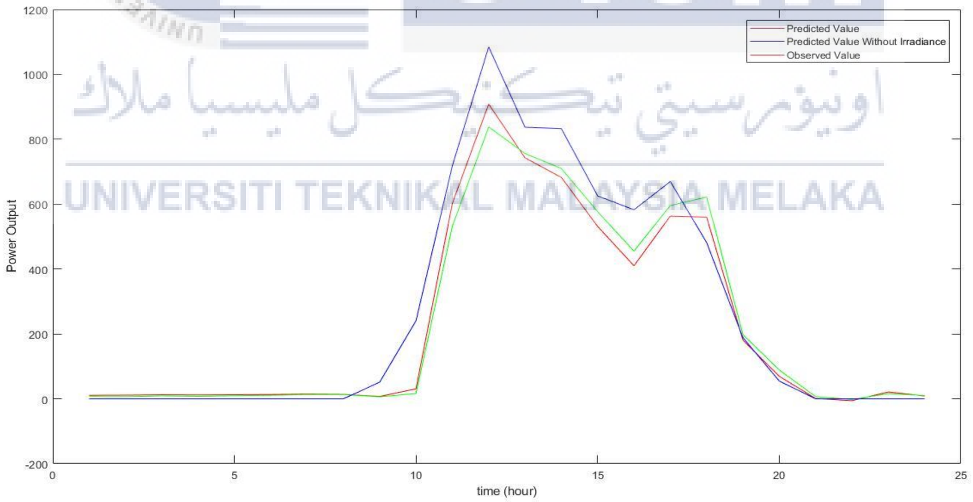

Figure 4.14 Comparison of Predicted Values with and without Irradiance

2) Without Mono Temperature

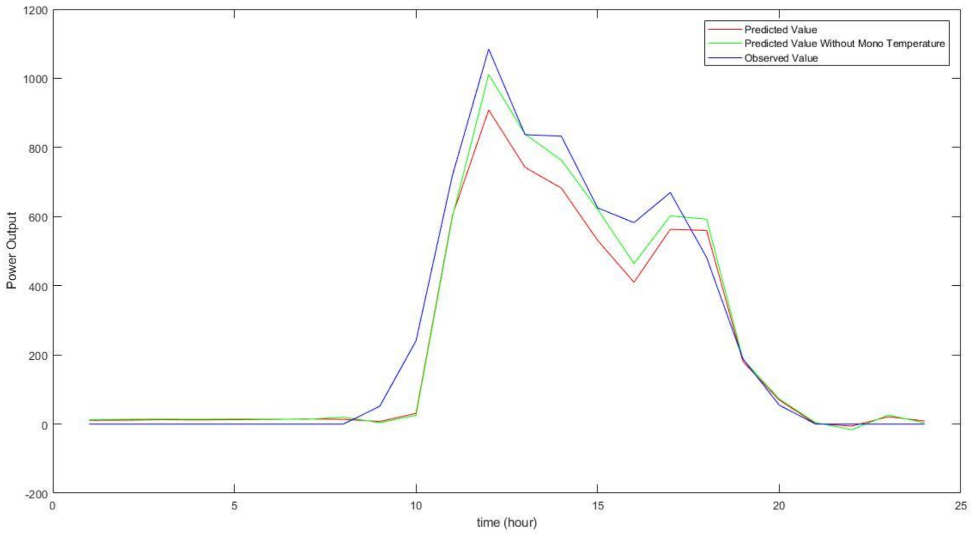

Figure 4.15 Comparison of Predicted Values with and without Mono Temperature

3) Without Ambient Temperature

