NIPES

Journal of Science and Technology Research

Journal homepage: www.nipesjournals.org.ng

A Comparative Study of Machine Learning Models for Solar Power Generation Forecasting Using Weather Parameters: A Case of Benin City

A. Obayuwana[a], Godwin O. Monica[b]

[a, b] Department of Computer Engineering, University of Benin, 1154, P.M.B, Benin City, Nigeria

Article Info

Received 18 October 2021
Revised 13 November 2021
Accepted 16 November 2021
Available online 12 December 2021

***Keywords*:**
Solar Power, Supervised Learning, Machine Learning, Regression Model, Model Score, Weather Changes, Data Logger

https://doi.org/10.37933/nipes/3.4.2021.15

https://nipesjournals.org.ng

Abstract

Solar power generation as a renewable energy source is one of the most used and highest in demand in recent times, solar systems totally depend on weather changes. Because of weather fluctuations, one cannot manually determine the amount of energy produce by a solar system due to the amount of data and computational power needed. Machine learning having this computational advantage and being able to perform almost accurate predictions was implemented in this study. A comparison of three different machine learning models (decision tree regression, support vector regression and random forest) used to predict solar power generation were carried out using dataset generated over a given period of time across different locations in Benin City. An IoT -based data logger was developed to generate the dataset using temperature, humidity and light-insensitive sensors interfaced with an Atmega 238p Microcontroller. The data acquired by sensors are sent to the cloud through a GSM module connected to the Microcontroller for storage. This dataset constitutes of weather parameters (temperature, humidity and light flux) and electrical parameters (voltage and current). The preferred model was selected using performance metrics and minimal error value. The results of the comparison shown that Random Forest regression model had a model score of 0.9506, decision tree regression and support vector regression with model score of 1.0594 and 3.7632 respectively. Random forest regression model gives the most accurate predictions of solar power over a given period. The model score was calculated by addition of the values gotten from root mean square error, mean absolute error, mean square error and R-squared value.

1. Introduction

Machine learning is a research field at the intersection of statistics, artificial intelligence, and computer science and is also known as predictive analytics or statistical learning [1]. It employs a variety of statistical, probabilistic and optimization methods to learn from past experience and detect useful patterns from large, unstructured and complex datasets. It's simply the practice of programming computers to learn from data refers as training sets by finding relations between inputs and outputs even if the representation is impossible. This characteristic allows the use of machine learning models in many applications such as pattern recognition, classification problems, spam filtering, data mining and forecasting problems [2].

Lately machine learning has found application in predicting solar power over a period of time [3,4, 5]. The use of machine learning for solar power forecast will ensure effective operation of the power grid, optimal management of the energy fluxes occurring in the solar system and this will also help for congestion management. The application of machine learning algorithm in solar power generation forecasting has been experimented in several research works. In Emil and Mikael, 2018 a comparison of different machine learning techniques and time series models is performed across five different sites in Sweden, it was discovered that employing time series models was a complicated procedure due to the non-stationary energy time series. however, machine learning techniques were more straight forward to implement. However, in Abrahamsen et al, 2006, the goal was to build models that can predict weather conditions faster than traditional meteorological models. Trying to manually calculate and predict solar power will yield wrong results and might be too ambiguous for human to do. Thus, the focus of this research, proposes a two-step modeling process that connects unannounced weather variables with announced weather forecasts. This project, when applied in a solar powered infrastructure makes it possible for solar grid operators to predict and balance energy generation and consumption.

2. Methodology

This section describes the method adopted in this project, it explains the mechanism and processes involved in the project implementation. The methodology articulates what data is required, what methods are going to be used to collect and analyze this data, and how all of this is going to answer the objectives stated earlier. This is depicted in Figure1.

