Analysis of Solar Radiation based on HI-SEAS Weather Station Meteorological Data

Thilina Jayathilaka[1], Kithmini Seneviratne[2]

Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
Email: [1]thilina.20@cse.mrt.ac.lk, [2]piyum.20@cse.mrt.ac.lk

***Abstract*— NASA HI-SEAS acts as a testbed and training ground for planetary exploration missions. A NASA Space App Challenge Hackathon wanted to use data collected from HI_SEAS site to predict solar radiation given a set of measurable meteorological conditions. In this paper we have used several ML algorithms to create models and predict the radiation values based on the given data.**

***Keywords*— solar radiation, solar irradiation, exploratory data analysis**

I. INTRODUCTION

Sun is by far the most important source of energy for life on Earth and its radiant energy is a convenient renewable source of energy that every country tries to exploit for almost all daily operations. The ***solar radiation*** is technically called as the radiant energy emitted by the sun from a nuclear fusion reaction that creates electromagnetic energy whereas ***solar irradiance*** is the power per unit area (watt per square meter, W/m^2), received from the Sun in the form of electromagnetic radiation as reported in the wavelength range of the measuring instrument [1].

Luckily, with the optimum position of the Earth in the solar system and because of the Earth's rotation and revolution times, every location on Earth receives sunlight at least part of the year. The amount of solar radiation that reaches any one spot on the Earth's surface varies according to many factors and some are as follows [2].

- Geographic location
- Time of day
- Season
- Local landscape
- Local weather.

Knowing the amount of solar radiation in different areas can help to get a general estimate of the maximum achievable amount of solar-electrical energy in that area which will ensure the use of resources by utilizing solar panels and increase the return of investment of the investor.

Our research question is to identify which factors that correlate with solar irradiation so that we can use them to predict solar irradiation in a given different situation.

For the first part of the report, we have described the methodology of our analysis by introducing the selected dataset and the pre-processing done on it. It will also describe the different ways we used to analyse the data. Next chapter will describe what we did to build and train a model to predict the solar irradiation for future data. For the results section, we have included the accuracy of our model and finally, we have described our conclusion with some our own feedback on the dataset and the analysis we implemented.

II. METHODOLOGY

2.1. DATASET AND TECHNOLOGY USED

In this analysis on solar radiation, we have used the meteorological data gathered by *The Hawaii Space Exploration Analog and Simulation (HI-SEAS)* weather station ranging four months from September to December, 2016. This HI-SEAS weather station dataset consists of the following fields.

1. UNIXTime: $seconds\ since\ Jan\ 1, 1970$
2. Date: $yyyy - mm - dd$
3. Time: $local\ time\ of\ day\ in\ hh:mm:ss$
4. Solar radiation: W/m^2
5. Temperature: F^o
6. Humidity: %
7. Barometric pressure: Hg
8. Wind direction: $degress$
9. Wind speed: $miles/hour$
10. Sunrise: $Hawaii\ time$
11. Sunset: $Hawaii\ time$

To do the analysis, we chose Python as the language and both Jupyter Notebook and PyCharm tools were used to implement operations on the dataset.

2.1. Pre-processing the data

As the first step, we pre-processed the data to make sure we can expect a good analysis out of it. First, we check for any *null* or *empty* values in columns and luckily, we found no such irregularities.

Secondly, we observed that the data was not sorted correctly. Although the data rows were sorted in ascending order by the Date, the rows belonging to a single value of Date column were sorted in descending order by the Time column. In order to get the data in correct order, we sorted the data by UNIXTime column.

Thirdly, when we started plotting graphs for the analysis, we observed that values in UNIXTime belong to the GMT zone but the actual Date and Time values belong to Pacific/Honolulu Time Zone (GMT -10:00). In order to fix that, we used a lambda function over the UNIXTime column to correct that time zone shift. Following graphs show how Solar Radiation varies on a single day in 5-minute intervals. Five days which are nearly 30 days apart were selected to draw 5 line charts before and after the pre-processing step on the UNIXTime time zone.

