SOLAR ENERGY FORECASTING FOR GRID-CONNECTED PV SYSTEMS

MUHAMMAD AIMAN BIN KHOSIM

BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

SOLAR ENERGY FORECASTING FOR GRID-CONNECTED PV SYSTEMS

MUHAMMAD AIMAN BIN KHOSIM

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

2021

DECLARATION

I declare that this thesis entitled "SOLAR ENERGY FORECASTING FOR GRID-CONNECTED PV SYSTEMS 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 : MUHAMMAD AIMAN BIN KHOSIM

Date : 1/7/2021

APPROVAL

I hereby declare that I have checked this report entitled "SOLAR ENERGY FORECASTING FOR GRID-CONNECTED PV SYSTEMS" 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 : DR. KYAIRUL AZMI BIN BAHARIN

Date : 5 JULAI 2021

DEDICATIONS

To my adoring mother and father, Khosim Bin Daud, and Sharifah Binti Hassan, my siblings and family. Special thanks to all my friends for the support and motivation.

ACKNOWLEDGEMENTS

As a first thank you to Allah for providing me with good health and the opportunity to complete my report at this difficult time, I would want to convey my thanks to him for the opportunity to do so. I would also want to express my appreciation to Dr Kyairul Azmi Bin Baharin, who served as my Final Year Project Supervisor and helped me a great deal with explanations, direction, and moral support during this semester. This Final Year Project would have been impossible to complete without the assistance and supervision of Dr Kyairul, especially during this very demanding semester. I admired my supervisor for being an excellent educator on managing the Simulation Project despite the fact that we did not have a face-to-face chat, which is difficult to achieve while communicating online. Without the instruction and assistance of my outstanding supervisor, it would have been impossible to complete this simulation project and submit this report in its current form.

I am also deeply proud to my FYP panels, Dr Mohd Hendra bin Hairi and Dr Aimie Nazmin Bin Azmi for their guidance, advices and encouragement. Without their continued feedbacks and attention, this project would not have been same as presented here. I would also like to express my gratitude to the University Technical Malaysia Malacca (UTeM) for providing a free library host webpage that has assisted me in my data collection and literature review for this project. Furthermore, Single Buyer deserves special recognition for their assistance in supplying the necessary literature and information.

My fellow FYP students should also be acknowledged for their help. Especially, thanks to Yusuf Bin Abdullah and Nurul Amanina as they are also helpful out of a number of my colleagues in aiding complete this report. My heartfelt thanks also goes to my parents Khosim Bin Daud and Sharifah Binti Hasan for their emotional support, encouragement, and recommendations in order to achieve my final year assignment. For those who have contributed aid at different moments in aid to finish this project, my deepest apology as sadly it is not feasible for me to include all of them in this limited area. To keep a long tale short, thank you to all who have helped me.

ABSTRACT

Renewable energy sources, such as solar energy, are an excellent investment for a future source of alternative electricity. However, the inherent characteristics of solar energy created several obstacles for renewable energy management systems in terms of power output management, monitoring, and scheduling. The challenge of sustaining the photovoltaic (PV) system's output for a grid-connected solar power plant must be resolved. Therefore, this paper will provides a novel way of forecasting modelling approaches for predict the power output of a photovoltaic systems based on the daily weather of historical data that generated by UTeM solar panel. This forecasting method is required to enhance the controllability and stability of photovoltaic power system performance and to ensure the power grid operates in a reliable and cost-effective manner. A comparative performance of several forecasting systems utilizing Linear Regression Toolbox approaches have also been documented. Generally, the raw data from the photovoltaic (PV) panel are acquired and organised into historical data with the assistance of Microsoft Excel Software. Subsequently, any disinformation or null data are filter out from the historical data before incorporate into the simulation. These historical data then provided to the MATLAB Toolbox that is used to simulate the data and create the appropriate forecasting model. In the simulation, several machine learning techniques are used to develop a PV power output prediction model that can be used to predict hourly, daily and weekly PV power output using machine learning algorithms such as linear regression, Support Vector Machines (SVM), and Gaussian Process Regression (GPR). The regression learner tool in MATLAB software version R2020b is used to create, train, and evaluate the prediction models. The findings show that complex regression models such as the Exponential GPR are more accurate in the long run when compared to linear regression methods such as the Interactions Linear Regression model. Additionally, by removing multiple criteria from being trained in the regression learner have an influence on the prediction models and the accuracy of these regression models may be further enhanced.