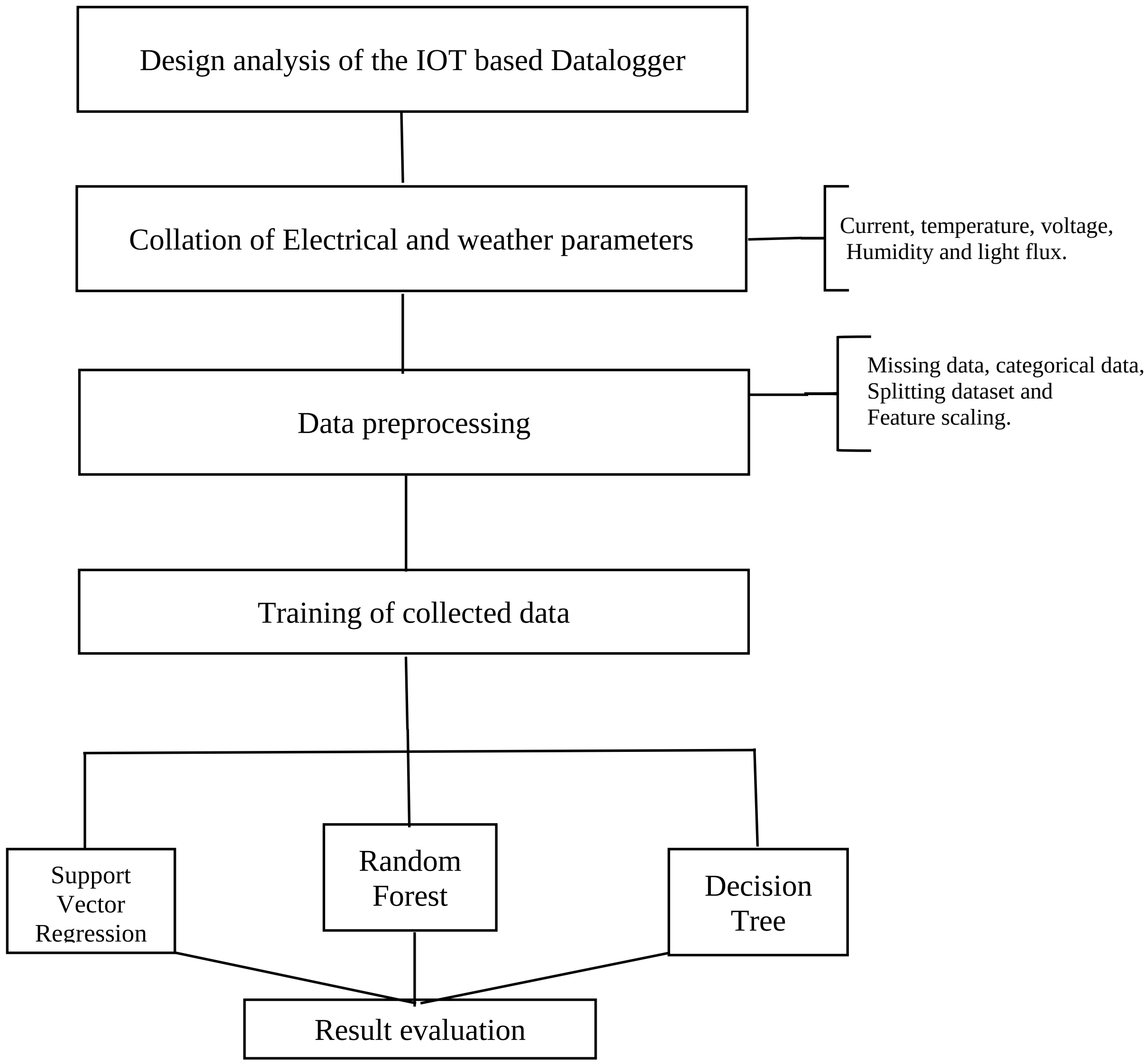


Figure1: Workflow process

2.1 Design Analysis of IoT-Based Datalogger

The IoT based datalogger design is presented in this section. Figure 2 shows the block diagram functional stages design of the tracker circuit and IoT data-logger and what they are made of and how they are interconnected.

1. Power source
2. Sunlight flux sensing stage
3. Temperature and humidity sensor
4. Load Voltage and current transducer
5. Arduino/Microcontroller stage
6. GSM Modem IOT interface
7. Software program