FIGURE I
SOLAR RADIATION VS. TIME OF DAY GRAPH WITH TIME ZONE ERROR

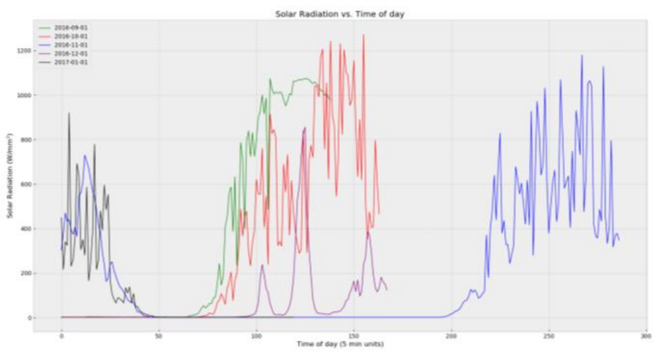
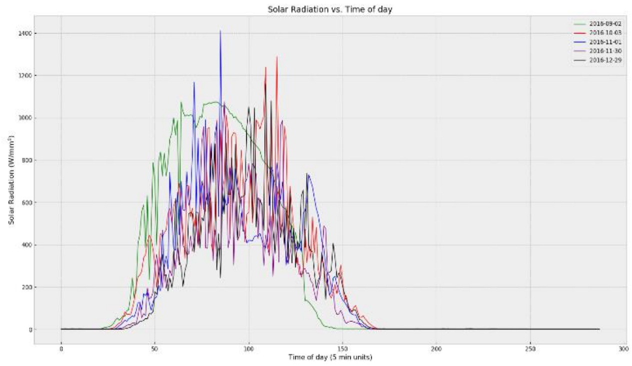


FIGURE II
SOLAR RADIATION VS. TIME OF DAY GRAPH WITH TIME ZONE ERROR FIXED



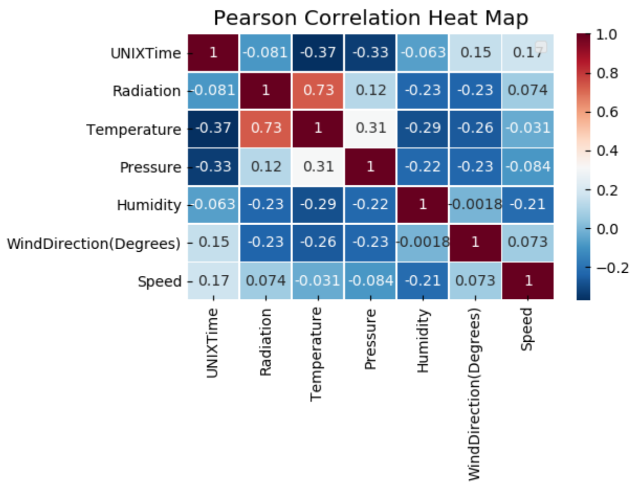
2.1. Correlation among features

The measurements were visualized and P*earson correlation coefficients* [3] were calculated to determine which parameters have the most impact on one another. The relevant equation is as follows.

$$r = \frac{\sum((x - \bar{x})(y - \bar{y}))}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

According to the Pearson correlation heatmap we generated for all numerical valued features, we observed that temperature has the highest positive correlation (+0.73) with solar radiation above all other features. The next closest feature which was positively correlated with radiation was Pressure (+0.12). We also observed that Humidity and Wind Direction has the same negative correlation with radiation (-0.23).

FIGURE III
PEARSON CORRELATION HEAT MAP BETWEEN NUMERICAL VALUED COLUMNS



In addition to this standard correlation measure, we have also checked the correlation between the above important features with two other measures as well. They are *Kendall Rank Correlation Coefficient* and *Spearman's Correlation Coefficient* methods

[4]. The corresponding values are given below in the table 1.