ABSTRAK

Sumber tenaga boleh diperbaharui, seperti tenaga suria, adalah pelaburan yang sangat baik untuk sumber elektrik alternatif masa depan. Walau bagaimanapun, ciri-ciri tenaga suria yang wujud menciptakan beberapa halangan untuk sistem pengurusan tenaga boleh diperbaharui dari segi pengurusan pengeluaran kuasa, pemantauan, dan penjadualan. Cabaran untuk mengekalkan pengeluaran sistem fotovoltaik (PV) untuk loji tenaga solar yang disambungkan ke grid mesti diselesaikan. Oleh itu, makalah ini akan memberikan kaedah baru untuk meramalkan pendekatan pemodelan untuk meramalkan kekuatan sistem fotovoltaik berdasarkan cuaca harian data sejarah yang dihasilkan oleh panel solar UTeM. Kaedah ramalan ini diperlukan untuk meningkatkan kebolehkendalian dan kestabilan prestasi sistem tenaga fotovoltaik dan untuk memastikan grid kuasa beroperasi dengan cara yang boleh diharap dan menjimatkan. Prestasi perbandingan beberapa sistem peramalan yang menggunakan pendekatan Linear Regression Toolbox juga telah didokumentasikan. Secara amnya, data mentah dari panel fotovoltaik (PV) diperoleh dan disusun menjadi data sejarah dengan bantuan Perisian Micorosft Excel. Selepas itu, sebarang maklumat atau data batal disaring dari data sejarah sebelum dimasukkan ke dalam simulasi. Data sejarah ini kemudian diberikan kepada Kotak Alat MATLAB yang digunakan untuk mensimulasikan data dan membuat model ramalan yang sesuai. Dalam simulasi, beberapa teknik pembelajaran mesin digunakan untuk mengembangkan model ramalan output daya PV yang dapat digunakan untuk meramalkan output daya PV setiap jam, harian dan mingguan menggunakan algoritma pembelajaran mesin seperti regresi linear, Mesin Vektor Sokongan (SVM), dan Gaussian Proses Regresi (GPR). Alat pembelajaran regresi dalam perisian MATLAB versi R2020b digunakan untuk membuat, melatih, dan menilai model ramalan. Hasil kajian menunjukkan bahawa model regresi kompleks seperti Exponential GPR lebih tepat dalam jangka masa panjang jika dibandingkan dengan kaedah regresi linear seperti model Interaksi Linear Regresi. Selain itu, dengan membuang beberapa kriteria daripada dilatih dalam pelajar regresi yang mungkin mempunyai pengaruh pada model ramalan dan ketepatan model regresi ini dapat ditingkatkan lagi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

PV	-	Photovoltaic
NN	-	Neural Network
SVR	-	Support Vector Regression
SVM	-	Support Vector Machines
GPR	-	Gaussian Process Regression
ANN	-	Artificial Neural Network
CNN	-	Convolutional Neural Network
RNN	-	Recurrent Neural Network
K-NN	-	K-Nearest Neighbor
K-SVM	-	Kernel Support Vector Machines
NBC	-	Naïve Bayes Classification
DT	-	Decision Trees
MAE	-	Mean Absolute Error
MSE	-	Mean Square Error
RMSE	-	Root Mean Square Error
RAM	-	Random Access Memory
PAC	-	Principle Component Analysis
TNB	-	Tenaga Nasional Berhad
MATLAB	-	Matrix Laboratory
UTeM	-	University Technical Malaysia Malacca
MPP	-	Maximum Posterior Probability
MLE	-	Maximum Likelihood Error
MLP	-	Multilayer Perceptron
IoT	-	Internet of Things
DNA	-	Deoxyribonucleic acid

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Renewable energy becomes increasingly significant as our earth ages over time while fossil fuels and natural gases are gradually diminishing. Renewable energy gathered from renewable resources like sunlight or wind gets progressively more important because of their minimal environmental effect. As one of the greatest choices among the renewables energy , solar energy is beginning to become a revelation in the sphere of power production as it not only create zero net carbon emissions but also extremely sustainable. However, renewable energy such as solar are quite erratic which in effect will provide a variable power production that poses some issues in regulating the power system. The same applies for solar PV since solar PV mainly relies on the sun itself as it turns sunlight into electrical energy. Solar insolation, temperature and weather conditions have a large effect in influencing the output of a PV power

Moreover, Malaysia's solar farm is progressively crucial as time goes by to generate a second energy category, apart from natural gas and fossil fuels. Another reason why Malaysia is similarly prone to developing a solar grid is because of Malaysia's geographical location on the earth equator. This implies that Malaysia would most of the time have just two seasons, rain and daylight which is suitable for generating solar electricity. However, even with geographical advantages, solar farm output power is still not substantial as a reliable power output to be consume to residential or consumer due to weather changing every day. Other reasons why solar farm is conceivable to be utilized as staple energy generating is because to its unpredictability in delivering power production consistently. Compare to fossils fuels, as long as the raw materials is still available and remains burning, the power production can be provided consistently with low loss. While solar farm, is greatly reliant on the weather and the quantity of sunshine that needed to reach the solar photovoltaic (PV) panel.