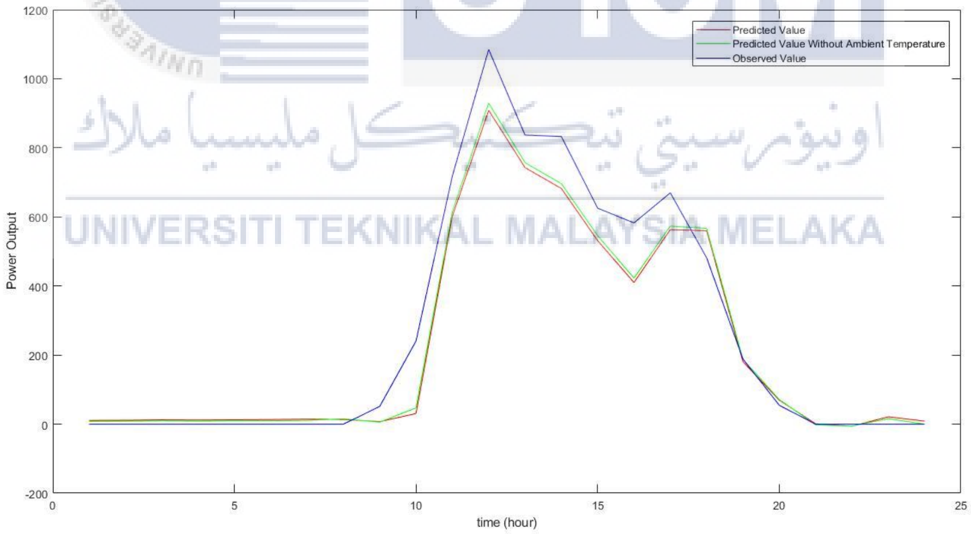

Figure 4.16 Comparison of Predicted Values with and without Ambient Temperature

4) Without Rain

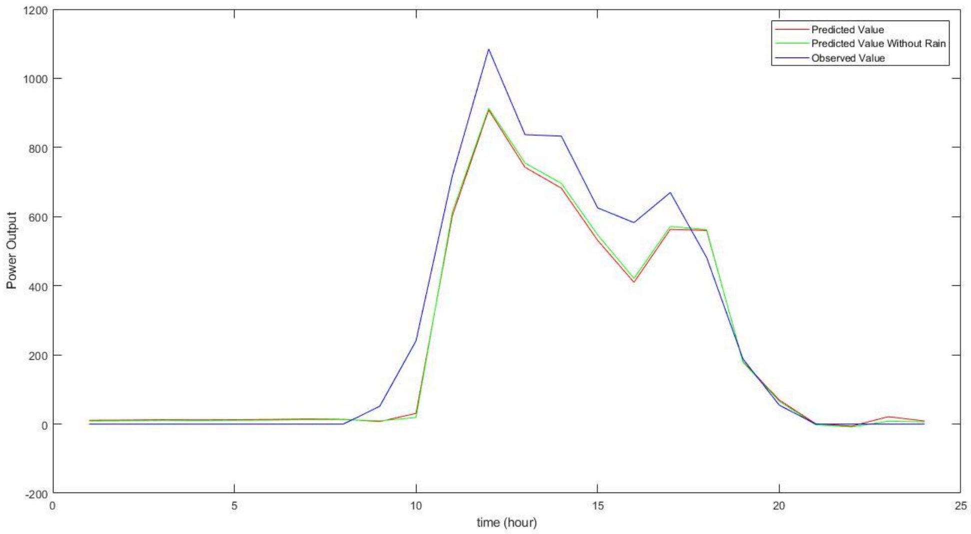

Figure 4.17 Comparison of Predicted Values with and without Rain

5) Without Wind

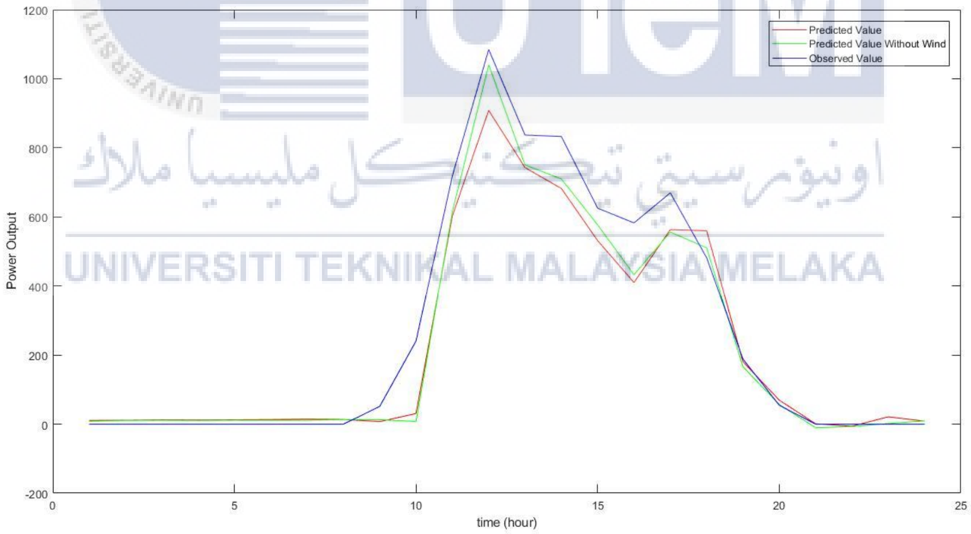