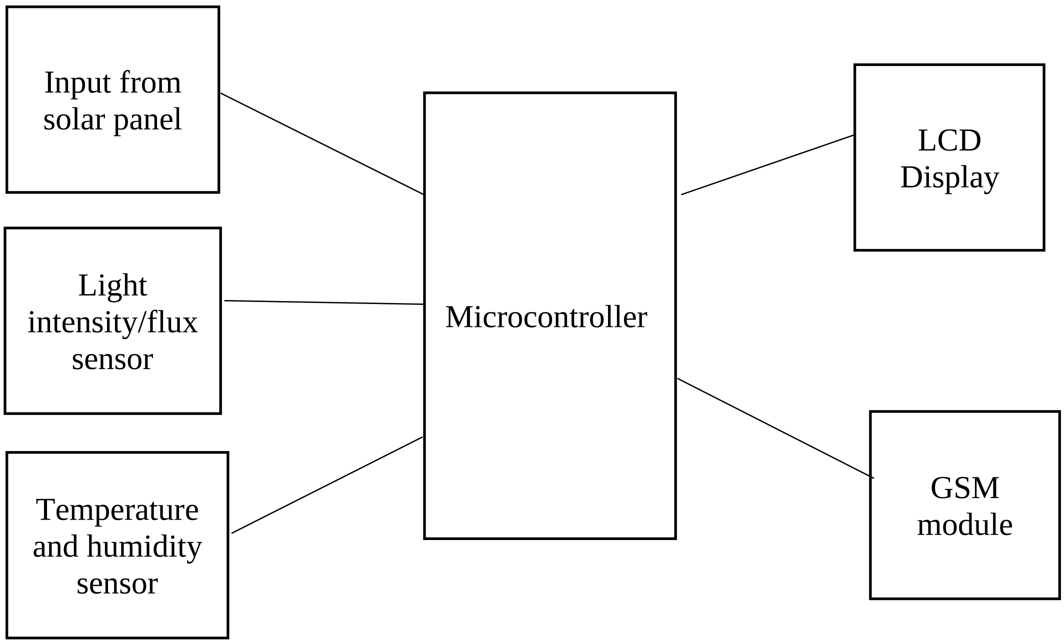


Figure 2: Block diagram of an IoT-based datalogger.

2.1.1. Power Supply Circuit Design

The power supply source for the control circuit is from the solar panel and from a back-up battery source. The circuit diagram is shown in Figure 3. The maximum power supply voltage from the solar panels is aimed at 22V based on the solar panel rating. The solar panel acquired is a monocrystalline 12V/150Watts solar panel. The supply expected output voltage for the control circuit is 5V and thus an IC voltage regulator 7805 was used which is a +5V voltage regulator. Due to the high dc voltage involved, a series regulator was used to regulate the dc to 14.3V dc before the 5V regulator for the Arduino control board. A 15V zener diode, limiting resistor and pass transistor make up the series pass regulator. The circuit consists of two stages consisting of the series regulator (RL,Vz, and Tr1) and IC voltage regulator.

The reference voltage is set up by 15V Zener diode model with a power rating of 300Mw (from datasheet).

Where, V_z is zener breakdown voltage,

Maximum Zener current, $Iz = P / Vz$ (1)

$(300\ x\ 10-3)/\ 15\ =\ 20mA$

This is the maximum current. A value less than the maximum is used so that it will operate within its capacity. A value of 70% of its maximum current is adequate for zener reference.

Chosen zener current

$$Iz\ =\ 0.7\ x\ Iz$$

$$Iz\ =\ 0.7\ x\ 20\ x\ 10-3\ =\ 14mA$$

This was necessary so that the zener diode would not fail. Current limiting resistor or biasing resistor RL value can thus be obtained from the relationship below

$$Rl=\ (V-\ VZ)/\ IZ \quad (2)$$

Where Vc = collector voltage = solar voltage = 37.V

$$Rl\ =\ 37-\ 15\ /\ 14\ x\ 10-3\ =\ 1.612\Omega$$

Used 1.8kΩ was used as closest standard

$$IZ\ =\ IB\ =\ Pass\ transistor\ base\ current\ =\ IC\ /\ ß1$$

$$Ic\ =\ ß1/IB \quad (3)$$

Where ß1 is pass transistor gain

The transistor chosen is TIP41C model for Tr1 as it satisfies the condition;

Vce (max) > $3V_c$

$$3Vc\ =\ 3\ x\ 22\ =\ 66V$$

TIP41 transistor Vce (max of 100V) > 66V

C1 was chosen as 47μF.

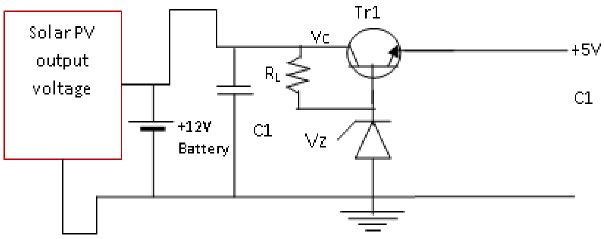


Figure 3: Power supply circuit

2.1.2. Light Flux Design Analysis

The sunlight intensity is a parameter to check out for as it relates to the amount of energy obtained from the solar PV module and the sunlight level sensor used for the work is the light dependent resistor (LDR) or photo-resistor. From test and observation, the following was detected;

Dark resistance = 1.5MΩ

Light resistance = 30Ω at full day light

LDR Type/model = Plastic case ORP12

The LDR resistance changes with the light intensity such that the more, light rays fall on it, the lower the resistance. This change must be converted to voltage change. This is achieved by connecting a fix resistance in series with the LDR to form a voltage divider. The voltage output is now light dependent.