TABLE I
CORRELATION MEASURES OF IDENTIFIED IMPORTANT FEATURES

Feature couple	**Correlation**		
	Pearson	**Kendall Rank**	**Spearman**
Radiation / Temperature	+0.73	+0.54	+0.72
Radiation / Pressure	+0.12	+0.032	+0.046
Radiation / Humidity	-0.23	-0.085	-0.12
Radiation / Wind Direction	-0.23	-0.21	-0.31

We have drawn scatter plots for the above features, and they are shown in the following figures.

FIGURE IV
SCATTER PLOT FOR SOLAR RADIATION VS. TEMPERATURE

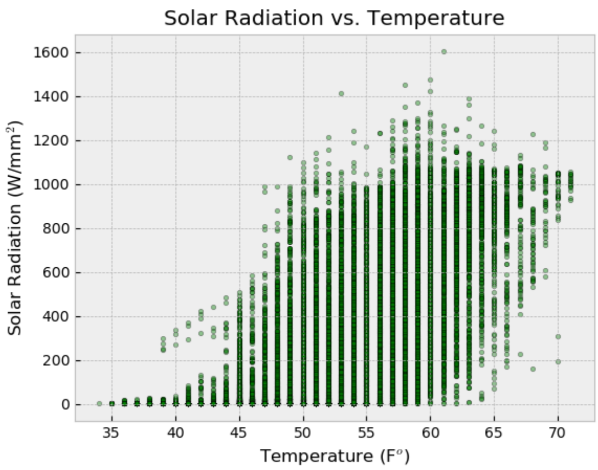


FIGURE V
SCATTER PLOT FOR SOLAR RADIATION VS. WIND DIRECTION

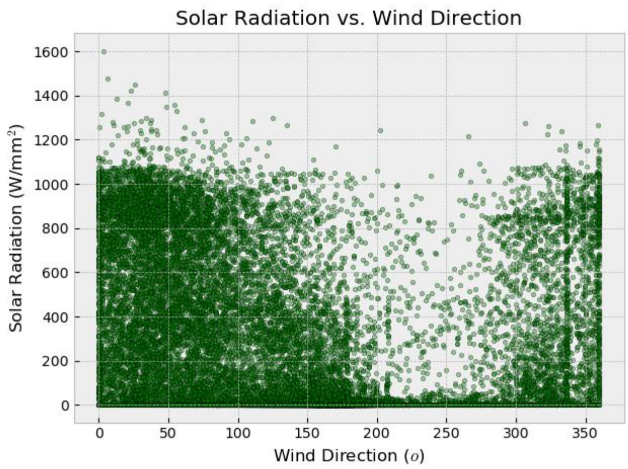


FIGURE VI
SCATTER PLOT FOR SOLAR RADIATION VS. HUMIDITY

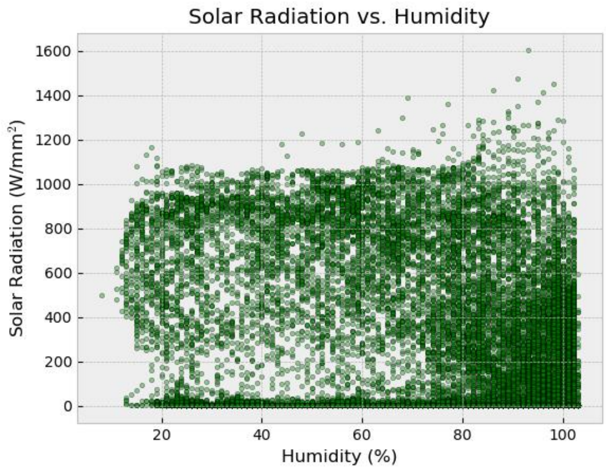
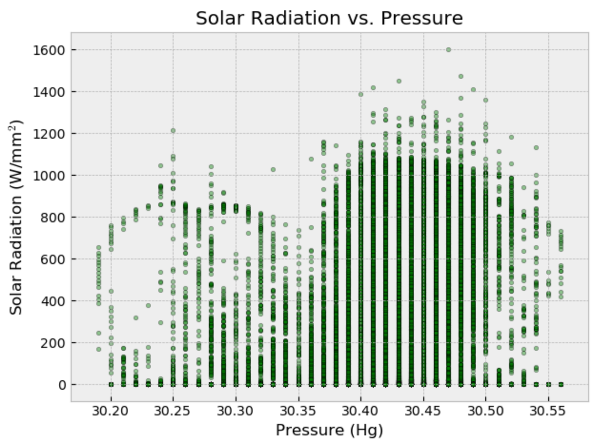