Figure 4.18 Comparison of Predicted Values with and without Wind

6) Without Relative Humidity

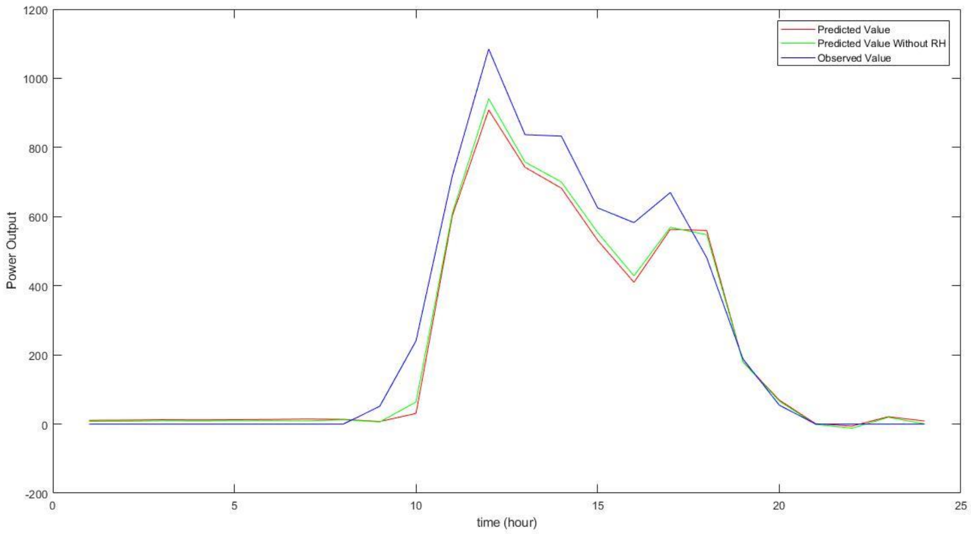

Figure 4.19 Comparison of Predicted Values with and without RH

Case 3: Trained Prediction Model without two or more Dominant Predictor Variables