Using Voltage Divider Rule;

$$Vout\ =\ R2/(\ RLDR\ +\ R2)\ x\ (Vcc) \quad (4)$$

Where V_S is Sensor Voltage and V_C is Voltage across series resistance.

Using daylight resistance of 30Ω and current of 1.2mA(chosen)

$$R2\ =\ Vcc-Vs\ =\ (Vcc-\ (IRLDR))/I \quad (5)$$

$$R2=\ (5-\ (1.2\ \times\ 10-3\ \times\ 30)\ )/1.2\times 10-33.3$$

$$R2 = 3266\Omega$$

Chose 3.3kΩ for R as the closest standard.

$Thus\ Vout\ (at\ full\ daylight) = 3300/(30 + 3300) \times 5 = 4.95V$

Thus Vout (at full darkness) = 3300/ (1.5 x 10^6 + 3300) × 5 = 0.01V

This voltage shows that the expected reference voltage be such, that it can be varied from a minimum of 0.1V to approximately 5V and the output shows that their voltage increases for increasing luminosity. This was converted by the microcontroller section to digital data for display and transmission to the cloud.

2.1.3. Voltage and Current Sensing Unit

The solar information of voltage and current is needed for the datalogger and so sensing circuits is required to achieve this. The sensing circuit function is to bring down the high analog voltage and current to a low value compatible with the microcontrollers onboard analog to digital converter (ADC) which is 2V maximum [3].

A voltage divider was used for dc voltage sensing and level measurement while a series resistor was used as the current sensor.

According to ohm's law the voltage developed across Rs is directly proportional to the load current.

$$Vs = IL\,Rs \tag{6}$$

Where Vs is sensor voltage, I_L is load current, and Rs is sensor resistance

For a sensor load voltage of 1V from Rs with load current of 10A

$$Rs = V/I = 1.0 / 10 = 0.1\Omega = 100m\Omega$$

Power rating of resistance Prs = I^2R = 10 x 10 x 0.1 = 10W

R1 and R2 form the voltage divider circuit for voltage level sensing. R2 was chosen to be a variable resistor for fine tuning and calibration purposes. For a current flow assumed to be 1.0mA flowing through R1 and R2 in series and for a peak dc voltage of 60V, then total resistance is;

$$Rt = Vdc / IT \tag{7}$$

Where Rt is the total resistance of the voltage divider of R1 and R2 and I_T is the current through them.

$$IT = 1.0mA\ (chosen)$$
$$Rt = 60 / 1.0\,x\,10 - 3 = 60k\Omega$$

This value was shared between both resistors. A value of 10kΩ variable resistor was used for R2 and 50kΩ for R1.

From the voltage divider expression,

$$Vout = Vin\,[R2 / (R1 + R2)] \tag{8}$$
$$Vout = 60\,[10\,000 / (50000 + 10000)]$$

Maximum adjustable ADC voltage, Vout = 10V

The operational amplifier was used as a voltage follower for its impedance matching properties to interface the current that was converted to voltage by Rs to the ADC input. The dual op-amp

LM358 was use for all the operational amplifier voltage follower circuitry because of economy and good slew rate and its unipolar operational capability.

2.1.4. Humidity and Temperature Sensing Unit

This sensor is used to sense humidity and it facilitates the conversion of analog inputs to digital output. The digital output pin is connected directly with the Arduino to Arduino's digital pin (pin 2). DHT11 has a full range temperature compensation, low power consumption, long term stability and calibrated digital signal. The DHT sensors are made of two parts, a capacitive humidity sensor and a thermistor. There is also an in-built chip that does analog to digital conversion and spits out a digital signal with the temperature and humidity.

2.1.5. Control Program and Software Design

The software is designed using arduino IDE specially for arduino based development boards and atmega controllers. The software was written in C language and was developed in sections for easy debugging and then later integrated. The integrated program was built and verified and the hex file was generated which was then loaded into the microcontroller.