This uncertainty creates far too many obstacles to the constant output of solar energy into the power grid, therefore a solution for precise solar energy production forecast is necessary. As a result, an accurate forecasting approach is required to optimize the operation and management of the integrated power grid system. These forecasting algorithms should be capable of accurately anticipating the PV power output one or more steps in advance, either directly or indirectly. For example, short-term forecasting, ranging from one minute to 24 hours in advance, is required for managing photovoltaic (PV) grid operations, performing maintenance assessments, and acting as a backup energy provider to consumers in the event of a defect or power outage. While long range forecasting is advantageous for planning, scheduling, and doing research or analysis in order to foresee future trends in use.

Numerous forecasting techniques, either hybridized or standalone, are used to anticipate photovoltaic (PV) power production, including neural networks (NN), support vector regression (SVR), and multiple linear regression. To choose an acceptable forecasting technique, users must evaluate and examine the prediction's accuracy over a range of relevant characteristics, including solar irradiance, meteorological conditions, and time interval. Not only may these short and long term forecasting strategies help minimize the uncertainty associated with the power output produced, but they are also beneficial for anticipating future load demand as more experiment data is collected.

1.2 Problem Statement

Malaysia's geographical location near the equator ensures that we get a plenty of sunlight each day. However, even with Malaysia's geophysical advantages, the solar farm is still unable to provide constant electricity. The difficulty in sustaining the output power of photovoltaic (PV) systems is mostly due to the variation of sunlight hitting the PV panel. The quantity of irradiance sunlight reaching the photovoltaic panel is insufficient because part of it is hindered by cloud formation and air resistance.

Due to the fact that photovoltaic systems are only accessible during the day, power generators must be timed precisely in order to preserve the power system's balance while also satisfying customer needs. This would result in an increase in the cost of maintaining and managing electricity systems. Moreover, forecasting on both the short and long term is necessary for a big grid-connected solar power plant, because it is disadvantageous to be unable to estimate power production, since this prevents any projected power loss or excess power from being controlled effectively. Without a forecasting model, we cannot foresee the amount of electricity we will need to give to the user. Solar farms also have faults in terms of supplying guaranteed solar output electricity on a constant basis to fulfil customer demand. This is because solar energy production varies according on the weather each day. As a result, a robust forecasting model with a small margin of error is necessary to provide consistent expected output power.

While several studies have offered different forecasting methodologies, none of them are generally applicable. All of the approaches stated above are examined and analyzed using factors specific to the researcher's locality. Means the forecasting methodologies derived from the research cannot be applied entirely in particular areas. Additionally, there are a huge variety machine learning models accessible, either standalone or customized for specific projects. These machine learning models use a variety of techniques and approaches to properly forecast photovoltaic power production. The complexity of the models dictates the forecasting accuracy, which means that various models will estimate PV power production with varying degrees of precision. However, the inherent unpredictability of sporadic characteristics such as solar insolation and climatic conditions complicates grid management, since solar energy sources are intermittent, resulting in variable power generation. It is critical to determine which factors have the most influence on the production of photovoltaic electricity.

1.3 Objectives

The main purpose of this project is to concentrate on the above-mentioned problems and conduct the following tasks:

I. To forecast daily, weekly and monthly PV power output using historical data and machine learning.

II. To compare and evaluate the accuracy of forecasting model that is available in the MATLAB.

III. To analyse the correlation between input variables which affect the accuracy of forecasting model.

1.4 Project Scope and Limitation

The goal of this work is to produce a forecasting model for solar energy production using the MATLAB software package. The forecasting model is constructed utilizing Machine Learning techniques in combination with multiple Linear Regression Models. And the simulation will be run using the MATLAB Machine Learning Toolbox software, which will be fed historical data from UTeM's photovoltaic solar network. Next, a viable forecasting model based on the desired output with little uncertainty and a low root mean square error (RMSE) will be developed utilizing several approaches in the MATLAB Machine Learning Toolbox. Finally, a comparison and analysis the accuracy of the forecasting model are analyze to conclude a suitable forecasting model.

This project limitations is limited by forecasting simulation is done by using regression learner application from MATLAB software. While selected forecasting methods are assessed in terms of accuracy by comparing the value of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) obtained by simulation in MATLAB software. Lastly, the forecasting method can only done by hourly, daily and weekly, due to insufficient raw data that required from the UTeM PV inverter panel.