1) Without Irradiance and Mono Temperature

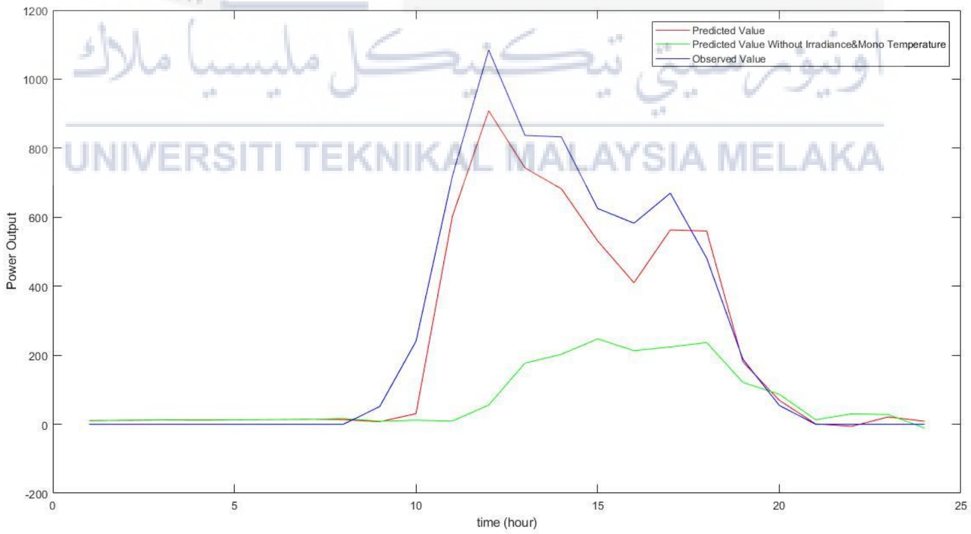

Figure 4.20 Comparison of Predicted Values with and without Irradiance and Mono Temperature

2) Without Irradiance and Ambient Temperature

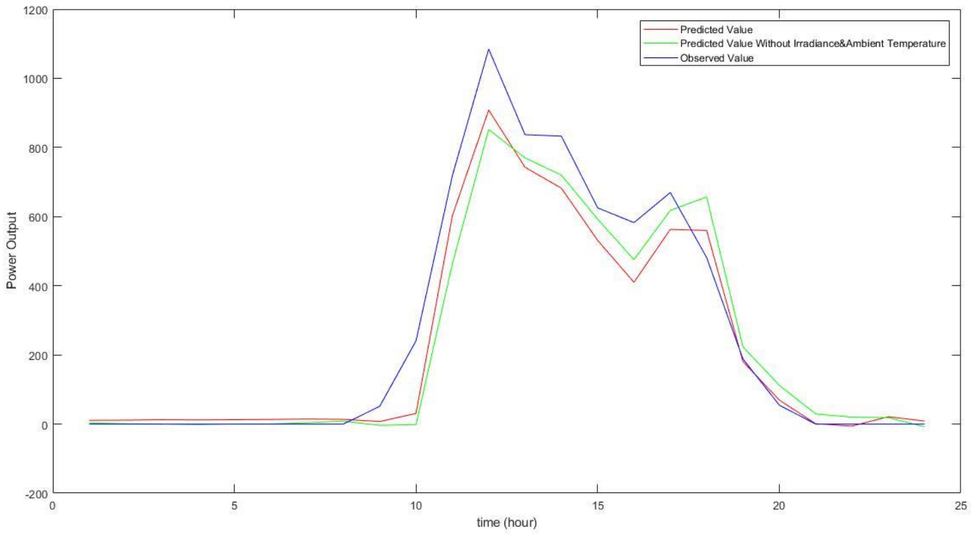

Figure 4.21 Comparison of Predicted Values with and without Irradiance and Ambient Temperature