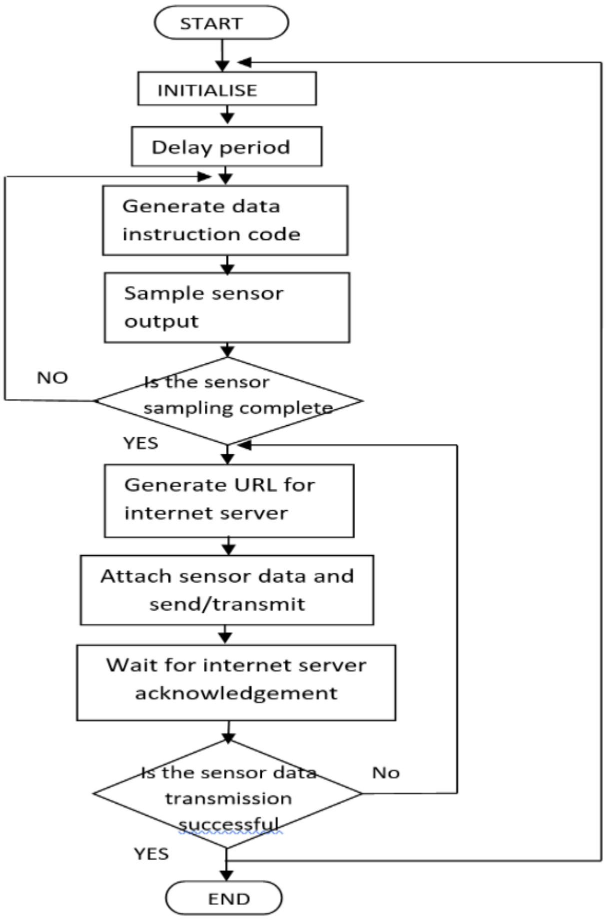


Figure 4: The flowchart of the code development

2.1.6. Operational Principle of the IoT-Based Datalogger

Figure 5 shows the complete circuit diagram of the Solar panel connected with the IoT-based Datalogger. The circuit operation is presented in two sections; component function description and operation. The LDR is the major sunlight sensing device used for checking the light level. When the system is powered up the first time, the microcontroller will first of initialize itself by setting all appropriate registers and internal hardware after which it initializes the GSM module and the Temperature sensor. This stages information is indicated on the liquid crystal display. After initialization, the microcontroller will then communicate with DHT11 to get the temperature and humidity data, convert the solar voltage and current read to digital information as well as that from the light intensity level stage. All this information is stringed in a for website HTTP protocol format and the GSM Module is then instructed by the microcontroller using AT commands to connect to MTN provider network and connect to IoT(Thingspeak) webserver. On successful connection sends the data to the webserver. After acknowledgement from the server, it closes the communication link and waits for the next hour interval to repeat the process.