1.5 Outline of Report

The report is divided into five chapters, the first of which contains a general introduction to the project, which covers the project's history, problem description, goals, and scope. Follow-up to Chapter 2, which is a literature review, in which I compile all of my reading and study of other publishers, as well as theories and legitimate rationale for doing this project, into a single document. While Chapter 3 will discuss the approach of this project, which will cover all of the methods and procedures for developing the forecasting model and ensuring its success, Chapter 4 will discuss the results of this project. In Chapter 4, the final outcomes of the project will be presented, as well as a review on the project's final development. Finally, in Chapter 5, I will provide a short explanation of the project's overall results and suggestions.

CHAPTER 2

LITERATURE REVIEW

2.1 Machine Learning

Machine learning is a field of study that is concerned with two interrelated questions: "How can one create computer systems that improve automatically with experience?" and "What are the basic theoretical principles that govern all learning systems, whether they are implemented in computers, individuals, or organizations?" Machine learning research is critical for solving these basic scientific and technical concerns, as well as for the extremely practical computer software that has been developed and deployed across a wide variety of applications.

Machine learning encompasses a broad range of learning activities, from classifying emails as spam to recognizing faces in photographs to controlling robots to accomplish specific objectives. Each machine learning issue may be accurately stated as the challenge of increasing some metric of performance P while doing some task T through some sort of training experience E. For instance, in order to train an email spam filter, the goal T is to discover a function that translates each given input email to the output label spam or not-spam. The performance parameter P that has to be improved might be described as the accuracy of this spam classifier, and the training experience E could be described as a collection of emails that have been classified as spam or not spam. Alternatively, one may create a separate performance measure P that imposes a greater penalty on non-spam that is identified as spam than on spam that is identified as non-spam. Additionally, one may establish a distinct sort of training experience by putting unlabeled emails with those identified as spam and not-spam. After clearly specifying the three components T, P, E, the learning issue is properly characterized.

2.1.1 Machine Learning as Optimization.

Frequently, machine learning tasks are framed as optimization issues. For instance, when training a neural network with millions of parameters, we typically frame the learning task as one of determining the parameter values that optimize a specific objective function, such as minimizing the sum of squared errors in the network outputs relative to the desired outputs specified by training examples. Similarly, when we train a Support Vector Machine classifier, we structure the task as a restricted optimization problem with the purpose of minimizing a function called the hinge loss. When optimization issues are used to frame machine learning tasks, the learning method is often itself an optimization algorithm. At times, we use general-purpose optimization techniques such as gradient descent (to train neural networks, or quadratic programming (e.g., to train Support Vector Machines). In other instances, we may develop and use more efficient strategies for the particular learning job at hand.

2.1.2 Machine Learning as Probabilistic Inference

A second view is that machine learning tasks are often probabilistic in nature, requiring inference of the learnt model from training data and prior probabilities. Indeed, the two primary principles for deriving learning algorithms are the probabilistic principles of Maximum Likelihood Estimation (in which the learner seeks the hypothesis that maximizes the probability of the observed training data) and Maximum a Posterior Probability (MPP) estimation (in which the learner seeks the most probable hypothesis, given the training data plus a prior probability distribution over possible hypotheses). Occasionally, the learnt hypothesis (i.e., model) will include explicit probabilities. In other circumstances, even if the model parameters do not correspond to particular probabilities, we may still regard the training procedure as doing probabilistic inference to determine the Maximum Likelihood or Maximum a Posterior probability values for the network parameters. Notably, this approach, in which machine learning algorithms execute probabilistic inference, is quite consistent with the preceding view, in which machine learning algorithms solve an optimization issue. In most cases, deriving a learning algorithm based on the MLE or MPP principle entails first defining an objective function in terms of the hypotheses' parameters and the training data, and then solving for the hypothesis parameter values that maximize or minimize this objective using an optimization algorithm.

2.1.3 Machine Learning as Evolutionary Search

It should be noted that certain types of learning do not lend themselves to simple formulations such as optimization or probabilistic inference problems. Natural evolution, for example, might be seen as a learning process in which organisms become more effective from generation to generation. However, it is not obvious if there is an explicit goal that is being maximized over time in natural evolution, or if there is an equivalent probabilistic inference issue in natural evolution. It is more likely that the concept of a "increasingly successful creature" would develop over time, as the organism's environment and the set of rivals develop along with it.

2.2 Supervised Learning

The supervised learning model is the most significant machine learning model. It is mostly used to address real-world problems [1]. This model is used to forecast outcomes given a set of inputs and two input/output instances. Each supervised training dataset has a pair of input goals, an input vector, and a desired output value referred to as a supervisory signal. These examples are used to train machine learning algorithms, which then infer a classifier function based on the training datasets. In supervised learning, the goal of training algorithms is to predict the value of one or more outcomes using a variety of input characteristics.

It is possible to train models by using a training set, which will provide the desired result. This training dataset contains both inputs and correct outputs, which allows the model to improve its performance over time. The loss function is used to assess the accuracy of the method, which is then adjusted until the error has been reduced to an acceptable level.