3) Without Mono Temperature and Ambient Temperature

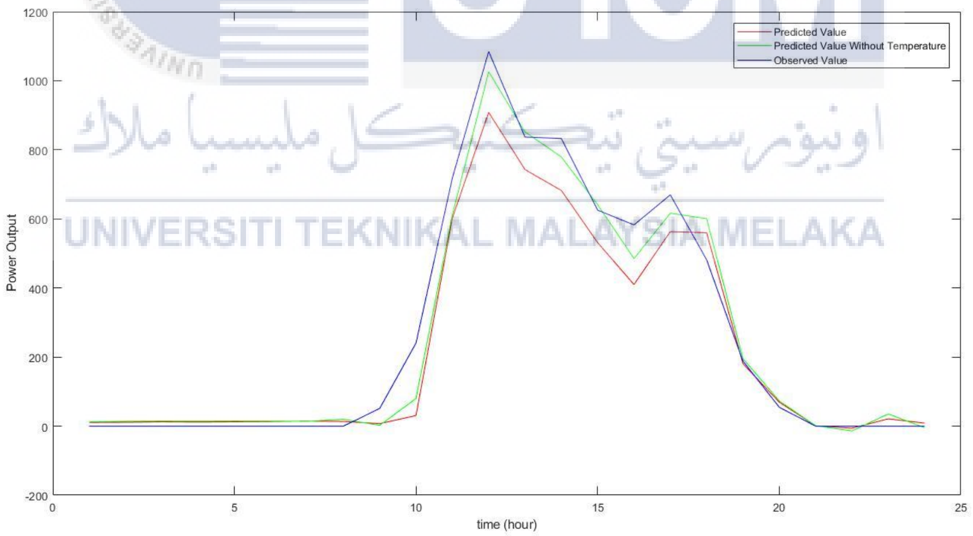

Figure 4.22 Comparison of Predicted Values with and without Mono Temperature and Ambient Temperature

4) Without Irradiance, Mono and Ambient Temperature

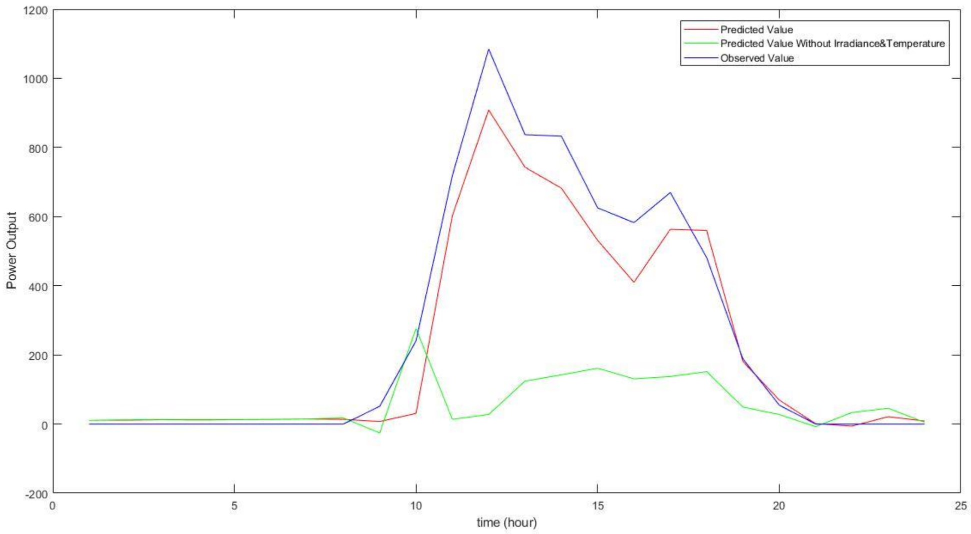

Figure 4.23 Comparison of Predicted Values with and without Irradiance, Mono Temperature and Ambient Temperature