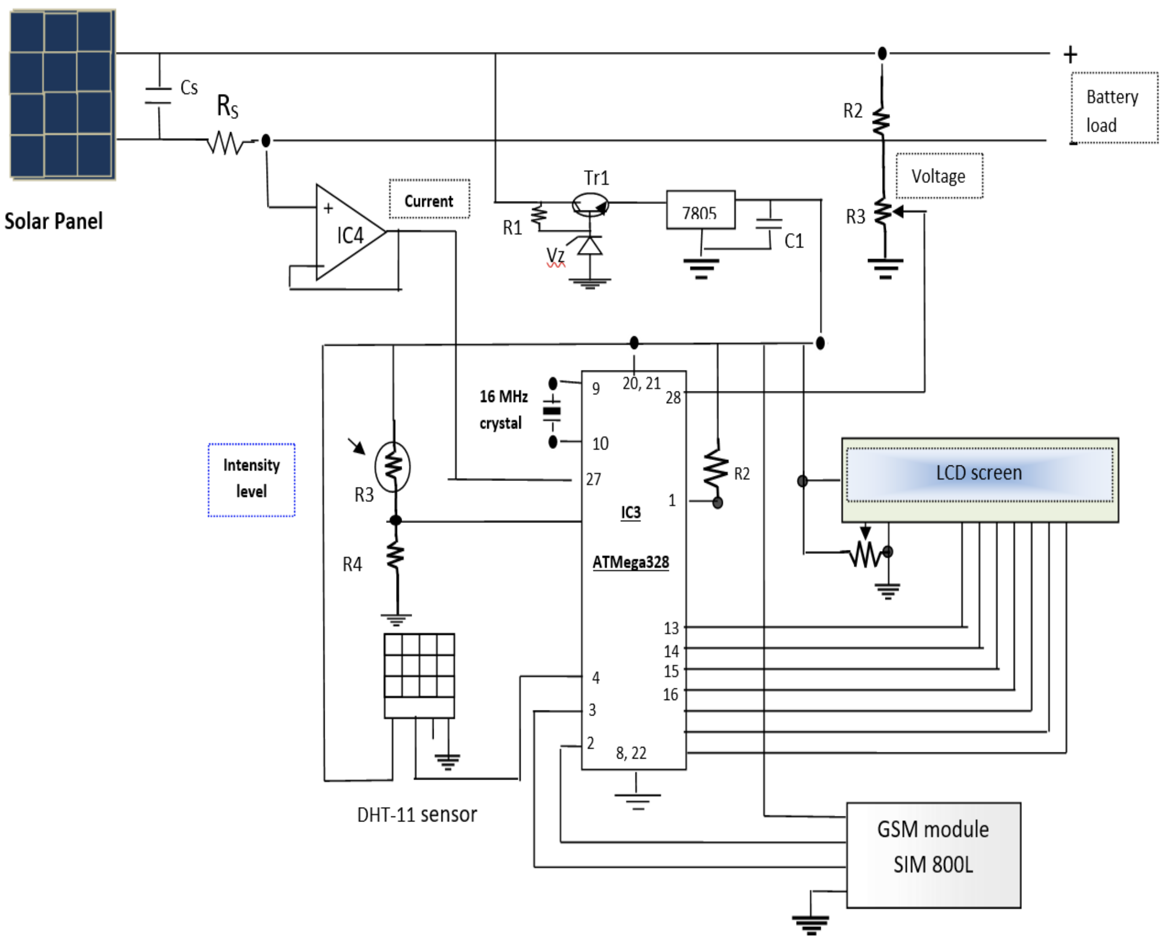


Figure 5: Complete circuit diagram of the IOT based Datalogger.

2.2 Development of Machine Models Solar Power Prediction

In building a machine learning model, a dataset is required to train the model. For this study the data obtained from the IoT-based data logger was used to train the models. The following procedures were followed to build the models.

2.2.1. Data Preprocessing

Data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing was done as a required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. This includes missing data and feature scaling.

2.2.2. Missing Data

In cases of missing data, columns in the dataset which had missing data were replaced with the mean, remaining values in the column. This method prevented the loss of data. Replacing the missing data with the mean is a statistical approach to handle the missing values as shown in Equation (9)

$$mean = \frac{N}{x} \tag{9}$$

Where N is the sum of the remaining values in the column and x is the number of values

2.2.3. Feature Scaling

Feature scaling which is one of the most critical steps during the pre-processing of data. Feature scaling was applied to the dataset, because of the vast difference in the range of values. The feature scaling approach chosen in this work is the Min-max scaler as shown in Equation (10)

$$X_{new} = \frac{X - Xmin}{Xmax - Xmin} \tag{10}$$

Where X_{new} is the new value of X, X_{min} is the minimum value in the column and X_{max} the maximum value in the column.

2.2.4. Data Splitting

Data splitting is commonly used in machine learning to split data into a training and test set, the training set was used to fit the model and test for the model evaluation. This approach allows us to find the model hyper-parameter and also estimate the generalization performance. For proper splitting of dataset in this work, the stratified shuffle split technique in the sklearn library with a test size of 20% was used.

2.2.5. Model Selection

In solar power forecasting, different supervised machine learning algorithms can be used for prediction. In the work, decision regression, support vector regression and random forest regression were used to train the dataset. These models were chosen in for this work because they are commonly used machine learning models. Also, it has been reported that these algorithms have good performance in solar power forecasting.

2.2.6. Decision Tree Regression

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

2.2.7. Support Vector Regression

In support vector regression, the best fit line is the hyperplane that has the maximum number of points. Unlike other Regression models that try to minimize the error between the real and predicted value, the SVR tries to fit the best line within a threshold value shown in equation (11). The threshold value is the distance between the hyperplane and boundary line.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha^*{}_i)\big(\varphi({x_i}^t), \varphi(x_i)\big) + b \qquad (12)$$

2.2.8. Random Forest Regression

Random forest is a Supervised Learning algorithm which uses ensemble learning method for classification and regression. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The predicted value is shown in Equation (13).

$$y_p = \frac{1}{n} \sum_{1}^{n} f_n(\dot{x}) \qquad (13)$$

Where y_p is the predicted values of y, n is number of datapoints and x is the training set.

3. Results and Discussion

3.1 Test Results of IoT Based Datalogger

This project was tested several times for the functionality of the design of the circuit. The power supply and the microcontroller were tested to ensure they are working as expected and other stages were also tested. The final tests carried out and recorded was that for the voltage, current, light flux temperature and humidity. Table 1 highlights the several tests carried out in the course of the work.