FIGURE VII
SCATTER PLOT FOR SOLAR RADIATION VS. PRESSURE



III. TRAINING AND TESTING

In order to predict the radiation against the given meteorological conditions several machine learning algorithms were implemented.

- Linear Regression
- Random Forest Regression
- Support Vector Machine

Performance of each algorithm with their accuracy is discussed below.

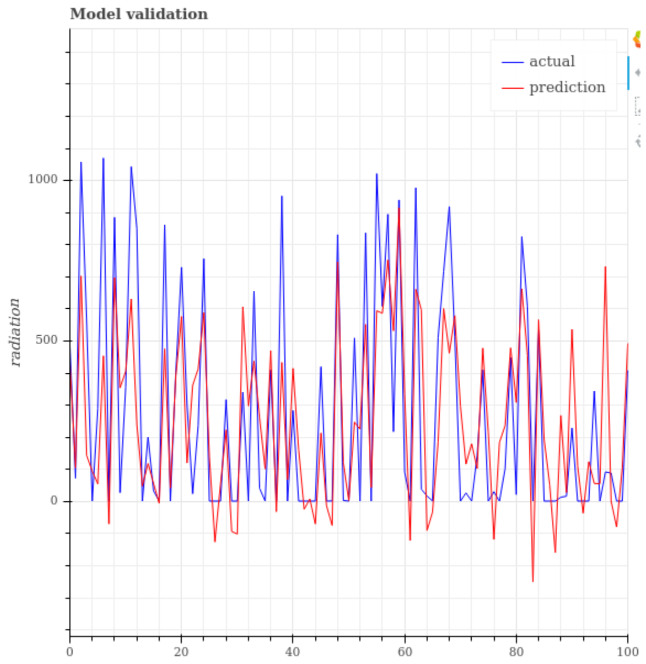
3.1. LINEAR REGRESSION

Linear Regression is an approach to modelling the relationship between a dependent variable and one or more independent variables. Although the characteristics seemed non linear, Linear Regression was implemented resulting in the following observations.

Accuracy : 59.09%

It can be observed that Linear Regression yields moderate accuracy rate.

FIGURE VIII
ACTUAL VS. PREDICTION - LINEAR REGRESSION

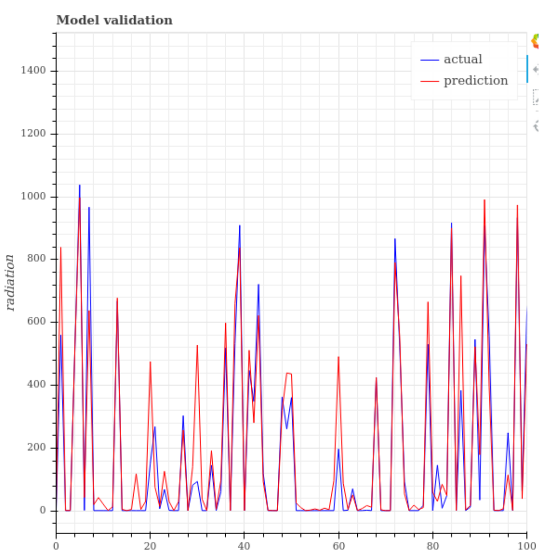


3.2. RANDOM FOREST REGRESSION

Random Forest is an ensemble technique which can be used to perform both classification and regression tasks. This technique utilizes multiple decision trees and a technique called Bootstrap Aggregation to achieve this. A random forest regressor was implemented with the data and following are the results.