5) Without Wind and Rain

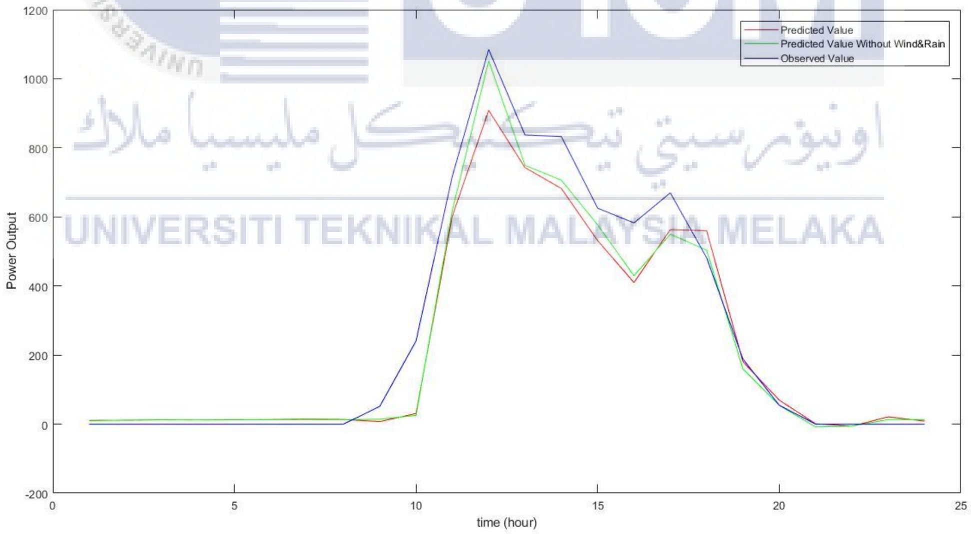

Figure 4.24 Comparison of Predicted Values with and without Wind and Rain

4.6.1 Error for each Predictive Model without one or more Dominant Predictor Variable

The table below shows the RMSE and MAE of each predictive model that can justify the accuracy of predictive model.

Table 4.7 Error for each Predictive Model for Partially Cloudy

Predictive Model	RMSE	MAE
With all variables	86.67	58.43
Without Irradiance	94.87	60.52
Without Mono Temperature	66.86	42.23
Without Ambient Temperature	78.86	52.39
Without Rain	83.31	54.60
Without Wind	73.98	46.29
Without Relative Humidity	74.80	50.26
Without Irradiance and Mono Temperature	353.26	209.31
Without Irradiance and Ambient Temperature	102.14	63.24
Without Mono Temperature and Ambient Temperature	56.43	38.63
Without Irradiance, Mono Temperature and Ambient Temperature	378.73	225.88
Without Wind and Rain	72.31	46.21

4.6.2 Summary

In partially cloudy category, the most accurate predictive model is the model shown in Figure 4.22 with the lowest RMSE and MAE value. Similar with the clear sky category, the irradiance and the module temperature are still the most dominant variables that cause the largest impact to the predictive model. Under section 4.6, by referring to most of the predictive model that included irradiance variable, the accuracy and error are convincing as a prediction model where the RMSE and MAE for each of the models are quite close to each other. This shows that the irradiance still play an important role to train a predictive model to perform day ahead solar PV power output forecasting.

Furthermore, the module temperature is the second most influence variable under partially cloudy category. Based on Figure 4.14 and Figure 4.21, the module temperature step up as a dominant variable that affect the accuracy of predictive model when the irradiance variable is absent. Thus, it can found that the module temperature give great influence on the predictive model aside from irradiance.

For this case of category, other variables had become less impactful to the predictive model. Unlike the clear sky category, wind variable does not influence the accuracy of the predictive model too much as the cloud and shading factor already has the information implicitly embedded in it so that the relation between irradiance and wind had become weaker until the wind variable less likely influence the accuracy of predictive model.

As a result, irradiance and module temperature are still the two most dominant variables that will affect the accuracy of the predictive model strongly while the other variables give less impact on the predictive model. Without irradiance and module temperature, the accuracy of the predictive will become drastically low, causing the model to perform poorly as shown in Table 4.7.

CHAPTER 5

CONCLUSION

5.1 Conclusion

In conclusion, the prediction model designed with SVM algorithm through regression learner in MATLAB R2017b can performed prediction and give an accurate and convincing day ahead hourly solar PV output forecasting. The SVM predictive model also shows prediction results that far better than the baseline error. Therefore, it shows the validity of using SVM to build a predictive model. The predictor variables used are irradiance, tilt irradiance, ambient temperature, relative humidity, rain, module measured temperature, module calculated temperature and wind. The predicted solar output power is depend on the predictor variables above while irradiance and module temperature gives the largest impact when training the predictive model. Thus, the data of irradiance and module temperature is necessary to train a predictive model compared to other variables.