3.2 Machine Learning Model Prediction Result

This section presents results predicted by each of the machine learning models. Table 1 shows the values of Y test for, test set, decision tree regression model, support vector regression model and Random Forest regression model.

Table 1 Machine learning model predictions

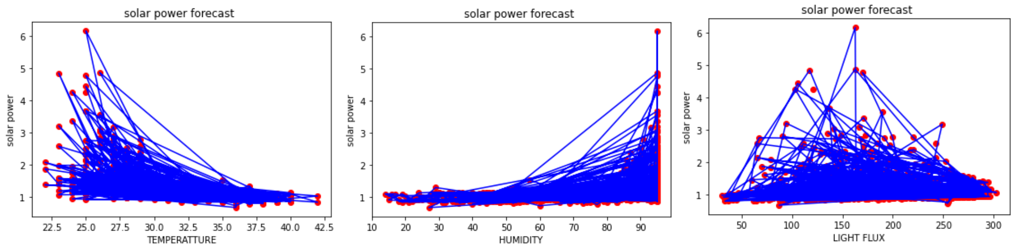
S/N	Y _test	Decision tree	Support vector	Random Forest
1	1.1622	1.1582	1.2244	1.2159
2	1.1622	1.1482	1.1823	1.1900
3	0.9200	0.8892	1.0818	0.9461
4	0.9440	0.1934	1.0483	1.0089
5	1.3489	1.3247	1.0202	1.3109
6	0.8715	0.7200	0.9285	0.8768

7	1.6119	1.4789	1.1425	1.5641
8	2.1307	2.5368	1.1891	2.0579
9	1.3559	1.4562	1.0822	1.2203
10	1.1622	1.8622	1.0586	1.1191

3.3. Visual Representation of Machine Learning Models

Plots For Decision Tree Regression Model

The figure below shows result gotten from decision tree regression model, plot showing; Figure 6(a) (solar power against temperature), Figure 6(b) (solar power against humidity) and Figure 6(c) (solar power against light flux).

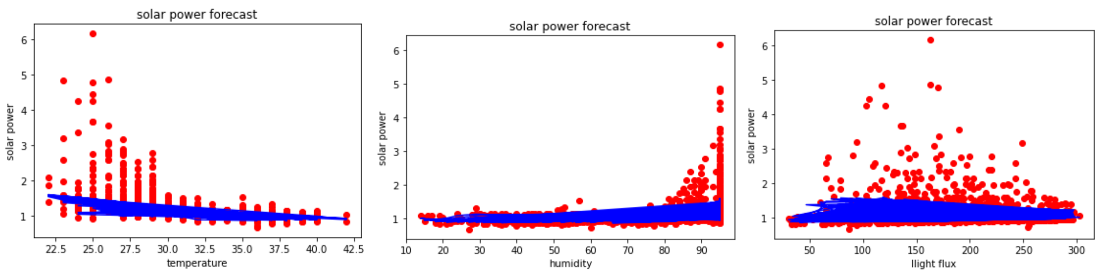


(a) (b) (c)

Figure 6: Plots for decision tree regression model result

3.4. Plots for support vector regression model

The figure below shows result gotten from decision tree regression model, plot showing; Figure 6(a) (solar power against temperature), Figure 6(b) (solar power against humidity) and Figure 6(c) (solar power against light flux).

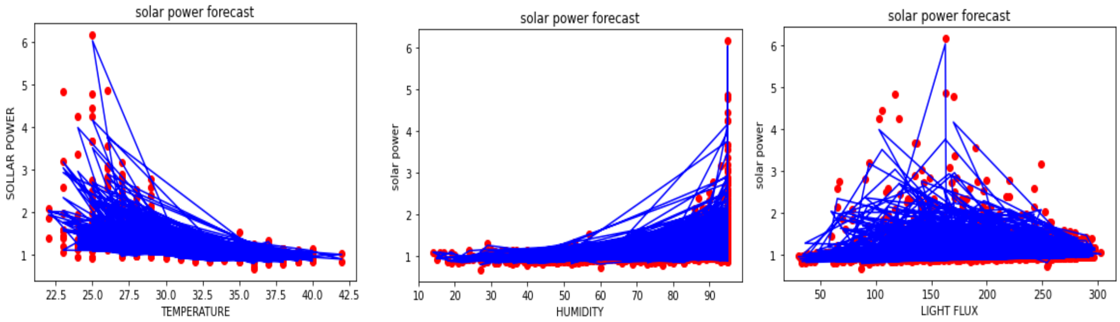


(a) (b) (c)

Figure 7: Plots for support vector regression model

3.5. Plots for random forest regression model

The figure below shows result gotten from decision tree regression model. The plots below shows; Figure 7 (a) (solar power against temperature), Figure 7(b) (solar power against humidity) and Figure 7(c) (solar power against light flux).