Accuracy: 87.28%

FIGURE IX
ACTUAL VS. PREDICTION – RANDOM FOREST REGRESSION



Random Forest Regression yields an acceptable accuracy at 87.28%. This was the highest accuracy obtained out of the tested algorithms.

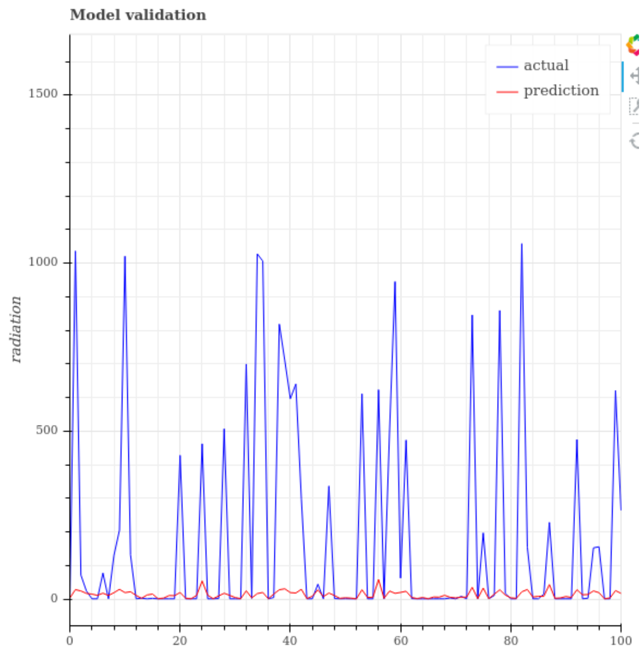
3.3. SUPPORT VECTOR REGRESSION

Support Vector Machines are another kind of learning algorithm which analyse data used for classification and regression analysis. Support Vector Regression implementation with the data produced the following observations.

Accuracy: -41.61%

In SVR, run time for training the algorithm is exceptionally longer than others. This method produced a negative accuracy rate of -41.61% which indicates the trained model produces incorrect predictions.

FIGURE X
ACTUAL VS. PREDICTION – SUPPORT VECTOR REGRESSION



IV. FURTHER IMPROVEMENTS

A more accurate model can be built by focussing on few other areas

A. Considering feature relationships

We know with our knowledge that certain parameters such as temperature, pressure and humidity are not completely independent of each other. If we could include their relationship among them to the model, we would have been able to create a far more realistic model.

B. *Considering more features*

Including data of more relevant features such as cloud cover and precipitation etc. We could create a better model if we have more data regarding the parameters which affects the transmission of light through the atmosphere.

V. Conclusions

With the observations from Random Forest Regressor a prediction of considerable accuracy (~ 87%) could be obtained.

Furthermore, from the initial correlation measurements it could be observed that temperature has the highest correlation with radiation output.

Acknowledgment

We would like to acknowledge Dr Shehan Perera and Dr. Uthayasanker Thayasivam for supervising this research project. Also, we would like to acknowledge all the staff members of the Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka for the support given to carry out this research.

.

References

[1] ResearchGate. 2020. What Is Different Between Solar Irradiation And Solar Radiation?. [online] Available at: <https://www.researchgate.net/post/What_is_different_between_solar_irradiation_and_solar_radiation> [Accessed 9 April 2020].

[2] "Solar Radiation Basics." Energy.Gov, 2013, www.energy.gov/eere/solar/articles/solar-radiation-basics. [Accessed 9 April 2020]

[3] Sedgwick, Philip. (2012). Pearson's correlation coefficient. BMJ. 345. e4483-e4483. 10.1136/bmj.e4483.

[4] Puth, Marie-Therese & Neuhäuser, Markus & Ruxton, Graeme. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. Animal Behaviour. 102. 10.1016/j.anbehav.2015.01.010.