Moreover, the data obtained from FKE Solar Lab is categorized according to clear sky index into three categories and used to train predictive model accordingly. The purpose of training two different predictive model accordingly is to increase the accuracy of PV output power forecasting based on the weather condition. The forecasted value of using correspond trained model based on weather condition is more precise.

5.2 Recommendation

From the results, it shows that SVM algorithms is validate to train a predictive model and the predictive model provided convincing and accurate prediction. However, the error of the predictive model still slightly higher and the predictive model still needed to improve to reduce the error. There are some ways to improve the predictive model and one of the method is to introduce hybrid SVM algorithms model to reduce the error. From the previous researches, it can found that the hybrid SVM algorithms had a higher performance compared to the normal SVM prediction model. It is also known that machine learning models perform better with more training data. Since the data in here is only limited to one year, future predictive models should incorporate data from other years as well to help improve the model accuracy.

REFERENCES

[1] I. Dr Fatih Birol, Executive Director, “Solar Energy,” *International Energy Agency*, 2018. [Online]. Available: https://www.iea.org/topics/renewables/solar/. [Accessed: 09-Nov-2018].

[2] “SEDA PORTAL,” *SEDA*, 2017. [Online]. Available: http://seda.gov.my/commissioned_plants.html. [Accessed: 15-Nov-2018].

[3] T. T. Teo, T. Logenthiran, and W. L. Woo, “Forecasting of photovoltaic power using extreme learning machine,” in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, 2015.

[4] SAURAV KAUSHIK, “An Introduction to Clustering & different methods of clustering,” *Analytics Vidhya*, 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/. [Accessed: 13-Nov-2018].

[5] C. Voyant *et al.*, “Machine learning methods for solar radiation forecasting: A review,” *Renewable Energy*. 2017.

[6] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, “Machine learning for solar irradiance forecasting of photovoltaic system,” *Renew. Energy*, 2016.

[7] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, “Predicting solar generation from weather forecasts using machine learning,” in *2011 IEEE International Conference on Smart Grid Communications, SmartGridComm 2011*, 2011.

[8] M. Rana, I. Koprinska, and V. G. Agelidis, “Univariate and multivariate methods for very short-term solar photovoltaic power forecasting,” *Energy Convers. Manag.*, 2016.

[9] J. Antonanzas, D. Pozo-Vázquez, L. A. Fernandez-Jimenez, and F. J. Martinez-de-Pison, “The value of day-ahead forecasting for photovoltaics in the Spanish electricity market,” *Sol. Energy*, 2017.

[10] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, “Forecasting power output of photovoltaic systems based on weather classification and support vector machines,” in *IEEE Transactions on Industry Applications*, 2012.

[11] S. Belaid and A. Mellit, “Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate,” *Energy Convers. Manag.*, 2016.

[12] R. Urraca, J. Antonanzas, M. Alia-Martinez, F. J. Martinez-De-Pison, and F.

Antonanzas-Torres, “Smart baseline models for solar irradiation forecasting,” *Energy Convers. Manag.*, vol. 108, pp. 539–548, 2016.

[13] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, and D. Heinemann, “Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data,” *Sol. Energy*, vol. 135, pp. 197–208, 2016.

[14] P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, “A benchmarking of machine learning techniques for solar radiation forecasting in an insular context,” *Sol. Energy*, 2015.

[15] F. Wang, Z. Zhen, B. Wang, and Z. Mi, “Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting,” *Appl. Sci.*, vol. 8, no. 1, p. 28, 2017.

[16] M. Abuella and B. Chowdhury, “Solar Power Forecasting Using Support Vector Regression,” *Proc. Am. Soc. Eng. Manag. Int. Annu. Conf.*, 2016.

[17] R. De Leone, M. Pietrini, and A. Giovannelli, “Photovoltaic energy production forecast using support vector regression,” *Neural Comput. Appl.*, vol. 26, no. 8, pp. 1955–1962, 2015.

[18] H. Y. Cheng, C. C. Yu, and S. J. Lin, “Bi-model short-term solar irradiance prediction using support vector regressors,” *Energy*, vol. 70, pp. 121–127, 2014.