(a) (b) (c)

Figure 8: Plots for support vector regression model

3.6. Evaluation Metrics Result for Machine Learning Models

Table 2 shows the results for each machine learning model score calculation. The model score was calculated by addition of the values gotten from root mean square error, mean absolute error, mean square error and R-squared value. Decision tree regression model has a model score of 1.0594, support vector regression has model score of 3.768 and Random Forest regression has a model score of 0.9506.

Table 2 Evaluation Metrics Result for Machine Learning Models

Machine Learning Models	Root Mean Square Error	Mean Absolute Error	Mean Square Error	R-Suared Value	Model Score
Decision Tree Regression	0.0825	0.01014	0.0068	0.9600	1.0594
Support Vector Regression	1.4588	1.3063	2.1281	-1.1300	3.7632
Random Forest Regression	0.1787	0.0808	0.0319	0.8200	0.9506

4. Conclusion

The IoT-based datalogger was designed and implemented, the dataset recorded was used to train three machine learning models. After the models were trained, an evaluation metric was done to calculate for root mean square error, mean absolute error, mean square error and R-squared value. The model with the lowest model score is the random forest regression model with model score of 0.9506. The completion of this project resulted the following advantages: it Provided a means of recording both weather and electrical parameters without being present at the location, a process of collating and representing data recorded online, how machine learning model are trained with dataset recorded with the datalogger and it shows comparison between three machine learning model.

References

[1] C.M Andreas, and G Sarah, (2017). Introduction to Machine Learning with Python, A Guide for Data Scientists, 2nd edition, Published by O'Reilly Media, Inc. 1005 Gravenstein Highway North, Sebastopol, CA 95472.

[2] D.F Keith (2019). A Brief History of Machine Learning, published by Keith D. Foote, viewed 25TH October 2020, < https://www.dataversity.net/a-brief-history-of-machine-learning/# >.

[3] A. Obayuwana and E.C Igbonoba, E.E.C. (2020). Design and Implementation of a Smart Prepaid Solar Inverter System. International Journal of Recent Engineering Science (IJRES) Volume 7 Issue 4, 7-13, Jul-Aug 2020 ISSN: 2349 – 7157 / doi: 10.1445/23497157/IJRES-V7I4P102
[4] I Emil, and K.C Mikael (2018). Solar Power forecasting with machine learning techniques, published by KTH Royal institute of technology school of engineering sciences, KTH SCI SE-100 44 Stockholm, Sweden URL: www.kth.se/sci.
[5] C Hayashi (1998). What is Data Science? Fundamental Concepts and a Heuristic Example. Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1_3. ISBN9784431702085.
[6] W Tang, T Zhang and E Sazonov (2013). The application of machine learning in monitoring physical activity with shoe sensors. Cyber-Physical Systems: Integrated Computing and Engineering Design. Sep26:187.
[7] J Gubbi, R Buyya, S Marusic and M Palaniswami, (2013). Internet of things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7):1645–1660.
[8] S Sudhanshu and D Ravindara (2012). Predicting daily incoming Solar Energy output using weather data Bhatt Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh.

Appendix

Table A1: Data Recorded with IOT Based Datalogger

Created_at	Entry_id	Voltage	Current	Temperature	Humidity	Light flux
2020-11-14 07:42:47	77	18.88	0.06	29	90	154.52
2020-11-14 09:17:36	78	19.85	0.06	30	84	248.53
2020-11-14 10:52:25	79	18.4	0.06	31	83	236.83
2020-11-14 12:27:14	80	19.37	0.06	33	71	292.5
2020-11-14 15:36:51	81	18.88	0.05	37	49	176.71
2020-11-14 17:11:41	82	18.88	0.05	36	59	315.9
2020-11-15 07:25:02	83	19.37	0.06	29	88	150
2020-11-15 08:59:51	84	19.37	0.06	30	83	165.82
2020-11-15 12:09:29	85	18.4	0.05	32	75	248.93
2020-11-15 13:44:18	86	18.88	0.05	34	64	291.54
2020-11-15 15:19:07	87	19.27	0.07	37	53	280.4
2020-11-15 16:53:56	88	17.43	0.05	38	46	129.91
2020-11-16 07:07:17	89	17.91	0.09	28	90	72.22
2020-11-16 08:42:06	90	19.37	0.11	29	86	196.64