[19] J. G. Da Silva Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, “Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan,” *Prog. Photovoltaics Res. Appl.*, 2012.

[20] R. Nageem and R. Jayabarathi, “Predicting the Power Output of a Grid-Connected Solar Panel Using Multi-Input Support Vector Regression,” *Procedia Comput. Sci.*, vol. 115, pp. 723–730, 2017.

[21] A. Alfadda, R. Adhikari, M. Kuzlu, and S. Rahman, “Hour-ahead solar PV power forecasting using SVR based approach,” in *2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2017*, 2017.

[22] J. M. Bakhashwain, “Prediction of global solar radiation using support vector machines,” *Int. J. Green Energy*, vol. 13, no. 14, pp. 1467–1472, 2016.

[23] F. Wang, Z. Zhen, Z. Mi, H. Sun, S. Su, and G. Yang, “Solar irradiance feature extraction and support vector machines based weather status pattern recognition

model for short-term photovoltaic power forecasting,” *Energy Build.*, vol. 86, pp. 427–438, 2015.

[24] R. Xu, H. Chen, and X. Sun, “Short-term photovoltaic power forecasting with weighted support vector machine,” in *IEEE International Conference on Automation and Logistics, ICAL*, 2012.

[25] J. Zeng and W. Qiao, “Short-term solar power prediction using a support vector machine,” *Renew. Energy*, 2013.

[26] K. P. Lin and P. F. Pai, “Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression,” *J. Clean. Prod.*, vol. 134, no. Part B, pp. 456–462, 2016.

[27] Y. Gala, Á. Fernández, J. Díaz, and J. R. Dorronsoro, “Hybrid machine learning forecasting of solar radiation values,” *Neurocomputing*, vol. 176, pp. 48–59, 2016.

[28] L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, and C. Sudheer, “A support vector machine-firefly algorithm-based model for global solar radiation prediction,” *Sol. Energy*, vol. 115, pp. 632–644, 2015.

[29] M. Bouzerdoum, A. Mellit, and A. Massi Pavan, “A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant,” *Sol. Energy*, vol. 98, no. PC, pp. 226–235, 2013.

[30] C. M. T. Huang, Y. C. Huang, and K. Y. Huang, “A hybrid method for one-day ahead hourly forecasting of PV power output,” *Proc. 2014 9th IEEE Conf. Ind. Electron. Appl. ICIEA 2014*, vol. 5, no. 3, pp. 526–531, 2014.

[31] M. Abuella and B. Chowdhury, “Random forest ensemble of support vector regression models for solar power forecasting,” in *2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2017*, 2017.

[32] M. G. De Giorgi, P. M. Congedo, M. Malvoni, and D. Laforgia, “Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate,” *Energy Convers. Manag.*, vol. 100, pp. 117–130, 2015.

[33] A. T. Eseye, J. Zhang, and D. Zheng, “Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information,” *Renew. Energy*, vol. 118, pp. 357–367, 2018.

[34] S. Sreekumar, K. C. Sharma, and R. Bhakar, “Optimized Support Vector Regression models for short term solar radiation forecasting in smart environment,” *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, pp. 1929–1932, 2017.

[35] Z. Dong, D. Yang, T. Reindl, and W. M. Walsh, "A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance," *Energy*, vol. 82, pp. 570–577, 2015.

APPENDICES

APPENDIX A GANTT CHART

Year	2018																	2019
Month	September				October					November				December				January
Activities	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Title Registration										Mid-Sem Break						Revision Week	Final Examination	
Identify the Problems																		
Identify the Objective																		
Project Research																		
Input Data Analysis																		
Software Simulation																		

Year	2019															
Month	February		March				April				May					June
Activities	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Project Research									Mid-Sem Break						Revision Week	Final Examination
Input Data Analysis																
Software Simulation																
Performance Analysis																
Troubleshooting																
Preparation of Final Report																